FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

POPH
PROGRAMA OPERACIONAL POTENCIAL HUMANO

QREN
QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL
PORTUGAL 2007.2013

Governo da República Portuguesa

UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional

Sandra Isabel Portas de Sousa **Four essays in Economics of Education** UMinho | 2016

Universidade do Minho
Escola de Economia e Gestão

Sandra Isabel Portas de Sousa

Four essays in Economics of Education

setembro de 2016

Universidade do Minho
Escola de Economia e Gestão

Sandra Isabel Portas de Sousa

Four essays in Economics of Education

Tese de Doutoramento em Economia

Trabalho realizado sob a orientação do
Professor Doutor Miguel Portela
e da
Professora Doutora Carla Sá

setembro de 2016

DECLARAÇÃO

Nome: Sandra Isabel Portas de Sousa

Endereço Eletrónico: sipsousa@hotmail.com

Título da tese: Four essays in Economics of Education

Orientadores:

Professor Doutor Miguel Portela

Professora Doutora Carla Sá

Ano de conclusão: 2016

Designação do doutoramento: Doutoramento em Economia

É AUTORIZADA A REPRODUÇÃO PARCIAL DESTA TESE APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE;

Universidade do Minho, 15 de setembro de 2016

Assinatura: Sandra Isabel Portas de Sousa

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração da presente tese. Confirmo que em todo o trabalho conducente à sua elaboração não recorri à prática de plágio ou a qualquer outra forma de falsificação de resultados.

Mais declaro que tomei conhecimento integral do Código de Conduta Ética da Universidade do Minho.

Universidade do Minho, 15 de setembro de 2016

Nome completo: Sandra Isabel Portas de Sousa

Assinatura: Sandra Isabel Portas de Sousa

"E quanto mais subimos no que desejaríamos saber,
mais descemos no que sabemos."
Fernando Pessoa, *Livro do Desassossego*

Acknowledgements

More than six years have passed since I started my PhD. This learning path started with two years of advanced Economics lectures, followed by more than four years of intensive research. These were split into hundreds of days of incubation in *Quadros de Pessoal* office, frequency of some courses, several meetings/discussions with my supervisors, presentations in conferences, workshops and seminars and, unavoidably, subsequent reviews of the essays. It is now time to acknowledge everyone and institutions, that somehow supported me while I was carrying out this work.

I would like to acknowledge my supervisors, Professor Carla Sá and Professor Miguel Portela, for their guidance, help, instructive comments and suggestions during the thesis process, which made me go further and grow as a researcher. In particular, I gratefully acknowledge Professor Miguel for his patience, endless availability, and valuable comments and suggestions concerning econometric models and specifications. I also thank Professor Miguel for all the financial support provided in conferences and courses. I express my sincere gratefulness to both.

I also thank Portuguese Ministry for Labour, Solidarity and Social Security for granting me access to *Quadros de Pessoal* the dataset used in Chapters 3 and 4. I am grateful to *Direção–Geral de Estatísticas da Educação e Ciência* (DGEEC) for making available the information used in the analysis in Chapters 5 and 6. A special note goes for its Director General, Luísa Castro Loura, and Subdirector General, Teresa Evaristo, for the invaluable role on the release of the data for scientific research and support provided. I also have to acknowledge *Ministério da Educação–Direção–Geral dos Estabelecimentos Escolares* for giving me the grant holder license.

Along this journey, I was getting important comments and suggestions to improve this work and my understanding towards research. I want to thank all participants in conferences and workshops who contributed with their opinions and suggestions. A special "thank you" goes to Professor João Cerejeira who contributed with valuable comments on the paper's early versions underlying Chapters 3, 4 and 5, as well to Professor Priscila Ferreira who contributed with important suggestions and comments to Chapter 4.

I have also made some good friends during my PhD, who directly or indirectly helped me along this journey and with whom I have spent very good moments. To all my classmates, Manuel, João, Gilson, Andreia, Susana, Sónia, Luís and Helder, I am grateful for the friendly and mutual support environment that existed, as well as the fun lunches and dinners in *Coimbra*. For all the nice conversations, lunches and coffee breaks, in *Universidade do Minho*, which often contributed for recharging my batteries, I would to thank Gilson, Andreia and Susana. I am also grateful to Maria José of the *Gabinete de apoio à investigação* for the availability and support.

A word of thanks to my co–workers, who are also my friends, who have always supported and encouraged me. In particular, I am grateful to Marta for her availability and help in text corrections. I thank Patrícia to hear my frustrations at the beginning of the thesis process.

I owe a special "thank you" to Henrique, who several times lend me his friend shoulder, where I often dumped my concerns and frustrations, and helped me through some difficult moments and stressful periods. Thank you for encouraging me to pursue my goal.

Finally, but certainly not least, I am gratefully acknowledge my parents, my sister and my nephews, Helena e Guilherme, for their patience and support.

This doctoral thesis was supported by the Foundation for Science and Technology (FCT) through Individual PhD Scholarship with reference: SFRH/BD/85985/2012, whom I am grateful.









Four essays in Economics of Education

Abstract

This thesis comprises four essays in economics of education. The first two essays are focused on the returns of accumulated human capital, in which the *Quadros de Pessoal* dataset is used.

The first essay examines the evolution of returns to education for Portuguese workers during the period 1986–2009. In order to address the possible estimate bias due to the endogeneity, the education is instrumented by the quarter of birth and by the changes in the Portuguese compulsory schooling laws. In addition, a new instrumental variable is proposed: average education by region in the year in which individual entered school.

Also in the context of the returns to education, in the second essay, the existence and size of human capital spillovers at different levels of analysis (firm, regional and inter-regional) is investigated, allowing to understand how the interactions among Portuguese workers can be determinant of differences in individual wages. Furthermore, the aggregated human capital is measured in three different ways: average education, share of high qualified workers and an adjusted multi–dimensional skill index. To control to the imperfect substitutability across workers, two qualification groups of workers are considered. Using a matched employer–employee dataset for the period 2002–2009 complemented by variables at county level from other sources, augmented Mincerian wage equations are estimated using different estimation methodologies.

The third and fourth essays focus on the human capital accumulation process and the MISI and JNE Statistics datasets are used. The third essay analyses the determinants that influence the students' achievement of secondary education and, in addition, estimates the school value–added. It is given greater emphasis to the analysis of the class size effect on students' achievement. For these purposes, a multilevel variance component model, with 3 levels: student, class and school, in a value added perspective, is applied. Additionally, the results obtained from the methodologies that are commonly used in academic work on performance assessment schools, Data Envelopment Analysis (DEA) and multilevel modelling, are compared.

Finally, the fourth essay examines which observable teachers' characteristics, who teach Mathematics and Portuguese in the secondary education, influence the achievement gains of their students, taking into account a whole set of other factors that influence its progression, such as student's characteristics and class size. A value–added approach that adjusts for teacher fixed–effects is applied.

Quatro ensaios em Economia da Educação

Resumo

Esta tese inclui quatro ensaios em Economia da Educação. Os dois primeiros ensaios centram–se na rendibilidade do capital humano agregado, nos quais é usada a base de dados "Quadros de Pessoal".

O primeiro ensaio analisa a evolução da rendibilidade da educação para os trabalhadores portugueses durante o período de 1986–2009. Com a finalidade de solucionar o possível enviesamento das estimativas devido à endogeneidade, as variáveis trimestre de nascimento e alterações na legislação portuguesa sobre a escolaridade obrigatória são utilizadas como instrumentos para a educação. Adicionalmente, é proposta uma nova variável instrumental: educação média por região no ano em que o individuo entra na escola.

Ainda no contexto da rendibilidade da educação, no segundo ensaio avalia–se a existência e a dimensão das externalidades do capital humano, em diferentes níveis de análise (empresa, regional e interregional), permitindo compreender de que forma a interação entre os trabalhadores portugueses podem ser determinantes nas diferenças salariais entre os trabalhadores. O capital humano agregado é medido de três formas diferentes: escolaridade média, percentagem de trabalhadores qualificados e um índice multidimensional de habilidade (*skill index*). Para controlar para a substituibilidade imperfeita entre os trabalhadores são considerados dois grupos de acordo com as suas qualificações. Usando uma base de dados empregador–empregado para o período 2002–2009 complementada com variáveis ao nível do concelho provenientes de outras fontes, é estimada uma equação salarial Minceriana à qual é acrescentado o termo do capital humano agregado, usando diferentes metodologias.

O terceiro e quarto ensaios centram–se no processo de acumulação de capital humano, nos quais são usados dados da MISI e do JNE. O terceiro ensaio analisa os fatores que influenciam o progresso dos alunos do ensino secundário. Para além disso, estima o valor acrescentado da escola. É dada maior enfase à análise do efeito da dimensão da turma no desempenho do aluno. Para tal, é utilizado um modelo multinível de componentes da variância com 3 níveis: aluno, turma e escola, numa perspetiva de valor acrescentado. Adicionalmente, são comparados os resultados obtidos a partir das metodologias geralmente usadas na investigação académica sobre a avalição do desempenho das escolas, DEA e modelo multinível.

Finalmente, o quarto ensaio avalia as características observáveis dos professores, de Matemática e Português do ensino secundário, que influenciam o desempenho dos seus alunos, tendo em conta um conjunto de outros fatores que influenciam o seu desem-

penho, tais como, as características do aluno e dimensão da turma. É usado o modelo em painel de efeitos–fixos para o professor, numa abordagem de valor acrescentado.

Contents

List of Figures **xvii**

List of Tables **xix**

1 Introduction **1**

2 A brief description of the Portuguese education system **7**

3 Characterization of returns to education in Portugal: 1986–2009 **11**

3.1 Introduction 11

3.2 Literature review 13

3.3 Theoretical framework 19

3.3.1 Mincerian model 19

3.3.2 Econometric framework 21

3.4 Methodology and data 30

3.4.1 Methodology 30

3.4.2 Data description 31

3.5 Empirical results 35

3.5.1 Ordinary least squares results 35

3.5.2 Firm specific effects 44

3.5.3 Instrumental variables results 47

3.5.4 Firm specific effects and instrumental variables 64

3.6 Concluding remarks 66

4 Educational spillovers in Portugal **69**

4.1 Introduction 69

4.2 Literature review 71

4.3 Theoretical framework 76

4.3.1 Perfect substitutability across workers 76

4.3.2 Imperfect substitutability across workers 79

4.4 Methodology and data . . . 82
4.4.1 Methodology . . . 82
4.4.2 Data description . . . 88
4.5 Empirical results . . . 94
4.5.1 Human capital externalities at county level . . . 94
4.5.2 Human capital externalities at regional and firm level . . . 108
4.5.3 Inter–regional human capital externalities . . . 113
4.6 Concluding remarks . . . 118

5 Factors that influence student achievement gains and performance assessment of the Portuguese public schools 121
5.1 Introduction . . . 121
5.2 The state of the art . . . 125
5.2.1 School assessment and the determinants of student achievement 125
5.2.2 The Portuguese setting . . . 132
5.3 Empirical approach . . . 138
5.3.1 Models . . . 138
5.3.2 Empirical specification . . . 150
5.3.3 Data and sample . . . 153
5.4 Empirical results . . . 157
5.4.1 Factors that influence student achievement gains: a multilevel approach . . . 158
5.4.2 Relative performance of schools: multilevel versus DEA approaches 170
5.5 Concluding remarks . . . 176

6 Teacher characteristics and student progress 179
6.1 Introduction . . . 179
6.2 Literature review . . . 180
6.3 Methodology and data . . . 183
6.3.1 Methodology . . . 183
6.3.2 Data description . . . 185
6.4 Empirical results . . . 187
6.5 Concluding remarks . . . 195

7 Conclusions 197

References 203

A Appendices to Chapter 3 **225**
A.1 Checks on the consistency of data and constraints imposed 225
A.2 Description of variables included in the empirical models and additional tables of descriptive statistics 227
A.3 Additional tables of OLS estimations 229
A.4 Additional tables of IV estimations 237

B Appendices to Chapter 4 **277**
B.1 Human Capital Index 277
B.1.1 Individual Skill Index 277
B.1.2 Average of the skill index of individuals within a county *vs.* the skill index of a county 278
B.2 Fixed–effects estimation: collinearity between "experience" and time dummies within worker's spell 279
B.3 Description of variables included in the empirical models 284
B.4 Additional tables of results 287

C Appendices to Chapter 5 **293**
C.1 Description of variables included in the empirical models 293
C.2 Additional tables of estimations 295
C.3 DEA efficiency scores 304

D Appendices to Chapter 6 **311**
D.1 Description of variables included in the empirical models 311

List of Figures

2.1 Organization of Portuguese Educational System 9

3.1 Average Schooling Attainment of Portuguese Workers, 1986–2009 . . . 34
3.2 Estimated Returns to Education by OLS, 1986–2009 36
3.3 Returns to different schooling levels by OLS, 1986–2009 38
3.4 Returns to education by district/region 43
3.5 Average Education by season of birth, Men born 1930–1949 48
3.6 Average Education by season of birth, Men born 1950–1966 49
3.7 Average Education by season of birth, Men born 1967–1980 49
3.8 Average Wages by Season of Birth, Men born 1930–1949 50
3.9 Average Wages by Season of Birth, Men born 1950–1966 51
3.10 Average Wages by Season of Birth, Men born 1967–1980 51
3.11 OLS and IV estimates of Return to Education, Men born 1930–1949 . . 54
3.12 OLS and IV estimates of Return to Education, Men born 1950–1966 . . 54
3.13 OLS and IV estimates of Return to Education, Men born 1967–1980 . . 55
3.14 Average Education, Men born 1957–1977 57
3.15 Average Education, Men born 1976–1984 58
3.16 OLS and IV estimates of Return to Education, Men born 1957–1977 . . 60
3.17 OLS and IV estimates using Average Schooling by Region and in Portugal, Men aged 18–60 64

4.1 Average schooling and average wages, by county, in 2002 and 2009 . . . 92
4.2 Educational attainment by county, as percentage of most qualified workers in 2002 and 2009 93
4.3 County skill index, in 2002 and 2009 94

5.1 Hierarchical data structure with three levels 140
5.2 Graphic illustration of the DEA projection 146
5.3 Concepts of input and output–oriented model and returns to scale . . . 148
5.4 Proportion of students by grade on national exams of Mathematics A and Portuguese 156

5.5 Mean national exams scores by discipline and by gender 157
5.6 District effects and approximate 95% confidence intervals versus ranking of districts in Mathematics and Portuguese achievement gains (district identifiers are shown on the top of the error bar) 169
5.7 School effects and approximate 95% confidence intervals versus ranking of 50 best schools in Mathematics and Portuguese achievement gains (school identifiers are shown on the top of the error bar) 171
5.8 School effects and approximate 95% confidence intervals versus ranking of 50 worst schools in Mathematics and Portuguese achievement gains (school identifiers are shown on the top of the error bar) 172

6.1 Experience in a 2^{nd} degree polynomial model, 2010–2012 190
6.2 Experience in a 4^{th} degree polynomial model, 2010–2012 191

A.1 Firm fixed-effects regressions with interactions education×year 234
A.2 Firm fixed-effects regressions with interactions education×district . . . 235
A.3 Firm fixed-effects regressions with interactions education×industry . . 236

C.1 Fitted values for the class size (in Mathematics) with a 2^{nd} degree polynomial model, in 2010, 2011 and 2012 303

List of Tables

2.1 Summary of the Portuguese Compulsory Schooling Laws 8

3.1 Descriptive statistics on workers 33
3.2 Descriptive statistics on firms 35
3.3 Statistics on sector of activity 35
3.4 Returns to education – OLS 40
3.5 First stage of IV2SLS estimation using CSL6 as instrument, Men born 1957–1977 59
3.6 First stage of IV2SLS estimation using CSL9 as instrument, Men born 1976–1984 59
3.7 OLS and IV estimates using CSL9 as instrument, Men born 1976 – 1984 61
3.8 Firm specific effects and instrumental variables 65

4.1 Descriptive statistics 90
4.2.1 Pooled OLS for 3 groups of workers: all, less qualified and most qualified, using average education of the county as a measure of human capital . 96
4.2.2 Pooled OLS for 3 groups of workers: all, less qualified and most qualified, using the share of highly qualified workers in a county as a measure of human capital 97
4.2.3 Pooled OLS for 3 groups of workers: all, less qualified and most qualified, using the skill index of county as a measure of human capital 98
4.3 First–stage regression of local human capital variable 104
4.4 IV results for 3 groups: all workers, less qualified and most qualified workers 106
4.5 Spillovers at regional and firm level 110
4.6 Regional and inter–regional externalities 116

5.1 Descriptive statistics on students 155
5.2 Descriptive statistics on classes and schools 156
5.3 Intraclass correlation 157
5.4.1 Multilevel model results (Year of examination: 2010) 160

5.4.2 Multilevel model results (Year of examination: 2011) 162
5.4.3 Multilevel model results (Year of examination: 2012) 164
5.5 Inputs and outputs specified in the DEA model 173
5.6 Descriptive statistics of aggregated data on public secondary schools . . 174
5.7 Schools with 100% efficiency score in 2010 and their ranking in subsequent years 175

6.1 Descriptive statistics 186
6.2 OLS regressions 188
6.3 Fixed–effects estimates of teacher characteristics 193

A.1 Description of the variables 227
A.2 Descriptive statistics on workers in 1986 and 2009 228
A.3 Descriptive statistics on firms in 1986 and 2009 228
A.4 Standard Mincer Equation (OLS) – All workers 229
A.5 Standard Mincer Equation (OLS) – Male workers 230
A.6 Standard Mincer Equation (OLS) – Female workers 231
A.7 Firm fixed–effects regressions 232
A.8 First Stage IV2SLS estimation using the 1^{st} quarter of birth (q_1) and the 3^{rd} quarter of birth (q_3) as instruments for education 237
A.9 OLS and IV estimates using q_3 and $q_3 \times year-of-birth$ as instruments, Men born 1930–1949 240
A.10 OLS and IV estimates using q_3 and $q_3 \times year-of-birth$ as instruments, Men born 1950–1966 245
A.11 OLS and IV estimates using q_3 and $q_3 \times year-of-birth$ as instruments, Men born 1967–1980 251
A.12 IV estimates using q_1 and $q_1 \times year-of-birth$ as instruments, Men born 1930–1949 256
A.13 IV estimates using q_1 and $q_1 \times year-of-birth$ as instruments, Men born 1950–1966 259
A.14 IV estimates using q_1 and $q_1 \times year-of-birth$ as instruments, Men born 1967–1980 263
A.15 OLS and IV estimates using CSL6 as instrument 266
A.16 First Stage of IV2SLS – Estimation using Average Education in Portugal and Average Education by Region 269
A.17 OLS and IV estimates using Average Education Country and Average Education by Region as instruments 270

B.1 Description of variables 284

B.2 Descriptive statistics in 2009 . . . 286
B.3 Additional results of estimation for pooled OLS . . . 287
B.4 Fixed–effects estimation . . . 288
B.5 Additional results of fixed–effects: Individuals who did not changed the educational level . . . 289
B.6 IV–FE results: All workers, less qualified and most qualified workers . . 290
B.7 Fixed–effects estimation including aggregated human capital at firm level 291
B.8 Fixed–effects estimation including aggregated human capital at inter–regional level . . . 292

C.1 Description of variables . . . 293
C.2.1 OLS regressions (Year of examination: 2010) . . . 295
C.2.2 OLS regressions (Year of examination: 2011) . . . 296
C.2.3 OLS regressions (Year of examination: 2012) . . . 297
C.3.1 Random effects estimations (Year of examination: 2010) . . . 298
C.3.2 Random effects estimations (Year of examination: 2011) . . . 300
C.3.3 Random effects estimations (Year of examination: 2012) . . . 302
C.4 Efficiency scores of public schools: DEA results . . . 304

D.1 Description of variables . . . 311

Chapter 1

Introduction

Human capital is the set of knowledge, skills, competencies and abilities of the individuals, acquired through education, training and experience (Schultz, 1961). According to the human capital theory, increases in education lead to increases in productivity of individuals by increasing their skills (Mincer, 1974). But the returns of an investment in education go beyond the individual sphere. Any investment in education has obvious benefits for the individual himself, but other individuals may profit as well due to spillover effects (Moretti, 2004a,c), and the whole society may gain by having more educated people.

Education is one of the most significant human capital investments. Investments in education are made when its benefits outweigh the costs. The economic benefits of education, either individual or social, have largely been explored. There is currently widespread consensus that education brings a wide-ranging benefits to the individual, as well as to the society as a whole. Researchers and governments have reported gains for the individuals that include positive effects on wages, occupation, unemployment, health. Results reported for the society include benefits on health, civic engagement, subjective well-being, economic growth, and other social variables.

Social benefits of education have justified huge investments by the governments of various countries (Hanushek, 2005b). Studies have shown that investments in human capital are essential for sustaining economic growth over time, as any change in the average level of schooling, at the country level, has a significant positive effect on income growth (Krueger and Lindahl, 2001).

During the second half of the last century, the question of educational expansion attracted much attention and the public debate has developed mostly around the evaluation of its socio-economic benefits and costs. The discussion was dominated by quantity, leaving quality concerns out of the scope of the debate. Nevertheless, the impact of the expansion of school attendance on better economic conditions has not been

homogeneous across countries, which contributed to a change in focus from quantity to quality of education. The discussion has been centered on the idea that spending on education *per se* is no guarantee of better quality of education, and that cognitive skills rather than mere school attainment may be of major importance in fostering the economic well–being. Discussions on educational quality bring policy issues to the debate.

Educational quality have been the focus of recent studies. Hanushek (2012) conclude that there is a positive and significant effect of cognitive skills, measured by students' performance, on individual earnings, on the distribution of income, and on economic growth. Individuals may acquire cognitive skills not only in schools through formal schooling, but also family and peers are a source of development of such skills. The important role of schools has been acknowledged (Hanushek, 2012). Several studies have looked at the impact of school resources, and it seems not obvious that putting more resources in schools will result in better student attainment. There is, however, no consensus among researchers, suggesting that more research is needed.

Portugal provides an interesting context to learn more about those education–related issues. The performance of Portuguese students in international tests (e.g. PISA) is slightly below the average of the OECD countries, despite some improvements that have accompanied large investments in education of the Portuguese government and families. Notwithstanding some progress, the educational level of the Portuguese labour force is still low when compared to other Organisation for Economic Co–operation and Development (OCDE) countries. Both quality and quantity issues are still identified and discussed, suggesting that there is a need for additional studies.

This discussion on private and social returns to education, as well as educational quality and quantity issues has motivated the choice of topics under analysis in this dissertation. It consists of four independent, self-contained contributions in the field of Economics of Education. The first two papers focus on the returns to human capital accumulation, whereas the other two essays focus on the determinants of the cognitive skills accumulation process.

The four empirical in this dissertation make use of Portuguese data. For a better understanding of some country specificities, as well as to provide some explanations of some of the empirical strategies and options made in the papers, the manuscript starts with a chapter on the Portuguese context and setting (Chapter 2). It is followed by four chapters, each one corresponding to a different essay.

The first essay on the"Characterization of returns to education in Portugal: 1986–2009" is presented in Chapter 3. It aims at analysing the evolution of returns to education between 1986 and 2009 for Portuguese men and women. Namely, its main

goal is to estimate the effect of the education on wages of Portuguese labour force, accounting for the endogeneity of schooling. In order to address the possible bias due to the endogeneity, individual education is instrumented by the quarter of birth, similar to the approach used by Angrist and Krueger (1991), and by the changes in the Portuguese compulsory schooling laws, similar to the approach used by Harmon and Walker (1995). In addition, the average education in the region in the year in which the individual entered school is also proposed as an instrument, which, to our knowledge, is used for the first time.

In line with the idea that returns to education go beyond the individual level, Chapter 4 looks at "Educational spillovers in Portugal", for the period 2002–2009. Both existence and size of human capital spillovers at different levels (region and firm) of analysis is investigated. This essay is an additional contribution to the literature on how interactions among Portuguese workers and regions may impact on individual wages, which is scarce for the Portuguese context. It differs from the existing empirical literature, by measuring external returns to human capital in three different perspectives: at the regional level (Rauch, 1993; Moretti, 2004a); at the firm and regional level (Bauer and Vorell, 2010) and, finally, at the regional and inter–regional level, considering the possibility of spatial externalities between regions (Rodríguez-Pose and Tselios, 2012). The aggregate human capital is proxied by three alternative variables: the average education, the share of high qualified workers, and an adjusted multi–dimensional skill index (Portela, 2001). Two groups of workers based on their qualifications are considered, and taken as imperfect substitutes. Augmented Mincerian wage equations are estimated using a matched employer–employee data set for Portugal. To address some methodological issues, different estimation methods and specification are applied. The IV method is applied and regional aggregate human capital is instrumented by means of the variable "presence of a Polytechnic Institute in the region", following Moretti (2004a), as well as the share of people who can read in the county in 1900. To our knowledge, this is the first empirical work that analyses for Portugal the role of imperfect substitutability between workers with different educational levels when human spillovers at the regional and inter–regional level are present, and makes use of these instruments.

These two studies use the "*Quadros de Pessoal*" database collected by the Ministry of Labour, Solidarity and Social Security. These are linked employer–employee data, and the information for the period 1986 to 2009 is used. This is an annual database obtained by mandatory surveys applied to all establishments with at least one employee. Public administration workers, armed forces, self–employed, as well as household workers, are not included in this database.

Differences on wages across individuals may be formed even before they enter the labour market, i.e. along the human capital accumulation process. Recently, public discussion on education acquisitions has moved from quantity issues to quality issues. The "Factors that influence student achievement gains and performance assessment of the Portuguese public schools" are the focus of Chapter 5. Its goal is twofold. First, it aims at analysing the determinants that influence the students' achievement and, in addition, to estimate the school value–added. There is some consensus in the literature on the positive relationship between students' background and their performance, but results regarding the role of the school, namely school resources, are not so consolidated. In this essay, it is emphasized the effect of the class size on students' achievement of secondary education. For this purpose, a multilevel variance component model, with three levels – student, class and school –, in a value added perspective, is applied.

Second, it is intended to shed some light on school rankings based on student performance. The problems of school rankings based on the average exam classifications have been largely discussed. A metric benchmarking that takes into account not only measures of output, such as national exam scores, but also input indicators, such students' prior achievement, socio–economic characteristics and school resources would be of major importance to assess the schools' performance (Sarrico et al., 2014). The methodologies that are commonly used in research work on performance assessment schools are Data Envelopment Analysis (DEA), a non–parametric approach, and multilevel modelling, a parametric approach, which results are to be compared.

This study improves the existing literature on several ways. To our knowledge, no other study does a similar type of analysis in the Portuguese context, as well as the international literature is mostly based on early grades data. Furthermore, Portuguese literature are only used models with 2 levels.

When discussion education quality issues, the role of the quality of teachers as a key determinant in student learning cannot be neglected. There is no other attribute of schools that comes close to having the same influence on student achievement. Moreover, parents, teachers, and policy-makers emphasize repeatedly the central role that teachers play in determining school quality. There is, however, little consensus in literature on the characteristics of a good teacher should have (Hanushek, 2011). Furthermore, in Portugal, characteristics such as teachers' education and experience are used in hiring and salary decisions, but there is no evidence that these characteristics are crucial for teacher quality.

These arguments set the stage for Chapter 6 on "Teacher characteristics and student progress". It examines and identify which observable characteristics of Portuguese teachers of secondary education, such as gender, experience, education level and the fact

that teachers are displaced from their residence area to work, influence the achievement gains of their students, taking into account a whole set of other factors that influence its progression, namely, student background and class size. A value–added approach that adjusts for teacher fixed–effects is applied. This type of analysis is essential to assess the effectiveness of any political or educational action directed at school and their teachers.

The data recently made available by the Ministry of Education made a key contribution for advances in research on Portugal's education system. Seldom data with such quality and richness of information is available worldwide. Two datasets for the period 2010–2012 were used for the latter two chapters, namely, MISI (*Sistema de Informação do Ministério da Educação*), managed by *Direção de Estatísticas da Educação e Ciência* (DGEEC) and Statistics published by *Júri Nacional de Exames – Direção Geral de Educação* (JNE). The data are obtained from administrative records. The first dataset provides information at the student and teacher level, with information on classes and schools, while the second one contains data at the student–level on the scores obtained in the national exams in both basic and secondary education. These two databases were combined in order to implement the analysis carried out in Chapter 6. These datasets were previously anonymized to secure private information on individuals, classes and schools.

Finally, the dissertation ends with a "Conclusion", in Chapter 7, which summarizes the main results of the research presented in the four previous chapters and discusses their main policy implications. It also addresses the main limitations of these studies, as well as points out some directions for future research.

Chapter 2

A brief description of the Portuguese education system

The Portuguese education system as it is known nowadays has been mostly designed during the last quarter of the twentieth century. Before, it had evolved in a quite erratic fashion, which cannot be dissociated from the very low level of education of the Portuguese population. With an average level of schooling of four years, Portugal arrived to the last quarter of the century among the OECD countries with the lowest level of education. Several facts and policy decisions have concurred to explain such situation. In the early years of the twentieth century there was a decrease in the compulsory schooling duration, which went from 5 years, in 1919, to 3 years, in 1930, unlike the situation in other European countries (Alves et al., 2010). In the 1950s, the need of a skilled and diversified workforce to respond to the requirements of the technical progress, which emerged after the World War II, was recognized and reflected in the education policy. Namely, the *Estado Novo* have extended compulsory schooling from 3 to 4 years although gender differences applied: in 1956, the Decree–Law no. 40936 determined that boys would stay for four years in school, and in 1960, with the Decree–Law no. 42994, this requirement was extended to girls. Starting then, boys and girls would stay in school from the age of 7 till 12 years old that had to be completed by the 31^{st} of December of the enrolment year.

Nevertheless, the problem of illiteracy subsisted, and, in the 1960s, the debate of education backwardness of the country is resumed. In 1964, the Decree–Law no. 45810, declares the extension of compulsory schooling to 6 years for both boys and girls that would enrol for the first time or for those repeating the 1^{st} grade in 1964/1965; the age period of compulsory schooling would be from 7 to 14 years old, completed until the 31^{st} of December. Only in 1968, with Law no. 48546 appears a mandatory complementary cycle (5^{th} and 6^{th} grade), which ensured an actual compulsory schooling of 6 years.

Since 1973, the Law no. 5/73 indicates that the age of school entry is 6 years old, which must be completed until 30^{th} September (Law no. 538/79). After the 1974 revolution there were significant transformations taking place in the education sector. The Law no. 538/79 indicates that for individuals born after 01/01/1967 the 6^{th} grade diploma is required. In 1986, the Law no. 46/86 sets the 9^{th} grade as compulsory schooling, starting for those students enrolled in the first year of the basic education in the school year 1987/1988. These students would have to stay in school from the age of 6 (completed before the 30^{th} September) to the age of 15.

Despite the significant improvement of the Portuguese education level in the last two decades, Portugal still ranks quite low when compared with the other OECD countries.[1] In 2009, the compulsory schooling has been raised to 12 years of education, through the Law no. 85/2009. As a result, currently, the school age is between 6 and 18 years old. According to the Law no. 12643/99, the compulsory schooling starting age depends on the individual date of birth, which main changes are summarized in the following table.

Table 2.1: Summary of the Portuguese Compulsory Schooling Laws

Law	**Years of compulsory school**	**Year of birth of the children covered by the change**	**Year of entering school**	**Year of entry into the labour market**
Law no. 42994/60	4	Born until 1966 (may start school if aged 7 years by 31^{st} December)		
Law no. 45810/64	6	Born between 1967 and 1980 (may start school if aged 7 years by 31^{st} December if was born until 1973, or 6 if was born after 1973)	1973/74	From 1981
Law no. 46/86	9	Born between 1981 and 1996 (may start school if aged 6 years by 30^{th} September)	1987/88	From 1996
Law no. 85/2009	12	Born in 1997 and from then on	2003/2004	From 2015

Source: Created by the author.

Nowadays, the Portuguese education system is organized in three levels: basic education, secondary education and tertiary education, as represented in the figure 2.1.

[1] In 2010, the proportion of 25–64 year–old individuals having attained at least upper secondary education was, on average, 74% for OECD countries, whereas for Portugal, the corresponding value was 32% (OECD, 2012).

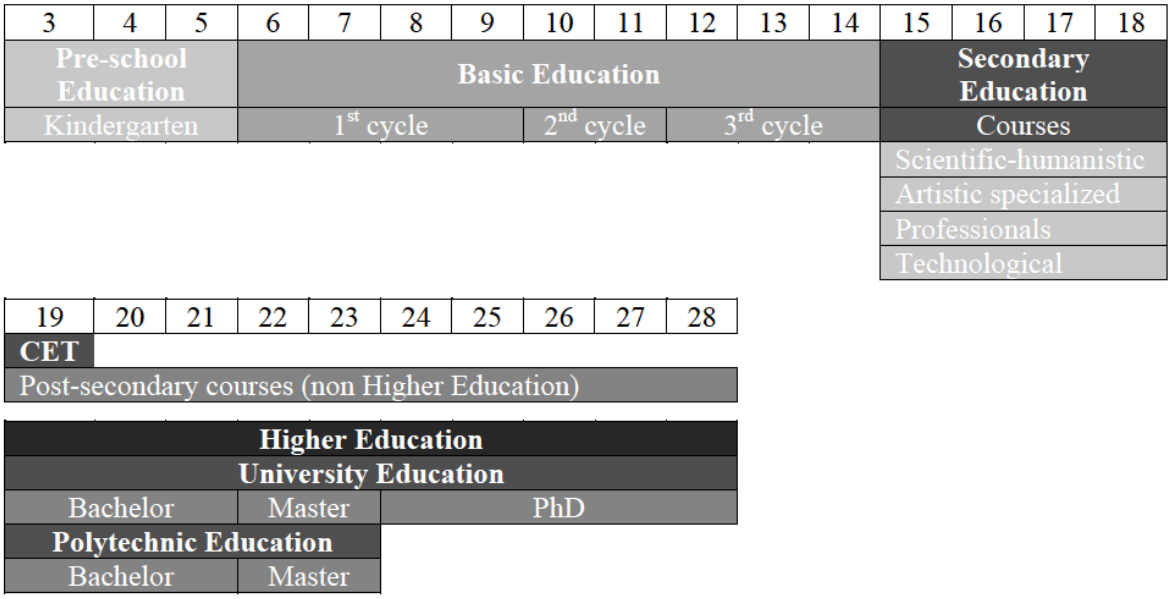


Figure 2.1: Organization of Portuguese Educational System

Although not mandatory, most children start pre–school education between 3 and 5 years old. Mandatory schooling starts at the age of 6, when they enter basic education. Basic education comprises the 1^{st}, 2^{nd} and 3^{rd} cycles (6–14 years old) and, it is followed by the secondary education (15–18 years old), which, currently, is the compulsory education. Those who get a secondary education diploma are entitled to apply to higher education.

Secondary education is structured in different tracks, targeted either for further study or for the labour market, namely students may opt from:

1. scientific–humanistic courses – aim mainly to the proceed to higher education studies;

2. technological courses – designed especially for students who wish to enter the labour market;

3. artistic specialized courses – aim at ensuring artistic specialized training in visual arts, audiovisual, dance or music;

4. professional courses – to prepare students for a job, and to to enter the the labour market.

In order to conclude the 3^{rd} cycle of basic education students have to take national exams of 9^{th} grade Mathematics and Portuguese subjects. Similarly, in certain subjects, the approval of students attending scientific–humanistic courses of secondary education is subject to a national final examination. Therefore, the approval of disciplines that

are subject to national final examination is obtained by frequency and mandatory final exam (internal students) or only by the realization of the national final exam (self–proposed students). Note that, in certain subjects, the national exams scores are also used as entrance examinations in higher education.

Higher education in Portugal is organized in three cycles. The first cycle takes 3 to 4 years and leads to the bachelor degree. The second cycle takes from one year and half to two years and allows students to get a master diploma. The third cycle is the PhD. The Portuguese higher education is a binary system, with universities and polytechnic institutes as the main providers, which can be either private or public. According to the Basic Law of the Portuguese Education System, polytechnic education differs from the university education because is more directed for cultural and top–level technique training, while university education is best suited for a scientific and cultural preparation. Universities offer bachelor, master and doctoral degrees, whereas polytechnic institutes can be responsible for bachelor and master programmes.

It is important to note that polytechnic institutes were introduced only in 1970s in Portuguese education system. In 1971, the Minister of Education *Veiga Simão* presented a project for the school system proposing deep structural changes. Its main purpose was the expansion and diversification of education. A new type of higher education was created, more oriented to professions which should be offered by polytechnic institutes. This followed from the need to train intermediate graduate technicians for industry, agriculture and other services, and the objective of assuring equity in the access to higher education, i.e. creating equal opportunities in education regardless of social origins and the region where people live. It was considered as short–duration higher education (Grácio, 1998). After the April 25^{th} revolution in 1974, the Law 513 – T79, of December 26^{th}, created a network of Polytechnic Institutes, covering all the national territory. The decision process for the creation of this network took into account the development needs of the regions that would harbour the Polytechnic Institutes, and those institutions were expected to train the technicians with practical preparation, in the scientific and economic areas that best suit the needs of each region.

Is should be noted, that these short–duration higher education courses offered by polytechnics led to the *bacharelato* degree, which was extinguished for the Portuguese higher education system with the Bologna process, in 2005. This degree distinguished from the bachelor degree because its duration was only 3 years, while the bachelor degree had a duration of five years.

Chapter 3

Characterization of returns to education in Portugal: 1986–2009

"The most valuable of all capital is that invested in human beings."
Alfred Marshall, *Principles of Economics*

3.1 Introduction

In 1986, the starting point of this analysis, Portugal joined the European Union. This event had a significant impact on the Portuguese labour market, following the integration into a global market, which increased exposure to the foreign competition. Consequently, many resources were directed at policies that had a significant effect in the wage structure, namely, policies that subsidized investment in modern technologies and in qualification of the population, which resulted in a change in the demand for skilled workers.

Over the past 25 years, there have been remarkable changes in the education of the Portuguese population, i.e. since the 1990s the fast expansion of secondary and tertiary education in Portugal has contributed to rapid chances in the educational structure of the workforce. According to *Instituto Nacional de Estatística*, between 1991 and 2011, the number of people with college degrees increased by about one million in tertiary education, quadrupling its original figure. In 1991 there were 284 thousand of degree graduates, in 2001 there were 674 thousand, and in 2011 there were 1260 thousand, representing 12% of the population and the majority being female. On the other hand, the number of people who completed the secondary level has also doubled in the last 20 years, from 643 thousand in the beginning of the 1990s to 1363 thousand currently, which represents 13% of the population.

The contemporary between changes in Portuguese workforce's schooling, high-

lighted by the huge increase in the supply of skilled individuals, and the size of the investments made[1] makes it crucial to assess the impact of these investments on the path of the returns associated with school attendance over these 25 years. This is particularly important because there is still a scarce knowledge about the variation in these returns since.

The rate of return to individual investments in schooling is an important factor in determining educational attainment, labour participation and wages, which makes it a relevant outcome measure to evaluate how beneficial were the investments described above. This fact made its estimation to be one of the predominant issues in applied economics over the last 50 years. Several authors have invested much energy in identifying the value of educational investment, aiming at determining whether governments and individuals are investing optimally. Much of this work stems from the work of Becker (1962) who introduced the concept of treating investment in education as a capital investment. Since then hundreds of papers have been published estimating the return to education investment.[2]

Over time, the empirical analysis of returns to education hinged upon a crucial debate around endogeneity issues involving schooling choices made by individuals. Endogeneity poses a key difficulty in the estimation of this economic parameter. The debate provided in the literature highlighted the Instrumental Variables (IV) approach as a useful tool to deal with this problem. When one reviews the research made for Portugal a conclusion comes about. Empirical studies of returns to education in the Portuguese labour market that make use of this solution are scarce. The current chapter aims, first, at a thorough analysis of private returns to education using IV, while, providing a comparison with the Ordinary Least Squares (OLS) counterparts. Second, it aims also to contribute to the knowledge of the impact of the investment in education in the productivity of Portuguese workers, by estimating the path of the private returns to education in Portugal over the period 1986–2009.

Following the literature, and accounting for the information available for the Portuguese labour market, different instruments will be used, namely quarter of birth, similar to the approach used by Angrist and Krueger (1991) and changes in the Portuguese compulsory schooling laws, similar to the approach used by Harmon and Walker (1995). Regarding the instruments, two novelties are introduced. First, changes in compulsory

[1] According to *Instituto Nacional de Estatística*, state expenditure on education have risen from 926.7 million Euro, in 1986, to 8507.4 million Euro, in 2009.

[2] See, for example, the reviews by Card, 1999; Harmon et al., 2003, and the meta analysis of Ashenfelter et al., 1999, for research on private returns to schooling; la Fuente and Ciccone, 2003, for research addressing the impact of education on the so–called "knowledge economy" through growth models; and Acemoglu and Angrist, 2001, or Oreopoulos and Salvanes, 2011, for research on wider externalities associated with education.

schooling introduced in 1964 and in 1986 will be used a instrument for education, an approach which was never used in the Portuguese literature. Second, a new instrument is proposed: the average education by region in the year in which individual entered school.[3]

This chapter proceeds in the following way. In Section 3.2, a literature review on the returns to education is carried out. Section 3.3 comprises a theoretical and econometric framework, followed by a discussion on the methodology and data, Section 3.4. Empirical results for the OLS and IV estimators are presented and discussed in Section 3.5. The chapter ends with some concluding remarks.

3.2 Literature review

In market economies wages reflect the productive features of both economy and workers. The human capital and technological knowledge are directly influenced by the investment decisions that individuals make throughout their lives. In this context, education is perceived as an investment (Mincer, 1958; Schultz, 1961; Becker, 1962) and, as such, it produces a return that may have social effects that have an impact in the population's welfare (Lochner and Moretti, 2004; Grossman, 2006), as well as, individual benefits (Card, 1999; Harmon et al., 2003; Psacharopoulos and Patrinos, 2004; Pereira and Martins, 2004b; Wang, 2013). Thus, the return to schooling, a wage premium associated to an additional year of education, has been the focus of research in economics in recent decades.

Mincer (1974) introduced a simplified strategy to estimate the rate of return to schooling. By taking a cross–section regression of individual earnings, the rate of return is estimated as the coefficient of years of schooling, after controlling for experience. Therefore, the rate of return measures how many additional units of income are generated by an extra unit of education, *ceteris paribus*.

Several studies have explored Mincer's research to estimate the impact of education on individuals' wages and have uniformly shown that more schooling is associated with higher individual earnings.[4] Psacharopoulos and Patrinos (2004), using data for 93 countries, suggests that the average rate of return to another year of schooling is 10%. In this line of research, Boarini and Strauss (2010), using data for 21 countries of the Organisation for Economic Co–operation and Development (OECD) for the period 1991–2005, suggest that the average internal rate of return is about 8%, ranging from 4% to 15%; a slightly growing trend could be observed. For the United States, there are

[3] To our knowledge, after reviewing in detail the available literature, this instrument has never been used.

[4] See Card (1999, 2001) for a literature review.

several studies on private returns to education. The results show estimated returns that range between 3% (Blackburn and Neumark, 1993) and 7% (Card, 1995a; Park, 2011). Additionally, the estimates obtained for the private returns to education for different European countries range between 8% and 10% (with an average value of about 9%). According to la Fuente and Jimeno (2009), the highest returns can be found in countries like the United Kingdom and Ireland, followed by Portugal and Finland (between 10% and 12%), while Sweden has the lowest returns. Psacharopoulos (1981) estimates for Portugal a rate of return of about 9% and Pereira and Martins (2004a) propose a rate of return of about 13%. This high value may be due to the low skill level of the Portuguese labour force. Portugal is one of the OECD countries with the lowest educational level and, thus, a greater return to such qualifications might be expected (Alves et al., 2010; OECD, 2012). Despite the diversity of estimates across countries, it should be noted that these rates are converging across European countries (Pereira and Martins, 2000).

Returns to education have been assessed, not only for average education, but also across different educational levels. It is shown that the wage benefits from an additional year of schooling is not similar across different education levels. Psacharopoulos (1985, 1994); Psacharopoulos and Patrinos (2004), based on cross–country evidence, shows decreasing returns, i.e. there are higher returns on primary education than on other schooling levels. Results show estimates of returns to an additional year of schooling in primary level of 27%, while the private return in higher educational level is 19%. In contrast, Trostel (2005), using cross–sectional individual data for the United Kingdom (UK), argues that during the first years of schooling the returns are practically null and reach their maximum at approximately 12 years of education. Similarly, Hungerford and Solon (1987) emphasize that returns are high in the year of diploma and the maximum is achieved when the 16 year old are completed. In turn, Song et al. (2008), focusing on the post–graduate education, found that the returns are not lower in the post–graduate level compared with the non–graduate level. In this context, several studies on the returns to tertiary education for different countries have emerged in the last decade. The estimates range between 7% and 19%. (Psacharopoulos and Patrinos, 2004, and Boarini and Strauss, 2010, for the OECD countries; Blondäl et al., 2002, and la Fuente and Jimeno, 2009, for European countries).

More recently, Carneiro et al. (2011), who developed a model to estimate the returns to education considering that returns are heterogeneous, estimate marginal returns to college for individuals who are led to enrol in university due to marginal changes of policies. They discuss how economic theory and local instrumental variables estimators can be combined to assess the efficacy of marginal policy changes, i.e., they discuss how

to use local instrumental variables to estimate the effects of marginal policy changes. Their empirical analysis, using data from the National Longitudinal Survey of Youth of 1979, provide evidence that the marginal return to college is often bellow the average return. They show that the returns are higher for the individuals who are more likely to attend university, and that the marginal policy changes which lead students to go to university produce lower returns to education.

Within the last decade, Park (2011) analysed the returns to schooling for individuals who returned to school after a certain period in the labour market. The author found that the reinvestment is associated with an increase of about 3.5 percentage points (p.p.) in the estimated return to an additional year of schooling.

Over the years a wide set of statistical and econometric techniques have been discussed in the literature to determine returns to education (Mincer, 1974; Card, 1999, 2001; Harmon et al., 2003; Boarini and Strauss, 2010). Schlotter et al. (2011) provide an up–to–date review of the econometric estimation procedures applied to earnings functions using individual level data. Mincer's standard model assumes that schooling is independent of individual's ability. However, in the standard formulation of the wage regression, the literature has being pointing out that schooling is most likely endogenous. If this is the case, the OLS estimate of average returns to education would be upward biased (Card, 1995a, 1999, 2001). The argument is if omitted ability is positively correlated with education, as, for example, more able individuals have lower costs of learning, and, at the same time, has a positive effect on wages, so the OLS estimate of the parameter of interest would capture both the effect of education and ability; once one filters out ability, the estimated return to education is expected to diminish.

Thus, when there is no ability measure, or at least a proxy, such as intelligence quotient tests, to reduce the ability bias, one of the solutions proposed in the literature is an instrumental variable (IV) approach (Harmon and Walker, 1995; Card, 1995a; Vieira, 1999; Psacharopoulos and Patrinos, 2004; Oreopoulos, 2006; Pischke and von T. Wachter, 2008; Devereux and Fan, 2011; Wang, 2013). The underlying econometric theory shows that, if one has valid instruments, this IV filters the bias originated in innate ability or, for example, motivation (Harmon and Walker, 1995). Theoretically, and within Mincer's framework, valid instruments are variables correlated with education, but not correlated with residuals.

Different instruments have been proposed in the literature. The main ones include compulsory schooling laws (Harmon and Walker, 1995; Vieira, 1999; Pereira and Martins, 2002a; Oreopoulos, 2006; Devereux and Hart, 2010; Dickson, 2012; Grenet, 2013), proximity from school to residence (Card, 1995b), season of birth of the individual (An-

grist and Krueger, 1991; Leigh and Ryan, 2008) and family background (Card, 1995a; Blackburn and Neumark, 1993; Pereira and Martins, 2002a; Wang, 2013). Devereux and Fan (2011) have exploited the expansion of tertiary education attendance as an instrument for education.

Results based on IV are diverse across countries and instruments used. Using compulsory schooling laws as an instrument for education, IV estimates of returns to education, found for some European countries, range from about 5% to 15% (see Brunello and Miniaci, 1999, for Italy; Levine and Plug, 1999, for Netherlands; Black et al., 2005, for Norway; Harmon et al., 2003 and Devereux and Hart, 2010, for the UK). However, Grenet (2013) and Pischke and von T. Wachter (2008), using data for France and Germany, respectively, found very small IV estimates, which, to some extent, suggest almost zero returns to schooling. Estimates for Great Britain show returns to education around 9% (Oreopoulos, 2006) and around 15% for UK (Harmon and Walker, 1995).

Evidence for Portugal has made a contribution for this branch of research. Portugal is one of the OECD countries with lower educational level and one of the European countries where returns to education are higher, which indicates that skills are particularly valued in the Portuguese labour market. Several studies conclude that the returns to formal education in Portugal are substantial (Vieira, 1999; Pereira and Martins, 2000, 2002a,b, 2004a,b; Budria and Nunes, 2005; Vieira et al., 2006; Budria and Pereira, 2011).

Using OLS, Pereira and Martins (2002a,b, 2004a,b) propose that, on average, an additional year of education raises wages by about 10% to 13%. However, this size of the rate of return is not consensual. Other results suggest that the private return to education in Portugal ranges between 5%–9% (Psacharopoulos, 1981; Vieira, 1999; Hartog et al., 2001; Machado and Mata, 2001; Carneiro, 2008). In addition, since the beginning of the 1980s until the mid–1990s, the private economic return to education took a growing trend and a reversion has been registered from then on; the peak was reached in 1995 (Pereira and Martins, 2002a, 2004b). It is also found that the gross returns are systematically higher than the net returns, which can be attributed to the progressive nature of the Portuguese tax system (Pereira and Martins, 2002a).

Following the international literature, in order to reduce the bias of the estimates caused by endogeneity of education variable, Vieira (1999) and Pereira and Martins (2002a) use the IV approach. Based on data only for men from the *Quadros de Pessoal*, and using the changes in the compulsory schooling law in 1956 (4 years of schooling) and in 1964 (6 years of schooling) as instruments to education, the authors found IV estimates for returns around 5%, which are lower than the results obtained using

OLS by about 3 p.p.. Additionally, Pereira and Martins (2002a), using European Community Household Panel (ECHP) data and the parental education as instrument, found that the IV estimate is 13%, which is 3 p.p. higher than the OLS estimate. The difference in results when using different instruments lead the authors to conclude that the IV results are inconclusive. It should be noted that, in the literature for Portugal on the private returns to education, the evidence obtained by IV methodology is still very scarce (see Pereira and Martins, 2004b's meta–analysis about returns to education in Portugal).

Over time, returns to education have been explored not only in an average return perspective. Thus, some Portuguese researches show that differences between education groups are substantial. Despite being relatively recent in Portugal, pre–school education could have a crucial role in the returns to education, since the investment in education has higher returns the earlier in the life cycle are realized (Carneiro, 2008). Within the same line of reasoning, Psacharopoulos (1981) suggests that the returns to education of the primary level are the highest when compared with the other levels. Contrary, Saraiva (1999) and Hartog et al. (2000) conclude that apparently there are no positive returns from the learning system and the vocational training schools, but they point out positive returns from the vocational training in higher education. In the same direction, other authors propose that a higher schooling level is associated with higher earnings (Pereira and Martins, 2002a; Budria and Pereira, 2011; Hartog et al., 2001). So, Budria and Pereira (2011) suggest that individuals with higher education earn wages that are at least 70 p.p. higher than the wages earned by individuals in the lowest educational category (lower secondary), 26%, and more than 40 p.p. higher than those earned by individuals in the upper secondary group. Similarly, Portugal (2004) reports that the wage premium of the workers with a graduation degree (*licenciatura*) is 63.1%. Consistent with these results, Pereira and Martins (2002a) reports, for example, that workers with a *Bacharelato* earn about 300% more than workers with no schooling. Additionally, Pereira and Martins (2002a) show that the average return to an extra year of schooling for the transition between the 9^{th} grade and 12^{th} grade increases the wages in 7%, while the increase associated to an additional year of education for the transition between secondary level and tertiary education varies from 12% to 18% (depending on the year considered).

Pereira and Martins (2002a), making the distinction between public and private sector, propose a higher return to education for the individuals employed in the private sector of about 1.5 p.p. than those employed in the public sector, where the rate of return is about 7.4%. This result is in line with that found by Psacharopoulos (1981), which suggests that the average return of an extra year of schooling is larger

in the private sector (6.2%) compared to the public sector (5.1%). By examining the distinction between genders, Carneiro (2008) suggests that the return to an additional year of schooling is about 7% for men and 9% for women, which is consistent with the work of Hartog et al. (2001), but contradicting the results obtained by Pereira and Martins (2002a), which show that there are no substantial differences between the genders. In addition, Vieira et al. (2005, 2006) suggest that the highest returns are found in region of Lisbon. According to the authors, the wage premium varies between -15.1% in *Bragança* and 8.9% in Lisbon, in 1996, and, in 2000 the differences amount to -13.8% in *Castelo Branco* and 8.8% in Lisbon. This result may be due to job supply characteristics. Due to the higher demand of more educated workers associated to technology gap (jobs complexity) required in Lisbon, the returns to education would be greater (Teulings and Vieira, 2004; Vieira et al., 2005, 2006).

Finally, aiming to explore the impact of the education on wages along the wage distribution, Quantile Regression (QR) has been applied over the last two decades (Machado and Mata, 2001, 2005; Hartog et al., 2001; Pereira and Martins, 2002a,b, 2004a; Budria and Pereira, 2011). With the QR approach, the wage equation is estimated at different points of the wage distribution, allowing to assess the influence of the covariates in these points. It is observed that the OLS estimation hides the differences in the returns to education (Lima and Pereira, 1999). For instance, an average return of about 96% to tertiary education masks a return of 75% in the bottom quantile and 104% in the top quantile (Hartog et al., 2001). So, comparing workers with similar characteristics, it is found that the returns to education of those who have high wages are much larger than those of workers with lower wages, suggesting that the dispersion of the earnings increases with the average educational level of the workers, i.e., the returns to education are greater at the top of the wage distribution than at the bottom (Machado and Mata, 2001; Hartog et al., 2001; Pereira and Martins, 2002a, 2004a). This can be due to the existence of an interaction between schooling and ability, due to the over–education or due to the difference in school quality or fields of study.[5] At the bottom of the wage distribution there should be a significant part of the workers with low levels of quality school, or who followed fields of study with lack of interest in Portuguese labour market (Pereira and Martins, 2004a).

[5] Pereira and Martins (2002a) suggest that the degrees of Engineering are highly compensated comparing with the degrees in Social and Human Sciences (except Economics), which attract lower returns. This results is consistent with Portugal (2004), who suggests a significant wage premium in technological degrees.

3.3 Theoretical framework

3.3.1 Mincerian model

The analysis of the demand for education stems from the work on human capital by the pioneers Gary Becker, Jacob Mincer and Theodore Schultz. In human capital theory, education is an investment of current resources (the opportunity cost of the time involved, as well as any direct costs) in exchange for future returns (Harmon et al., 2003).

The reference model for the development of empirical estimation of the returns to education is the key relationship derived by Mincer (1974). The Mincerian model assumes that individuals choose an education level that maximizes their expected present value of the stream of future incomes, up to their retirement date, R; years of work is defined as $(T = R - S)$. The optimum level of schooling, S, is obtained by equating marginal benefits and marginal costs, where benefits are given by the wage differential obtained by an extra year of education. The costs incurred by the individual are the opportunity costs, i.e. the postponed income from employment during the period of school enrolment (it is assumed that during this time individuals do not work).

Consider that $W(S,t)$ represents the earnings in time t of an individual with S years of education, where it is assumed that earnings are constant during individual's lifetime. If one uses discrete time, the present value of income associated with the level of schooling S is given by

$$V(S) = W(S) \sum_{t=S+1}^{R} \left(\frac{1}{1+r}\right)^{t}, \tag{3.1}$$

while in continuous time it is defined as

$$V(S) = W(S) \int_{S}^{R} e^{-rt}\, dt = \frac{W(S)}{r}\left(e^{-rS} - e^{-rR}\right), \tag{3.2}$$

where r denotes the discount rate, which is exogenous.

The ratio of individual's income with S years of schooling, $V(S)$, to those who do not have any level of schooling, $V(0)$, is obtained by

$$\frac{W(S)}{r}\left(e^{-rS} - e^{-rR}\right) = \frac{W(0)}{r}\left(e^{-r\times 0} - e^{-rR}\right),$$

or

$$\frac{W(S)}{W(0)} = \frac{1 - e^{-rR}}{e^{-rS} - e^{-rR}}. \tag{3.3}$$

Applying logs to equation (3.3) one gets

$$\log W(S) = \log W(0) + \log\left(\frac{1 - e^{-rR}}{e^{-rS} - e^{-rR}}\right). \tag{3.4}$$

Under the assumption that individuals live infinitely, and never withdraw from the labour market, $R = \infty$, equation (3.4) becomes

$$\log W(S) = \log W(0) + rS \tag{3.5}$$

The econometric formulation of equation (3.5) is defined as

$$\log W(S) = \alpha + rS + \varepsilon, \tag{3.6}$$

where $\alpha = \log W(0)$ and ε is, for the time being, a white noise error term.

However, Mincer (1974) addresses the same model from a different perspective, by assuming that agents can also invest in human capital after school through experience in order to acquire and improve their skills. If after the end of the school period no additional investment takes place, and there is no depreciation of human capital, income during the life of the worker is given by $W(S)$. Additionally, Mincer assumes that time invested by individuals in its qualifications tend to diminish over time. Thereby, if h_t represents the proportion of period t spent to acquire human capital (during full–time education $h_t = 1$), it is assumed that h_t declines linearly with experience, so

$$h_t = h_0 - \frac{h_0}{R}t$$

Thus, for x years of post–school work experience, earnings can be written as

$$W(x, S) = W(S)e^{r\int_0^x h_t dt} \tag{3.7}$$

wherein

$$\int_0^x h_t\, dt = h_0 x - \frac{h_0}{2R}x^2. \tag{3.8}$$

Substituting, it is possible to write the expression for income after x years of experience and S years of schooling as

$$W(x, S) = W(S)e^{r\left(h_0 x - \frac{h_0}{2R}x^2\right)}. \tag{3.9}$$

Taking logs it becomes

$$\log W(x,S) = \log W(S) + rh_0x - r\frac{h_0}{2R}x^2 \tag{3.10}$$

and replacing equation (3.5) in equation (3.10) we obtain

$$\log W(x,S) = \log W(0) + rS + rh_0x - r\frac{h_0}{2R}x^2 \tag{3.11}$$

Knowing that actual earnings are given by $W_x = (1-h_x)W(x,S)$, or $\log W_x = \log(1-h_x) + \log W(x,S)$, the conventional Mincer specification is defined as

$$\log W = \log(1-h_x) + \log W(0) + rS + rh_0x - \frac{rh_0}{2R}x^2 \tag{3.12}$$

where $\alpha = \log(1-h_x) + \log W(0)$, $\beta = r$, $\gamma_1 = rh_0$ and $\gamma_2 = -\frac{rh_0}{2R}$. This specification is known by Mincerian standard equation and it is a log–linear relationship between earnings and schooling and experience.

Therefore, the empirical approximation of the human capital theoretical framework is the following wage equation which is an empirically testable relationship:

$$\log W_i = \alpha + X_i\delta + \beta S_i + \gamma_1 x_i + \gamma_2 x_i^2 + \varepsilon_i \tag{3.13}$$

where W_i represents earnings of individual i; S_i denotes a measure of his schooling; x_i is an experience measure;[6] X_i is a set of other variables assumed to affect earnings; and ε_i is a disturbance term denoting other variables not controlled for, which are assumed to be independent of both X_i and S_i (for now we assume $\varepsilon_i \sim i.i.d.(0,1)$). Note that, experience is included as a quadratic term to capture the concavity of the experience earnings profile. Under the assumptions made (particularly no tuition costs), β can be considered the private return to education as this measures how much wages increase due to an additional year of schooling.

3.3.2 Econometric framework

If the Mincerian equation is estimated by Ordinary Least Squares (OLS) method, the estimates of the returns to education may be biased due to the endogeneity of education (if $\text{Cov}(S_i, \varepsilon_i) \neq 0$), i.e. the OLS estimates of the return to education will not give the causal effect of schooling on earnings, because the schooling coefficient captures some of the effects that would otherwise be attributed to the omitted variables. For

[6] Typically age minus schooling attainment years minus age at which schooling started.

instance, if the omitted variable is ability or motivation, and if both schooling and earnings are positively correlated with these omitted variables, OLS estimation ignores that more able or motivated people are likely to earn more than less able or motivated people even when they have similar schooling levels. Thus, it can be said that the years of schooling are not sufficient to explain the return to education. So, as in Verbeek (2012); Wooldridge (2010), the endogeneity of education can be understood as an omitted variable bias. Under this setup, excluding a relevant variable (e.g., innate ability or motivation) will result in an overestimation of β when using the OLS estimator applied to equation (3.13).

For instance, supposing that wage is determined by

$$W_i = \alpha + \beta S_i + \lambda A_i + u_i \tag{3.14}$$

where W_i represents wages, S_i denotes education and A_i is innate ability. Since innate ability is not observed, we instead estimate the model:

$$W_i = \alpha + \beta S_i + \varepsilon_i \tag{3.15}$$

where $\varepsilon_i = \lambda A_i + u_i$.

The estimator of β from the regression of wage on education, denoted by $\widetilde{\beta}$, is given by

$$\begin{aligned}
\widetilde{\beta} &= \frac{\sum_{i=1}^{n} (s_i - \overline{s})(w_i - \overline{w})}{\sum_{i=1}^{n} (s_i - \overline{s})^2} \\
&= \frac{\sum_{i=1}^{n} (s_i - \overline{s})(\beta s_i + \lambda a_i + u_i - \beta\overline{s} - \lambda\overline{a} - \overline{u})}{\sum_{i=1}^{n} (s_i - \overline{s})^2} \\
&= \beta + \lambda \frac{\sum_{i=1}^{n} (s_i - \overline{s})(a_i - \overline{a})}{\sum_{i=1}^{n} (s_i - \overline{s})^2} + \frac{\sum_{i=1}^{n} (s_i - \overline{s})(u_i - \overline{u})}{\sum_{i=1}^{n} (s_i - \overline{s})^2}.
\end{aligned} \tag{3.16}$$

Under the assumption that $E(u_i) = 0$, one has

$$E\left(\widetilde{\beta}\right) = E\left[\beta + \lambda \frac{\sum_{i=1}^{n}(s_i - \overline{s})(a_i - \overline{a})}{\sum_{i=1}^{n}(s_i - \overline{s})^2} + \frac{\sum_{i=1}^{n}(s_i - \overline{s})(u_i - \overline{u})}{\sum_{i=1}^{n}(s_i - \overline{s})^2}\right]$$
$$= \beta + \lambda \frac{\sum_{i=1}^{n}(s_i - \overline{s})(a_i - \overline{a})}{\sum_{i=1}^{n}(s_i - \overline{s})^2} \neq \beta. \tag{3.17}$$

Thus, because the variables education and ability are correlated, the OLS estimator of β overestimates the impact of education on wages and the ability bias is given by:

$$\text{plim}\widetilde{\beta} = \beta + \lambda \frac{\text{Cov}(A, S)}{V(S)}$$

or

$$\text{plim}\widetilde{\beta} - \beta = \lambda \frac{\text{Cov}(A, S)}{V(S)} \tag{3.18}$$

If $\lambda > 0$, and A and S are positively correlated, the asymptotic bias is positive (Verbeek, 2012; Wooldridge, 2010).

3.3.2.1 Instrumental variables method

A useful tool to deal with the problem of endogeneity of education is the Instrumental Variables (IV) method. At the core of this estimation method is the identification of one or a set of variables which affect the individual's choice of education, but do not independently enter into the earnings equation, and are uncorrelated with the error term. As was previously mentioned, and because OLS does not address the non-null covariance between education and the error term, $\text{Cov}(S_i, \varepsilon_i) \neq 0$ in equation (3.15), it generally results in inconsistent estimates of β (Verbeek, 2012; Wooldridge, 2010).

The conventional IV approach estimates the following two–equation system:

$$\log W_i = \beta_1 X_i^* + \beta_2 S_i + \varepsilon_i \tag{3.19}$$

and

$$S_i = \theta Z_i + v_i \tag{3.20}$$

where equation (3.19) is an extended version of equation (3.15) and X_i^* is a matrix with exogenous regressors, $E(X_i^*, \varepsilon_i) = 0$, of dimension $N \times k$, where k is the number

of variables and N is the number of observations.[7]

The schooling equation (3.20) is a reduced form in which Z is a vector of exogenous variables that influence the schooling decision, but not wages. Z_i is a matrix that defines the set of instruments, with dimension $N \times l$, where l is the number of instruments. A suitable instrumental variable included in Z must satisfy two conditions: relevance and exogeneity. The relevance condition requires that the instrument(s) be correlated with the number of years of schooling of the individual, $E(Z_i, S_i) \neq 0$. The exogeneity condition requires that the instrument affect wage only through the channel of schooling and, thus, the instrument must be uncorrelated with the error term in earnings equation, $E(Z_i, \varepsilon_i) = 0$ (Verbeek, 2012; Wooldridge, 2010; Leigh and Ryan, 2008).

Given the conditions above, the population moments can be written as:

$$E(X^*, \varepsilon) = E\left[X'^*(\log W - \beta_1 X^* - \beta_2 S)\right] = 0$$

$$E(Z, \varepsilon) = E\left[Z'(\log W - \beta_1 X^* - \beta_2 S)\right] = 0$$

and their sampling analogues:

$$\frac{1}{N}\sum_{i=1}^{N}\left[X_i^*\left(\log W_i - \widehat{\beta}_{1,IV} X_i^* - \widehat{\beta}_{2,IV} S_i\right)\right] = 0$$

$$\frac{1}{N}\sum_{i=1}^{N}\left[Z_i\left(\log W_i - \widehat{\beta}_{1,IV} X_i^* - \widehat{\beta}_{2,IV} S_i\right)\right] = 0$$

Defining $X_i = (X_i^*, S_i)$, and solving the above equations, we obtain the expression for the IV estimator

$$\widehat{\beta}_{IV} = \frac{\sum_{i=1}^{N}(Z_i - \overline{Z})(\log W_i - \overline{\log W})}{\sum_{i=1}^{N}(Z_i - \overline{Z})(X_i - \overline{X})} \tag{3.21}$$

where $\widehat{\beta}_{IV} = (\widehat{\beta}_{1,IV}, \widehat{\beta}_{2,IV})$.

So, by the law of large numbers

$$p\lim \widehat{\beta}_{IV} = \frac{\mathrm{Cov}(Z, \log W)}{\mathrm{Cov}(Z, X)} = \beta,$$

[7] Estimation of equation (3.19) by OLS will only yield an unbiased estimate of the parameters, and one can interpret β_2 as the return to schooling, if S_i is exogenous, $E(S_i, \varepsilon_i) = 0$ (Harmon and Walker, 1995).

where $\widehat{\beta}_{IV}$ is a consistent estimator of β.

The empirical implementation of the IV estimator can be done by the Two–Stage Least Squares (2SLS) estimator. Considering $X_i = (X_i^*, S_i)$ and $\beta = (\beta_1, \beta_2)$, we can write equation (3.19) as

$$\log W_i = \beta X_i + \varepsilon_i. \tag{3.22}$$

According to the 2SLS method, in order to model the relationship between schooling and wages in a first stage one starts by using the schooling equation to compute the predicted value of schooling. In the actual estimation of equation (3.20) one uses not only Z_i as a set of regressors, but also X_i^*. The predicted values of schooling will replace the schooling variable in the wage equation (3.22).

Consequently, in the first stage it is carried out the regression of S on Z, where one obtains consistent estimates for the fitted values

$$\widehat{S}_i = \widehat{\theta} Z_i$$

where Z is a matrix of dimension $N \times l$, $l \geq k$ of instruments and the estimator of θ is given by:

$$\widehat{\theta} = (Z'Z)^{-1} Z'S$$

where $Z = (z_1, \ldots, z_N)$ and $X = (s_1, \ldots, s_N)$. Thus,

$$\widehat{S} = Z(Z'Z)^{-1} Z'S$$

In the second stage, equation (3.22) is estimated applying OLS, but the values of S_i are replace by their predicted values obtained in the first stage, $\widehat{S}_i$. The coefficient on $\widehat{X}_i$ in the regression of $\log W_i$ on $\widehat{X}_i$ is the two–stage least squares estimator of β. We define $\widehat{X}_i = (X_i^*, \widehat{S}_i)$.

For the sake of simplicity, in the following discussion we will consider that equation (3.22) has just one explanatory variable, so $\widehat{X}_i = \widehat{S}_i$. We now can define the second stage equation as

$$\log W = \beta \widehat{X} + \varepsilon \tag{3.23}$$

The form of the 2SLS estimator is:

$$\begin{aligned} \widehat{\beta}_{2SLS} &= \left(\widehat{X}'\widehat{X}\right)^{-1} \widehat{X}' \log W \\ &= \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1} X'Z(Z'Z)^{-1}X' \log W \\ &= (X'P_Z X)^{-1} X'P_Z \log W \end{aligned} \tag{3.24}$$

where $P_Z = Z(Z'Z)^{-1}Z'$ denotes the projection matrix. If $l > k$ the model is overidentified. When $l = k$, the 2SLS is exactly identified and is reduced to the IV estimator, so

$$\widehat{\beta}_{2SLS} = (Z'X)^{-1}Z' \log W = \widehat{\beta}_{IV} \tag{3.25}$$

Although both the theory and the empirical literature suggest that schooling is endogenous, this hypothesis should be tested. If schooling were not endogenous, the OLS approach would yield more efficient estimates than the IV approach. Literature discuss the issue of OLS *versus* IV in the context of the Durbin–Wu–Hausman (DWH) test. This test involves the adjustment of the model for both approaches and comparing the resulting coefficient vectors, i.e. the idea is to check whether the difference $\left(\widehat{\beta}_{IV} - \widehat{\beta}_{OLS}\right)$ is significantly different from zero. The test is distributed as χ^2 with degrees of freedom equal to the number of regressors being tested for endogeneity. The null hypothesis of the DWH test is that the potential endogeneity of the variable does not bias the estimated coefficients, i.e., it allows us to reject, or not, the null hypothesis that the OLS estimate is consistent $\widehat{\beta}_{IV} = \widehat{\beta}_{OLS}$. A rejection of the null indicates that endogenous regressors' effects on the estimates are meaningful, and appropriate estimation methods, among which IV, are required (for instance, Davidson and MacKinnon, 1999; Baum et al., 2003, 2007).

3.3.2.2 Weak instruments

The IV estimation sometimes is difficult to implement. If the instruments used are weakly correlated with endogenous variables, it might yield biased estimates (Garcia-Mainar and Montuenga-Gómez, 2005). Carneiro and Heckman (2002) suggest that most of the empirical studies on return to education uses invalid instruments, which tend to have high correlation with the omitted individual unobserved abilities. Therefore, the weak instruments have received considerable attention by several authors (Bound et al., 1995; Staiger and Stock, 1997; Stock and Yogo, 2005; Baum et al., 2003, 2007). According to Bound et al. (1995), instruments' quality is insured only if there is a strong correlation between the instrument and the endogenous variable. Otherwise, the weak correlation with instruments and dependent variable can yield a large bias in the IV estimates, which can be even more inconsistent than those obtained by OLS estimation, even if the sample size is large.

To illustrate the problem, let us consider the IV estimator for the case of a single

regressor,

$$p\lim \widehat{\beta}_{IV} = \beta + p\lim \frac{\frac{1}{N}\sum_{i=1}^{N}\left(Z_i - \overline{Z}\right)\varepsilon_i}{\frac{1}{N}\sum_{i=1}^{N}\left(Z_i - \overline{Z}\right)\left(S_i - \overline{S}\right)}$$
$$= \beta + \frac{\mathrm{Cov}(Z,\varepsilon)}{\mathrm{Cov}(Z,S)}$$

but $\mathrm{Cov}(Z,\varepsilon) = \dfrac{\mathrm{Corr}(Z,\varepsilon)}{\sigma_Z\,\sigma_\varepsilon}$ and $\mathrm{Cov}(Z,S) = \dfrac{\mathrm{Corr}(Z,S)}{\sigma_Z\,\sigma_S}$, thus,

$$p\lim \widehat{\beta}_{IV} = \beta + \frac{\mathrm{Corr}(Z,\varepsilon)}{\mathrm{Corr}(Z,S)} \times \frac{\sigma_S}{\sigma_\varepsilon} \tag{3.26}$$

The inconsistency in the IV estimator can be larger if $\mathrm{Corr}(Z,\varepsilon) \neq 0$ and $\mathrm{Corr}(Z,S) \approx 0$, and the asymptotic distribution of $\widehat{\beta}_{IV}$ deviates substantially from a normal distribution. So, the instrument is weak if there is some correlation between Z and S, but not enough to make the asymptotic normal distribution provide a good approximation in finite samples, even if potentially very large (Verbeek, 2012).

Comparing with OLS estimator, one has:

$$p\lim \widehat{\beta}_{OLS} = \beta + \frac{\mathrm{Cov}(S,\varepsilon)}{\mathrm{Cov}(S,S)}$$

but $\mathrm{Cov}(S,\varepsilon) = \dfrac{\mathrm{Corr}(S,\varepsilon)}{\sigma_S\,\sigma_\varepsilon}$ and then,

$$p\lim \widehat{\beta}_{OLS} = \beta + \mathrm{Corr}(S,\varepsilon) \times \frac{\sigma_S}{\sigma_\varepsilon} \tag{3.27}$$

Comparing equation (3.26) with equation (3.27), the OLS estimator will have smaller asymptotic bias than the IV estimator whenever

$$\mathrm{Corr}(S,\varepsilon) < \frac{\mathrm{Corr}(Z,\varepsilon)}{\mathrm{Corr}(Z,S)}$$

As an example, Bound et al. (1995) re–estimates the results from Angrist and Krueger (1991) and finds that the hundreds of instruments used in that study are mostly uncorrelated with S, which can result in IV being more biased that OLS (Harmon et al., 2001).

Due to this problem, there have emerged in econometrics literature various tests to

examine the hypothesis of weak instruments (Bound et al., 1995; Staiger and Stock, 1997; Stock and Yogo, 2005; Baum et al., 2007). Bound et al. (1995) propose that the researcher check the quality of IV estimates in two ways. First, the F–statistic on excluded instruments in the reduced form schooling equation needs to indicate statistical significance, i.e., as a simple rule–of–thumb it is suggested that if the F–statistics of the reduce form exceeds 10 there in no weak instruments problem. Secondly, they argue that the addition of the instrument to the reduced form equation needs to improve substantially its R^2. However, the statistic proposed by the authors is only able to diagnose the instrument's relevance only in the presence of a single endogenous regressor (Baum et al., 2003).

Stock and Yogo (2005) have helped to establish that even when an instrument is significant at conventional levels, it may still be weak and lead to the problems of bias and unreliable inference outlined above. The Stock and Yogo (2005) test is calculated based in the F–statistic of the Cragg–Donald, and it is defined by the null hypothesis that the selected instruments are weak and the bias checked is unacceptably large. The authors tabulate critical values that are determinate by the IV estimator the research is using, the number of instruments, the number of included endogenous regressors and how much bias or size distortion the research is willing to tolerate. The weak instrument test rule proposed by the authors is, if the Cragg–Donald F–statistic exceed the tabulated critical value one can concluded that the instruments are strong, otherwise conclude that they are weak. Thus, it is fundamental to establish that there is a strong relationship between the instrument and the endogenous regressor (schooling), i.e. that the instrument is relevant, and that it passes the various tests to determine that it is not a weak instrument (Baum et al., 2003, 2007).

In addition, to test the relevance of the instrument, it should be tested the condition of validity of the instrument. In order to test the exogeneity of the instrument, i.e. to ensure that the potential instruments have no relationship with the unobserved error term of the wage equation ($E(Z_i, \varepsilon_i) = 0$), the over–identification tests should be implemented. However, this implementing this test is only feasible if we have more instruments than endogenous variables.

The advantage of having more instruments is that they help us to reduce the variance of the estimator. In accordance with econometric literature, the degree of over–identification of an over–identified linear regression model is defined to be $(l - k)$, i.e. a model that implicitly incorporates $(l - k)$ over–identifying restrictions (Davidson and MacKinnon, 1999). A test of over–identifying restrictions regresses the residuals, $\widehat{\varepsilon}$, from instrumental variables two stage least squares (IV2SLS) estimation of the original model on the full set of instruments, Z. Under the null hypothesis that all

instruments are uncorrelated with the error term, ε, which states that the validity of the instruments is not rejected, the test is asymptotically distributed as a $\chi^2(l-k)$ distribution where $(l-k)$ is the number of over–identifying restrictions (Davidson and MacKinnon, 1999; Baum et al., 2003, 2007). A test based on this statistic is often called a Sargan test, which is computed as $(N \times R^2)$. But this test is inconsistent when heteroscedasticity of the errors is present. So, in this case, the Hansen's J test is recommended (the test computes the statistic from a two–stage estimate). Thus, to the efficient Generalized Method of Moments (GMM)[8] estimator, the Hansen's J statistic is the minimized value of the objective function. Baum et al. (2003) claim that the Sargan's statistic is a specific case of Hansen's J statistic under the assumption of conditional homoscedasticity. Sargan–Hansen test should be performed routinely in any over–identified model estimated with instrumental variables techniques. One final note is appropriate. Instrumental variables techniques are powerful, but if a strong rejection of the null hypothesis of the Sargan–Hansen test is checked, it should be questioned the validity of the estimates (Baum et al., 2003, 2007).

[8]Considering the set of l moment conditions, corresponding to the l instruments, the underlying concept of GMM is to choose an estimator for β that solves the following equations:

$$\overline{g}(\widehat{\beta}) = \frac{1}{N}\sum_{i}^{N} Z'(W_i - \widehat{\alpha} - \widehat{\beta}S_i) = 0 \tag{3.28}$$

According to Baum et al. (2003, 2007), if the equation to be estimated is exactly identified, l=k, then one can exactly solve the l moment conditions for the k coefficients in $\widehat{\beta}$. In this case, there is a unique $\widehat{\beta}_{GMM}$ that solves the equation (3.28), and this GMM estimator is in fact the IV estimator of (3.25). If the equation is over–identified, $l > k$, there are more equations (l moment conditions) than unknowns (k coefficients of $\widehat{\beta}$) and it will not be possible to find a k–vector $\widehat{\beta}$ that will set all l sample moment conditions to exactly zero. Thus, the aim is to choose $\widehat{\beta}$ so that the elements of $\overline{g}(\widehat{\beta})$ are as close to zero. In this case, we need to take a weight matrix Ω $(l \times l)$ that accounts for the correlations among $\overline{g}(\widehat{\beta})$ when the errors are not i.i.d., and use it to construct a quadratic form in the moment conditions. So, the GMM estimator chooses the $\widehat{\beta}$ that minimizes $J(\widehat{\beta}) = N\overline{g}(\widehat{\beta})'\Omega\overline{g}(\widehat{\beta})$. A GMM estimator for β is the $\widehat{\beta}$ that minimizes $J(\widehat{\beta})$. Solving the k first order conditions $\dfrac{\partial J(\widehat{\beta})}{\partial \widehat{\beta}} = 0$ yields the GMM estimator of an over–identified equation:

$$\widehat{\beta}_{GMM} = (S'Z\Omega Z'S)^{-1}S'Z\Omega Z'W$$

Consistency is ensured for any positive symmetric weight matrix Ω and, so, there are as many GMM estimators as there are choices of weighting matrix Ω. However, efficiency is not ensured for an arbitrary Ω. Thus, the efficient estimator is the GMM estimator with optimal weighting matrix Ω which minimizes the asymptotic variance of the estimator.

3.4 Methodology and data

3.4.1 Methodology

Initially, and following Pereira and Martins (2002a), the returns to education are estimated, by OLS, with respect to log wages ($\log W$) as in the standard Mincer's wages function,

$$\log W_i = \alpha + \beta S_i + \gamma_1 \, exper_i + \gamma_2 \, exper_i^2 + \varepsilon_i \tag{3.29}$$

where W is hourly real wages, S denotes years of schooling, *exper* represents potential experience in the labour market and ε is a random error term that captures other unobservable determinants of wages, assumed to be uncorrelated with S and other explanatory variables. The index i refers to individuals ($i = 1, 2, 3, \ldots, n$).

Additional regressors will be added to equation (3.29) to control for other individuals' characteristics (gender, tenure, occupation). Variables that characterize firms, such as region–district, sector of activity, logarithm of real sales and logarithm of number of workers in firm, will also be added. To better understand the returns to education, the specification will be extended by including interaction terms, *gender* × *education*, *district* × *education*, *industry* × *education* and *year* × *education*. These interactions are included to capture gender, location and industry differences in returns to education. By interacting education with a time trend, or time dummies, allows us to test if there are significant long–run differences in the returns to schooling over time. Finally, in order to capture the heterogeneity of the firms in terms of productivity, the firm fixed–effects will also be considered in the specification.

OLS estimates of returns to education β are not consistent, either because the error term in equation (3.29) includes unobservable factors, such as, individual ability that are correlated with schooling, or because of the existence of measurement errors in the schooling variable (Card, 1995b, 1999, 2001).

Following data features, namely available information that could allow us to estimate causal effects from education to wages, we will follow the IV solution proposed in the literature. Identification is achieved using three different instruments: the season of birth, as in Angrist and Krueger (1991), changes in the schooling system introduced in 1964 and in 1986, Law no. 45810/64 and Law no. 46/86, respectively, as in Harmon and Walker (1995). Finally, the average education by region in the year of worker's entry into school is proposed as a new instrument.

As discussed above, to test if schooling is endogenous, the Durbin–Wu–Hausman test is applied. Furthermore, we will assess the validity of the IV estimates by evaluating the quality of the selected instruments, and if they are uncorrelated with the unobserved error term. The orthogonality condition applied to the instruments will be

tested using, when feasible, the over–identifying restrictions test.

Despite the fact that the consistency of the IV coefficient estimates is not affected by the presence of heteroscedasticity, the standard IV estimates of the standard errors are inconsistent, even in large samples, preventing valid inference. The usual forms of the diagnose tests for endogeneity and over–identifying restrictions will also be invalid if heteroscedasticity is present. In order to test the validity of homoscedasticity assumption, it will be applied the Pagan–Hall test.[9] A rejection of the null hypothesis implies that one should use Generalized Method of Moments (GMM) estimator. The GMM method produces consistent and efficient estimates in the presence of arbitrary heteroscedasticity. The method is efficient when one has more instruments than parameters to be estimated (Baum et al., 2003, 2007).

3.4.2 Data description

The data used in this essay was constructed from *Quadros de Pessoal* (Personal Records) collected by the Ministry of Labour, Solidarity and Social Security. *Quadros de Pessoal* (QP) is an annual database, obtained since 1982, by mandatory surveys applied to all establishments with at least one employee. Public administration workers, armed forces, self-employed, as well as household workers, are not included in this database. Due to data limitations, namely representativeness and quality, the current analysis excluded the first four years, and covers the period from 1986 to 2009. Additionally, the years 1990 and 2001 are also not included as data for workers was not made available.

The existing information comprises three levels: company, establishment and worker. For each company, and year of observation, the location, the economic activity, the sales amount, the legal nature, the capital ownership (national, foreign or public), and its the size, measured by the number of workers, is known. At the establishment level the data provided information about the location and economic activity. The data has detailed information about each worker, namely, gender, age, date of birth, schooling, date of admission in the company, occupation, category and employment status, monthly base salary, overtime compensation, number of monthly hours worked, among others. The database contains an anonymized unique identification number for each worker and each company, which allows one to follow them over time. Furthermore, the data allows a match between the company, the establishment and each employee.

Although the quality of the data is very high, still a set of checks and corrections had to be performed in order to assemble a consistent panel of workers and their firms.

[9]The test of Pagan and Hall (1983) is designed specifically for detecting the presence of heteroscedasticity in IV estimation without restrictions. Under the null hypothesis of homoscedasticity in the IV regression, the Pagan–Hall statistic is distributed as chi–square with degrees of freedom equal to the sum of the number of instruments included and excluded from that equation.

A description and discussion of the data assembling process is provided in Appendix A.1.

The sample used in the following analysis has full time employees, aged between 16 and 64 years old. Workers whose wage is below the national minimum wage were excluded.[10] The sample used in the current analysis comprises almost 34 million observations worker–year on 5,867,131 unique workers, and 666,343 unique firms. The data set contains 1.5 million workers on average each year and from 2000, there are more than 200 thousand firms in each year. Descriptive statistics for workers are presented in Table 3.1.[11]

Between 1986 and 2009, the Portuguese labour market registered substantial changes. First, women's participation rate has grown considerably. In our sample, while in 1986 about two thirds of the workers were male, in 2009 the female share in the labour force increased to 44%, i.e. an increase of approximately 11 percentage point (p.p.), or 35% (see Table A.2).[12]

Second, real wages have registered a boost. Within this context, one should highlight that, over the entire sample, men's average real hourly wage was 5.7 Euro, while women's was 4.4 Euro, 23% less (see Table 3.1). There are substantial variations along the wage distribution for both genders. 5% of women earn an hourly wage up to 2 Euros, while at the top of the distribution 5% of women earn per hour above 11 Euros. Similarly, in the tails of the wage distribution 5% of men earn a wage per hour lower than 2.25 Euros, while 5% of men earn more than 14.5 Euros. Although some of the gap may be explained by the initial low participation rate of women, or because of differences across occupations (for example, leadership positions are more common for men), which is another form of differentiating men from women, the available evidence indicates that women's discrimination in the labour market cannot be rejected.[13] Wage dispersion is considerably high, particularly for men – the standard deviation of wages is about the same as the average, compared to about 0.83 for females. The wage dispersion has increased substantially between 1986 and 2009, despite its substantial real increase (Table A.2).

Third, considering education, about 80% of the men and 71% of women have only up to 9 years of schooling (3^{rd} cycle), and about 7% of the men and 10% of women

[10] According to Article no. 275 of the Portuguese Labour Code, employers can pay a wage below the national minimum wage for apprentices, practitioners, trainees or participating in certified training, as well as for workers with reduced work capacity.

[11] See Table A.1, in Appendix A.2, for a description of the variables. Detailed descriptives for workers in 1986 and 2009 are available in Table A.2.

[12] For the whole labour market female share was 42% in 1986 and 48% in 2009 (National Statistics Institute – INE).

[13] For a discussion of gender discrimination in the labour market, see, for instance, Oaxaca, 1973; Neumark, 1988; Portela et al., 2007a.

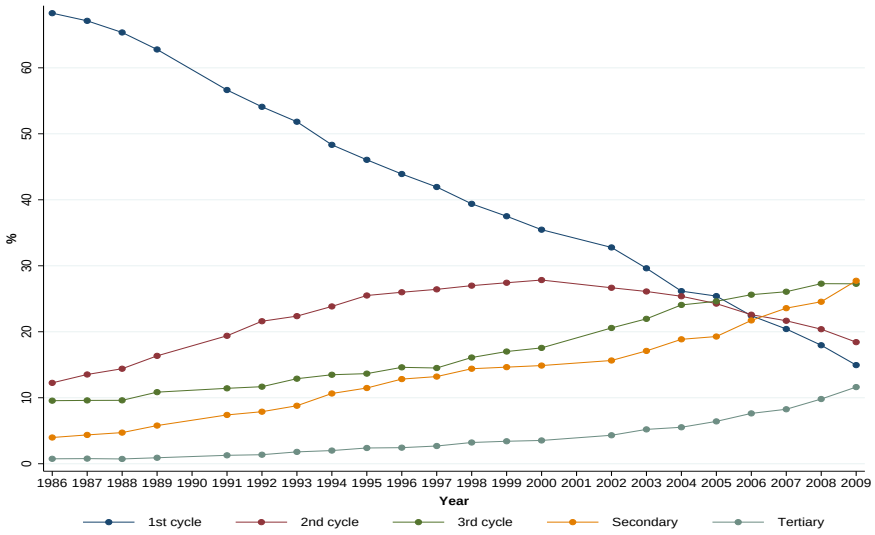
Table 3.1: Descriptive statistics on workers

Variable	Obs.	Mean	Std. Dev.	Min	Max
Men					
Real hourly wage (Euro)	20,224,043	5.660	5.666	0.319	1390
Age	20,224,043	38.143	11.345	16	64
Experience	20,221,307	21.471	11.553	0	48
Tenure	20,224,043	8.857	8.926	0.083	53.667
Education	20,224,043	7.113	3.748	0	16
≤ 1st cycle		0.417			
2nd cycle		0.205			
3rd cycle		0.174			
Secondary		0.134			
Tertiary		0.071			
Occupation	18,995,741				
Executive/managerial		0.054			
Intermediate managerial/executive		0.041			
Low managerial		0.108			
Technicians highly specialized		0.115			
Sales, administ. and precision prod.		0.093			
Administ. support and production		0.010			
Unskilled		0.320			
Apprentice		0.156			
Unknown		0.103			
Women					
Real hourly wage (Euro)	13,670,833	4.411	3.663	0.319	452
Age	13,670,833	35.738	10.469	16	64
Experience	13,667,983	18.804	10.816	0	48
Tenure	13,670,833	7.759	7.949	0.083	52.583
Education	13,670,833	7.868	3.981	0	16
≤ 1st cycle		0.339			
2nd cycle		0.203			
3rd cycle		0.171			
Secondary		0.190			
Tertiary		0.097			
Occupation	12,826,666				
Executive/managerial		0.025			
Intermediate managerial/executive		0.05			
Low managerial		0.084			
Technicians highly specialized		0.228			
Sales, administ. and precision prod.		0.215			
Administ. support and production		0.008			
Unskilled		0.217			
Apprentice		0.045			
Unknown		0.128			

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.

has a university degree (Table 3.1). The significant increase in average education, as well as the share of workers with tertiary education, is a remarkable feature of the Portuguese labour following integration into the European Union. Table A.2 shows that average education in our sample went from 5.3 to 9.5 years; the share of workers

with secondary and tertiary education increased from 6.8% in 1986 to 41.4% in 2009. It is also noticeable that, the share of workers with higher education passed from 2.3% in 1986 to 16% in 2009. Figure 3.1 highligths the low educational qualifications of the Portuguese labour force. Until 2004 more than 50% of the workers only had 4 or 6 years of schooling. Note that, in 1986, almost 70% of this group had at most the 4^{th} grade of schooling. Only in 2003, this share is below the line of 30% (about 29%), but it is still the dominant schooling category. One can also notice a substantial upward trend in the 2^{nd} cycle of schooling, particularly, during the end of 1980s and first half of 1990s, and in secondary and tertiary levels since the 1990s. In the end of 2000, the dominant schooling categories are the 3^{rd} cycle and the secondary level. In spite of the improvement in the workers' schooling, Portugal is still one of the OECD countries with lower education (OECD, 2012).



Source: Created by author based on *Quadros de Pessoal*, 1986–2009.
Note: Average schooling attainment reported in the graph for the years 1990 and 2001 is interpolated.

Figure 3.1: Average Schooling Attainment of Portuguese Workers, 1986–2009

Fourth, the average age is higher for men than women, 38 and 36 years old, respectively and the mean seniority at the firm is nearly 9 for men and about 8 years for women. While average age has increased by about 3 years, average tenure decreased by 1.5 years. Decreasing tenure is one of the changing features of the Portuguese labour market.

Table 3.2 presents descriptives on the firms used in the analysis. While building the data no restrictions were imposed on firm size. From 1986 to 2009, the number of firms more than tripled and, in 2009, the entrepreneurial structure is composed by more than 250,000 firms, from which 99% are small or medium size. Average size

Table 3.2: Descriptive statistics on firms

Variable	Obs.	Mean	Std. Dev.	Min	Max
Size	667,376	13.297	115.249	1	30,288
Real sales (1000 Euro)	540,949	1,459,057	2.28e+07	2.441	6.34e+09
Region	667,376				
North		0.295			
Centre		0.161			
Lisbon and Tagus Valley		0.356			
Inland		0.137			
Algarve		0.051			

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.

decreased from 22 employees to 11 employees (its standard deviation has decreased by half).[14] The mean of real sales volume of firms in the sample is about 1.5 million Euro and firms employ, on average, 13 workers. Firms are concentrated mainly in the North and Lisbon and Tagus Valey regions, and almost 50% of the firms belong to "whole sale and retail trade" and "construction" (see Table 3.3).

Table 3.3: Statistics on sector of activity

Sector	Percent	Sector	Percent
2 – Mining	0.32%	12 – Whole sale and retail trade	31.77%
3 – Food, beverage, tobacco	1.85%	13 – Restaurants, hotels	12.71%
4 – Textiles, dressing, leather	5.11%	14 – Transport, communications	3.44%
5 – Wood, cork, paper	2.65%	15 – Post and telecommunications	0.09%
6 – Manuf. Non–metallic products	1.61%	16 – Financial intermediation	0.69%
7 – Manuf. metals products	3.90%	17 – Real estate, renting and business	9.76%
8 – Manuf. of furniture	1.73%	19 – Education	1.20%
10 – Electriciy, gas and water supply	0.05%	20 – Health and social work	3.08%
11 – Construction	14.91%	21 – Other services	5.13%

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: The number of observations is 640,084.

3.5 Empirical results

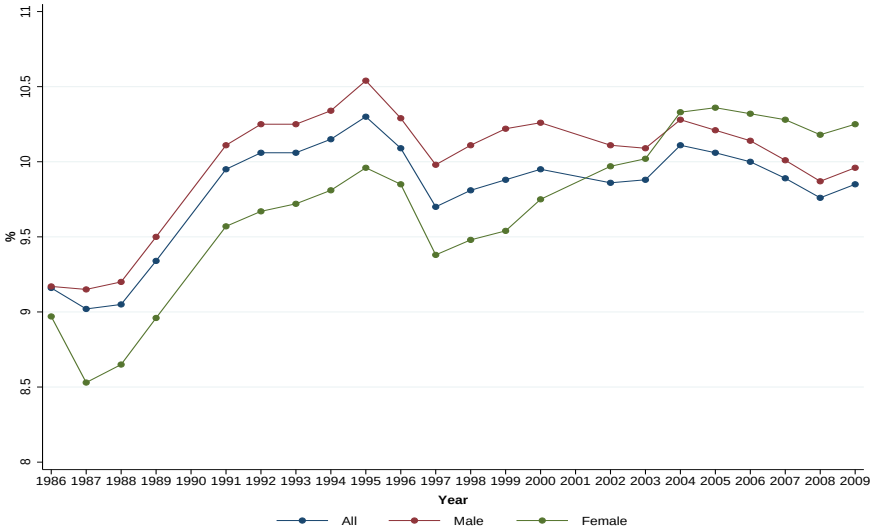
3.5.1 Ordinary least squares results

In a first stage, standard Mincerian earnings functions are estimated using OLS for all workers in the sample, for men and for women separately, and for each year between 1986 and 2009.[15] The estimated function is the one defined in equation (3.29). Figure

[14] See Table A.3 in the Appendix A.2.

[15] See Tables A.4–A.6, in Appendix A.3 for results of OLS estimations. These are not provided for 1990 and 2001, as worker data was not collected for these two years. See the discussion on the data above.

3.2 shows the corresponding point estimates.



Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.
Note: Returns to education reported in the graph for the years 1990 and 2001 are interpolated.

Figure 3.2: Estimated Returns to Education by OLS, 1986–2009

The results suggest that education has a significant and positive impact on wages. For the period 1986–2009, on average, the estimated return to an additional year of education is about 10%. Earnings increase with education, and this positive linkage is consistent with previous research on Portugal (Psacharopoulos, 1981; Vieira, 1999; Machado and Mata, 2001; Pereira and Martins, 2002a, 2004b). The estimates are also in line with the ones published for other countries (for instance, Mincer, 1974; Card, 1995a,b; Harmon et al., 2003; Devereux and Fan, 2011; Grenet, 2013; Wang, 2013). Moreover, it is found that the returns to education show an increasing trend in the period from 1986 to 1995, they pass from 9%, in 1986, to 10.3%, in 1995 (its peak for the period under analysis), as in Pereira and Martins (2002a, 2004b). In 1996 and 1997 there is a slight decrease in the size of the return, and from 1998 until 2004, it can be observed again a slightly increasing trend. Then, until the end of the period under study, the returns to education remain practically constant and close to 10%.

Differences in returns across gender are also considered. The average return to education for men, over the entire sample period, is 10.1%, while for women it is 9.8%. Between 1986 and 2000, the trend for both is practically the same, but the returns to an extra year of schooling for men are slightly higher than for women. This result is in line with other studies (for instance, Lima and Pereira, 1999; Harmon et al., 2001; Pereira and Martins, 2002a). From 2000 to 2003 the trend of the returns to education

for both genders is opposite.[16] While the returns for men decreased slightly, for women continued to increase until 2005, reaching 10.4%. For women, returns have noticeably increased from 8.5%, in 1987, to 10.3%, in 2009, an increase of 1.8 p.p. (or 21%); for men it went from 9.2% to 9.9%, a variation of 0.7 p.p. (or 7.6%).[17] From 2004 the returns are higher for women than for men, which is in line with some of the literature (Psacharopoulos, 1994; Psacharopoulos and Patrinos, 2004). In 2009 the difference in returns is about 0.3 p.p., favourable to women.

The results previously obtained assume a constant return rate to education. Now, dropping this assumption, education is disaggregated by schooling levels: 1^{st} cycle (4 years); 2^{nd} cycle (6 years); 3^{rd} cycle (9 years); secondary (12 years); tertiary (15 years corresponding to "*bacharelato*") and tertiary (greater or equal to 16 years corresponding to "*licenciatura*"). Figure 3.3 reports the OLS estimates of returns to specific educational levels from 1986 to 2009, in which the base educational level is the 2^{nd} cycle (6 years).[18] So, the estimated coefficients of the educational dummies should be interpreted as differences to the baseline return accruing to individual with the 2^{nd} cycle of schooling.

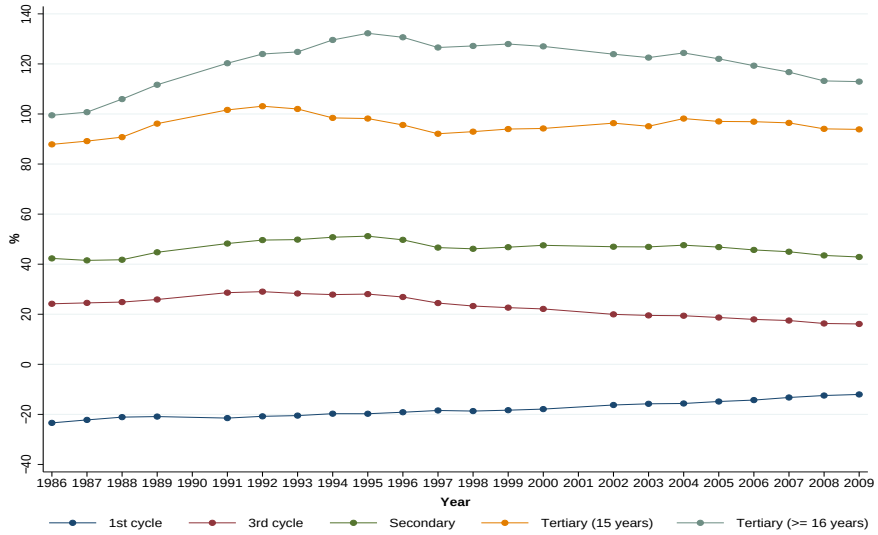
As expected, it can be observed that the higher the schooling level the greater wage premium of the individual. The individuals with tertiary education ("*licenciatura*") earn about 120% more than the individuals with only the 2^{nd} cycle (base–schooling).[19] Workers with just the 1^{st} cycle of education earn about 20% less when compared with individuals that have the 2^{nd} cycle. The results also show that higher educational qualifications yield higher returns for both men and women. There are no significant differences between males and females in the lower levels, but these differences are

[16] Keep in mind that we do not have data for 2001.

[17] Figure 3.2 reveals that, starting in 2002, both male and female average returns to education are above overall average returns within each period. Although such result might seem counter-intuitive, it can occur if: (1) returns to education are heterogeneous for different levels of education, (2) if such pattern is different between males and females and, (3), if the distribution of education is different by gender. By estimating average returns to education using equation (3.29) one imposes a constant return for all levels of education, and the same return for males and females. Once we allow for different slopes by gender, estimating equation (3.29) for males and females separately, one still gets average returns, which now may differ between the two groups, and be different from the overall average return. If returns to education are constant within gender groups, its average return does not depend upon education distribution, and the overall average return will, indeed, be a weighted average of average returns for males and females. However, if returns are heterogeneous over the education distribution, the gender specific education distribution becomes relevant for the comparison of average returns. This implies that average returns will be conditional on male/female's distribution of education and on how returns to education vary over that distribution. So, the combination of varying returns to education, both across levels of education, and by gender, may, thus, generate average returns for males and females which are above average overall returns.

[18] Once again we estimate equation (3.29), but now education enters the regression as dummies for each level, where the base category is the 2^{nd} cycle.

[19] The true return to education estimated is given by $\exp(\beta) - 1$ so in this case the return is 232%.



Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.
Note: Returns to education reported in the graph in the years 1990 and 2001 are interpolated. Returns to education reported in this graph are not exact. The 'precise' return to education is given by $\exp(\beta)-1$. However, as the transformation $\exp(\beta)-1$ is monotonic, the only change the correction implies is the shifting of the lines, maintaining the trends.

Figure 3.3: Returns to different schooling levels by OLS, 1986–2009

significant in the higher levels.[20] Rates of return to schooling for men ranged between 16% (3^{rd} cycle) and 126.4% (tertiary – "*licenciatura*"); corresponding values for women are 17.5% and 135.3%.

Over the period 1986–2009, the trend is practically constant in all education levels, except in tertiary ("*licenciatura*") where there is a remarkably increasing trend from the beginning of the period until 1995 (Pereira and Martins, 2002a, also suggest that this peak is reached in 1995), but this trend is slightly decreasing thereafter. Thus, the increase in wages caused by return of an extra year of education is higher in 1995 compared to 2009. For example, in 1995, the workers who complete the secondary or tertiary ("*licenciatura*") levels have an increase in their wage, respectively, of around 51.2% and 131.7%, while in 2009 the corresponding values are about 42.9% and 113%, respectively. According to the literature, this reduction in the amount of the wage premium of people with higher education in 2009 compared with 1995, might have been caused by the significant expansion of tertiary education.

Estimated coefficients for experience and its square have the expected sign and significance, over all regressions and periods, i.e., positive and negative, respectively, confirming the theory predictions of decreasing marginal returns to experience. Taking the results in 2009, see Table A.4, one observes that $\widehat{\gamma}_1 = 0.0345$ and $\widehat{\gamma}_2 = -0.0003$. These estimates indicate that the return to the first year of experience is about 3.5%;

[20] See Tables A.4–A.6, in Appendix A.3, for results of OLS estimations.

after 20 years of experience, approximately average experience in the sample, the return to an additional year of experience is still around 2.3%.[21] Corresponding values for 1986 are 5.3% and 2.1%, respectively; in 1995 estimates indicate initial returns to experience of about 4.3%, while after 20 years of experience they have decreased to 2.3%. The discussion on returns to experience is just indicative of positive and decreasing marginal returns, as returns to experience and tenure cannot be disentangled under the current formulation of the model (see, for example, Altonji and Shakotko, 1987; Topel, 1991; Buhai et al., 2014).

In order to test the robustness of the estimated results, and to analyse if the inclusion of other covariates changes the estimated coefficients of education, alternative specifications are experimented, i.e. additional controls for individuals' and firm's characteristics are included. The estimations of each specification are made for 1986, 1992, 1998, 2004, 2009 (dividing the survey period into equal parts) and for the whole period 1986–2009. The results are reported in Table 3.4. The results have the expected sign and are statistically significant at 1%. The inclusion of the additional controls improves the explanatory power of the model, as can be seen from the comparison of the R–squared and the RMSE between Table 3.4 and Table A.4 in the Appendix. After including the additional controls, returns to education vary between 4.2% and 7.2%. These returns are smaller than the ones discussed above for the simple specification of the Mincer equation. This decrease is expected following the inclusion of variables that control for tenure and firm size. The result is in accordance with the literature; e.g., including control variables related to employment, firm size, as well as industry, several authors found a substantial decrease in returns to schooling (Lima and Pereira, 1999; Pereira and Martins, 2002a, 2004b).

Within each year, the models that include firm size, either through sales, or employment, second and third columns of a given year in Table 3.4, show estimated coefficients that are close in magnitude, but smaller than the ones estimated without considering this information.[22] The result seems to indicate that, to some extent, larger firms tend to recruit workers with higher levels of education. The estimation of a simple specification, taking education as the dependent variable, and using gender, age, and a set of controls for occupation, industry and location, does show a positive correlation between education and firm size.[23]

[21] The computation is given by $0.0225 = 0.0345 - 0.0003 \times 2 \times 20$.

[22] An additional specification has been estimated where both sales and size are included simultaneously; it would correspond to a "crude" measure of productivity, as it would consider as an explanatory variable sales per worker. The estimated returns to investments in education do not change substantially, so we made the option presented in Table 3.4.

[23] See the discussion in Section 3.5.2 on how firm specificities, not controlled in typical wage regressions, may affect wages, and bias the estimate of returns to education. The argument put forward is that unobserved firm "productivity" may explain both wages and education, which could explain why

Table 3.4: Returns to education – OLS

VARIABLES	**1986**			**1992**			**1998**		
Education	0.0643***	0.0477***	0.0496***	0.0716***	0.0530***	0.0549***	0.0624***	0.0419***	0.0438***
	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Female	-0.2068***	-0.1418***	-0.1535***	-0.2441***	-0.1792***	-0.1898***	-0.2275***	-0.1676***	-0.1728***
	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0008)	(0.0007)	(0.0006)	(0.0006)	(0.0006)
Experience	0.0290***	0.0212***	0.0220***	0.0253***	0.0206***	0.0208***	0.0239***	0.0204***	0.0208***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Experience sq	-0.0005***	-0.0003***	-0.0004***	-0.0004***	-0.0003***	–0.0003***	-0.0003***	-0.0003***	-0.0003***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Tenure	0.0173***	0.0067***	0.0076***	0.0163***	0.0067***	0.0064***	0.0178***	0.0104***	0.0108***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Tenure sq	-0.0002***	-0.0001***	-0.0001***	-0.0002***	-0.0001***	-0.0001***	-0.0002***	-0.0002***	-0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Log sales		0.0456***			0.0443***			0.0492***	
		(0.0001)			(0.0002)			(0.0001)	
Log size			0.0468***			0.0506***			0.0517***
			(0.0002)			(0.0002)			(0.0001)
Observations	927,245	871,373	912,636	1,169,586	1,072,054	1,151,789	1,474,374	1,291,030	1,447,142
R–squared	0.500	0.653	0.638	0.513	0.615	0.609	0.572	0.661	0.670
RMSE	0.319	0.263	0.270	0.364	0.323	0.327	0.357	0.310	0.314

Continued on next page

Table 3.4 – continued from previous page

VARIABLES	2004			2009			1986–2009		
Education	0.0670***	0.0464***	0.0496***	0.0641***	0.0472***	0.0498***	0.0664***	0.0476***	0.0502***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
Female	-0.2157***	-0.1633***	-0.1702***	-0.1972***	-0.1474***	-0.1519***	-0.2208***	-0.1634***	-0.1705***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0001)	(0.0001)	(0.0001)
Experience	0.0243***	0.0205***	0.0209***	0.0215***	0.0204***	0.0203***	0.0241***	0.0203***	0.0205***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
Experience sq	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Tenure	0.0222***	0.0141***	0.0160***	0.0228***	0.0163***	0.0177***	0.0195***	0.0111***	0.0123***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
Tenure sq	-0.0003***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Log sales		0.0545***			0.0483***			0.0488***	
		(0.0001)			(0.0001)			(0.0000)	
Log size			0.0522***			0.0427***			0.0479***
			(0.0001)			(0.0001)			(0.0000)
Observations	1,720,727	1,529,398	1,684,492	1,883,554	1,631,409	1,840,372	31,817,613	28,296,787	31,167,549
R–squared	0.553	0.645	0.653	0.534	0.607	0.626	0.563	0.648	0.652
RMSE	0.384	0.330	0.337	0.387	0.341	0.347	0.371	0.325	0.331

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.

Note: Standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include a set of dummies for occupation, industry and district/region. The specification for 1986–2009 includes a set of year dummy variables.

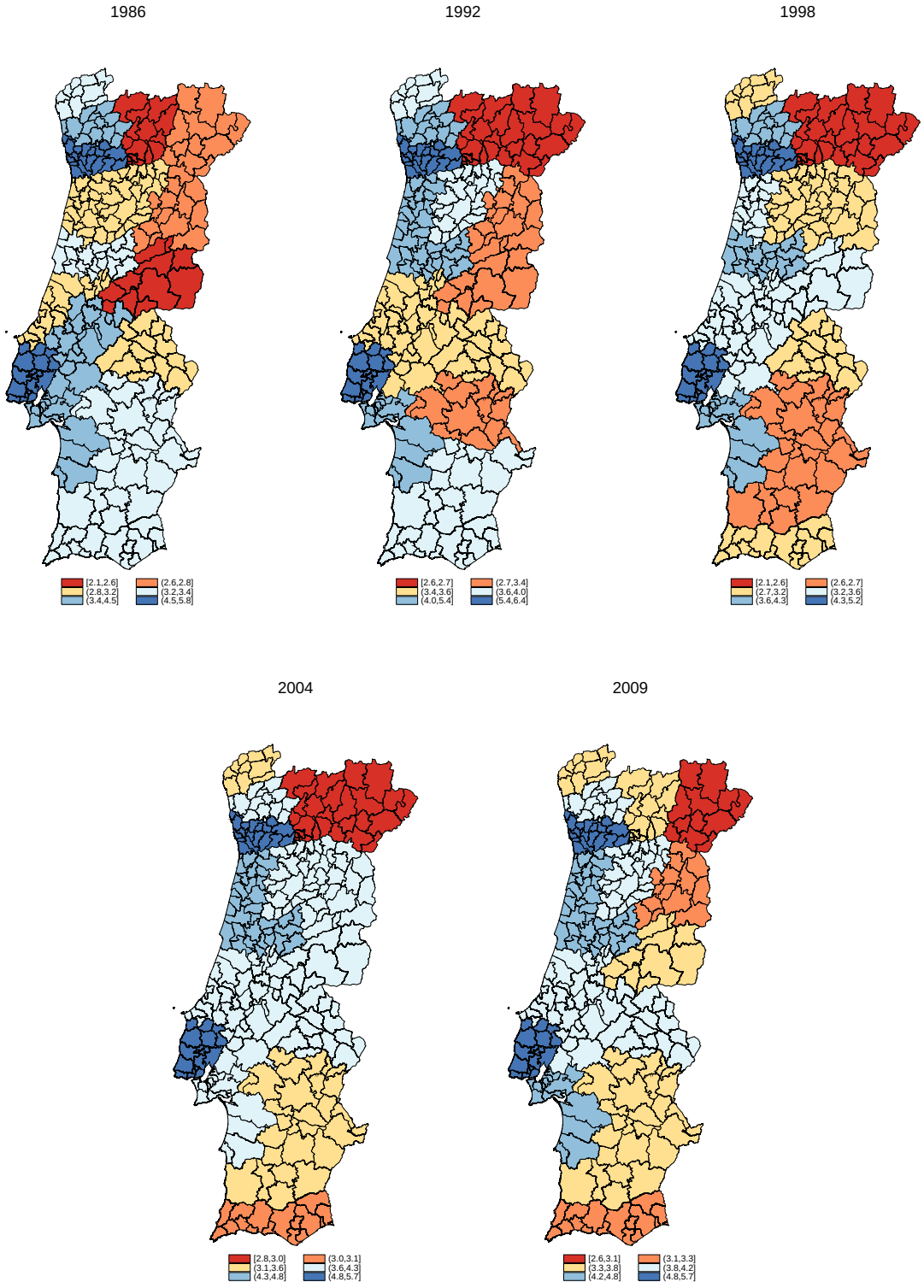
Regarding the remaining regressors, on average, women earn less than men between 14% and 24%. The logarithm of the number of workers, *Logsize*, and the logarithm of real sales, *Logsales*, show significant and positive effects on wages. On average, an increase of 1% in firm size leads to a wage increase bounded between 4.3% and 5.2%; firms with a higher number of workers pay higher wages, as in Machado and Mata (2001). The results also indicate that an increase of 1% in real sales leads to an increase between 4.4% and 5.5% in wages. This last result is consistent with the literature on rent sharing,[24] which presents evidence of strong positive correlations between different measures of profitability and wages and refer the importance of the rent sharing as a component of the functioning of the labour markets (Martins, 2009b; Card et al., 2014).

Figure 3.4 presents the private returns by district, based on one of the specification discussed above, the one presented for each year in the second column, Table 3.4. The estimates of the coefficients of district dummies are jointly statistically significant; most of the 17 estimated coefficients for district dummies are also individually, economically and statistically, significant. This results occurs throughout the different years of the sample. This result highlights the wage heterogeneity across regions.

In the maps presented in Figure 3.4, the distribution of returns to education is divided into six categories, through the percentiles 10, 25, 50, 75 and 90, from which the respective legends are formed in each year. It can be seen that education is more valued in the coastal districts and less valued in the interior districts of Portugal. The returns to education in the districts of Oporto and Lisbon are higher than in the other districts (as in, Lima and Pereira, 1999; Pereira and Martins, 2004b; Vieira et al., 2006), following the districts of *Aveiro*, *Braga*, *Coimbra* and *Setúbal*. It is also observed that the districts where the returns to education are lower are *Bragança* and *Vila Real*. These results are stable over time. However, some of the districts, namely, *Castelo Branco* and *Viseu*, reveal a growing value for education. This may be the result, among other key factors, of industrial changes, as well as the installation of universities and polytechnics in these regions.

in the current results estimates on returns to education decrease once we include additional firm variables, i.e., this variables may act as proxies for those unobserved characteristics and, thus, attenuate the possible bias in the estimated return to education.

[24] "Rent sharing" occurs when rents, defined in the literature as profits above the level that would result from paying all production factors their competitive market rates, are shared by the firm with its employees (Martins, 2007).



Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.

Figure 3.4: Returns to education by district/region

3.5.2 Firm specific effects

The results discussed in Section 3.5.1 may be skewed due to unobservable characteristics both of workers and their firms. Here we will discuss the bias in the estimate of returns to education that my occur following unobserved heterogeneity on the firm side. Endogeneity due to worker features will be discussed in Section 3.5.3.

So far we have not addressed firm specificities like management quality. Management models adopted by the firm, namely, regarding the organization of its activities, the working conditions and the methods of recruiting workers, may result in some firms being more efficient and productive, and pay higher wages. Such specificities generally promote the evaluation of features with impacts on labour productivity that are not captured by the explanatory variables used (Martins, 2004). Thus, it is relevant to assume that there are some firm's unobserved attribute differentials captured in the error term of the regression that may be correlated with the variables related with the firms, particularly education.

One of the possible ways to tackle this source of endogeneity is to include a firm–specific effect in the wage regression. These terms are bound to capture the heterogeneity across firms, which includes, for example, unobserved differences in productivity. Taking the specifications defined in Section 3.5.1 we now include firm specific effects. Additionally, in some of the specifications interaction terms were included: $gender \times education$, $year \times education$, $district \times education$ and $industry \times education$. These interactions are used to capture gender differences in returns over different education levels, the differences in returns to education over the district and industry; interacting education with a time trend allows us to test if there are significant differences in the returns to schooling over time. Firm specific effects are controlled for by absorbing the categorical variable that defines firms' id, i.e., it is equivalent to include in the OLS regression a set of dummies for the firms included in the sample.

The results of these estimations are reported in Table A.7 and Figures A.1–A.3 in Appendix A.3. First, firm specific effects are jointly significant, which highlights its relevance for the wage determination. Second, estimates for key parameters are statistically significant and show the expected sign.

One expects that controlling for firms' unobservable characteristics should isolate more effectively the influence of the different explanatory variables. Looking to the first estimates, column (1) in Table A.7, average return to education is now estimated to be about 3.9%, which is significantly lower than the estimates reported in Table 3.4. One plausible interpretation is that high wage firms are recruiting workers with higher education. As such, the results above on the return to education are capturing not only the true effect of education on wages, but also firms' recruiting policies.

The results also indicate that this is a labour market where gender wage differentials are substantial and still the norm. Although one does not address explicitly gender discrimination, the consistency of the results across different model specifications, methodologies and time periods, seems to indicate that it occurs in the Portuguese labour market. Looking to column (2) in Table A.7, men earn more than women with similar levels of educational attainment. The largest wage differences occur in lower educational levels. Female workers with the 1^{st} cycle and with 2^{nd} cycle of education earn less about 17.6% and 16.6%, respectively, than male workers with same education levels. For higher education levels this difference is attenuated. For instance, female workers with the 3^{rd} cycle earn less about 12.8% than male workers with the same education level. Females' returns to the 3^{rd} cycle of education are larger by 3.7 p.p. when compared with returns for males in the same educational level; females benefit 9% when compared with women with just the 2^{nd} cycle, while men benefit 5.4% to males with the 2^{nd} cycle.

From column (3) in Table A.7, assuming education as a continuous variable, one concludes that, on average, female's returns to education are higher 0.5 p.p. when compared to male's returns to an additional year of education. Overall, these results are consistent with previous Portuguese literature (Lima and Pereira, 1999; Machado and Mata, 2001, 2005; González et al., 2007; Portela et al., 2007a; Budria and Pereira, 2011) and with international literature (for instance, Li et al., 2012 and Rodríguez-Pose and Tselios, 2012).

The left graph in Figure A.1 reveals a variation over time of the relative returns to different levels of education. One can read the graph in the following way. The 2^{nd} cycle is the base category for education. The estimates for the parameters associated with each level of education reported in column (4) in Table A.7 show conditional average wage difference between each level of education and the 2^{nd} cycle in 1986. For example, in the first year of observation, workers without formal education earn less 20% when compared with those with the 2^{nd} cycle. Figure A.1 shows how those differences evolved over time. Looking to the orange line one concludes that returns to the 3^{rd} cycle, when compared with 2^{nd} cycle of formal education, decreased gradually over time. By 2009 the returns to the 3^{rd} cycle were still substantial, almost 6%, but decreased by about 3 p.p., when compared with the 8.6% observed in 1996. The graph highlights that the benefits accrued to both the 3^{rd} cycle and secondary education have decreased relative to the returns in the 2^{nd} cycle, while the advantage of the 2^{nd} cycle over the 1^{st} cycle, 2^{nd} grade and no formal education has decreased. At the same time, *bacharelato* and *licenciatura* have increased their returns (the corresponding lines in Figure A.1 stay positive over the period). This shows that the returns to education

up to the secondary were compressed over the sample period, while returns to higher levels of education have expanded. This outcome follows, one can say naturally, from the extreme low levels of education in the 1980's and the expansion that occurred since.

It is in the higher educational levels the main changes took place. The advantage of tertiary education occurred in mid–1990's, having decreased since, while keeping substantial positive returns. As reported in Figure A.1, there is an increase in the returns to education of about 17.4 p.p and 12.4 p.p. in tertiary education, *bacharelato* and *licenciatura*, respectively, compared with the base year (1986) and the base educational level (2^{nd} cycle). In 1995 the return to education for workers with Tertiary (*licenciatura*) level is about 61% (45% for *bacharelato*). After 1995 the trend reverses. By 2009 the return to *licenciatura* was about 50%, while for *bacharelato* the return was 41%. Looking to the graph on the right in Figure A.1, corresponding to column (5) in Table A.7, one concludes that average returns to education have increased between 1986 and 1995 by 0.6 p.p., but decreased, reaching 3.9% in 2009. While the graph on the right seems to show a very irregular path, one should consider the scale of the graph, i.e., the extremes are -0.1 p.p. and 0.6 p.p, with an average variation of about 0.3 p.p. around the 3.6% return in 1986.

Comparing the return rate across mainland Portuguese districts, taking Aveiro as the base district, and sorting the districts in ascending order of returns to Tertiary (*licenciatura*), higher education is more valued in the districts of *Braga* (average return to *licenciatura* of about 61,1%, 2.3 p.p. above the return in *Aveiro* – 58.8%, column (6), Table A.7), *Castelo Branco*, 60.3%, and *Portalegre*, 60.2%. *Faro* is the district where *licenciatura* is less valued, 44.1% over the 2^{nd} cycle, *Leiria*, 45.5%, and *Bragança*, 46%. Under the specification formulated in column (7), Table A.7, *Aveiro* has an average return to education of about 4%, while returns in *Bragança* and *Faro* are lower by about 1.2 p.p. Compared to *Aveiro*, average returns to an additional year of education in *Braga*, *Porto* and *Lisbon* are higher by 0.25, 0.17 and 0.09 p.p., respectively. In *Braga* the return to education is about 4.2%.

Finally, regarding industry, Figure A.3, where "Mining" is the base category, one observes significant variation across the different industries. Looking to Table A.7, column (8), conditional on firms' specificities captured by the firm fixed–effect, one observes that in "Mining" workers with Tertiary (*licenciatura*) level earn more 57%, when compared with workers with the 2^{nd} cycle. The highest return to *licenciatura* is achieved in "Education", with a return of 83%, followed by "Textiles, dressing, leather" and "Food, beverage and tobacco", with returns of about 73% and 69%, respectively – see the light blue line in the left graph of Figure A.3. Overall, while average returns to education in "Mining" are about 3.4%, column (9), Table A.7, the higher average

return is achieved in "Education", 6.8%, 3.4 p.p. above the one reported for "Mining", and "Health and Social Work", 4.9%.

3.5.3 Instrumental variables results

As previously mentioned, if the residual in the wage equation, e.g. the one defined in equation (3.13), is correlated with years of education due to omitted variables, OLS estimates of return to education will be biased (Angrist and Krueger, 1991; Card, 1999; Wooldridge, 2010, 2006). So, to deal with endogeneity of education the IV approach is used and the variable education will be instrumented with three different instruments: (1) quarter of birth, (2) changes in compulsory schooling in 1964 and 1986, and (3) average education by region in the year of entry into the school.

In order to avoid issues underlying modelling female labour market participation, one restricted the sample to men. Such decision follows from two main aspects. First, the nature of the data prevents any correction for potential biases arising from decisions relating to participation in the labour market (Vieira, 1999). Second, mens' participation in the labour market is more stable over time and, consequently, the group of men is a more homogeneous cohort.

3.5.3.1 Quarter of birth

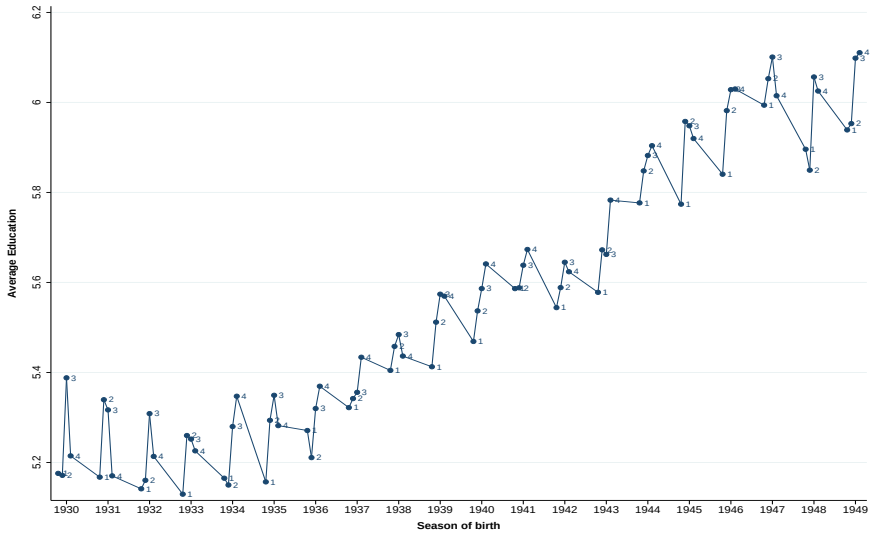
A first instrument we will use is the quarter of birth of the individual, which was first implemented by Angrist and Krueger (1991). The authors showed that students' quarter of birth, combined with education laws, generates exogenous variation in the years of schooling. The authors documented significant associations between the quarter of birth and age at school entry, educational attainment and earnings, for cohorts of men born between 1930 and 1959.

In the United States, the educational law requires a student to enter school in the fall of the calendar year in which he or she turns 6 years old, i.e. students who are born in the first months of the calendar year are older when they enter school than children who are born in the last months of the year (Angrist and Krueger, 1991). This induces a relationship between quarter of birth and schooling attainment, because children who are born in the first months of the year are allowed to leave school after having completed fewer years of schooling than children who enter school being younger. If the quarter of birth is exogenously correlated with the years of schooling, then it may be a valid instrument for education (e.g., Bound et al., 1995, among others).

The Portuguese educational system similarly, which makes it suitable for the implementation of the same estimation strategy. Its key features are the following. First,

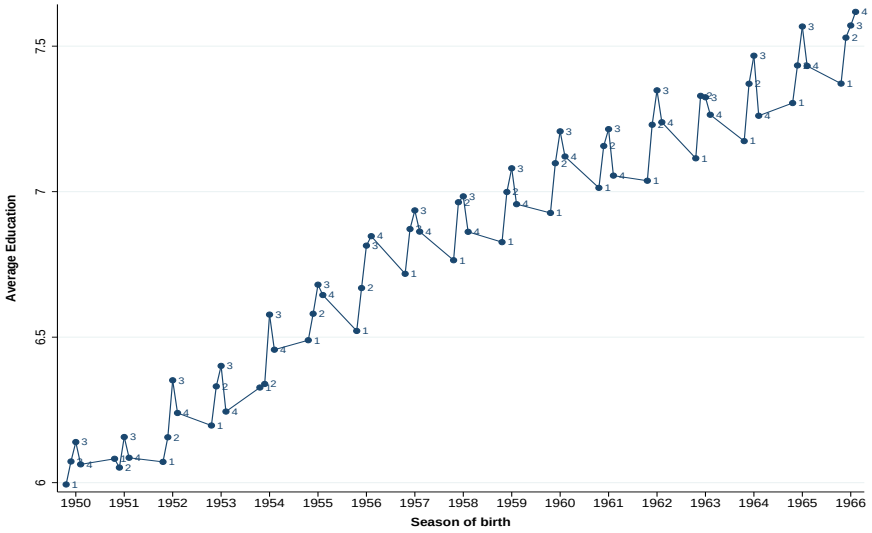
students born until 1979 started their first year of schooling in the autumn of the calendar year in which they turned seven years; thereafter, school age begins at age six. Second, if the student was born until 1966, he would have to attended school until he turned twelve years old; for those born in 1967 and afterwards, they would have to stay in school until becoming fourteen years old. Since 1991 the legal minimum working age is 16 years. Third, based on compulsory schooling laws, which relates the date of birth to his compulsory schooling, three cohorts can be defined: (1) those born before 1950 whose compulsory education is three years; (2) those born between 1950 and 1966 whose compulsory education is four years; and (3) those born between 1967 and 1980 whose schooling is compulsory six years.

Combining these conditions, it follows that an individual born at the beginning of the year, usually enters to school when he is close to 8 years old, completes the mandatory school age 12 (or 14) years old shortly after achieving the 4^{th} (or 6^{th}) grade, at which point he is allowed to leave school. This implies that individuals who do not want, or cannot, proceed with their studies, can abandon school before completing the 5^{th} (or 7^{th}) grade. However, an individual born in the third or fourth quarter, who started school in a month nearing completion of 7 years, completes the 4^{th} (or 6^{th}) grade one year before completing 12 (or 14) years old. This implies that he cannot leave school before completing an additional year of schooling. As such, these individuals begin, and probably complete, a new cycle of studies (2^{nd} or 3^{rd} cycle). Summing up, everything else equal, and at the margin, being born latter in the year induces one additional year of education.



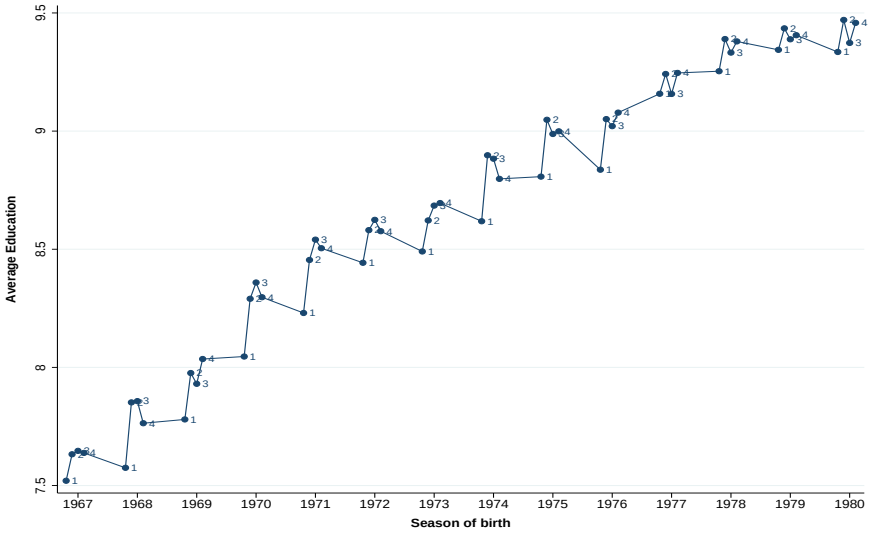
Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.
Note: Quarter of birth is listed next each observation.

Figure 3.5: Average Education by season of birth, Men born 1930–1949



Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.
Note: Quarter of birth is listed next each observation.

Figure 3.6: Average Education by season of birth, Men born 1950–1966



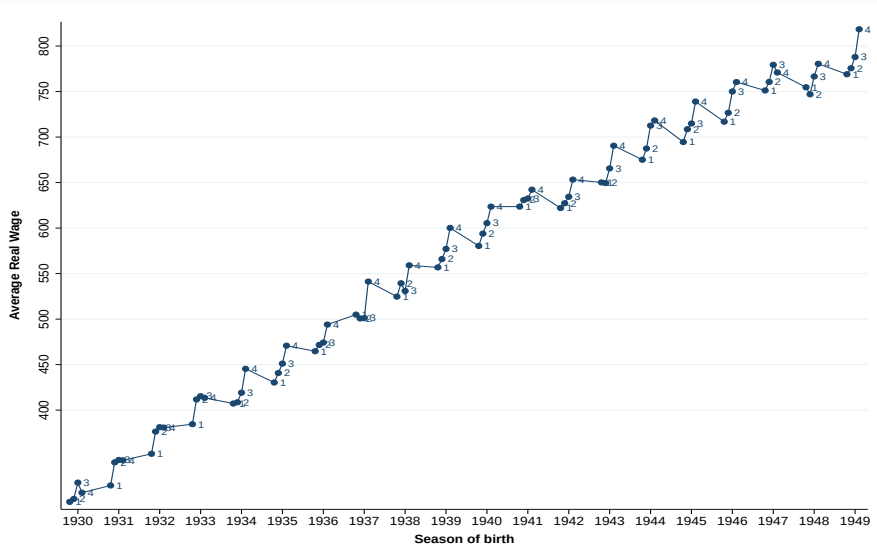
Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.
Note: Quarter of birth is listed next each observation.

Figure 3.7: Average Education by season of birth, Men born 1967–1980

Figures 3.5, 3.6 and 3.7, present the relationship between education and quarter of birth for Portuguese men born between 1930 and 1980, i.e., each graph depicts the average years of schooling by quarter and year of birth for the three cohorts. The Figures show a generally increasing trend in average education for all cohorts. A detailed observation of the graphs indicates that the average education is generally higher for individuals born in the third or fourth quarter, when compared with individuals born early in the same year. Furthermore, individuals born in the third or fourth quarter

of the year tend to have more education than those born in the first quarter of the following year. In addition, it can be verified that the average education increased to about the double between the beginning of the 1930s and the end of the 1970s; average education of the men included in the sample increased from about 5 years to around 9.4 years of education. The pace of the change in average education has increased for those born since 1957, which matches the year of birth of the first Portuguese men covered by the compulsory schooling of the 4^{th} grade. This is another piece of evidence that seems to show that compulsory schooling laws were effective in increasing the education.

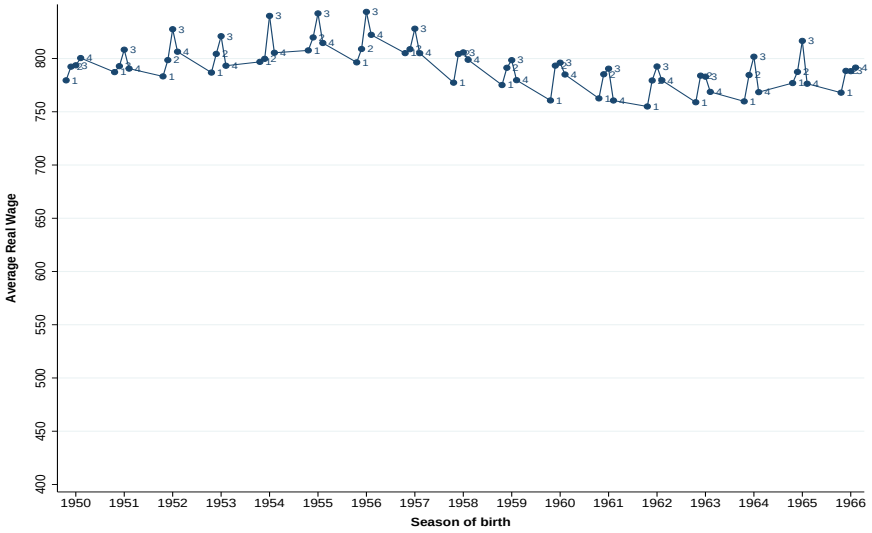
A preliminary check on how these differences in average education by quarter of birth are translated into differences in average wages *per* quarter of birth can be made by depicting the average real wage by quarter and year of birth of men born between 1930 and 1980 (see Figures 3.8, 3.9 and 3.10, respectively). This is a naive and oversimplified representation, whose sole propose is to have a first glimpse on to what extent workers born in the first quarter have lower average wages. Consistently, on average men born on the third or fourth quarter earn more than men born in the first two quarters of the year. This regularity fades over time, as the quarter of birth becomes less and less stringent on the education choices. This is particularly true for recent cohorts; Figure 3.10 shows that after mid–1970s there is no distinctive wages differences across cohorts within a given birth year.



Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.
Note: Quarter of birth is listed next to each observation.

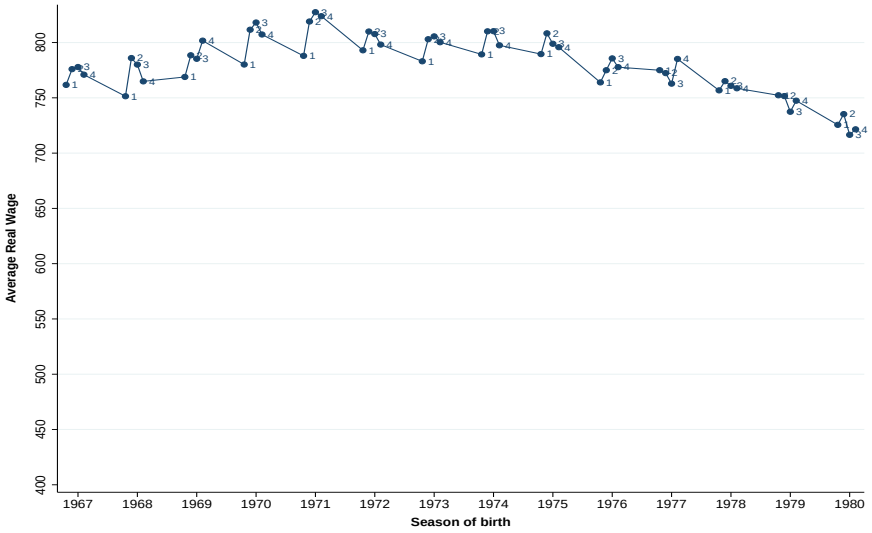
Figure 3.8: Average Wages by Season of Birth, Men born 1930–1949

We will now estimate returns to education using instrumental variables, where the specific instrument will be the quarter of birth. Estimates for both OLS and IV applied



Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.
Note: Quarter of birth is listed next to each observation.

Figure 3.9: Average Wages by Season of Birth, Men born 1950–1966



Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.
Note: Quarter of birth is listed next to each observation.

Figure 3.10: Average Wages by Season of Birth, Men born 1967–1980

to the men's sample will be compared in the following analysis. The first stage of the instrumental variables two stage least squares (IV2SLS) estimator is given by:

$$S_i = \rho X_i + \theta(q_{1,3})_i + v_i \tag{3.30}$$

where S_i is the level of schooling of men i; $q_{1,3}$ is a dummy variable which takes value equal to 1 if the individual was born in the first (or third) quarter, and zero otherwise;

X_i is a vector of covariates, including experience, experience squared, occupation, sector of activity, region and a set of dummies for the year of birth. Then, another instrument is created, which is an interaction between quarter of birth, $q_{1,3}$, and year of birth, i.e., $q_1 \times year\text{–}of\text{–}birth$ and $q_3 \times year\text{–}of\text{–}birth$, and the linear combination between q_1 and $q_1 \times year\text{–}of\text{–}birth$, as well between q_3 and $q_3 \times year\text{–}of\text{–}birth$ are used (following Angrist and Krueger, 1991; Leigh and Ryan, 2008). Thus, the second stage IV equation is given by:

$$\log W_i = \delta X_i + \beta \widehat{S}_i + \varepsilon_i \tag{3.31}$$

where the dependent variable is the logarithm of real hourly wage, $\widehat{S}_i$ is estimated education from equation (3.30) and the coefficient β is the return to education; X_i includes the regressors experience and its square, occupation, industry, region and a set of dummies for the year of birth.

In these IV estimations, only the cohorts whose number of observations is greater than 65,000 for each year of study are included. From the IV2SLS estimation, Table A.8 in Appendix A.4 presents the results from the first stage regression for the three cohorts, i.e. the estimates of the regression of schooling on first quarter of birth and on third quarter of birth. It can be observed that the first quarter of birth presents a positive and significant effect on schooling attainment, as in Angrist and Krueger (1991), but the sign of third quarter of birth is negative and significant at 1% level. The instruments used have coefficients significantly different from zero, which indicates that the instrument is probably strong. The fact that one of the instruments presents a positive sign and the other a negative sign results, to some extent, from the way one constructs the variable experience. Take, for instance, two men born in the same year, but one of them born in the first quarter and the other born in the third quarter. Keeping all the other variables in the model constant, when the two men have the same experience the one born later must necessarily have less education, as he certainly had to leave school earlier.

To check the quality and validity of the instruments, q_1 and q_3, it is taken into account the tests suggested by Bound et al. (1995). Table A.8 in Appendix A.4 shows that for all years in all men cohorts. The first stage R^2 ranges between 0.57 and 0.84. These values are quite higher than first stage R^2 regressions found in many other studies (for instance, Harmon and Walker, 1995; Leigh and Ryan, 2008; Dickson and Smith, 2011). Therefore, the results show that the instruments q_1 and q_3 are significantly related to endogenous schooling, which points out that the used instruments are relevant. Furthermore, for each year in all men cohorts the first stage F–statistic on the exclusion of instruments exceeds the rule–of–thumb value of 10 for non–weak instruments, and are, at the same time, statistically significant at the 1% significance

level. So, the high degree of statistical significance of the excluded instruments makes it unlikely that we face the so–called "weak instruments" problem (see Bound et al., 1995 and Staiger and Stock, 1997).

In order to test the validity of homoscedasticity assumption, as well as that schooling is actually endogenous, Pagan–Hall and Durbin–Wu–Hausman (DWH) tests, respectively, are applied. The results show that the failure of homoscedasticity of errors assumption is evident.[25] The highly significant values of the Pagan–Hall test statistic for heteroscedasticity reject the null that the error variance is constant. Regarding endogeneity of education, in the first cohort (men born 1930–1949), the null is not rejected, i.e. the high p–value suggests that there is no difference between IV2SLS and OLS estimates, hence, in this particular cases, the OLS should be preferred as being more efficient than the IV2SLS method. This result is expected, since the schooling of men born between 1930 and 1949 is not too different; most individuals have only the 4^{th} grade and, therefore, education is an exogenous variable. In the two remaining cohorts, the results suggest that the exogeneity of schooling is rejected by the DWH test in almost all of the years under study (except in 1987, 1989, 1992, 1994 and 1995 in second cohort using q_3 as instrument). Consequently, the OLS estimates are biased.

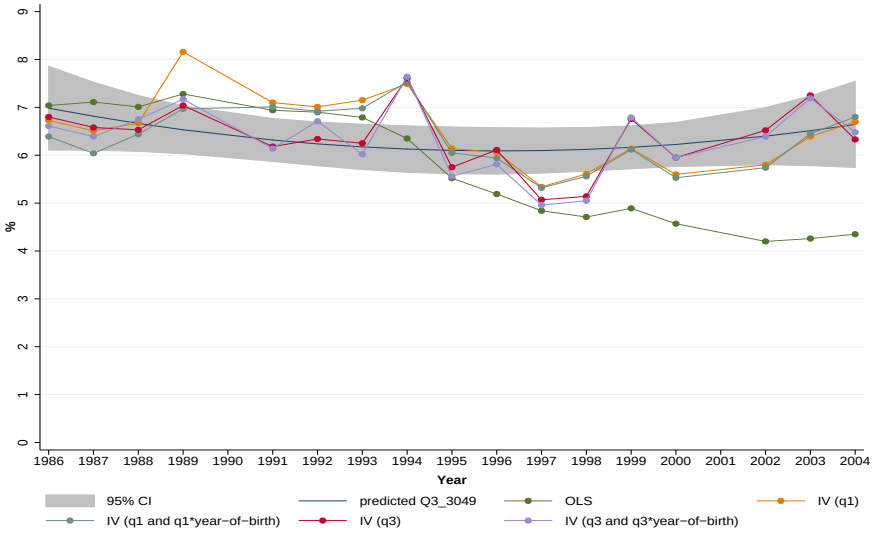
The results indicate that using either q_1 and $q_1 \times$ *year–of–birth*, or q_3 and $q_3 \times$ *year–of–birth* as instruments, the IV satisfies both relevance and exogeneity conditions for almost all years in all men cohorts. The Hansen J–test statistic indicates that one should not reject the null hypothesis that the instruments are valid, thus, it indicates that the instruments are appropriately uncorrelated with the disturbance process. The first stage R^2 is high, ranges between 0.28 and 0.62, and the F–statistic on the exclusion of the instruments is greater than 10 in almost every year in all men cohorts.[26] Thus, the Hansen J–test provides compelling statistical evidence for the validity of the q_1 and $q_1 \times$ *year–of–birth*, and q_3 and $q_3 \times$ *year–of–birth* instruments and the high R^2 and F statistics suggest that the inherent bias in IV estimation will be absent, or at least smaller, than the OLS bias.

By comparison, Figures 3.11, 3.12 and 3.13 present the results of the OLS and second stage of IV2SLS estimates of returns to education for each cohort. The graphs show the results for IV estimates when the selected instrument is the first or the third quarter of birth, q_1 or q_3, and also show the results of the returns to education when the interaction of both quarter of birth with year of birth are included as instrument. Note that, standard errors are corrected to take into account for heteroscedasticity related to the fact that many workers are present more than one period in the sample.

[25] On this particular set the detailed estimations and statistics are not reported to avoid overloading the thesis with the size of the appendices. Naturally, they will be provided when requested.

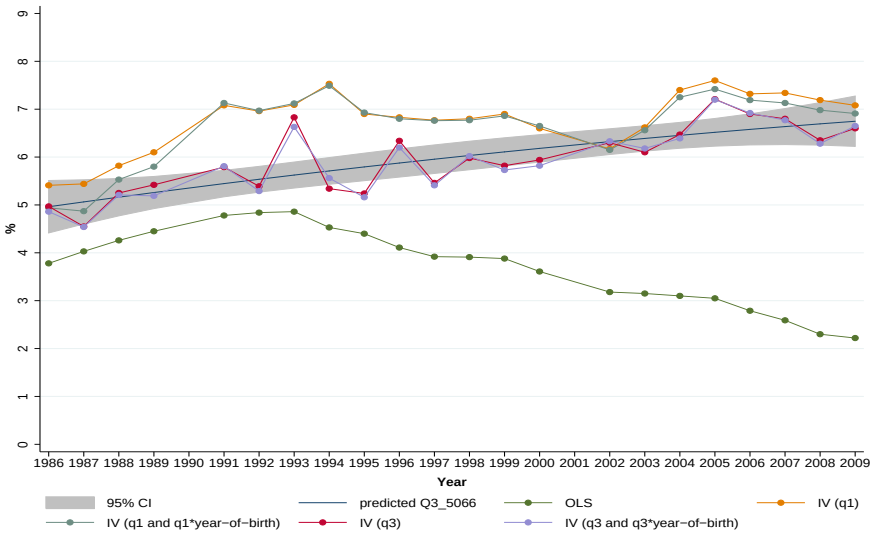
[26] See Tables A.9–A.14 in Appendix A.4 which report the results of the tests

As such, the GMM estimation is used.



Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.
Note: The shaded area corresponds to the 95% confidence interval for the predicted mean. Returns to education reported in the graph in the years 1990 and 2001 are interpolated. CI stands for confidence interval, q1 and q3 for 1^{st} and 3^{rd}quarter of birth, respectively. "predicted Q3–3049" identifies the blue line around which one builds the shaded 95% confidence interval.

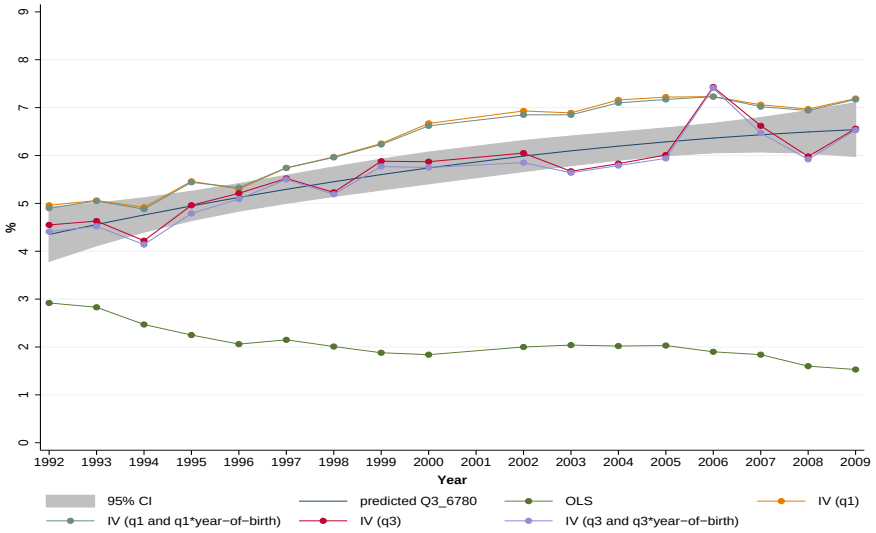
Figure 3.11: OLS and IV estimates of Return to Education, Men born 1930–1949



Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.
Note: The shaded area corresponds to the 95% confidence interval for the predicted mean. Returns to education reported in the graph in the years 1990 and 2001 are interpolated. See note to Figure 3.11.

Figure 3.12: OLS and IV estimates of Return to Education, Men born 1950–1966

Results obtained by OLS estimation of the wage equation shows that return rates to schooling range between around 1.8% and 7.3%, which are close to OLS results from



Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.
Note: The shaded area corresponds to the 95% confidence interval for the predicted mean. Returns to education reported in the graph in the years 1990 and 2001 are interpolated. See note to Figure 3.11.

Figure 3.13: OLS and IV estimates of Return to Education, Men born 1967–1980

other studies (see Levine and Plug, 1999; Dickson, 2012; Wang, 2013), and lower that the ones reported above (see Table 3.4). In all cohorts, in general, the OLS estimates of the returns have decreasing trend. The results also suggest that the IV estimates range between 4.1% and 8.2%. Over the period analysed, and for each cohort, the IV estimates' trend is more stable than the trend of OLS estimates. The return to education is statistically identical when the quarter of birth is used as instrument as compared to when using the quarter of birth together with the set of dummies interacting quarter of birth with the year of birth.[27] From the results one sees that, over the three cohorts, the IV estimated returns converge to about 7%, i.e., on average, an additional year of education leads to an increase in wage of about 7%, *ceteris paribus*.

Another note is that the IV estimates, for the first cohort from 1994 onwards, and for the other cohorts over the entire sample period, are higher than the OLS estimates, regardless of the instrument used. These findings are in line with most of the literature (for instance, Angrist and Krueger, 1991; Card, 1995b; Harmon and Walker, 1995; Harmon et al., 2003; Oreopoulos, 2006; Dickson, 2012; Wang, 2013). In Section 3.6 we will add more on this discussion.

[27] See Table A.9–A.14 in Appendix A.4

3.5.3.2 Changes of the compulsory schooling laws

In the literature the changes in laws affecting leaving school decisions are an alternative instrument for education (see Harmon and Walker, 1995 and Devereux and Hart, 2010, to the United Kingdom (UK); Vieira, 1999, to Portugal; Brunello and Miniaci, 1999, to Italy, among others). Econometric identification of the key parameter is usually achieved by the inclusion of dummy variables that record the exogenous change. Harmon and Walker (1995), for United Kingdom, were the pioneers to use the minimum legal school leaving age as an instrument. Minimum school leaving age has changed in the UK in 1947, from 14 to 15 years old, therefore authors define a dummy that equals one if the individual is aged 14 in the year in which the change takes effect. In this case, the individual faced a minimum school leaving age of 15, i.e. one year more than the previous cohort. The idea is that the cohort of people who faced different legislation of minimum school leaving age will have different levels of education, everything else equal, which one expects to be free of ability and any other effects related to wage in the labour market. The authors reported a substantial difference between OLS and IV estimates, where IV returns to schooling are of about 15%, compared to 6% from OLS estimation.

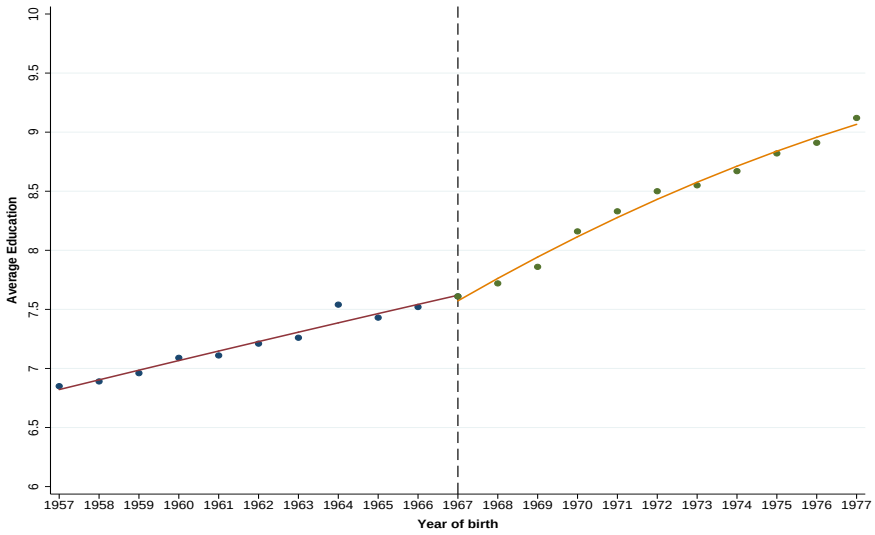
For Portugal, Vieira (1999) uses change in compulsory school attainment in 1964 which established the compulsory education of 6 years. The author found that OLS estimates range between 7.5% and 8.2%, while the IV estimates are around 5%. This result is also found by Pereira and Martins (2002a), who replicated the work of Vieira (1999).

In this section, a similar approach to Harmon and Walker (1995) and Vieira (1999) is followed. As mentioned in Chapter 2, Law no. 45810/64 sets up the compulsory schooling of six years for individuals born between 1967 and 1980, and Law no. 46/86 establishes an increase in compulsory schooling to 9 years for individuals born after 1980. The instruments used in this section take into account these two changes in the educational laws. Therefore, the definition of the identifying variables is as follows: (1) change in the compulsory schooling to 6 years (CSL6) is a dummy variable taking value 1 if the individual born was from 1967 (January) until 1980 (December) and 0 otherwise; (2) change in the compulsory schooling to 9 years (CSL9) is dummy variable taking value 1 if the individual was born from 1981 (January) and 0 otherwise.

These two dummy variables, CSL6 and CSL9, are expected to pick up the effects of the 1964 and 1986 reforms, respectively. We have defined a sub–sample to analyse the information around the first change of legislation. In this case, we focused on cohorts of workers born ten years before and 10 years after the implementation of the change (1967) – from 1957 to 1977. We will use labour market data for the period 1991–2009,

which means that this cohorts include men aged 18 to 53. Concerning the second law change one considers the cohort of individuals born five years before and four years after the implementation of the change (1981) – from 1976 to 1984 –, and use QP data for the period 2004–2009, i.e., men aged 18 to 34. The option for "four years after" results from the fact in 2009, the most recent data available in the current sample, workers latter cohorts would be younger than the minimum age in the sample, 18 years old.

The question that arises is if the change in minimum legal school leaving age leads to an increase in the educational attainment. Most studies generally found that the increase in minimum school leaving age boosts the schooling attainment of cohorts facing different compulsory education law (for instance, Harmon and Walker, 1995; Vieira, 1999; Meghir and Palme, 2003; Oreopoulos, 2006; Devereux and Hart, 2010).

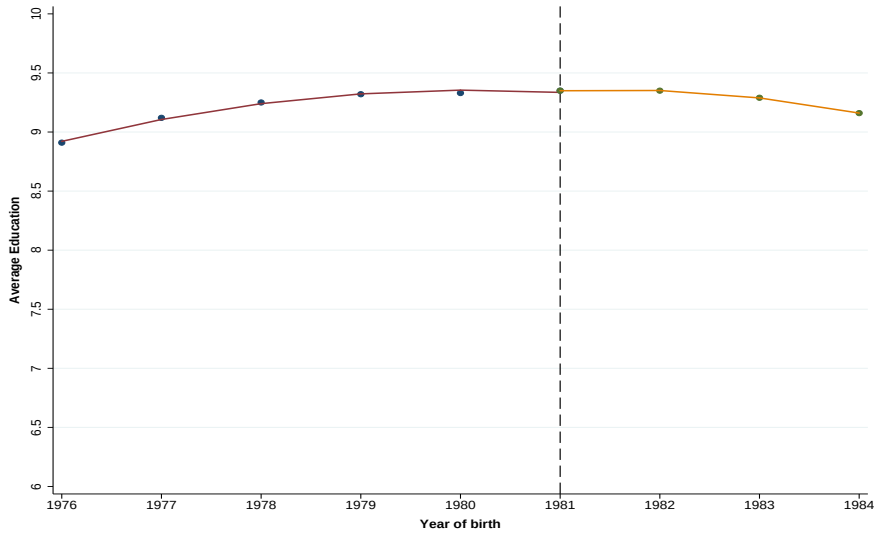


Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.

Figure 3.14: Average Education, Men born 1957–1977

In Portugal, the reform of 1964, which has effects on individuals who were born from 1967 onwards, had a significant impact on school leaving behaviour, as it can be seen in Figure 3.14.

When comparing the slopes on average education achieved before and after 1967, it is verified that the slope after the change in compulsory schooling seems to be more pronounced, i.e., the average education had significant enhancements from 1967 on. This improvement does not occur with the reform of 1986, since the average education is almost constant for individuals born before and after 1981, as it can be seen in Figure 3.15. Perhaps the individuals born before 1981 have chosen to invest in their education before the change of the compulsory schooling law.



Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.

Figure 3.15: Average Education, Men born 1976–1984

To use the compulsory schooling laws as instruments for educational attainment, the relationship between the change in the law and the variable of interest (schooling) is defined as

$$S_i = \rho X_i + \theta(Compulsory\,Schooling\,Law)_i + v_i$$

which is the first stage regression of the IV2SLS approach, where the explanatory variables are the same as in equation (3.30), except the variable *Compulsory Schooling Law* which is a dummy variable indicating the law change, CSL6 or CSL9, that will be used as instruments. The second stage equation is equation (3.31).

Tables 3.5 and 3.6 present the regression results of the schooling equation using educational law changes as instruments, CSL6 and CSL9, respectively. Firstly, it can be seen that CSL6 and CSL9 present a negative and significant effect. As discussed in Section 3.5.3.1, this occurs under the very definition of equal experience.

Tables 3.5 and 3.6 report statistics on the validity of the instruments and regressions results, when using CSL6 and CSL9 as instruments, respectively. These tables show that the first stage R^2 ranges between 0.44 and 0.70. Despite being lower than those of the previous estimates, these values are still quite higher than the R^2 first stage regressions found in many other studies (for instance, Harmon and Walker, 1995; Leigh and Ryan, 2008; Dickson and Smith, 2011). The results clearly show that the instrumental variables, CSL6 and CSL9, are significantly related to endogenous schooling, which points out that the instruments used are relevant. The validity of the instruments is also tested by computing the F–statistic on the exclusion of instruments in the reduced form schooling equation and for each year in the analysis. One concludes

Table 3.5: First stage of IV2SLS estimation using CSL6 as instrument, Men born 1957–1977

Year	CSL6	R^2	F–test on exc. variables	Year	CSL6	R^2	F–test on exc. variables
1991	-1.8420***	0.4398	2184***	**2000**	-1.7913***	0.6252	6099***
	(0.0394)				(0.0229)		
1992	-1.7423***	0.4465	2580***	**2002**	-1.9373***	0.6260	7311***
	(0.0343)				(0.0227)		
1993	-1.7040***	0.4582	2767***	**2003**	-2.0483***	0.6171	7544***
	(0.0306)				(0.0236)		
1994	-1.7057***	0.4797	3834***	**2004**	-2.1559***	0.6111	7964***
	(0.0275)				(0.0242)		
1995	-1.5601***	0.5742	2055***	**2005**	-2.2039***	0.6145	8179***
	(0.0337)				(0.0244)		
1996	-1.6522***	0.5762	3294***	**2006**	-2.2201***	0.6134	9083***
	(0.0288)				(0.0233)		
1997	-1.6890***	0.5855	4608***	**2007**	-2.2546***	0.6100	7192***
	(0.0249)				(0.0266)		
1998	-1.7982***	0.5927	5739***	**2008**	-2.2887***	0.6089	8842***
	(0.0237)				(0.0243)		
1999	-1.8211***	0.6062	6118***	**2009**	-2.3418***	0.6076	8608***
	(0.0232)				(0.0254)		

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 3.6: First stage of IV2SLS estimation using CSL9 as instrument, Men born 1976–1984

Year	CSL9	R^2	F–test on exc. variables	Year	CSL9	R^2	F–test on exc. variables
2004	-3.282***	0.5751	25703***	**2007**	-3.331***	0.6668	44556***
	(0.0204)				(0.0158)		
2005	-3.321***	0.6116	32365***	**2008**	-3.330***	0.6900	48053***
	(0.0185)				(0.0152)		
2006	-3.326***	0.6430	35904***	**2009**	-3.353***	0.7046	38056***
	(0.0176)				(0.0172)		

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

that the values of F–statistic are quite high for two instruments at 1% significance level, indicating that the two instruments are not weak instruments.

As before, using both CSL6 and CSL9 as instruments, it is applied the DWH test for differences between IV2SLS and OLS coefficients and the Pagan–Hall test to verify the homoscedasticity. The results show that the null hypothesis of DWH test is strongly reject, indicating that education variable is endogenous and the OLS estimates are biased. The Pagan–Hall test reject the null of homoscedasticity of the error terms,

and, consequently, the GMM method is used.[28]



Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.
Note: The shaded area corresponds to the 95% confidence interval for the predicted mean. Returns to education reported in the graph in the years 1990 and 2001 are interpolated.

Figure 3.16: OLS and IV estimates of Return to Education, Men born 1957–1977

Figure 3.16 and Table 3.7 present the estimation results from the OLS method and from the second stage of IV2SLS approach, which uses CSL6 and CSL9 as instruments for schooling, respectively.[29]

The results show that the OLS estimates are similar among them in both cohorts. A similar conclusion applies to IV estimates. It is found that the percentage increase in real hourly earnings attributed to an additional year of schooling ranges between 4% and 5.6% with OLS estimation, while it ranges between 5.8% and 8.1% with IV estimation. These results are in line with other studies (for instance, Angrist and Krueger, 1991; Brunello and Miniaci, 1999; Hartog et al., 2001; Oreopoulos, 2006; Dickson, 2012; Grenet, 2013). As it happens in most of the international literature, and in the previous subsection where the quarter of birth is used as instrument, the OLS results underestimate the returns to investments in education. Two key arguments that explain this result, which are discussed at length in Card (1999, 2001), hinge upon the possibility of heterogeneous returns to education and measurement error. We postpone the discussion of these arguments to Section 3.6.

A final note. Our results are at odds with Vieira (1999) and Pereira and Martins (2002a), as these authors found higher OLS estimates for the returns to education,

[28]We do not report the full table with the statistics to avoid dimensionality issue with the size of the appendices. These statistics will be provided once requested.
[29]Results in Figure 3.16 are reported in the Table A.15 in Appendix A.4.

Table 3.7: OLS and IV estimates using CSL9 as instrument, Men born 1976 – 1984

	2004		2005		2006	
	OLS	**IV**	**OLS**	**IV**	**OLS**	**IV**
VARIABLES	**Dependent variable: log real hourly wage**					
Education	0.0390***	0.0586***	0.0417***	0.0625***	0.0431***	0.0637***
	(0.0003)	(0.0013)	(0.0003)	(0.0013)	(0.0003)	(0.0013)
Experience	0.0378***	0.0457***	0.0362***	0.0466***	0.0328***	0.0436***
	(0.0015)	(0.0020)	(0.0014)	(0.0022)	(0.0013)	(0.0025)
Experience sq	-0.0008***	-0.0008***	-0.0007***	-0.0008***	-0.0005***	-0.0007***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Centre	0.0178***	0.0125***	0.0116***	0.0063***	0.0210***	0.0163***
	(0.0021)	(0.0091)	(0.0021)	(0.0088)	(0.0021)	(0.0090)
Lisbon and Tagus Valley	0.1180***	0.1041***	0.1190***	0.1053***	0.1290***	0.1165***
	(0.0017)	(0.0083)	(0.0017)	(0.0081)	(0.0017)	(0.0082)
Inland	-0.0408***	-0.0445***	-0.0449***	-0.0487***	-0.0462***	-0.0505***
	(0.0026)	(0.0094)	(0.0025)	(0.0094)	(0.0026)	(0.0095)
Algarve	0.0730***	0.0633***	0.0651***	0.0559***	0.0660***	0.0578***
	(0.0039)	(0.0110)	(0.0037)	(0.0106)	(0.0037)	(0.0095)
Observations	202,550	202,550	237,355	237,355	249,961	249,961
R–squared	0.432	0.420	0.431	0.418	0.435	0.423
RMSE	0.308	0.311	0.323	0.326	0.332	0.335
	2007		**2008**		**2009**	
	OLS	**IV**	**OLS**	**IV**	**OLS**	**IV**
VARIABLES	**Dependent variable: log real hourly wage**					
Education	0.0442***	0.0645***	0.0447***	0.0655***	0.0471***	0.0676***
	(0.0003)	(0.0013)	(0.0003)	(0.0013)	(0.0003)	(0.0013)
Experience	0.0292***	0.0404***	0.0334***	0.0444***	0.0337***	0.0442***
	(0.0013)	(0.0025)	(0.0012)	(0.0024)	(0.0012)	(0.0024)
Experience sq	-0.0004***	-0.0006***	-0.0006***	-0.0008***	-0.0006***	-0.0007***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Centre	0.0145***	0.0104	0.0154***	0.0115	0.0073***	0.0032
	(0.0022)	(0.0088)	(0.0022)	(0.0088)	(0.0022)	(0.0086)
Lisbon and Tagus Valley	0.1281***	0.1171***	0.1224***	0.1123***	0.1102***	0.1007***
	(0.0017)	(0.0091)	(0.0017)	(0.0091)	(0.0017)	(0.0094)
Inland	-0.0550***	-0.0586***	-0.0563***	-0.0594***	-0.0525***	-0.0562***
	(0.0026)	(0.0100)	(0.0026)	(0.0100)	(0.0026)	(0.0092)
Algarve	0.0740***	0.0674***	0.0719***	0.0655***	0.0596***	0.0538***
	(0.0037)	(0.0109)	(0.0037)	(0.0109)	(0.0038)	(0.0109)
Observations	266,476	266,476	279,683	279,683	275,351	275,351
R—squared	0.431	0.421	0.438	0.429	0.450	0.441
RMSE	0.343	0.346	0.347	0.350	0.349	0.352

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: Robust standard errors in parentheses (IV estimates columns), clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include a set of dummies for occupation and sector of activity.

around 8%; reported IV estimates were about 5%. The gap between the results found in this research and the results obtained by these two Portuguese papers could, at least in part, be explained by the substantial sample difference, both in the time frame as well as the number of observations. In the current analysis the number of observations in each year is over 250,000, while Pereira and Martins (2002a), for example, used less than 30,000 observations.

3.5.3.3 Average education by region

Decisions about the education of an individual, i.e. whether or not one chooses to own more education, are likely to be affected by several factors. It may be reasonable to believe that one of the factors that can influence the education of an individual is the general trend for schooling of previous generations that surround him. Taken this factor into account, in this section a new instrument for education is proposed, which is the "average education by region in the year in which individual entered school". This instrument is suitable as it is based on factors at the time the individual enters school, so these factors are not correlated with their earnings.

On should make it clear from the beginning of this section that the validity of the instrument relies heavily on the assumption that individuals work in the same region as they went to school. One needs to make such assumption as we do not have information in QP on the location of the worker while he was in school. The Portuguese labour market is characterized by a low mobility rate, which would support such assumption. The instrument would be reasonable reliably as long as on average workers do work currently where they lived in the past.

We now estimate the returns to education using the average education by region in the year in which individual entered school, and in the region he works currently, as instrument to the endogenous schooling. In addition, the "average education in Portugal in the year in which individual entered school" is also used as instrument. The instruments are computed from QP data as averages of education of each worker's cohort.[30]

As in both previous sections, to use this new instrument for educational attainment, the previous specification is estimated, i.e. the relationship between the average education by region and the variables of interest (schooling and wages), using a IV2SLS approach. Thus, the first stage regression estimated is given by:

$$S_i = \rho X_i + \theta(Average\,Schooling\,by\,Region)_i + v_i$$

where the explanatory variables are the same as in equation (3.30), except the variable *Average Schooling by Region*. The second stage equation is again equation (3.31). In addition, it is also used the average schooling in Portugal in the year of entry school. Reinforcing the information, the sample of men aged 18 till 60 is used. Table A.16 shows the first–stage regressions of the schooling equation using the average schooling

[30] An alternative would be to use census data, as it is representative of the whole population. However, census information includes, for example, public servants who are not representative of the sample at hand. As such, one has made the option to use averages across the groups that are closer to the workers we are analysing, i.e., employees in the private sector.

in Portugal and average schooling in region as instruments.

Concerning the quality and validity of both instruments, Table A.16 in Appendix A.4 indicates that the first stage R^2 ranges between 0.36 and 0.71 and this range of values being similar to those of the previous IV regressions of this essay. Thus, for each year under study, both instrumental variables are significantly related to endogenous education variable, which points out that the instruments used are relevant. The results also includes values for F–statistic on the exclusion instruments in the reduced form schooling equation and, similarly to the previous IV regressions, they are quite high for both instruments. As such, it cannot be rejected the significance of the excluded instruments in the schooling equation, in this way the instruments are considered to be valid.

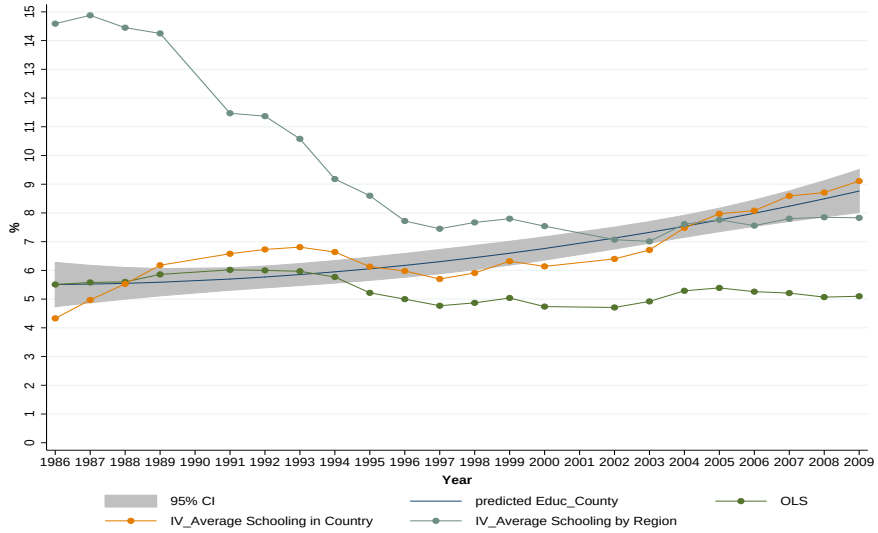
As in previous sections, the DWH and Pagan–Hall tests are applied. The results suggest that the OLS estimates of the returns to education may be biased due the endogeneity of the schooling variable, as the DWH test statistic rejects the null hypothesis of exogeneity of education. In addition, the Pagan–Hall test for heteroscedasticity rejects the null hypothesis that the variance of the unobservable error, conditional on the explanatory variables, is constant. Therefore, the GMM approach is applied as it is the appropriate estimation procedure under the current set-up.

The results show that the instrument has a negative and significant correlation with education. The first–stage regression under the IV2SLS includes as explanatory variables for education the instruments as well as the exogenous variables defined in equation (3.31), namely experience and its square, occupation, industry and location. This means that, conditional on worker's experience and occupation, and the industry and location of his employer, the higher the education level of his cohort at this county the lower is his education.

Figure 3.17 presents the estimation results from the OLS estimation and from the second stage of IV2SLS approach that uses both instruments for schooling.[31]

The results are in line with the ones described in the two previous sections. It is found that the returns to schooling obtained from OLS estimations range between 4.7% and 6% for each additional year of schooling, while IV2SLS estimates of returns to education range between 5.7% and 9.1% if the average education in Portugal in the year in which individual entered school is used as instrument if the instrument is average education by region in the year in which individual entered school the return are in the range 7%–14.9%. Furthermore, from 2004 onwards the IV estimates are statistically similar when the different instruments are used. While in the first years of the sample the difference in returns after each estimation choice is high, it decreases

[31] Results in Figure 3.17 are reported in the Table A.17 in Appendix A.4.



Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.
Note: The shaded area corresponds to the 95% confidence interval for the predicted mean. Returns to education reported in the graph in the years 1990 and 2001 are interpolated.

Figure 3.17: OLS and IV estimates using Average Schooling by Region and in Portugal, Men aged 18–60

substantially over time. The difference is particularly high when one uses as instrument the average schooling by region, where estimates go from about 15% in 1986 to about 8% in 2009. By the end of the sample time frame, the estimates of returns to schooling are closer and bounded between 8% and 9% for IV, while OLS shows a return of about 5%.

Once again, IV estimates are generally higher than the OLS ones. We provide further discussion on this in Section 3.6.

3.5.4 Firm specific effects and instrumental variables

Taking advantage of the fact that one has an employer–employee panel data, in this section we go a step further in the analysis of returns to education and combine fixed-effects procedures with an instrumental analysis approach. We will control for firm heterogeneity using firm specific effects, while addressing workers' endogenous education by using as instruments quarter of birth, changes in the compulsory schooling laws and average schooling in region / country in year that individual entered school. Given the large number of firms included in the sample, one cannot use the usual procedures, but, instead, use a strategy based on high dimensional fixed-effects. Specifically, we will make use of the work developed by Guimarães and Portugal (2010).[32] In order

[32]To run the regressions discussed in this section one uses specifically the solution implemented by Correia (2014), namely the Stata command "reghdfe". This is an improved version of the procedure

Table 3.8: Firm specific effects and instrumental variables

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Instrument: quarter of birth					
	Cohort of men born 1930–1949		Cohort of men born 1950–1966		Cohort of men born 1967–1980	
	Year: 2004	1986–2004	Year: 2009	1986–2009	Year: 2009	1992–2009
Education	0.0638***	0.0613***	0.0477***	0.0539***	0.0471***	0.0453***
	(0.0141)	(0.0016)	(0.0054)	(0.0010)	(0.0050)	(0.0012)
Observations	45,771	3,030,494	300,576	7,833,080	390,094	5,549,012
R–squared	0.844	0.776	0.826	0.784	0.794	0.770
RMSE	0.285	0.273	0.290	0.283	0.277	0.264
	Instrument: compulsory schooling laws					
	CSL: 6^{th} grade		CSL: 9^{th} grade			
	Year: 2009	1991–2009	Year: 2009	2004–2009		
Education	0.0512***	0.0525***	0.0496***	0.0477***		
	(0.0008)	(0.0002)	(0.0006)	(0.0002)		
Observations	531,228	9,738,817	226,172	1,476,854		
R–squared	0.805	0.771	0.771	0.767		
RMSE	0.288	0.280	0.245	0.228		
	Instrument: Average schooling in region / country in year that individual entered school					
	Av. region		Av. country			
	Year: 2009	1986–2009	Year: 2009	1986–2009		
Education	0.0616***	0.0851***	0.0653***	0.0757***		
	(0.0008)	(0.0002)	(0.0006)	(0.0001)		
Observations	921,539	17,593,265	921,539	17,593,265		
R–squared	0.784	0.725	0.782	0.737		
RMSE	0.286	0.301	0.288	0.295		

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: Standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. The specifications that regressions are run for a period, include a set of dummies for year.

to compare the estimates of returns to education, for each instrument one uses the period of the cohort previously considered, as well the most recent year within each cohort. The validity of the estimates reported below is not rejected, as the Stock and Yogo (2005) test shows that the instruments are strong – the Cragg–Donald F–statistic exceeds the tabulated critical value. The results of these regressions are reported in Tables 3.8.

Overall, the magnitude of the return to education is quite stable across the regres-

developed by Guimarães and Portugal (2010), which allows for the use of instrumental variables and linear GMM.

sions. Controlling for the unobservable characteristics of the workers and their firms, on average, an additional year of schooling yields a wage premium between 4.5% and 8.5%. Comparing with the results discussed above, one arrives to the key message of this chapter. Addressing endogeneity and different sources of heterogeneity, and working with distinct specifications, returns to education are significant and substantial. Once one takes into account for endogenous education choices, estimated returns become larger, when compared with OLS. However, firms' heterogeneity plays a relevant role in explaining wage differences, which can be seen by the decrease in the IV estimates after controlling for firms specificities. It is as if "better firms", taken as those who pay higher wages, recruit more educated workers. Clearly, the typical IV approach does not control for this source of endogeneity.

Summing up, a contribution of this chapter might be the combination of different instruments, while proposing the instruments based on average schooling, with estimation strategies, namely the ones that address firms' heterogeneity, to estimate consistently the path of returns to education for the Portuguese labour market.

3.6 Concluding remarks

In this chapter, the returns to education in Portugal are examined, because despite the considerable educational upgrading of the labour force in the last years, little has been analysed about the changes in the returns to education over time for Portugal. Using the standard formulation of the Mincer equation, for the period 1986–2009, the results clarify the existence of significant and large returns to investments in education measured by years of schooling, i.e. during the period under study, the impact of schooling on wages is, on average, around 10% and men receive, marginally, and over the whole period, higher returns than women, 10.1% and 9.8%, respectively. Furthermore, it is found that the trend of the returns to education seems to be growing from 1986 to 1995, reaching a peak in 1995 (10.3%), and reverting after that. However, from 1997 to 2004, there is a slightly increasing trend again and at the end of the period under study, the returns remain practically constant and close to 10%.

When the rate of the return for both genders is compared, it can be observed that the trend of the returns for men and for women is opposed, since from 2000 this trend is slightly decreasing for men, but for women this trend keeps increasing until 2009, reaching a return rate of 10.3%. Considering the different educational levels, the return rate to all levels is fairly stable, except to the tertiary level (*"licenciatura"*), which has a remarkably increasing trend from 1986 to 1995. The results also propose that the higher educational qualifications leads to higher returns for both genders.

Additionally, OLS and IV estimates of the returns to education is also compared in this analysis, i.e. the IV method is used and to provide accurate estimates of the returns a representative sample of Portuguese male workers is explored. Changes in the educational legislation and season of birth are considered as instrument to education, in particularly, the laws of 1964 and 1986, and furthermore, the first and third quarter of birth, as well the interaction between the quarter of birth and the year of birth are used as instrument. Besides that, the "average education by region in the year in which individual entered school" is proposed as an innovative instrument.

Concerning the results of the first stage of IV show that all instruments used present a strong correlation with the schooling variable and, after performing the tests suggested by the empirical literature, they are found valid.

The results show, firstly, that, regarding the magnitude of the estimates, when using the standard Mincer equation together with additional controls, and quarter of birth as an instrument, the rate of return to schooling by IV varies between 4.1% and 8.2%. Such returns are generally higher than the OLS estimates (1.8%–7.3%). Secondly, using the compulsory schooling laws as instruments, the results show returns to schooling ranging between 5.8% and 8.1%. For comparable cohorts, the OLS estimates range between 4% and 5.6%, i.e. IV estimates are, once more, higher than the OLS estimates. The results discussed above on the use of this instrument highlights the influence of the legal changes on the school attainment. Thirdly, using "average education in Portugal in the year in which individual entered school" as an instrument, estimated returns range between 5.7% and 9.1%, which is in line with the ones reported by IV based on quarter of birth. Finally, when the "average education by region in the year in which individual entered school" is used as instrument, the estimated return to education ranges between 7.8% in 2009 and 14.9% in 1986. These results are somehow larger than the previous IV results in the 1980s, but align with them in the second half of the period under analysis. A key result is that the IV estimates converge to about 7% by the end of 2000s.

The fact that the results of the OLS estimates are lower than the IV estimates, which is at odds with the findings proposed by the Portuguese empirical literature, deserves a discussion on its own. From the outset, one should note that the current results are consistent with most of the international literature, which generally finds that the OLS approach underestimates the return to education. Card (1995a, 1999, 2001) discuss different explanations that may reconcile the different results. On the one hand, the difference between the two estimators can occur because of endogeneity. This argument predicts that if one does not control for unobserved heterogeneity the education parameter will capture, both the true effect of education, as well as the

ability effect on wages; the argument being that more able workers invest more in education. Using IV to filter out the endogeneity source would lead to a decrease in the estimated parameter associated with education, as now we would be capturing the true return to education.

On the other hand, a downward bias can be expected as the result of measurement error on education or heterogeneous returns to education. First, education is a bounded variable, which implies that mismeasurement of lower values of education cannot be downward, while errors for highly educated workers can only be downwards. Combined, this would lead to an underestimation of the true return to education. A similar would be reached if one assumes orthogonality between the measurement error and earnings (see Card, 2001).

Second, a downward bias on the OLS estimator can also be expected following heterogeneous returns to education. This heterogeneity can result from, for example, the existence of workers with lower education that, in the absence of an intervention, will may have higher marginal returns than the average marginal returns to education in the population as a whole. Their lower level of education may reflect their high marginal costs of schooling, due to increasing marginal discount rates, taste factors, or financial constraints, rather than their low ability, which would limit their aggregate return to education, i.e., if the main factor determining low formal education is higher than average costs of schooling, rather than because lower than average returns to education, and if the relative variation in ability is small, some institutional features, like compulsory schooling, may affect the schooling choices of individuals who would otherwise have relatively low schooling. In other words, the marginal returns to schooling for certain groups of individuals with low education, which are the subgroups of population whose schooling decisions are most affected by structural innovations in the schooling system, are somewhat higher than the average marginal returns to education in the population as a whole. This way, IV estimates of return to education based on institutional changes that raise schooling levels among workers with lower education may exceed the corresponding OLS estimates.

Overall, a possible explanation, following the arguments above, is that the negative bias from measurement error and heterogeneous returns, overcome the positive bias associated with ability, leading to greater IV estimates of the return to education, when compared with the OLS counterpart.

Chapter 4

Educational spillovers in Portugal

> "Most of what we know we learn from other people. We pay tuition to a few of these teachers, either directly or indirectly by accepting lower pay so we can hand around them, but most of it we get for free, and often in ways that are mutual - without a distinction between student and teacher."
>
> Lucas, 1988, p.38

4.1 Introduction

Private returns to education have long been analysed, and there is consensus among researchers around the positive impact of education on individual earnings. Nevertheless, positive effects of education go beyond the individual level, as education may have spillover effects to other individuals, which may lead to social returns.

Educational externalities may arise if workers become more productive by interacting with more educated workers. Recently, research has been looking at those externalities at the local labour market level. Empirical results on both the existence and the magnitude of educational externalities are not consensual. Rauch (1993) and Moretti (2004a), in studies for the US, find evidence for the existence of a relationship between individual earnings and the regional concentration of human capital that can be explained partly by productivity gains that workers may obtain by interacting with more skilled workers in locations with a high concentration of human capital. However, more recent papers suggest opposite results, providing no evidence of local educational externalities (Ciccone and Peri, 2006).

The idea that educational benefits are not fully appropriated by the individual, may justify public intervention in education provision, and as such, better knowledge on educational spillovers is of major interest. Furthermore, education spillovers have been extensively studied for the US, but there is lack of studies for other countries.

The main goal of this study is to contribute to the literature by analysing the ed-

ucation spillovers in Portugal. Namely, this essay intends to contribute to a better understand of the relationship between interactions among Portuguese workers and individual wages. It aims at answering the following questions: Do human capital externalities matter for individual earnings in Portugal? If there are human capital externalities, to what extent are they more or less relevant than the individual educational attainment?

External returns to human capital are analysed by means of an augmented Mincerian wage equations, which is estimated using a matched employer–employee data set. The aggregated human capital is measured by means of three different indicators: the average formal education level, the share of highly qualified workers, and an adjusted multi–dimensional skill index (Portela, 2001), which are computed at different aggregation levels (firm and region). The existing literature is improved by computing those indicators in three different perspectives, which is an improvement to the existing literature: at the regional level (Rauch, 1993; Moretti, 2004a); at the firm and regional level (Canton, 2009; Bauer and Vorell, 2010) and, finally, at the regional and inter–regional level, considering the possibility of spatial externalities between regions (Rodríguez-Pose and Tselios, 2012).

In order to account for the unobserved time–invariant individual's characteristics, the change in compulsory school attainment in 1981 that established the compulsory education of 9 years is used as instrument to individual education. Additionally, the endogeneity of aggregate human capital is suggested in literature. Following Moretti (2004a), the presence of a polytechnic institute in the county is used as instrument to the aggregated human capital at regional level and a new instrument is proposed: the share of people who could read in the county according to the 1900 census. A fundamental assumption in the Mincerian regression is that workers with different educational levels are perfect substitutes in production, however, empirical evidence indicates that the causal impact of local human capital concentration on wages is that workers with different levels of education are imperfect substitutes in production (Moretti, 2004a,b; Ciccone and Peri, 2006). As such, to control to the imperfect substitutability across workers, two qualification groups of workers are considered.

The remainder of this essay is organized as follows. The next Section provides a literature review on the external effects of education. Section 4.3 briefly presents the theoretical framework. Section 4.4 describes the empirical identification strategy and the data used for models' estimation. Section 4.5 reports the estimation results for three aggregation levels and for three groups of workers: all workers, the most qualified workers and the less qualified workers. Section 4.6 reports the main concluding remarks.

4.2 Literature review

Over several decades, research on the effects of human capital on individual earnings have been the focus of a huge amount of research. Consensus has been reached on the result that the greater the educational level of the individual, the higher the private gains he/she gets. Furthermore, there are some findings in the empirical literature indicating that social returns to education exceed private returns, which may be due to the presence of externalities. The central idea is that the cumulative knowledge by agents in a local environment may benefit the other agents; that is, the average stock of human capital in a region could have an effect on individual wages of the workers in this region (Rauch, 1993; Acemoglu and Angrist, 2001; Moretti, 2004a,b; Ciccone and Peri, 2006). Thus, the strength of human capital externalities is defined as the difference between the marginal social return and the private return to an additional unit of human capital (Garcia-Fontes and Hidalgo, 2009).

Empirical literature has been using three main strategies to evaluate the external effects of the human capital: (i) the assessment of the effects of the aggregate human capital in a city/region (e.g. Rauch, 1993; Moretti, 2004b) on individual wages; (ii) the examination of the effects of the aggregate human capital on productivity indicators using data at firm level or at regional level (e.g. Liberto, 2008); (iii) the evaluation of the effect of the education on social outcomes, such as, crime rates (e.g. Lochner and Moretti, 2004), living standards, occupation and political participation (e.g. Wantchekon et al., 2015).

The existence of human capital spillovers has long been documented in economic literature. Marshall (1890) was the first to recognize that the interactions among workers create learning opportunities which may yield increases in productivity. The most recent literature has been built upon this Marshall's vision and some models where the human capital externalities are the fundamental engine to economic growth have been proposed (e.g. Lucas, 1988). In a prominent article, Lucas (1988) suggests that the externalities of the human capital in the form of learning spillovers may be large enough to explain the differences in the long run income between poor and rich countries. Furthermore, the issue of externalities of the human capital is also extremely relevant, since it determines the magnitude of human capital accumulation investment (Heckman and Carneiro, 2003; Heckman, 2000; Isacsson, 2005; Garcia-Fontes and Hidalgo, 2009).

Despite significant political implications, the empirical evidence of the existence and the magnitude of the external returns to education is very scarce when compared with the extensive literature on the private returns to education. Some authors attempted to estimate the magnitude of the education spillovers by comparing the wages of similar people who work in cities or states with different average education levels (for instance,

Rauch, 1993; Acemoglu and Angrist, 2001; Moretti, 2004a,b; Ciccone and Peri, 2006; Heuermann, 2011).

The methodology used is mostly the estimation of Mincerian wage equations augmented with an aggregate level human capital term at region level. This methodology estimates the strength of human capital externalities by observing the effect of regional human capital level on individual wages; the aggregate human capital is measured by the average schooling (e.g. Rauch, 1993; Acemoglu and Angrist, 2001; Czaller, 2016) or by the share of high skilled workers in the region (e.g. Moretti, 2004b; Muravyev, 2008; Bakis et al., 2010; Heuermann, 2011). The underlying idea is that human capital externalities should appear in the individual wages, when the relevant individual characteristics, such as education, gender, experience, are controlled for (Rauch, 1993; Rudd, 2000; Acemoglu and Angrist, 2001; Moretti, 2004b; Ciccone and Peri, 2006; Garcia-Fontes and Hidalgo, 2009).

In a pioneer work, using cross–sectional data from 1980 for Standard Metropolitan Statistical Areas (SMSA) in United States (US) and using the average level of formal education and work experience as proxies for the average level of human capital, Rauch (1993) finds significant and positive effects of the average level of schooling on individual wages; an increase of an year in the average schooling in the SMSA raises the workers' wages from 3% to 5%, on average. Concerning the average work experience, a weak and insignificant coefficient is found.

Several authors have followed that Rauch's work of 1993. Acemoglu and Angrist (2001), using pooled data from the United States Census for the period 1960–1980, find that a variation of one year in the average schooling of the US states yields an increase on workers' wages of about 7%, on average. Using the share of skilled workers at the city level as a measure of regional human capital, Moretti (2004a) concludes that an increase of 1 percentage point (p.p.) in the share of skilled workers yields externalities of about 1.5%. Research on countries outside the US confirms the existence of positive and significant human capital externalities. For countries such as Sweden (Isacsson, 2005), Italy (Dalmazzo and de Blasio, 2007), China (Liu, 2007), Russia (Muravyev, 2008), Netherlands (Canton, 2009), Spain (Garcia-Fontes and Hidalgo, 2009), Germany (Bauer and Vorell, 2010), Turkey (Bakis et al., 2010) and Hungary (Czaller, 2016), the OLS estimates of educational externalities on average wages lie between 1% and 9%, on average.

Two main criticisms of the Rauch's work have been made in the literature. First, the author doesn't discuss the endogeneity of the human capital variable. Second, he doesn't distinguish between external effect and complementary effect among skilled and unskilled workers (Moretti, 2004b; Ciccone and Peri, 2006).

By ignoring the endogeneity of the human capital variable, the dimension of the coefficient of human capital spillovers may be biased due to the presence of unobservable factors that are correlated with the average stock of human capital and wages among the regions (Moretti, 2004a; Dalmazzo and de Blasio, 2007). The key feature in the estimation of human capital externalities at the regional level is that changes in levels of aggregate human capital are endogenous due to the migration among regions, because regions with higher productivity and wages may attract more skilled workers. Moreover, high income regions may have amenities that are especially attractive for the highly skilled workers. Therefore, in order to reduce this bias, the implementation of instrumental variables (IV) approach is desirable (Moretti, 2004a; Dalmazzo and de Blasio, 2007; Garcia-Fontes and Hidalgo, 2009).

In the context of the IV methodology, a good instrument for average schooling must affect the schooling of the majority of workers in a location without being correlated with the local wages in another way (Duranton, 2007). The literature proposes a variety of instruments such as changes in child labour and compulsory education laws at the state level (Acemoglu and Angrist, 2001), the lagged demographic structure of the city/region (Moretti, 2004a; Dalmazzo and de Blasio, 2007; Garcia-Fontes and Hidalgo, 2009), the existence of colleges in the cities (Moretti, 2004a,b), the regional share of workers with a college degree 20 years before the observed period (Bauer and Vorell, 2010), and the number of high schools in the region, as well the number of students attending them (Heuermann, 2011).

In empirical studies, the IV results confirm the existence of positive and significant external returns to education, however, the range of the return rate is notably large in size. Acemoglu and Angrist (2001), using pooled data from 1960 to 1980 Census, find little evidence of the effects of the external returns to education, but when the 1990 data are added, the effect becomes significant and is about 1%–2%, indicating the rise in importance of human capital. Some other studies found similar IV estimates, such as Dalmazzo and de Blasio (2007); Garcia-Fontes and Hidalgo (2009); Bakis et al. (2010); Czaller (2016). Nevertheless, higher IV estimates of external returns to education are found in the literature. For instance, Liu (2007) indicates that an increase of one year in the average education at the city level could increase individual earnings by between 11% and 13%. Similarly, Ciccone and Peri (2006) suggest an external effect of an additional year of average schooling of 10%.

A key assumption of Mincerian regressions is that workers with different human capital levels are perfect substitutes in production (Moretti, 2004a,b; Ciccone and Peri, 2006). It assumes that changes in the relative supply of human capital do not affect the relative wages of the different human capital groups, keeping the total factor pro-

ductivity constant. Nevertheless, empirical evidence indicates that different levels of human capital are imperfect substitutes (Angrist, 1995; Moretti, 2004a; Ciccone and Peri, 2006), that is, an increase in the aggregate supply of skilled workers tends to increase the unskilled workers' wages and diminish the wages of high level education workers. Consequently, the effects of any change in the relative supply of human capital have on workers with a given level of human capital reflects on the total factor productivity and should be interpreted as externalities (Moretti, 2004b; Ciccone and Peri, 2006). When different levels of human capital are imperfect substitutes, the Mincerian approach yields a positive effect of aggregate human capital on individual wages (Ciccone and Peri, 2006).

Moretti (2004a,b) investigates the role of imperfect substitution by estimating Mincerian regressions for separated educational groups, and develops a theoretical model to capture the external effects of education on wages. Using individual data from the United States and taking into account the endogeneity of the location choice by instrumenting the share of high skilled workers by the existence of land–grant colleges in a region, he finds a wage premium of 0.4% for the college educated workers if the share of the high educated workers increases by 1 p.p., while for low skilled workers the estimate is considerably higher varying between 1.6%–1.9%. In line with these results, other authors found higher educational externalities for low skilled workers than for more educated workers. For instance, Liu (2007) conclude that the effect for low skilled workers is 13% and 10% for high skilled workers. According to Czaller (2016) the low skilled workers' effect is 16.4% whereas high skilled workers' effect is 14.9%. Bauer and Vorell (2010) shows that 1 p.p. increase in the share of high skilled in a region increases the wage of high skilled workers by 0.2% and those of low skilled workers by 0.6%. Heuermann (2011) conclusions go in the opposite direction: the wage effects of regional human capital are smaller for non highly qualified than for highly qualified workers, that is, a 1 p.p. of increase in regional human capital raises the wages of the former by 0.6% and wages of the latter by 1.79%.

Ciccone and Peri (2006) have introduced a constant–composition approach, an alternative to Mincerian approach. This approach aims to estimate the gross spillovers to correct the bias and allows for the identification of the human capital externalities even when the workers with different educational levels are imperfect substitutes. As such, it estimates the strength of the human capital externalities as the marginal effect of the aggregate levels of human capital on average wages holding constant the composition of the workforce, even if the skilled and unskilled workers are imperfect substitutes (Garcia-Fontes and Hidalgo, 2009). The authors argue that the external effects of education on wages are identified by the variation of the average educational level over

time in a given geographical area, keeping constant the weights of skills composition in the geographical area. Ciccone and Peri (2006) use pooled data at the regional level, for the period between 1970 and 1990, and find no evidence of external returns to the average years of education in US cities and states.

It is worth a note that the size of the educational externalities found in literature depends on the aggregate human capital measure used, the time period under study, the level spatial aggregation, the country and the model specification.

The proximity of the interactions across workers may determine differences in individual earnings as well. The literature suggests that external effects can occur at both levels, firm and regional, and that the impact of the regional average education and the regional share of skilled workers on wages can be extinguished when the human capital stock of the firms is included as a control variable. Only two works that analyse these two effects together are known, Canton (2009) and Bauer and Vorell (2010). The latter uses an employer–employee matched panel data set for Germany and finds that high–skilled workers experience positive regional and intra-firm spillover effects of education, contrary to low–skilled workers that seems to benefit only from the regional externalities. The results indicate that an increase of the share of high–skilled workers in a region by 1 p.p. leads to an increase of the wage of high skilled workers by 0.2% and those of low-skilled workers by 0.6%. For the high–skilled workers, a 1 p.p. increase of the share of high-skilled in the firm increases the wages by about 3%. Canton (2009), using Dutch survey data, concludes that the regional fraction of high–skilled workers and the region average schooling have a positive and statistically significant effect on individual wages. Nevertheless, the impact of both average educational attainment in the region and the regional fraction of high–skilled workers on individual wages completely vanishes when the firm's human capital stock is included as a control variable, suggesting that human capital externalities predominate within firms.

According to Rodríguez-Pose and Tselios (2012), interregional educational externalities may also affect wages. Rodríguez-Pose and Tselios (2012) made an attempt to measure the relationship between individual wages and interregional educational externalities across the regions of the European Union. A model including the educational attainment of the individual and the aggregate educational attainment at the level of the household, the region and the neighbouring regions was estimated by OLS. Results show that educational externalities are 3%, between 1.7% and 6.6%, and between 7% and 13% at the household, regional and interregional level, respectively. It suggests that individual earnings are determined not only by the individual's education, but also by the external returns to schooling, which incorporates household, regional, and interregional externalities.

There are some studies documenting the existence of substantial wage gap between different regions in Portugal (for instance, Cardoso, 1991; Teulings and Vieira, 2004; Vieira et al., 2005, 2006), which point out that the wages are higher in the region of Lisbon and Tagus Valley comparatively to other regions of the country. Nevertheless, externalities at the regional level are almost unexplored. An exception is the working paper of Silva (2003). Employing an empirical approach with data for workers who were displaced between different counties during the period from 1989 to 1999, Silva (2003) finds positive evidence of educational spillovers at the county level, which range from 1.7% to 8.4%. However, controlling for individual fixed effects, the results suggest that an increase of one year in the average schooling in a county implies a small positive increase on individual wages of about 1.7%. When the firm controls are included, this result flocks to zero. The author also proposes that, after 5 years of displacement, the wages of college graduate workers increase by more than 3% comparing to workers with basic schooling, if there is an increase of one year in the average education in their county.

Another study looking at educational spillovers issue, but at firm level, in the Portuguese context is the that of Martins and Jin (2010). They find evidence of high returns that range between 14% and 23% implying that the social returns to education are higher than the private returns, which may due to a transfer of skills from the workers with more formal education to the workers with less formal education. They also find evidence of significant wage spillovers to less–educated workers (i.e., their wages increase by 2% to 3% *per* extra year of education of workers in the same firm), but not for their more qualified colleagues.

In summary, there is lack of empirical literature on the external effect of the human capital, namely for Portugal, which justifies the objective of the present essay in assessing the effects of the aggregate human capital in a city/region on individual wages.

4.3 Theoretical framework

4.3.1 Perfect substitutability across workers

The research carried out in this chapter builds upon research by Rauch (1993), Acemoglu and Angrist (2001) and Dalmazzo and de Blasio (2007). The discussion below follows closely their formulation of the model and arguments. At the model's core resides a Mincerian wage equation augmented with a local human capital term.

The model is defined as follows. Each region j, where $j = 1, \ldots, J$, is treated as a competitive economy where a single output is produced, Y. Because Y is traded on

a global market, its price is the same everywhere and set equal to 1. The typical firm operating in region j is assumed to produce under a Cobb-Douglas technology, with constant returns to scale, employing capital, K, and effective labour units, L, such that its output is defined as

$$Y = A\left(HC_j\right) K^{1-\alpha} L^{\alpha} \tag{4.1}$$

with $0 < \alpha < 1$. The term $A\left(HC_j\right)$ is a function of local human capital HC_j and captures the spillovers effects of human capital in the productivity of region j; $A\left(HC_j\right)$ measures the productivity advantage that benefits a firm which operates on region j. One assumes that firms across regions share the production technology, $A\left(.\right) K^{1-\alpha} L^{\alpha}$.

The theoretical formulation assumes that $\dfrac{dA\left(HC_j\right)}{dHC_j} \geq 0$. Effective labour units are defined as

$$L = \sum_{i=1}^{N} \psi\left(h_i\right), \quad with \quad \psi'(h_i) > 0, \tag{4.2}$$

i.e., the firm hires N workers and each worker i provides ψ effective labour units. The effectiveness ψ of worker i is increased in his/her individual education, h_i. The additive to L implies perfect substitutability among the workers with different education levels (see Rauch, 1993; Ciccone and Peri, 2006). In each region j, the price of one effective labour unit is designated by w_j. The capital is rented in the market at a rate, r.

Each firm in the region j maximizes the profits, choosing (K, L), given the local level of human capital HC_j. As such, profit function is defined by:

$$\Pi = Y - rK - w_j L \tag{4.3}$$

considering equation (4.1) and substituting:

$$\Pi = A\left(HC_j\right) K^{1-\alpha} L^{\alpha} - rK - w_j L \tag{4.4}$$

Determining the first order conditions which maximize this problem and solve both in order to K and L, respectively, one obtains:

$$\frac{\partial \Pi}{\partial K} = 0 \Leftrightarrow (1-\alpha) A\left(HC_j\right) K^{-\alpha} L^{\alpha} - r = 0 \Leftrightarrow$$

$$\Leftrightarrow (1-\alpha) A\left(HC_j\right) L^{\alpha} = rK^{\alpha}$$

$$\Leftrightarrow K^{\alpha} = \frac{(1-\alpha) A\left(HC_j\right) L^{\alpha}}{r}$$

Multiplying both sides of the expressions by $K^{1-\alpha}$:

$$K = \frac{(1-\alpha)A\left(HC_j\right)K^{1-\alpha}L^{\alpha}}{r} \Leftrightarrow$$

$$\Leftrightarrow K = \frac{(1-\alpha)Y}{r} \tag{4.5}$$

$$\frac{\partial \Pi}{\partial L} = 0 \Leftrightarrow \alpha A\left(HC_j\right)K^{1-\alpha}L^{\alpha-1} - w_j = 0 \Leftrightarrow$$

$$\Leftrightarrow \alpha A\left(HC_j\right)K^{1-\alpha} = w_j L^{1-\alpha}$$

$$\Leftrightarrow L^{1-\alpha} = \frac{\alpha A\left(HC_j\right)K^{1-\alpha}}{w_j}$$

Multiplying both sides of the expressions by L^{α}:

$$L = \frac{\alpha A\left(HC_j\right)K^{1-\alpha}L^{\alpha}}{w_j} \Leftrightarrow$$

$$\Leftrightarrow L = \frac{\alpha Y}{w_j} \tag{4.6}$$

Substituting the expressions (4.5) and (4.6) in Cobb–Douglas technology function (4.1) is found the equilibrium value of w_j:

$$Y = A\left(HC_j\right)\left[\frac{(1-\alpha)Y}{r}\right]^{1-\alpha}\left(\frac{\alpha Y}{w_j}\right)^{\alpha} \Leftrightarrow$$

$$\Leftrightarrow Y = A\left(HC_j\right)\left(\frac{1-\alpha}{r}\right)^{1-\alpha}Y^{1-\alpha}Y^{\alpha}\left(\frac{\alpha}{w_j}\right)^{\alpha}$$

$$\Leftrightarrow {w_j}^{\alpha} = A\left(HC_j\right)\left(\frac{1-\alpha}{r}\right)^{1-\alpha}\alpha^{\alpha}$$

$$\Leftrightarrow w_j = \alpha\left(\frac{1-\alpha}{r}\right)^{\frac{1-\alpha}{\alpha}}\left[A\left(HC_j\right)\right]^{\frac{1}{\alpha}} \tag{4.7}$$

So, the equilibrium value of w_j, i.e., the local price of an effective labour unit is given by:

$$w_j = \mu\left[A\left(HC_j\right)\right]^{\frac{1}{\alpha}} \tag{4.8}$$

where $\mu = \alpha\left(\frac{1-\alpha}{r}\right)^{\frac{1-\alpha}{\alpha}}$.

If a greater level of local human capital increases the price of an effective labour unit in the region then $\frac{dw_j}{dHC_j} > 0$; if this is not the case we would have $\frac{dw_j}{dHC_j} = 0$. Substituting the equilibrium expressions for K and L in the profit function, Π, and

assuming that each firm will have zero profit in equilibrium, they have no incentives to change region.

The wage earned by the individual i in region j, denoted by w_{ij}, is defined by $w_j \times \psi(h_i)$, hence, applying logs, one gets

$$\log w_{ij} = \log \mu + \log \psi(h_i) + \frac{1}{\alpha} \log \left[A \left(HC_j \right) \right] \tag{4.9}$$

Assuming the log effective labour, ψ, and the externalities of the production $A\left(HC_j\right)$ are linear functions of individual's education and local human capital, respectively, one has:

$$\log \psi(h_i) = \phi + \beta h_i \tag{4.10}$$

and

$$\log A\left(HC_j\right) = \theta + \gamma HC_j \tag{4.11}$$

If private returns to education exist then $\beta > 0$; if local human capital generates positive spillovers on productivity then $\gamma > 0$, otherwise $\gamma = 0$. Substituting these two expressions in equation (4.9), a Mincerian wage equation augmented with a local human capital term is obtained:

$$\log w_{ij} = \delta + \beta h_i + \eta HC_j \tag{4.12}$$

where δ and η are constants, and are defined as $\delta = \log \mu + \phi + \frac{\theta}{\alpha}$ and $\eta = \frac{\gamma}{\alpha}$.

Similarly to Rauch (1993), Acemoglu and Angrist (2001) and Moretti (2004a), if the local human capital generates positive spillovers on productivity then $\eta > 0$. If this parameter is equal to zero, the model becomes the standard Mincerian equation, where the wage differences only depend on individual's education (plus other productivity variables, like experience, which are not explicitly discussed here for the sake of simplicity).

Finally, despite the incentives for workers to migrate, following wage differentials across regions, they do not do so, unless, for example, they are compensated for differences in local amenities or in the price of house renting.(Dalmazzo and de Blasio, 2007) In other words, underlying the model we have imperfect mobility of workers.

4.3.2 Imperfect substitutability across workers

A key assumption of the Mincerian approach is that workers with different levels of human capital are perfect substitutes in production. Perfect substitutability simplifies identification, because it implies that changes in the relative supply of human capital

does not affect the relative wages of different human capital groups, holding total factor productivity constant. The model proposed by Moretti (2004a) complements the previous approach, as it adopts the assumption that there is imperfect substitutability between the different types of workers. The paper adds that an increase in the share of skilled workers in the region raises the productivity of unskilled workers, and consequently, the level of wages. Such increase may be due to the existence of human capital spillovers. The model identifies the effect on wages of individuals with different education levels caused by an increase in relative supply of skilled workers in a city/region. Within this context, standard demand and supply considerations suggest that wages of unskilled workers benefit both from imperfect substitution, as well as, from the human capital spillovers. On the other hand, the wages of skilled workers are impaired by the increase supply of unskilled workers, but benefit from the externalities that are caused by higher aggregate human capital (Moretti, 2004a).

Moreti's model assumes that each city/region has a competitive economy that produces a single output Y traded in the national market. The model is based on a production function that uses two types of labour and physical capital inputs. Output in the region is assumed to be produced under a Cobb–Douglas technology that employs capital and skilled and unskilled workers,

$$Y = (\theta_0 N_0)^{\alpha_0} (\theta_1 N_1)^{\alpha_1} K^{1-\alpha_1-\alpha_0} \tag{4.13}$$

where N_0 is the number of workers with low education in the region; N_1 is the number of workers with high education in the region; K is capital; and the θ's are productivity shifters for workers type 0 and 1, respectively.[1] We futher assume constant returns to scale. The imperfect substitutability results from the multiplicative form of introducing the different worker types in equation (4.13).

It is assumed that the productivity of skilled and unskilled workers, θ_j, depends on their own human capital, ϕ_j (j=0, 1), with $\phi_1 > \phi_0$, as well as, on the ratio of skilled workers to total workforce in the region, which is given by:

$$\log(\theta_j) = \phi_j + \eta \left(\frac{N_1}{N_0 + N_1} \right), \qquad j = 0, 1 \tag{4.14}$$

where $s_{HE} = \dfrac{N_1}{N_0 + N_1} < 1$ is the share of skilled workers in region. With $\eta = 0$, the model is the standard Mincerian model without spillovers, i.e., individual's productivity depends only on individual's human capital, with skilled workers being by definition

[1] The Cobb–Douglas technology is chosen only for its simplicity; results do not change if a more general CES technology is used (Moretti, 2004a).

endowed with higher level of human capital; if $\eta > 0$ there are positive spillovers, as the increase of one unit in aggregate human capital impacts on worker's productivity. One can take η as a quantification of the effect on worker's productivity of an increase in the share of skilled workers.

If wages are equal to the marginal product for each type of labour and if θ is exogenous for firms (spillover is external to individual firms in the region, but internal to the region as a whole), firms' profit maximization derivation, similar to the one discussed in the previous section, combining the information from equations (4.13) and (4.14) leads to log wages for skilled and unskilled workers expressed, respectively, as

$$\begin{aligned}\log(w_1) = \log(\alpha_1) + \alpha_1 \log(\theta_1) + (1 - \alpha_1 - \alpha_0) \log\left(\frac{K}{N}\right) + \\ + (\alpha_1 - 1) \log(s_{HE}) + \alpha_0 \log(\theta_0(1 - s_{HE}))\end{aligned} \tag{4.15}$$

and

$$\begin{aligned}\log(w_0) = \log(\alpha_0) + \alpha_0 \log(\theta_0) + (1 - \alpha_1 - \alpha_0) \log\left(\frac{K}{N}\right) + \\ + (\alpha_0 - 1) \log(1 - s_{HE}) + \alpha_1 \log(\theta_1 s_{HE})\end{aligned} \tag{4.16}$$

where $N = N_0 + N_1$.

To discuss the effect of changes in region's share of skilled workers on worker's wage one combines the information from equations (4.13) and (4.14) and takes partial first order derivatives, obtaining

$$\frac{\partial \log(w_1)}{\partial s_{HE}} = \frac{\alpha_1 - 1}{s_{HE}} - \frac{\alpha_0}{1 - s_{HE}} + (\alpha_1 + \alpha_0)\eta \tag{4.17}$$

$$\frac{\partial \log(w_0)}{\partial s_{HE}} = \frac{1 - \alpha_0}{1 - s_{HE}} + \frac{\alpha_1}{s_{HE}} + (\alpha_1 + \alpha_0)\eta \tag{4.18}$$

According to these expressions, an increase in the share of skilled workers impacts on the wages of the unskilled workers in two ways. On the one hand, through neoclassical supply effects arising from the imperfect substitution between skilled and unskilled workers, $\frac{1 - \alpha_0}{1 - s_{HE}} + \frac{\alpha_1}{s_{HE}} > 0$, and on the other hand, through human capital spillover $(\alpha_1 + \alpha_0)\eta$, if $\eta > 0$. The impact of an increase in the supply of skilled workers on their own wage is determined by two opposing forces, i.e., the effect of the imperfect substitutability is negative, $\frac{\alpha_1 - 1}{s_{HE}} - \frac{\alpha_0}{1 - s_{HE}} < 0$, and the spillover that raises productivity is positive $(\alpha_1 + \alpha_0)\eta$, if $\eta > 0$. Therefore, for the skilled workers the sign of the effect on productivity and, consequently on wages, depends on the size of the spillover, but

to the unskilled workers, wages raise even without the presence of spillovers ($\eta = 0$). Finally, the formulation of the model implies that both skilled and unskilled workers benefit equally from the spillover (see equation (4.14)).

To conclude, the wage difference between worker types, specified in equations (4.15) and (4.16), depends on regions' specific human capital, summarized in s_{HE}, and on worker's own productivity, which is determined by his education type.

4.4 Methodology and data

4.4.1 Methodology

From the discussion of the theoretical framework outlined in Section 4.3, one can build a Mincerian wage equation augmented with a local human capital term, here defined as

$$\log w_{irt} = \alpha + \theta s_{irt} + \beta X_{irt} + \eta \bar{S_{rt}} + \delta Z_{rt} + d_{sa} + d_t + \varepsilon_{irt} \tag{4.19}$$

with w_{irt} denoting the hourly wage of individual i, in county r at time t. s_{irt} stands for individual's schooling, X_{irt} is a vector of worker's observable characteristics, which includes gender, experience and tenure, and their squares, a set of dummies that control for occupation, and firm size, measured by the number of employees and sales; $\bar{S_{rt}}$ represents regional human capital in the county r in time t. Z_{rt} is a matrix of county characteristics that may affect the concentration of human capital and wages, and d_{sa} and d_t control for the sector of activity and time specific effects, capturing the unobserved time–invariant factors of sector of activity and time that may influence wages; α is a constant and ε_{irt} is a white noise error term, for which we assume the standard assumptions of the classic regression model. The variable of interest is $\bar{S_{rt}}$ and its estimated coefficient will provide evidence on the existence, or not, of regional spillover effects of education. So, our main goal is to estimate η, i.e., the impact of local human capital on individual wages.

Different regional human capital measures can be characterized by different spillovers. In the discussion that follows, we will consider three specifications to measure the regional human capital. Two of these measures have been generally used in the literature, namely average years of education in the region (as in Rauch, 1993; Acemoglu and Angrist, 2001; Silva, 2003; Dalmazzo and de Blasio, 2007; Liu, 2007) and share of highly skilled workers in the region (as in Moretti, 2004a; Garcia-Fontes and Hidalgo, 2009; Bauer and Vorell, 2010).

In addition, we will bring a contribution to the existing literature by using a measure

for the aggregate human capital based on the skill index proposed by Portela (2001). So, the third formulation of $\bar{S_{rt}}$ hinges upon an adjusted skill index at county level, SI_r.[2] Taking into account different dimensions of the productivity, the skill index of a given county with n workers is given by:

$$SI_r = mschool\left(0.5 + \frac{e^{\frac{\overline{E_r} - mschool}{sschool}}}{1 + e^{\frac{\overline{E_r} - mschool}{sschool}}}\right) \times \left(0.5 + \frac{e^{\frac{\overline{exper_r} - mexper}{sexper}}}{1 + e^{\frac{\overline{exper_r} - mexper}{sexper}}}\right) \tag{4.20}$$

where $\overline{E_r} = \frac{\sum_{i=1}^{n_r} s_{ir}}{n_r}$ is the average schooling of county r and s_{ir} is the school level of worker i located in county r, evaluated as years of schooling, *mschool* represents the average of school level in the population and *sschool* denote its standard deviation; $\overline{exper_r} = \frac{\sum_{i=1}^{n_r} exper_{ir}}{n_r}$ is the average experience of the county and $exper_{ir}$ is the experience of the worker i in county r, *mexper* and *sexper* represent the average and the standard deviation of experience for the whole economy, respectively. In one of the regression models discussed below, this skill index will be adapted at firm level, in which case firms 'stand' for county.

Concerning the control variables at county level, and following the empirical literature, a proxy for unemployment rate in the county is included, as the correlation of education with earnings might be affected by the distribution of unemployment across counties. If the most skilled individuals are less likely to be unemployed, then the aggregate human capital variable may capture the effect of the unemployment rate. This variable also captures the local conditions of the labour market in different periods, which will be useful once we explore the panel dimension of the data.

A control variable for the average firm–size at county level is also included, which can be taken as a proxy for local competitiveness (Dalmazzo and de Blasio, 2007). The compensatory effect of amenities on wages is controlled in the model by including the attractiveness index of the county, which is correlated with the welfare of the individuals; the greater the positive amenities of a county, the higher its attractiveness.[3] Regions with better amenities tend to attract more skilled workers with higher potential and those may be more able to learn, which makes them more productive (Rauch, 1993; Moretti, 2004a,b; Combes et al., 2011). Detailed information on the variables is provided in Appendix, Table B.1.

Furthermore, the impact of local human capital on wages could also reflect agglom-

[2] See Appendix B.1 for more details about this index.

[3] E.g., Heuermann (2011) uses the number of hotel beds in a county, as a proxy for a region's amenity endowment, since a large number of hotel beds indicates that it is attractive for people to travel into that region, be it out of leisure or business motives.

eration effects (Dalmazzo and de Blasio, 2007), since the larger the dimension of the local population, the most benefits of agglomeration associated with better market linkages are presented and, therefore, might make more productive workers (Heuermann, 2011). So, to control for local characteristics that may induce economies of agglomeration, the log population of the county is included. Finally, to control county's price level, the ratio of the mean price of evaluation of accommodation per square meter between each county and the county of Lisbon is included. This variable is expected to measure the influence of changes in the cost of living at the county level in the period considered.

Having defined the empirical model in equation (4.19), we now will discuss different estimation strategies that can be used to estimate the parameters of interest. The natural starting point is to estimate by OLS the effect of the aggregate human capital at regional level on individual log hourly wage, conditional on worker's own education, and the regressors discussed above.

However, there are methodological issues that must be addressed. On the one hand, there may be unobserved factors that are correlated with individual wages and education, i.e., one source of bias is associated with the omission of unobservable characteristics of workers that may influence their productivity and, therefore, their wages. On the other hand, due to the correlation between wages and human capital in a county, for example, workers born in regions with the best schools and universities and, thus, integrated into an environment favourable to accumulation of human capital, are more productive (Combes et al., 2011), leading to migration across counties, which implies that regional schooling levels may be endogenous (Moretti, 2004a,b). Under such conditions, the OLS estimation of Mincerian equation (4.19) for the identification of human capital externalities leads to biased estimates of the coefficients of interest, which implies that one needs alternative solutions to retrieve the parameters.

First, the panel structure of data allows to control for individual time–invariant unobserved factors that may be correlated with education, eliminating part of the bias which can influence the coefficients associated with human capital variables. The benefits accrued to longitudinal data include higher variability of the regressors, improved efficiency, as well as additional degrees of freedom, and less collinearity among the variables (Baltagi, 2008). We can now re–define the Mincer equation described in equation (4.19) as build the following fixed–effects (FE) model

$$\log w_{irt} = \theta s_{irt} + \beta X_{irt} + \eta \bar{S}_{rt} + d_{sa} + \alpha_i + \varepsilon_{irt} \tag{4.21}$$

An important characteristic of the FE model, the specific solution implemented in the empirical analysis, is that some of the explanatory variables may be correlated with

worker's ability, captured under this specification by the individual–specific effect (α_i). Despite such source of endogeneity, the FE estimators guarantees consistent estimates for the parameters of interest. For instance, even if the unobserved individual's ability α_i is simultaneously correlated with individual's schooling and wage, the estimates will still be correct, as long as the correlation between education and the error goes through the individual–specific fixed effect and one assumes that unobserved ability doesn't vary over time.

If, additionally, one assumes that all variables included in the model are uncorrelated with α_i, and, naturally, with ε_{irt}, the estimates obtained by random–effects estimation are both consistent and efficient – such solution corresponds to a Feasible Generalized Least Squares (FGLS). However, taking into account that some variables may be endogenous, e.g., individual educational level and regional human capital, the application of the fixed–effects estimator is the appropriate one, as the correlation between some of the X's the specific effect renders the random–effects estimator inconsistent.[4] In other words, a fixed-effects model becomes one of the appropriate methodologies to deal with workers', counties' or firms' unobserved heterogeneity.

In implementation of the fixed–effects estimation, control variables for the characteristics of the county will not included, because in the sample only a negligible number of workers moved from their county.[5] This implies that Z_{rt} in equation (4.19) is not included in equation (4.21). Panel data allows one to control for units' – individuals and counties – unobserved time–invariant characteristics. However, under this data restriction, we will not be able to disentangle individuals' and counties' specific effects.

Experience is an individual specific time trend, which implies that one cannot combine fixed–effects, experience and time dummies. As such, in this model the set of time dummy variables is dropped, so d_t from equation (4.19) does not show up in equation (4.21).[6]

Second, an alternative method discussed in the literature to deal with endogeneity of education is the IV (instrumental variables). Endogeneity of both own education and county's education can be jointly tackled when there is an instrument for local human capital and another for individual education. The advantage of the IV method is that a valid instrument isolates the effect of exogenous changes in human capital levels on wages (Moretti, 2004a,b).

As previously mentioned, the OLS estimates might be biased due to endogenous

[4] In the empirical analysis discussed below appropriate statistical tests will be implemented to evaluate the validity of the assumptions underlying each estimation strategy. In this particular case, a Hausman test will allow us to evaluate to what extent the random-effects model could be an appropriate estimator.

[5] Workers moving across counties represent only about 5.7% of the sample.

[6] See the discussion on collinearity between *experience* and time dummies in the Appendix B.2.

location choices of individuals and unobservable characteristics of the county, as well as worker's unobserved heterogeneity. A key issue when estimating human capital externalities at regional level is that changes in aggregate human capital levels are endogenous as counties with higher productivity and higher wages may attract more high skilled workers (Moretti, 2004a,b; Dalmazzo and de Blasio, 2007; Garcia-Fontes and Hidalgo, 2009). Another factor that may work in the same direction is that high income regions may have amenities that are especially attractive for high skilled workers. In this case, the coefficient of local human capital should be upward biased.

An instrument that is correlated with human capital in a region and uncorrelated with unobserved factors that affect wages directly is needed. The instrument must take into account for the observed variation in the local human capital, but not correlated with the residual of the earnings equation, i.e., should be orthogonal to current individual wages (Moretti, 2004a).

In this essay, two instruments are considered. Following Moretti (2004a), the first instrument proposed is the presence of a Polytechnic Institute in a county. A second instrument is the share of people who can read in the county based on the data of the 1900 Census, which is something new in empirical literature.

According to the literature, the presence of a Polytechnic Institute in a county tends to increase the education of this county, since the costs to attend higher education are smaller (there are no costs for housing, for example). At the same time, people tend to fix themselves to work, where they live and study (Moretti, 2004a).

The presence of a Polytechnic Institute in a county was established by the so–called educational reform of *Veiga Simão*, in the 1970s.[7] The validity of the exogeneity of the instrument occurs as long as workers in a county where there is a Polytechnic Institute are not systematically different from other workers with the same education in a county where there is no Polytechnic Institute. At the same time, decisions for the installation process of the polytechnic educational system were political and took place about three decades before our sample period. For these reasons, the presence of a Polytechnic Institute is likely to be uncorrelated with unobservable factors that affect wages in the period under study.

Concerning the second instrument proposed, the underlying idea is that there is an intergenerational persistence of taste for studying. This is similar to the argument in favour of using family background characteristics as an instrument for an individual's education. This variable affects the education in the county, because in counties where there were more literate people in 1900 and where there were probably a greater number of families able to provide education for its members through financial support, making

[7]See Chapter 2 for more details about educational reform of *Veiga Simão*.

the county more educated. In other words, the current education of the county, in principle, was affected by the influence of the valuation of education that passes from generation to generation. At the same time, the share of people who can read in a county in 1900 is likely uncorrelated with unobservable factors that affect wages during 2002–2009. Thus, the share of the people who can read in the county in 1900 may, theoretically, be used as a reliable instrument for human capital in the county.

In addition, individuals' education is also treated as an endogenous variable and the change in the compulsory school attainment in 1981, which established the compulsory education of 9 years, is used as an instrument. Changes in compulsory schooling are common instruments used in the empirical analysis (see Angrist and Krueger, 1991; Acemoglu and Angrist, 2001).

As previously, the specification to estimate by IV method includes the individual and firms characteristics, control dummies for sector of activity and time effects (see equation (4.19), excluding county's controls).

Third, to improve estimates of our parameters, one further explores the panel structure of the data and combines IV with panel data methodologies. Particularly, we have used the instrumental variables fixed–effects estimation (IV–FE), which is applied to circumvent the potential endogeneity problem that remained after controlling for fixed–effects. The instruments are the ones discussed above, namely, the presence of a Polytechnic Institute in a county, and the share of people who can read in the county based on the data of the 1900 Census. Note that, control variables for the county are not used due to its collinearity with these instruments. One should add that the instrumental variables are included in order to control for the impact of unobservable time–region specific factors (Heuermann, 2011). The estimated model is defined in equation (4.21).

Fourth, taking into account the advantages mentioned above for longitudinal data, and possible uncertainty associated with the choice of instruments for the regional human capital variables, the IV–type model proposed by Hausman and Taylor (1981) is also used in our empirical analysis. The motivation for the selection of this procedure is, first, the fact that fixed–effects may not capture fully the endogeneity in the education variables. Second, one does not have to use external instruments, since the means of the strictly exogenous regressors within each county are used as instruments for the time–invariant regressors.[8]

[8]The Hausman–Taylor estimation is only possible if there are at least as many exogenous time–varying variables as endogenous time–invariant variables(Baltagi et al., 2003; Verbeek, 2012; Wooldridge, 2010). If is the case, the choice of the strictly exogenous time–varying variables is a testable hypothesis. A Hausman test can be computed to compare the Hausman–Taylor and the fixed–effects estimator. This test compares the consistent but inefficient fixed–effects estimates of parameters with the Hausman–Taylor estimates, which under the null hypothesis are consistent and efficient (Baltagi et al., 2003).

In the empirical analysis discussed below, these four procedures are applied only to the Mincerian wage regression as defined in equations (4.19) and (4.21). When one includes the firm's aggregate human capital or when one includes neighbour's aggregate human capital we only consider OLS and FE estimation procedures.

A final issue must be addressed. A difficulty related with the identification of the causal impact of local human capital concentration on wages arises when workers with different levels of education are imperfect substitutes within the production function (Moretti, 2004a,b; Ciccone and Peri, 2006). As mentioned in Section 4.3.2, under imperfect substitutability, an increase in the number of skilled workers implies an increase of the wages of unskilled workers, as they benefit from county's education level, and, at the same time, are combined with more skilled workers within the same firm. However, for skilled workers imperfect substitutability implies that their productivity relies on the lower productivity of less educated workers, so the overall effect of aggregate education will depend on how positive external effects offsets the decrease in the wage of skilled workers due to their combination with less productive workers (Moretti, 2004a,b).

Whenever workers with different levels of education are imperfect substitutes in production, the parameter η will comprise effects that can be obtained by "composition effects", i.e., due to a larger share of skilled workers which increases the productivity of both types of workers and by genuine spillovers, so, due to human capital externalities. As a consequence, the literature indicates that the estimated value of η must still be positive, even if spillovers are zero (Dalmazzo and de Blasio, 2007). According to the empirical literature, the role of imperfect substitution can be studied by estimating regressions for separate education groups, the less qualified workers and highly qualified workers (Moretti, 2004a).

4.4.2 Data description

The data used in this essay is taken from four distinct Portuguese sources. Individual data on wages, age, gender, education, experience and further controls, as well data at the firm level, such as sector of activity, geographic location, real sales volume and firm size, are provided by *Quadros de Pessoal*. *Quadros de Pessoal* (QP) is an annual database, collected by the Ministry of Labour, Solidarity and Social Security since 1982, by mandatory surveys applied to all establishments with at least one employee. Public administration workers, armed forces, self-employed, as well as household workers, are not included in this database.[9]

[9] A description and discussion of the data assembling process is provided in Appendix A.1. In the current chapter we only use data for the period 2002–2009.

Information from *Quadros de Pessoal* was augmented by regional variables of the 278 mainland Portuguese counties, that may affect the concentration of human capital, as well as the wages, as previously discussed. The unemployment rates at county level are not available, thus a proxy for unemployment rates is computed using information for the working population (2001 and 2011 Census) and the number of unemployed people registered at Portuguese Employment Centres. These data are provided by *Instituto Nacional de Estatística* (INE) and *Instituto de Emprego e Formação Profissional – Ministério do Trabalho, Solidariedade e Segurança Social* (IEFP), respectively. The data at county level, such as population by county and the general development index – index of attractiveness – are provided by Marktest Consulting – Sales Index 2013.[10] Finally, data from the 1900 Census and information of the mean price of evaluation of accommodation per square meter by county is provided by INE.

An unbalanced panel of workers, their firms and counties where they work for the years 2002 to 2009 is used. The descriptive statistics of the variables are reported in Table 4.1.[11] The final sample comprises full time employees, aged between 16 and 64 years old. Workers whose wage is below the national minimum wage were excluded.[12] The sample used in the current analysis includes more than 14 million observations worker–year on 3,668,129 unique workers, working in 458,493 unique firms distributed across 278 Portuguese counties.

Over the sample period the average real hourly wage was 5.83 Euros, but there are substantial variations along the wage distribution. 5% of workers earn a hourly wage lower than 2.45 Euros, while at the top of the distribution 5% of workers earn per hour above 15 Euros. 57% of the workers in our sample are males, the average age is 38 years and the mean seniority at the firm is nearly 8 years. Considering education, 66% of the workers have only up to 9 years of schooling and about 13% of the workforce has a university degree. Noting that, in the latest year of the sample, 2009, the average real hourly wage was 6.23 Euros, and 5% of workers earn a hourly wage lower than 2.68 Euros, while at the top of the distribution 5% of workers earn per hour above 16 Euros. In 2009, 56% of the workers are males, the average age is 39 years and the mean seniority at the firm is nearly 8 years. In addition, 61% of the workers have only up to

[10]Salex Index 2013 is a database updated annually, organized at county level, which Marktest Consulting has collected and provided since 1992. This database results from continuous contact with official entities, as well as with several private entities, whose collaboration offers relevant county level indicators, namely related to macroeconomics, demographics, business structure, economic activity and the regional development index created exclusively by Marktest Consulting.

[11]See Table B.1 in Appendix B.3 for a description of the variables.

[12]According to Article no. 275 of the Portuguese Labour Code, employers can pay a wage below the national minimum wage for apprentices, practitioners, trainees or participating in certified training, as well as for workers with reduced work capacity.

Table 4.1: Descriptive statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
Individual–level variables					
Real hourly wage (Euro)	14,569,858	5.831	5.828	1.262	1,390
Female	14,569,858	0.429			
Education	14,569,858	8.689	3.968	0	16
≤ 1st cycle		0.235			
2nd cycle		0.209			
3rd cycle		0.213			
Secondary		0.216			
Tertiary		0.127			
Age	14,569,858	38.17	10.72	16	64
Tenure	14,569,858	7.918	8.325	0.083	52.58
Experience	14,566,873	21.00	11.21	0	48
Occupation	14,480,711				
Executive/managerial		0.041			
Intermediate managerial/executive		0.060			
Low managerial		0.118			
Technicians highly specialized		0.164			
Sales, administ. and precision prod.		0.162			
Administ. support and production		0.012			
Unskilled		0.218			
Apprentice		0.112			
Unknown		0.113			
Firm–level variables					
Average schooling	458,493	7.932	3.258	0	16
Share of most qualified workers	458,493	26.13	39.34	0	100
Firm Skill Index	159,414	7.076	2.716	2.172	20.09
Size	458,493	8.082	67.94	1	17,504
Real sales (1000 Euros)	319,017	973,021	1.59e+07	1745	2.44e+09
County–level variables					
Population	278	1,164	29,427	872	267138
Average schooling	278	7.608	0.893	5.553	10.88
Share of most qualified workers	278	23.08	8.331	6.424	58.08
County Skill Index	278	6.846	1.278	4.378	12.42
Attractiveness Index (log form)	278	6.984	0.3291		
Housing price (ratio)	278	46.63	11.25	32.49	100
Unemployment rate	278	4.332	1.963	0.666	11.44
Average size of firms	278	7.857	3.466	1.75	24.54

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting – Sales Index 2013, INE and, *Instituto de Emprego e Formação Profissional – Ministério do Trabalho, Solidariedade e Segurança Social*, 2002–2009.

9 years of schooling and about 16% of the workforce has tertiary education.[13]

Regarding the observed characteristics of firms, the mean of the real sales volume of firms in our sample is about 973,021 thousands Euro and firms employ, on average, 8 workers. Regarding the educational level within firms, the average schooling of firms

[13]See Table B.2 in Appendix B.3.

is about 8 years and the mean of the share of most educated workers within firms is 26%. The mean skill index is about 7 and this index ranges between 2 and 20.

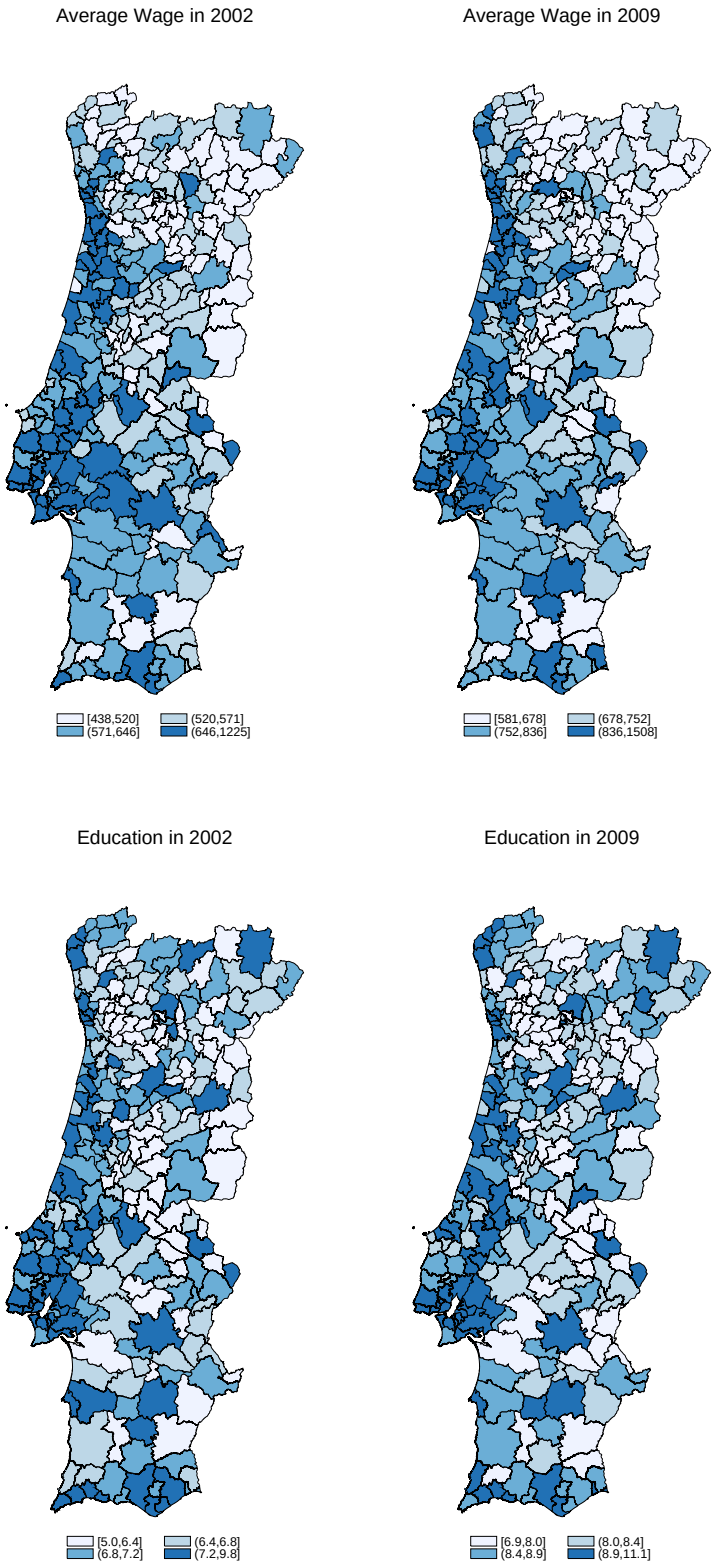
At county level, highlighting the human capital variables, over the period 2002–2009, the mean of education in the counties is about 7.6 years of schooling and the average of the share of workers who have at least the 12^{th} grade is about 23%. The mean of the skill index across counties is about 7 and this index varies between 4 and 12.

Figure 4.1 shows both wage and educational inequalities, between Portuguese counties, at the beginning and end of the period under study, 2002 and 2009, respectively. This figure shows that there are marked educational and wage asymmetries between the coastline and the interior of Portugal, which remains throughout the years. Portuguese counties positioned in the lower quartile of the distribution of education have an average schooling up to 6.4 in 2002, for instance, *Cinfães, Fornos de Algodres, Freixo de Espada à Cinta, Idanha–a–Nova, Penedono, Tarouca* and *Vizela*. In 2009, 25% of the mainland Portuguese counties had average schooling up to 8.0, such as *Alter do Chão, Crato, Fornos de Algodres, Gavião, Marvão, Penamacor* and *Felgueiras*. Noting that, except *Felgueiras*, these counties are located in the interior of the country. On the contrary, those positioned in the highest quartile, in which the average schooling is higher than 7.2 (in 2002) and 8.9 (in 2009) are located in the coastline, for instance, Lisbon and *Oeiras*.

With such differences in education levels at the county level, it does not come as a surprise that the average wages diverge across Portuguese regions, since practically all the counties have the same trend of average education. Counties where the mean wage is low are, for instance, *Freixo de Espada à Cinta, Gavião, Penedono, Penamacor* and *Vizela* and, in contrast, Lisbon and *Oeiras* are the counties where the mean wage is higher.

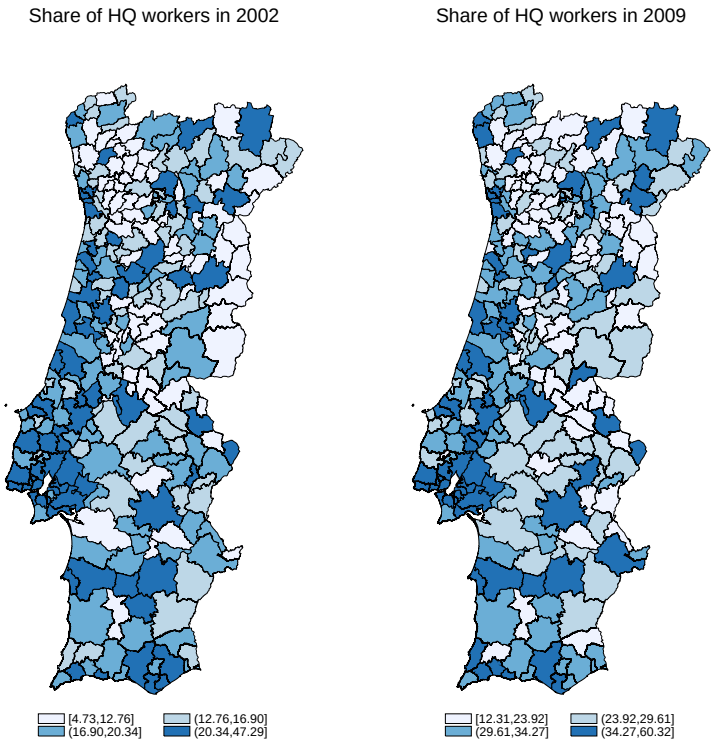
Figures 4.2 and 4.3 present a perspective of the distribution of human capital by county as measured by the share of highly qualified workers within a county's workforce, who have 12 years of schooling or more, and by the skill index, as defined in equation (4.20), which takes into account different dimensions of the productivity of the worker, respectively.

Looking to Figure 4.2, the share of the most qualified workers increases with the number of workers in a county (the linear correlation coefficient of Pearson is 0.81 in 2002, 0.76 in 2009 and 0.74 from 2002 to 2009), and this share ranges from 4.7% to 47.2% in 2002 and from 12.3% and 60.3% in 2009. Note that, counties with low values of share of the most qualified workers are mostly located in the interior of Portugal, such as, *Arronches, Cinfães, Freixo de Espada à Cinta, Gavião, Paços de Ferreira,*



Source: Created by the author based on *Quadros de Pessoal.*

Figure 4.1: Average schooling and average wages, by county, in 2002 and 2009



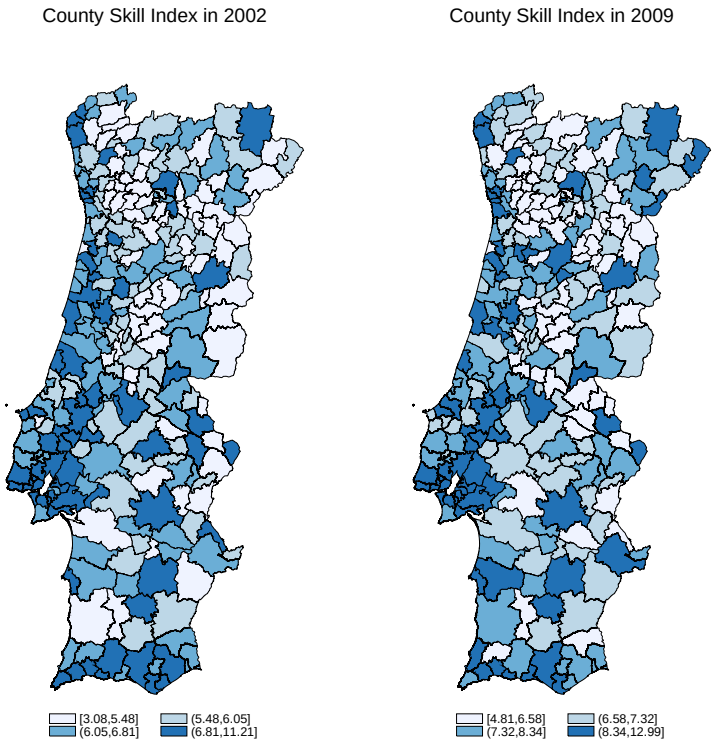
Source: Created by the author based on *Quadros de Pessoal*

Figure 4.2: Educational attainment by county, as percentage of most qualified workers in 2002 and 2009

Tarouca and *Vizela* in 2002, and *Felgueiras, Lousada, Marvão, Pampilhosa da Serra, Penamacor* and *Vizela* in 2009. In contrast, the counties with the highest value of share of the most qualified workers are Lisbon followed by *Oeiras*, in both years 2002 and 2009.

With respect to Figure 4.3, the skill index by county ranges between 3.50 and 11.05 in 2002, and between 5.42 and 12.15 in 2009. The counties which have a value for the index below the 1^{st} percentile are *Baião*, *Barrancos*, *Cinfaẽs*, *Felgueiras*, *Manteigas*, *Tarouca* and *Vizela*, in both years 2002 and 2009. As before, the counties with the highest skill index are *Oeiras* and Lisbon. There is also a positive correlation between this skill index and the number of workers in a county (the linear correlation coefficient of Pearson is 0.80 in 2002, 0.74 in 2009 and 0.75 in the period 2002–2009).

Note that, the county's mean wage, as well as the concentration of human capital, are higher on the coastline than on the interior of Portugal and this asymmetry remains throughout the years, regardless of the measure of aggregate human capital used.



Source: Created by the author based on *Quadros de Pessoal*

Figure 4.3: County skill index, in 2002 and 2009

4.5 Empirical results

4.5.1 Human capital externalities at county level

4.5.1.1 Ordinary least squares results

In a first stage, Mincerian equation augmented with a local human capital term is estimated by pooled OLS for all workers in the sample for the period 2002–2009. The estimated function is the one defined in equation (4.19). Additionally, to address imperfect substitutability using the Portuguese data, as discussed in Moretti (2004a), in our empirical analysis two groups of workers will be formed. The less qualified workers are those with schooling equal or below 12 years. Taking into account for the information available in the data, specifically, reduced variability in the education level within workers with tertiary education, the group of the most qualified workers will include both those with 12 years of schooling, and the ones with tertiary education.

First, in order to test the validity of homoscedasticity assumption, the Pagan–Hall test was applied and the null hypothesis that the error variances are all equal was rejected. For the purpose to control heteroscedasticity and possible serial correlation

between firms, in all estimations robust standard errors, clustered at the firm level, are applied.

For the three groups of workers, the results by pooled OLS are reported in Tables 4.2.1 – 4.2.3 and each table relates to each aggregate human capital measure, average years of education in the county, share of most qualified workers and the skill index of the county, respectively. Columns with odd number report OLS estimates of the specification which includes the individual and firms characteristics and columns with an even number report the OLS estimates of the specification that includes regional variables.

The results show that the coefficients obtained, for both individual and firm characteristics, present the expected sign with statistical significance at the 1% significance level. The positive coefficient of individual education is in accordance with the human capital theory, indicating that each year of schooling increases wages by about 4.5%, on average, and, as expected, returns to education are higher for the most qualified workers; on average, an additional year of schooling of an individual yields an increase of about 3% on wages for less qualified workers and of about 10% for most qualified workers.

Tenure and experience display the usual non–linear impact on wages, confirming the theory predictions of decreasing marginal returns to tenure and experience. Taking the results in Table 4.2.1, column (1), one observes that the estimated coefficient of experience is 0.0205 and the coefficient of its squared is -0.0003, indicating that, on average, the return to the first year of experience is about 2.1%, as well that after 20 years of experience, the return to an additional year of experience is just about 0.9%.[14] The discussion on returns to experience is just indicative of positive and decreasing marginal returns, as returns to experience and tenure cannot be disentangled under the current formulation of the model (see, for example, Altonji and Shakotko, 1987; Topel, 1991; Buhai et al., 2014).

As expected, the results also show that probably we are facing a labour market where gender discrimination is still the norm, because women's wages are around 15% lower than men's wages, conditional on other labour market characteristics, such as experience or firm size. But this discrimination is higher for less qualified women; the most qualified women earn about 13.5% less than men and the less qualified female workers earn about 16% less than men. Furthermore, the results show that "Financial intermediation" is the only sector of activity with statistically significant wage premium comparing with the base–industry "Mining" (this result is not reported in the tables; it will be provided upon request). In addition, the results suggest that an increase of 10%

[14]The computation is given by $0.0085 = 0.0205 - 0.0003 \times 2 \times 20$.

Table 4.2.1: Pooled OLS for 3 groups of workers: all, less qualified and most qualified, using average education of the county as a measure of human capital

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	All workers		Less qualified		Most qualified	
Education	0.0450***	0.0449***	0.0308***	0.0305***	0.0976***	0.0977***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0013)	(0.0013)
Av. education	0.0379***	0.0253***	0.0375***	0.0259***	0.0467***	0.0376***
	(0.0018)	(0.0033)	(0.0018)	(0.0033)	(0.0024)	(0.0053)
Female	-0.1551***	-0.1528***	-0.1620***	-0.1598***	-0.1354***	-0.1344***
	(0.0029)	(0.0033)	(0.0029)	(0.0033)	(0.0037)	(0.0040)
Experience	0.0205***	0.0205***	0.0175***	0.0175***	0.0294***	0.0295***
	(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.0005)	(0.0005)
Experience sq.	-0.0003***	-0.0003***	-0.0002***	-0.0002***	-0.0004***	-0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Tenure	0.0141***	0.0141***	0.0131***	0.0133***	0.0208***	0.0206***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0007)
Tenure squared	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0004***	-0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Log size	-0.0277***	-0.0282***	-0.0294***	-0.0298***	-0.0231***	-0.0236***
	(0.0044)	(0.0045)	(0.0044)	(0.0044)	(0.0053)	(0.0053)
Log sales	0.0724***	0.0720***	0.0707***	0.0706***	0.0768***	0.0763***
	(0.0031)	(0.0031)	(0.0032)	(0.0032)	(0.0036)	(0.0036)
Log aver. size	No	0.0082**	No	-0.0003	No	0.0133**
		(0.0042)		(0.0040)		(0.0062)
Log population	No	0.0018	No	0.0037*	No	-0.0034
		(0.0022)		(0.0021)		(0.0035)
Log attrac. index	No	-0.0261***	No	-0.0334***	No	-0.0156**
		(0.0047)		(0.0047)		(0.0068)
Housing price	No	0.0015***	No	0.0018***	No	0.0010***
(ratio)		(0.0002)		(0.0002)		(0.0003)
Unemployment	No	-0.0037***	No	-0.0037***	No	-0.0022
rate		(0.0009)		(0.0009)		(0.0014)
Observations	12,584,353	12,562,901	11,135,257	11,114,763	4,137,302	4,133,676
R-squared	0.632	0.633	0.547	0.549	0.636	0.637
RMSE	0.333	0.333	0.307	0.306	0.389	0.389

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting – Sales Index 2013, INE and, *Instituto de Emprego e Formação Profissional–Ministério do Trabalho, Solidariedade e Segurança Social*, 2002–2009.
Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include a set of dummies to control for occupation, sector of activity and year.

in sales per worker is associated with an average increase of 0.5% on wages, i.e., one may argue that more competitive firms pay better wages to their workers. Alternative hypothesis may be pointed out, e.g., rent sharing.

With respect to the variable of interest, local human capital, strong evidence for the existence of human capital externalities is found, as the coefficients of regional human capital variables are positive and highly significant in all regressions. Controlling for

Table 4.2.2: Pooled OLS for 3 groups of workers: all, less qualified and most qualified, using the share of highly qualified workers in a county as a measure of human capital

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	All workers		Less qualified		Most qualified	
Education	0.0450***	0.0449***	0.0309***	0.0305***	0.0976***	0.0977***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0013)	(0.0013)
Share of HQ workers	0.0034***	0.0026***	0.0034***	0.0026***	0.0042***	0.0037***
	(0.0002)	(0.0003)	(0.0002)	(0.0003)	(0.0002)	(0.0005)
Female	-0.1550***	-0.1529***	-0.1619***	-0.1598***	-0.1354***	-0.1346***
	(0.0029)	(0.0033)	(0.0029)	(0.0033)	(0.0037)	(0.0040)
Experience	0.0205***	0.0205***	0.0175***	0.0174***	0.0294***	0.0294***
	(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.0005)	(0.0004)
Experience sq.	-0.0003***	0.0003***	-0.0002***	-0.0002***	-0.0004***	-0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Tenure	0.0141***	0.0141***	0.0131***	0.0133***	0.0208***	0.0206***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0007)	(0.0007)
Tenure squared	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0004***	-0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Log size	-0.0277***	-0.0281***	-0.0295***	-0.0297***	-0.0232***	-0.0236***
	(0.0045)	(0.0045)	(0.0044)	(0.0045)	(0.0053)	(0.0053)
Log sales	0.0724***	0.0719***	0.0708***	0.0705***	0.0768***	0.0762***
	(0.0031)	(0.0031)	(0.0032)	(0.0032)	(0.0036)	(0.0036)
Log aver. size	No	0.0083**	No	-0.0002	No	0.0136**
		(0.0042)		(0.0041)		(0.0062)
Log population	No	0.0027	No	0.0045**	No	-0.0022
		(0.0022)		(0.0021)		(0.0035)
Log attrac. index	No	-0.0309***	No	-0.0374***	No	-0.0219***
		(0.0048)		(0.0049)		(0.0070)
Housing price (ratio)	No	0.0014***	No	0.0017***	No	0.0010***
		(0.0002)		(0.0002)		(0.0003)
Unemployment rate	No	-0.0036***	No	-0.0036***	No	-0.0019
		(0.0009)		(0.0009)		(0.0014)
Observations	12,584,353	12,562,901	11,135,257	11,114,763	4,137,302	4,133,676
R-squared	0.632	0.633	0.547	0.549	0.636	0.637
RMSE	0.333	0.333	0.307	0.306	0.389	0.389

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting – Sales Index 2013, INE and, *Instituto de Emprego e Formação Profissional–Ministério do Trabalho, Solidariedade e Segurança Social*, 2002–2009.
Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include a set of dummies to control for occupation, sector of activity and year.

workers' and firms' characteristics, column 1 in Tables 4.2.1 – 4.2.3, on average, an extra year in average education of the county is associated with an increase of about 3.8% on wages, or an increase in the share of highly qualified workers in the county by 1 p.p. increases wages by 0.3%. These results are in line with the empirical literature (e.g. Rauch, 1993; Moretti, 2004a; Isacsson, 2005; Canton, 2009; Heuermann, 2011). The county skill index also shows that an additional year of calibrated education in

Table 4.2.3: Pooled OLS for 3 groups of workers: all, less qualified and most qualified, using the skill index of county as a measure of human capital

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	All workers		Less qualified		Most qualified	
Education	0.0456***	0.0453***	0.0314***	0.0309***	0.0982***	0.0982***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0013)	(0.0013)
Skill index	0.0206***	0.0137***	0.0203***	0.0139***	0.0256***	0.0208***
(county)	(0.0010)	(0.0017)	(0.0010)	(0.0017)	(0.0013)	(0.0032)
Female	-0.1548***	-0.1525***	-0.1618***	-0.1596***	-0.1353***	-0.1340***
	(0.0029)	(0.0033)	(0.0029)	(0.0033)	(0.0037)	(0.0040)
Experience	0.0204***	0.0204***	0.0174***	0.0174***	0.0294***	0.0294***
	(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.0005)	(0.0005)
Experience sq.	-0.0003***	-0.0003***	-0.0002***	-0.0002***	-0.0004***	-0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Tenure	0.0141***	0.0141***	0.0131***	0.0134***	0.0208***	0.0206***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0007)	(0.0007)
Tenure squared	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0004***	-0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Log size	-0.0277***	-0.0282***	-0.0295***	-0.0298***	-0.0231***	-0.0236***
	(0.0045)	(0.0045)	(0.0044)	(0.0045)	(0.0053)	(0.0053)
Log sales	0.0725***	0.0721***	0.0708***	0.0706***	0.0769***	0.0763***
	(0.0031)	(0.0031)	(0.0032)	(0.0032)	(0.0036)	(0.0036)
Log aver. size	No	0.0088**	No	0.0002	No	0.0143**
		(0.0042)		(0.0044)		(0.0063)
Log population	No	0.0023	No	0.0044**	No	-0.0040
		(0.0022)		(0.0021)		(0.0035)
Log attrac. index	No	-0.0266***	No	-0.0342***	No	-0.0162***
		(0.0047)		(0.0047)		(0.0068)
Housing price	No	0.0015***	No	0.0018***	No	0.0010***
(ratio)		(0.0002)		(0.0002)		(0.0003)
Unemployment	No	-0.0037***	No	-0.0038***	No	-0.0026*
rate		(0.0009)		(0.0009)		(0.0014)
Observations	12,584,009	12,562,557	11,134,955	11,114,461	4,137,260	4,133,634
R-squared	0.632	0.633	0.547	0.549	0.636	0.636
RMSE	0.333	0.333	0.307	0.306	0.389	0.389

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting – Sales Index 2013, INE and, *Instituto de Emprego e Formação Profissional–Ministério do Trabalho, Solidariedade e Segurança Social*, 2002–2009.
Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include a set of dummies to control for occupation, sector of activity and year.

the county, as defined in equation (4.20), leads to a wage premium of about 2.1%, but can not be compared with other results in literature.[15]

As previously discussed, these estimated external returns only present the average

[15]In order to interpret the parameter estimate associated with county's skill index reported in Table 4.2.3, one has to combine the information available in Table 4.1 with the definition of the skill index provided in equation (4.20). From Table B.2, one observes that in 2009 the average skill index is about 8, with standard–deviation of 1.3, and bounded between 5.1 and 12.8.

external effect of human capital across different educational groups and may just be a reflection of the imperfect substitutability, because they are obtained with all individuals who have different educational qualifications. In both qualification groups of workers, there is evidence that the externalities effect is positive and significant and there is also evidence that local human capital has a marginally larger effect on the wages of the most qualified workers than on wages of the less qualified workers (see columns 3 and 5 in Tables 4.2.1 – 4.2.3). The OLS estimates for the less qualified group reflect the sum of positive complementarity and spillover effects and for the most qualified group the external returns reflect the net impact of two opposing effects, i.e., a positive estimate indicates that the positive spillover effect more than offsets the negative effect of the increase in their relative supply. An additional year in average education of the county is associated with an increase of about 4.7% on wages of the most qualified workers and about 3.8% on wages of the less qualified workers; an increase in the share of highly qualified workers in the county by 1 p.p. yields an increase on wages of the most qualified workers and of the less qualified workers of about 0.4% and 0.3%, respectively. Alternatively, an additional year of calibrated education in the county leads to a wage premium of about 2.6% on wages of the most qualified workers, while the effect is about 2% on wages of the less qualified workers.

These estimates may be potentially biased or inconsistent, as one does not control adequately for unobserved county–specific characteristics that may be correlated with both individual and aggregate education of the county, as mentioned in Section 4.4.1. To mitigate the omitted variable bias, explanatory variables measuring county–specific characteristics that could affect both the concentration of human capital and wages are included. The results are reported in columns with even number in Tables 4.2.1–4.2.3. The introduction of these additional controls slightly reduces the estimated external returns to education, but they remain highly significant, which confirms the existence of human capital spillovers. This result suggests that there is a positive bias in estimations without the inclusion of county controls, indicating unobservable local idiosyncrasies, that affect the demand curve of the labour market, which are relevant to the OLS estimation of local human capital spillovers. The estimated spillover effects indicate that an additional year on average education of the county increases the wage of the most qualified workers by 3.8% and those of the less qualified workers by 2.6% (this represents a decrease of about 1 p.p. in the returns to aggregate education, when compared with estimates that do not control for county characteristics). On the other hand, an increase in the share of highly qualified workers in the county by 1 p.p. increases the wages of the most qualified workers by 0.4% and about 0.3% those the less qualified workers, or else an additional year of calibrated education in the county leads

to a wage premium of about 2.1% for the most qualified workers and of about 1.4% for the less qualified workers. These results provide evidence on the existence of positive pure educational spillovers, and they are in line with results proposed in the empirical literature (Acemoglu and Angrist, 2001; Moretti, 2004a; Isacsson, 2005; Dalmazzo and de Blasio, 2007; Liu, 2007; Heuermann, 2011; Rodríguez-Pose and Tselios, 2012).

Furthermore, an interaction term between individual education and local education measure (using the three measures of local human capital used previously) was added. Estimates are reported in Table B.3, columns 1–3, in the Appendix B.4. The results indicate that if an individual does not have schooling then he/she does not benefit from educational externalities, since the coefficient of local human capital variable alone is not statistically significant in any of the specifications. Looking to column 1, and taking a partial derivative with respect to aggregate education, one gets an estimated interaction term of 0.0037. This means that, *ceteris paribus*, the marginal effect of average education at the county level increases by 0.4 p.p. with each additional year of own education, i.e., the workers benefit more from externalities, the higher their level of education. A similar interpretation can be made for the regressions in column 2 and 3.

Additionally, even columns in Tables 4.2.1–4.2.3 show that most of the county–specific variables are significant determinants of individual earnings.[16] As expected, county unemployment significantly reduces wages (as in Dalmazzo and de Blasio, 2007; Heuermann, 2011) and the coefficient of the proxy variable to local amenities, Log Attractiveness Index, also presents a significant and negative sign. This last result is in agreement with the theory of compensatory wage differentials (Roback, 1982), which emphasizes that more amenable locations affect positively the individual welfare and negatively the production costs (e.g., polluted air), and consequently the wages should be lower; it seems that workers are willing to accept a lower wage in exchange for higher welfare. Conversely, earnings tend to be higher in counties where the index of local competition is higher, i.e., in counties where the average firm–size is higher. The coefficient of the county house price variable is also positive and significant, because the demand of workers for more attractive regions is reflected in an increase in the rental price of land (Roback, 1982) and, consequently, in the housing price. Finally, regarding the size of the county, it can be observed that the coefficients are not statistically significant. The estimates for the set of parameters associated with county variables are relatively stable across specifications and aggregate human capital measures.[17]

Finally, in the case of a high dispersion of education within the county the average

[16]The county–specific variables are jointly statistically significant in all regressions.
[17]These estimates are equally stables over the different estimation methodologies.

education may not be a good measure for local education. As such, new specifications are tested and in specification, which includes the average education of the county, is augmented by the coefficient of variation (in logarithmic form).[18] In other specification the share of the most qualified workers is replaced by the Hirschman–Herfindahl Index of education,[19] which addresses the idea that human capital externalities are rooted in the local concentration of workers of the same educational background (Heuermann, 2011). The results are reported in Table B.3, columns 4 and 5, in the Appendix B.4. The results show that, conditional to the average education within county the heterogeneity (dispersion) of education has a positive effect on wages, i.e., the wage premium is higher in counties with higher dispersion of education.

On the other hand, following Heuermann (2011), the Hirschman–Herfindahl index measures the concentration of workers who have at most the 12^{th} year of schooling, which constitutes about 85% of the Portuguese workforce.[20] Thus, the higher the concentration of these workers in the local labour force, the smaller the fraction of most qualified workers therein. So, the negative sign of this index might stem from a lower share of most qualified workers in the workforce of the county. Nevertheless, the impact is close to zero, -0.00002.[21] Furthermore, given that an increase of this index (higher concentration) has a negative effect on wages, it means that the two types of workers are complementary and not substitutes.

4.5.1.2 Worker fixed–effects results

As discussed in Section 4.4.1, if controls for unobservable characteristics of workers are not included in the wage equation, and there exists correlation between the unobserved ability and the regressors, namely education, the OLS produces biased estimates of η, consequently, the fixed–effects estimation will applied. The results underlying the following analysis are reported in Tables B.4 to B.5 in the Appendix B.4.

The use of the fixed-effects model is based on the following evidence. For all the regressions one rejects the null hypothesis of absence of workers specific effects; the F–test statistics is bounded between 16 and 18, and the null is rejected at the 1% significance level (see Table B.4). Additionally, when comparing fixed–effects and random–effects models, the null hypothesis, defined under the Hausman test, that the random–effects

[18] Coefficient of variation is given by the ratio between standard deviation of education in the county and average education in the county.

[19] Hirschman–Herfindahl Index of education is determined by the sum between the squared of share of non–qualified workers and squared of share of high–qualified workers.

[20] According to *Instituto Nacional de Estatística*, in 2009, 85% of the working population has the secondary education or less. This share is decreasing over time and, in 2015, 76% of the working population has the secondary education or less.

[21] The mean value of the Hirschman–Herfindahl index is 9066, out of a maximum possible of 10000 ($=100^2$).

estimator is consistent and efficient is rejected. So, one concludes that the unobserved individual effects are correlated with the explanatory variables, which renders pooled OLS as an inappropriate solution to estimate equation (4.21). The fixed–effects estimation is, thus, a suitable solution.[22]

For the group of workers that includes individuals who increased their schooling level over the period under study, Table B.4, fixed-effects results show that the private returns to education is 1.3%, on average, and the return for the most qualified workers are about five times larger than those for the less qualified; an extra year of schooling for a qualified worker who has improved his/her education leads to an increase on wages of about 5%, while this increase is about 0.8% for the less qualified worker.

The results show a positive and significant effect of local human capital on individual wages. Controlling for the imperfect substitutability, the existence of positive and significant externalities of local human capital is confirmed, whatever the measure of human capital used (compare columns (1) to (3), as well as (4) to (6) and (7) to (9)). The most qualified workers benefit marginally more of external returns to human capital than the less qualified workers. The estimated spillover effects indicate that an additional year in the average education of the county increases the wage of the most qualified workers by 1.1%, while for the less qualified workers is by 0.7% (columns (4) and (7)); an increase in the share of highly qualified workers in the county by 1 p.p. increases the wages of the most qualified workers by 0.1% and about 0.08% the less qualified workers (columns (5) and (8)), or else an additional year of calibrated education in the county leads to a wage premium of about 0.7% for the most qualified workers and of about 0.4% for the less qualified workers (columns (6) and (9)).

From Table B.5 in the Appendix B.4 one sees that the spillover effects are similar for the sample of workers who did not improve their education during the period 2002–2009.

4.5.1.3 Instrumental variables results

As previously discussed, another alternative approach to deal with the problem of endogeneity is the instrumental variable (IV) estimation. The advantage of this method is that if valid instruments exist they will allow the identification of the effect of exogenous changes in human capital levels on wages (Moretti, 2004a,b). In this analysis the presence of a Polytechnic Institute in a county is used as an instrument for local

[22] As mentioned in Section 4.4.1, control variables for the characteristics of the county are not included in fixed–effects model, because in the sample only a negligible number of workers moved from the county. Individual fixed–effects estimates allow to control for unobserved time–invariant characteristics of individuals, as well as counties where individuals work, because the control variables at county level do not vary across individuals and are stable over time. Because gender is a time–invariant variable, the variable is dropped.

human capital, following Moretti (2004a). A second instrument is proposed, the share of people who could read in the county in 1900, which is something new in empirical literature on external returns (see the discussion in Section 4.4.1). Because in 1900 the number of counties was lower than 278, the number of observations in the sample decreased by about 7%, so the sample used in this IV estimation contains about 13.5 million observations worker–year distributed across 261 Portuguese mainland counties.

Individual education will also treated as an endogenous variable and the change in compulsory school attainment in 1981 that established the compulsory education of 9 years is used as instrument, which is a common instrument in empirical economics (see Angrist and Krueger, 1991; Acemoglu and Angrist, 2001).

As mentioned in Section 4.4.1, the specification to estimate by IV only includes the individual and firms controls, a set of dummies for sector of activity and time effects. Because the instrumental variables are included in order to control the impact of unobservable time–region specific factors, the control variables for the county are not used due to its collinearity with the instrument for the aggregate human capital (Heuermann, 2011). As above, different ways to measure human capital in the county are used, namely, average years of education, share of highly skilled workers and the skill index.

When evaluating the exogeneity assumption on individual and local human capital variables, using the Durbin–Wu–Hausman (DWH) test, the null hypothesis of exogeneity is systematically rejected.[23] Table 4.3 reports the regression results of the local human capital equation using the presence of a Polytechnic Institute and the share of people who could read in the county in 1900 as instruments for local human capital.

The first–stage results show that both instruments present a positive and significant impact on the human capital in the county, which points out that the instruments used are relevant. This supports the idea that the share of people who could read in the county in 1900 has a positive and significant impact on aggregated human capital in the county, as well as the presence of a Polytechnic Institute in a county. According to these, one notes that both the presence of a Polytechnic Institute and the share of people who can read in the county have a pronounced weight in local human capital improvement. The results also indicate that the R^2 of first–stage regression capital ranges between 0.67 and 0.82 and these values are higher than those that appear in the empirical literature. F–statistics on the excluded instruments in the reduced form local human capital equation are quite high for both instruments, exceeding the rule–of–thumb value of 10 for non–weak instruments (see Bound et al., 1995 and Staiger and Stock, 1997). As such, one can not reject the significance of the excluded instruments

[23]The detailed statistics are not reported, but will be provided upon request.

Table 4.3: First–stage regression of local human capital variable

Local Human Capital Variable	Average Education	Share of HQ workers	Skill index
Polytechnic Institute	1.5838***	17.8921***	2.9675***
	(0.0241)	(0.2766)	(0.0422)
R^2	0.6988	0.6894	0.6717
F–test on excluded variables	2773***	2717***	2874***
Share 1900	0.0660***	0.7422***	0.1204***
	(0.0003)	(0.0040)	(0.0007)
R^2	0.8172	0.8142	0.7914
F–test on excluded variables	19488***	19287***	16798***
Observations	13,420,048	13,420,048	13,419,674

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting – Sales Index 2013, INE and, *Instituto de Emprego e Formação Profissional–Ministério do Trabalho, Solidariedade e Segurança Social*, 2002–2009.
Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

in the local human capital equation. Additionally, the Stock and Yogo (2005) test shows that the instruments are strong, since the Cragg–Donald F–statistic exceed the tabulated critical value. Consequently, given these results, together with discussion in Section 4.4.1, the instruments proposed for local human capital are likely to satisfy relevance and exogeneity properties, and as such are considered to be valid.

Table 4.4 presents the estimation results from the second stage of IV approach that uses both mentioned instruments for aggregated human capital, 'Share 1900' and 'Polytechnic Institute', respectively; the compulsory schooling of 9 years is used as the instrument for individual education. Results from the OLS estimation are also reported in order to facilitate comparison. Each regression method is identified at the top of the table. Columns 1–3 report the estimates of the external returns to education that are obtained by pooling the two qualification groups, 'All workers', and, therefore, represent an average effect across groups, that may not mean the existence of spillovers (Moretti, 2004a,b). As such, to control for imperfect substitutability, in Table 4.4, the estimates for the two groups of workers, the 'Less qualified', columns 4–6, and the 'Most qualified', columns 7–9, are reported. Horizontally, Table 4.4 is organized in three sections, one for each measure of local human capital, 'Average education', 'Share of Highly Qualified workers' and the 'Skill index', respectively.

For the sample with all workers, IV results show that the estimated private returns to education of about 6% (comuns (2) and (3)), on average, while the corresponding return by OLS is about 4.5% (column (1)). As expected, the most qualified workers benefit from higher private returns to education. IV estimates of the returns to education are higher than the OLS estimates for the less qualified workers, about 5% and 3%, respectively. But for the most qualified workers the reverse is verified, the IV

estimated return to education is about 8%, while the OLS counterpart is about 10%. This finding may depend on their individual characteristics, consisting on the ability of the workers with lower schooling to translate benefits from their education into higher individual productivity.[24]

The estimates of educational externalities decrease in size regardless of the measure of local human capital when IV estimation is used (as in Acemoglu and Angrist, 2001; Iranzo and Peri, 2009). According to Moretti (2004a,b), the OLS bias is positive if positive shocks are associated with increases in aggregated human capital and individual income in a region. Thus, if variation on most qualified workers across counties is driven by unobserved demand for skilled workers factors, OLS estimates are biased upward. Despite this decrease in the effect size, the IV results indicate that statistically significant human capital externalities effects exist, regardless the different measures of human capital used. Considering all workers of the sample, on average, an increase of the average education in a county by one year increases the wage of workers by around 1.1%–2.5%, or an increase in the share of the most qualified workers in the county by 1 p.p. raises wages by about 0.1%–0.2%. Finally, coefficients of county skill index also show that an additional year of calibrated education in the county leads to a wage premium of about 0.6%–1.4%.

Controlling for the imperfect substitutability across workers, within the group of the most qualified workers is noted that the coefficient of the aggregate human capital variables appears with no statistical significance, when the presence of a Polytechnic Institute in a county is used as instrument, regardless of the aggregated human capital measure used. This situation can occur due to the low variability of the instrument, since there are Polytechnic Institutes in only 6.5% of the mainland Portuguese counties, which distributed by more than 13 million of observations yields a low standard–deviation (0.480) of the instrument.

IV estimation suggest that the most qualified workers receive a higher wage premium for working in counties with high human capital than the less qualified workers, who also benefit from a significant wage premium. This result is in disagreement with some literature, e.g., Moretti (2004a) who suggests that unskilled workers receive higher wage premium than skilled workers. But, on the other hand, Heuermann (2011) and Silva (2003) suggest the opposite and the last one proposes that, after 5 years of being displaced, the wages of a college graduate worker raise by more than 3% compared to a displaced worker with basic education if the average education of their county increases by one year. As such, both groups of workers obtain productivity gains in

[24] As in previous essay the IV estimated private return to education is higher than the corresponding OLS estimate. See Section 3.6 in Chapter 3 for a detailed discussion on this issue.

Table 4.4: IV results for 3 groups: all workers, less qualified and most qualified workers

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All workers			Less qualified			Most qualified		
	OLS	**IV**		**OLS**	**IV**		**OLS**	**IV**	
		Share 1900	Polyt. Inst.		Share 1900	Polyt. Inst.		Share 1900	Polyt. Inst.
Education	0.0450***	0.0610***	0.0602***	0.0308***	0.0478***	0.0485***	0.0973***	0.0760***	0.0773***
	(0.0005)	(0.0015)	(0.0015)	(0.0005)	(0.0025)	(0.0025)	(0.0013)	(0.0025)	(0.0025)
Av. education	0.0379***	0.0111***	0.0253***	0.0370***	0.0240***	0.0147***	0.0481***	0.0378***	0.0056
	(0.0019)	(0.0041)	(0.0029)	(0.0018)	(0.0030)	(0.0040)	(0.0025)	(0.0035)	(0.0055)
Observations	11,712,773	11,712,773	11,712,773	10,348,805	10,348,805	10,348,805	3,883,300	3,883,300	3,883,300
R-squared	0.634	0.627	0.629	0.549	0.542	0.540	0.637	0.635	0.631
RMSE	0.334	0.337	0.336	0.307	0.310	0.311	0.389	0.391	0.39
Education	0.0450***	0.0610***	0.0602***	0.0334***	0.0477***	0.0485***	0.0973***	0.0760***	0.0773***
	(0.0005)	(0.0015)	(0.0015)	(0.0006)	(0.0025)	(0.0025)	(0.0013)	(0.0025)	(0.0025)
Share of HQ workers	0.0034***	0.0010***	0.0022***	0.0039***	0.0021***	0.0013***	0.0043***	0.0033***	0.0005
	(0.0002)	(0.0004)	(0.0003)	(0.0002)	(0.0003)	(0.0004)	(0.0002)	(0.0003)	(0.0005)
Observations	11,712,773	11,712,773	11,712,773	11,492,864	10,348,805	10,348,805	3,883,300	3,883,300	3,883,300
R-squared	0.634	0.627	0.629	0.562	0.542	0.540	0.637	0.635	0.631
RMSE	0.334	0.337	0.336	0.314	0.310	0.311	0.389	0.391	0.393
Education	0.0456***	0.0613***	0.0607***	0.0314***	0.0483***	0.0489***	0.1001***	0.0763***	0.0774***
	(0.0005)	(0.0015)	(0.0015)	(0.0005)	(0.0025)	(0.0025)	(0.0020)	(0.0025)	(0.0025)
Skill index	0.0204***	0.0059***	0.0138***	0.0199***	0.0130***	0.0078***	0.0347***	0.0208***	0.0031
	(0.0010)	(0.0022)	(0.0016)	(0.0010)	(0.0016)	(0.0021)	(0.0015)	(0.0019)	(0.0030)
Observations	11,712,438	11,712,438	11,712,438	10,348,511	10,348,511	10,348,511	4,581,165	3,883,259	3,883,259
R-squared	0.634	0.627	0.629	0.549	0.542	0.540	0.637	0.634	0.631
RMSE	0.334	0.337	0.336	0.308	0.310	0.311	0.394	0.391	0.393

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting – Sales Index 2013, INE and, *Instituto de Emprego e Formação Profissional–Ministério do Trabalho, Solidariedade e Segurança Social*, 2002–2009.

Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include individual controls, firm controls and a set of dummies to control for occupation, sector of activity and year.

counties with the greatest concentration of human capital which demonstrates the existence of productivity externalities generated by the interaction with more qualified people.

The IV results reported in Table 4.4 show that an increase by one year in the average education of the county raises the wages of the most qualified workers about 3.8% or an increase in the share of most qualified workers in the county by 1 p.p. increases their wages by 0.3%. Additionally, an additional year of calibrated education in the county leads to a wage premium for the most qualified workers of about 2.1%. Regarding the less qualified workers, an extra year in the average education yields a wage premium of 2.4%; an increase in the share of most qualified workers by 1 p.p. leads to an increase on wages of about 0.2%, and an increase of one year of calibrated education raises wages about 1.3%. We should highlight that the IV estimates of human capital externalities are higher than OLS estimates for the most qualified workers and for less qualified workers the opposite occurs.

These findings imply that the existence of human capital spillovers cannot be rejected for both qualification groups of workers. In addition, these results show that in both groups of workers the private returns to education are higher than the external returns to education. Using IV estimation, and using 'Share 1900' as the instrument for aggregate human capital, column (8), Table 4.4, for the most qualified workers an extra year in individual education leads to an increase on wages by about 8%, which compares to the fact that an increase in average education in the county leads to an increase of about 3.8%. Regarding the less qualified workers, column (5), an extra year in individual education leads to an increase on wages of about 5%, while an increase in the average education in the county leads to an increase of about 2.4%. Overall, the returns to education are substantial.

Since the panel structure is available, in order to complete this empirical analysis the instrumental variables fixed–effects estimation (IV–FE) is applied to circumvent the potential endogeneity problem of human capital in the county. Because the sample used is too large the software used is unable to perform the IV–FE estimation. Consequently, in the IV–FE analysis only a random sub–sample of about 6 million of observations is used. The estimates are reported in Table B.6 in the Appendix B.4. The organization of columns and rows is similar to the one discussed above for Table 4.4. Given the low variation in the instrument 'Polytechnic Institutes' described above, and the fact that the estimates for the spillover parameter resulting from the use of this instrument are consistently negative, lead us to cast reasonable doubt on validity of the results. Nevertheless, we have opted to report the estimates, even though we will not discuss them below.

The IV–FE estimates reported in Table B.6 show that the private returns to education are higher than the fixed effects estimates previously analysed, suggesting that measurement error has biased the fixed–effects estimates downward (Li et al., 2012). In both groups of workers, high and low qualified, the private returns to education are higher than the external returns to education. The spillover effect for the most qualified workers is higher than for the less qualified. For the most qualified workers, an extra year in individual education leads to an increase on wages by about 5.4% and an increase in average education in the county leads to a wage increase of about 1.1%. Regarding the less qualified workers an extra year in individual education leads to an increase on wages by about 1.6% and an increase in the average education in the county leads to an increase of about 0.2%. Additionally, an increase in the share of the most qualified workers in the county by 1 p.p. raises their wages by about 0.09% and the wages of the less qualified workers about 0.02%. Finally, coefficients of county skill index also show that an additional year of calibrated education in the county leads to a wage premium of about 0.6% and for the less qualified 0.1%.

Finally, as discussed in Section 4.4.1, the Hausman–Taylor estimation was applied. However, we made the option not to present the estimates. First, a Hausman test that compares the fixed–effects and Hausman–Taylor models, strongly rejects the null hypothesis, i.e., the Hausman–Taylor estimator efficient and consistent, while the fixed–effects estimator is consistent regardless of the assumptions on correlation between regressors and the composite error term. Second, the Hausman–Taylor estimates for the specification studied are inconsistent. On the one hand, the estimated results of the effect of education on individual wages are in complete disagreement with the theory of human capital and, on the other hand, the results are not in line with results obtained for other estimation methods in this essay.

4.5.2 Human capital externalities at regional and firm level

4.5.2.1 Ordinary least squares results

The analysis performed in this section follows a path that, to our knowledge, has not yet been explored in the existing literature for Portugal. Following Canton (2009) and Bauer and Vorell (2010), both local and firm human capital effects on individual wages are analysed, allowing to assess whether local spillover effects of human capital still exist and comparing them with firm spillovers effects.

So, to go further in the investigation about externalities of education, the previous specification is augmented with a firm human capital term. Following a similar path to the one discussed in previous sections, this firm's component is measured within

firms in three different ways, : (1) average years of education, (2) share of highly skilled workers and, (3) the skill index at the firm level. So the wage equation has the following way:

$$\log s_{irt} = \alpha + \theta s_{irt} + \beta X_{irt} + \eta_1 \bar{S_{rt}} + \eta_2 \bar{S_{ft}} + \delta Z_{rt} + d_t + \varepsilon_{irt} \tag{4.22}$$

with w_{irt} denoting the hourly wage of individual i in county r at time t; s_{irt} denotes the individual's schooling; X_{irt} is a vector of observable characteristics of individual i in county r at time t, including gender, experience and tenure, and their corresponding squares, as well a set of dummies for occupation and sector of activity, a firm size control and a control to the sales of the firm; $\bar{S_{rt}}$ represents regional human capital in the county r in time t; $\bar{S_{ft}}$ represents human capital in the firm f in time t; Z_{rt} is a vector of county characteristics; and d_t is a time fixed effects; α is a constant and ε_{irt} is white noise error term. The variables of interest are $\bar{S_{rt}}$ and $\bar{S_{ft}}$, and the coefficients of these variable, η_1 and η_2, respectively, provide evidence on the existence of regional– and firm–level spillover effects of education.

In order to meaningfully calculate human capital stock at firm level, a minimum of five workers is defined to select the firms to include in the sample. As in the previous sections, robust standard errors, clustered at firm level, are applied, aiming to control for heteroscedasticity and possible serial correlation within firms.

Results from pooled OLS estimation of equation (4.22) are reported in Table 4.5; the estimates are done using the whole sample following the restriction defined in the previous paragraph. As above, evidence of a strong relationship between aggregated human capital and earnings does not necessarily, in general, imply positive externalities, due to the imperfect substitution between skilled and unskilled workers which may induce such result. So, additional estimations are performed for separate qualification groups of workers, which are reported in the same table.

Results reported in Table 4.5 show that the estimated coefficient of human capital variables appear to be positive and statistically significant at 1% level. Human capital externalities exist both at county level and at firm level, regardless of the human capital measure used. Furthermore, it appears that the inclusion of aggregated human capital at firm level makes the coefficient of aggregated human capital of the county slightly lower, comparing with OLS results in the previous section. In addition, the magnitude of the coefficients of human capital at county level also decrease when controls for county characteristics are included, but the coefficients of human capital variables at firm level remain similar.

Considering the whole sample, columns (1) and (2) in Table 4.5, the results suggest that an increase of one year in the average of formal education in the county yields an

Table 4.5: Spillovers at regional and firm level

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	All workers		Less qualified		Most qualified	
Education	0.0383***	0.0383***	0.0224***	0.0231***	0.0909***	0.0910***
	(0.0005)	(0.0005)	(0.0004)	(0.0004)	(0.0013)	(0.0013)
Average county education	0.0295***	0.0166***	0.0304***	0.0167***	0.0270***	0.0174***
	(0.0019)	(0.0036)	(0.0019)	(0.0036)	(0.0026)	(0.0059)
Average firm education	0.0308***	0.0307***	0.0348***	0.0346***	0.0422***	0.0425***
	(0.0012)	(0.0013)	(0.0013)	(0.0013)	(0.0017)	(0.0017)
County controls	No	Yes	No	Yes	No	Yes
Observations	10,741,731	10,723,750	9,407,188	9,390,088	3,627,386	3,624,297
R–squared	0.656	0.657	0.572	0.574	0.652	0.653
RMSE	0.326	0.327	0.302	0.302	0.376	0.376
Education	0.0398***	0.0397***	0.0255***	0.0252***	0.0947***	0.0948***
	(0.0005)	(0.0005)	(0.0004)	(0.0004)	(0.0014)	(0.0014)
Share of HQ workers (county)	0.0028***	0.0021***	0.0028***	0.0020***	0.0027***	0.0024***
	(0.0002)	(0.0003)	(0.0002)	(0.0003)	(0.0002)	(0.0006)
Share of HQ workers (firm)	0.0025***	0.0024***	0.0027***	0.0027***	0.0031***	0.0031***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)
County controls	No	Yes	No	Yes	No	Yes
Observations	10,741,731	10,723,750	9,407,188	9,390,088	3,627,386	3,624,297
R–squared	0.656	0.657	0.571	0.573	0.650	0.650
RMSE	0.327	0.327	0.303	0.303	0.376	0.377
Education	0.0433***	0.0430***	0.0282***	0.0277***	0.0991***	0.0991***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0014)	(0.0014)
Skill index (county)	0.0164***	0.0089***	0.0164***	0.0084***	0.0182***	0.0147***
	(0.0011)	(0.0021)	(0.0011)	(0.0021)	(0.0015)	(0.0040)
Skill index (firm)	0.0240***	0.0239***	0.0270***	0.0267***	0.0304***	0.0304***
	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0017)	(0.0017)
County controls	No	Yes	No	Yes	No	Yes
Observations	10,101,727	10,085,127	8,895,630	8,879,710	3,375,818	3,373,169
R–squared	0.662	0.663	0.575	0.577	0.654	0.654
RMSE	0.326	0.326	0.303	0.303	0.377	0.376

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting – Sales Index 2013, INE and, *Instituto de Emprego e Formação Profissional–Ministério do Trabalho, Solidariedade e Segurança Social*, 2002–2009.

Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include individual controls, firm controls and a set of dummies to control for occupation, sector of activity and year.

increase on wages of about 1.7%–3%, on average, and an increase of one year in the average education within firm leads to an increase of about 3.1% on the wages of their workers. Also the estimated coefficients indicate that the wages of the workers rise by about 0.25% when the share of the most skilled workers in a firm increases by 1 p.p., and an increase by 1 p.p. in the share of the most skilled workers in the county leads an increase of about 0.2%–0.3%. An additional year of calibrated education in the firm

leads to an spillover effect of about 2.4%, while this effect ranges between 0.9%–1.6% with an additional year of calibrated education in the county. As such, the effect of the firm's educational externalities is equally significant as the local human capital of the county, and the size of the effect is slightly higher at firm level than at county level. These results are in agreement with the literature. A possible interpretation for these findings is that human capital spillovers are strongest within the firm because productive social interactions are probably more intense among colleagues than among the people in a county (Canton, 2009).

Controlling for imperfect substitutability across workers, columns (3) to (6), the results confirm the previous findings that the aggregated human capital outside the firm exerts a significant influence on the wages of both groups of workers and the human capital spillovers within firm are higher than at county level. The estimates also confirm that, generally, the most qualified workers benefit more from higher human capital externalities than the less qualified. OLS estimates indicate that, on average, an extra year in average education within the firm yields an increase on the wages of the most qualified workers (less qualified) of about 4.3% (3.5%), while the corresponding value at county level is of about 1.7%–2.7% (1.7%–3.04%). It also shows that an increase by 1 p.p. of the share of the most qualified workers within the firm increases the wages of the most qualified (less qualified) by about 0.31% (0.27%), and the corresponding estimate at county level is about 0.31% (0.27%). Furthermore, an additional year of calibrated education in the firm increases the wages of the most qualified workers (less qualified) by about 3% (2.7%), whereas the corresponding estimate at county level is of about 1.5%–1.8% (0.8%–1.6).

The size of spillover effects at both aggregation levels, firm and county, is individually lower than the private return to education and this difference is higher for the groups of workers with more qualifications. On average, the return rate of an additional year in worker's education is about 4%, but the return to education for the less qualified workers is about 2.3%–2.8% and for the most qualified workers is about 9%–9.1%.

Our estimates are significantly lower when compared with the pooled OLS estimates of Martins and Jin (2010), who find evidence of firm–level social returns of about 18%. Their fixed–effects estimated return falls considerably, being about 5%, which is relatively close to the ones we obtain.

4.5.2.2 Worker fixed–effects results

As previously mentioned, pooled OLS delivers upward–biased effects of both local and firm human capital on the individual wages, thus the estimated effects of aggregated

human capital appear to be biased by unobserved time–invariant effects that may be correlated with $\bar{S}_{rt}$ and $\bar{S}_{ft}$. Therefore, fixed–effects estimation is also applied. The fixed–effects estimation seems to be the most suitable model, because the F–test indicates that there are significant unobserved individual effects. Additionally, the Hausman test's comparing the fixed–effects and the random–effects (RE) rejects the null hypothesis of valid RE estimates. Both results imply that pooled OLS would be inappropriate.

Results are reported in Table B.7 in the Appendix B.4. Controlling for time–invariant individual fixed–effects, the results propose that the estimated effect of the aggregated human capital on wages is lower than the effect obtained by pooled OLS, but still positive and significant at the 1% level, providing evidence of the existence of significant and positive human capital externalities. This result applies to the three measures of human capital used. The external return at firm level is higher than the external return at county level for workers who have improved their schooling in the period under study. In addition, in each level of aggregation the spillover effect is identical for both groups of workers. An extra year in the average education at firm level yields a wage premium of about 1.3% and 1.5% for the less and most qualified workers, respectively. At the county level the corresponding returns are 0.6% and 0.7%. An increase in the share of most qualified workers by 1 p.p. within firm leads to an increase on wages of about 0.1% for both qualification groups and at county level the spillover effect is about 0.06% and 0.09% for the less and most qualified workers. Finally, an increase of one year of calibrated education at firm level raises wages about 1% for both groups of workers and the corresponding external return at county level is about 0.36% and 0.54% for the less and most qualified workers.

These findings are partially in agreement with international literature. On one hand, Canton (2009) finds an important relationship between individual wages and human capital indicators, when the fraction of high–skilled workers is used as a measure for aggregated human capital. In this case, the size of human capital spillovers at firm level are higher than the human capital spillovers at county level. However, if aggregated human capital is measured by the average education in the region/firm no evidence for human capital spillovers at regional level is found and at firm level an external return of about 1.3% is found. When controlling for imperfect substitution between workers no evidence for human capital spillovers for high skilled workers is found, but results suggest that the lower the educational level the greater the human capital spillovers at firm level. At regional level the coefficient is not statistically significant. Opposite results are found by Bauer and Vorell (2010). These authors, using the regional share of high–skilled workers as a measure of human capital, and

pooled OLS estimation, find evidence of human capital spillovers only for the high–skilled workers (at the 10% significance level); external returns to regional human capital is higher than external return to human capital at firm level. Controlling for individual fixed–effects, the authors find evidence of human capital externalities only at regional level for low–skilled workers and external returns to human capital only at firm level for high–skilled workers. Martins and Jin (2010), using a measure aggregated human capital at firm level, provide evidence of significant spillovers for the less–educated workers, as their wages increase by 2% to 3% following an additional year of education of workers in the same firm. However, the group of educated workers does not seem to benefit from such spillovers.

4.5.3 Inter–regional human capital externalities

The test for external returns to education can be expanded by explicitly taking into account for the spatial dimension of the information available. This seems a natural extension of what has been discussed so far. So, in this section we will evaluate to what extent the results we have reported so far are robust to an adaptation of the analysis designed above by including location, particularly distance between economic unities, in the formulation of the Mincerian wage equation.

Over a long period of time, variables related to the location were not considered relevant to the explanation of many phenomena resulting from economic dynamics. However, with the emergence of adequate data and the development of techniques of spatial econometrics, the influence through effects of spatial spillovers of neighbouring regions in the development of a given region has lately been considered in the empirical literature (Anselin, 2003).

The use of locational data may lead to spatial dependence between observations, which is an inherent characteristic to the representation of data through territorial subdivisions (Anselin, 1988). The spatial dependence can be defined as a lack of independence between observations (Anselin, 1988; LeSage, 2008; LeSage and Pace, 2009). A factor associated with the spatial dependence is the possible existence of externalities between observations, for instance, regions with high levels of human capital tend to a have positive spillovers to another (Anselin, 1988; LeSage, 2008). This assumption is directly related to the so–called first law of geography enunciated by Tobler (1970, p.236) “Everything is related to everything else, but near things are more related than distant things”.

The presence of spatial dependence violates the basic assumptions of Gauss–Markov, specifically when it is assumed that observations, in this case regions, are independent and identically distributed. In contrast, spatial dependence takes into account for the

fact that values observed at one region depend on the values of observations at nearby regions (LeSage and Pace, 2009). Within this framework the spatial weight matrix, W, which summarizes the spatial correlation structure, becomes a critical concept (Anselin, 1988).

While the literature on inter–regional knowledge spillovers has emerged, namely, the contribution of inter–regional spillovers in the aggregate regional wage (Rodríguez-Pose and Tselios, 2012), the research to measure the relationship between individual wages and inter–regional educational externalities seems to be made only by Rodríguez-Pose and Tselios (2012). The authors use microeconomic data from the European Community Household Panel and resort to spatial economic analysis for the purpose of determining to what extent differences in individual earnings are the result of the educational attainment of the individual, the educational attainment of the members of the individual household, the educational endowment of the region where the individual lives, or the educational endowment of the neighbouring regions. The results show that, in addition to the expected positive private returns, place–based regional and supra–regional educational externalities generate significant benefits for workers (Rodríguez-Pose and Tselios, 2012).

Therefore, following Rodríguez-Pose and Tselios (2012), the goal in this section is to determine if it exists supra–regional externalities of human capital on individual wages in Portugal. This external return will occur if an improvement in the educational level of the people who work in a county contributes to an increase in the individual wages of an individual who works in a neighbouring county. One wants to learn if workers with similar educational characteristics working in different counties may have different wages just because the neighbouring counties have a different education level (Rodríguez-Pose and Tselios, 2012).

We now define the model to estimate as

$$\log w_{irt} = \alpha + \theta s_{irt} + \beta X_{irt} + \eta_1 \bar{S_{rt}} + \eta_2 [W\bar{S}_t]_r + \delta_1 Z_{rt} + \delta_2 [WZ_t]_r + d_t + \varepsilon_{irt} \quad (4.23)$$

where W is the spatial weight matrix, $[W\bar{S}_t]_r$ represents the average education of neighbouring counties of county r in time t; $[WZ_t]_r$ includes control variables of neighbouring counties. As already discussed, w_{irt} denotes the hourly wage; s_{irt} is the individual's schooling; X_{irt} is a vector of observable characteristics of the worker. $\bar{S_{rt}}$ represents regional human capital in the county, while Z_{rt} is a vector of county characteristics. The variables of interest are $\bar{S_{rt}}$ and $[W\bar{S}_t]_r$. Their coefficients, η_1 and η_2, respectively, provide an indication of the existence of regional and inter–regional externalities of human capital.

Matrix W is the spatial weighting matrix and quantifies the connections between

regions, i.e., the matrix W aims to capture the possible spatial correlation structure present in the data. Its choice is one of the most difficult and controversial methodological issues, since the incorrect choice of W may lead to inconsistent estimates of the parameters in the wage equation (Anselin, 1988; LeSage, 2008). Consequently, the choice of the spatial weight matrix should be supported by theoretical considerations.

In order to construct the spatial weights matrix W, a neighbour set is defined for each location, i.e., in our case W is a 278×278 matrix where each element ω_{ij} represents the proximity between spatial element i (in the row of the matrix) and j (column). The elements of the matrix can be understood as the strength of the potential interaction between units i and j. In this sense, it can be interpreted as the presence and the strength of a linkage between spatial points (the regions) in a network representation that matches the spatial weights structure (Anselin, 1988; Anselin et al., 2008; Arbia and Baltagi, 2009). The choice of the spatial weights is generally conducted by a geographic criteria, namely distance between units (Anselin et al., 2008).

In our analysis one assumes that a county is not considered its own neighbour, which implies that the diagonal elements of the matrix W are equal to zero, $\omega_{ii} = 0$. The assignment of weights ω_{ij} is based on information on the contiguity among mainland Portuguese counties, assigning $\omega_{ij} = 1$ when county i and j share a common border or vertex (contiguity matrix of queen type); $\omega_{ij} = 0$ otherwise. Explicitly, one assumes that contiguous neighbouring counties present a greater degree of dependence when compared with distant counties. The current option implies that W is a symmetric matrix.

Following most of the applied literature on spatial econometrics, for computational simplicity and to simplify the interpretation of the spatial variables, the spatial weights are standardized, thus the elements in each row sum to 1 $\left(\omega_{ij}^{S} = \dfrac{\omega_{ij}}{\sum_j \omega_{ij}}\right)$ and, although the original matrix W is defined to be symmetrical, the row–normalized matrix may not be symmetric (Anselin et al., 2008). In this sense, $[W\bar{S}_t]_r = \sum_{j \neq r} \omega_{ir} \bar{S_{jt}}$ is a linear combination of the observations of all counties, representing the average education of neighbouring counties of county r at time t, and it is called spatial lag variable.

4.5.3.1 Ordinary least squares results

Table 4.6 reports the pooled OLS results of the Mincerian equation (4.23) and, as in the previous sections, regional human capital is measured in three ways: average years of education, share of highly skilled workers and the skill index in the county. These three measures of human capital are also used to determine the human capital endowment of the neighbouring counties. Similarly to the analysis performed above, and to control

Table 4.6: Regional and inter–regional externalities

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	All workers		Less qualified		Most qualified	
Education	0.0449***	0.0453***	0.0305***	0.0306***	0.0977***	0.0979***
	(0.0005)	(0.0005)	(0.0001)	(0.0005)	(0.0013)	(0.0013)
Average county	0.0249***	0.0177***	0.0258***	0.0203***	0.0367***	0.0197***
education	(0.0032)	(0.0028)	(0.0032)	(0.0028)	(0.0051)	(0.0045)
Education	-0.0010	0.0643***	0.0002	0.0536***	-0.0025	0.1201***
neighbours	(0.0012)	(0.0102)	(0.0012)	(0.0099)	(0.0020)	(0.0137)
County controls	No	Yes	No	Yes	No	Yes
Neighbouring controls	No	Yes	No	Yes	No	Yes
Observations	12,562,924	11,192,659	11,114,765	9,863,479	4,870,546	4,427,817
R–squared	0.633	0.632	0.5491	0.547	0.6353	0.6369
RMSE	0.333	0.334	0.306	0.307	0.394	0.394
Education	0.0449***	0.0453***	0.0305***	0.0306***	0.0978***	0.0979***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0013)	(0.0013)
Share of HQ	0.0026***	0.0017***	0.0026***	0.0019***	0.0038***	0.0016***
(county)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0005)	(0.0016)
Share of HQ	0.0008**	0.0074***	0.0009***	0.0063***	0.0016***	0.0139***
(neighbours)	(0.0003)	(0.0011)	(0.0003)	(0.0011)	(0.0005)	(0.0015)
County controls	No	Yes	No	Yes	No	Yes
Neighbouring controls	No	Yes	No	Yes	No	Yes
Observations	12,562,924	11,192,659	11,114,765	9,863,479	4,870,546	4,427,817
R–squared	0.633	0.632	0.549	0.547	0.6355	0.6370
RMSE	0.333	0.334	0.306	0.307	0.394	0.394
Education	0.0451***	0.0455***	0.0309***	0.0309***	0.0978***	0.0977***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0013)	(0.0013)
Skill index	0.0137***	0.0098***	0.0141***	0.0111***	0.0208***	0.0101***
(county)	(0.0018)	(0.0016)	(0.0017)	(0.0015)	(0.0033)	(0.0030)
Skill index	-0.0001	0.0445***	0.0010	0.0363***	-0.0003	0.0860***
(neighbours)	(0.0013)	(0.0075)	(0.0012)	(0.0072)	(0.0021)	(0.0096)
County controls	No	Yes	No	Yes	No	Yes
Neighbouring controls	No	Yes	No	Yes	No	Yes
Observations	10,896,194	9,525,962	9,654,249	8,402,990	4,393,982	3,951,260
R–squared	0.635	0.633	0.551	0.549	0.6346	0.6360
RMSE	0.332	0.334	0.307	0.307	0.395	0.394

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting – Sales Index 2013, INE and, *Instituto de Emprego e Formação Profissional–Ministério do Trabalho, Solidariedade e Segurança Social*, 2002–2009.
Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include individual controls, firm controls and a set of dummies to control for occupation, sector of activity and year.

for the imperfect substitution between workers with different educational levels, the estimation for separate education groups is applied and the results are also reported in Table 4.6.

In line with the human capital theory, as well as with our previous results, pooled

OLS estimation, considering all workers, shows that the return to an additional year in own education is positive, statistically significant, and about 4.5%; the return to education for the less qualified workers is about 3.1%, while for the most qualified workers the return is about 9.8%.

In specifications which include control variables for characteristics of county and for characteristics of the neighbouring counties, the dimension of the coefficients for human capital externalities at county level decrease, but they remain positive and significant. The results show a positive and significant effect of externalities at both levels, county and extra–county, on individual wages, supporting the importance of the influence through educational spatial spillovers effects of the neighbouring regions. Considering imperfect substitution between workers, columns (3) to (6) in Table 4.6, it is found that results still provide evidence of the existence of pure positive spillover effects for both qualifications groups of workers.

As in Rodríguez-Pose and Tselios (2012), we also conclude that inter–regional geographical spillovers human capital exist. Therefore, an individual that works in a county with a high human capital level, whose neighbouring counties have higher levels of human capital, would tend to have higher wages than an individual with similar characteristics but working in a county with lower human capital level, surrounded by other counties with similar human capital level (Rodríguez-Pose and Tselios, 2012).

4.5.3.2 Worker fixed–effects results

Following possible bias in OLS estimates due the endogeneity of human capital variables, we will use the panel structure of the data. The statistical analysis does reveal that there are significant unobserved individual effects, and that RE and OLS are unsuitable to estimate consistently the parameters of interest.

Table B.8 in the Appendix B.4 reports estimations of the individual fixed–effects models. We find evidence of positive human capital externalities, at different levels of aggregation. Thus, the evidence implies that the wage of any worker is affected by the conjunction of their individual education, the human capital level of the county where the individual works and the human capital level of the neighbouring counties, being that the individual educational level is the most relevant factor.

Note that, if aggregated human capital is defined in terms of average educational attainment or skill index, the effect of externalities at county level is higher than the effect of externalities at extra–county level, but if aggregated human capital is defined by the share of qualified workers the opposite is verified. In addition, there is evidence of the existence of pure positive spillover effects for both qualifications groups of workers. However, the magnitude of the spillover is lower. The effect of

human capital externalities is higher for the most qualified workers than for the less qualified workers, which confirms previous results. Comparing the spillover effects of the aggregated human capital at different aggregation levels, if aggregated human capital is defined in terms of average educational attainment or skill index, the effect of externalities at county level is higher than the effect of externalities at extra–county level in both groups of workers, but if aggregated human capital is defined by the share of qualified workers the spillover effects are identical.

The results for both qualification groups of workers can not be compared with other studies, since the only known research on the study of the relationship between the potential effect of inter–regional educational spillovers and individual wages is the work of Rodríguez-Pose and Tselios (2012) and the authors did not analyse the role of imperfect substitutability between workers with different educational levels.

Summarizing, it seems evident that the regression results show that the role of educational attainment at the three levels of analysis – individual, regional and inter–regional – matters for the wages of both the less and the most qualified workers.

4.6 Concluding remarks

In this essay, the existence and size of human capital spillovers at different levels of analysis, in the Portuguese counties in the period 2002–2009, is examined. The estimation framework is the Mincerian wage regression augmented with the aggregated human capital variable applied to a matched employer–employee panel set and instrumenting both individual and aggregated human capital using the compulsory schooling of nine years of education, and the presence of a Polytechnic Institute in a county and the share of people who could read in the county based on the data of the 1900 Census, respectively. In this analysis, three forms of aggregated human capital are distinguished, namely average years of education in the county, share of highly qualified workers in the county workforce and, finally, a skill index in the county, because different forms of measure of regional human capital can be characterized by different spillovers. Initially, the external returns to education are estimated by pooling all education groups, but subsequently two groups are defined to take into account for the imperfect substitutability. Furthermore, since there are some methodological issues that must be addressed, different estimation strategies are applied, i.e., in addition to the OLS estimation, the fixed–effects, the instrumental variables and instrumental variables fixed–effects are used.

The results suggest a positive spillover effect of the human capital of the county on workers' wages. First, defining regional human capital as average educational attain-

ment, one finds that an additional year in county's education increases, on average, worker's wage by about 1%–3%. This finding is in line with empirical literature, namely, Moretti (2004a,b); Dalmazzo and de Blasio (2007); Garcia-Fontes and Hidalgo (2009); Heuermann (2011); Rodríguez-Pose and Tselios (2012) – returns to individual's education amount roughly to 4.5%–6%. Second, taking counties' human capital as the share of highly qualified workers (12 or more years of schooling), the estimates indicate that an increase of 1 p.p. in this share is associated with a 0.1% and 0.3% increase in individual wages. Third, an additional year of calibrated education in the county leads to an external return of about 0.6% and about 2%.

When one controls for imperfect substitutability, by splitting the sample according to workers' education, the external return to education ranges between 3.8%–7.6%, for most qualified workers, and between 2.6%–4.8%, for the less qualified workers. Taking counties' human capital as the share of highly qualified workers (12 or more years of schooling), the estimates indicate that an increase of 1 p.p. in this share is associated with a 0.2% and 0.3% increase in individual wages, respectively, of the most qualified and the less qualified workers. In addition, an additional year of calibrated education in the county leads to an external return of about 2.1% and about 1.3% for the most qualified and the less qualified workers, respectively. So, one concludes that the effects of the aggregated human capital on wages increase with the individual education in Portugal.

Throughout the different formulations of counties' aggregate human capital, returns to workers' own education range between 5% and 6%. When splitting the sample, these returns vary between 3% and 5% for least skilled workers, and 7% and 10% for high skill workers.

In a second stage, following Canton (2009) and Bauer and Vorell (2010), firm's human capital has been included in the original spillovers regression. This allows to assess whether local spillover effects of human capital on individual wages still exist after controlling for aggregated skills within workers' firm. Such analysis has never been made for the Portuguese economy. The inclusion of aggregated human capital at firm level decreases the coefficient of aggregated human capital at county level, but it remains positive and significant.

The estimated spillover effects indicate that, on average, an increase of one p.p. in the share of highly qualified workers in a firm increases the wages, for both the most qualified and the less qualified workers in the firm, by about 0.3%; an extra year in average education of the firm yields an increase on wages of the most qualified workers of about 4.3% and of about 3.5% on the wages of the less qualified workers. Finally, an additional year of calibrated education in the firm leads to an external

return of about 3% and about 2.7% for the most qualified and the less qualified workers, respectively. Note that, the human capital spillover effect within the firm is stronger than the spillover effect in a county. A possible cause might be that productive social interactions are more intense among colleagues than among the people in a county. Finally, this analysis assesses the relationship between geographical (regional and inter–regional) human capital externalities across the mainland Portuguese counties and individual wages. One would expect that regions with high levels of human capital tend to have a positive spillover to another (Anselin, 1988; LeSage, 2008). The results indicate that beyond individual education and human capital level in the county, the human capital level of the neighbouring counties also matters.

In summary, individual wages are determined not only by an individual's own education – private returns to education –, but also by the knowledge gained through interaction with other workers within firms, regions, and inter–regional externalities – external returns to human capital. These benefits accrue to both high and low skill workers.

The positive human capital spillovers identified above provide an additional and strong argument for policies aiming at reducing early school leaving. To the extent that an increase in the drop–out rate negatively influences the regional educational level, its negative impact spills over to workers' and firm's productivity. Additionally, policies that promote on–the–job–training, by increasing average skills within firms, will play a significant role in the whole economy productivity. Passing on the idea that better qualified workers promote overall productivity, both within regions and counties, will, to some extent, convince employers to invest in own workers human capital, either through training or formal education. The key argument is that by raising firms and/or regions skills benefits individual workers, which in turn increases overall productivity (firm, county and country). The discussion by Krueger and Lindahl (2001) add evidence to the discussion by showing that an increase in aggregate education leads to economic growth.

As in Acemoglu and Angrist (2001), the current findings of external returns to education of the magnitude discussed above are justify significant public, as well as private, investments in education. It follows from the current discussion that local development policies based on investment in local human capital would generate positive effects beyond the private return of the workers themselves. Furthermore, the public support of education leads to greater equality in education, between poor and wealthy families, and hence in individual wages. Increased wage equality tends to increase social cohesion, providing a further efficiency argument in favour of public support of education.

Chapter 5

Factors that influence student achievement gains and performance assessment of the Portuguese public schools

> "It's not just going to school, but learning something while there that matters."
> Hanushek et al., 2008, p.62

5.1 Introduction

The decision to invest in education is taken jointly by the family and the student, who compare the benefits and costs of such investment. That decision benefits, on the one hand, the student himself, through the wage premium (Card, 2001; Harmon et al., 2003; Psacharopoulos and Patrinos, 2004), and, on other hand, the society as a whole. In other words, education of each individual has the possibility to make the others better through educational externalities (Rudd, 2000; Moretti, 2004a; Rodríguez-Pose and Tselios, 2012). Furthermore, a more educated society may lead to higher rates of innovation and a higher level of introduction of new technologies; workers become more productive and firms can introduce new and improved production methods. Higher levels of education at the country level are associated with higher economic growth rates (Krueger and Lindahl, 2001).

In most countries, education is one of the main services provided by governments, and justifies very large investments. For instance, in 2010, the expenditure on educa-

tion in Portugal was 4.8% of GDP, representing 8,559.2 millions of Euros.[1] Recently, it has been recognized that such high spending in education *per se* may not have similar, strong impacts in all countries, which has brought quality issues to the centre of the debate (Hanushek, 1989; Hanushek et al., 1996; Hanushek, 2005a). There are now new political challenges that are henceforth related to the quality rather than the quantity of education. The actual discussion is around the argument that economic well-being is maybe more determined by cognitive skills rather than schooling attainment (Hanushek, 2005a,b; Hanushek et al., 2008; Hanushek and Woessman, 2010b), and the educational assessment issues assume a prominent place in public discussions and in designing public policy.

Countries use a range of techniques for the evaluation and assessment of students, teachers, schools and education systems. International tests on student performance on cognitive skills (as, for instance, PISA) have been developed to evaluate educational systems worldwide, providing external comparators. Evaluation and assessment is seen more and more as pivotal in the building of better and stronger school systems. Assessment is increasingly viewed not as an end in itself, but as an important tool for improving student outcomes. There is also the believe that measuring and publicising student outcomes by school will stimulate schools to take actions to improve their relative performance.

In such context, in most European countries, the assessment of schools has been an instrument of a guarantee of productivity and efficiency of the education systems and individual schools, and used to improve the quality of education (Kirjavainen, 2012). Those assessments make use of school performance data, based on which informed judgements are provided and school improvement is stimulated. Trends of poor performance can lead to different types of actions being taken, like recommendations for school improvement, obligations for school to produce formal plans of improvement, and even sanctions (Dumay et al., 2014).

In Portugal, some efforts to the end of school assessment have been made, namely: *Observatório da Qualidade da Escola* (1992–1999), *Projeto Qualidade XXI* (1999–2002), *Programa Avaliação Integrada das Escolas* (1999–2002), *Projeto Melhorar a Qualidade* (2000–2004), *Programa Aves – Avaliação de Escolas Secundárias* (started in 2000), *Projeto de Aferição da Efetividade da Autoavaliação das Escolas* (2004–2006), *Programa de Avaliação Externa das Escolas* (started in 2006), *Projeto Eficácia Escolar no Ensino da Matemática, 3EM* – University of *Beira Interior*, and *Projeto BESP* – secondary schools benchmarking (Azevedo, 2005; Coelho et al., 2008).

Recently, the national press has published school annual rankings that make a hi-

[1] Information provided by *Instituto Nacional de Estatística*.

erarchy of schools based on the average performance of their students in the national exams. In the earlier exercises, the rankings were produced based solely on the grades, without any controls for the socio–economic context in which schools operate. The literature on student performance has shown, however, that student and family characteristics are major determinants of student achievement (e.g. Carneiro, 2008). Those rankings have been publicly criticized for not considering the student economic and social background, which made them unfair to most public schools. In principle, public schools can not select their students, and all low income students, as well as those from low social backgrounds, attend public schools, which makes the access to top positions in those rankings much more difficult for most of public schools than for the privates. Despite of that, those rankings have had a great impact on public opinion. In an attempt to minimize the criticism of lack of contextualization of student outcomes, more recently, an additional indicator on the social background of the school, based on both the proportion of students benefiting from social support and the parents education, has been introduced, making the analysis of the rankings more accurate. Nevertheless, it has not been possible to collect comparable data for private schools, which makes comparisons still difficult.

There are, however, several issues to improve in those rankings in order to make them a better instrument of school performance analysis. Such context indicator is an aggregate variable, computed at the school level. More interesting conclusions would emerge from an analysis at the student level. Furthermore, by looking at final exam results, but ignoring the knowledge level of the student at the entrance moment, can be very erroneous and may lead to the wrong conclusions. An added value approach would allow to compare across schools the improvement their students achieve as a result of their experience in that school.

This essay aims at contributing to the ongoing discussion by providing an analysis of school rankings at the individual level, that is, based on individual level data, performed in the context of an added value approach. The objective of the essay is twofold: (1) it aims at shedding some light on the determinants of school performance; (2) school rankings are build based on two different approaches, which results are compared to show that the approach is not innocuous.

Namely, attention is given to the answer to a relevant question: What are the factors that determine school performance? School performance is usually assessed through the average student achievement. Several factors have been found to have an influence on student performance. Although there is consensus in the literature about the impact of the student background, the same is not true with respect to class size. Most of the international evidence on class–size effect is for the early grades. Furthermore, class size

effects are under studied for Portugal, as there is no evidence about the class size effect on student performance. In addition, school rankings based on just exam classifications are not particularly useful, as previously mentioned, thus, metric benchmarking that takes into account not only measures of output, namely national exam scores, but also indicators, such students' prior achievement, socio–economic characteristics and school resources would be fundamental to assess the schools' performance (Sarrico et al., 2014), and to better understand which schools add more value to their students. The outcomes will be measured through the results in the final exams of Portuguese and Mathematics.

In a second step, the study will compare the methodologies that are commonly used in academic work on performance assessment of schools, namely the Data Envelopment Analysis (DEA) and the multilevel modelling. More specifically, the essay will compare the results of a non–parametric approach (DEA) with those of a parametric approach (multilevel modelling), in the context of student achievement gains. It seems that there is no other study that compares the DEA with multilevel modelling, in the context of secondary schools. Concerning the parametric models, a multilevel model with 3 levels (student, class, school) is used, through which a ranking of Portuguese schools based on school added value is generated. Note that previous empirical literature, including studies on the Portuguese students and schools, has applied multilevel models with only 2 levels (student, school).

This study benefits from the use of a unique, comprehensive database. It is managed by the Ministry of Education, and contains information at the student level for all students attending Portuguese public schools. Most of the previous studies have used the PISA (Programme for International Student Assessment) data or smaller size databases that were built in the context of some ongoing projects (for instance, 3EM, BESP or AVES).

The results suggest that the most relevant factors for students' achievement gains in both subjects, Portuguese and Mathematics, are the characteristics of students, including the family characteristics. It points out that, as expected, students from lower socio–economic status show worse academic performance. Furthermore, peers' characteristics appear not to have any influence on the students' progress. At the class level, the characteristic that influences the most significantly the student performance is the class size, and it appears that the optimum class size is 23 students. In addition, multilevel model results suggest that most schools have shown no effect on students' achievement gains, mainly in Mathematics. Results also show that there is some similarity between efficiency measures provided by both methodologies, DEA and multilevel modelling, but the choice of methodology to analyse the performance

of schools should be carefully made, as this choice, in principle, affects the conclusions on efficiency and, consequently, the efficiency rankings.

The remainder of the chapter is organized as follows. The next section provides a literature review on the determinants of students performance and efficiency of schools, and an overview of the Portuguese context. Section 5.3 presents the econometric theoretical framework and describes the empirical identification strategy and the data used for estimation. It is followed by Section 5.4 that reports the results divided in two subsections – factors that influence student achievement gains: a multilevel approach and relative performance of schools: multilevel versus DEA approaches. Section 5.5 reports the Concluding remarks.

5.2 The state of the art

School performance assessment has given reason to a huge number of studies that look at the student achievement and its major determinants. In Portugal, some programmes and its evaluation have given good reason for studies that look at student achievement and try to identify which are the factors that lie behind different performances.

5.2.1 School assessment and the determinants of student achievement

An extensive, multidisciplinary empirical literature examines the determinants of student achievement, including studies in Economics (see for a literature review, Hanushek, 1986, 1997, 2005a; Haveman and Wolfe, 1995; Worthington, 2001; Todd and Wolpin, 2003; Hanushek and Woessman, 2010a; Grosskopf et al., 2014), as well as in other social sciences (see Haveman and Wolfe, 1995; Sammons et al., 1995; Rutter and Maughan, 2002; Scheerens, 2004, 2013; Reynolds et al., 2014).

According to that literature, parents' education, individual ability, and peers undoubtedly contribute to student outcomes. Nevertheless schools have a special place in those analyses because they are most directly affected by public policies. As a result, the main purpose of most studies in that body of literature is either to estimate the school effects magnitude and analyse their properties, including consistency, size, stability, the different outcomes of education, and compositional effects (school effects research), or, to identify classroom, school and context factors that make an effective school (effective schools research) (Torrencilla, 2003; Scheerens, 2013; Reynolds et al., 2014). Although most studies are focused on academic achievement, some researchers have looked at non–cognitive outcomes, such as, the student well–being and achieve-

ment motivation (see Reynolds et al., 2014), attendance, attitudes and behaviour (see Sammons et al., 1995). It is noteworthy that the literature suggests that schools tend to have a larger effect on student achievement than on non–cognitive outcomes (Reynolds et al., 2014).

A key concept that arises in literature is the concept of school effectiveness based on the idea that "schools matter, that schools do have major effects upon children's development and that, to put it simply, schools do make a difference" (Reynolds and Creemers, 1990, p.1). Therefore, an effective school is "a school in which students progress further than might be expected from a consideration of its intake" (Mortimore, 1991, p.1) or, in another way, "*una escuela es eficaz si consigue un desarrollo integral de todos y cada uno de sus alumnos mayor de lo que sería esperable teniendo en cuenta su rendimiento previo y la situación social, económica y cultural de las familias*" (Torrencilla, 2003, p.2). This school effectiveness perspective emphasizes the learning outcomes actually achieved in the education process. The literature provides a variety of models for assessing the educational quality, that can be grouped into three main categories, depending on whether the emphasis is put on results, or on school improvement processes, or on the organizational aspects (Scheerens, 2004, 2013).

In the economics of education literature, one of the most common conceptual frameworks employed takes the form of a production function, also referred to as "input–output" analysis (Hanushek, 1986, 1997; Worthington, 2001). A school or education system can be considered as the production units of the service "education", in which the inputs are many students' and schools' characteristics; the output corresponds to the results achieved by the students at the end of a cycle of studies. The transformation process corresponds to the set of teaching methods, curriculum and organizational conditions that allow students to acquire the knowledge (Hanushek, 1986, 1997; Scheerens, 2013). In this sense, most of the empirical studies define productivity as the main criterion, since the effectiveness is a causal concept, in which the relationship between means and ends is equivalent to the cause–effect relationship (Scheerens, 2004).

Several research works emerged as a reaction to the seminal study by Coleman, known as Coleman report and published in 1966, where the results show that the influence of school internal factors (such as characteristics of schools, class and teacher) on the outcomes of the students is nearly zero as compared with the effect of other variables such as the socio–economic context, ethnicity, and the cultural level of students and families (Coleman, 1966). That conclusion was highly publicized, and become known as the "Schools don't matter!" statement.

Since then, a vast literature on the subject has been developed, but the findings on the relationship between school resources and the student achievement are not

consensual. On the one hand, in a review of 400 studies on student achievement, Hanushek (1997) concludes "that there is no a strong or consistent relationship between student performance and school resources" (Hanushek, 1997, p.141), after controlling for family characteristics. On the other hand, Lee and Barro (2001) suggest that more school resources enhance educational outcomes, i.e., their "findings are consistent with a view that inputs from schools, families and communities are all important in improving school quality" (Lee and Barro, 2001, p.485).

In order to analyse the school effectiveness, different measures of education outcomes have been used. Most of the economics of education studies have been looking at the test scores, in particular, the maths, reading and science scores (e.g. Hanushek, 1986, 1997; Lee and Barro, 2001; Berlinski et al., 2009; Abdulkadiroglu et al., 2011; Denny and Oppedisano, 2013; Brunello and Rocco, 2013). Other studies, however, have employed alternative quantitative measures, such as the students' success rates (e.g. Kirjavainen and Loikkanen, 1998; Mancebón and Bandrés, 1999; Berlinski et al., 2008), attendance rates (e.g. Berlinski et al., 2008), repetition rates (e.g. Lee and Barro, 2001; Stevens and Schaller, 2011), the percentage of students who have not dropped out of school (Lee and Barro, 2001), and subsequent labour market earnings (Hanushek, 1986, 1997, 2005a).

Three groups of factors emerge as the main aspects that may influence the students' outcomes, and, consequently, school performance: student and family, peers and school characteristics (see Hanushek, 1986, 1997; Hanushek and Woessman, 2010a; Haveman and Wolfe, 1995; Grosskopf et al., 2014, for literature reviews).

Student characteristics, oftenly used as control variables, are gender, age, ethnicity, special educational needs, eligibility for a free or reduced price lunch and years of pre–schooling education (e.g. Thanassoulis and Portela, 2002; Hanushek et al., 2003; Cullen et al., 2005; Berlinski et al., 2008; Husain and Millimet, 2009; Abdulkadiroglu et al., 2011; Agasist and Vittadini, 2012; Denny and Oppedisano, 2013; Thieme et al., 2013; Goldhaber and Hansen, 2013). Family inputs tend to be measured by socio–demographic characteristics, such as parents' education level, income, parents' occupation, family size and number of books in the household (e.g. Summers and Wolfe, 1977; Kirjavainen and Loikkanen, 1998; Feinstein and Symons, 1999; Lee and Barro, 2001; Ladd and Walsh, 2002; Woessman, 2003; Cullen et al., 2005; Clotfelter et al., 2007; Perelman and Santin, 2011; Stevens and Schaller, 2011; Kirjavainen, 2012; Brunello and Rocco, 2013; Denny and Oppedisano, 2013; Thieme et al., 2013; Goldhaber and Hansen, 2013).

The positive and significant effects of the individual and family characteristics, in particular, the prior student achievement, as well as the family background, seem to

be consensual in the literature (Hanushek, 1986). This means that the higher the prior achievement scores, the higher the final achievement scores (Hanushek et al., 2003; Cullen et al., 2005; Kirjavainen, 2012). Growing up in a poor or low–income family tend to have a negative impact on the educational outcomes (Hanushek, 1986; Lee and Barro, 2001; Haveman and Wolfe, 1995). The parents' education level appears to positively influence the student performance (Hanushek, 1986; Worthington, 2001; Lee and Barro, 2001; Woessman, 2003; Fertig and Wright, 2005; Kirjavainen, 2012). The structure of the family, however, shows a negative impact on the student outcomes. Namely, growing up in a single–parent (Kirjavainen, 2012) or in a step–parent family has a negative effect on the educational attainment (Haveman and Wolfe, 1995). It has been pointed out that family background and socio–economic factors are more important determinants of student performance as compared to school resources (Lee and Barro, 2001).

Empirical evidence on peer effects is rather mixed. Several existing studies recognise a positive and strong effect of the students' peer group on the educational outcomes. The average peer group achievement (Hanushek et al., 2003), the average education of mothers of other students in the same class (McEwan, 2003), and a high proportion of girls (Kirjavainen, 2012) have a highly significant effect on student performance. Simultaneously, some sets of results provide little evidence that the growth of students' achievement is affected by the average education of the fathers, the average income, or the heterogeneity of peers in terms of variation in achievement levels (Hanushek et al., 2003; McEwan, 2003). There is also evidence that the higher the share of immigrant pupils in schools, the lower the performance of native students, especially those with a disadvantaged parental background (Brunello and Rocco, 2013).

The most controversial results are related to school resources. Measures of school resources usually include the financial aggregates of resources, such as expenditure per student (e.g. Woessman, 2003; Chiang, 2009; Kirjavainen, 2012; Brunello and Rocco, 2013) and measures of other resources in schools, such as administrative inputs, class size, teacher–pupil ratio and facilities (e.g. Akerhielm, 1995; Kirjavainen and Loikkanen, 1998; Lee and Barro, 2001; Woessman, 2003; Krueger, 2003; Fertig and Wright, 2005; Chiang, 2009; Thieme et al., 2013; Deutsch et al., 2013; Denny and Oppedisano, 2013). In general, the results suggest only weak relationships between school expenditures and student performance. According to Hanushek (1986, 1997) and Woessman (2003) there is no strong or consistent relationship between school resources and student results, whereas Krueger (2003) and (Lee and Barro, 2001) conclude that increases in school expenditure lead to modest gains in student performance. Furthermore, some studies suggest that the higher the educational expenditure per student (measured

at the country level) the lower the student performance in mathematics and science (Woessman, 2003). Kirjavainen (2012) concludes that teaching expenditures per student have a negative and significant effect on tests scores, as well.

In the context of school resources, some literature provide evidence that smaller classes have a positive effect on student achievement (Lee and Barro, 2001; Krueger, 2003; Akerhielm, 1995; Berlinski et al., 2009; Krassel and Heinesen, 2014); pupil–teacher ratio is negatively correlated with the test scores confirming that smaller classes are associated with better performing students (Lee and Barro, 2001). For example, a ten student reduction in class size is typically associated with 0.1–0.3 standard deviation increase in student test scores (Hanushek and Rivkin, 2010). Nevertheless, that result is not consensual. Denny and Oppedisano (2013) concludes that increases in class size lead to improvements in student's mathematics scores and Coupé et al. (2016) suggest a negligible class size effect on test scores.

That set of variables also include variables to characterise the school context, namely, the type of school (public or private) (e.g. Woessman, 2003; Cullen et al., 2005; Berlinski et al., 2008; Abdulkadiroglu et al., 2011; Kirjavainen, 2012; Krassel and Heinesen, 2014), school size (e.g. Benabou et al., 2009; Kirjavainen, 2012; Coupé et al., 2016) and its location (e.g. Akerhielm, 1995; Agasist and Vittadini, 2012; Kirjavainen, 2012; Coupé et al., 2016). School size appears to have a small effect on student achievement (Kirjavainen, 2012; Coupé et al., 2016), although some studies suggest that this characteristic is negligible (Kirjavainen and Loikkanen, 1998). Public schools are found to be more efficient than private schools (Kirjavainen and Loikkanen, 1998; Thieme et al., 2013), but the result is not consistent across studies (Kirjavainen, 2012). According to Woessman (2003) some institutional features of schooling systems are found to be positively related to student performance, such as, central examinations, centralized control mechanisms in curricular and budgetary affairs; school autonomy in process and personnel decisions; individual teachers having both incentives and powers to select appropriate teaching methods; encouragement of parents to take interest in teaching matters, among others.

Furthermore, concerning the environmental context, peer–group inputs at the school–level, when included, are typically aggregate summaries of socio–demographic characteristics of other students in the school, such as the percentage of boys/girls (e.g. Kirjavainen, 2012; Brunello and Rocco, 2013), the share of immigrant students (e.g. Brunello and Rocco, 2013), the proportion of students whose families have economic difficulties or the proportion of students eligible for a free or reduced price lunch (e.g. Hanushek et al., 2003; Agasist and Vittadini, 2012), the average level of education of the mothers (e.g. Perelman and Santin, 2011), the percentage of black or/and the

proportion Hispanic (e.g. Hanushek et al., 2003).

Early studies used to include only contemporaneous inputs (Hanushek, 1986). Nevertheless, education is a cumulative process, and, as such, both historical and contemporaneous inputs influence current performance (Hanushek, 1997). The concept of effective school is attached to the concept of the added value of schools, which is a measure of school performance. Value–added specifications differ from the contemporaneous specifications only in the fact that the latter include a measure of the prior achievement of the student (Todd and Wolpin, 2003).

According to the literature, the value–added indicator is the conceptually appropriate indicator for measuring school performance and it is defined as the contribution of the school to the educational outcome (Meyer, 1997); in other words, value–added is defined as the progress that schools help their students to make relative to their different starting points (Portela et al., 2011). The key idea is to isolate the contribution of the school to the student achievement from all other factors that may have a role in student outcomes (Meyer, 1997). Studies that adopt the value–added approach are, for example, Ladd and Walsh (2002); Cullen et al. (2005); Buddin and Zamarro (2009); Kukla-Acevedo (2009); Agasist and Vittadini (2012); Kirjavainen (2012) and Goldhaber and Hansen (2013).

In most studies in the economics of education, least squares regression techniques have been used (for instance Akerhielm, 1995; Woessman, 2003; Cullen et al., 2005; Berlinski et al., 2008; Brunello and Rocco, 2013). Nevertheless, starting in the 1980s, methodological advances combined with the availability of pupil–level data have allowed for significant improvements in the studies that have shown the scientific properties of the school effects (Thanassoulis and Portela, 2002; Reynolds et al., 2014).

Multilevel models and large–scale longitudinal research recognize the complex and hierarchical structure of most educational systems (Reynolds et al., 2014). Multi–level modelling can be used to estimate the expected outcomes at different hierarchical levels, for instance, considering that pupils are taken as level 1, cohorts (classes) as level 2, schools as level 3 and so on (Thanassoulis and Portela, 2002). The value–added perspective taken in the context of a multilevel approach allows for a framework much closer to the reality of schools and students. An estimate of the expected performance of a school, given the initial characteristics of the enrolled students (Goldstein, 1995), is an outcome of multilevel models. It is noteworthy that most multilevel model studies only combine student level and school level factors that impact on student achievement, making use of two level models (e.g. Agasist and Vittadini, 2012; Thieme et al., 2013).

Note that other estimation techniques are used to estimate education production functions, namely, instrumental variables (for instance Ladd and Walsh, 2002; Brunello

and Rocco, 2013; Denny and Oppedisano, 2013), panel data models: fixed–effects (for instance Lee and Barro, 2001; Hanushek et al., 2003; Fertig and Wright, 2005; Clotfelter et al., 2007; Kirjavainen, 2012) and random–effects (for instance Kirjavainen, 2012), among others (see Reynolds et al., 2014). Differences in the findings are due to variations in sample inclusion criteria and the choice of control variables (Hanushek, 1986, 1997).

Alternatively, frontier models have been estimated. Frontier models can be classified into two groups: parametric models, which base its calculations on regression techniques, as Stochastic Frontier Approach (SFA); and non–parametric models which rest on mathematical programming, as Data Envelopment Analysis (DEA) (Mancebón and Bandrés, 1999). Models based on DEA, originally developed by Charnes et al.(1978), have been the most popular non–parametric methodology to evaluate schools (Worthington, 2001, for a literature review), because do not impose any functional form as opposed to the parametric approach which assumes a theoretical functional form of the production function, given the existing technology, to estimate the efficiency frontier, representing a limitation in the education sector where it is difficult to specify the appropriate model that links the inputs and the outputs (Mancebón and Bandrés, 1999). A less–constrained alternative to DEA sometimes employed in the analysis of efficiency is known as Free Disposal Hull (FDH), which also is a non–convex non–parametric model. Both DEA and FDH assume all deviations from the frontier to be the result of inefficiency (Worthington, 2001). The advantages claimed for this methodology in the assessment of school efficiency have been reinforced by its intensive use (Worthington, 2001).

Summing up, there is a huge literature on the analysis of the education production function, in which the output is the student achievement, whereas the inputs are student and family, peers and school characteristics. There is some consensus on the effects of individual variables on student performance. The most controversial results are those on the effects of school resources. As education is a cumulative process, previous performance cannot be ignored, which has been the argument for several studies taking an added value approach. Data Envelopment Analysis and Multilevel models are the main methodologies in use in these type of analyses. Parametric and non–parametric methods have been applied and compared in the context of the assessment of higher education institutions (e.g. Johnes, 2006), but there seems not to exist any study comparing the DEA with the multilevel modelling in the context of secondary schools. Furthermore, multilevel models have considered two levels (student, and school); three–level models, considering the student, the class and the school are much less common, although more adequate to most of the data structures.

5.2.2 The Portuguese setting

In Portugal, there have been some studies on the quality of schools and projects of external evaluation of schools; the first school assessment project started in the beginning of the 90s (Azevedo, 2005; Coelho et al., 2008). There have been carried out some projects (as in Azevedo, 2005; Coelho et al., 2008), namely:

- *Observatório da Qualidade da Escola:* it was an initiative of the Ministry of Education that was developed between 1992 and 1999 and covered the different educational levels. With the aim of measuring gain and loss of quality from year to year, the project based its evaluative action in the identification of 15 main indicators related to the family background of students, educational resources and its management, the context and functioning of the school, and the educational outcomes.

- *Projeto Qualidade XXI:* it was started by the *Instituto da Inovação Educacional* (Ministry of Education) in 1999 and it came up under the *Projeto–Piloto Europeu sobre a Avaliação da Qualidade na Educação Escolar*. The main purpose of the project is to foment strategies that contribute to the improvement of educational quality of schools, covering four key dimensions: learning outcomes, processes at the classroom and at the school level, and relationships with the context. It covered schools of the 2^{nd} and the 3^{rd} cycle, as well as, secondary schools.

- *Programa Avaliação Integrada das Escolas:* it is an external evaluation project by the *Inspeção–Geral da Educação (IGE)*, that began in the school year 1999/2000 and took a period of 3 school years. It considers four levels of analysis: organization and management, academic student performance, climate school and education, and teaching and learning. All educational levels were included. The project had as main goals to provide performance accountability of the education system and encourage schools to develop self–evaluation processes.

- *Projeto Melhorar a Qualidade:* it was based on the European for Quality Management model and it was developed between 2000 and 2004, in order to provide the possibility of performing a voluntary exercise of self–evaluation of 46 private schools. The schools were evaluated based on the idea that the performance of schools – students, teachers, non–teaching staff and community – is achieved through leadership that energizes the institutional policy and the strategy of people management, resources, partners and educational processes.

- *Projeto de Aferição da Efetividade da Autoavaliação das Escolas:* it was an activity of external evaluation in the meta–evaluation modality developed, between

2004 and 2006, by the *Inspeção–Geral da Educação (IGE)*. It aimed to develop a critical attitude and self–questioning of the work done in schools and to obtain an overview of the status of self–assessment dynamic. It covered schools of all different educational levels.

- *Programa de Avaliação Externa das Escolas:* this project intends to develop an external evaluation model in coordination with the schools self–evaluation. Since 2006, the Ministry of Education has appointed a working group – *Inspeção–Geral da Educação e da Ciência (IGEC)*, whose jurisdiction is based on the development of reference for the external evaluation and self–evaluation of schools. Currently, only this one is still open.

The Law no.31/2002, December 20, approved the evaluation system of public establishments of the pre–school education, as well as the basic and secondary education, by defining general guidelines for both self–evaluation and external evaluation. Since 2006, this program has been developed and the first evaluation cycle closed in 2011 (Almeida et al., 2011). Thereafter, *IGEC* was instructed to continue the external evaluation program, for a new cycle of external evaluation, as a result of the model proposed by the working group. Every academic year, *IGEC* selects the schools that will be visited by a team of two inspectors and an external evaluator, which produces a qualitative report, on which the school gets a classification[2] in three domains: outcomes (educational and social outcomes, community recognition), provision of educational services (planning and articulation, teaching practices, monitoring and evaluation of teaching and learning) and leadership and management (leadership, management, self–evaluation and improvement) (Nunes et al., 2013). Despite the external evaluation carried out by the *IGEC* is a valuable tool for schools to the extent that provides a guide for self–evaluation, it does not allow them to compare their performance with other schools with the same conditions (Portela et al., 2012), because each school follows its own evaluation program just following the general guidelines in the dimensions of the assessment provided by law (Portela et al., 2011).

There are, however, other school assessment programs that have been developed by other Portuguese entities, namely:

- *Projeto BESP*: Secondary Schools Benchmarking, developed by School of Economics and Management of the Catholic University of Porto – it is an innovative web platform in Portugal that supports schools in internal and external evaluation. In the *BESP* there is a set of indicators of the performance of schools at

[2]These classifications result from the application of a five–level scale: Excellent, Very Good, Good, Sufficient and Insufficient (Nunes et al., 2013)

various levels: outcomes of students, school context, resources and processes. To determinate these indicators the outcomes of national tests are used, as well as, data from the schools. The indicators provide an overview of the evolution of school performance over time (internal evaluation) and they allow schools to see graphically their position relative to other schools involved in this project (external evaluation). The evaluation of schools is performed by the DEA method, adopting a value–added perspective (Portela et al., 2011), but the use of aggregated results of students at school level as the school value–added is an imperfect measure (Portela, 2011).

- *Programa AVES*: Secondary Schools Evaluation developed by *Fundação Manuel Leão*[3] – it is an external evaluation program, which started in 2000. It is not a nationwide program; in 2000 the project was comprised only 13 schools, public and private, whereas in 2009 that number has increased to 60 schools. The key purpose is to identify factors that promote or inhibit the quality of their performance with the actions and projects that they can mobilize in order to improve this same performance. External evaluation is based in a set of own tests (academic and non–academic), of which data at the student level are collected and analysed. A non–parametric method – DEA (value–added perspective) – is used. It is organized into four levels: entrance (initial results of the students at the beginning of the study cycle); context (sociocultural involvement and type of school); processes (institutional and classroom); outcomes (mainly focuses on students (learning of some curriculum areas – Mathematics, Portuguese Language, History and Science). The project produces reports to schools that present their performance and compares each school with the average of schools considered under the program (Azevedo, 2005; Coelho et al., 2008; Portela et al., 2007b, 2012).

- *Projeto Eficácia Escolar no Ensino da Matemática, 3EM* developed by University of *Beira Interior* – it was designed to assess students' Mathematics ability and skills at any grade of compulsory education. Its development was based on a reference matrix that includes the purposes of the assessment, the contents and descriptors, and some operational specifications. This is a longitudinal study, in which the data collection at the student level takes place at the beginning and in the end of each academic year and the target population is defined by the

[3] *Fundação Manuel Leão* is a private institution founded in January 1996 by *Manuel Valente Leão*, a priest of Porto. Its main purposes are to promote the public good in education, culture, artistic activity and socio–charitable action. It has been developing and supporting a number of specific projects in the fields of Education, Arts and Culture.

students of the 1^{st}, 2^{nd} and 3^{rd} cycles of basic education, in the region of *Cova da Beira*.[4] The target population is 7,727 students (3,418 in 1st cycle, 1,709 in 2nd cycle and 2,602 in 3rd cycle). The total number of schools involved is 105 (84 in 1^{st} cycle, 9 in 2^{nd} cycle and 12 in 3^{rd} cycle). The data were collected in the academic years 2005/6, 2006/7 and 2007/8 (Dias et al., 2008; Ferrão, 2012; Ferrão and Couto, 2014).

Those projects have given reason to academic work. Under the *Programa AVES*, Portela et al. (2007b) compares two methods for the value–added determination in a sample of 11 Portuguese schools: DEA and the method used by Department for Education and Skills (DfES) in UK. The value–added of schools is determined for each student and variables used are: results at the entrance of the 3^{rd} cycle (classifications of some subjects in 7^{th} grade) and secondary (classification of Portuguese and Mathematics in 9^{th} grade), and the results on the exit of the 3^{rd} cycle, as well results on the exit of secondary (Portuguese and Mathematics in 11^{th} grade). The survey results show substantial differences between the methods used. DEA analyses the best results observed at the exit for a given level of initial results, while the DfES method analyses the average performance of students. Thus, authors emphasize that DfES method confounds the student effects with the school practices effects, while in the DEA approach school and student effects are separately considered (Portela et al., 2007b).

Under the *Projeto Eficácia Escolar no Ensino da Matemática, 3EM*, Ferrão (2012); Ferrão and Couto (2014) have investigated the issue of the assessment of the school performance. The study applies multilevel models to a longitudinal dataset, gathered from a representative sample of students in primary education collected under the project *3EM* over 3 academic years. A traditional value–added model (a two–level random intercept model) with pupils as level 1 units and schools (in Ferrão and Couto, 2014) or classes (in Ferrão, 2012) as level 2 units is estimated. The student's score in Maths is the response variable (3EMat score at the end of academic year) and the independent variables are prior achievement (3EMat score at the beginning) and socio–economic status. Results show that both prior achievement and socio–economic status are positively associated with students' outcomes in mathematics.[5]

[4] *Cova da Beira* is a NUTS III (Nomenclature of Territorial Units for Statistics, level III) region that includes the counties of *Covilhã*, *Fundão* and *Belmonte*

[5] The marginal effect of prior achievement is, at least, two and half times stronger than the effect of socio–economic status (both variables were standardised). In addition, the intra–class correlation varies between 0.13 and 0.15, and the variance partition coefficient (based on the value–added model) varied between 0.05 and 0.13 (Ferrão and Couto, 2014). Furthermore, these works present evidence about the stability of school value–added indicators: when classes are the level 2 units (Ferrão, 2012), 65% of them maintained the level of performance for, at least, two years, and 12% maintained a stable level for 3 years. When schools are the level 2 units (Ferrão and Couto, 2014), the school value–added

In addition to the academic work carried out under these projects, recently the evaluation of the performance of Portuguese schools has deserved a growing concern among researchers, in line with the international and national concerns and public debates. Studies on Portuguese schools consider cognitive outcomes as the dependent variable: score tests (mainly in math and reading) (Dias and Ferrão, 2006; Pereira and Moreira, 2007; Portela et al., 2012); success rates (Oliveira and Santos, 2005; Sarrico and Rosa, 2009; Sarrico et al., 2010); completion rates and non drop–out rates (Sarrico and Rosa, 2009; Sarrico et al., 2010).

Three main groups of factors may have an effect on students' performance: students characteristics, such as, ability at entrance or prior achievement, gender, grade attended, age (Dias and Ferrão, 2006; Sarrico and Rosa, 2009; Sarrico et al., 2010; Portela et al., 2012); family aspects, for instance, socio–economic level, parents' education, household composition, cultural and financial level of the family (Dias and Ferrão, 2006; Sarrico and Rosa, 2009; Sarrico et al., 2010); and school/teachers characteristics, such as, the average socio–economic and cultural level of the school, percentage of girls, location of school, conditions of teaching and learning in school, average age of teachers, school climate, private/public and number of students, teachers and classes (Oliveira and Santos, 2005; Dias and Ferrão, 2006; Sarrico and Rosa, 2009; Sarrico et al., 2010; Pereira and Moreira, 2007). Oliveira and Santos (2005) and Pereira and Moreira (2007) consider also environmental variables, namely, GDP per capita, unemployment rate, social development index in 1998, divorce rate, number of physicians per 1000 people (in Oliveira and Santos, 2005), the average domestic electricity consumption, the illiteracy rate and a Health Conditions Index of the municipality where the school is located (in Pereira and Moreira, 2007) to devise the factors that most impacted the efficiency scores of schools.

The findings for Portugal on the impact of student and family characteristics show that the observable factor that contributes the best to the inequality in student performance is the family background, and the school resources have a limited role on student results (Carneiro, 2008; Pereira, 2010; Pereira and Reis, 2012). Socio–economic background has a strong effect on test scores and the number of books at home is the most important predictor of student performance (Pereira, 2010; Pereira and Reis, 2012). Parents' education (secondary and university education) has a positive impact on student achievement (Pereira and Reis, 2012) and, particularly, Carneiro (2008) suggests that better educated fathers provide better home and school environments for their children, which then translate into better test scores. However, Pereira (2010) sug-

indicators suggest that at least 26% of the schools present positive levels of value–added effects for 3 years, 15% of schools present negative levels for the same period, and 85% of schools maintained at the same performance level for at least 2 adjacent years.

gests this inputs is no statistically significant. Regarding parents' occupation, only the intellectual/specialized categories have a positive influence on test scores, but the effect is small (Pereira, 2010; Pereira and Reis, 2012). Male students perform better in mathematics and female students have better reading performance (Pereira, 2010). Furthermore, the relationship between prior achievement and student performance is, at least two and a half times stronger than the relationship between socio–economic status and student performance (Ferrão, 2012).

There is, however, some differences in results between studies. Pereira and Moreira (2007) and Portela et al. (2012) suggest that private schools are more efficient than public schools, whereas Dias and Ferrão (2006) propose that, when the classroom climate and the teaching and learning conditions are controlled for, the private versus public effect on academic achievement is not statistically significant. Furthermore, it is found that 42% of the schools evaluated changed ranks between 2008 and 2009 by more than 50 places, and private schools can show more stable results over time than public schools (Portela et al., 2012). Portela et al. (2013) explores the changes in value–added of schools for those cohorts of students finishing secondary education and suggest a volatile behaviour of schools performance over time, reflected also in catch–up and frontier shift components that do not show clear trends over time. Pereira and Moreira (2007) points to the existence of technical inefficiency, i.e., for the resources expended the exams results could be higher on average 10% and "results indicate that the 'quality' of the teachers have more effect on production than on the 'quantity' ".

School environment characteristics (e.g., unemployment rate, access to health care services, adult education and living infrastructures) are determinants of school efficiency (Oliveira and Santos, 2005) and suggest the existence of performance inequality between regions (Oliveira and Santos, 2005; Pereira and Moreira, 2007; Pereira and Reis, 2012; Portela et al., 2012), namely, there are differences between the coastline and inland Portugal (Oliveira and Santos, 2005); as an example, the average added value of Portuguese schools is the highest in the districts of *Braga*, *Lisboa*, *Porto*, *Aveiro* and *Santarém* (Portela et al., 2012). In other words, "...schools located in cities with better economic conditions and higher educational levels reach comparatively better performance" (Pereira and Moreira, 2007, p.116).

Finally, Sarrico and Rosa (2009) compare their results with those obtained by the panels of evaluators within the national evaluation programme of schools. The authors conclude that the performance of the schools under the metric benchmarking exercise does vary substantially. However, the results do not correlate with the assessment made by the *Inspeção–Geral da Educação*. The authors suggest that this might happen that panel of evaluators "find it difficult to take a value–added approach and instead assess

schools mostly on output measures, such as exam classifications and graduation rates, having difficulty in taking into account pupils' socio–economic background and other variables not under the control of the school" (Sarrico and Rosa, 2009, p.767).

Several methodologies are used in those studies on the Portuguese schools, namely, the more commonly used are the Stochastic Frontier Approach (SFA) (Pereira and Moreira, 2007), DEA (Portela et al., 2007b; Sarrico and Rosa, 2009; Sarrico et al., 2010; Portela et al., 2012, 2013), Free Disposal Hull (FDH) which is a variant of DEA (Oliveira and Santos, 2005) and multilevel models with two levels (students and schools) (Dias and Ferrão, 2006). In several cases, two or more models considered within the same analysis. Pereira and Reis (2012) following Carneiro (2008), using PISA data, apply a regression approach to analyse regional differences in the performance level of inequality of schools in Portugal. Oliveira and Santos (2005) use a two–stage approach: in the first stage, output–based technical efficiency scores and slacks are generated using FDH, and, in the second stage, a truncated dependent variable model and the single bootstrapping procedure are used to identify environmental variables that influence the school performance. Sarrico et al. (2010) use two models of efficiency of schools which differ on the type of inputs: school choice model and school management model, and also in Portela et al. (2012) two types of DEA analysis are performed: an output–oriented model that restricts output (exam scores) weights to be linked to the number of students that have done an exam and a model that restricts factor weights to be equal for all schools (restrictions are linked to the total number of exams done at national level).

In brief, in Portugal, preoccupation on quality issues has been present in peoples minds for the last decades. Several projects have been implemented aiming at promoting and encouraging schools evaluation. Several research work has been conducted on the results of those projects. Actually, a good number of studies has been applied to small samples of Portuguese schools or PISA data; the works of Pereira and Moreira (2007) and Portela et al. (2012) are exceptions, as they evaluate almost the full sample of secondary schools. There is still lack of studies that look at the performance of the overall population of Portuguese schools, mainly due to data unavailability (Portela et al., 2012).

5.3 Empirical approach

5.3.1 Models

The literature on school assessment has highlighted the relevance of measuring school performance by taking an added value approach. Two main types of models have been

used to look at the main determinants of school performance in that context, as well as school rankings, namely, the multilevel regression models and the data envelopment analysis.

5.3.1.1 Multilevel regression models

Early parametric studies measuring efficiency in education applied the method of Ordinary Least Squares (OLS) to estimate a school production function. However, those models rely on the assumptions of independence of individuals/observations, which seems not to be consistent with the nature of the data they use. The data in use is hierarchical, i.e., the individuals/students under study (corresponding to the lowest–level units, level 1) are grouped into classes (level 2), which in turn are grouped into schools (level 3). It follows that students in the same class are subject to identical conditions, as well as students from the same school who tend to have more in common with each other as compared to those attending other schools. Therefore, the random errors of the students attending the same school will not be independent if the school has some impact on student performance, and the assumption of independence of observations is possibly violated.

When the assumption of independence of observations is violated, i.e., the units of the same level are correlated, the multilevel regression models are the most suitable (Goldstein, 1995). When the method ignores the hierarchical structure of the data it leads to underestimation of the standard–errors of the regression coefficients, and, consequently, the results may falsely be significant. According to the literature, the estimates obtained from the multilevel regression models are correct given the data structure, as well as the multilevel models are highly suitable to isolate the variation in students achievement gains due to effects of class/teachers and school effects. By decomposing the error variance according to hierarchical levels, the multilevel regression model allows for a better understanding/explanation of the educational process (Ferrão et al., 2002; Raudenbush and Bryk, 2002; Goldstein, 1995, 1997, 2011; Hox, 2010).

The multilevel regression model assumes that there is a hierarchical data structure, constituted by a single dependent variable, which is measured at the lowest level, and several explanatory variables that can exist at all hierarchical levels (Hox, 2010). Multilevel models accounts for the shared variance in hierarchically structured data; the technique accurately estimates lower level slope coefficients (e.g., student level) and their implementation in estimating higher–level effects (e.g., classroom level; school level).

In the empirical literature, multilevel regression models (by Goldstein, 1995), also

known as hierarchical linear models (by Raudenbush and Bryk, 2002) can be subdivided into two types: random coefficient models (by Leeuw and Kreft, 1986, cited by Hox (2010); Longford, 1995) and variance component models (Longford, 1987, cited by Hox, 2010), which differ only in a few specific aspects. The variance component model decomposes the variance of the dependent variable in different components for each hierarchical level. It is assumed that the intercept varies between macro units, but regression coefficients are considered fixed, whereas in the random coefficient model the intercept and the slope coefficients are assumed to vary across macro units (Goldstein, 2011).

For the purpose of the present essay, a three–level regression model is specified. The model divides the variance in the dependent variable into three components: variance between students (level 1), variance between classes (level 2), and variance between schools (level 3). Figure 5.1 presents the structure of the educational data according to a multilevel model with three levels.

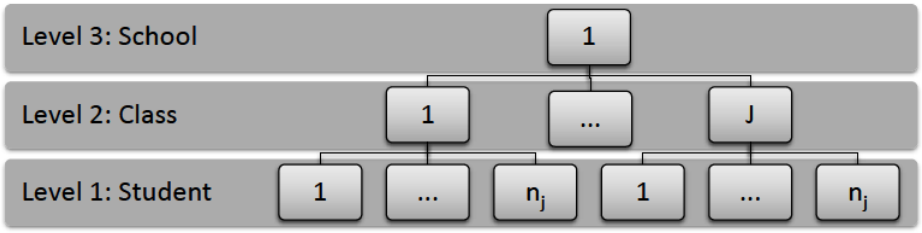


Figure 5.1: Hierarchical data structure with three levels

According to the Raudenbush and Bryk (2002) and considering a simple multilevel model, at level 1, the specification describes the dependent variable, the academic performance of the student, as a function of the student and family characteristics. That is,

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk} X_{ijk} + e_{ijk} \tag{5.1}$$

where specifications at level 2 are the following:

$$\beta_{0jk} = \gamma_{00k} + u_{0jk} \tag{5.2}$$

$$\beta_{1jk} = \gamma_{10k} + u_{1jk} \tag{5.3}$$

and the level 3 specification is given by:

$$\gamma_{00k} = \pi_{000} + r_{00k} \tag{5.4}$$

where Y_{ijk} is the achievement of student i $(i = 1, \ldots, n_j)$ of class j (or taught by teacher j) $(j = 1, \ldots, J)$ in school k $(k = 1, \ldots, K)$; X_{ijk} represents the control variables for

student and family characteristics; β_{0jk} is the average achievement of students of class j in school k; β_{1jk} are the coefficients at level 1 and measure the effects of student and family characteristics on individual student achievement; γ_{00k} denotes the average of intercepts for all classes of school k, i.e., is the average achievement of students of classes in school k; γ_{10k} represents the average of coefficients for all classes of school k; π_{000} is the average achievement of students across k schools. Finally, e_{ijk}, u_{0jk} and u_{1jk}, and r_{00k} are the random error at level 1, 2 and 3, respectively. Note that, u_{0jk} and u_{1jk} are the error terms associated to the class j in school k associated to the intercept and to the slope, respectively, $cov(u_{0jk}, u_{1jk}) = \sigma_{u_{01}}$. The random error terms are independent, i.e., the errors of the different hierarchical levels are uncorrelated:

$$cov(e_{ijk}, u_{0jk}) = cov(e_{ijk}, u_{1jk}) = cov(e_{ijk}, r_{00k}) = 0,$$

$$cov(u_{0jk}, r_{00k}) = cov(u_{1jk}, r_{00k}) = 0$$

and

$$\begin{gathered} e_{ijk} \sim N(0, \sigma_e^2) \\ u_{0jk} \sim N(0, \sigma_{u_0}^2) \text{ and } u_{1jk} \sim N(0, \sigma_{u_1}^2) \\ r_{00k} \sim N(0, \sigma_r^2) \end{gathered}$$

where σ_e^2 is the variance of the error term and represents the variance of the student achievement in the same class; $\sigma_{u_0}^2$ and $\sigma_{u_1}^2$ represent the variability across classes relating to intercept and the regression coefficient, respectively, and σ_r^2 is the variance inter–schools.

The three–level random coefficients regression model can easily be generalized by adding explanatory variables to the specifications of all hierarchical levels. Thus, the following specification applies:

$$Y_{ijk} = \beta_{0jk} + \sum_{p=1}^{P} \beta_{pjk} X_{pijk} + e_{ijk} \tag{5.5}$$

$$\beta_{pjk} = \gamma_{p0k} + \sum_{q=1}^{Q} \gamma_{pqk} W_{qjk} + u_{pjk}, \qquad p = 0, \ldots, P \tag{5.6}$$

$$\gamma_{pqk} = \pi_{pq0} + \sum_{s=1}^{S} \pi_{pqs} Z_{sk} + r_{pqk}, \qquad p = 0, \ldots, P; q = 0, \ldots, Q \tag{5.7}$$

where W_{qjk} and Z_{sk} are vectors of explanatory variables at level 2 and level 3, representing the class characteristics and school attributes, respectively. In addition, P, Q and S denote the number of explanatory variables in the vectors X, W and Z, in first, second and third level, respectively.

This model can be written by means of a single equation in which student achievement is a function of the explanatory variables related to the student, the class and the school. By replacing equations (5.6) and (5.7) in equation (5.5), the model equation is given by:

$$Y_{ijk} = \left(\gamma_{00k} + \sum_{q=1}^{Q} \gamma_{0qk} W_{qjk} + u_{0jk}\right) + \sum_{p=1}^{P} \left(\gamma_{p0k} + \sum_{q=1}^{Q} \gamma_{pqk} W_{qjk} + u_{pjk}\right) X_{pijk} + e_{ijk}$$

$$\begin{aligned} Y_{ijk} = & \left(\pi_{000} + \sum_{s=1}^{S} \pi_{00s} Z_{sk} + r_{00k}\right) + \sum_{q=1}^{Q} \left(\pi_{0q0} + \sum_{s=1}^{S} \pi_{0qs} Z_{sk} + r_{0qk}\right) W_{qjk} + u_{0jk} + \\ & + \sum_{p=1}^{P} \left[\left(\pi_{p00} + \sum_{s=1}^{S} \pi_{p0s} Z_{sk} + r_{p0k}\right) + \sum_{q=1}^{Q} \left(\pi_{pq0} + \sum_{s=1}^{S} \pi_{pqs} Z_{sk} + r_{pqk}\right) W_{qjk} + u_{pjk}\right] \times \\ & \times X_{pijk} + e_{ijk} \end{aligned}$$

by manipulating and simplifying the terms of that expression one obtains

$$\begin{aligned} Y_{ijk} = & \pi_{000} + \sum_{p=1}^{P} \pi_{p00} X_{pijk} + \sum_{q=1}^{Q} \pi_{0q0} W_{qjk} + \sum_{s=1}^{S} \pi_{00s} Z_{sk} + \sum_{p=1}^{P} \sum_{q=1}^{Q} \pi_{pq0} W_{qjk} X_{pijk} + \\ & + \sum_{p=1}^{P} \sum_{s=1}^{S} \pi_{p0s} Z_{sk} X_{pijk} + \sum_{q=1}^{Q} \sum_{s=1}^{S} \pi_{0qs} Z_{sk} W_{qjk} + \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{s=1}^{S} \pi_{pqs} Z_{sk} W_{qjk} X_{pijk} + e_{ijk} + \\ & + u_{0jk} + \sum_{p=1}^{P} u_{pjk} X_{pijk} + r_{00k} + \sum_{p=1}^{P} r_{p0k} X_{pijk} + \sum_{q=1}^{Q} r_{0qk} W_{qjk} + \sum_{p=1}^{P} \sum_{q=1}^{Q} r_{pqk} W_{qjk} X_{pijk} \end{aligned} \tag{5.8}$$

This model includes the explanatory variables of the three levels, represented by X_{pijk}, W_{qjk} and Z_{sk}, respectively, as well as it includes interaction terms between levels, $W_{qjk}X_{pijk}$, $Z_{sk}X_{pijk}$, $Z_{sk}W_{qjk}$ and $Z_{sk}W_{qjk}X_{pijk}$, representing the cross–level interactions. For instance, $W_{qjk}X_{pijk}$ is an interaction term that appears in the model as a consequence of modelling the varying regression slope β_{pjk} of student level variable X_{pijk} with the class level variable W_{qjk}. In this sense, the moderator effect of W on the relationship between the dependent variable Y and the explanatory variable X is expressed in equation (5.8) as a cross–level interaction (Hox, 2010).

In the empirical literature, equation (5.8) is often called mixed model as it includes both the fixed and the random effect parts of the model. The deterministic component of the model includes the constant and the slope coefficients of the different explanatory

variables of the model and it is given by:

$$\pi_{000} + \sum_{p=1}^{P} \pi_{p00} X_{pijk} + \sum_{q=1}^{Q} \pi_{0q0} W_{qjk} + \sum_{s=1}^{S} \pi_{00s} Z_{sk} + \sum_{p=1}^{P} \sum_{q=1}^{Q} \pi_{pq0} W_{qjk} X_{pijk} +$$
$$+ \sum_{p=1}^{P} \sum_{s=1}^{S} \pi_{p0s} Z_{sk} X_{pijk} + \sum_{q=1}^{Q} \sum_{s=1}^{S} \pi_{0qs} Z_{sk} W_{qjk} + + \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{s=1}^{S} \pi_{pqs} Z_{sk} W_{qjk} X_{pijk}$$

whereas the random or stochastic component is composed by the error terms of the school, class and individual levels, and it is given by:

$$e_{ijk} + u_{0jk} + \sum_{p=1}^{P} u_{pjk} X_{pijk} + r_{00k} + \sum_{p=1}^{P} r_{p0k} X_{pijk} + \sum_{q=1}^{Q} r_{0qk} W_{qjk} + \sum_{p=1}^{P} \sum_{q=1}^{Q} r_{pqk} W_{qjk} X_{pijk}$$

which is interpreted as a linear model with a complex form error.

As previously mentioned, multilevel models are adequate to grouped data, in which observations from the same group are generally more similar to each other than observations from different groups, and this violates the assumption of independence of all observations. This lack of independence can be expressed by means of a correlation coefficient: the intraclass correlation. The intraclass correlation indicates the proportion of the variance explained by the grouping structure in the population and can be interpreted as the expected correlation between two randomly chosen units that are in the same group (Hox, 2010). To estimate this correlation coefficient, the literature suggests the use of a model that contains no explanatory variables at all, the so–called intercept–only model. This model is for the purpose of estimating the proportion of the variance in the achievement outcomes at each level of the model – within classroom, between classes within schools, and between school levels (Raudenbush and Bryk, 2002; Hox, 2010).

Based on the previous model and keeping up all its assumptions, the equations defining intercept–only model are given by:

$$Y_{ijk} = \beta_{0jk} + e_{ijk} \tag{5.9}$$

$$\beta_{0jk} = \gamma_{00k} + u_{0jk} \tag{5.10}$$

$$\gamma_{00k} = \pi_{000} + r_{00k} \tag{5.11}$$

i.e., rewriting the model by a single equation it follows that:

$$Y_{ijk} = \pi_{000} + r_{00k} + u_{0jk} + e_{ijk} \tag{5.12}$$

As this model does not include explanatory variables of any level, it allows us to estimate the overall average of the dependent variable, π_{000}, thus, it does not explain any variance in Y but decomposes it into three independent terms: (1) the intra–group variability, i.e., the variance of the error to the student level (σ_e^2); (2) the variability between classes, i.e., the variance of the error to the class level within schools ($\sigma_{u_0}^2$); and (3) the variability between groups, i.e., the variance of the error to the school level (σ_r^2). So,

$$\begin{aligned} var(Y_{ijk}) &= var(r_{00k} + u_{0jk} + e_{ijk}) \\ &= \sigma_r^2 + \sigma_{u_0}^2 + \sigma_e^2 \end{aligned}$$

which represents the total variance. According to Siddiqui et al. (1996), the intraclass correlation coefficient for the students within classes is defined as:

$$\rho_{class} = \frac{\sigma_r^2 + \sigma_{u_0}^2}{\sigma_r^2 + \sigma_{u_0}^2 + \sigma_e^2} \tag{5.13}$$

and intraclass correlation coefficient for the students within schools is defined as:

$$\rho_{school} = \frac{\sigma_r^2}{\sigma_r^2 + \sigma_{u_0}^2 + \sigma_e^2} \tag{5.14}$$

Since the students within a classroom are more likely to be similar than the students within a school, ρ_{class} are always expected to be higher than or equal to ρ_{school}. They are equal if $\sigma_{u_0}^2 = 0$. ρ_{class} is the expected correlation between two students within the same class, and it takes in account that two students that are in the same class/teacher must also be in the same school ($\sigma_r^2 + \sigma_{u_0}^2$). ρ_{school} represents the proportion of variance of the dependent variable explained by the variability between schools (Hox, 2010). ρ varies between 0 and 1. If $\rho = 0$, it means that the groups are similar, i.e., the micro units are homogeneously distributed between the macro units, so, there is no multilevel structure. If ρ is not zero, it means that OLS estimation is inapplicable, since it leads to incorrect inferences, because the variability in Y is due to the difference between the macro units (Goldstein, 2011).

5.3.1.2 Data Envelopment Analysis – DEA

DEA was first introduced as a managerial and performance measurement tool in 1978. Since then, new applications with more variables and more complex models have been introduced, as well as its use has been extended to different fields. DEA, originally proposed by Charnes et al. (1978), is a non–parametric linear programming method

used to assess the relative efficiency of similar organizations, each known as Decision Making Unit (DMU), wherein each one uses multiple resources (inputs) to produce multiple results (outputs). Its idea is to compare the productive efficiency of DMUs, allowing for the identification of efficient DMUs, that would constitute an efficient frontier. That is, under the DEA, each DMU is separately analysed and a maximum performance measure is calculated for each unit. The identification of efficient DMUs arise from the comparison between similar DMUs, so the DEA score is a measure of relative efficiency. Furthermore, this methodology allows to measure the level of efficiency of the DMUs that are not in the efficiency frontier, and to identify points through which inefficient units can be compared (Cook and Seiford, 2009; Cook et al., 2014).

The main concept in the context of the DEA methodology is efficiency. According to the literature, efficiency is the "degree to which a production process reflects "best practices" either in an engineering sense ("technical efficiency") or in an economic sense ("allocative efficiency")" (OECD, 2001, p.89). From the combination of both technical efficiency and allocative efficiency it results the "total economic efficiency". If an organization uses its resources completely allocatively and technically efficiently, then it can be said to have achieved total economic efficiency.[6]

Technical or productive efficiency characterises a production process where the maximum possible output has been achieved, given a fixed set of outputs and given a certain technology. Within the context of education, technical efficiency may then refer to the physical relationship between the resources used and some education outcome. Full allocative efficiency is an extension of technical efficiency, since it involves not only the physical aspects, but also the monetary aspects. Namely, allocative efficiency refers to an organization's ability to use inputs in optimal proportions given a set of input prices, or to produce outputs in optimal proportions given a set of output prices (OECD, 2001; Worthington, 2001). This decomposition efficiency in technical efficiency and allocative efficiency was proposed by Farrel (1957), who developed the contributions of Pareto, Debreu and Koopmans.

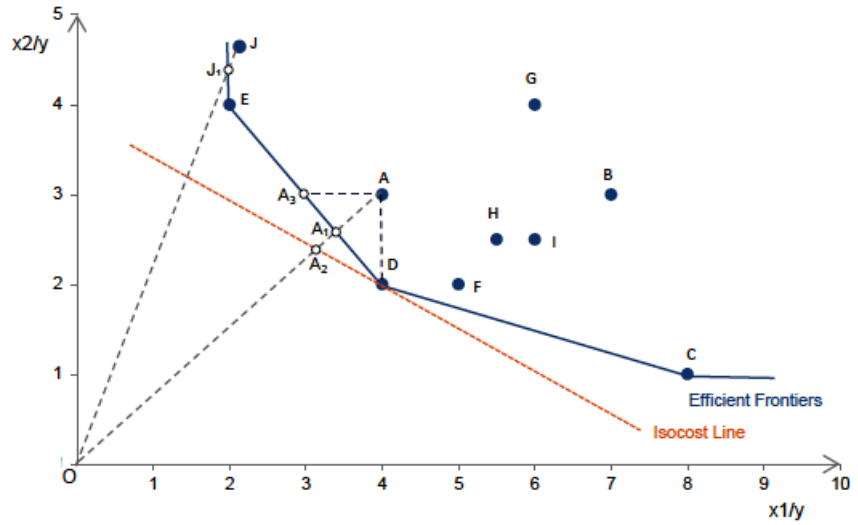
Charnes et al. (1978) generalize the study of Farrel (1957) and develop a performance indicator that is based on the concept of Pareto–Koopmans efficiency, namely: "a DMU (decision–making unit) is fully efficient if and only if it is not possible to improve any input or output without worsening some other input or output"(Cooper et al., 2007, p.153). In accordance with this concept, Charnes and Cooper (1985) provide the formal definition of efficiency saying that 100% efficiency is attained for a

[6]Farrel (1957) considers the total efficiency as the product between technical efficiency and allocation efficiency.

DMU only when: none of the outputs can be increased without, either increasing one or more of the inputs, or decreasing some of its other outputs; none of its outputs can be decreased without, either decreasing some of its inputs, or increasing some of its other outputs.

However, some adaptation of the definition is required, so that it refers to levels of relative efficiency to known levels attained elsewhere in similar circumstances. Charnes and Cooper (1985) provide such definition suggesting that 100% relative efficiency is attained by any unit only when comparisons with other relevant units do not provide the evidence of inefficiency in the use of any input or output.

Figure 5.2 illustrates the concept of efficiency in the DEA methodology using two inputs (x_1, x_2) and one output (y) for nine DMUs, from A to J.



Source: Adapted from Coelli et al. (2005, p.167) and Cooper et al. (2006, p.7)

Figure 5.2: Graphic illustration of the DEA projection

The technical efficiency frontier is defined by the line connecting the DMUs C, D and E. There is no point in this frontier line in which one of its input values can be increased without worsening the other. All data points can be enveloped within the region enclosed by the frontier line; that region is called "production possibility set". (Cooper et al., 2006). The efficiency of DMUs that are not on the frontier line can be measured by referring to the frontier point as follows. For example, DMU A is inefficient, then, the efficiency of A can be evaluated by:

$$\text{Technical efficiency of } A = \frac{\overline{OA_1}}{\overline{OA}}$$

$$\text{Allocative efficiency of } A = \frac{\overline{OA_2}}{\overline{OA_1}}$$

$$\text{Total economic efficiency of } A = \frac{\overline{OA_1}}{\overline{OA}} \times \frac{\overline{OA_2}}{\overline{OA_1}} = \frac{\overline{OA_2}}{\overline{OA}}$$

A given DMU is fully efficient when the DEA score equals to 1 and all slacks are equal to zero.[7] If only the first condition is satisfied, we talk about "radial", "technical" or "weak" efficiency of the the DMU. If both conditions are satisfied, the DMU is called efficient in terms of "Pareto–Koopmans", or, in other words, there is "strong" efficiency. Note that, the efficient frontier is constituted at point E with a vertical extension and at point C by a horizontal extension, which represents inefficiencies of the "non–radial" type, i.e., the inputs reduction means changes in the ratio of the inputs used.

For example, in Figure 5.2 the efficient point A_1 is Pareto–Koopmans efficient, whereas the point J_1 is Pareto–Koopmans inefficient and technically efficient. The technical efficiency of J is given by $J = \frac{\overline{OJ_1}}{\overline{OJ}}$. For the DMU J to be technically efficient there should be a decrease in its inputs to the levels observed in point J_1; nevertheless, for the DMU J to be Pareto–Koopmans efficient it should decrease their inputs to the levels observed in DMU E. In other words, point J_1 is inefficient, because it is still possible to reduce one of the inputs (in this case x_2) until the input level of DMU E without increasing the other input (x_1), and keeping the same output level.

The optimal amount of inputs that make the DMU A efficient (point A_1) can be obtained considering the linear combination of the inputs used by the DMUs E and D, given that the point A_1 is on the straight line that connects those two points. DMUs E and D are called "reference set" or "benchmarks" of A. Note that DMU D contributes more than DMU E for the determination of the optimal level of the inputs, because it is closer to A_1. Thus, the benchmarks of an inefficient DMU may differ from DMU to DMU and they provide useful information to the inefficient DMU on the best practices to produce the same level of output using less resources (Cooper et al., 2006).

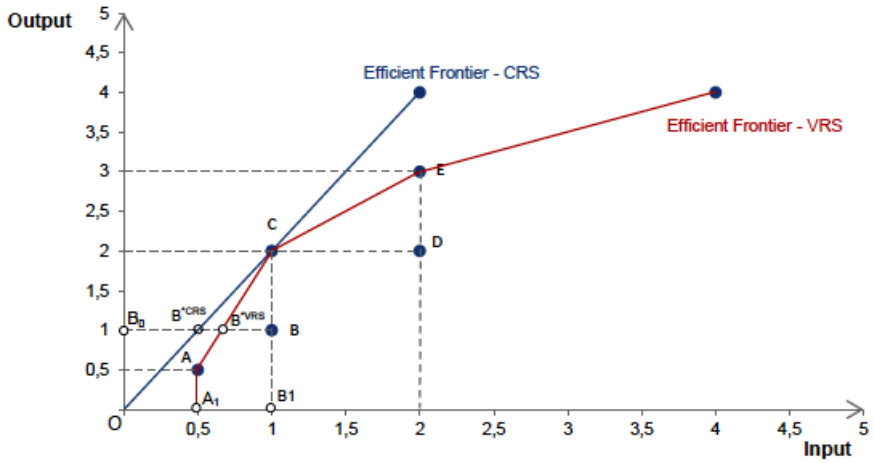
DEA allows multiple inputs–outputs to be considered at the same time, without any assumption on data distribution. In each case, efficiency is measured in terms of a proportional change in inputs or outputs.

A DEA model can be either an input–oriented model, which is used when the aim is to minimize inputs to achieve a given level of outputs; or, an output–oriented model, which is used when the aim is to maximize outputs for a fixed level of inputs (Cook et al., 2014).

[7] Figure 5.2 indicates that J_1 is a technical efficient point. In that point the DMU J can produce the same output using the same amount of input x_1, but the input x_2 can be further reduced until the value of x_2 observed in DMU E. This individual input reduction ($\overline{EJ_1}$) is called input slack. Mathematically, to obtain the DEA slacks, one needs to solve a two–stage linear programming problem. If the slack of the DMU J is neglected and residually calculated it becomes a single stage DEA model (Cooper et al., 2006; Sherman and Zhu, 2006).

Another important issue regarding the application of this methodology is the selection of the type of returns to scale in the production in the examined sector, which gives reason to different types of DEA models. The CCR model, originally presented by Charnes, Cooper, and Rhodes (1978), also known as the CRS – Constant Returns to Scale model, proposes the efficiency measurement of the DMUs by considering constant returns to scale, that is, any change in the input produces a proportional change in the output. Thus, all DMUs are operating at their optimal scale. The BCC model, presented later by Banker, Charnes, and Cooper (1984), proposes the Variable Returns to Scale (VRS) efficiency measurement, i.e., it does not assume proportionality between inputs and outputs (this assumption requires that the efficiency frontier is convex). As such, this model allows for the breakdown of efficiency into technical and scale efficiencies in DEA. Furthermore, the BCC model allows that DMUs operating with low input amounts have increasing returns to scale and DMUs operating with high input amounts have decreasing returns to scale (Cook and Seiford, 2009; Ji and Lee, 2010).

Figure 5.3 shows five DMUs (A–E), by considering only an input and an output, as well, efficiency frontiers with different returns to scale.



Source: Adapted from Ji and Lee (2010, p.269)

Figure 5.3: Concepts of input and output–oriented model and returns to scale

In Figure 5.3, the CRS frontier is represented by the straight line passing through the point (0, 0), coinciding with O, and the DMU C, and the VRS frontier consists in the segments connecting the points A, C and E. If the CRS is assumed, only DMU C would be efficient; DMUs A, C, and E are efficient when the VRS is assumed. The portion of the frontier from point A up to point C (not including the point) constitutes the increasing returns to scale portion of the frontier. So, DMU A exhibits increasing returns to scale, i.e., a change in the input level causes a more than proportional change in the output, and DMU C is experiencing constant returns to scale. All points on the

frontier to the right of C, for example E, are part of the decreasing returns to scale portion of the frontier. That is, for the DMU E a change in the input amount level causes a less than proportional change in the output amount. The remaining DMUs, which are inefficient, can be classified as increasing returns to scale, if the sum of the reference weights is less than unity, and as decreasing returns to scale otherwise (Cook and Seiford, 2009; Ji and Lee, 2010).

The CCR (CRS) model efficiency score is an overall productivity measure, known as "technical efficiency", "global technical efficiency", or "productive efficiency". By considering the orientation of the model, in the input–oriented CRS model, the efficiency of DMU B is defined by:

$$E_{B,input,CRS} = \frac{\overline{B_0 B^{*CRS}}}{\overline{B_0 B}}$$

and implies that the same output can be obtained by reducing the input by the ratio of $1 - E_{B,input,CRS}$. The efficiency in the output–oriented CRS model is defined as:

$$E_{B,output,CRS} = \frac{\overline{B_1 B}}{\overline{B_1 C}}$$

and traduces the fact that same input can be obtained by increasing the output by the ratio $1 - E_{B,output,CRS}$.

The measure of efficiency relative to the VRS frontier is designated by "pure technical efficiency" and in the input–oriented model is defined as:

$$E_{B,input,VRS} = \frac{\overline{B_0 B^{*VRS}}}{\overline{B_0 B}}$$

According to the literature, it is possible to decompose the CRS technical efficiency in "scale efficiency" and "pure technical efficiency". The "Scale Efficiency" (SE) is a measure of the impact of the DMU operating scale, reflecting the ability to obtain the maximum productivity, and it is defined by the ratio between technical efficiency and pure technical efficiency. Considering again Figure 5.3 one has:

$$SE = \frac{\overline{B_0 B^{*CRS}}}{\overline{B_0 B^{*VRS}}} \quad \Leftrightarrow \quad E_{B_{CRS}} = SE \times E_{B_{VRS}}$$

In this case, $\overline{B^{*VRS} B}$ contributes to the pure technical efficiency (VRS) of the B and $\overline{B^{*CRS} B}$ contributes to its technical efficiency (CRS). Note that, the intersection between the CRS and VRS boundaries identifies the DMU that shows constant returns to scale, i.e., those that have the dimension that maximizes productivity, which, in this case, is the DMU C. All the efficiency measures of DMU C are the same regardless of

the orientation of the model.

5.3.2 Empirical specification

The first goal of this analysis is to shed some light on the school performance assessment. Such assessment is based on the students performance, and as such, a better understanding of the factors that influence student performance are of major importance. Among those factors, special attention would be given to the role of school specific effect, which captures common factors across the school that might influence student progress. Based on those results, a ranking of schools will be built. The analyses are to be performed using either the DEA or the multilevel modelling techniques, in order to analyse to what extent methodological choices may justify differences in the position in which each school ranks.

As explained above, economists typically approach the challenge of measuring school effectiveness in the context of a standard education production function. This education production is going to be estimated for each year, and for both Mathematics and Portuguese subjects.[8] In order to assess and disentangle the effects of the student background and school resources on student educational performance, the education production function might take the following form:

$$\log A_{ijk} = \lambda \log {A_i}^9 + \beta X_{ijk} + \delta C_{jk} + \theta S_k + \varepsilon_{ijk} \tag{5.15}$$

where A_{ijk} is the student's achievement gain in Portuguese or Mathematics, measured by the 12^{th} grade national exam score, for student i in class j in school k; ${A_i}^9$ is the student's achievement in the 9^{th} grade exam in the same subject (Portuguese and Mathematics). The inclusion of prior achievement means that the focus is on the impact of the school on student's progress, that is, a added value perspective is taken (Ladd and Walsh, 2002). X_{ijk} is a vector that captures the student and family characteristics that affect his/her achievement; C_{jk} is a vector of measurable class characteristics; S_k is a vector of measurable school characteristics and ε_{ijk} is an error term representing unobserved factors influencing school quality.

The choice of the variables that may have an influence on student performance, and consequently, on school performance, was based in the previous literature. At the individual level, several student characteristics such as age, gender, degree of kinship of the legal–guardian and nationality. The socio–economic status of the students is captured by four variables: whether the student benefits from social support, whether

[8] In Portugal there are two Mathematics exams, A and B. Mathematics A includes more advanced topics, whereas the Mathematics B exam usually is taken by students in non–academic tracks. For the purpose of the present analysis, only Mathematics A has been considered.

there is internet at home, whether the parents'/legal–guardians hold the secondary education diploma or less, and their employment status. Furthermore, the student prior achievement is included in order to capture the role of students own ability and motivation, since such characteristics of the students have constant effects on achievement over time (Ladd and Walsh, 2002). Students are distributed across 18 districts/regions.

Class–level variables include class size and some variables that characterize the classroom peer–group, such as the share of female students, the share of students with social support, the share of immigrants and the share of parents/legal–guardians holding no more than the secondary education.

At the school level, the model includes a resource measure: teaching expenditures per student (where teaching expenditures refer to the gross wages of teachers). It also includes the size of the school, measured by the number of students who attend it. The share of female students, the share of students with social support, the share of immigrants, and the share of parents/legal–guardians with at most the high school diploma, at the school–level, are included as well. Furthermore, for a better understanding of the class size effect and how it may differ across different groups of students (namely, males versus females, and low income versus high income students), two interaction terms are included: *class size*×*beneficiary of social support* and *class size*×*gender*.

Under the assumption that the explanatory variables are not correlated with the error term, equation (5.15) could be estimated by OLS. Nevertheless, some inputs in the education production function might not be exogenous, to their educational performance, and as a result, the OLS estimation yields consistent but inefficient estimates of the coefficients. For instance, one could argue that student ability and motivation are omitted variables, but in the context of the present model, those aspects are controlled for through the inclusion of prior achievement. Another point that can be made is that, at least for some students, the peer group is related to the family decision about which school the child attends, implying that school inputs are correlated with the error term of the production function (Woessman, 2003).

In order to avoid those endogeneity problems, school specific effects are included to control for the fact that families whose children attend a given school have made similar choices regarding their children's education and potential peer group. Families may be similar in other ways, including regarding the unobserved characteristics that influence their children's achievement (McEwan, 2003). The inclusion of school specific effects implicitly controls for student heterogeneity that is constant across a given school and also controls for other unobservable school characteristics that vary by school, such as the effectiveness of the school administration, resources and facilities, disciplinary

policy and selection policy, which may all influence student outcomes. Consequently, in the present analysis, specifications that replace controls for school characteristics with school specific effects are also estimated. The school value–added is estimated by the following education production function:

$$\log A_{ijk} = \lambda \log A_i^{9} + \beta X_{ijk} + \delta C_{jk} + \mu_k + \varepsilon_{ijk} \tag{5.16}$$

where μ_k is a school specific effect that provides a measure of school value–added for school k.

As discussed in Subsection 5.3.1.1, the OLS method assumes the independence of the individuals, which may be questioned in a context of hierarchical data. Students of the same class/school tend to have more in common with each other compared to others who attend different classes/schools, i.e., the random errors of the students belonging to the same school will not be independent if the school has some impact on student performance. Thus, the assumption of independence is probably violated, because the units at the same level are correlated. By ignoring the hierarchical structure of the data can lead to problems such as aggregation bias, estimation error of standard errors and parameters, and the unit of analysis problem, which may lead to inference errors. Multilevel models avoid such problems (Raudenbush and Bryk, 2002; Johnes, 2006), and provide efficient maximum likelihood estimates of the coefficients (Raudenbush and Bryk, 2002), which makes them more appropriate. For the present analysis, a multilevel model with three levels is used: student (level 1), class (level 2) and school (level 3), as discussed in Subsection 5.3.1.1.

The multilevel model is a parametric non–frontier method, which allows observations to lie above or below the line of best fit. A view to deepen this analysis the DEA methodology is also used, since it is a non–parametric frontier method, as discussed in subsection 5.3.1.2. The comparison between these two methods is of particular interest, since it is likely that these alternative approaches can provide different efficiency measures. The advantage of performing the statistical analysis firstly is that the variables found to be significant determinants on student achievement can be used as inputs in the DEA, thereby overcoming the problem of specification usually encountered in DEA, since increasing the number of inputs can increase the proportion of efficient DMUs in the data set and leads to a higher overall average efficiency (Johnes, 2006).

To sum up, the empirical strategy will be divided in two parts. First, a three-level model on the determinants of student achievement is estimated. The model specification is also estimated by OLS, in order to facilitate comparisons, as well as justifying the option for the multilevel model. Second, the school rankings are built based on the results of both models: the multilevel model and the data envelopment

analysis.

5.3.3 Data and sample

Both methods, multilevel and DEA, are implemented on a dataset that was built based on two distinct databases managed by the Portuguese Ministry of Education: the MISI (*Sistema de Informação do Ministério da Educação*), managed by *Direção–Geral de Estatísticas da Educação e Ciência* (DGEEC), and the Statistics published by *Júri Nacional de Exames – Direção Geral de Educação* (JNE). MISI is a very detailed administrative database that contains data on pre–school education, as well as basic and secondary education, in public schools, overseen by the Ministry of Education and Science. The data collection relies on information obtained through the school management programs that exist in schools. This database contains information at student–level, such as gender, nationality, academic outcome, grade, social support eligibility, type of student, type of education, course, residence, availability of computer and internet at home, kinship of legal–guardians, legal–guardians'/parents' employment situation and legal–guardians'/parents' education, class and school. Information at school–level includes localization and school resources. The data provided by the *Júri Nacional de Exames* include information on the scores in national exams taken by students attending Portuguese schools. Those results refer to all the disciplines of basic and secondary education subjected to examination, however, only the disciplines of Portuguese and Mathematics A are considered in this study. This database also includes some student characteristics, such as gender, age, school. As students are identified by the same number in both datasets, it is possible to follow the academic trajectory of students over time.

The two datasets, MISI and the exam data from JNE, we had access for this research, have been previously anonymized to absolutely secure private information on individuals, classes and schools. All the information on individuals, namely students and professors, cannot be individually traced by the researcher.

In our empirical analysis only the 12^{th} grade students who took the Mathematics and/or Portuguese examination, in 2010–2012, were selected. Furthermore, only internal students and students who enrolled in the national exam but who attended the discipline throughout the school year are included, to ensure that the student attends a school. Students with special educational needs and schools located in the Islands were not included. Note that, students have the opportunity to perform the same national test more than once, either to successfully pass the exam or to improve their scores. For the latter only the best score that the student has obtained in each exam was considered.

A careful look at the data resulted in some data corrections, as well as, a selection of the available information. Due to inconsistencies in student ID number, some observations have been lost when merging the two data sources. The outliers regarding student age, that correspond to a small part of the original database, were eliminated. All observations that had missing data on the variables used in the present analysis were also excluded, namely, students who had no information about both Mathematics and Portuguese 9^{th} grade exam scores, which happened with just a small number of observations. In addition, due to the lack of information in the variables related to parents/legal–guardians, about 28% of the students were eliminated from the sample. Students attending schools where the number of students who took those exam was less than 20 were also eliminated, implying that about 8% of students were eliminated. The final sample includes information from the period 2010–2012, including almost 36,000 students, who attended the scientific–humanistic courses, clustered in more than 5,800 classes, which in turn are clustered into about 351 schools.

The variables selected for the following analysis combine the available information with the type of information used in existing empirical literature. In addition, the explanatory variables were grouped in each of the different levels: student, class and school. Some descriptive statistics on the variables used in this essay are reported in Tables 5.1 and 5.2.[9]

The average age of the students is about 18 years old.[10] About 24.5% of the students have social support, and about 26.3% have no Internet access at home. Mothers are in general the responsible for the student school life (about 70.8% of the legal guardians), and only in about 20.6% of the cases the father is the responsible for the student education. In addition, it is observed that schooling of about 53% of the person responsible for the school education is at most the 3^{rd} cycle and only about 24% of the guardians have a higher education diploma.

Figure 5.4 shows the distribution of the classifications obtained in the national exams for the different subjects under study.[11] The exams' results do not suggest field differences (the results in Portuguese were about 11.6 out of 20, on average, whereas 11.5 is the average performance in Mathematics). The share of students who score less than 10 (out of 20) is about 34% in Mathematics A and about 25% in Portuguese. Note that, one can consider these percentages as very relevant, since the students in the sample are internal students who are admitted to the national exam only if the

[9]See also Table C.1 in Appendix C.1 for a description of the variables

[10]The values for this variable go from 17 to 20 but a large proportion of the students are 18 years old, which is the normal age for carrying out the national secondary school examinations.

[11]Note that, for the elaboration of the graphs, the scores of the exams were converted to a scale of 0–20 values; the original values vary between 0 and 200 points.

Table 5.1: Descriptive statistics on students

Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
	2010		2011		2012	
Mathematics	125.34	46.164	108.33	47.256	106.833	44.743
Portuguese	116.95	29.499	105.93	31.347	112.418	30.261
Maths 9^{th}	95.00	43.401	124.28	47.051	133.712	38.589
Portuguese 9^{th}	138.37	24.046	138.60	25.401	129.029	26.927
Female	0.596		0.588		0.579	
Age	18.11	0.401	18.280	0.551	18.27	0.528
Portuguese student	0.991		0.995		0.995	
Internet	0.587		0.806		0.832	
Beneficiary s.s.	0.267		0.249		0.221	
Parent/legal–guardian						
Father	0.219		0.203		0.195	
Mother	0.712		0.700		0.709	
Own	0.045		0.076		0.073	
Other	0.025		0.021		0.023	
Parent/legal–guardian education						
Tertiary	0.249		0.227		0.241	
Secondary	0.221		0.237		0.244	
3^{rd} cycle	0.209		0.239		0.235	
2^{nd} cycle	0.174		0.165		0.157	
1^{st} cycle or less	0.147		0.133		0.122	
Parent/guardian employment status						
Worker for others	0.640		0.636		0.633	
Self–employed	0.114		0.101		0.104	
Unemployed	0.049		0.053		0.061	
Student	0.048		0.079		0.075	
Domestic/retired	0.140		0.125		0.119	
Other	0.009		0.006		0.008	

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2010–2012.
Note: The number of observations in all variables, except Mathematics and Portuguese variables, is 12,782, 9,984 and 13,131 in 2010, 2011 and 2012, respectively. The corresponding values for Mathematics variable are 7,800, 6,109 and 8,454, and for Portuguese variable are 12,773, 9,550 and 12,583, respectively.

internal grade is equal or above 10.

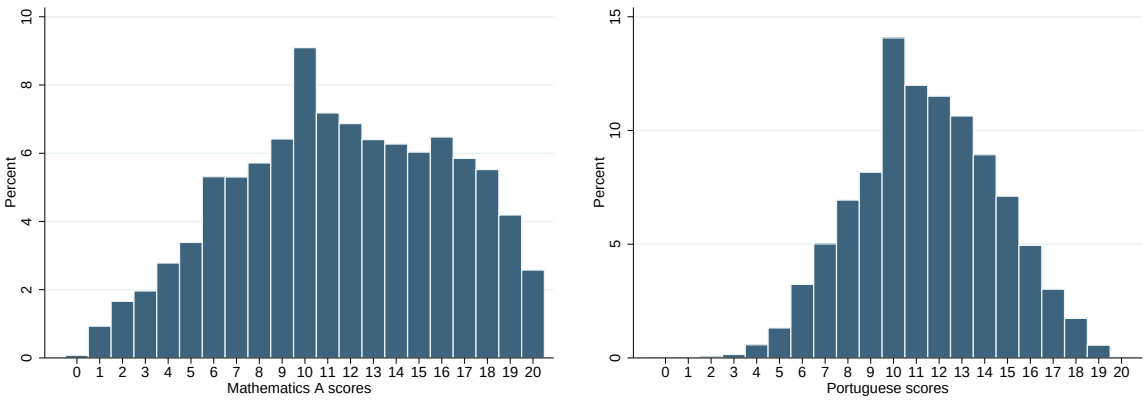
In addition, boys and girls have different performances, which are more evident in the Portuguese exam results. Figure 5.5 shows that the medium ranking in both disciplines is 11 and 12 for boys and girls, respectively, and the distribution of classifications is slightly different for boys and girls. Furthermore, the average performance of girls is slightly better than the boys in Portuguese (11.8 for girls versus 10.9 for bys), however, this difference is negligible in Mathematics (11.7 for girls versus 11.5 for boys).

As mentioned above the multilevel model approach to be applied in the next section is explainable in the context of the hierarchical structure of the available data. As a

Table 5.2: Descriptive statistics on classes and schools

Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
	2010		2011		2012	
Class level variables						
Class size	23.80	4.961	25.11	5.194	25.57	5.453
% economically disadvantaged	24.46	17.31	23.96	16.74	21.67	15.45
% of girls	57.05	17.91	56.55	17.55	55.98	17.74
% of immigrants	3.608	5.876	3.667	5.666	3.335	5.166
% less educated parent/guardian	83.06	14.88	82.94	14.50	82.49	14.25
School level variables						
School size	969.7	315.9	1,056	325.3	1,030	342.0
% economically disadvantaged	31.87	14.67	29.67	14.24	28.90	14.03
% of girls	51.51	3.755	50.59	3.878	50.05	3.655
% of immigrants	4.496	4.844	4.890	5.413	4.590	4.941
% less educated parent/guardian	86.40	9.389	86.53	9.869	86.37	9.460
Expenditure–student ratio	567.1	229.1	518.1	204.5	477.6	165.2

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2010–2012.
Note: The number of classes observed is 1,817, 1,928 and 2,115 in 2010, 2011 and 2012, respectively; the number of schools is 284, 283 and 322 in 2010, 2011 and 2012, respectively.

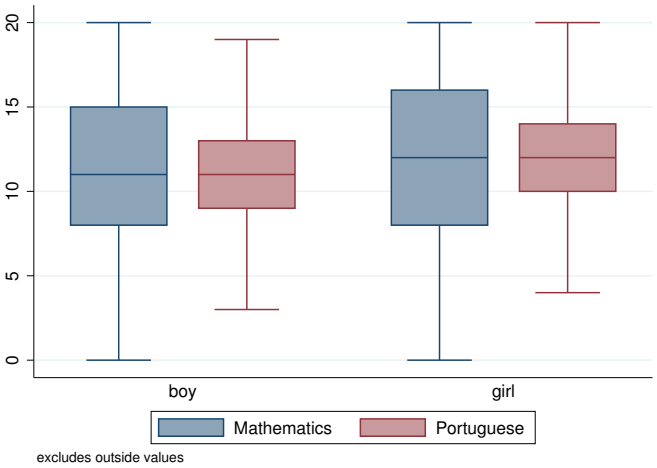


Source: Created by the author based on *MISI* and *JNE Statistics*, 2010–2012

Figure 5.4: Proportion of students by grade on national exams of Mathematics A and Portuguese

first step, considering all students and schools in each year and in each subject, based on the null model, the coefficients of partition of the estimated variance, are presented in Table 5.3.

The values in Table 5.3 confirm the adequacy of the multilevel regression models to the data (Goldstein, 2011). The intraclass correlation results suggest that 15.3%



Source: Created by the author based on *MISI* and *JNE Statistics*, 2010–2012

Figure 5.5: Mean national exams scores by discipline and by gender

Table 5.3: Intraclass correlation

	2010		2011		2012	
Level	Maths	Portuguese	Maths	Portuguese	Maths	Portuguese
school	0.0668	0.0452	0.0652	0.0521	0.0405	0.0423
class/school	0.1557	0.1362	0.1636	0.1531	0.1531	0.1365

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2010–2012.

to 16.4% of the variance in student performance in Mathematics A are due to the variability between classes and the corresponding values for the discipline of Portuguese vary between 13.6% and 15.3%.

5.4 Empirical results

This section is divided into two parts which are related to the research objectives proposed for this essay. Thus, in a first part, the factors influencing the student achievement gains are analysed, using the multilevel modelling approach. In the second part, the school–effect is analysed and for this purpose a ranking of Portuguese schools based on school added value is created, using the multilevel regression results and DEA efficiency scores.[12]

[12]Keep in mind that the anonymized data does not contain information on the identity of the school.

5.4.1 Factors that influence student achievement gains: a multilevel approach

As explained in Section 5.3.2, the specification (5.15) was first estimated by OLS and by means of random effects, for each subject, Mathematics and Portuguese, and for each year between 2010 and 2012. Tables C.2.1–C.2.3 in Appendix C.2 show the estimation results for both subjects. The OLS results will be a basis of comparison for the multilevel results.[13]

Just briefly, the OLS results show that prior achievement is strongly associated with Mathematics and Portuguese achievement. Both family and student characteristics are strongly associated with achievement gains in both Mathematics and Portuguese. Female students perform better than male students in both Mathematics and Portuguese, as do younger student. Low income students performance worse than high income students, mostly in Mathematics. Parent/legal–guardian education has an important influence on achievement.

Those results are a mere starting point, as the factors that influence student achievement gains are to be analysed in the context of a three-level model. First, a multilevel model defined by equations (5.9)–(5.11) with no explanatory variables (null model), which serves as a benchmark for the other model specifications, is estimated. Then, the variables for the student and family characteristics, and class–level variables are considered. Finally, the variables at school–level, as well the cross–level interaction terms are included. The inclusion of the explanatory variables allows for the interpretation of the estimated residuals r_{00k} from this model as indicators of each school effectiveness in terms of added value, that is, by taking into account the inter–school variations in the input variables (Raudenbush and Bryk, 2002). These specifications are estimated for each of the two subjects, Mathematics and Portuguese, and for each year between 2010–2012.

Note that, the OLS and random–effects results had showed that some of the explanatory variables had no effect on student achievement. Those variables were included one by one in the multilevel model to check whether there was any change in the value of the *deviance* is a model adjustment measure;[14] and in the absence of any change, the variable would be excluded from the model. In addition, the variable

[13] In each subject set of results, the first column reports the estimates of the specification that only includes control variables for the student characteristics and for the family background, including their home location (district, included by means of a set of dummy variables). Additionally, some specifications includes school specific effects. The magnitudes of the coefficient differences between specifications, with and without school specific effects, are not large and, in general, the significance level remains the same.

[14] *Deviance* $= -2$ *log likelihood* the lower its value, the better the model (Ferrão, 2003).

District is not included in this multilevel model, because it is a macro variable, since students are nested within schools, which in turn are nested nested within districts.

The results for the three level model are reported in Tables 5.4.1–5.4.3.

The unconditional model results (columns 1 and 5 in Tables 5.4.1–5.4.3) for the period 2010–2012 show that in the national secondary school examinations between 22.7% and 27.1% of the variance in Mathematics achievement and between 6.7% and 8.61% of the variance in Portuguese achievement is between students within classes, whereas only 1.8%–3.6% and 0.7%–1% of the variance in Mathematics and Portuguese achievement, respectively, is between classes within schools. Only 0.6%–1.8% and 0.3%-0.5% of the variance in Mathematics and Portuguese achievement, respectively, is due to differences between schools. Even controlling for student's and family characteristics, classes and school characteristics, the results show that there is more variability in student achievement gains between students within classes than between classes within schools or between schools; this variability is much higher in Mathematics.

By considering the variation of achievement gain parameters across classes and schools, it seems that family and student characteristics account for a small variation in achievement between classes within schools, as the variance does not decrease substantially. This result suggests that classes are formed not taking into account these student's and family characteristics. Thus, the inclusion of those explanatory variables at student level reduces both the class–effect and the school–effect in the achievement gains in Portuguese and Mathematics (compared with the null model), but not significantly. So, in the specification of column (2) of Tables 5.4.1–5.4.3 5.6%–7.4 and 2.9%–3.9% of the variance in student achievement gains in Mathematics is due to the variability between classes and schools, respectively, depending on the year under study. And in column (6) about 6% and 3.7%–5.6% of the variance in student achievement gains in Portuguese is due to the variability between classes and schools, respectively. In addition, the inclusion of the student's and family variables explains 10%, 23% and 27% of the variance model in Mathematics in 2010, 2011 and 2012, respectively. And, the same variables explain 29%, 31% and 33% of the variance model in Portuguese in 2010, 2011 and 2012, respectively.[15]

The results suggest that students' prior achievement scores are the strongest predictor of their current academic performance, meaning that better prepared students in Mathematics and Portuguese in the 3^{rd} cycle are more able to learn new concepts in the context of those disciplines in secondary education. In both disciplinary fields, gender has a clear influence on the student's achievement gains, with girls outperforming boys; the size of the effect is the highest in Mathematics in 2010 and 2012.

[15]The explained variance is calculated by: $R^2 = \frac{(\widehat{\sigma}^2_{r_0}+\widehat{\sigma}^2_{u_0}+\widehat{\sigma}^2_{e_0})-(\widehat{\sigma}^2_{r_1}+\widehat{\sigma}^2_{u_1}+\widehat{\sigma}^2_{e_1})}{\widehat{\sigma}^2_{r_0}+\widehat{\sigma}^2_{u_0}+\widehat{\sigma}^2_{e_0}}$.

Table 5.4.1: Multilevel model results (Year of examination: 2010)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable: Log of Mathematics scores				**Dependent variable: Log of Portuguese scores**			
Fixed–effects Parameters								
Log Maths 9^{th}		0.5877***	0.5834***	0.5836***				
		(0.0127)	(0.0127)	(0.0127)				
Log Portuguese 9^{th}						0.6420***	0.6414***	0.6414***
						(0.0117)	(0.0117)	(0.0117)
Female		0.0951***	0.0962***	0.0964***		0.0627***	0.0630***	0.0631***
		(0.0099)	(0.0099)	(0.0099)		(0.0042)	(0.0043)	(0.0043)
Age		-0.1987***	-0.1993***	-0.1990***		-0.0972***	-0.0974***	-0.0973***
		(0.0169)	(0.0169)	(0.0169)		(0.0054)	(0.0054)	(0.0054)
Beneficiary S.S.		-0.0238	-0.0177	-0.0175		-0.0082	-0.0092	-0.0092
		(0.0128)	(0.0131)	(0.0131)		(0.0051)	(0.0052)	(0.0052)
Internet		0.0259*	0.0220	0.0226		0.0161**	0.0157**	0.0158**
		(0.0120)	(0.0120)	(0.0120)		(0.0051)	(0.0051)	(0.0051)
Portuguese student		-0.0167	-0.0223	-0.0224		0.0464*	0.0460*	0.0460*
		(0.0517)	(0.0520)	(0.0520)		(0.0214)	(0.0215)	(0.0215)
Class size			0.0266*	0.0270*			0.0058	0.0058
			(0.0112)	(0.0112)			(0.0043)	(0.0043)
Class size sq			-0.0006**	-0.0006**			-0.0001	-0.0001
			(0.0002)	(0.0002)			(0.0001)	(0.0001)
% of economically disadvantaged (class)			-0.0008	-0.0009			0.0003	0.0003
			(0.0005)	(0.0005)			(0.0002)	(0.0002)
% of female (class)			-0.0019	-0.0020			-0.0001	-0.0001
			(0.0015)	(0.0015)			(0.0006)	(0.0006)
% less educated parents/guardians (class)			-0.0010*	-0.0013*			-0.0007**	-0.0007**
			(0.0005)	(0.0006)			(0.0002)	(0.0003)
% less educated parents/guardians (school)				0.0010				0.0003
				(0.0010)				(0.0005)

Continued on next page

Table 5.4.1 – continued from previous page

Random–effects Parameters								
Level–three variance:								
$\sigma^2_{r_0}$ Intercept variance	0.0179***	0.0073***	0.0071***	0.0072***	0.0034***	0.0031***	0.0030***	0.0030***
	(0.0031)	(0.0016)	(0.0016)	(0.0016)	(0.0006)	(0.0005)	(0.0005)	(0.0005)
Level–two variance:								
$\sigma^2_{u_0}$ Intercept variance	0.0238***	0.0121***	0.0117***	0.0116***	0.0070***	0.0035***	0.0035***	0.0035***
	(0.0029)	(0.0018)	(0.0018)	(0.0018)	(0.0006)	(0.0004)	(0.0004)	(0.0004)
Level–one variance:								
σ^2_e Residual variance	0.2265***	0.1671***	0.1670***	0.1670***	0.0667***	0.0484***	0.0483***	0.0483***
	(0.0039)	(0.0029)	(0.0029)	(0.0029)	(0.0009)	(0.0006)	(0.0006)	(0.0006)
Deviance	11328.27	8760.39	8737.78	8736.86	2816.39	-1481.09	-1493.72	-1494.09
Observations	7800	7800	7800	7800	12773	12773	12773	12773
LR test (χ^2)	517.74	253.42	240.56	241.25	724.83	608.54	570.67	570.71
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2010–2012.
Note: Standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. All regressions include a set of dummies for the parents'/legal–guardians' education, except in the null model.

Table 5.4.2: Multilevel model results (Year of examination: 2011)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable: Log of Mathematics scores				**Dependent variable: Log of Portuguese scores**			
Fixed–effects Parameters								
Log Maths 9^{th}		0.6807***	0.6736***	0.6737***				
		(0.0199)	(0.0199)	(0.0200)				
Log Portuguese 9^{th}						0.7032***	0.6987***	0.6987***
						(0.0151)	(0.0151)	(0.0151)
Female		0.0643***	0.0655***	0.0656***		0.0755***	0.0766***	0.0766***
		(0.0120)	(0.0120)	(0.0120)		(0.0055)	(0.0056)	(0.0056)
Age		0.0147	0.0133	0.0134		-0.0770***	-0.0761***	-0.0760***
		(0.0160)	(0.0160)	(0.0160)		(0.0057)	(0.0057)	(0.0057)
Beneficiary S.S.		-0.0405*	-0.0354*	-0.0353*		-0.0022	-0.0038	-0.0038
		(0.0161)	(0.0166)	(0.0166)		(0.0068)	(0.0070)	(0.0070)
Internet		0.0219	0.0161	0.0161		0.0003	-0.0005	-0.0004
		(0.0168)	(0.0167)	(0.0167)		(0.0075)	(0.0075)	(0.0075)
Portuguese student		0.0995	0.0915	0.0915		-0.0287	-0.0386	-0.0386
		(0.0805)	(0.0806)	(0.0806)		(0.0365)	(0.0366)	(0.0366)
Class size			0.0190	0.0191			0.0031	0.0031
			(0.0122)	(0.0122)			(0.0049)	(0.0049)
Class size sq			-0.0004	-0.0004			-0.0001	-0.0001
			(0.0002)	(0.0002)			(0.0001)	(0.0001)
% of economically disadvantaged (class)			-0.0002	-0.0002			0.0004	0.0003
			(0.0006)	(0.0006)			(0.0003)	(0.0003)
% of female (class)			-0.0012	-0.0012			-0.0021**	-0.0021**
			(0.0016)	(0.0016)			(0.0007)	(0.0007)
% less educated parents/guardians (class)			-0.0026***	-0.0026***			-0.0008*	-0.0008*
			(0.0006)	(0.0007)			(0.0003)	(0.0003)
% less educated parents/guardians (school)				0.0001				0.0002
				(0.0011)				(0.0006)

Continued on next page

Table 5.4.2 – continued from previous page

Random–effects Parameters								
Level–three variance:								
$\sigma^2_{r_0}$ Intercept variance	0.0060***	0.0073***	0.0049***	0.0049**	0.0053***	0.0036***	0.0033***	0.0033***
	(0.0019)	(0.0016)	(0.0018)	(0.0018)	(0.0009)	(0.0006)	(0.0006)	(0.0006)
Level–two variance:								
$\sigma^2_{u_0}$ Intercept variance	0.0178***	0.0121***	0.0168***	0.0168***	0.0102***	0.0044***	0.0042***	0.0042***
	(0.0029)	(0.0018)	(0.0029)	(0.0029)	(0.0010)	(0.0006)	(0.0006)	(0.0006)
Level–one variance:								
σ^2_e Residual variance	0.2567***	0.1974***	0.1975***	0.1975***	0.0861***	0.0623***	0.0623***	0.0623***
	(0.0051)	(0.0040)	(0.0040)	(0.0040)	(0.0014)	(0.0010)	(0.0010)	(0.0010)
Deviance	9749.54	7956.76	7927.38	7927.36	4741.10	1362.78	1345.38	1345.25
Observations	6109	6109	6109	6109	9550	9550	9550	9550
LR test (χ^2)	346.21	142.53	119.34	119.35	505.31	336.91	297.32	296.75
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2010–2012.
Note: Standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. All regressions include a set of dummies for the parents'/legal–guardians' education, except in the null model.

Table 5.4.3: Multilevel model results (Year of examination: 2012)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable: Log of Mathematics scores				**Dependent variable: Log of Portuguese scores**			
Fixed–effects Parameters								
Log Maths 9^{th}	1.0110***	1.0000***	1.0006***					
	(0.0246)	(0.0246)	(0.0246)					
Log Portuguese 9^{th}					0.5869***	0.5873***	0.5874***	
					(0.0098)	(0.0099)	(0.0099)	
Female	0.0742***	0.0773***	0.0785***		0.0488***	0.0484***	0.0483***	
	(0.0106)	(0.0106)	(0.0106)		(0.0043)	(0.0044)	(0.0044)	
Age	-0.1037***	-0.1039***	-0.1026***		-0.1025***	-0.1018***	-0.1019***	
	(0.0128)	(0.0128)	(0.0128)		(0.0045)	(0.0045)	(0.0045)	
Beneficiary S.S.	-0.0571***	-0.0533***	-0.0534***		-0.0062	-0.0080	-0.0080	
	(0.0142)	(0.0144)	(0.0144)		(0.0054)	(0.0055)	(0.0055)	
Internet	0.0217	0.0166	0.0190		-0.0009	-0.0003	-0.0004	
	(0.0158)	(0.0158)	(0.0158)		(0.0063)	(0.0063)	(0.0064)	
Portuguese student	-0.0120	-0.0161	-0.0152		-0.0418	-0.0482	-0.0483	
	(0.0726)	(0.0727)	(0.0726)		(0.0283)	(0.0284)	(0.0284)	
Class size		0.0281**	0.0303**			0.0123***	0.0122***	
		(0.0106)	(0.0106)			(0.0034)	(0.0034)	
Class size sq		-0.0006**	-0.0006**			-0.0002***	-0.0002***	
		(0.0002)	(0.0002)			(0.0001)	(0.0001)	
% of economically disadvantaged (class)		-0.0004	-0.0007			0.0003	0.0003	
		(0.0006)	(0.0006)			(0.0002)	(0.0002)	
% of female (class)		-0.0009	-0.0008			-0.0015*	-0.0015*	
		(0.0017)	(0.0017)			(0.0006)	(0.0006)	
% less educated parents/guardians (class)		-0.0029***	-0.0040***			0.0004	0.0004	
		(0.0006)	(0.0007)			(0.0002)	(0.0003)	
% less educated parents/guardians (school)			0.0031**				-0.0002	
			(0.0011)				(0.0005)	

Continued on next page

Table 5.4.3 – continued from previous page

Random–effects Parameters								
Level–three variance:								
$\sigma^2_{r_0}$ Intercept variance	0.0130***	0.0067***	0.0075***	0.0073***	0.0036***	0.0021***	0.0020***	0.0020***
	(0.0028)	(0.0017)	(0.0018)	(0.0017)	(0.0006)	(0.0004)	(0.0003)	(0.0003)
Level–two variance:								
$\sigma^2_{u_0}$ Intercept variance	0.0361***	0.0174***	0.0148***	0.0147***	0.0079***	0.0034***	0.0033***	0.0033***
	(0.0040)	(0.0025)	(0.0023)	(0.0023)	(0.0007)	(0.0004)	(0.0004)	(0.0004)
Level–one variance:								
σ^2_e Residual variance	0.2714***	0.2101***	0.2101***	0.2101***	0.0727***	0.0506***	0.0507***	0.0507***
	(0.0046)	(0.0035)	(0.0035)	(0.0035)	(0.0010)	(0.0007)	(0.0007)	(0.0007)
Deviance	13877.12	11463.36	11414.94	11406.04	3910.46	-935.15	-959.05	-959.18
Observations	8454	8454	8454	8454	12583	12583	12583	12583
LR test (χ^2)	430.80	217.47	196.72	192.84	598.56	384.73	351.22	349.85
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2010–2012.
Note: Standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. All regressions include a set of dummies for the parents'/legal–guardians' education, except in the null model.

During the period under study, the achievement gains of girls are higher than for boys in approximately 6.6%–9.6% in Mathematics and in about 4.8%–7.7% in Portuguese, depending on the year, and on the specification considered. In addition, the age of the students negatively impacts the performance; the effect is higher in Mathematics than in Portuguese (the effect for Mathematics is not statistically significant in 2011). An one year increase in age leads to a decline in student achievement of about 10% for both disciplines in 2012, suggesting a considerable effect on student achievement. This result presumably reflects a grade repetition effect; students that are older that the usual age of a given grade usually had some retention over his/her school life. Although the negative effect of age is more or less consensual in the empirical literature (for instance, Dias and Ferrão, 2006; Denny and Oppedisano, 2013), the same is not true for the gender effect. Some authors suggest that male students perform better in Mathematics and female students have a better performance in reading (Pereira, 2010; Pereira and Reis, 2012), whereas, other authors suggest that girls outperform boys in Mathematics (Denny and Oppedisano, 2013). The student's nationality seems to have no influence on his/her achievement.

A close look at the socio–economic status variables shows that family background has a positive and significant effect on student achievement. Namely, the educational level of the parents/legal–guardians has a strong positive impact to the student achievement gains.[16] Thus, students whose parents/legal–guardians have secondary or higher education better perform at both disciplines compared to students whose parents/guardians have at most the 3^{rd} cycle of schooling; such impact is higher in Mathematics. It is noteworthy that some results show that in the case of parents/legal–guardians who have completed at most the 1^{st} cycle of schooling a negative effect on student performance appears, in Mathematics in 2010 and in Portuguese in 2011. Furthermore, what seems to matter is the parents'/legal–guardians' schooling and not their employment status, as most of the estimates of coefficients of the employment situation variable are not statistically significant.

Economically disadvantaged students perform worse in Mathematics. In the period 2010–2012, the performance in Mathematics of a low income student (beneficiary of social support) is about 4%–8% lower, as compared to students from high income families. In Portuguese subject, no performance differences between the students that benefit from social support and those who do not benefit from these subsidies show up. Differences in Portuguese performance between low income and the other students don't seem to be so remarkable.

Including the measurable characteristics at the class (columns 3 and 7) level and

[16]The results relating to parents/guardians are not reported in Tables 5.4.1–5.4.3.

school level (columns 4 and 8), the results on individual–level variables do not change very much. In addition, controlling for prior achievement and student's and family characteristics, results show that almost all the coefficients of the variables do not have statistical significance and when coefficients are statistically significant the effect size is too small, as do the results in the OLS estimation. As such, it appears that classes with a higher proportion of female students negatively influence student achievement in Portuguese, but this effect is very small and close to zero. The proportion of immigrants at class level has no influence on student achievement. As expected, the share of parents/legal–guardians in class who have no more than the secondary education has a negative effect in achievement in both subjects and this effect is substantial; for instance, in 2012, increasing this by 1 percentage point leads to a decrease of about 0.3% on the Mathematics student achievement gains.

Among the variables that characterize classes and schools, the class size is of great interest. The class size effect is positive and statistically significant, mainly in the Mathematics discipline, suggesting that students in larger classes perform better than students in smaller classes. Although it seems to be counter–intuitive, as there is the idea that a teacher allocates more time to each student when there are fewer students, this result is common in the empirical literature (for instance, Hanushek, 1986; Woessman, 2003; Fertig and Wright, 2005). Nevertheless, the class size effect is non-linear. The negative coefficient of the variable "class size squared" suggests that the effect of class size on student achievement gains is convex, indicating that increases in the student performance are subject to diminishing returns to scale, reaching a peak at 22–23 students per class; and thereafter increasing class size adversely affects the performance of students in the class.[17]

Taking the results for Mathematics, column 4 in Table 5.4.3, estimated coefficient for class size is 0.0303 and for its square is -0.0006, indicating that an increase of 5 units in a class with 20 students yields a benefit on achievement of about 1.7%, but increasing this class size at 10 students the return is only about 0.3%.[18] Corresponding values Portuguese subject are 1.6% and 2.2%, respectively. In order to further explore the class size effect, cross–level interaction terms in the model were included, namely,

[17]This is in line with the OLS and random effects results previously obtained. A positive class size effect emerges, which is stronger in Mathematics results. Such effect, however, is convex, meaning that a peak is reached in classes with 22 or 23 students (depending on the model specification), and thereafter increasing class size adversely affects the performance of students in the class. This non–linear relationship between class size and student achievement is shown in Figure C.1 in the Appendix C.2. Considering the matched student–school panel, the random–effects estimation is used, and previous OLS results are confirmed. See Table C.3.1–C.3.3, in Appendix C.2. Noting that, these results also suggest that the effect of size class is stronger in Mathematics and that the optimal size is 23 students per class.

[18]The computation is given by $0.003 = (0.0303 \times 30 - 0.0006 \times 30^2) - (0.0303 \times 20 - 0.0006 \times 20^2)$.

class size×*beneficiary of social support* and *class size*×*gender*. The coefficients appear without statistical significance (results are not reported in Tables 5.4.1–5.4.3), indicating that there is no different effect of the class size on the performance of the low income students, as well as there is no gender difference in the class size effect.

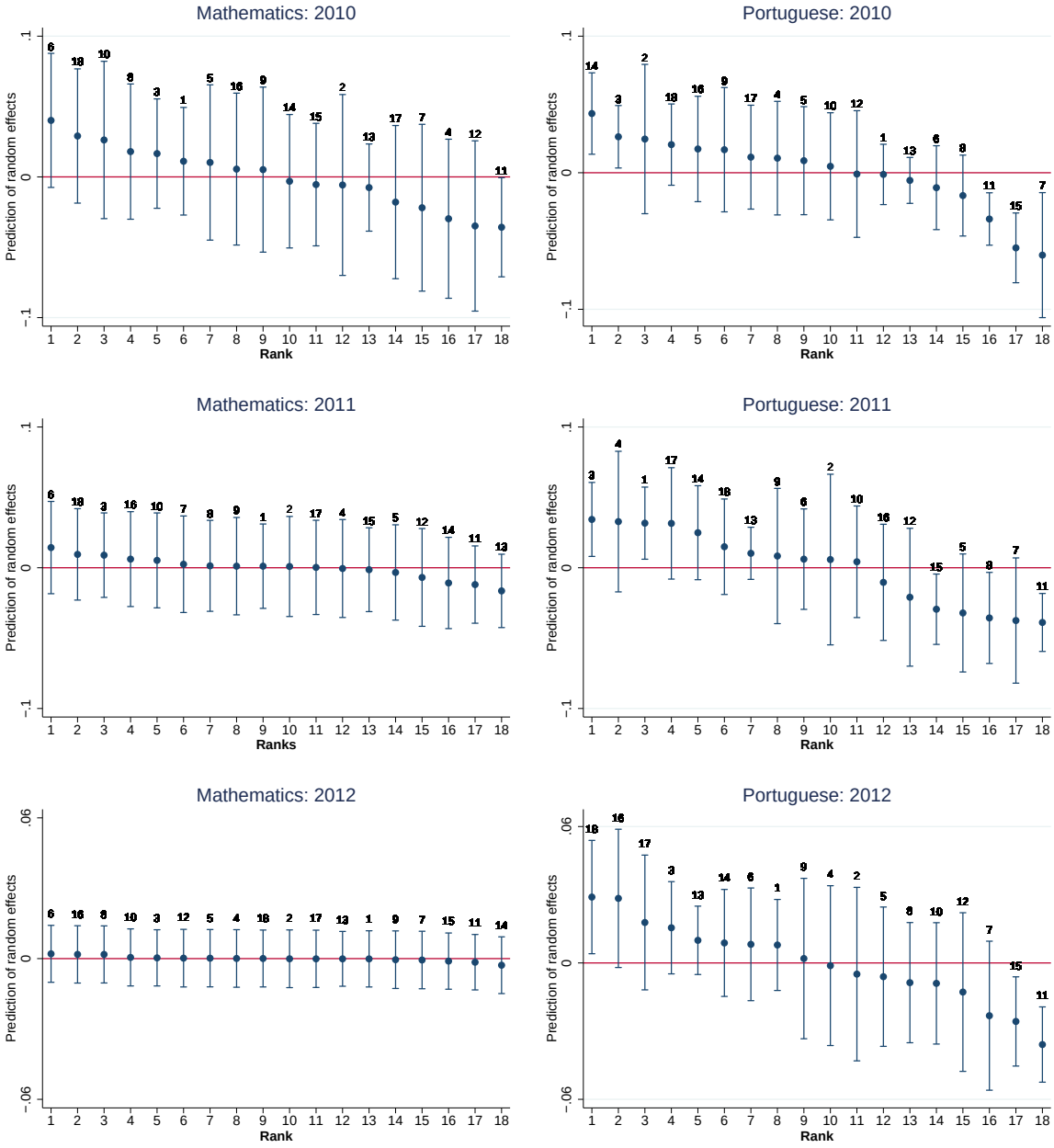
Summarizing, the main conclusion of the Coleman report seems to hold for Portugal, i.e., the individual and family characteristics of the students are the main observable determinants of inequality in educational achievement. However, one should pay attention to the class size, since the weight of this variable in student achievement can not be neglected.

Those results, however, do not account for the student's district of residence/school localization, as explained before. Therefore, a small exercise was performed to examine possible district–effects. A multilevel model with three levels is considered (level 1 – student; level 2 – school and level 3 – district). Most results at the student and the school level do not differ very much from the main model. Interesting enough is to look at the region–level results. Figure 5.6 shows the predicted random intercepts at the district level with approximate 95% confidence intervals based on the comparative standard errors and rank districts.

In general, the results show that, it is not possible to identify statistically significant differences between most of the districts, because the confidence intervals overlap. In Figure 5.6, it appears that the district–effect in achievement gains in Mathematics is not statistically different from zero, for all districts during the period 2010–2012. However, it may be noted that, in 2010, the district with students with the best performance is *Coimbra* (identified with number 6) and the district with the worst performance is Lisbon (number 11). In the case of district–effect on achievement gains in Portuguese, some differences between districts are identifies. The most efficient districts on achievement gains in Portuguese are: *Santarém* (number 14) followed by *Braga* (number 3), in 2010; *Braga* followed by *Bragança* (number 4) and *Aveiro* (number 1); and finally, *Viseu* (number 18) followed by *Viana do Castelo* (number 16), in 2012. Surprisingly, students in the districts of Lisbon and *Setúbal* (number 15) have a poor performance which takes place over the three years under study. Also, students in the district of *Évora* have a poor performance in 2010 and 2011. In part, these results are in line with those obtained in the OLS estimation with regard to the effect of location.

It is emphasized that, the district of Lisbon is poorly ranked in both disciplines and throughout the whole period under study. Given that almost half of the schools with secondary education in the district of Lisbon are private,[19] there is a very significant

[19] According to *Pordata* data, in the district of Lisbon, the percentage of private secondary schools



Source: Created by the author based on *MISI* and *JNE Statistics*, 2010–2012.
Note: Each number represents a district: 1 – Aveiro; 2 – Beja; 3 – Braga; 4 – Bragança; 5 – Castelo Branco; 6 – Coimbra; 7 – Évora; 8 – Faro; 9 – Guarda; 10 – Leiria; 11 – Lisboa; 12 – Portalegre; 13 – Porto; 14 – Santarém; 15 – Setúbal; 16 – Viana do Castelo; 17 – Vila Real; 18 – Viseu.

Figure 5.6: District effects and approximate 95% confidence intervals versus ranking of districts in Mathematics and Portuguese achievement gains (district identifiers are shown on the top of the error bar)

part of students attending them and, in principle, students with high socio–economic status. As the sample under study is constituted only by public schools, students living in the Lisbon district belonging to the sample are probably the most disadvantaged,

is 45%, 48% and 47%, in 2010, 2011 and 2012, respectively.

and which could justify the poor students' performance of the district.

5.4.2 Relative performance of schools: multilevel versus DEA approaches

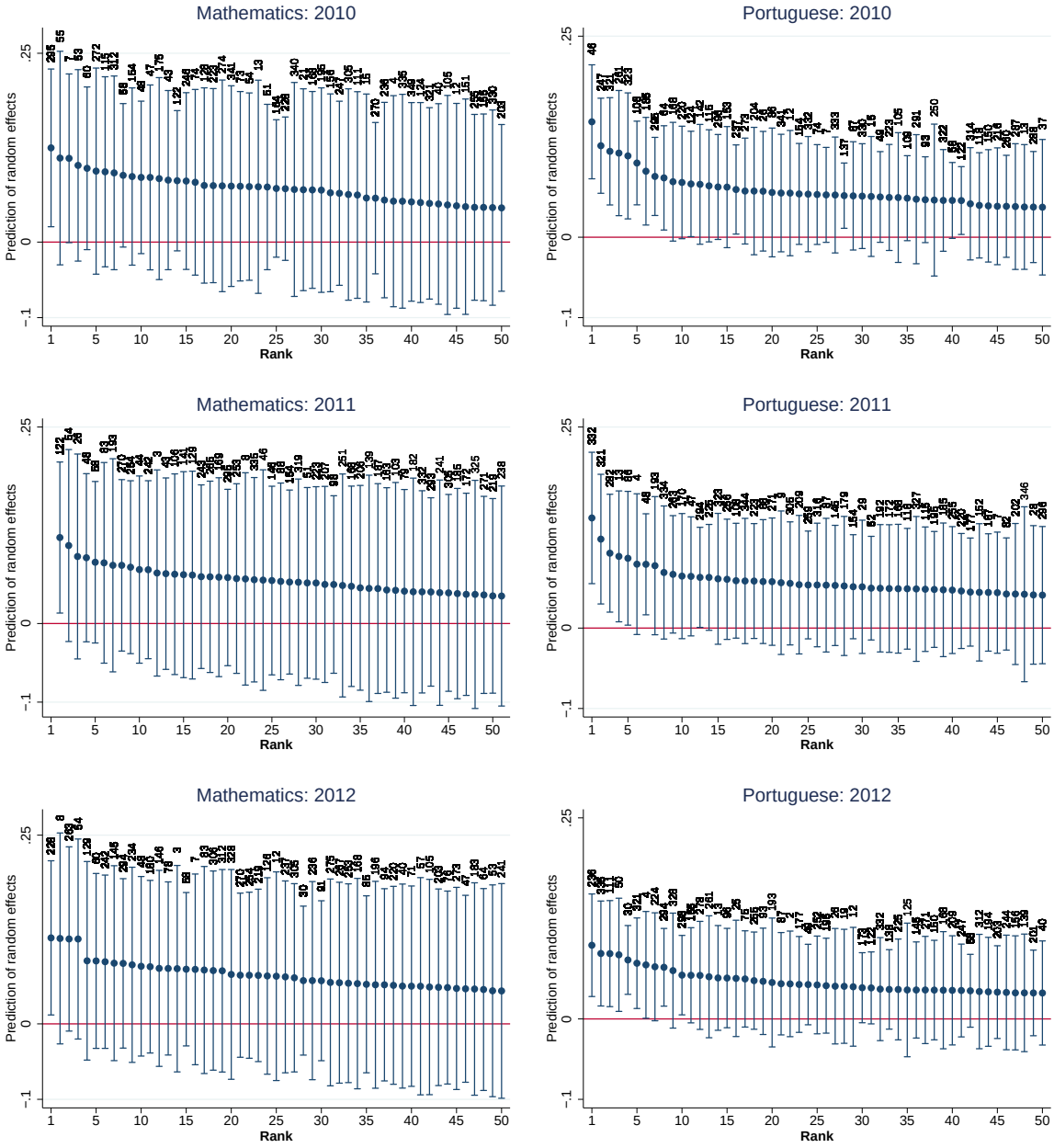
After having looked at the main determinants of student achievement gains, the second purpose of the present work is to look at the relative performance of schools, that is, to use performance to rank schools. Two alternative approaches will be used and results compared: the multilevel modelling and the data envelopment analysis.

The starting point is the main multilevel regression set of results (a three–level model, with students, classes and schools), whose random–intercept predictions can be viewed as measures of the school added value, i.e., how much value each school adds to its students. In order to compare schools in terms of added value, one can plot the predicted random intercepts at the school level ($\widehat{r}_{00k}$) and their 95% confidence intervals based on the comparative standard errors, and rank schools according to their institutional effectiveness.[20] Any two schools have significantly different residuals at 5%, if and only if their error bars do not overlap (Ferrão, 2003).

A graph has been plotted with the ranking of all schools in the sample based on residual estimates for each subject, Mathematics and Portuguese, and for each year under study. Each school is identified by a unique anonymous identification number, which preserves schools anonymity. The wide confidence intervals in these figures suggest that the rankings may be not very precise (figures show substantial overlap of the confidence intervals, suggesting that most of the schools can not be distinguished in terms of added value (as in Ferrão, 2012)). But these graphs are hard to read due to the number of schools represented. To facilitate the reading of the figures, the pool of schools has been divided into groups and only the top 50 best and worst schools were represented. Figures 5.7 and 5.8 show the added value of these schools (random intercept predictions) and the associated 95% confidence intervals versus their rank position, for each year from 2010 to 2012 and for both disciplines, Mathematics and Portuguese.

Looking at the top 50 best schools ranking in Figure 5.7, the number of schools showing an added value that is different from zero is higher in Portuguese than in Mathematics: according to the 2010 ranking, only one school in Mathematics and thirteen schools in Portuguese have a confidence interval that does not contain the zero. In 2010, the school with the best performance in Mathematics is also one of the best performing schools in Portuguese (in 8^{th} position). The number of schools in the

[20]Note that the school value added is derived from the model that includes student and family characteristics as explanatory variables.

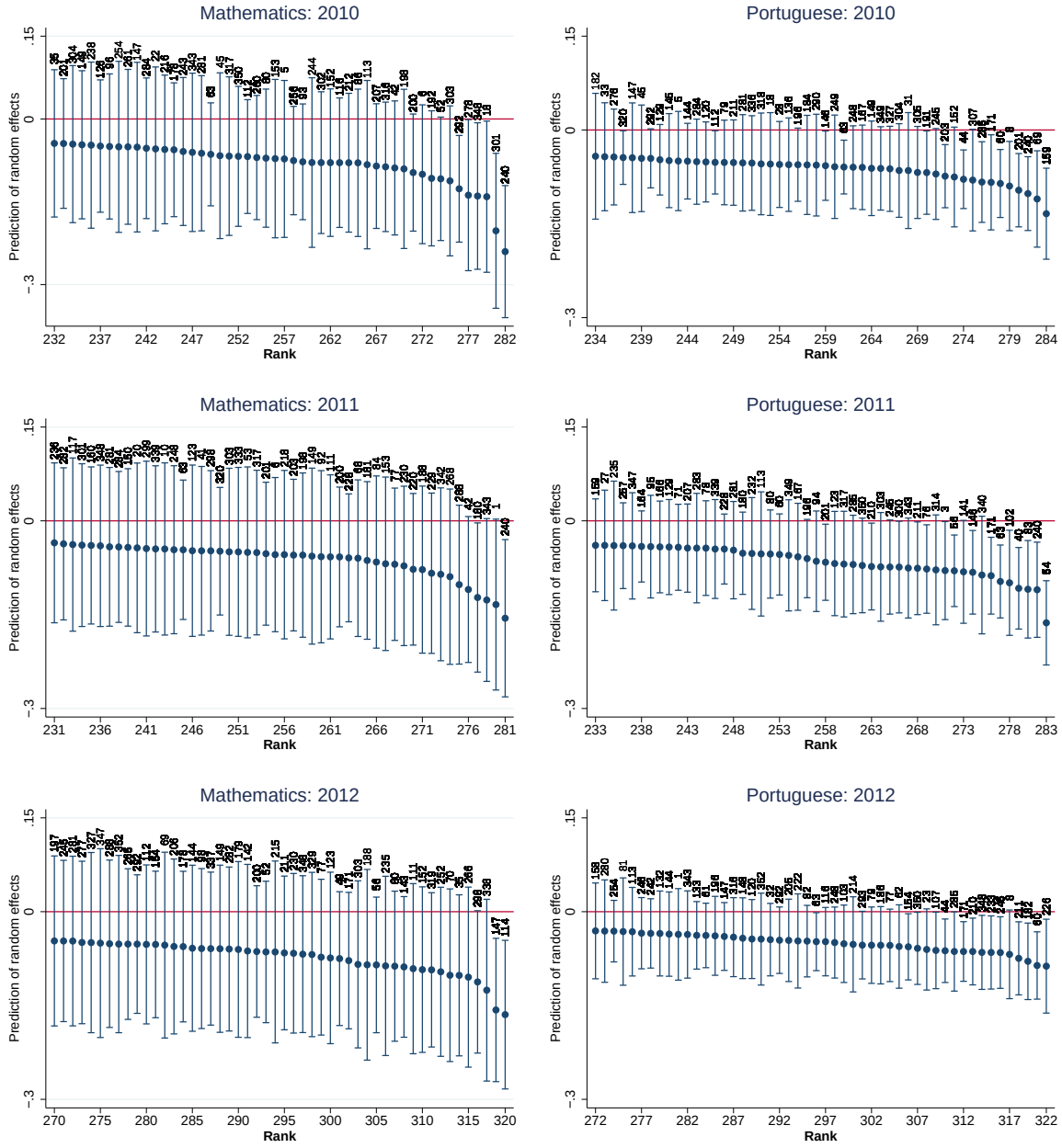


Source: Created by the author based on *MISI* and *JNE Statistics*, 2010–2012

Figure 5.7: School effects and approximate 95% confidence intervals versus ranking of 50 best schools in Mathematics and Portuguese achievement gains (school identifiers are shown on the top of the error bar)

top 50 of both disciplines has been reducing over time: there were 16 schools in 2011, 12 in 2011, and 9 in 2012. It is noteworthy that the persistency over time of the rank of schools is not the rule; only one school is in the top 10 best schools in Portuguese in the three years under analysis.

Figure 5.8 shows the top 50 worst schools. These schools have negative added values in both disciplines, but the performance appears to be worse in Portuguese than



Source: Created by the author based on *MISI* and *JNE Statistics*, 2010–2012.

Figure 5.8: School effects and approximate 95% confidence intervals versus ranking of 50 worst schools in Mathematics and Portuguese achievement gains (school identifiers are shown on the top of the error bar)

in Mathematics. Among the top 50 worst schools, the persistency of the rank over time is more frequent. Several schools rank in the top 10 worst schools in two out of the three years analysed, and there are one schools in the top 10 worst schools three years.

Alternatively, the literature has been applying the DEA to rank schools according to their efficiency, as described in section 5.3.1.2. The analysis has been applied to

both subjects, Mathematics and Portuguese, for each year in the period 2010–2012. The unit of observation is the school (DMU) and the data are aggregated at the school level, that is, the performance indicators measure the achievement of pupils in each school. The BCC model is applied, with variable returns to scale (VRS), that is, the proportionality between inputs and outputs is not assumed. For instance, if the average score of national 9^{th} grade exams of school A is n times higher than the average score of school B, there is no reason to expect that the average score of national 12^{th} grade exams of school A will be n times higher than of school B (Johnes, 2006). In addition, an output–oriented approach is used, implying that the school efficiency is maximized by maximizing the outputs subject to its given level of inputs.

The selection of the appropriate inputs and outputs has been extensively discussed in the literature on the DEA methodology. According to the literature, a large number of inputs and outputs compared with the number of DMUs may diminish the discriminatory power of the DEA, as the results generate a large number of efficient DMUs. A "rule–of–thumb" suggested for selecting the variables is that the number of inputs and outputs altogether should be smaller than one third of the number of DMUs (Cook et al., 2014).

The inclusion of inputs that reflect the student academic potential, as well as the socio–economic and cultural environment of the school is consensual in the literature. The inputs included were chosen based on the literature, as well as on the multilevel model results presented before (i.e., the statistically significant variables are considered). Dummy variables were not considered, as they are not appropriate in the context of the DEA (Johnes, 2006). The set of inputs and outputs used are described in Table 5.5.

Table 5.5: Inputs and outputs specified in the DEA model

Outputs	**Inputs**
Average 12th exam scores in Mathematics	Average 9th exam scores in Mathematics % of disadvantaged students in school % of parents with secondary education or less in school Weighted average class size
Average 12th exam scores in Portuguese	Average 9th exam scores in Portuguese Language % of disadvantaged students in school % of parents with secondary education or less in school Weighted average class size

Source: Created by the author based on *MISI* and *JNE Statistics*, 2010–2012.

First, the DEA has been performed using only one input and one output, such that, the distance measure computed by means of the DEA is not indeed an efficiency measure of schools, as school resources are not included in the analysis, but resembles

an added value model at the aggregate school level (Portela et al., 2012). Second, the weighted average class size is also included as an input, since the class size is only partially controlled by schools. The choice of the weighted average class size rather than its arithmetic mean is justified by the fact that the weighted average better characterizes the class size of a school and it is not as sensitive to small size classes. Third, to allow for comparisons, factors that are not under the control of students and schools, but may affect student attainment, are required. Namely, contextual variables such as the share of disadvantaged students in the school and the share of parents with no more than the secondary education in the school are included.

The descriptive statistics for the input and output variables for each year between 2010–2012 are reported in Table 5.6. Note that the scores of national exams for both 9^{th} and 12^{th} grades and for each subject vary between 0 and 200 points.

Table 5.6: Descriptive statistics of aggregated data on public secondary schools

Inputs and Outputs Variables	**Obs.**	**Mean**	**Std. Dev.**	**Min.**	**Max.**
			2010		
Average 12^{th} exam scores in Mathematics	282	125.4	18.58	69.75	180
Average 12^{th} exam scores in Portuguese	284	117.2	10.45	87	175
Average 9^{th} exam scores in Mathematics	284	95.13	16.26	47.14	188
Average 9^{th} exam scores in Portuguese	284	138.0	7.707	104	172
Weighted average class size	284	23.98	2.640	15.90	31.94
% of disadvantaged students in school	284	31.87	14.67	0	77.04
% of parents with secondary education or less in school	284	86.40	9.389	36.80	99.19
			2011		
Average 12^{th} exam scores in Mathematics	279	107.8	20.24	51.58	174
Average 12^{th} exam scores in Portuguese	281	106.4	10.97	76	149
Average 9^{th} exam scores in Mathematics	281	125.2	18.11	82.19	188
Average 9^{th} exam scores in Portuguese	281	138.7	8.517	94	174
Weighted average class size	281	24.61	2.350	18.92	30.93
% of disadvantaged students in school	281	29.84	14.14	0	69.63
% of parents with secondary education or less in school	281	86.53	9.902	36.55	99.56
			2012		
Average 12^{th} exam scores in Mathematics	317	104.9	18.04	25	166
Average 12^{th} exam scores in Portuguese	319	111.9	10.56	74.67	169
Average 9^{th} exam scores in Mathematics	319	132.6	14.58	74	179
Average 9^{th} exam scores in Portuguese	319	128.8	8.649	102.5	174
Weighted average class size	319	24.56	2.503	17.77	30.48
% of disadvantaged students in school	319	29.04	14.02	0	70.24
% of parents with secondary education or less in school	319	86.45	9.418	35.55	98.52

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2010–2012.

The DEA efficiency scores, for each year under study and for each subject, Mathe-

matics and Portuguese, are reported in Table C.4 in the Appendix C.3, which includes three rankings, depending on the inputs included, as described above. Besides that, only the top 20 best and worst schools are reported.

In all the three rankings presented for each subject, it appears that when more inputs are added to the objective function, the number of 100% efficient schools also increases, as expected. In order to compare these results with the school–effect found in the multilevel model, only the results of the last ranking in each subject will be discussed, i.e., column 7 (referring to schools in column 6 and rank position in column 1) and column 14 (referring to schools in column 13 and rank position in column).

The hierarchy of the efficiency scores shows that there are non negligible differences in performance betweens schools. In both subjects and every year, efficiency scores for the top 20 best schools are generally high, ranging between 0.91 and 1 in 2010, 0.93 and 1 in 2011, and 0.93 and 1 in 2012. The worst efficiency scores are, in Mathematics and Portuguese, respectively, 0.48 and 0.63 in 2010; 0.43 and 0.61 in 2011; and 0.19 and 0.58 in 2012. In general, the school performance is better ranked in Portuguese than in Mathematics.

In 2010, the number of benchmark schools (100% DEA score) in Mathematics and Portuguese, respectively, is 11 and 12 in 2010; 15 and 14 in 2011; and 10 and 20 in 2012. Table 5.7 shows the changes in 2011 and 2012 rankings of benchmark schools in 2010.

Table 5.7: Schools with 100% efficiency score in 2010 and their ranking in subsequent years

Mathematics			**Portuguese**		
Benchmark schools in 2010	Rank in 2011 (efficient score)	Rank in 2012 (efficient score)	Benchmark schools in 2010	Rank in 2011 (efficient score)	Rank in 2012 (efficient score)
242	benchmark	benchmark	242	benchmark	benchmark
305	47 (0.8534)	23 (0.9256)	86	benchmark	benchmark
51	60 (0.8364)	40 (0.8648)	48	benchmark	benchmark
105	103 (0.7830)	98 (0.7533)	105	60 (0.8484)	benchmark
314	155 (0.7121)	103 (0.7492)	323	67 (0.8350)	156 (0.8607)
41	259 (0.5747)	195 (0.6530)	146	227 (0.7140)	115 (0.8876)
2	115 (0.7714)	154 (0.6955)	108	43 (0.8778)	193 (0.8399)
55	152 (0.7155)	72 (0.7860)			
323	263 (0.5554)	240 (0.6058)			

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2010–2012.
Note: The remaining benchmark schools in 2010 are not included because they are not in the ranking during the three years under analysis.

It seems that there is no stability of the benchmark schools over the three years in both disciplines. Only one school kept is status as a benchmark school in both disciplines and in the three years under analysis; this is a school located in Lisbon. The

total number of benchmark schools during for the period 2010–2012 is 36 in Mathematics and 46 in Portuguese. On the one hand, about 34% of the benchmark schools in Mathematics and 28% of the benchmark schools in Portuguese are located in *Porto* and Lisbon. On the other hand, 42.5% of schools in the top 10 worst schools in the Mathematics ranking are located in *Porto* and Lisbon, whereas in the Portuguese ranking that proportion is 40% (most of which are located in in Lisbon).

In order to compare the results of both methodologies, the DEA and multilevel modelling, Table C.4 is compared with the Figures 5.7 and 5.8. By comparing the set of schools with a positive school–effect (Figure 5.7) with the set of schools with 100% efficiency score, there is not much overlapping. In Mathematics, only one school is in both rankings. In Portuguese, there are three schools in each year that are efficient regardless of the methodology used. The overlapping in the lower part of the rankings confirms the same very low level of coincidence of the rankings that result from the application of the two proposed methods. This suggests a very careful choice of the methodology to analyse the performance of schools should be careful, since that choice, in principle, affects the conclusions on efficiency (as in, Johnes, 2006).

5.5 Concluding remarks

This study was guided by two objectives: to improve the understanding of the factors that influence student achievement gains and school efficiency, and to shed some light on school rankings by comparing the multilevel model results with those of the data envelopment analysis. In pursuing those objectives, the study benefited from the availability of a unique matched school–student dataset on all public Portuguese secondary schools and their students.

The results provide evidence that student achievement depends on a whole range of factors, including student and family characteristics, and peers and school characteristics. The students' characteristics and their families are the main factors influencing student achievement gains, as expected. In accordance with previous literature, student prior achievement is the strongest predictor of current performance in both subjects, Mathematics and Portuguese. Female students perform better than males students in both fields and age has a negative and significant influence on student achievement gains, which probably reflects a grade repetition effect. The student nationality seems to have no influence on performance. Student background characteristics, in particular, the parent/legal–guardian education has an important, positive influence on student achievement gains, unlike his/her employment status and his/her kinship with the student, which do not matter. Low income students perform worse than the higher income

ones, mainly in Mathematics.

School resources, measured by the teaching expenditure–student ratio, have no effect on student achievement gains, as do not seem to be at work any peer–group effects at both the class and the school level. In addition, school size also has no impact on students' performance. It appears that schools located in the districts of *Braga*, *Santarém* and *Viseu* are the most efficient schools.

At the school level, the class size gets the highest weight in students' performance, mainly in Mathematics, in which, by increasing the number of students up to 23, the student achievement increases, probably due to spillover effects. Nevertheless, classes with more than 23 students seem to hinder students' performance. Results also suggest that class size effect does not differ between boys and girls, neither depends on the student income. Such class size effect deserves to be carefully analysed. More research is needed to see how it interacts with other factors such as teacher quality. It is important to use several different measures to determine the effects of making classes smaller, rather than the exam performance, and its (possible different) impact across disciplines should be explored. These results should be discussed in the context of recent policy measures. In fact, class size is subject to national legislative action.

Additionally, multilevel results, on the one hand, show that added value estimates overlaps zero for most schools, which is the same to say that in most cases the school–effect is null. On the other hand, there is more variability in student achievement gains within classes than between classes within schools or between schools, which confirms that the weight of school influence on student performance is very small. In general, the school–effects size on student achievement gains is small as compared to family background effects, as in the Coleman report. However, the class size effect on student performance can not be neglected.

Another purpose of this essay is to compare the results of two methods often used to assess the performance of Portuguese schools, multilevel modelling and DEA. An added value approach is used. In addition, another contribution of this essay is the fact that it is the first time that a database managed by the Portuguese Ministry of Education is used and, furthermore it covers all the students attending high school in Portuguese public schools.

As previously mentioned, the choice of analytical technique is important to perform a school ranking. In this sense, another purpose of this study is to compare two techniques for constructing performance indicators and, consequently, efficiency rankings: DEA and multilevel modelling. Results show that there is some similarity between efficiency measures provided by both methodologies, but the choice of methodology to analyse the performance of schools should be careful, since this choice, in principle,

affects the conclusions on efficiency and, consequently, the efficiency rankings.

Chapter 6

Teacher characteristics and student progress

"A teacher affects eternity;
he can never tell where his influence stops."
Henry Adams, *The Education of Henry Adams*

6.1 Introduction

Policy makers, school administrators, parents and students themselves support the notion that teacher quality is among the most significant determinants of academic success. A majority of education policy discussions focus on the role of teachers, and, as such, the challenge of empirical literature is to solve the teacher quality puzzle. That is, there is evidence that teacher quality is a key determinant of student learning and no other attribute of schools comes close to having the same influence on student achievement, but there is a lack of consensus about which observable teacher characteristics can account for this impact (Rivkin et al., 2005; Hanushek and Rivkin, 2010; Hanushek, 2011). The identification of such characteristics would inform education policies, namely it may have a role in the effectiveness of hiring policies. For example, in Portugal, the teacher education and experience are accounted for hiring and salary decisions, but there is no evidence that those characteristics are crucial for teacher quality.

Important literature on how teachers affect the performance of the students has emerged in the recent years (e.g. Rockoff, 2004; Rivkin et al., 2005; Clotfelter et al., 2006; Buddin and Zamarro, 2009; Goldhaber and Hansen, 2013; Guarino et al., 2015; Walsh et al., 2015). The most consensual finding is that teacher experience has a positive effect on student test scores (Rockoff, 2004; Rivkin et al., 2005; Clotfelter et al.,

2006). Despite the lack of consistency of the results, other observable characteristics of teachers are discussed in the literature, such as, the education level, teacher test scores, credentials and salary (Hanushek and Rivkin, 2006).

Considering that the socio–economic background of students does not explain everything and that teachers are the most important resource of a school, this essay aims at answering the following research questions: Do traditional human capital measures like experience and education explain differences in productivity among teachers? What other teachers characteristics influence students' achievement gains? How much of the variation in student achievement is explained by the characteristics of teachers? The objective of the study is to evaluate which observable characteristics of secondary education teachers influence the performance of their students, taking into account a whole set of other factors that may influence their progression, namely, student background and class size. This type of evaluation is essential to assess the effectiveness of any policy measures addressing schools and their teachers.

The empirical analysis uses a Portuguese matched student–teacher panel data set, for the period 2010 – 2012. The fixed–effects method in a value added perspective is applied. Note that, due to the fact that only very recently these data have become available for research, this essay is the first empirical work which analyses the quality of the teacher using Portuguese data at the student–level.

Results show that students taught by female teachers perform better than those attending classes with male teachers. Working away from home have a negative and significant effect on students' achievement. Furthermore, having an advanced degree seems to have no relationship with teacher quality as a measured by student achievement gains, i.e. teachers with postgraduate, master or PhD diplomas do not do better than those teachers who stop studying after the bachelor degree. Finally, more experienced teachers are more effective in increasing student achievement gains than those with less experience.

This chapter proceeds in the following way. In Section 6.2, a literature review on the on the determinants of teacher performance is carried out and Section 6.3 comprehends the Methodology and data, followed by Empirical results. The chapter ends with a section of Concluding remarks.

6.2 Literature review

Empirical literature underlines that teacher quality is a key factor in the academic performance of students and, therefore, the challenge is to identify the observable characteristics of teachers that signal the quality of teaching. In this sense, a large

body of literature that examines teacher quality characteristics and the relationship of indicators of those characteristics to teacher effectiveness has emerged.

According to Hanushek (2011), two key findings have emerged from this literature. On the one hand, teachers are very important, and no other measured aspect of schools is as important as teachers in student achievement. In this sense, Hanushek and Rivkin (2010) find that the average standard deviation of the teacher fixed–effect for reading and maths is 0.11 and 0.15, respectively, and Rockoff (2004) concludes that a one standard deviation increase in teacher quality results in a 0.11 standard deviation increase in reading and writing test results. On the other hand, some studies have tried to analyse how the observable characteristics of teachers influence student performance, but it has not been possible to identify any specific teacher characteristic that are reliably correlated with student performance (Hanushek, 2011).

Literature on teacher quality has focused on measurable and observable teacher's characteristics, such as, years of teaching experience, education level, teacher test scores, certification and salary (e.g. Hanushek and Rivkin, 2006, 2010; Clotfelter et al., 2007, 2010; Buddin and Zamarro, 2009; Kukla-Acevedo, 2009; Goldhaber and Hansen, 2013). Other teacher's characteristics are analysed, namely, gender (Woessman, 2003; Clotfelter et al., 2010; Leigh, 2010, e.g.) and race/ethnicity (Egalite et al., 2015, e.g.). More recently, studies have been focused on teaching activities (e.g. Schwerdt and Wuppermann, 2011; Witte and Klaveren, 2014; Lavy, 2015).

Both teaching experience and education level have received a prominent place in the literature on the determinants of teacher quality. Although some studies suggest that the association between teaching experience and student performance is weak, in general, it is found that teacher experience have a significant positive effect on maths and reading test results (Akerhielm, 1995; Woessman, 2003; Rockoff, 2004; Rivkin et al., 2005; Clotfelter et al., 2007; Croninger et al., 2007; Leigh, 2010). This positive effect appears to be of non–linear nature as demonstrated by substantial improvements in teaching skills during the first 3–5 years in the classroom (Rivkin et al., 2005; Kukla-Acevedo, 2009; Buddin and Zamarro, 2009; Hanushek, 2011). High levels of teacher experience may have important benefits for schools (Buddin and Zamarro, 2009).

The literature also reports that teacher education level has no effect on mathematics and reading scores in elementary and middle school (e.g. Rivkin et al., 2005; Clotfelter et al., 2007; Kane et al., 2008; Buddin and Zamarro, 2009) and proposes that having a master degree has no systematic relationship to teacher quality as measured by student outcomes (Hanushek and Rivkin, 2006). But, for instance, Woessman (2003), Carrell and West (2010) and Croninger et al. (2007) conclude that the educational level of teachers is positively related to students performance.

Also related with the teacher characteristics, the findings in the literature show that teacher credentials matter; teacher licensure test scores have positive effects on student achievement; and the effects are particularly large for achievement in Mathematics (Hanushek and Rivkin, 2006; Clotfelter et al., 2006, 2007). This result, however, is contradicted by Buddin and Zamarro (2009). Moreover, the effects of teachers' credentials appear to be quite large comparing to the estimated effects of changes in class size or to the socio–economic characteristics of students, particularly, in maths (Clotfelter et al., 2007). In addition, empirical research shows no strong evidence that salaries are a good measure of teacher quality and several studies show that salaries are more likely to be positively related to student achievement than negatively (Figlio, 1997; Hanushek and Rivkin, 2006). There is no consensus on the effect of the teacher gender, as well. In some works it appears that students of female teachers have a performance statistically significantly higher than students of male teachers (Woessman, 2003; Clotfelter et al., 2010; Leigh, 2010), but there are studies which contradict this result (Akerhielm, 1995).

From the available research on this topic it emerges that, in general, these variety of teacher attributes exhibits an effect on student achievement and that the effects are larger for student achievement in maths than in reading.

Looking at the methodology in use in those studies, early literature relied on cross–sectional data aggregated at the level of schools or even school districts, i.e. the average school test scores are related to aggregate measures of teacher proficiency (Buddin and Zamarro, 2009). Consequently, such data did not allow for complete control of the students characteristics, such as prior achievement, and the allocation of students to teachers. Recent literature has improved those aspects. New analyses have been made possible by the availability of administrative data and the emergence of empirical approaches which increase the quality of research. Thus, the most recent literature on teacher quality uses panel data to better control for student heterogeneity and in some cases teacher heterogeneity (Buddin and Zamarro, 2009; Slater et al., 2012). Furthermore, the so–called value added models have been used to measure the importance of teacher quality to the educational production (e.g. Rivkin et al., 2005; Aaronson et al., 2007; Clotfelter et al., 2007; Buddin and Zamarro, 2009; Carrell and West, 2010; Goldhaber and Hansen, 2013).

The estimation methodologies employed in these studies are diversified, but the most common estimation methods in the literature are least squares regression techniques (e.g. Aaronson et al., 2007), fixed–effects model (e.g. Rivkin et al., 2005; Buddin and Zamarro, 2009), random–effects model (e.g. Carrell and West, 2010) and multilevel modelling (e.g. Croninger et al., 2007).

Note that, most of these studies relate to students and teachers in the United States (Figlio, 1997; Rockoff, 2004; Clotfelter et al., 2006, 2007, 2010; Aaronson et al., 2007; Kane et al., 2008; Buddin and Zamarro, 2009; Carrell and West, 2010; Goldhaber and Hansen, 2013). Countries outside the United States, have received little attention in research on the measurement of teacher performance (Leigh, 2010; Slater et al., 2012). In Portugal, literature on the characteristics of teachers who influence the performance of secondary school students is practically absent. It is known the work of Pereira and Moreira (2007) that suggest that the teacher's age, used as a proxy for the experience, is an important determinant of the student achievement. Martins (2009a) examines the effects of teacher performance–related pay and tournaments in public schools implemented in Portugal in 2006–2007. Using schools in the Portuguese Islands, Azores and Madeira, as well as private schools as controls, there is no evidence of achievement gains induced by the program and, in addition, the results indicate that the increased focus on individual teacher performance caused a significant and sizeable relative decline in student achievement, as measured by national exams.

The present study aims at filling that gap in the literature, and will provide results on the teacher characteristics that determine student performance.

6.3 Methodology and data

6.3.1 Methodology

In the economics literature, the main empirical strategy used to assess the importance of teachers and teacher characteristics is the estimation of education production functions, that recognises education as a cumulative process (value–added model), which generally take the following form:

$$\log A_{ijt} = \lambda \log {A_i}^9 + \beta X_{it} + \delta T_{jt} + \gamma_t + \varepsilon_{ijt} \tag{6.1}$$

where i, j and t refer to students, teachers, and time, respectively. A_{ijt} is the student outcome in the grade 12^{th} national exams Portuguese or Mathematics A, measured by the national exam score. The lagged test score ${A_i}^9$ is the score of the i^{th} student in the national exam of 9^{th} grade (for simplicity, the lagged achievement term refers to the same subject as grade used as the dependent variable). It is included in the equation to reflect the cumulative nature of the education process and it is intended to capture the effects of prior achievement. Namely, there are unobserved characteristics of students, such their ability and motivation, that have effects on achievement that are constant over time, which are controlled by prior achievement (Clotfelter et al.,

2006). Additionally, teachers are non–randomly assigned to students, by schools, and the common practice in the literature to deal with non–random sorting is to control for students' prior achievement (Hanushek and Rivkin, 2010). X_{it} is a vector of student and family background characteristics; vector T_{jt} represents measurable teacher characteristics and includes the class size, as well. ε_{ijt} is a random error. The model also controls for year of the examination, γ_t.

Regarding the student and family background characteristics only the explanatory variables that in the previous chapter showed a statistically significant effect in achievement gains are included, i.e. gender, age, beneficiary of social support, internet at home and a set of dummies to control for the district of residence. Although, in the previous chapter, a relationship between parental/legal–guardian education and the students' achievement gains is found, in this analysis this variable is not included, because it contains many missing observations and its inclusion would imply the loss of about 6,000 students/observations. The teacher characteristics included are the traditional measures of human capital, such as experience and education; gender; and a proxy to the teacher motivation (i.e. weather the teacher is working away from your home, based on the argument that the closer to the family the teacher, the more motivated). Note that, the average salary of teacher was not taken into consideration, given the obvious collinear with seniority, since in Portugal, the teacher wage tables are based on years of teaching experience. Because it does not consider weather the teacher is effective or not due to lack of variability in the data, it seems that most of the teachers teaching the 12^{th} grade are teachers with tenure.

This analysis starts with an estimation of the equation (6.1) by Ordinary Least Squares (OLS). According to the empirical literature, any study analysing the effect of teachers on student achievement has to deal with important potential identification problems that might bias the conventional OLS estimates. Even when students are randomly assigned to teachers, there may be unobserved teacher traits that are correlated with student outcomes which again may bias the conventional OLS estimates. This may be the case, for instance, when there are unobserved gender specific differences across teachers' quality. As mentioned above, the inclusion of prior achievement eliminates any bias associated with the non–random matching of teachers and students and, as would be the case in longitudinal studies, that the effects of the T_{jt} variables are estimated within students. In this case, it means that they are based only on the variation in teacher characteristics across the subjects for each individual student (Clotfelter et al., 2007, 2010).

To deal with the omitted time–invariant variables, fixed–effects are included in the model, which can be expressed as:

$$\log A_{ijt} = \lambda \log A_i{}^9 + \beta X_{it} + \delta T_{jt} + \tau_j + \varepsilon_{ijt} \tag{6.2}$$

where τ_j is a teacher fixed–effect. If some teachers perform better than others, due to their tastes, ability, or other time–invariant factors, teacher fixed effects will account for those differences. Note that, the school fixed–effect is omitted, because most teachers change schools throughout their career. Time fixed–effects are not included due the collinearity with the variable *experience* (See discussion on this issue in the Appendix B.2).

Finally, it will be implemented the Wooldridge test for autocorrelation in panel-data (Wooldridge, 2010). Auto-correlation in linear panel-data models biases the standard errors and causes the results to be less efficient.[1] Additionally, a Breusch–Pagan test for heteroscedasticity, as well as a Wald test for groupwise heteroscedasticity in fixed-effect model will be applied (Greene, 2012).

6.3.2 Data description

The Data used in this study are managed and arranged by the Portuguese Ministry of Education: MISI (*Sistema de Informação do Ministério da Educação*), managed by *Direção de Estatísticas da Educação e Ciência* (DGEEC) and Statistics published by *Júri Nacional de Exames – Direção Geral de Educação* (JNE). The data are obtained from the administrative records of all teachers and students in Portugal. The first dataset provides information at the student and teacher level, and the second one contains data at the student–level on the scores obtained in the national exams in both basic and secondary education. These two databases were merged using the student's identification.[2]

Thus, a matched student–teacher dataset was created, which includes the national test scores of students in Mathematics and Portuguese, student background information, and teacher information, from school years 2006–2007 to 2011–2012. There is information at the student–level, such as gender, date of birth, nationality, academic outcomes, year of schooling, social support eligibility, residence, availability of computer and internet at home, parents' employment situation and parents' education,

[1] Wooldridge test for autocorrelation under the null hypothesis that there is no serial correlation, i.e. in the regression of the first-differenced variables should have an autocorrelation of -0.5, implying that the coefficient on the lagged residuals in a regression of the lagged residuals on the current residuals should be -0.5 (See Wooldridge, 2010; Drukker, 2003).

[2] The two datasets, MISI and the exam data from JNE, we had access for this research, have been previously anonymized to absolutely secure private information on individuals, classes and schools. All the information on individuals, namely students and professors, cannot be individually traced by the researcher. See additional discussion in Chapter 5.

class and school, among others. At the teacher–level, it provides information about gender, date of birth, education, teaching experience, disciplinary group, salary, county and district of residence, among others.

From the original data the Portuguese students of high–school who made national test in Mathematics and Portuguese, during 2010–2012, and their corresponding teachers were selected. Note that, only the period 2010–2012 is considered, because the 9^{th} national exam scores are used as control to the prior achievement and students who made national tests in 2010, 2011 and 2012 made national 9^{th} grade exams in 2007–2009.

Table 6.1 provides descriptive statistics for all the explanatory variables included in the regressions.[3] The outcome variable is the log of 12^{th} national exam score of Mathematics and Portuguese, that measures the teacher productivity, and the explanatory variables were grouped in each of the different levels: student and teacher.

Table 6.1: Descriptive statistics

Variable	**Mean**	**Std. Dev.**	**Min**	**Max**
Student–level variables				
12^{th} national exam score	112.7	37.46	0	200
9^{th} national exam score	137.1	31.18	2	200
Female student	0.566			
Age	18.08	0.349	17	20
Beneficiary social support	0.228			
Internet	0.729			
Teacher–level variables				
Female teacher	0.750			
Advanced degree	0.082			
Experience	24.70	6.613	5	40
Commuting	0.351			
Class size	25.87	4.843	2	40

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2010–2012.
Note: The sample includes 21,549 observations.

The sample contains 21,549 student observations and 4,817 unique teachers, over three years, 2010–2012. These students attended the scientific–humanistic courses of secondary education and they performed the 12^{th} grade national exams of Mathematics and Portuguese, of which 38% took the 12^{th} mathematics exam. Their ages vary between 17 and 20 years, and 91% of students are 18 years old. About 23% of the students benefit from social support and about 73% of the students have internet

[3]See Table D.1 in Appendix D.1 for a description of the variables.

access at home. Students are distributed across 18 districts/regions.

This dataset only includes teachers of Mathematics and Portuguese subjects, working in 446 Portuguese public secondary schools. All these teachers are hired by the Ministry of Education and about 35.1% are working outside their county of residence. It is observed that 91.8% of the teachers have a *bacharelato* or bachelor degree as the highest level of education and 75% are female. Experience is a continuous variable that measures the number of years the individual has been teaching; half of the teachers have at least 25 years of experience.

6.4 Empirical results

Equation 6.1 is estimated using pooled OLS for the all sample and for each subject, Mathematics and Portuguese, separately. Estimations are performed with robust standard errors to account for heteroscedasticity, once Breusch–Pagan test rejects the null hypothesis of homoscedasticity at 1% level. Table 6.2 shows the estimation results for three alternative specification that use different sets of variables, although all specifications include both students' and teachers' characteristics. Note that, results reported in columns 2, 5 and 8 include the class size and its square. In order to examine the non–linearity of the experience effects, the experience squared variable is also included in columns 3, 6 and 9.

The results reported in the Table 6.2 show some differences in achievement gains across different students, as expected. So, students' prior achievement score is the strongest predictor of their current academic performance. Female students have higher growth rates in achievement than male students in both Mathematics and Portuguese. Age has a negative and significant effect on achievement gains and this effect is two times higher in Mathematics than in Portuguese. Socio–economic status, proxied by beneficiary social support status of the student variable, has a negative impact on achievement gains, i.e. beneficiary students have a lower increase in their results in Mathematics and Portuguese than other students. This impact is higher in Mathematics; beneficiary of social support students have a Mathematics performance of about 8% worse than other students and the corresponding value in Portuguese subject is about 3%.

As indicated in the previous chapter ("Factors that influence Student Achievement gains and Performance Assessment of the Portuguese public schools"), the positive coefficients of class size and the negative coefficients of class size squared indicate a non–linear relationship between the class size and the performance. This relationship is stronger in Mathematics than in Portuguese. Taking the results for Mathematics,

Table 6.2: OLS regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All students			**Mathematics**			**Portuguese**		
Log of 9^{th} exam scores	0.7110***	0.7101***	0.7101***	0.9534***	0.9503***	0.9505***	0.6972***	0.6959***	0.6959***
	(0.0129)	(0.0129)	(0.0129)	(0.0242)	(0.0241)	(0.0241)	(0.0133)	(0.0133)	(0.0133)
Female student	0.0598***	0.0597***	0.0598***	0.0594***	0.0589***	0.0589***	0.0602***	0.0604***	0.0604***
	(0.0050)	(0.0050)	(0.0050)	(0.0103)	(0.0103)	(0.0103)	(0.0044)	(0.0044)	(0.0044)
Age	-0.1211***	-0.1210***	-0.1210***	-0.1865***	-0.1845***	-0.1841***	-0.0974***	-0.0968***	-0.0968***
	(0.0077)	(0.0077)	(0.0077)	(0.0234)	(0.0232)	(0.0232)	(0.0066)	(0.0066)	(0.0066)
Beneficiary S.S.	-0.0494***	-0.0489***	-0.0488***	-0.0853***	-0.0842***	-0.0844***	-0.0301***	-0.0295***	-0.0296***
	(0.0062)	(0.0062)	(0.0062)	(0.0141)	(0.0141)	(0.0141)	(0.0053)	(0.0053)	(0.0053)
Internet	0.0196***	0.0195***	0.0196***	0.0175	0.0177	0.0179	0.0157***	0.0157***	0.0157***
	(0.0057)	(0.0057)	(0.0057)	(0.0126)	(0.0126)	(0.0126)	(0.0049)	(0.0049)	(0.0049)
Female teacher	0.0219***	0.0219***	0.0222***	0.0288**	0.0278**	0.0281**	0.0120**	0.0120**	0.0120**
	(0.0060)	(0.0060)	(0.0060)	(0.0117)	(0.0117)	(0.0117)	(0.0053)	(0.0053)	(0.0053)
Advanced degree	0.0157*	0.0161*	0.0165*	-0.0082	-0.0089	-0.0080	0.0254***	0.0253***	0.0252***
	(0.0089)	(0.0089)	(0.0089)	(0.0198)	(0.0198)	(0.0197)	(0.0071)	(0.0071)	(0.0071)
Experience	0.0013***	0.0013***	-0.0010	0.0012	0.0012	-0.0061	0.0007**	0.0007**	0.0011
	(0.0004)	(0.0004)	(0.0022)	(0.0008)	(0.0008)	(0.0046)	(0.0003)	(0.0003)	(0.0018)
Experience sq			0.0000			0.0002			-0.0000
			(0.0000)			(0.0001)			(0.0000)
Commuting	-0.0081	-0.0078	-0.0081	-0.0162	-0.0152	-0.0163	-0.0084*	-0.0080*	-0.0079*
	(0.0053)	(0.0053)	(0.0053)	(0.0114)	(0.0113)	(0.0114)	(0.0046)	(0.0046)	(0.0046)
Class size		0.0156***	0.0156***		0.0355***	0.0356***		0.0064**	0.0065**
		(0.0035)	(0.0035)		(0.0087)	(0.0087)		(0.0028)	(0.0028)
Class size sq		-0.0003***	-0.0003***		-0.0007***	-0.0007***		-0.0001**	-0.0001**
		(0.0001)	(0.0001)		(0.0002)	(0.0002)		(0.0001)	(0.0001)
Observations	21,549	21,549	21,549	8,100	8,100	8,100	13,449	13,449	13,449
R-squared	0.251	0.252	0.252	0.298	0.300	0.300	0.329	0.329	0.329
RMSE	0.245	0.245	0.245	0.245	0.245	0.245	0.245	0.245	0.245

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2010–2012.

Note: Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is log 12^{th} grade national exam score. All regressions include a set of dummies to control for district/region and year.

column 6, estimated coefficient for class size is 0.0356 and for its square is -0.0007, indicating that an increase 5 units in a class with 20 students yields a benefit on achievement of about 2.1%, but increasing this class size at 10 students the return is only about 0.6%.[4] Corresponding values for Portuguese subject are 1% and 1.5%, respectively. The results also indicate that the optimal number of students per class is 25 for Mathematics, which is very close to the number of students per class that emerged from the results in the previous chapter.

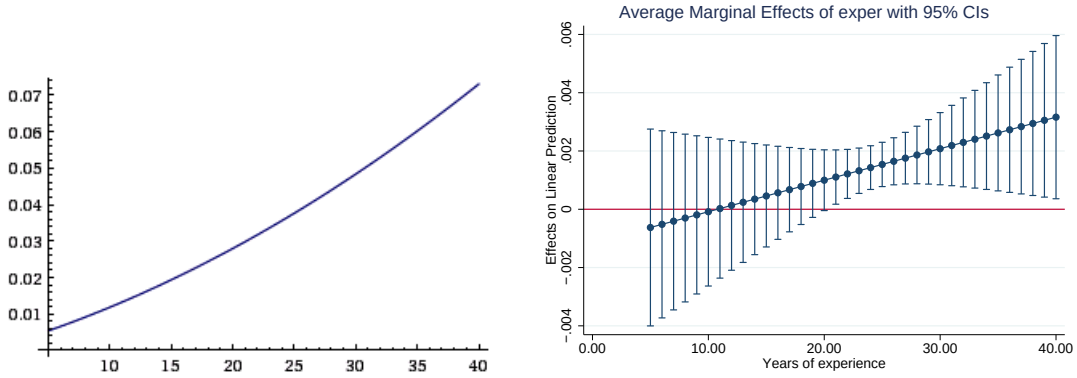
Looking at teachers' characteristics, it appears that females have a stronger effect on student achievement gains than male teachers, in both subjects. Thus, student achievement gain is about 3% and about 1% higher, in Mathematics and Portuguese, respectively, when they have female teachers as compared to males. Although only few studies account for the gender of the teacher, this result has been found in the empirical literature (e.g. Buddin and Zamarro, 2009).

The results for the variable commuting, which is a proxy for teacher motivation, as explained in Section 6.3.1, show that teachers working away from your home have similar performance when compared to teachers who work close to their place of residence. It is found that commuting Portuguese teachers negatively influence the student achievement gains, but this effects is small and it is statistically significant only in the Portuguese outcomes model. This is possibly due to lower motivation of commuting teachers, whose incentives for a better performance may be affected by the fact that they are away form home and family and earning the same wage as if they were working in a school close to home. It may also be related with the tiredness of commuting to school every day. There are no differences in perform of female commuting teachers and male commuting teachers. It is not possible to compare this result with those in the existing literature, since, as far as we know, that teacher characteristic has not appeared in previous literature.

In general, teacher qualifications seem to have no relationship with teacher quality as a measured by student performance, i.e. teachers with more qualifications (postgraduate, master or PhD) do not perform differently from those teachers with a licentiate degree. This result is confirmed by the existing empirical literature (Hanushek and Rivkin, 2006; Croninger et al., 2007; Buddin and Zamarro, 2009). Only one exception applies in the Portuguese case, that is, teachers with an advanced degree positively appear to have some positive influence on the Portuguese exam results as compared to those teachers holding a *bacharelato* or licentiate diploma (i.e., the effect is about 3% higher for teachers who have a postgraduate, master or PhD degrees than teachers with the minimum qualification for teaching).

[4]The computation is given by $0.006 = (0.0356 \times 30 - 0.0007 \times 30^2) - (0.0356 \times 20 - 0.0007 \times 20^2)$.

Finally, the results of the linear regression reported in columns 1 and 2 in the Table 6.2 show that, on average, teaching experience positively influences the students' performance, although this effect is small (i.e., an additional year of teaching experience results in an increase on student results of about 0.13%). Furthermore, the non–significance of the coefficients of both *experience* and its squared suggests that the relationship between teacher experience and student performance is not non–linear of degree 2. On the one hand, those results only partly confirm the human capital theory, which states that additional experience has a positive effect on the worker productivity, but diminishing marginal returns apply, whereas previous research on student achievement is in accordance with the theory of human capital (e.g. Hanushek and Rivkin, 2006; Clotfelter et al., 2006, 2007). On the other hand, when applying the Wald test to those variables, experience and experience squared, the null hypothesis that both coefficients are equal to zero is rejected indicating that the teaching experience is a relevant characteristics on students' achievement. Therefore, it is of all interest to calculate the marginal effects along the distribution of the experience variable, since the marginal effects provide a good approximation to the amount of change on students' achievement that will be produced by one unit change in teaching experience. Marginal effects of the *experience* variable relating to the 2^{nd} degree polynomial model, above mentioned, are presented on the right side of the Figure 6.1. On the left side of the same figure, it is represented the graph of the polynomial function in the variable experience.



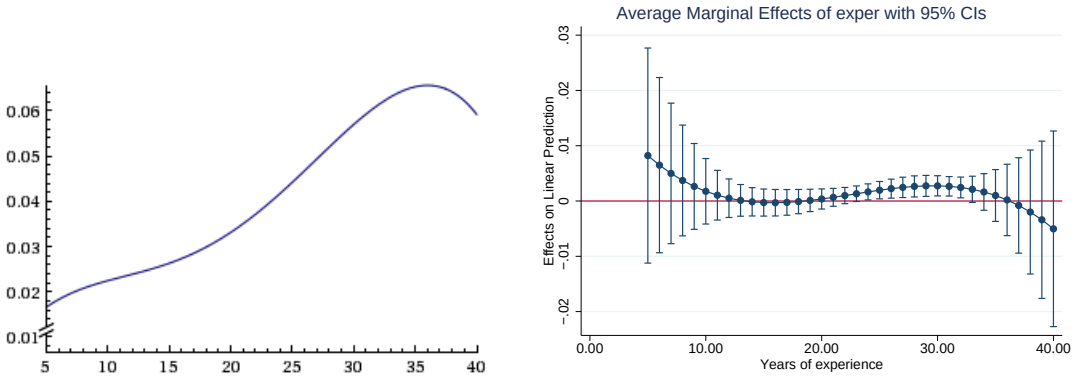
Source: Created by author based on *MISI* and *JNE Statistics*

Figure 6.1: Experience in a 2^{nd} degree polynomial model, 2010–2012

It can be seen that teachers with 20 or more years of experience have a positive and statistically significant effect on students' achievement, confirming that this effect is different from zero. For example, a teacher with 30 years of experience performs better than a teacher with 20 years of experience in about 1.5%.[5] However, as previously discussed, the 2^{nd} degree polynomial does not seem to be the model that best fits to

[5]The computation is given by $0.015 = (-0.0010 \times 30 + 0.00004 \times 30^2) - (-0.0010 \times 20 + 0.00004 \times 20^2)$.

the distribution of the *experience* variable. In order to get a better model fit, it seems to be necessary to consider a higher degree polynomial. For that purpose, a polynomial of 4^{th} degree is used, and its results are shown in Figure 6.2. It represents the marginal effects of *experience*, on the right side, and the polynomial function of the variable experience, on the left side.



Source: Created by author based on *MISI* and *JNE Statistics*

Figure 6.2: Experience in a 4^{th} degree polynomial model, 2010–2012

As proposed by the human capital theory, teaching experience shows diminishing returns. The Wald test for the 4^{th} order polynomial on experience is applied and the null hypothesis that the four coefficients are equal to zero is rejected (i.e., teaching experience is a relevant teacher characteristic on students' achievement). Figure 6.2 shows that teaching experience has a positive and significant impact on students' achievement when teachers have between 23 and 33 years of experience. In addition, it can be observed that, for example, comparable achievement gains is about 1.8% higher with teachers who have 30 years of experience than with teachers who have 20 years of experience. Thus, teachers with more teaching experience perform better than novice teachers.[6]

Given the unobserved heterogeneity of teachers, the OLS estimates may be biased and inconsistent if the quality of teachers is correlated with any of their observed characteristics. To control for unobserved heterogeneity of teacher, the specification (6.2), a fixed–effects model, is estimated. Note that, the inclusion of the teacher fixed–effects yields more robust estimates, since characteristics such as motivation and ability to transmit knowledge are controlled. An F–test rejects the null hypothesis of absent specific effects, which is evidence that there are significant unobserved individual effects, so pooled OLS would be inappropriate. In addition, the Hausman–test rejects the null hypothesis that random–effects estimator is consistent at the 1% significance level,

[6]The computation is given by $0.0182 = (0.0208 \times 30 - 0.0016 \times 30^3 + 0.0001 \times 30^3 - 5.80e-07 \times 30^4) - (0.0208 \times 20 - 0.0016 \times 20^3 + 0.0001 \times 20^3 - 5.80e-07 \times 20^4)$.

so the unobservable individual effects are correlated with the explanatory variables. Combing both results, existence of unobserved heterogeneity and inconsistency of the random-effects model, implies that the fixed–effects estimation is the most suitable model, among the ones discussed above. Note that, estimations are performed with robust standard errors to account for possible heteroscedasticity. A modified Wald test for groupwise heteroscedasticity in the fixed-effect regression model rejects the null hypothesis of homoscedasticity at the 1% significance level.

Additionally, the Wooldridge test for autocorrelation was performed, but one cannot reject the null hypothesis of no serial the residuals in the error term of equation (6.2). Both OLS and fixed–effects models are run on the full sample of students, as well as, for two subgroups (the subjects Mathematics and Portuguese are analysed separately). Note that, as previously discussed, the inclusion of students' prior achievement controls for the unobserved heterogeneity of students.

Results from the fixed–effects estimations are reported in the Table 6.3. Naturally, the variable *gender* of the teacher is not reported in this table, because fixed–effects regression does not provide estimates for time–invariant variables. As mentioned above, time fixed–effects are also not included, because experience and year are collinear within teachers.

Table 6.3 shows that 49% of the variance in students' achievement is due to differences across teachers. This variability is two times higher for Mathematics teachers, 79%, than for Portuguese teachers, 19%.

By comparing with the previous findings, the results reported in Table 6.3 show that there are no significant changes with regard to observable characteristics of students, i.e., in general, both level of significance and magnitude of the coefficients of the explanatory variables at the student level are very similar to those obtained before. The student's prior achievement remains as the strongest predictor of their current academic performance. Furthermore, female students have higher growth rates in results than male students in both subjects and, older students and beneficiary of social support students perform worse than other students in Mathematics and Portuguese. The size of those effects is higher in Mathematics than in Portuguese.

Fixed–effect results confirm that the signal of the coefficients of class size and its square are positive and negative, respectively, indicating that the effect of class size on students' performance is convex. Thus, there are increases on students' achievement caused by increasing the number of students in class, probably due to spillover effects, but they are subject to diminishing returns, reaching a peak at 24 students per class, when considering all students (column 2). Such class size effect is higher in Mathematics than in Portuguese results (as in, Rivkin et al., 2005). That is, an additional student

Table 6.3: Fixed–effects estimates of teacher characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
	All students		**Mathematics**		**Portuguese**	
Log of 9^{th} exam scores	0.7067***	0.7066***	0.8347***	0.8341***	0.6466***	0.6463***
	(0.0149)	(0.0148)	(0.0297)	(0.0296)	(0.0151)	(0.0151)
Female student	0.0544***	0.0544***	0.0661***	0.0661***	0.0607***	0.0607***
	(0.0055)	(0.0055)	(0.0126)	(0.0126)	(0.0054)	(0.0054)
Age	-0.1112***	-0.1105***	-0.1983***	-0.1963***	-0.0932***	-0.0925***
	(0.0087)	(0.0088)	(0.0289)	(0.0289)	(0.0078)	(0.0078)
Beneficiary S.S.	-0.0443***	-0.0442***	-0.0706***	-0.0714***	-0.0327***	-0.0325***
	(0.0072)	(0.0072)	(0.0188)	(0.0187)	(0.0064)	(0.0064)
Internet	0.0141*	0.0142*	0.0017	0.0024	0.0226***	0.0225***
	(0.0080)	(0.0080)	(0.0191)	(0.0190)	(0.0073)	(0.0073)
Advanced degree	-0.0223	-0.0218	-0.0957	-0.0913	0.0505	0.0501
	(0.0507)	(0.0505)	(0.1136)	(0.1107)	(0.0411)	(0.0411)
Experience	0.0565***	0.0567***	0.2238***	0.2195***	0.0076	0.0084
	(0.0211)	(0.0209)	(0.0492)	(0.0486)	(0.0179)	(0.0179)
Experience sq	-0.0018***	-0.0018***	-0.0065***	-0.0063***	-0.0003	-0.0003
	(0.0004)	(0.0004)	(0.0010)	(0.0010)	(0.0004)	(0.0004)
Commuting	-0.0216**	-0.0215**	-0.0136	-0.0136	-0.0214**	-0.0212**
	(0.0100)	(0.0100)	(0.0255)	(0.0254)	(0.0090)	(0.0090)
Class size		0.0142**		0.0315*		0.0103**
		(0.0060)		(0.0169)		(0.0051)
Class size sq		-0.0003**		-0.0007**		-0.0002*
		(0.0001)		(0.0003)		(0.0001)
Observations	21,549	21,549	8,100	8,100	13,449	13,449
No. of teachers	4,817	4,817	2,868	2,868	3,828	3,828
σ_u	0.334	0.333	0.7885	0.775	0.192	0.192
σ_e	0.341	0.341	0.453	0.453	0.243	0.243
ρ	0.490	0.488	0.750	0.746	0.384	0.386

Source: Computations of the author based on *MISI* and *JNE Statistics*.
Note: Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is log 12^{th} grade national exam score. All regressions include a set of dummies to control for district/region.

in a Mathematics class with 15 students leads to an increase achievement of about 1.1% and in a class with 20 students the corresponding return is about 0.35%.[7] For Mathematics, the optimal size class is 22 students. The corresponding benefits for Portuguese subject are 0.43% and 0.23%, reaching a peak at 26 students per class.

Regarding the academic education of teachers, those with more qualifications (post-graduate, masters or PhDs) do no better or worse compared to teachers with the minimum qualification to teach. In addition, the previous finding that Portuguese teachers with an advanced degree have a positive influence on achievement is not confirmed, since in fixed–effects results the teacher education level has no statistical significance.

Commuting teachers have a negative and significant effect on students achievement,

[7]The computation is given by $0.0035 = 0.0315 - 0.0007 \times 2 \times 20$.

i.e. students of those teachers working away from home get exam results 2% lower than those taught by teachers who work close to home. Considering the two subjects separately, being away from home does not have any impact on student achievement in Mathematics; as in previous set of results, it is found that commuting teachers negatively influence the student achievement gains in the Portuguese exam, but this effect is small.

The positive coefficient of experience and the negative coefficient of experience squared indicate a non–linear relationship with student achievement, so this result is according to the discussion previously conducted. The negative sign of the coefficient of the variable experience squared, reveal that the effect of experience on students' achievement is convex, indicating that experience increases student results and it shows diminishing returns, reaching a peak at about 16 years of experience. Considering all sample, the results in column 2 in Table 6.3 show that estimated coefficients for experience and its square are 0.0567 and -0.0018, respectively, indicating that the return to the first year of experience is about 6%; after 10 years of experience, approximately average experience in the sample, the return to an additional year of experience is still around 2.1%.[8]

Analysing separately the teachers of Mathematics and Portuguese, it is found that experience of Portuguese teachers is not statistically significant, but teaching experience has a positive and significant impact on Mathematics achievement. The large estimate one obtains for the parameter on linear experience deserves additional discussion. First, estimating a fixed-effects model on professors, where experience is only reported for three periods, diminishes, to some extent, our ability to identify with precision this parameter. Second, the estimate of the parameter on the second order polynomial on experience indicates the presence of marginal diminishing returns to experience. Finally, in our sample experience varies between 5 and 40 years. While at five years of experience an additional year of experience raises grades by about 15%, such marginal effect decreases until it becomes null at 15 years of experience. This marginal effect becomes negative after 25 years of experience. This result indicates that students benefit significantly from an increase in the experience of younger professors, but such benefit decreases over time.

This result compares with the literature, there is reported evidence that teachers with more experience are more effective in increasing student achievement gains than those with less experience (as in, Clotfelter et al., 2006, 2007).

[8]The computation is given by $0.0207 = 0.0567 - 0.0018 \times 2 \times 10$.

6.5 Concluding remarks

Teachers have a key role in the teaching–learning process, and consequently they are a central issue in political discussions about the quality of education and schools. The role of the teacher as a major determinant of school quality, namely in basic and secondary education, has been emphasized. Nevertheless, little is known on which teacher characteristics contribute the most to improve the process. In fact, several teacher characteristics can make a difference in the teaching–learning process, however, most of those are difficult to measure, as they are unobserved. The purpose of this chapter was to analyse the impact of teachers' observable characteristics on student results in two subjects, Mathematics and Portuguese, which are correlated with teaching effectiveness.

Although teacher experience and education are the most discussed observable characteristics in the literature, in this chapter other characteristics have also been considered, such as gender and motivation. Controlling for students' characteristics, the starting point is the OLS estimation and, then, to control the heterogeneity of teachers the fixed–effects estimation is applied. The results are based on a matched student–teacher panel data relating to the Portuguese public schools, for the period 2010–2012, using a value–added approach.

The main result found in this study is that teacher quality is important for student achievement, i.e., empirical results support the idea that raising teacher quality may be a key instrument in improving student outcomes. In this sense, it is found that 49% of the variance in students' achievement is due to differences across teachers. When looking at the teacher attributes that contribute for such quality differences, the results show that female teachers have better performance on student achievement gains than males, in both subjects. Teachers working away from home have negative and significant effects on student results. This issue should be taken into consideration when designing the teacher allocation process.

Another result relates to the level of education the teacher attained; holding advanced degree diplomas seem to have no relationship to teacher quality as a measured by student achievement gains, i.e., teachers with more qualifications (postgraduate, masters or PhDs) do not show better performance than those with a *bacharelato* or licentiate diploma. This result does not confirm some of the existing literature, as some authors suggest that teacher qualifications have a significant effect, particularly in Mathematics (Clotfelter et al., 2006, 2007).

Teachers with more experience are more effective in increasing student performance than those with less experience. Controlling for the teacher fixed–effects and considering all students and teachers, the results show that an extra year of experience leads

to a return to the first year of experience of about 6% and after 10 years of experience and the return to an additional year of experience is about 2.1%. They also indicates that these increases are subject to diminishing returns, reaching a peak at about 16 years of teaching experience. Note that, it is found that teaching experience is not statistically significant when it comes to the results in the Portuguese exam.

Apart from teacher specific characteristics, the study also controls for class size and some student attributes. Results show that there are increases on students' achievement caused by increasing the number of students in class, probably due to spillover effects, but they are subject to diminishing returns. This effect is significant only in Mathematics classes. As expected, the prior achievement of students is the strongest predictor of performance in their current academic performance. Furthermore, female students perform better than male students in both subjects and, older students and disadvantaged students perform worse than other students in Mathematics and Portuguese. The size of these effects is higher in Mathematics discipline than in Portuguese discipline.

This study is a first attempt to analyse the impact of teacher characteristics on student performance. There are several ways in which it can be extended. Gender role models, for instance, explore the effect that the teacher's gender may have on student achievement gains. In this context, it has been empirically tested the hypothesis that the same–sex teacher may improve student outcomes. Some contribution to this literature could inform school administrators when assigning teachers to classes.

Chapter 7

Conclusions

This dissertation has been organized in four independent essays on Economics of Education. The first essay, Chapter 3, aimed at analysing private returns to education in Portugal. According to the human capital literature, people who invest more in education and accumulate more human capital, enhance their own productivity, and receive higher earnings throughout their working lives. Estimates of private rates of returns to education in Portugal, for the period 1986–2009, indicate that, by controlling for observable characteristics, as well as for unobservable individual characteristics, such as ability or motivation, the return resulting from an additional year of schooling is about 7%, on average. Starting in 2002, returns to education are higher for females than for male workers. There are differences across Portuguese districts; those working on the coastal districts benefit from higher returns than those working in inland districts of Portugal.

Nevertheless, returns to education go beyond the individual returns. Each individual's educational choice affects his own individual economic returns, and it impacts on the other workers, i.e., there are educational externalities as well. The second essay, Chapter 4, has looked at education spillovers. It provides evidence that investments in education lead to positive spillovers, that is, an additional year in the average education of the county increases individual wages by about 3%, on average. This is corroborated by other indicators of aggregate human capital. An increase in the share of the most qualified workers in the county by one percentage point raises wages by about 0.2%–0.3%. The existence of spillover effects is confirmed by the coefficients on the proposed county skill index as well. Using this measure, the estimates show that an additional year of calibrated education in the county leads to a wage premium of about 1.2%–1.8%. When the human capital at the firm level is included, the spillover effect decreases but remains positive and significant, suggesting that an additional year in the average education of the county and of the firm causes wages to increase by about

1.5% and 4%, respectively. Furthermore, the wages of the workers rise by about 0.3% when the share of the most skilled workers in a firm increases by one percentage point, while an increase by one percentage point in the share of the most skilled workers in the county leads to an increase in returns of about 0.2%. These findings seem to suggest that human capital spillovers are stronger within the firm, as productive social interactions are probably more intense among colleagues than among the people in a county.

There appear to be neighbouring effects of human capital as well. An increase by one percentage point in the share of the most skilled workers in the neighbouring counties leads to an increase of about 0.1% on worker wages. In addition, both the least and the most qualified workers benefit from these external returns to human capital at the three aggregation levels, but the most qualified workers receive higher private and external returns to education. Therefore, individual wages of Portuguese workers are determined by the individual own education (private returns to education), but also by the knowledge gained by interacting with other workers (external returns to human capital), i.e., by firm, regional and inter–regional externalities.

Those rates of returns emphasize the importance of investments in education, as a process of accumulation of human capital, as they cause *per capita* income to increase and are a key factor for the aggregate performance of the economy and, consequently, for the economic growth, as well. The existence of educational externalities on top of the substantial private returns to education provides justification for significant public support and funding for education. Such public support leads to a greater equality in education, between poor and rich families, and hence more equal individual wages. Thereafter, increasing wage equality tends to increase social cohesion, providing a further efficiency argument that supports public funding for education.

Earning inequalities may, however, be related not only with years of schooling, but also with school quality. Furthermore, differences in learning achievements are better predictors of cross–countries differences in economic development and economic growth. Several factors may determine variations in student performance, namely, the characteristics of students and families, peer effects, school and teacher characteristics. These facts gave reason to the third and fourth essays.

The starting point for the third essay, Chapter 5, was to identify the factors that influence student achievement gains and, additionally, to estimate the school added–value. It seems that there is evidence that student achievement gains depend on a whole range of factors including student's and family characteristics, and school related aspects such as class size. Student characteristics as well as his/her family features are the main factors influencing student achievement gains.

Student's prior achievement is the strongest predictor of current academic performance in both subjects, Mathematics and Portuguese. Female students perform better than male students of about 5% – 6%, on average, but this effect is different between the two subjects analysed, i.e., the girls performance is higher than that of the boys by about 6% – 9.9% in Mathematics and by about 5% – 7.6% in Portuguese. Student age has a negative and significant influence of about 11% – 12% on achievement, on average, which probably reflects a grade repetition effect. Note that, this negative effect has twice the size of the effect in Mathematics when compared to the Portuguese discipline. Students who benefit from social support performance worse compared to other students, mainly in Mathematics. Family background, in particular, the education of the parents or the legal–guardian has a positive influence on student performance, but their employment status does not show up relevant. Since family background and the characteristics of the students are the factors with the greatest influence on student performance, therefore, more attention should be given to the role of external factors to the school that affect the level of school performance.

The performance of a student beneficiary of social support is lower by about 4%–5%, on average, but this effect is higher in Mathematics achievement than in Portuguese achievement, about 5% – 7% and 3%, respectively. The parents'/legal–guardians' education effect is higher in the Mathematics achievement, i.e., students whose parents/legal–guardians have higher education perform better in about 8% – 12% and 5% – 7% in Mathematics and Portuguese, respectively, compared to students whose parents/legal–guardians have only the 3^{rd} cycle of schooling. The corresponding values for secondary education are of about 3% – 4% and 2% – 3% in Mathematics and Portuguese achievement, respectively.

In most cases the school–effect is null. There is more variability in student performance between students within classes rather than between classes within the same school, or even between schools, which confirms that the small influence of the school on student performance. Nevertheless, class size appears to matter for student performance. There is evidence of increases on student achievement by increasing the number of students in a class, probably due to the spillover effects, but they are subject to diminishing returns, reaching a peak when there are 23–25 students per class. This effect is more relevant in Mathematics classes.

Class size provides a mechanism for improving student outcomes. The implicit assumption that smaller classes are better in general is based in the fact that a teacher allocates more time per student when there are fewer students or they may find it easier to exercise discipline and, this will, in turn, most probably lead to higher student achievement. According to the current Portuguese legislation, in the scientific–

humanistic courses at the secondary level of education, the minimum number of students for opening a class is 26 students with a maximum of 30 students[1], which are above the optimal number of students per class of 23–25 suggested above. Reducing class size is an educational policy that is popular among students, parents, teachers, and policy–makers, and hence easy to argue for.

The results on performance of Portuguese schools of the multilevel model were compared to those resulting from the DEA approach in an added value perspective. From the performance indicators that were built and the resulting efficiency rankings it emerges that although there is some similarity between efficiency measures provided by both methodologies, the methodology of analyse of the school performance should be carefully chosen and justified as, in principle, conclusions on efficiency and, consequently, the efficiency rankings depend on that choice.

In fact, many factors contribute to a student's academic performance, including individual characteristics and family and school experiences. But research suggests that, among school–related factors, teachers matter most. This observation set the stage for the fourth essay, Chapter 6, which aims at looking to what extent teachers' quality determines student progress. In this study the observable characteristics of the teachers that influence student achievement were analysed in a value added perspective. Increasing teacher quality may be a key instrument for improving student's outcomes. It is found that 49% of the variance in student achievement gains is due to differences across teachers.

Teachers with more experience are more effective in promoting student achievement gains than those with less experience, but these increases are subject to diminishing returns, reaching a peak at about 16 years of teaching experience. It has a higher impact on Mathematics than on Portuguese achievement. The idea behind it is that the experience gained over time promotes effectiveness, enhances knowledge, skills and productivity of the teachers, but that effectiveness declines after some point.

There is a variety of observable factors that influence the performance of students and that these effects are higher in Mathematics achievement. The impact of experience is stronger than the effect of other observable teacher characteristics. Female teachers impact more on student achievement gains than male teachers, in both subjects. The teacher academic education, namely, advanced degrees, seem to have no relationship to teacher quality as measured by student achievement gains. Finally, having teachers working away from home negatively influences achievement. Furthermore, according to the results presented, it seems that this is an important issue for policy relating to the placing of teachers, since there seems to be a large number of teachers far from

[1]Legislative Order n.º 7–B/2015.

their home and families.

From a policy point of view, the results provided in this last part of the thesis should lead to relevant reflections and implications, in particular, with regard to educational policies that focus on tackling inequality in academic performance, since the equality in education, on the future, will be reflected through greater economic and social equality. From the results, it appears that family background and the characteristics of the students are the factors with the greatest influence on student performance, therefore, more attention should be given to the role of external factors to the school that affect the level of school performance. In this sense, based on the social, cultural and economic differences of Portuguese students who attend public schools, it is necessary to promote a family–school relationship.

Currently in Portugal, the Ministry of Education has a policy of incentives to schools that consists of assigning additional credit–hours, mainly based on external student outcomes.[2] However, it also seems important to provide additional resources to schools that are inserted in more disadvantaged socio–economic contexts, enabling them to develop specific initiatives/programmes of prevention and intervention tailored to the needs of their students, in order to improve student achievement, as well as to combat school drop–out. These programmes should be focused on students who benefit from social support, students whose parents have low education and students with retentions in the previous school years. Additionally, it is extremely important to implement projects that sensitize parents to be more involved in the school life of their children and to allow them to be aware of the size of the social and economic returns to education.

In parallel with the discussion on reducing class size, there is a need for a discussion on teacher quality. Teachers differ in their performance with consequences for students. Thus, discussing educational policies that ignore this issue is belittling the main actors of an educational system, i.e., teacher effectiveness has important implications for the development of policies designed to raise the quality of instruction and to reduce unequal access to high quality teachers. The results presented in this thesis show that teachers with more experience are more effective in increasing student achievement gains, but their performance declines after a certain period of experience. The decline in the performance of teachers at the end of their career may be explained by the fact that teachers are exhausted/tired or probably because experienced teachers may not be updated on curricular and pedagogical advances. Incentive strategies of teacher professional development should be considered. Nevertheless, the size of the experience effects on student achievement gains is not very large. This can be explained, at least

[2]Legislative Order of Organization of the school year 2015–2016.

partly, by the current pay structure for teachers in Portugal, which is based, mainly, on teaching experience rather than on teacher performance *per se*. More experienced teachers may inherently be better teachers, but they may not apply all their efforts and skills in the classroom, since this system provides too little incentive for the "best" to deliver their best performance in the classroom. However, it is important to provide other incentives (apart from the salary) to teachers so that they are motivated and empowered.

Improving the state of education in Portugal can probably be based on the one hand, by giving incentives to parents and, secondly, providing incentives to teachers/schools to promote student success. Nevertheless, more research is needed to address the unanswered questions raised in this thesis. So, for future research, it would be important to further investigate why female students/teachers perform better than male students/teachers. Furthermore, it would certainly contribute to a better understanding of student and school performance if special attention is devoted to other factors that may influence the quality of education in Portugal. Issues such as teacher and student absences, the school teaching body stability, the characteristics of the "head–teacher" of the class, as well as the extension of this analysis to other years of schooling should be explored in future research. The use of school drop–out rates as outcome would certainly bring interesting insights to the research.

References

Aaronson, D., Barrow, L., and Sander, W. (2007). "Teachers and student achievement in the Chicago public high schools." *Journal of Labor Economics*, 25(1), 95–135.

Abdulkadiroglu, A., Angrist, J., Dynarski, S., Kane, T., and Pathak, P. (2011). "Accountability and flexibility in public schools: Evidence from Boston's charters and pilots." *The Quarterly Journal of Economics*, 126(2), 699–748.

Acemoglu, D. and Angrist, J. (2001). "How large are the social returns to education: Evidence from compulsory schooling laws." *NBER Macroeconomics Annual 2000, Volume 15*, Bernanke and Rogoff, eds.

Agasist, T. and Vittadini, G. (2012). "Regional economic disparities as determinants of students' achievement in Italy." *Research in Applied Economics*, 4(1), 33–54.

Akerhielm, K. (1995). "Does class size matter?" *Economics of Education Review*, 14(3), 229–241.

Almeida, V., Curado, A. P., Sarrico, C., Nunes, J., Azevedo, J. J., Azevedo, J. M., Duarte, M. L., and Teixeira, P. N. (2011). *Propostas para um novo ciclo de avaliação externa de escolas - Relatório Final*. Grupo de Trabalho para a Avaliação Externa das Escolas. Ministério da Educação, Lisboa.

Altonji, J. and Shakotko, R. (1987). "Do wages rise with job seniority?" *Review of Economic Studies*, 54(3), 437–459.

Alves, N., Centeno, M., and Novo, A. (2010). "O investimento em educação em Portugal: retornos e heterogeneidade." *Boletim Económico Primavera 2010, Departamento de Estudos Económicos, Banco de Portugal*, 16(1), 1–129.

Angrist, J. (1995). "The economic returns to schooling in the West Bank and Gaza Strip." *American Economic Review*, 85(5), 1065–1087.

Angrist, J. and Krueger, A. (1991). "Does compulsory school attendance affect schooling and earnings?" *Quarterly Journal of Economics*, 106(4), 979–1014.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, Vol. 4. Springer, Netherlands.

Anselin, L. (2003). "Spatial econometrics." *International Regional Science Review*, 26(2), 7–152.

Anselin, L., Gallo, J., and Jayet, H. (2008). *Spatial panel econometrics*. Springer–Verlag, Berlin.

Arbia, G. and Baltagi, B. (2009). *Spatial Econometrics: Methods and Applications Studies in Empirical Economics*. Physica–Verlag A Springer Company, Heidelberg.

Ashenfelter, O., Harmon, C., and Oosterbeek, H. (1999). "A review of estimates of the schooling/earnings relationship, with tests for publication bias." *Labour Economics*, 6(4), 453–470.

Azevedo, J. (2005). *Avaliação das escolas: fundamentar modelos e operacionalizar processos*. Comissão Nacional de Educação, Lisboa.

Bakis, O., Davutyan, N., Levent, H., and Polat, S. (2010). "External returns to higher education in Turkey." *Report No. 517*, Economic Research Forum.

Baltagi, B. (2008). *Econometric Analysis of Panel Data*. John Wiley and Sons Ltd, Chichester, 4th edition.

Baltagi, B., Bresson, G., and Pirotte, A. (2003). "Fixed effects, random effects or Hausman–Taylor? A pretest estimator." *Economics Letters*, 79(3), 361–369.

Bauer, T. and Vorell, M. (2010). "External effects of education: Human capital spillovers in regions and firms." *Report No. 195*, Ruhr Economic Papers.

Baum, C., Schaffer, M., and Stillman, S. (2003). "Instrumental variables and GMM: Estimation and testing." *Stata Journal*, 3(1), 1–31.

Baum, C., Schaffer, M., and Stillman, S. (2007). "Enhanced routines for instrumental variables/GMM estimation and testing." *Stata Journal*, 7(4), 465–506.

Becker, G. (1962). "Investment in human capital: A theoretical analysis." *Journal of Political Economy*, 70(5), 9–49.

Benabou, R., Kramarz, F., and Prost, C. (2009). "The French zones d'education prioritaire: Much ado about nothing?" *Economics of Education Review*, 28(3), 345–356.

Berlinski, S., Galiani, S., and Gertler, P. (2009). "The effect of pre–primary education on primary school performance." *Journal of Public Economics*, 93(1), 219–234.

Berlinski, S., Galiani, S., and Manacorda, M. (2008). "Giving children a better start: Preschool attendance and school–age profiles." *Journal of Public Economics*, 92(5), 1416–1440.

Black, S., Devereux, P., and Salvanes, W. (2005). "Why the apple doesn' t fall far: understanding intergenerational transmission of human capital." *American Economic Review*, 95(1), 437–449.

Blackburn, M. and Neumark, D. (1993). "Omitted–ability bias and the increase in the return to schooling." *Journal of Labor Economics*, 11(3), 521–544.

Blondäl, S., Field, S., and Giroard, N. (2002). "Investment in human capital through upper second and tertiary education." *OECD Economic Studies*, 2002(1), 41–89.

Boarini, R. and Strauss, H. (2010). "What is the private return to tertiary education? New evidence from 21 OECD countries." *OECD Journal: Economic Studies*, 2010, 7–31.

Bound, J., Jaeger, D., and Baker, R. (1995). "Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak." *Journal of the American Statistical Association*, 90(430), 443–450.

Brunello, G. and Miniaci, R. (1999). "The economic return to schooling for Italian men. An evaluation based on instrumental variables." *Labour Economics*, 6(4), 509–519.

Brunello, G. and Rocco, L. (2013). "The effect of immigration on the school performance of natives: Cross country evidence using PISA test scores." *Economics of Education Review*, 32(C), 234–246.

Buddin, R. and Zamarro, G. (2009). "Teacher qualifications and student achievement in urban elementary schools." *Journal of Urban Economics*, 66(2), 103–115.

Budria, S. and Nunes, C. (2005). "Education and wage inequality in Portugal." *Education and Wage Inequality in Europe: A Literature Review*, R. Asplund and E. Barth, eds.

Budria, S. and Pereira, P. (2011). "Educational qualifications and wage inequality: evidence for Europe." *Revista de Economia Aplicada*, XIX(56), 5–34.

Buhai, I. S., Portela, M., Teulings, C., and van Vuuren, A. (2014). "Returns to tenure or seniority?" *Econometrica*, 82(2), 705–730.

Canton, E. (2009). "Human capital externalities and proximity: evidence from repeated cross–sectional data." *De Economist*, 157(1), 79–105.

Card, D. (1995a). "Earnings, schooling, and ability revisited." *Research in Labor Economics*, S. Polachek, ed., Vol. 14, JAI Press, Greenwich Connecticut.

Card, D. (1995b). "Using geographic variation in college proximity to estimate the return to schooling." *Aspects of Labour Market Behaviour: Essays in Honour of JohnVanderkamp*, E. K. G. Louis Christofides and R. Swindinsky, eds.

Card, D. (1999). "The causal effect of education on earnings." *Handbook of Labor Economics*, O. C. Ashenfelter and D. Card, eds., Vol. 3, Part A.

Card, D. (2001). "Estimating the return to schooling: Progress on some persistent econometric problems." *Econometrica*, 69(5), 1127–1160.

Card, D., Devicienti, F., and Maida, A. (2014). "Rent–sharing, hold–up, and wages: Evidence from matched panel data." *The Review of Economic Studies*, 81(1), 84–111.

Cardoso, A. (1991). "Desigualdades salariais em Portugal: uma perspectiva regional, 1982–1989." M.S. thesis, Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa.

Carneiro, P. (2008). "Equality of opportunity and educational achievement in Portugal." *Portuguese Economic Journal*, 7(1), 17–41.

Carneiro, P. and Heckman, J. (2002). "The evidence on credit constraints in postsecondary schooling." *Economic Journal*, 112(482), 705–734.

Carneiro, P., Heckman, J., and Vytlacil, E. (2011). "Estimating marginal returns to education." *American Economic Review*, 101(6), 2754–2781.

Carrell, S. and West, J. (2010). "Does professor quality matter? Evidence from random assignment of students to professors." *Journal of Political Economy*, 118(3), 409–432.

Charnes, A. and Cooper, W. (1985). "Preface to topics in data envelopment analysis." *Annals of Operations Research*, 2(1), 59–94.

Charnes, A., Cooper, W., and Rhodes, E. (1978). "Measuring the efficiency of decision making units." *European Journal of Operational Research*, 2(6), 429–444.

Chiang, H. (2009). "How accountability pressure on failing schools affects student achievement." *Journal of Public Economics*, 93(9), 1045–1057.

Ciccone, A. and Peri, G. (2006). "Identifying human–capital externalities: Theory with applications." *Review of Economic Studies*, 73(2), 381–412.

Clotfelter, C., Ladd, H., and Vigdor, J. (2006). "Teacher–student matching and the assessment of teacher effectiveness." *The Journal of Human Resources*, 41(4), 778–820.

Clotfelter, C., Ladd, H., and Vigdor, J. (2007). "Teacher credentials and student achievement: Longitudinal analysis with student fixed effects." *Economics of Education Review*, 26(6), 673–682.

Clotfelter, C., Ladd, H., and Vigdor, J. (2010). "Teacher credentials and student achievement in high school: A cross–subject analysis with student fixed effects." *The Journal of Human Resources*, 45(3), 655–681.

Coelho, I., Sarrico, C., and Rosa, M. (2008). "Avaliação de escolas em Portugal: Que futuro?" *Revista Portuguesa e Brasileira de Gestão*, 7(2), 56–67.

Coelli, T., Rao, D., O'Donnell, C., and Battese, G. (2005). *An Introduction to Efficiency and Productivity Analysis*. Springer, New York, 2nd edition.

Coleman, J. (1966). *Equality of Educational Oportunity*. National Center For Eduactional Statistics, Washington D.C.

Combes, P., Duranton, G., and Gobillon, L. (2011). "The identification of agglomeration economies." *Journal of Economic Geography*, 11(2), 253–266.

Cook, W. and Seiford, L. (2009). "Data envelopment analysis (DEA) – Thirty years on." *European Journal of Operational Research*, 192(1), 1–17.

Cook, W., Tone, K., and Zhu, J. (2014). "Data envelopment analysis: Prior to choosing a model." *Omega*, 44(1), 1–4.

Cooper, W., Seiford, L., and Tone, K. (2006). *Introduction to Data Envelopment Analysis and Its Uses: With DEA–Solver Software and References*. Springer Science & Business Media.

Cooper, W., Seiford, L., Tone, K., and Zhu, J. (2007). "Some models and measures for evaluating performances with DEA: past accomplishments and future prospects." *Journal of Productivity Analysis*, 28(3), 151–163.

Correia, S. (2014). "REGHDFE: Stata module to perform linear or instrumental–variable regression absorbing any number of high–dimensional fixed effects." *Statistical Software Components*, Boston College Department of Economics.

Coupé, T., Olefir, A., and Alonso, J. (2016). "Class size, school size and the size of the school network." *Education Economics*, 24(3), 329–351.

Croninger, R., Rice, J., Rathbun, A., and Nishio, M. (2007). "Teacher qualifications and early learning: Effects of certification, degree, and experience on first–grade student achievement." *Economics of Education Review*, 26(3), 312–324.

Cullen, J., Jacob, B., and Levitt, S. (2005). "The impact of school choice on student outcomes: an analysis of the Chicago public schools." *Journal of Public Economics*, 89(5–6), 729–760.

Czaller, L. (2016). "Increasing social returns to human capital: Evidence from Hungarian regions." *Regional Studies*, DOI: 10.1080/00343404.2015.1112898.

Dalmazzo, A. and de Blasio, G. (2007). "Social returns to education in Italian local labor markets." *Annals of Regional Science*, 41(1), 51–69.

Davidson, R. and MacKinnon, J. (1999). *Econometric Theory and Methods*. Oxford University Press, New York.

Denny, K. and Oppedisano, V. (2013). "The surprising effect of larger class sizes: Evidence using two identification strategies." *Labour Economics*, 23(C), 57–65.

Deutsch, J., Dumas, A., and Silber, J. (2013). "Estimating an educational production function for five countries of Latin America on the basis of the PISA data." *Economics of Education Review*, 36(C), 245–262.

Devereux, P. and Fan, W. (2011). "Earnings returns to the British education expansion." *Economics of Education Review*, 30(6), 1153–1166.

Devereux, P. and Hart, R. (2010). "Forced to be rich? Returns to compulsory schooling in Britain." *Economic Journal*, 120 (549), 1345–1364.

Dias, V. and Ferrão, M. (2006). "Modelo multinível do desempenho escolar de alunos socialmente desfavorecidos em escolas públicas/privadas – aplicação aos dados portugueses do PISA 2000." *Psicologia e Educação*, V(2), 63–77.

Dias, V. D., Navio, V. M., and Ferrão, M. (2008). "3EM – Eficácia escolar no ensino da Matemática, um projecto de investigação." *REICE–Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 6(1), 42–57.

Dickson, M. (2012). "The causal effect of education on wages revisited." *Oxford Bulletin of Economics and Statistics*, 75(4), 477–498.

Dickson, M. and Smith, S. (2011). "What determines the return to education: an extra year or hurdle cleared?" *Economics of Education Review*, 30(6), 1167–1176.

Drukker, D. (2003). "Testing for serial correlation in linear panel–data models." *The Stata Journal*, 3(2), 1–10.

Dumay, X., Coe, R., and Anumendem, D. (2014). "Stability over time of different methods of estimating school performance." *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 25(1), 64–82.

Duranton, G. (2007). "Human capital externalities in cities: identification and policy issues." *A Companion to Urban Economics*, Arnott and McMillen, eds., Blackwell Publishing Ltd.

Egalite, A., Kisida, B., and Winters., M. (2015). "Representation in the classroom: The effect of own–race teachers on student achievement." *Economics of Education Review*, 45, 44–52.

Farrel, M. (1957). "The measurement of productive efficiency." *Journal of the Royal Statistical Society. SeriesA (General)*, 120(3), 253–290.

Feinstein, L. and Symons, J. (1999). "Attainment in secondary school." *Oxford Economic Papers*, 51(2), 300–321.

Ferrão, M. (2012). "On the stability of value added indicators." *Quality & Quantity*, 46(2), 627–637.

Ferrão, M., Beltrão, K., and Santos, D. (2002). "Políticas de não–repetência e a qualidade da educação: evidências obtidas a partir da modelagem dos dados da 4ª série do SAEB–99." *Estudos em avaliação educacional*, 26, 47–73.

Ferrão, M. and Couto, A. (2014). "The use of a school value–added model for educational improvement: a case study from the Portuguese primary education system." *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 25(1), 174–190.

Ferrão, M. E. (2003). *Introdução aos modelos de regressão multinível em educação*. Komedi, Campinas.

Fertig, M. and Wright, R. (2005). "School quality, educational attainment and aggregation bias." *Economics Letters*, 88(1), 109–114.

Figlio, D. (1997). "Teacher salaries and teacher quality." *Economics Letters*, 55(2), 267–271.

Garcia-Fontes, W. and Hidalgo, M. (2009). "Estimating human capital externalities: The case of spanish regions." *Report No. 09.17*, Ruhr Economic Papers.

Garcia-Mainar, I. and Montuenga-Gómez, V. (2005). "Education returns of wage earners and self-employed workers: Portugal vs. Spain." *Economics of Education Review*, 24(2), 161–170.

Goldhaber, D. and Hansen, M. (2013). "Is it just a bad class? Assessing the long–term stability of estimated teacher performance." *Economica*, 80(319), 589–612.

Goldstein, H. (1995). *Multilevel Statistical Models*. Edward Arnold, London.

Goldstein, H. (1997). "Methods in school effectiveness research." *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 8(4), 369–395.

Goldstein, H. (2011). *Multilevel statistical models*, Vol. 922. John Wiley and Sons Ltd, United Kindgom, 4th edition.

González, P., Santos, M., and Santos, L. (2007). "Education and gender wage differentials in Portugal: What can we learn from an age cohort analysis?" *DP 2007- 01, CETE: Centro de Estudos de Economia Industrial, do Trabalho e da Empresa*.

Grácio, S. (1998). *Ensinos Técnicos e Política em Portugal 1910/1990*. Instituto Piaget, Lisboa, Chapter 11 and 12, 175–202.

Greene, W. (2012). *Econometric Analysis*. Pearson Education, Boston, 7th edition.

Grenet, J. (2013). "Is it enough to increase compulsory education to raise earnings? Evidence from French and British compulsory schooling laws." *Scandinavian Journal of Economics*, 115(1), 176–210.

Grosskopf, S., Hayes, K., and Taylor, L. (2014). "Applied efficiency analysis in education." *Economics and Business Letters*, 3(1), 19–26.

Grossman, M. (2006). "Education and nonmarket outcomes?" *Handbook of the Economics of Education*, M. Grossman, ed., Vol. 1.

Guarino, C., Reckase, M., Stacy, B., and Wooldridge, J. (2015). "A comparison of growth percentile and value–added models of teacher performance." *Statistics and Public Policy*, 2(1), 1–11.

Guimarães, P. and Portugal, P. (2010). "A simple feasible procedure to fit models with high–dimensional fixed effects." *Stata Journal*, 10(4), 628–649.

Hanushek, E. (1986). "The economics of schooling: Production and efficiency in public schools." *Journal of Economic Literature*, 24(3), 1141–1177.

Hanushek, E. (1989). "The impact of differential expenditures on school performance." *Educational Researcher*, 18(4), 45–62.

Hanushek, E. (1997). "Assessing the effects of school resources on student performance: An update." *Educational evaluation and policy analysis*, 19(2), 141–164.

Hanushek, E. (2005a). "The economics of school quality." *German Economic Review*, 6(3), 269–286.

Hanushek, E. (2005b). "Why quality matters in education." *Finance and Development*, 42(2), 15–19.

Hanushek, E. (2011). "The economic value of higher teacher quality." *Economics of Education Review*, 30(3), 466–479.

Hanushek, E. (2012). *The 4 percent solution: Unleashing the economic growth America needs*. Crown Business, New York.

Hanushek, E., Jamison, D., Jamison, E., and Woessmann, L. (2008). "Education and economic growth: It's not just going to school but learning that matters." *Eduction Next*, 8(2), 62–70.

Hanushek, E., Kain, J., Markman, J., and Rivkin, S. (2003). "Does peer ability affect student achievement?" *Journal of Applied Econometrics*, 18(5), 527–544.

Hanushek, E. and Rivkin, S. (2006). "Teacher quality." *Handbook of the Economics of Education*, E. Hanushek and F. Welch, eds., Vol. 2.

Hanushek, E. and Rivkin, S. (2010). "Generalizations about using value–added measures of teacher quality." *American Economic Review*, 100(2), 267–271.

Hanushek, E., Rivkin, S., and Taylor, L. (1996). "Aggregation and the estimated effects of school resources." *The Review of Economics and Statistics*, 78(4), 611–627.

Hanushek, E. and Woessman, L. (2010a). "The economics of international differences in educational achievement." *Report No. 15949*, National Bureau of Economic Research.

Hanushek, E. and Woessman, L. (2010b). *The High Cost of Low Educational Performance: The Long–Run Economic Impact of Improving PISA Outcomes*. OECD, Paris.

Harmon, C., Oosterbeek, H., and Walker, I. (2003). "The returns to education: microeconomics." *Journal of Economic Surveys*, 17(2), 115–155.

Harmon, C. and Walker, I. (1995). "Estimates of the economic return to schooling for the United Kingdom." *American Economic Review*, 85(5), 1278–1286.

Harmon, C., Walker, I., and Westergaard-Nielsen, N. (2001). *Education and earnings in Europe: A cross country analysis of the returns to education*. Edward Elgar Publishing, Cheltenham.

Hartog, J. and Oosterbeek, H. (2007). *Human capital : advances in theory and evidence*. Cambridge University Press, Cambridge.

Hartog, J., Pereira, P., and Vieira, J. (2000). "Vocational training and earnings in portugal." *Economia*, XXIV, 35–52.

Hartog, J., Pereira, P., and Vieira, J. (2001). "Changing returns to education in Portugal during the 1980s and early 1990s: OLS and quantile regression estimators." *Applied Economics*, 33(8), 1021–1037.

Haveman, R. and Wolfe, B. (1995). "The determinants of children's attainments: A review of methods and findings." *Journal of Economic Literature*, 33(4), 1829–1878.

Heckman, J. (2000). "Policies to foster human capital." *Research in Economics*, 51(1), 3–56.

Heckman, J. and Carneiro, P. (2003). "Human capital policy." *Report No. 9495*, National Bureau of Economic Research.

Heuermann, D. (2011). "Human capital externalities in Western Germany." *Spatial Economic Analysis*, 6(2), 139–165.

Hox, J. (2010). *Multilevel Analysis: Techniques and Applications, second Edition (Quantitative Methodology Series)*. Routledge, New York.

Hungerford, T. and Solon, G. (1987). "Sheepskin effects in the return to education." *Review of Economic and Statistics*, 69(1), 175–177.

Husain, M. and Millimet, D. (2009). "The mythical "boy crisis"?" *Economics of Education Review*, 28(1), 38–48.

Iranzo, S. and Peri, G. (2009). "Schooling externalities, technology, and productivity: Theory and evidence from U.S. States." *The Review of Economics and Statistics*, 91(2), 420–431.

Isacsson, G. (2005). "External effects of education on earnings: Swedish evidence using matched employee–establishment data." *Report No. 10*, Working Paper, IFAU–Institute for Labour Market Policy Evaluation.

Ji, Y. and Lee, C. (2010). "Data envelopment analysis." *The Stata Journal*, 10(2), 267–280.

Johnes, J. (2006). "Measuring efficiency: A comparison of multilevel modelling and data envelopment analysis in the context of higher education." *Bulletin of Economic research*, 58(2), 75–104.

Kane, T., Rockoff, J., and Staiger, D. (2008). "What does certification tell us about teacher effectiveness? Evidence from New York City." *Economics of Education Review*, 27(6), 615–631.

Kirjavainen, T. (2012). "Efficiency of Finnish general upper secondary schools: an application of stochastic frontier analysis with panel data." *Education Economics*, 20(4), 343–364.

Kirjavainen, T. and Loikkanen, H. (1998). "Efficiency differences of Finnish senior secondary schools: An application of DEA and tobit analysis." *Economics of Education Review*, 17(4), 377–394.

Krassel, K. and Heinesen, E. (2014). "Class–size effects in secondary school." *Education Economics*, 22(4), 412–426.

Krueger, A. (2003). "Economic considerations and class size." *The Economic Journal*, 113(485), F34–F63.

Krueger, A. and Lindahl, M. (2001). "Education for growth: Why and for whom?" *Journal of Economic Literature*, 39(4), 1101–1136.

Kukla-Acevedo, S. (2009). "Do teacher characteristics matter? New results on the effects of teacher preparation on student achievement." *Economics of Education Review*, 28(1), 49–57.

la Fuente, A. D. and Ciccone, A. (2003). *Human capital in a global and knowledge–based economy*. European Communities.

la Fuente, A. D. and Jimeno, J. (2009). "The private and fiscal returns to schooling in the European Union." *Journal of the European Economic Association*, 7(6), 1319–1360.

Ladd, H. and Walsh, R. (2002). "Implementing value–added measures of school effectiveness: getting the incentives right." *Economics of Education Review*, 21(1), 1–17.

Lavy, V. (2015). "What makes an effective teacher? Quasi–experimental evidence." *CESifo Economic Studies*, forthcoming.

Lee, J. and Barro, R. (2001). "Schooling quality in a cross–section of countries." *Economica*, 68(272), 465–488.

Leigh, A. (2010). "Estimating teacher effectiveness from two–year changes in students' test scores." *Economics of Education Review*, 29(3), 480–488.

Leigh, A. and Ryan, C. (2008). "Estimating returns to education using different natural experiment techniques." *Economics of Education Review*, 27(2), 149–160.

LeSage, J. (2008). "An introduction to spatial econometrics." *Revue d'économie industrielle*, 123, 19–44.

LeSage, J. and Pace, R. (2009). *Introduction to spatial econometrics*. Taylor and Francis Group, London.

Levine, J. and Plug, E. (1999). "Instrumenting education and the returns to schooling in the Netherlands." *Labour Economics*, 6(4), 521–534.

Li, H., Liu, P., and Zhang, J. (2012). "Estimating returns to education using twins in urban China." *Journal of Development Economics*, 97(2), 494–504.

Liberto, A. D. (2008). "Education and Italian regional development." *Economics of Education Review*, 27(1), 94–107.

Lima, F. and Pereira, P. (1999). "Wages and human capital: Evidence from the Portuguese data." *Returns to Human Capital in Europe*, ETLA, The Research Institute of the Finnish Economy: Taloustieto Oy.

Liu, Z. (2007). "The external returns to education: Evidence from Chinese cities." *Journal of Urban Economics*, 61(3), 542–564.

Lochner, L. and Moretti, E. (2004). "The effect of education on crime: evidence from prison inmates, arrests and self–reports." *American Economic Review*, 94(1), 155–189.

Longford, N. (1995). "Random coefficient models." *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, Arminger, Clogg, and Sobel, eds.

Lucas, R. J. (1988). "On the mechanics of economic development." *Journal of Monetary Economics*, 22(1), 3–42.

Machado, J. and Mata, J. (2001). "Earning function in Portugal 1982–1994: evidence from quantile regressions." *Empirical Economics*, 26(1), 115–134.

Machado, J. and Mata, J. (2005). "Counterfactual decomposition of changes in wage distributions using quantile regression." *Journal of Applied Econometrics*, 20(4), 445–465.

Mancebón, M. and Bandrés, E. (1999). "Efficiency evaluation in secondary schools: the key role of model specification and of ex post analysis of results." *Education Economics*, 7(2), 131–152.

Martins, P. (2004). "Do foreign firms really pay higher wages? Evidence from different estimators." *Report No. 409*, Centre for Research into Industry, Enterprise, Finance and the Firm.

Martins, P. (2007). "Rent sharing and wages." *Reflets et Perspectives*, 2(XLVI), 23–31.

Martins, P. (2009a). "Individual teacher incentives, student achievement and grade inflation." *Report No. 4051*, IZA Discussion Paper.

Martins, P. (2009b). "Rent sharing before and after the wage bill." *Applied Economics*, 41(17), 2133–2151.

Martins, P. and Jin, J. (2010). "Firm–level social returns to education." *Journal of Population Economics*, 23(2), 539–558.

McEwan, P. (2003). "Peer effects on students achievement: evidence from Chile." *Economics of Education Review*, 22(2), 131–141.

Meghir, C. and Palme, M. (2003). "Ability parental background and education policy: Empirical evidence a social experiment." *Report No. 03/05*, Institute for Fiscal Studies Working Paper.

Meyer, R. (1997). "Value-added indicators of school performance: A primer." *Economics of Education Review*, 16(3), 283–301.

Mincer, J. (1958). "Investment in human capital and personal income distribution." *Journal of Political Economy*, 66(4), 281–302.

Mincer, J. (1974). *Schooling, Experience and Earnings*. Columbia University Press, New York.

Moretti, E. (2004a). "Estimating the social return to higher education: evidence from longitudinal and repeated cross–sectional data." *Journal of Econometrics*, 121(1–2), 175–212.

Moretti, E. (2004b). "Human capital externalities in cities." *Handbook of Regional and Urban Economics*, Henderson and Thisse, eds., Vol. 4.

Moretti, E. (2004c). "Workers' education, spillovers, and productivity: Evidence from plant–level production functions." *American Economic Review*, 94(3), 656–690.

Mortimore, P. (1991). "School effectiveness research: Which way at the crossroads?" *School effectiveness and school improvement*, 2(3), 213–229.

Muravyev, A. (2008). "Human capital externalities: Evidence from the transition economy of Russia." *Economics of Transition*, 16(3), 415–443.

Neumark, D. (1988). "Employers' discriminatory behavior and the estimation of wage discrimination." *The Journal of Human Resources*, 23(3), 279–295.

Nunes, J. M., Duarte, M. L., Gravitoand, A. P., Monteiro, A., Roque, C., Oliveira, E., Afonso, H. C., Duarte, M. L., Paulo, M. M., Valadares, P. M., and Santos, U. Q. (2013). *Avaliação Externa das Escolas 2011–2012 – Relatório*. Inspeção–Geral da Educação e Ciência. Ministério da Educação, Lisboa.

Oaxaca, R. (1973). "Male–female wage differentials in urban labor markets." *International Economic Review*, 14(3), 693–709.

OECD (2001). *Measuring Productivity: Measurement of Aggregate and Industry-level Productivity Growth: OECD Manual.*

OECD (2012). *Education at a Glance 2012: OECD Indicators.*

Oliveira, M. and Santos, C. (2005). "Assessing school efficiency in Portugal using FDH and bootstrapping." *Applied Economics*, 37(8), 957–968.

Oreopoulos, P. (2006). "Estimating average and local average treatment effects of education when compulsory schooling laws really matter." *American Economic Review*, 96(1), 152–175.

Oreopoulos, P. and Salvanes, K. G. (2011). "Priceless: The nonpecuniary benefits of schooling." *Journal of Economic Perspectives*, 25(1), 159–84.

Park, S. (2011). "Returning to school for higher returns." *Economics of Education Review*, 30(6), 1215–1228.

Pereira, M. (2010). "Desempenho educativo e igualdade de oportunidades em Portugal e na Europa: O papel da escola e a influência da família." *Boletim Económico, Inverno*, Banco de Portugal, 25–48.

Pereira, M. and Moreira, S. (2007). "Eficiência das escolas secundárias portuguesas: uma análise de fronteira de produção estocástica." *Boletim Económico, Primavera*, Banco de Portugal, 101–118.

Pereira, M. and Reis, H. (2012). "Diferenças regionais no desempenho dos alunos portugueses: evidência do programa PISA da OCDE." *Boletim Económico, Inverno*, Banco de Portugal, Banco de Portugal, 59–83.

Pereira, P. and Martins, P. (2000). "Educação, salários e desigualdade: a situação europeia e o caso português." *Economia Pura.*

Pereira, P. and Martins, P. (2002a). "Education and earnings in Portugal." *Desenvolvimento Económico Português no Espaço Europeu: Determinantes e Políticas*, Bank of Portugal Conference Proceedings 2002.

Pereira, P. and Martins, P. (2002b). "Is there a return–risk link in education?" *Economics Letters*, 75(1), 31–37.

Pereira, P. and Martins, P. (2004a). "Does education reduce wage inequality? Quantile regression evidence from 16 countries." *Labour Economics*, 11(3), 355–371.

Pereira, P. and Martins, P. (2004b). "Returns to education and wage equations." *Applied Economics*, 36(6), 525–531.

Perelman, S. and Santin, D. (2011). "Measuring educational efficiency at student level with parametric stochastic distance functions: An application to Spanish PISA results." *Education Economics*, 19(1), 29–49.

Pischke, J. and von T. Wachter (2008). "Zero returns to compulsory schooling in Germany: Evidence and interpretation." *Review of Economics and Statistics*, 90(3), 592–598.

Portela, M. (2001). "Measuring skill: a multi–dimensional index." *Economics Letters*, 72(1), 27–32.

Portela, M., Vieira, J., and Cardoso, A. (2007a). "Gender segregation and the wage gap in Portugal: an analysis at the establishment level." *The Journal of Economic Inequality*, 3(2), 145–168.

Portela, M. C. (2011). "Avaliação: Como medir o valor acrescentado de escolas e professores?" *Encontro sobre Reformas Educativas de Sucesso do Fórum para a Liberdade de Educação*, Fundação Calouste Gulbenkian, Lisboa.

Portela, M. C., Camanho, A., and Azevedo, J. (2007b). "Análise do valor acrescentado de escolas portuguesas." *Revista Portuguesa de Investigação Educacional*, 5, 129–163.

Portela, M. C., Camanho, A., and Borges, D. (2011). "BESP – benchmarking of Portuguese secondary schools." *Benchmarking: An International Journal*, 18(2), 240–260.

Portela, M. C., Camanho, A., and Borges, D. (2012). "Performance assessment of secondary schools: the snapshop of a country taken by DEA." *Journal of Operational Reasearch Society*, 63(8), 1098–1115.

Portela, M. C., Camanho, A., and Keshvari, A. (2013). "Assessing the evolution of school performance and value–added: trends over four years." *Journal of Productivity Analysis*, 39(1), 1–14.

Portugal, P. (2004). "Mitos e factos sobre o mercado de trabalho português: a trágica fortuna dos licenciados." *Boletim Económico*, Banco de Portugal.

Psacharopoulos, G. (1981). "Education and structure of earnings in Portugal." *De Economist*, 129(4), 532–545.

Psacharopoulos, G. (1985). "Returns to education: a further international update and implications." *Journal of Human Resources*, 20(4), 583–604.

Psacharopoulos, G. (1994). "Returns to investment in education: A global update." *World Development*, 22(9), 1325–1343.

Psacharopoulos, G. and Patrinos, H. (2004). "Returns to investment in education: a further update." *Education Economics*, 12(2), 111–134.

Rauch, J. (1993). "Productivity gains from geographic concentration of human capital: Evidence from the cities." *Journal of Urban Economics*, 34(3), 380–400.

Raudenbush, S. and Bryk, A. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications, Thousand Oaks, 2nd edition.

Reynolds, D. and Creemers, B. (1990). "School effectiveness and school improvement: A mission statement." *School Effectiveness and School Improvement*, 1(1), 1–3.

Reynolds, D., Sammons, P., Fraine, B. D., Damme, J. V., Townsend, T., Teddlie, C., and Stringfield, S. (2014). "Educational effectiveness research (EER): a state–of–the–art review." *School Effectiveness and School Improvement*, 25(2), 197–230.

Rivkin, S., Hanushek, E., and Kain, J. (2005). "Teachers, schools, and academic achievement." *Econometrica*, 73(2), 417–458.

Roback, J. (1982). "Wages, rents, and the quality of life." *The Journal of Political Economy*, 90(6), 1257–1278.

Rockoff, J. (2004). "The impact of individual teachers on student achievement: Evidence from panel data." *American Economic Review*, 94(2), 247–252.

Rodríguez-Pose, A. and Tselios, V. (2012). "Individual earnings and educational externalities in the European Union." *Regional Studies*, 46(1), 39–57.

Rudd, J. (2000). "Empirical evidence on human capital spillovers." *Report No. 2000-46*, Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System (U.S.).

Rutter, M. and Maughan, B. (2002). "School effectiveness findings 1979–2002." *Journal of School Psychology*, 40(6), 451–475.

Sammons, P., Hillman, J., and Mortimore, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research*. Office for Standards in Education, London.

Saraiva, A. (1999). "Incidência e impacto salarial da formação profissional em Portugal." *Economia*, XXII, 27–43.

Sarrico, C., Costa, C., Machado, I., Rosa, M., Teixeira, P., Cardoso, M., and Sá, C. (2014). *Desempenho de Escolas Portuguesas – Que Fatores Fazem a Diferença?* CIPES, Centro de Investigação de Políticas do Ensino Superior.

Sarrico, C. and Rosa, M. (2009). "Measuring and comparing the performance of Portuguese secondary schools: A confrontation between metric and practice benchmarking." *International Journal of Productivity and Performance Management*, 58(8), 767–786.

Sarrico, C., Rosa, M., and Coelho, I. (2010). "The performance of Portuguese secondary schools: an exploratory study." *Quality Assurance in Education*, 18(4), 286–303.

Scheerens, J. (2004). *Review of school and instructional effectiveness research.* UNESCO.

Scheerens, J. (2013). "The use of theory in school effectiveness research revisited." *School Effectiveness and School Improvement: An International Journal of Research*, 24(1), 1–38.

Schlotter, M., Schwerdt, G., and Woessman, L. (2011). "Econometric methods for causal evaluation of education policies and practices: a non–technical guide." *Education Economics*, 19(2), 109–137.

Schultz, T. (1961). "Investment in human capital." *The American Economic Review*, 51(1), 1–17.

Schwerdt, G. and Wuppermann, A. (2011). "Is traditional teaching really all that bad? A within–student between–subject approach." *Economics of Education Review*, 30(2), 365–379.

Sherman, H. and Zhu, J. (2006). *Service Productivity Management: Improving Service Performance using Data Envelopment Analysis (DEA).* Springer Science & Business Media, New York.

Siddiqui, O., Hedeker, D., Flay, B., and Hu, F. (1996). "Intraclass correlation estimates in a school–based smoking prevention study, outcome and mediating variables, by sex and ethnicity." *American Journal of Epidemiology*, 144(4), 425–433.

Silva, J. (2003). "Local human capital externalities or sorting? Evidence from a displaced workers sample." *Report No. 9/2003*, Núcleo de Investigação em Políticas Económicas Working Paper.

Slater, H., Davies, N., and Burgess, S. (2012). "Do teachers matter? Measuring the variation in teacher effectiveness in England." *Oxford Bulletin of Economics and Statistics*, 74(5), 629–645.

Song, M., Orazem, P., and Wohlgemuth, D. (2008). "The role of mathematical and verbal skills on the returns to graduate and professional education." *Economics of Education Review*, 27(6), 664–675.

Staiger, D. and Stock, J. (1997). "Instrumental variables regression with weak instruments." *Econometrica*, 65(3), 557–586.

Stevens, A. H. and Schaller, J. (2011). "Short–run effects of parental job loss on children's academic achievement." *Economics of Education Review*, 30(2), 289–299.

Stock, J. and Yogo, M. (2005). "Testing for weak instruments in linear IV regression." *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, D. Andrews, J. Stock, and T. Rothenberg, eds.

Summers, A. and Wolfe, B. (1977). "Do schools make a difference?" *The American Economic Review*, 67(4), 639–652.

Teulings, C. and Vieira, J. (2004). "Urban versus rural return to human capital in Portugal: a cookbook recipe for applying assignment models." *Labour Economics*, 18(2), 265–291.

Thanassoulis, E. and Portela, M. (2002). "School outcomes: Sharing the responsibility between pupil and school." *Education Economics*, 10(2), 183–207.

Thieme, C., Prior, D., and Tortosa-Ausina, E. (2013). "A multilevel decomposition of school performance using robust nonparametric frontier techniques." *Economics of Education Review*, 32(2013), 104–121.

Tobler, W. (1970). "A computer movie simulating urban growth in the Detroit region." *Economic geography*, 46, Supplement: Proceedings. International Geographical Union. Commission on Quantitative Methods (Jun., 1970), 234–240.

Todd, P. and Wolpin, K. (2003). "On the specification and estimation of the production function for cognitive achievement." *The Economic Journal*, 113(485), F3–F33.

Topel, R. (1991). "Specific capital, mobility and wages: Wages rise with job seniority." *Journal of Political Economy*, 99(1), 145–176.

Torrencilla, F. (2003). "Una panorámica de la investigación Iberoamericana sobre eficácia escolar." *REICE - Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 1(1), 1–14.

Trostel, P. (2005). "Nonlinearity in the return to education." *Journal of Applied Economics*, 8(1), 191–202.

Verbeek, M. (2012). *A Guide to Modern Econometric*. Wiley, New York, 4th edition.

Vieira, J. (1999). "Returns to education in Portugal." *Labour Economics*, 6(4), 535–541.

Vieira, J., Couto, J., and Tiago, M. (2005). "Inter–regional wage dispersion in Portugal." *Report No. 1664*, Institute for the Study of Labor Discussion Paper.

Vieira, J., Couto, J., and Tiago, M. (2006). "Regional differences in returns to education in Portugal." *46th Congress of the European Regional Science Association University of Thessaly, Greece, August 30th–September 3rd*.

Walsh, E., Liu, A. Y., and Dotter, D. (2015). *Measuring Teacher and School Value Added in Oklahoma, 2013–2014 School Year*. Mathematica Policy Research.

Wang, L. (2013). "Estimating returns to education when the IV sample is selective." *Labour Economics*, 21, 74–85.

Wantchekon, L., Klašnja, M., and Novta, N. (2015). "Education and human capital externalities: evidence from colonial Benin." *The Quarterly Journal of Economics*, 130(2), 703–757.

Witte, D. and Klaveren, C. (2014). "How are teachers teaching? A nonparametric approach." *Education Economics*, 22(1), 3–23.

Woessman, L. (2003). "Schooling resources, educational institutions and student performance: the international evidence." *Oxford Bulletin of Economics and Statistics*, 65(2), 117–170.

Wooldridge, J. (2006). *Introductory Econometrics, A modern approach*. Thomson Corporation, South–Westerm, 3rd edition.

Wooldridge, J. (2010). *Econometric analysis of cross section and panel data*. The MIT Press, Massachusetts, 2nd edition.

Worthington, A. (2001). "An empirical survey of frontier efficiency measurement techniques in education." *Education Economics*, 9(3), 245–268.

Appendix A

Appendices to Chapter 3

A.1 Checks on the consistency of data and constraints imposed

Quadros de Pessoal (QP) is an annual mandatory employment survey collected by the Portuguese Ministry of Labour, Solidarity and Social Security, since 1982, which gathers information on all firms with wage-earners in Portugal. Until 1993, these surveys were carried out in March and, from the year 1994 onwards, they were collected in October. Given the legal binding nature of the inquiry, the response rate is extremely high. At the same time, and because the information on earnings is reported by the employer, it is likely to be subject to less measurement error than worker-provided earnings data. The fact that the information included in the QP survey needs, by law, to be permanently displayed in a public space at the establishment, it contributes to the reliability of the data, reducing measurement errors.

This longitudinal database includes information on firms (such as location, industry, sales, number of employees, date of constitution, legal setting, ownership type), establishments (location, industry, number of workers) and workers (gender, age, education, hours of work, labour earnings, date of admission into the firm, among others). Civil servants and domestic service are not covered, and the coverage of agriculture is low given its low share of wage-earners over most of the period under analysis.

While assembling the data, several inconsistencies in key variables, at worker and firm level, were detected. In order to guarantee data consistency, a thorough analysis of the data was carried out. The data assembling task had at its core the goal to ensure a high quality of the data used in the empirical analysis. While not being exhaustive on all the checks performed underlying the data preparation, one must emphasise the thorough diligence that characterized the data preparation. We will highlight the main

tasks performed.

In a first stage, duplicated lines in each year, as well as the workers whose identification number is lower than 100,000, were eliminated (workers whose identification number were duplicated have also been eliminated). As such, in each year between 9% and 20% of the observations were eliminated. In the first years of the data the information was inserted manually, originating, in some cases, typing errors. So, a second set of checks was performed. Workers who registered changes in gender or date of birth, or, for example, with unrealistic date of birth, were given special attention. Similarly, education, following its critical relevance for the research project, has been one of the variables that has deserved a detailed consistency checks.

As an example, making use of the longitudinal dimension of the data, missing information, or inconsistent values reported over time for a given worker, were evaluated in order to perform a correction or elimination. The underlying rule has been that imputed values would only occur if the information registered over time more than half the periods was considered as true. While upward changes in education were considered valid, a decrease in education, or unrealistic increases in a short period, were discarded.

If some inconsistencies within a worker were detected, his whole information was dropped (2.2%, 0.5% and 6.8% of the observations, respectively for gender, birth date and schooling). Workers who remain with missing data for gender, birth date or education were dropped corresponding to about 7% of the workers. Initially, only workers aged between 15 and 75 years old remained in the data, which accounted for only a small elimination of observations (0.15%). Also workers workers with more than 3 simultaneous jobs in a given year were eliminated. Additionally, in order to eliminate outliers on wage were eliminated. The criteria has been log difference in real wages either greater than 2 or smaller than -0.5 between two consecutive years. This condition implied only a small elimination of observations.

Appending the full original worker data between 1986 and 2009 resulted in 55,841,335 observations. After the corrections outlined above, about 39% of the observations was eliminated. The final sample comprises almost 34 million observations worker–year on 5,867,131 unique workers, and 667,343 unique firms. The data set contains 1.5 million workers on average each year and from 2000, there are more than 200 thousand firms in each year.

The final data has information on full-time employees, aged between 16 and 64 years old, and whose wages are not less than the national minimum wage . Due to specific characteristics of the Islands' labour market, for example, regarding the minimum wage definition, only workers and firms for Portugal's mainland were kept in the final data.

A.2 Description of variables included in the empirical models and additional tables of descriptive statistics

Table A.1: Description of the variables

Variable	**Description**
Real hourly wage (Euro)	Real hourly wage (Euro) (wages adjusted using the IPC)
Education	Years of schooling and it were calculated by imputing the nominal number of years required to complete the level of schooling
Female	Dummy variable: 0 for male and 1 for female
Age	Years of age
Tenure	Years of tenure (years of experience in the current firm) which is equal to the date of the questionnaire minus date of admission, converted to years
Experience	Years of experience which is calculated by: $exper = age - years\,of\,education - 6$ if $(years\ of\ education\ +6) \geq 16$ $exper = age - 16$ if $(years\ of\ education\ +6) < 16$
Occupation	Occupation level inside the firm is a categorical variable 1 = Executive and managerial 2 = Intermediate managerial and executive 3 = Low managerial 4 = Technicians highly specialized 5 = Sales, administrative and precision production 6 = Administrative support, and production 7 = Unskilled 8 = Apprentice
Size	Number of workers in firm
Region	Location of the firm is a set of dummies 5 regions 100 = North; 200 = Centre; 300 = Lisbon and Tagus Valley 400 = Inland; 500 = Algarve
Real sales (1000 Euro)	Real sales in thousands of euros (sales adjusted using the GDP deflator)

Source: Created by the author based on *Quadros de Pessoal*, 1986–2009.

Table A.2: Descriptive statistics on workers in 1986 and 2009

	1986			2009		
Variable	**Obs.**	**Mean**	**Std. Dev.**	**Obs.**	**Mean**	**Std. Dev.**
Real hourly wage (Euro)	1,066,019	3.566	2.490	1,885,151	6.228	6.188
Female	1,066,019	0.328		1,885,151	0.442	
Age	1,066,019	36.034	11.425	1,885,151	38.99	10.60
Experience	1,065,943	19.815	11.465	1,883,591	21.539	11.229
Tenure	1,066,019	9.745	8.263	1,885,151	8.180	8.372
Education	1,066,019	5.239	2.945	1,885,151	9.473	3.967
$\leq 1^{st}$ cycle		0.683			0.168	
2^{nd} cycle		0.132			0.183	
3^{rd} cycle		0.113			0.235	
Secondary		0.045			0.251	
Tertiary		0.023			0.163	
Occupation	927,281			1,884,114		
Executive/managerial		0.054			0.046	
Intermediate managerial/executive		0.030			0.075	
Low managerial		0.046			0.123	
Technicians highly specialized		0.153			0.164	
Sales, administ. and precision prod.		0.113			0.173	
Administ. support and production		0.000			0.012	
Unskilled		0.414			0.194	
Apprentice		0.088			0.102	
Unknown		0.102			0.108	

Source: Computations of the author based on *Quadros de Pessoal*, 1986 and 2009.

Table A.3: Descriptive statistics on firms in 1986 and 2009

	1986			2009		
Variable	**Obs.**	**Mean**	**Std. Dev.**	**Obs.**	**Mean**	**Std. Dev.**
Size	77,558	22.304	205.29	250,072	11,208	103.362
Real sales (1000 Euro)	69,444	1,629,188	2.89e+07	224,600	1,284,602	1.72e+07
Region	77,558			250,042		
North		0.287			0.297	
Centre		0.149			0.166	
Lisbon and Tagus Valley		0.417			0.339	
Inland		0.116			0.141	
Algarve		0.031			0.058	

Source: Computations of the author based on *Quadros de Pessoal*, 1986 and 2009.

A.3 Additional tables of OLS estimations

Table A.4: Standard Mincer Equation (OLS) – All workers

VARIABLES	**1986**	**1987**	**1988**	**1989**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**
Education	0.0916***	0.0902***	0.0905***	0.0934***	0.0995***	0.1006***	0.1006***	0.1015***	0.1030***	0.1009***	0.0970***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Experience	0.0525***	0.0498***	0.0485***	0.0470***	0.0439***	0.0427***	0.0417***	0.0424***	0.0429***	0.0416***	0.0405***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Experience sq	-0.0008***	-0.0007***	-0.0007***	-0.0006***	-0.0006***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Observations	1,065,943	1,090,330	1,127,276	1,215,832	1,322,819	1,369,619	1,362,959	1,400,904	1,460,372	1,448,800	1,544,887
R-squared	0.449	0.440	0.446	0.448	0.439	0.431	0.415	0.410	0.434	0.418	0.410
RMSE	0.362	0.360	0.358	0.370	0.406	0.416	0.430	0.443	0.428	0.431	0.421
VARIABLES	**1998**	**1999**	**2000**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Education	0.0981***	0.0988***	0.0995***	0.0986***	0.0988***	0.1011***	0.1006***	0.1000***	0.0989***	0.0976***	0.0985***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Experience	0.0402***	0.0378***	0.0375***	0.0360***	0.0369***	0.0386***	0.0382***	0.0376***	0.0363***	0.0344***	0.0345***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Experience sq	-0.0005***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0003***	-0.0003***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Observations	1,567,140	1,656,395	1,674,304	1,665,287	1,675,406	1,727,619	1,857,626	1,878,142	1,932,095	1,961,684	1,883,591
R–squared	0.416	0.411	0.409	0.400	0.393	0.394	0.390	0.380	0.370	0.364	0.363
RMSE	0.418	0.422	0.424	0.431	0.437	0.447	0.449	0.451	0.453	0.450	0.452

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: Standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage.

Table A.5: Standard Mincer Equation (OLS) – Male workers

VARIABLES	**1986**	**1987**	**1988**	**1989**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**
Education	0.0917***	0.0915***	0.0920***	0.0950***	0.1011***	0.1025***	0.1025***	0.1034***	0.1054***	0.1029***	0.1001***
	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Experience	0.0564***	0.0543***	0.0534***	0.0517***	0.0500***	0.0491***	0.0483***	0.0493***	0.0500***	0.0487***	0.0475***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Experience sq	-0.0009***	-0.0008***	-0.0008***	-0.0008***	-0.0007***	-0.0007***	-0.0007***	-0.0007***	-0.0007***	-0.0006***	-0.0006***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Observations	715,819	730,395	744,516	786,152	837,292	852,502	845,403	851,092	879,784	867,851	917,822
R–squared	0.451	0.448	0.455	0.455	0.450	0.445	0.429	0.430	0.452	0.435	0.429
RMSE	0.366	0.365	0.361	0.374	0.408	0.418	0.430	0.440	0.428	0.431	0.424
VARIABLES	**1998**	**1999**	**2000**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Education	0.1012***	0.1024***	0.1028***	0.1014***	0.1011***	0.1030***	0.1023***	0.1016***	0.1003***	0.0989***	0.0998***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Experience	0.0473***	0.0450***	0.0446***	0.0424***	0.0432***	0.0451***	0.0451***	0.0448***	0.0436***	0.0416***	0.0418***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Experience sq	-0.0006***	-0.0006***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Observations	925,377	966,918	973,834	975,765	973,087	1,002,073	1,067,840	1,066,367	1,089,538	1,100,113	1,051,600
R–squared	0.438	0.432	0.430	0.415	0.406	0.404	0.398	0.388	0.376	0.368	0.369
RMSE	0.419	0.424	0.427	0.436	0.443	0.453	0.457	0.460	0.463	0.460	0.462

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: Standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage.

Table A.6: Standard Mincer Equation (OLS) – Female workers

VARIABLES	**1986**	**1987**	**1988**	**1989**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**
Education	0.0897***	0.0853***	0.0865***	0.0896***	0.0957***	0.0967***	0.0972***	0.0981***	0.0996***	0.0985***	0.0939***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Experience	0.0436***	0.0405***	0.0381***	0.0370***	0.0321***	0.0315***	0.0305***	0.0320***	0.0326***	0.0322***	0.0314***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Experience sq	-0.0007***	-0.0006***	-0.0005***	-0.0005***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Observations	350,124	359,935	382,760	429,680	485,527	517,117	517,556	549,812	580,588	580,949	627,065
R–squared	0.471	0.450	0.463	0.473	0.457	0.448	0.431	0.416	0.460	0.448	0.445
RMSE	0.313	0.306	0.305	0.313	0.348	0.361	0.377	0.400	0.374	0.379	0.365
VARIABLES	**1998**	**1999**	**2000**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Education	0.0949***	0.0956***	0.0976***	0.0999***	0.1003***	0.1034***	0.1038***	0.1034***	0.1030***	0.1020***	0.1027***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Experience	0.0313***	0.0291***	0.0294***	0.0285***	0.0296***	0.0309***	0.0305***	0.0294***	0.0284***	0.0265***	0.0265***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Experience sq	-0.0004***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0002***	-0.0002***	-0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Observations	641,763	689,477	700,470	689,522	702,319	725,546	789,786	811,775	842,557	861,571	831,991
R–squared	0.453	0.453	0.459	0.465	0.459	0.460	0.459	0.450	0.444	0.438	0.432
RMSE	0.362	0.363	0.366	0.375	0.381	0.391	0.394	0.396	0.397	0.396	0.401

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: Standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage.

Table A.7: Firm fixed–effects regressions

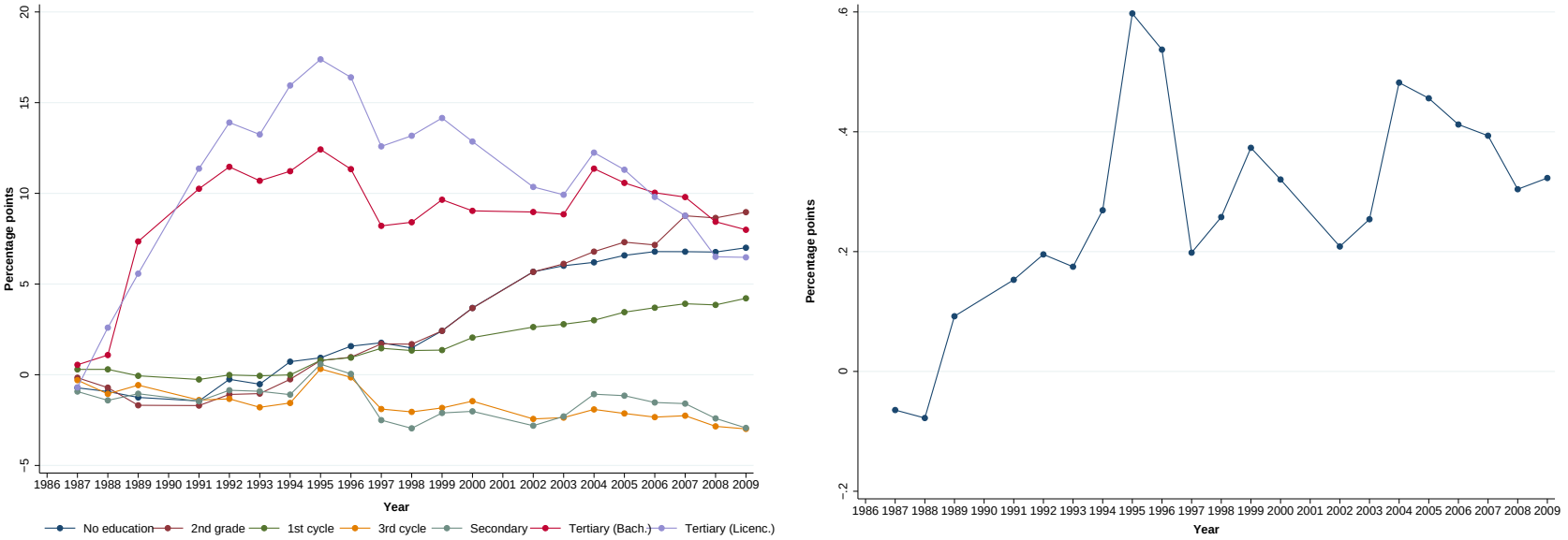
VARIABLES	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Education	0.0386***		0.0365***		0.0355***		0.0395***		0.0339***
	(0.0000)		(0.0000)		(0.0001)		(0.0001)		(0.0002)
Female	-0.1480***	-0.1658***	-0.1897***	-0.1526***	-0.1480***	-0.1521***	-0.1482***	-0.1491***	-0.1467***
	(0.0001)	(0.0003)	(0.0003)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Experience	0.0182***	0.0193***	0.0184***	0.0194***	0.0182***	0.0192***	0.0181***	0.0191***	0.0182***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Experience sq	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Tenure	0.0110***	0.0110***	0.0109***	0.0110***	0.0109***	0.0111***	0.0110***	0.0111***	0.0110***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Tenure sq	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Log size	0.0079***	0.0071***	0.0076***	0.0089***	0.0077***	0.0074***	0.0085***	0.0073***	0.0073***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
No education		-0.2040***		-0.2010***		-0.1840***		-0.2066***	
		(0.0007)		(0.0018)		(0.0021)		(0.0039)	
2^{nd} grade		-0.1591***		-0.1678***		-0.1555***		-0.1842***	
		(0.0006)		(0.0014)		(0.0015)		(0.0039)	
1^{st} cycle		-0.0705***		-0.0888***		-0.0718***		-0.0749***	
		(0.0002)		(0.0009)		(0.0005)		(0.0017)	
3^{rd} cycle		0.0535***		0.0860***		0.0682***		0.0369***	
		(0.0002)		(0.0011)		(0.0006)		(0.0026)	
Secondary		0.1368***		0.1736***		0.1560***		0.0883***	
		(0.0002)		(0.0015)		(0.0007)		(0.0028)	
Tertiary (Bach.)		0.4100***		0.3255***		0.4517***		0.4183***	
		(0.0006)		(0.0041)		(0.0016)		(0.0073)	
Tertiary (Lic.)		0.5338***		0.4329***		0.5877***		0.5743***	
		(0.0004)		(0.0031)		(0.0011)		(0.0044)	
No education×Female		0.0452***							
		(0.0013)							
2^{th} grade×Female		0.0127***							
		(0.0009)							

Continued on next page

Table A.7 – continued from previous page

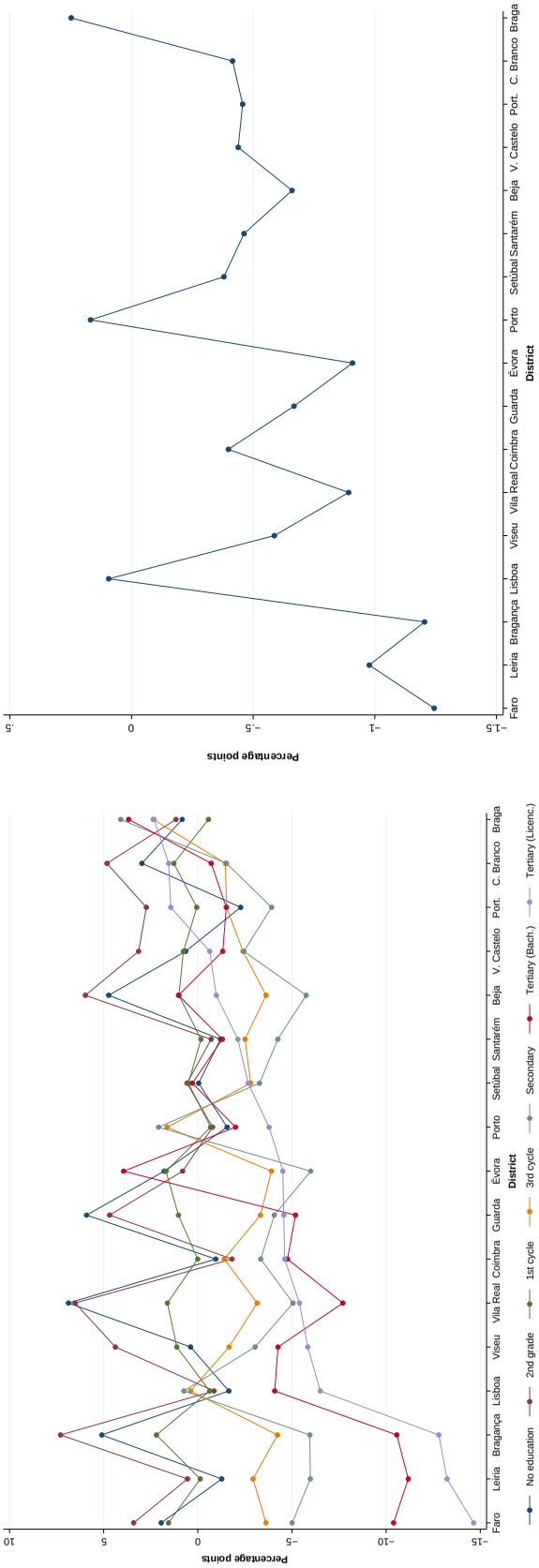
1^{st} cycle×Female		-0.0100***							
		(0.0003)							
3^{rd} cycle×Female		0.0367***							
		(0.0003)							
Secondary×Female		0.0424***							
		(0.0004)							
Tertiary (Bach.)×Female		0.0187***							
		(0.0008)							
Tertiary (Lic.)×Female		-0.0002							
		(0.0005)							
Education×Female			0.0051***						
			(0.0000)						
R–squared	0.775	0.781	0.775	0.782	0.775	0.782	0.775	0.784	0.777
RMSE	0.269	0.265	0.269	0.265	0.269	0.265	0.269	0.264	0.268

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: Standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. The sample includes 31,167,549 observations. The wage regressions include interaction terms: education×gender in columns (2) and (3); education×year in columns (4) and (5); education×district in columns (6) and (7); education×industry in columns (8) and (9). All regressions control for firm specific effects and include a set of dummies to control for occupation, industry, region/district and year. 6^{th} grade is the omitted category in education.



Source: Created by author based on *Quadros de Pessoal*, 1986–2009.

Figure A.1: Firm fixed-effects regressions with interactions education×year



Source: Created by author based on *Quadros de Pessoal*, 1986–2009.
Note: On the horizontal axis V. Castelo is Viana do Castelo, Port. is Portalegre and C. Branco is Castelo Branco.

Figure A.2: Firm fixed-effects regressions with interactions education×district

Source: Created by author based on *Quadros de Pessoal*, 1986–2009.
Note: Industry labels are: 2–Mining; 3–Food, beverage, tobacco; 4–Textiles, dressing, leather; 5–Wood, cork, paper; 6–Manuf. non metallic products; 7–Manuf. metals products; 8–Manuf. of furniture; 10–Electriciy, gas and water supply; 11–Construction; 12–Whole sale and retail trade; 13–Restaurants, hotels; 14–Transport, communications; 15–Post and telecommunications; 16–Financial intermediation; 17–Real estate, renting and business; 19–Education; 20–Health and social work; 21–Other services .

Figure A.3: Firm fixed-effects regressions with interactions education×industry

A.4 Additional tables of IV estimations

Table A.8: First Stage IV2SLS estimation using the 1^{st} quarter of birth (q_1) and the 3^{rd} quarter of birth (q_3) as instruments for education

Year/Cohort	q_1	R^2	***F*-test on excluded variables**	q_3	R^2	***F*-test on excluded variables**
			Cohort 1930-1949			
1986	0.8555***	0.6263	12332***	-0.2631***	0.6060	1078***
	(0.0077)			(0.0080)		
1987	0.8591***	0.6355	11999***	-0.2627***	0.6153	1153***
	(0.0078)			(0.0077)		
1988	0.8608***	0.6285	12553***	-0.2539***	0.6080	907***
	(0.0077)			(0.0008)		
1989	0.8660***	0.6410	13587***	-0.2599***	0.6213	1227***
	(0.0074)			(0.0074)		
1991	0.8442***	0.6914	9261***	-0.2548***	0.6756	990***
	(0.0088)			(0.0081)		
1992	0.8542***	0.6909	8504***	-0.2610***	0.6749	1006***
	(0.0093)			(0.0082)		
1993	0.8470***	0.6975	8257***	-0.2625***	0.6821	1028***
	(0.0093)			(0.0082)		
1994	0.8682***	0.7172	5495***	-0.2720***	0.7025	911***
	(0.0117)			(0.0090)		
1995	0.7840***	0.7847	4498***	-0.2522***	0.7750	910***
	(0.0117)			(0.0084)		
1996	0.7790***	0.7890	5459***	-0.2482***	0.7796	902***
	(0.0105)			(0.0083)		
1997	0.7774***	0.7898	4857***	-0.2371***	0.7804	795***
	(0.0112)			(0.0084)		
1998	0.8043***	0.7879	4874***	-0.2543***	0.7782	859***
	(0.0115)			(0.0087)		
1999	0.8191***	0.7851	4586***	-0.2579***	0.7751	807***
	(0.0121)			(0.0091)		
2000	0.8333***	0.7863	4272***	-0.2759***	0.7761	810***
	(0.0127)			(0.0097)		
			Cohort 1950-1966			
1986	0.8730***	0.6315	8410***	-0.2949***	0.6139	1281***
	(0.0095)			(0.0082)		
1987	0.8745***	0.6440	9994***	-0.2866***	0.6267	1723***

Continued on next page

Table A.8 – continued from previous page

	(0.0087)			(0.0069)		
1988	0.8852***	0.6500	10511***	-0.2855***	0.6329	1490***
	(0.0086)			(0.0074)		
1989	0.8744***	0.6725	10753***	-0.2876***	0.6570	1738***
	(0.0084)			(0.0069)		
1991	0.8513***	0.7138	9078***	-0.2823***	0.7013	1582***
	(0.0089)			(0.0071)		
1992	0.8560***	0.7216	8500***	-0.2782***	0.7092	1586***
	(0.0093)			(0.0070)		
1993	0.8572***	0.7298	7966***	-0.2925***	0.7178	1780***
	(0.0096)			(0.0069)		
1994	0.8408***	0.7538	6318***	-0.2782***	0.7432	1733***
	(0.0106)			(0.0069)		
1995	0.7855***	0.8006	8558***	-0.2620***	0.7928	1534***
	(0.0085)			(0.0067)		
1996	0.7771***	0.8015	7482***	-0.2542***	0.7939	1566***
	(0.0090)			(0.0064)		
1997	0.7780***	0.8006	8170***	-0.2525***	0.7930	1566***
	(0.0086)			(0.0064)		
1998	0.7792***	0.8010	8596***	-0.2543***	0.7936	1615***
	(0.0084)			(0.0063)		
1999	0.7784***	0.8009	8668***	-0.2514***	0.7934	1646***
	(0.0084)			(0.0062)		
2000	0.7770***	0.7996	7876***	-0.2566***	0.7922	1650***
	(0.0088)			(0.0063)		
2002	0.8006***	0.7886	6788***	-0.2692***	0.7807	1669***
	(0.0097)			(0.0066)		
2003	0.8059***	0.7846	6228***	-0.2624***	0.7766	1537***
	(0.0102)			(0.0067)		
2004	0.8047***	0.7827	6604***	-0.2673***	0.7749	1655***
	(0.0099)			(0.0066)		
2005	0.8053***	0.7834	7404***	-0.2622***	0.7756	1582***
	(0.0094)			(0.0066)		
2006	0.8110***	0.7796	6022***	-0.2643***	0.7718	1451***
	(0.0105)			(0.0069)		
2007	0.8168***	0.7746	6478***	-0.2687***	0.7666	1479***
	(0.0101)			(0.0070)		
2008	0.8041***	0.7711	6591***	-0.2607***	0.7634	1322***
	(0.0099)			(0.0072)		
2009	0.7934***	0.7722	6102***	-0.2576***	0.7649	1199***
	(0.0102)			(0.0074)		

Continued on next page

Table A.8 – continued from previous page

Cohort 1967-1980						
1992	1.0623***	0.6198	5519***	-0.2782***	0.7092	1586***
	(0.0143)			(.0070)		
1993	1.0250***	0.6524	5182***	-0.2925***	0.7178	1780***
	(0.0142)			(.0069)		
1994	0.9762***	0.6932	5389***	-0.2782***	0.7432	1733***
	(0.0133)			(.0069)		
1995	0.8707***	0.7359	4631***	-0.2620***	0.7928	1534***
	(0.0128)			(.0067)		
1996	0.8398***	0.7486	4497***	-0.2542***	0.7939	1566***
	(0.0125)			(.0064)		
1997	0.8159***	0.7614	5682***	-0.2525***	0.7930	1566***
	(0.0108)			(.0064)		
1998	0.7906***	0.7751	6599***	-0.2543***	0.7936	1615***
	(0.0097)			(.0063)		
1999	0.7823***	0.7829	7299***	-0.2514***	0.7934	1646***
	(0.0092)			(.0062)		
2000	0.7571***	0.7947	7093***	-0.2566***	0.7922	1650***
	(0.0090)			(.0063)		
2002	0.7421***	0.8054	7942***	-0.2692***	0.7807	1669***
	(0.0083)			(.0066)		
2003	0.7339***	0.8089	8510***	-0.2624***	0.7766	1537***
	(0.0080)			(.0067)		
2004	0.7266***	0.8148	9298***	-0.2673***	0.7749	1655***
	(0.0075)			(.0066)		
2005	0.7236***	0.8228	10012***	-0.2622***	0.7756	1582***
	(0.0072)			(.0066)		
2006	0.7198***	0.8278	9984***	-0.2643***	0.7718	1451***
	(0.0072)			(.0069)		
2007	0.7118***	0.8308	10998***	-0.2687***	0.7666	1479***
	(0.0068)			(.0070)		
2008	0.7007***	0.8348	11571***	-0.2607***	0.7634	1322***
	(0.0065)			(.0072)		
2009	0.6900***	0.8390	10865***	-0.2576***	0.7649	1199***
	(0.0066)			(.0074)		

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.

Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, ** p<0.05, * p<0.1.

Table A.9: OLS and IV estimates using q_3 and $q_3 \times year - of - birth$ as instruments, Men born 1930–1949

	1986			1987			1988			1989		
VARIABLES	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$
Education	0.0704***	0.0680***	0.0661***	0.0711***	0.0658***	0.0639***	0.0701***	0.0653***	0.0675***	0.0728***	0.0704***	0.0717***
	(0.0004)	(0.0070)	(0.0066)	(0.0004)	(0.0068)	(0.0067)	(0.0004)	(0.0072)	(0.0069)	(0.0005)	(5.1515)	(0.0072)
Experience	0.0467***	0.0422***	0.0403***	0.0456***	0.0359***	0.0291**	0.0385***	0.0297**	0.0167	0.0157***	-0.0110	0.0038
	(0.0038)	(0.0136)	(0.0132)	(0.0039)	(0.0139)	(0.0145)	(0.0040)	(0.0147)	(0.0152)	(0.0043)	(13.8759)	(0.0155)
Experience sq	-0.0008***	-0.0008***	-0.0008***	-0.0007***	-0.0007***	-0.0006***	-0.0006***	-0.0006***	-0.0003*	-0.0003***	0.0003	-0.0001
	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.3248)	(0.0002)
Centre	-0.0520***	-0.0520***	-0.0524***	-0.0675***	-0.0674***	-0.0660***	-0.0598***	-0.0598***	-0.0610***	-0.0435***	-0.0434	-0.0416***
	(0.0025)	(0.0170)	(0.0164)	(0.0025)	(0.0163)	(0.0161)	(0.0025)	(0.0162)	(0.0159)	(0.0026)	(0.0283)	(0.0145)
Lisbon and	0.1577***	0.1584***	0.1497***	0.1564***	0.1578***	0.1490***	0.1439***	0.1452***	0.1425***	0.1589***	0.1592	0.1634***
Tagus Valley	(0.0017)	(0.0220)	(0.0201)	(0.0018)	(0.0212)	(0.0203)	(0.0018)	(0.0194)	(0.0181)	(0.0019)	(0.2362)	(0.0185)
Inland	-0.0788***	-0.0789***	-0.0777***	-0.1009***	-0.1012***	-0.1010***	-0.1160***	-0.1162***	-0.1160***	-0.1002***	-0.1007	-0.0971***
	(0.0035)	(0.0173)	(0.0169)	(0.0034)	(0.0168)	(0.0165)	(0.0034)	(0.0149)	(0.0146)	(0.0034)	(0.2001)	(0.0142)
Algarve	0.0376***	0.0380	0.0347	0.0462***	0.0469*	0.0457*	0.0939***	0.0948***	0.0923***	0.0843***	0.0835	0.0885***
	(0.0062)	(0.0258)	(0.0252)	(0.0062)	(0.0262)	(0.0260)	(0.0060)	(0.0281)	(0.0277)	(0.0061)	(0.7400)	(0.0227)
Observations	244,451	244,451	244,451	242,365	242,365	242,365	239,442	239,442	239,442	240,928	240,928	240,928
R–squared	0.527	0.527	0.525	0.530	0.530	0.485	0.518	0.518	0.304	0.510	0.399	0.420
RMSE	0.325	0.325	0.326	0.329	0.329	0.344	0.326	0.326	0.392	0.344	0.344	0.344
Hansen J stat.			11.59			14.18			9.71			18.12
p–value of HJ			0.903			0.773			0.960			0.515

Continued on next page

Table A.9 – continued from previous page

VARIABLES	1991			1992			1993			1994		
	OLS	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$
Education	0.0694***	0.0618***	0.0614***	0.0690***	0.0517	0.0671***	0.0679***	0.0625***	0.0602***	0.0635***	0.0761***	0.0764***
	(0.0005)	(0.0075)	(0.0071)	(0.0005)	(5.3083)	(0.0079)	(0.0005)	(0.0079)	(0.0077)	(0.0006)	(0.0082)	(0.0080)
Experience	0.0187***	0.0058	0.0002	0.0303***	0.0017	0.0270	0.0209***	0.0116	0.0033	0.0390***	0.0584***	0.0561***
	(0.0046)	(0.0160)	(0.0155)	(0.0051)	(8.5955)	(0.0166)	(0.0058)	(0.0197)	(0.0188)	(0.0064)	(0.0201)	(0.0192)
Experience sq	-0.0004***	-0.0004**	-0.0003**	-0.0006***	-0.0006	-0.0006***	-0.0005***	-0.0005**	-0.0004*	-0.0009***	-0.0008***	-0.0008***
	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0008)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0002)
Centre	-0.0437***	-0.0440**	-0.0392**	-0.0298***	-0.0332	-0.0263*	-0.0354***	-0.0355**	-0.0344**	-0.0381***	-0.0377***	-0.0347**
	(0.0027)	(0.0174)	(0.0169)	(0.0028)	(1.4815)	(0.0142)	(0.0030)	(0.0139)	(0.0135)	(0.0031)	(0.0139)	(0.0137)
Lisbon and	0.1491***	0.1509***	0.1506***	0.1641***	0.1648	0.1747***	0.1638***	0.1653***	0.1629***	0.1504***	0.1472***	0.1479***
Tagus Valley	(0.0019)	(0.0211)	(0.0202)	(0.0020)	(0.3112)	(0.0172)	(0.0021)	(0.0180)	(0.0172)	(0.0022)	(0.0145)	(0.0143)
Inland	-0.1029***	-0.1037***	-0.0970***	-0.0894***	-0.0946	-0.0855***	-0.1134***	-0.1140***	-0.1131***	-0.0904***	-0.0890***	-0.0858***
	(0.0035)	(0.0155)	(0.0151)	(0.0036)	(2.1049)	(0.0148)	(0.0038)	(0.0141)	(0.0139)	(0.0039)	(0.0141)	(0.0139)
Algarve	0.0562***	0.0569**	0.0619**	0.0560***	0.0536	0.0630**	0.0503***	0.0508**	0.0556***	0.0134**	0.0120	0.0124
	(0.0059)	(0.0277)	(0.0274)	(0.0062)	(1.3004)	(0.0248)	(0.0065)	(0.0214)	(0.0213)	(0.0066)	(0.0190)	(0.0189)
Observations	245,029	245,029	245,029	232,054	232,054	232,054	213,901	213,901	213,901	190,929	190,929	190,929
R–squared	0.548	0.548	0.548	0.551	0.157	0.536	0.535	0.535	0.533	0.559	0.557	0.557
RMSE	0.363	0.363	0.363	0.368	0.504	0.374	0.378	0.378	0.379	0.380	0.380	0.380
Hansen J stat.			25.15			14.31			25.48			9.103
p–value of HJ			0.121			0.645			0.062			0.824

Continued on next page

Table A.9 – continued from previous page

VARIABLES	1995			1996			1997		
	OLS	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$
Education	0.0552***	0.0575***	0.0556***	0.0519***	0.0611***	0.0582***	0.0484***	0.0507***	0.0496***
	(0.0006)	(0.0088)	(0.0086)	(0.0006)	(0.0094)	(0.0091)	(0.0006)	(0.0100)	(0.0098)
Experience	0.0123**	0.0154	0.0118	0.0134**	0.0249	0.0180	0.0022	0.0049	0.0032
	(0.0057)	(0.0165)	(0.0161)	(0.0064)	(0.0170)	(0.0163)	(0.0071)	(0.0203)	(0.0193)
Experience sq	-0.0003***	-0.0003*	-0.0003*	-0.0004***	-0.0003*	-0.0003*	-0.0003***	-0.0003	-0.0003
	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0002)
Centre	-0.0451***	-0.0450***	-0.0440***	-0.0473***	-0.0468***	-0.0458***	-0.0293***	-0.0292**	-0.0270*
	(0.0031)	(0.0153)	(0.0152)	(0.0032)	(0.0151)	(0.0150)	(0.0032)	(0.0147)	(0.0146)
Lisbon and	0.1554***	0.1548***	0.1566***	0.1451***	0.1429***	0.1450***	0.1487***	0.1482***	0.1494***
Tagus Valley	(0.0022)	(0.0163)	(0.0161)	(0.0023)	(0.0156)	(0.0153)	(0.0024)	(0.0148)	(0.0146)
Inland	-0.1034***	-0.1032***	-0.1015***	-0.1096***	-0.1085***	-0.1061***	-0.1056***	-0.1054***	-0.1020***
	(0.0039)	(0.0152)	(0.0151)	(0.0041)	(0.0161)	(0.0161)	(0.0041)	(0.0160)	(0.0158)
Algarve	0.0077	0.0076	0.0098	0.0063	0.0055	0.0085	0.0109	0.0108	0.0145
	(0.0065)	(0.0227)	(0.0227)	(0.0068)	(0.0221)	(0.0221)	(0.0068)	(0.0225)	(0.0224)
Observations	194,143	194,143	194,143	177,799	177,799	177,799	169,451	169,451	169,451
R–squared	0.597	0.597	0.597	0.587	0.587	0.587	0.589	0.589	0.589
RMSE	0.380	0.380	0.380	0.382	0.383	0.383	0.380	0.380	0.380
Hansen J statistic			7.906			5.740			6.618
p–value of HJ statistic			0.850			0.929			0.829

Continued on next page

Table A.9 – continued from previous page

VARIABLES	1998			1999			2000		
	OLS	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$
Education	0.0471***	0.0514***	0.0505***	0.0489***	0.0676***	0.0679***	0.0457***	0.0595***	0.0595***
	(0.0006)	(0.0093)	(0.0092)	(0.0006)	(0.0098)	(0.0098)	(0.0007)	(0.0093)	(0.0092)
Experience	0.0093	0.0143	0.0090	-0.0046	0.0153	0.0188	0.0026	0.0157	0.0218
	(0.0077)	(0.0210)	(0.0202)	(0.0092)	(0.0236)	(0.0232)	(0.0109)	(0.0274)	(0.0269)
Experience sq	-0.0004***	-0.0004*	-0.0004	-0.0003**	-0.0001	-0.0002	-0.0003**	-0.0002	-0.0003
	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0003)	(0.0003)	(0.0001)	(0.0003)	(0.0003)
Centre	-0.0290***	-0.0288*	-0.0261*	-0.0278***	-0.0269*	-0.0243*	-0.0206***	-0.0201	-0.0172
	(0.0032)	(0.0148)	(0.0147)	(0.0035)	(0.0145)	(0.0144)	(0.0037)	(0.0130)	(0.0129)
Lisbon and	0.1566***	0.1557***	0.1569***	0.1567***	0.1529***	0.1572***	0.1740***	0.1712***	0.1730***
Tagus Valley	(0.0024)	(0.0151)	(0.0146)	(0.0026)	(0.0154)	(0.0151)	(0.0028)	(0.0168)	(0.0164)
Inland	-0.0933***	-0.0928***	-0.0904***	-0.1041***	-0.1018***	-0.1005***	-0.0917***	-0.0905***	-0.0858***
	(0.0042)	(0.0152)	(0.0151)	(0.0044)	(0.0141)	(0.0141)	(0.0046)	(0.0150)	(0.0150)
Algarve	0.0133*	0.0129	0.0167	0.0249***	0.0225	0.0237	0.0278***	0.0261	0.0326*
	(0.0068)	(0.0204)	(0.0203)	(0.0071)	(0.0181)	(0.0181)	(0.0073)	(0.0199)	(0.0198)
Observations	159,173	159,173	159,173	147,103	147,103	147,103	129,497	129,497	129,497
R–squared	0.608	0.608	0.608	0.601	0.598	0.598	0.616	0.614	0.614
RMSE	0.371	0.371	0.371	0.382	0.383	0.383	0.381	0.382	0.382
Hansen J statistic			11.55			13.70			21.50
p–value of HJ statistic			0.316			0.133			0.006

Continued on next page

Table A.9 – continued from previous page

VARIABLES	2002			2003			2004		
	OLS	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$
Education	0.0420***	0.0652***	0.0639***	0.0426***	0.0725***	0.0719***	0.0435***	0.0633***	0.0648***
	(0.0008)	(0.0123)	(0.0120)	(0.0009)	(0.0144)	(0.0142)	(0.0010)	(0.0163)	(0.0163)
Experience	-0.0140	-0.0007	-0.0001	-0.0470**	-0.0390	-0.0322	-0.1960***	-0.2007***	-0.1905***
	(0.0166)	(0.0364)	(0.0359)	(0.0205)	(0.0376)	(0.0375)	(0.0265)	(0.0453)	(0.0450)
Experience sq	-0.0002	0.0001	0.0000	0.0001	0.0007	0.0005	0.0019***	0.0024***	0.0023***
	(0.0002)	(0.0004)	(0.0004)	(0.0003)	(0.0005)	(0.0005)	(0.0003)	(0.0006)	(0.0006)
Centre	-0.0028	-0.0020	-0.0034	-0.0024	-0.0010	0.0005	-0.0031	-0.0019	-0.0023
	(0.0044)	(0.0143)	(0.0142)	(0.0049)	(0.0137)	(0.0137)	(0.0055)	(0.0137)	(0.0137)
Lisbon and	0.1807***	0.1759***	0.1738***	0.1664***	0.1598***	0.1593***	0.1686***	0.1641***	0.1631***
Tagus Valley	(0.0034)	(0.0159)	(0.0158)	(0.0038)	(0.0157)	(0.0157)	(0.0043)	(0.0154)	(0.0154)
Inland	-0.0903***	-0.0878***	-0.0881***	-0.1026***	-0.0996***	-0.1008***	-0.0982***	-0.0956***	-0.0967***
	(0.0055)	(0.0152)	(0.0151)	(0.0061)	(0.0159)	(0.0159)	(0.0068)	(0.0152)	(0.0152)
Algarve	0.0158*	0.0128	0.0121	0.0263***	0.0212	0.0243	0.0091	0.0059	0.0046
	(0.0083)	(0.0192)	(0.0190)	(0.0090)	(0.0205)	(0.0203)	(0.0099)	(0.0195)	(0.0195)
Observations	92,529	92,529	92,529	78,654	78,654	78,654	65,242	65,242	65,242
R–squared	0.584	0.580	0.581	0.571	0.564	0.564	0.569	0.567	0.566
RMSE	0.397	0.398	0.398	0.411	0.414	0.414	0.424	0.426	0.426
Hansen J statistic			4.932			6.622			11.41
p–value of HJ statistic			0.553			0.250			0.022

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include a set of dummies to control for occupation and sector of activity. North is the omitted category in region.

Table A.10: OLS and IV estimates using q_3 and $q_3 \times year - of - birth$ as instruments, Men born 1950–1966

	1986			**1987**			**1988**			**1989**		
VARIABLES	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$
Education	0.0378***	0.0497***	0.0486***	0.0403***	0.0455***	0.0454***	0.0426***	0.0525***	0.0521***	0.0445***	0.0542***	0.0519***
	(0.0003)	(0.0046)	(0.0044)	(0.0003)	(0.0047)	(0.0044)	(0.0003)	(0.0050)	(0.0047)	(0.0003)	(0.0050)	(0.0049)
Experience	0.0599***	0.0827***	0.0802***	0.0632***	0.0732***	0.0719***	0.0554***	0.0746***	0.0749***	0.0426***	0.0607***	0.0568***
	(0.0023)	(0.0093)	(0.0090)	(0.0021)	(0.0096)	(0.0091)	(0.0021)	(0.0104)	(0.0096)	(0.0020)	(0.0102)	(0.0097)
Experience sq	-0.0029***	-0.0030***	-0.0030***	-0.0029***	-0.0029***	-0.0029***	-0.0025***	-0.0025***	-0.0025***	-0.0021***	-0.0022***	-0.0021***
	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0002)
Centre	-0.0190***	-0.0204**	-0.0159*	-0.0217***	-0.0223**	-0.0220**	-0.0122***	-0.0133	-0.0109	0.0042**	0.0033	0.0046
	(0.0017)	(0.0087)	(0.0085)	(0.0016)	(0.0092)	(0.0089)	(0.0017)	(0.0089)	(0.0088)	(0.0017)	(0.0089)	(0.0088)
Lisbon and	0.1251***	0.1195***	0.1205***	0.1190***	0.1165***	0.1147***	0.1207***	0.1158***	0.1126***	0.1272***	0.1224***	0.1216***
Tagus Valley	(0.0013)	(0.0120)	(0.0118)	(0.0013)	(0.0122)	(0.0118)	(0.0013)	(0.0114)	(0.0111)	(0.0014)	(0.0109)	(0.0107)
Inland	-0.0516***	-0.0531***	-0.0510***	-0.0700***	-0.0707***	-0.0699***	-0.0857***	-0.0871***	-0.0864***	-0.0759***	-0.0770***	-0.0742***
	(0.0023)	(0.0128)	(0.0127)	(0.0021)	(0.0105)	(0.0104)	(0.0021)	(0.0092)	(0.0091)	(0.0021)	(0.0101)	(0.0100)
Algarve	0.0781***	0.0732***	0.0770***	0.0750***	0.0727***	0.0760***	0.1220***	0.1178***	0.1182***	0.1185***	0.1148***	0.1172***
	(0.0043)	(0.0181)	(0.0180)	(0.0041)	(0.0176)	(0.0174)	(0.0040)	(0.0192)	(0.0191)	(0.0041)	(0.0168)	(0.0167)
Observations	275,909	275,909	275,909	294,018	294,018	294,018	300,608	300,608	300,608	327,995	327,995	327,995
R–squared	0.530	0.527	0.528	0.530	0.529	0.529	0.511	0.510	0.509	0.489	0.487	0.487
RMSE	0.271	0.271	0.271	0.273	0.273	0.273	0.282	0.283	0.283	0.304	0.305	0.305
Hansen J stat.			16.13			19.26			10.22			21.92
p–value of HJ			0.444			0.255			0.855			0.146

Continued on next page

Table A.10 – continued from previous page

	1991			1992			1993			1994		
VARIABLES	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$
Education	0.0478***	0.0579***	0.0581***	0.0484***	0.0539***	0.0529***	0.0486***	0.0683***	0.0663***	0.0453***	0.0534***	0.0556***
	(0.0004)	(0.0052)	(0.0051)	(0.0004)	(0.0052)	(0.0051)	(0.0004)	(0.0055)	(0.0058)	(0.0004)	(0.0057)	(0.0059)
Experience	0.0065***	0.0235**	0.0261***	-0.0013	0.0076	0.0073	-0.0083***	0.0225**	0.0206**	-0.0200***	-0.0082	-0.0046
	(0.0019)	(0.0103)	(0.0100)	(0.0019)	(0.0097)	(0.0095)	(0.0020)	(0.0101)	(0.0103)	(0.0020)	(0.0097)	(0.0099)
Experience sq	-0.0011***	-0.0011***	-0.0011***	-0.0009***	-0.0009***	-0.0009***	-0.0007***	-0.0006***	-0.0006***	-0.0005***	-0.0005***	-0.0004***
	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Centre	-0.0008	-0.0015	0.0000	0.0135***	0.0130	0.0163*	0.0086***	0.0068	0.0088	0.0026	0.0020	0.0028
	(0.0020)	(0.0097)	(0.0094)	(0.0020)	(0.0100)	(0.0099)	(0.0021)	(0.0104)	(0.0104)	(0.0021)	(0.0100)	(0.0100)
Lisbon and	0.1347***	0.1295***	0.1298***	0.1452***	0.1424***	0.1483***	0.1539***	0.1437***	0.1465***	0.1472***	0.1433***	0.1412***
Tagus Valley	(0.0015)	(0.0139)	(0.0134)	(0.0016)	(0.0138)	(0.0136)	(0.0017)	(0.0148)	(0.0147)	(0.0016)	(0.0133)	(0.0131)
Inland	-0.0823***	-0.0830***	-0.0829***	-0.0824***	-0.0830***	-0.0813***	-0.0886***	-0.0907***	-0.0866***	-0.0796***	-0.0805***	-0.0790***
	(0.0024)	(0.0135)	(0.0132)	(0.0025)	(0.0124)	(0.0122)	(0.0027)	(0.0135)	(0.0133)	(0.0027)	(0.0129)	(0.0127)
Algarve	0.0693***	0.0659***	0.0671***	0.0937***	0.0919***	0.0946***	0.0961***	0.0885***	0.0916***	0.0508***	0.0479***	0.0474***
	(0.0044)	(0.0175)	(0.0173)	(0.0045)	(0.0179)	(0.0178)	(0.0048)	(0.0180)	(0.0181)	(0.0047)	(0.0165)	(0.0165)
Observations	343,191	343,191	343,191	344,325	344,325	344,325	339,556	339,556	339,556	341,560	341,560	341,560
R–squared	0.496	0.495	0.494	0.491	0.491	0.490	0.478	0.473	0.472	0.506	0.505	0.422
RMSE	0.347	0.347	0.347	0.360	0.360	0.360	0.376	0.377	0.378	0.381	0.381	0.412
Hansen J stat.			5.028			22.03			22.85			12.13
p–value of HJ			0.996			0.142			0.118			0.735

Continued on next page

Table A.10 – continued from previous page

VARIABLES	1995			1996			1997			1998		
	OLS	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$
Education	0.0440***	0.0524***	0.0516***	0.0411***	0.0634***	0.0619***	0.0392***	0.0546***	0.0541***	0.0391***	0.0598***	0.0602***
	(0.0004)	(0.0059)	(0.0058)	(0.0004)	(0.0062)	(0.0061)	(0.0004)	(0.0060)	(0.0059)	(0.0004)	(0.0058)	(0.0057)
Experience	-0.0207***	-0.0098	-0.0096	-0.0240***	0.0039	0.0032	-0.0284***	-0.0094	-0.0093	-0.0352***	-0.0106	-0.0097
	(0.0017)	(0.0085)	(0.0083)	(0.0017)	(0.0087)	(0.0086)	(0.0017)	(0.0084)	(0.0083)	(0.0017)	(0.0080)	(0.0078)
Experience sq	-0.0003***	-0.0002*	-0.0002**	-0.0003***	-0.0001	-0.0001	-0.0001***	-0.0000	-0.0000	-0.0001	0.0001	0.0001
	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)
Centre	0.0000	-0.0006	0.0004	0.0038*	0.0026	0.0054	0.0174***	0.0159	0.0151	0.0079***	0.0064	0.0071
	(0.0020)	(0.0104)	(0.0103)	(0.0020)	(0.0103)	(0.0101)	(0.0020)	(0.0100)	(0.0099)	(0.0019)	(0.0097)	(0.0095)
Lisbon and	0.1495***	0.1455***	0.1461***	0.1471***	0.1368***	0.1404***	0.1566***	0.1488***	0.1467***	0.1530***	0.1430***	0.1411***
Tagus Valley	(0.0015)	(0.0132)	(0.0129)	(0.0016)	(0.0136)	(0.0133)	(0.0015)	(0.0112)	(0.0110)	(0.0015)	(0.0106)	(0.0104)
Inland	-0.0932***	-0.0939***	-0.0933***	-0.0840***	-0.0858***	-0.0841***	-0.0843***	-0.0858***	-0.0874***	-0.0764***	-0.0778***	-0.0758***
	(0.0025)	(0.0124)	(0.0122)	(0.0026)	(0.0149)	(0.0146)	(0.0025)	(0.0131)	(0.0129)	(0.0024)	(0.0119)	(0.0116)
Algarve	0.0562***	0.0536***	0.0533***	0.0642***	0.0577***	0.0625***	0.0645***	0.0598***	0.0595***	0.0538***	0.0482***	0.0483***
	(0.0044)	(0.0180)	(0.0179)	(0.0045)	(0.0167)	(0.0165)	(0.0043)	(0.0156)	(0.0156)	(0.0042)	(0.0149)	(0.0148)
Observations	375,529	375,529	375,529	368,409	368,409	368,409	383,575	383,575	383,575	381,686	381,686	381,686
R–squared	0.572	0.571	0.571	0.562	0.558	0.559	0.565	0.563	0.563	0.587	0.583	0.583
RMSE	0.374	0.374	0.374	0.379	0.381	0.381	0.375	0.375	0.375	0.368	0.369	0.369
Hansen J stat.			13.47			17.61			11.35			13.68
p–value of HJ			0.638			0.347			0.787			0.623

Continued on next page

Table A.10 – continued from previous page

	1999			2000			2002			2003		
VARIABLES	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$
Education	0.0388***	0.0582***	0.0573***	0.0361***	0.0594***	0.0582***	0.0318***	0.0630***	0.0633***	0.0315***	0.0610***	0.0618***
	(0.0058)	(0.0057)	(0.0004)	(0.0058)	(0.0057)	(0.0004)	(0.0058)	(0.0057)	(0.0004)	(0.0060)	(0.0059)	(0.0060)
Experience	-0.0468***	-0.0242***	-0.0255***	-0.0447***	-0.0184**	-0.0175**	-0.0544***	-0.0208***	-0.0200**	-0.0603***	-0.0296***	-0.0298***
	(0.0018)	(0.0079)	(0.0078)	(0.0019)	(0.0085)	(0.0083)	(0.0021)	(0.0079)	(0.0078)	(0.0022)	(0.0082)	(0.0081)
Experience sq	0.0001	0.0002**	0.0002**	0.0001	0.0003**	0.0002*	0.0001***	0.0004***	0.0004***	0.0001**	0.0004***	0.0004***
	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)
Centre	0.0186***	0.0171*	0.0166*	0.0283***	0.0262***	0.0259***	0.0313***	0.0289***	0.0334***	0.0261***	0.0234**	0.0318***
	(0.0019)	(0.0101)	(0.0099)	(0.0019)	(0.0101)	(0.0100)	(0.0020)	(0.0108)	(0.0106)	(0.0021)	(0.0110)	(0.0107)
Lisbon and	0.1562***	0.1471***	0.1448***	0.1762***	0.1654***	0.1640***	0.1749***	0.1611***	0.1671***	0.1686***	0.1554***	0.1636***
Tagus Valley	(0.0015)	(0.0109)	(0.0108)	(0.0015)	(0.0122)	(0.0119)	(0.0016)	(0.0117)	(0.0113)	(0.0017)	(0.0119)	(0.0116)
Inland	-0.0807***	-0.0823***	-0.0832***	-0.0741***	-0.0762***	-0.0765***	-0.0741***	-0.0757***	-0.0707***	-0.0776***	-0.0805***	-0.0718***
	(0.0024)	(0.0117)	(0.0115)	(0.0024)	(0.0122)	(0.0120)	(0.0025)	(0.0118)	(0.0114)	(0.0026)	(0.0123)	(0.0120)
Algarve	0.0548***	0.0491***	0.0501***	0.0651***	0.0569***	0.0558***	0.0367***	0.0243*	0.0290**	0.0470***	0.0335**	0.0428***
	(0.0042)	(0.0141)	(0.0141)	(0.0040)	(0.0147)	(0.0146)	(0.0039)	(0.0147)	(0.0144)	(0.0039)	(0.0154)	(0.0151)
Observations	395,205	395,205	395,205	391,504	391,504	391,504	377,627	377,627	377,627	375,745	375,745	375,745
R–squared	0.591	0.588	0.588	0.606	0.602	0.602	0.595	0.587	0.587	0.579	0.572	0.571
RMSE	0.373	0.375	0.375	0.371	0.373	0.373	0.385	0.389	0.389	0.396	0.400	0.400
Hansen J stat.			20.30			16.97			13.44			22.10
p–value of HJ			0.207			0.388			0.640			0.140

Continued on next page

Table A.10 – continued from previous page

VARIABLES	2004			2005			2006		
	OLS	IV q_3	IV q_3, $q_3 \times yb$	OLS	IV q_3	IV q_3, $q_3 \times yb$	OLS	IV q_3	IV q_3, $q_3 \times yb$
Education	0.0310***	0.0647***	0.0639***	0.0305***	0.0721***	0.0720***	0.0279***	0.0690***	0.0692***
	(0.0004)	(0.0061)	(0.0060)	(0.0004)	(0.0061)	(0.0060)	(0.0004)	(0.0062)	(0.0061)
Experience	-0.0699***	-0.0365***	-0.0372***	-0.0758***	-0.0357***	-0.0368***	-0.0768***	-0.0393***	-0.0402***
	(0.0023)	(0.0081)	(0.0080)	(0.0024)	(0.0077)	(0.0076)	(0.0025)	(0.0081)	(0.0080)
Experience sq	0.0000	0.0004***	0.0004***	0.0001*	0.0005***	0.0006***	0.0000	0.0005***	0.0006***
	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)
Centre	0.0247***	0.0219**	0.0246**	0.0170***	0.0138	0.0150	0.0221***	0.0180*	0.0185*
	(0.0021)	(0.0111)	(0.0109)	(0.0022)	(0.0112)	(0.0110)	(0.0022)	(0.0105)	(0.0104)
Lisbon and	0.1731***	0.1585***	0.1619***	0.1764***	0.1586***	0.1595***	0.1861***	0.1684***	0.1679***
Tagus Valley	(0.0017)	(0.0117)	(0.0114)	(0.0017)	(0.0122)	(0.0118)	(0.0018)	(0.0113)	(0.0111)
Inland	-0.0801***	-0.0834***	-0.0814***	-0.0754***	-0.0795***	-0.0785***	-0.0732***	-0.0784***	-0.0789***
	(0.0026)	(0.0122)	(0.0120)	(0.0026)	(0.0125)	(0.0122)	(0.0027)	(0.0119)	(0.0117)
Algarve	0.0285***	0.0123	0.0158	0.0147***	-0.0066	-0.0048	0.0294***	0.0090	0.0099
	(0.0040)	(0.0149)	(0.0146)	(0.0040)	(0.0147)	(0.0144)	(0.0040)	(0.0138)	(0.0137)
Observations	385,692	385,692	385,692	394,777	394,777	394,777	386,044	386,044	386,044
R–squared	0.577	0.568	0.568	0.565	0.551	0.551	0.563	0.549	0.549
RMSE	0.410	0.414	0.414	0.420	0.426	0.426	0.423	0.429	0.429
Hansen J statistic			16.79			10.60			9.572
p–value of HJ statistic			0.400			0.833			0.888

Continued on next page

Table A.10 – continued from previous page

VARIABLES	2007			2008			2009		
	OLS	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$
Education	0.0259***	0.0680***	0.0677***	0.0230***	0.0635***	0.0628***	0.0222***	0.0660***	0.0665***
	(0.0004)	(0.0063)	(0.0062)	(0.0004)	(0.0066)	(0.0065)	(0.0004)	(0.0069)	(0.0068)
Experience	-0.0889***	-0.0524***	-0.0538***	-0.0870***	-0.0542***	-0.0556***	-0.0843***	-0.0512***	-0.0516***
	(0.0026)	(0.0081)	(0.0080)	(0.0027)	(0.0081)	(0.0080)	(0.0028)	(0.0080)	(0.0080)
Experience sq	0.0002***	0.0007***	0.0007***	0.0001***	0.0006***	0.0007***	0.0001**	0.0007***	0.0007***
	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)
Centre	0.0182***	0.0143	0.0178*	0.0267***	0.0231**	0.0251***	0.0176***	0.0134	0.0137
	(0.0022)	(0.0107)	(0.0105)	(0.0023)	(0.0097)	(0.0096)	(0.0023)	(0.0103)	(0.0102)
Lisbon and	0.1922***	0.1750***	0.1777***	0.1881***	0.1715***	0.1723***	0.1852***	0.1672***	0.1637***
Tagus Valley	(0.0018)	(0.0121)	(0.0117)	(0.0018)	(0.0108)	(0.0106)	(0.0018)	(0.0118)	(0.0115)
Inland	-0.0762***	-0.0808***	-0.0819***	-0.0677***	-0.0721***	-0.0739***	-0.0691***	-0.0734***	-0.0762***
	(0.0027)	(0.0126)	(0.0124)	(0.0027)	(0.0116)	(0.0115)	(0.0028)	(0.0119)	(0.0117)
Algarve	0.0349***	0.0128	0.0156	0.0402***	0.0192	0.0206	0.0304***	0.0099	0.0100
	(0.0040)	(0.0140)	(0.0138)	(0.0040)	(0.0130)	(0.0130)	(0.0043)	(0.0137)	(0.0136)
Observations	383,973	383,973	383,973	375,137	375,137	375,137	351,151	351,151	351,151
R–squared	0.552	0.538	0.538	0.550	0.536	0.537	0.559	0.544	0.543
RMSE	0.427	0.434	0.433	0.425	0.432	0.431	0.426	0.434	0.434
Hansen J statistic			10.88			19.98			16.81
p–value of HJ statistic			0.817			0.221			0.398

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.

Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include a set of dummies to control for occupation and sector of activity. North is the omitted category in region.

Table A.11: OLS and IV estimates using q_3 and $q_3 \times year - of - birth$ as instruments, Men born 1967–1980

	1992			1993			1994			1995		
VARIABLES	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$
Education	0.0292***	0.0455***	0.0441***	0.0283***	0.0463***	0.0452***	0.0247***	0.0422***	0.0414***	0.0225***	0.0496***	0.0479***
	(0.0006)	(0.0065)	(0.0063)	(0.0006)	(0.0067)	(0.0069)	(0.0005)	(0.0060)	(0.0059)	(0.0004)	(0.0058)	(0.0056)
Experience	0.0550***	0.0851***	0.0832***	0.0552***	0.0862***	0.0817***	0.0521***	0.0822***	0.0807***	0.0333***	0.0746***	0.0716***
	(0.0050)	(0.0137)	(0.0134)	(0.0045)	(0.0137)	(0.0149)	(0.0036)	(0.0113)	(0.0111)	(0.0024)	(0.0098)	(0.0094)
Experience sq	-0.0052***	-0.0051***	-0.0051***	-0.0055***	-0.0052***	-0.0050***	-0.0051***	-0.0049***	-0.0049***	-0.0042***	-0.0038***	-0.0038***
	(0.0004)	(0.0006)	(0.0006)	(0.0003)	(0.0005)	(0.0006)	(0.0002)	(0.0003)	(0.0003)	(0.0001)	(0.0003)	(0.0003)
Centre	0.0396***	0.0377***	0.0386***	0.0353***	0.0326***	0.0332***	0.0207***	0.0184**	0.0175**	0.0303***	0.0273***	0.0269***
	(0.0029)	(0.0070)	(0.0070)	(0.0030)	(0.0079)	(0.0078)	(0.0028)	(0.0083)	(0.0082)	(0.0023)	(0.0088)	(0.0087)
Lisbon and	0.1186***	0.1101***	0.1131***	0.1324***	0.1226***	0.1246***	0.1290***	0.1195***	0.1184***	0.1327***	0.1176***	0.1202***
Tagus Valley	(0.0025)	(0.0104)	(0.0103)	(0.0025)	(0.0120)	(0.0118)	(0.0023)	(0.0126)	(0.0123)	(0.0018)	(0.0123)	(0.0120)
Inland	-0.0411***	-0.0428***	-0.0407***	-0.0452***	-0.0478***	-0.0465***	-0.0555***	-0.0578***	-0.0580***	-0.0587***	-0.0607***	-0.0599***
	(0.0037)	(0.0077)	(0.0077)	(0.0038)	(0.0101)	(0.0100)	(0.0035)	(0.0093)	(0.0092)	(0.0029)	(0.0092)	(0.0091)
Algarve	0.1158***	0.1102***	0.1129***	0.1239***	0.1168***	0.1186***	0.0780***	0.0702***	0.0694***	0.0695***	0.0577***	0.0597***
	(0.0066)	(0.0152)	(0.0151)	(0.0068)	(0.0153)	(0.0152)	(0.0061)	(0.0147)	(0.0146)	(0.0049)	(0.0150)	(0.0148)
Observations	102,047	102,047	102,047	120,693	120,693	120,693	150,832	150,832	150,832	194,573	194,573	194,573
R–squared	0.320	0.315	0.316	0.327	0.322	0.317	0.360	0.355	0.356	0.478	0.468	0.469
RMSE	0.310	0.311	0.311	0.340	0.341	0.342	0.346	0.347	0.347	0.319	0.322	0.322
Hansen J stat.			14.40			7.763			8.150			10.93
p–value of HJ			0.072			0.354			0.700			0.535

Continued on next page

Table A.11 – continued from previous page

	1996			1997			1998			1999		
VARIABLES	**OLS**	**IV** q_3	**IV OLS** q_3, $q_3 \times yb$	**IV**	**IV** q_3	**OLS** q_3, $q_3 \times yb$	**IV**	**IV** q_3	**OLS** q_3, $q_3 \times yb$	**IV**	**IV** q_3	q_3, $q_3 \times yb$
Eduaction	0.0206***	0.0521***	0.0509***	0.0215***	0.0552***	0.0550***	0.0201***	0.0523***	0.0519***	0.0188***	0.0588***	0.0577***
	(0.0004)	(0.0060)	(0.0059)	(0.0004)	(0.0053)	(0.0052)	(0.0004)	(0.0051)	(0.0049)	(0.0003)	(0.0053)	(0.0051)
Experience	0.0296***	0.0769***	0.0745***	0.0304***	0.0792***	0.0786***	0.0208***	0.0654***	0.0631***	0.0122***	0.0657***	0.0638***
	(0.0021)	(0.0095)	(0.0093)	(0.0017)	(0.0082)	(0.0080)	(0.0015)	(0.0075)	(0.0072)	(0.0013)	(0.0074)	(0.0072)
Experience sq	-0.0039***	-0.0035***	-0.0035***	-0.0036***	-0.0032***	-0.0032***	-0.0034***	-0.0029***	-0.0028***	-0.0031***	-0.0026***	-0.0026***
	(0.0001)	(0.0003)	(0.0003)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0002)
Centre	0.0230***	0.0199**	0.0199**	0.0359***	0.0324***	0.0318***	0.0211***	0.0188***	0.0205***	0.0291***	0.0247***	0.0260***
	(0.0022)	(0.0094)	(0.0094)	(0.0020)	(0.0077)	(0.0077)	(0.0019)	(0.0068)	(0.0067)	(0.0018)	(0.0064)	(0.0063)
Lisbon and	0.1379***	0.1201***	0.1217***	0.1388***	0.1175***	0.1168***	0.1339***	0.1141***	0.1103***	0.1329***	0.1099***	0.1121***
Tagus Valley	(0.0018)	(0.0129)	(0.0128)	(0.0016)	(0.0106)	(0.0105)	(0.0015)	(0.0104)	(0.0101)	(0.0015)	(0.0089)	(0.0088)
Inland	-0.0561***	-0.0598***	-0.0591***	-0.0579***	-0.0622***	-0.0625***	-0.0595***	-0.0628***	-0.0609***	-0.0671***	-0.0713***	-0.0688***
	(0.0028)	(0.0106)	(0.0105)	(0.0025)	(0.0086)	(0.0085)	(0.0024)	(0.0089)	(0.0089)	(0.0022)	(0.0085)	(0.0084)
Algarve	0.0814***	0.0675***	0.0687***	0.0801***	0.0654***	0.0642***	0.0794***	0.0664***	0.0636***	0.0696***	0.0526***	0.0533***
	(0.0048)	(0.0149)	(0.0149)	(0.0042)	(0.0134)	(0.0133)	(0.0039)	(0.0128)	(0.0127)	(0.0037)	(0.0120)	(0.0119)
Observations	213,515	213,515	213,515	263,581	263,581	263,581	283,475	283,475	283,475	322,209	322,209	322,209
R–squared	0.472	0.458	0.459	0.491	0.475	0.475	0.520	0.506	0.506	0.537	0.516	0.517
RMSE	0.330	0.334	0.334	0.322	0.327	0.327	0.316	0.321	0.321	0.321	0.328	0.327
Hansen J stat.			9.806			7.015			21.33			21.20
p–value of HJ			0.710			0.901			0.067			0.069

Continued on next page

Table A.11 – continued from previous page

VARIABLES	2000			2002			2003		
	OLS	IV	IV OLS	IV	IV	OLS	IV	IV	
		q_3	q_3, $q_3 \times yb$		q_3	q_3, $q_3 \times yb$		q_3	q_3, $q_3 \times yb$
Education	0.0184***	0.0587***	0.0575***	0.0200***	0.0605***	0.0585***	0.0204***	0.0567***	0.0564***
	(0.0003)	(0.0051)	(0.0050)	(0.0003)	(0.0052)	(0.0051)	(0.0003)	(0.0053)	(0.0053)
Experience	0.0096***	0.0609***	0.0595***	0.0174***	0.0650***	0.0621***	0.0184***	0.0593***	0.0585***
	(0.0012)	(0.0069)	(0.0068)	(0.0011)	(0.0064)	(0.0063)	(0.0011)	(0.0063)	(0.0062)
Experience sq	-0.0029***	-0.0024***	-0.0024***	-0.0030***	-0.0024***	-0.0023***	-0.0028***	-0.0023***	-0.0023***
	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)
Centre	0.0415***	0.0363***	0.0362***	0.0404***	0.0364***	0.0370***	0.0351***	0.0307***	0.0318***
	(0.0018)	(0.0067)	(0.0066)	(0.0018)	(0.0064)	(0.0064)	(0.0018)	(0.0072)	(0.0071)
Lisbon and	0.1519***	0.1289***	0.1301***	0.1516***	0.1315***	0.1326***	0.1503***	0.1333***	0.1330***
Tagus Valley	(0.0014)	(0.0090)	(0.0089)	(0.0014)	(0.0084)	(0.0083)	(0.0014)	(0.0088)	(0.0088)
Inland	-0.0506***	-0.0552***	-0.0549***	-0.0529***	-0.0567***	-0.0552***	-0.0527***	-0.0570***	-0.0558***
	(0.0021)	(0.0087)	(0.0086)	(0.0021)	(0.0076)	(0.0075)	(0.0022)	(0.0088)	(0.0088)
Algarve	0.0808***	0.0632***	0.0619***	0.0698***	0.0550***	0.0556***	0.0745***	0.0598***	0.0590***
	(0.0035)	(0.0113)	(0.0113)	(0.0032)	(0.0105)	(0.0104)	(0.0032)	(0.0111)	(0.0110)
Observations	347,462	347,462	347,462	385,309	385,309	385,309	399,331	399,331	399,331
R–squared	0.549	0.529	0.530	0.549	0.532	0.533	0.540	0.526	0.526
RMSE	0.322	0.329	0.329	0.338	0.345	0.344	0.348	0.353	0.353
Hansen J statistic			17.14			15.79			10.73
p–value of HJ statistic			0.193			0.261			0.633

Continued on next page

Table A.11 – continued from previous page

VARIABLES	2004			2005			2006		
	OLS	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$	**OLS**	**IV** q_3	**IV** q_3, $q_3 \times yb$
Education	0.0202***	0.0583***	0.0579***	0.0203***	0.0601***	0.0594***	0.0190***	0.0743***	0.0741***
	(0.0003)	(0.0054)	(0.0053)	(0.0003)	(0.0053)	(0.0053)	(0.0004)	(0.0056)	(0.0056)
Experience	0.0194***	0.0604***	0.0601***	0.0232***	0.0640***	0.0632***	0.0195***	0.0731***	0.0729***
	(0.0011)	(0.0060)	(0.0060)	(0.0011)	(0.0060)	(0.0060)	(0.0011)	(0.0058)	(0.0058)
Experience sq	-0.0029***	-0.0023***	-0.0023***	-0.0030***	-0.0024***	-0.0023***	-0.0027***	-0.0018***	-0.0018***
	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)
Centre	0.0336***	0.0286***	0.0296***	0.0266***	0.0220***	0.0230***	0.0337***	0.0277***	0.0288***
	(0.0018)	(0.0075)	(0.0074)	(0.0018)	(0.0075)	(0.0075)	(0.0018)	(0.0075)	(0.0075)
Lisbon and	0.1494***	0.1322***	0.1328***	0.1520***	0.1354***	0.1352***	0.1635***	0.1419***	0.1413***
Tagus Valley	(0.0014)	(0.0088)	(0.0088)	(0.0014)	(0.0088)	(0.0088)	(0.0014)	(0.0084)	(0.0083)
Inland	-0.0499***	-0.0558***	-0.0541***	-0.0513***	-0.0578***	-0.0572***	-0.0522***	-0.0616***	-0.0619***
	(0.0022)	(0.0092)	(0.0092)	(0.0022)	(0.0101)	(0.0100)	(0.0022)	(0.0096)	(0.0096)
Algarve	0.0665***	0.0500***	0.0491***	0.0599***	0.0435***	0.0422***	0.0659***	0.0439***	0.0438***
	(0.0032)	(0.0126)	(0.0126)	(0.0032)	(0.0123)	(0.0123)	(0.0032)	(0.0104)	(0.0104)
Observations	419,834	419,834	419,834	449,972	449,972	449,972	453,254	453,254	453,254
R–squared	0.542	0.529	0.529	0.533	0.519	0.520	0.530	0.504	0.505
RMSE	0.358	0.364	0.364	0.370	0.375	0.375	0.376	0.387	0.387
Hansen J statistic			11.03			14.93			15.42
p–value of HJ statistic			0.609			0.312			0.282

Continued on next page

Table A.11 – continued from previous page

VARIABLES	2007			2008			2009		
	OLS	IV q_3	IV q_3, $q_3 \times yb$	OLS	IV q_3	IV q_3, $q_3 \times yb$	OLS	IV q_3	IV q_3, $q_3 \times yb$
Education	0.0184***	0.0662***	0.0647***	0.0160***	0.0598***	0.0592***	0.0153***	0.0656***	0.0653***
	(0.0004)	(0.0057)	(0.0057)	(0.0004)	(0.0056)	(0.0056)	(0.0004)	(0.0060)	(0.0060)
Experience	0.0207***	0.0646***	0.0627***	0.0172***	0.0555***	0.0546***	0.0142***	0.0560***	0.0555***
	(0.0011)	(0.0058)	(0.0057)	(0.0011)	(0.0055)	(0.0054)	(0.0012)	(0.0056)	(0.0056)
Experience sq	-0.0027***	-0.0019***	-0.0019***	-0.0025***	-0.0018***	-0.0018***	-0.0023***	-0.0015***	-0.0015***
	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)
Centre	0.0271***	0.0223***	0.0224***	0.0266***	0.0229***	0.0223***	0.0205***	0.0161**	0.0167**
	(0.0019)	(0.0081)	(0.0080)	(0.0019)	(0.0080)	(0.0080)	(0.0019)	(0.0081)	(0.0081)
Lisbon and	0.1616***	0.1441***	0.1439***	0.1549***	0.1395***	0.1380***	0.1481***	0.1307***	0.1297***
Tagus Valley	(0.0014)	(0.0095)	(0.0094)	(0.0014)	(0.0098)	(0.0096)	(0.0015)	(0.0102)	(0.0101)
Inland	-0.0615***	-0.0690***	-0.0687***	-0.0611***	-0.0674***	-0.0684***	-0.0596***	-0.0676***	-0.0671***
	(0.0022)	(0.0104)	(0.0104)	(0.0023)	(0.0108)	(0.0107)	(0.0023)	(0.0104)	(0.0103)
Algarve	0.0707***	0.0518***	0.0512***	0.0708***	0.0541***	0.0531***	0.0624***	0.0439***	0.0435***
	(0.0032)	(0.0126)	(0.0126)	(0.0032)	(0.0125)	(0.0124)	(0.0034)	(0.0131)	(0.0131)
Observations	461,434	461,434	461,434	465,057	465,057	465,057	448,227	448,227	448,227
R–squared	0.522	0.503	0.504	0.521	0.506	0.507	0.528	0.509	0.509
RMSE	0.386	0.393	0.393	0.387	0.393	0.393	0.389	0.396	0.396
Hansen J statistic			21.42			9.382			5.850
p–value of HJ statistic			0.065			0.743			0.951

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include a set of dummies to control for occupation and sector of activity. North is the omitted category in region.

Table A.12: IV estimates using q_1 and $q_1 \times year - of - birth$ as instruments, Men born 1930–1949

	1986		**1987**		**1988**		**1989**		**1991**		**1992**	
VARIABLES	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$
Education	0.0673***	0.0639***	0.0651***	0.0604***	0.0667***	0.0644***	0.0816	0.0697***	0.0710***	0.0701***	0.0701***	0.0692***
	(0.0045)	(0.0037)	(0.0042)	(0.0034)	(0.0042)	(0.0036)	(5.1494)	(0.0035)	(0.0037)	(0.0033)	(0.0042)	(0.0042)
Experience	0.0410***	0.0302***	-0.0120	0.0243***	0.0322***	0.0285***	-0.0129	0.0165*	0.0215*	0.0221**	0.0323***	0.0276**
	(0.0097)	(0.0086)	(19.0785)	(0.0083)	(0.0098)	(0.0089)	(14.5825)	(0.0094)	(0.0117)	(0.0106)	(0.0122)	(0.0117)
Experience sq	-0.0008***	-0.0006***	0.0004	-0.0006***	-0.0006***	-0.0006***	0.0004	-0.0003**	-0.0004**	-0.0004***	-0.0006***	-0.0005***
	(0.0002)	(0.0001)	(0.4459)	(0.0001)	(0.0001)	(0.0001)	(0.3250)	(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0002)
Centre	-0.0520***	-0.0500***	-0.0678	-0.0715***	-0.0598***	-0.0589***	-0.0434	-0.0436***	-0.0437**	-0.0455***	-0.0297**	-0.0271*
	(0.0170)	(0.0162)	(0.1571)	(0.0156)	(0.0162)	(0.0156)	(0.0268)	(0.0143)	(0.0173)	(0.0168)	(0.0144)	(0.0145)
Lisbon and	0.1586***	0.1558***	0.1574	0.1509***	0.1448***	0.1381***	0.1586	0.1482***	0.1488***	0.1455***	0.1638***	0.1672***
Tagus Valley	(0.0219)	(0.0206)	(0.2763)	(0.0199)	(0.0192)	(0.0179)	(0.1241)	(0.0178)	(0.0209)	(0.0198)	(0.0177)	(0.0173)
Inland	-0.0789***	-0.0768***	-0.1032	-0.1049***	-0.1161***	-0.1142***	-0.1007	-0.1036***	-0.1027***	-0.1030***	-0.0893***	-0.0836***
	(0.0173)	(0.0169)	(0.8123)	(0.0163)	(0.0149)	(0.0146)	(0.2274)	(0.0141)	(0.0153)	(0.0151)	(0.0149)	(0.0150)
Algarve	0.0380	0.0371	0.0439	0.0393	0.0945***	0.0955***	0.0829	0.0787***	0.0560**	0.0545**	0.0559**	0.0614**
	(0.0257)	(0.0252)	(1.2658)	(0.0257)	(0.0280)	(0.0276)	(0.6772)	(0.0225)	(0.0275)	(0.0270)	(0.0249)	(0.0247)
Observations	244,451	244,451	242,365	242,365	239,442	239,442	240,928	240,928	245,029	245,029	232,054	232,054
R–squared	0.527	0.525	0.530	0.526	0.518	0.511	0.510	0.507	0.548	0.548	0.551	0.476
RMSE	0.325	0.326	1.116	0.330	0.326	0.329	0.783	0.345	0.363	0.363	0.368	0.397
Hansen J stat.		20.44		19.50		18.30		15.11		13.55		11.16
p–value of HJ		0.369		0.426		0.503		0.715		0.758		0.848

Continued on next page

Table A.12 – continued from previous page

	1993		**1994**		**1995**		**1996**		**1997**		**1998**	
	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$
Education	0.0715***	0.0698***	0.0749***	0.0753***	0.0614***	0.0605***	0.0606***	0.0594***	0.0534***	0.0532***	0.0561***	0.0556***
	(0.0043)	(0.0038)	(0.0039)	(0.0037)	(0.0039)	(0.0038)	(0.0040)	(0.0039)	(0.0040)	(0.0039)	(0.0037)	(0.0036)
Experience	0.0270*	0.0219	0.0566***	0.0584***	0.0205*	0.0192	0.0242*	0.0236*	0.0081	0.0050	0.0196	0.0200
	(0.0155)	(0.0136)	(0.0162)	(0.0150)	(0.0124)	(0.0120)	(0.0135)	(0.0129)	(0.0163)	(0.0155)	(0.0179)	(0.0169)
Experience sq	-0.0005**	-0.0004**	-0.0008***	-0.0009***	-0.0003*	-0.0003*	-0.0003*	-0.0003*	-0.0002	-0.0002	-0.0004	-0.0004*
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Centre	-0.0353**	-0.0342**	-0.0377***	-0.0378***	-0.0448***	-0.0417***	-0.0468***	-0.0456***	-0.0292**	-0.0292**	-0.0286*	-0.0290**
	(0.0138)	(0.0135)	(0.0139)	(0.0137)	(0.0152)	(0.0150)	(0.0151)	(0.0149)	(0.0146)	(0.0146)	(0.0148)	(0.0147)
Lisbon and	0.1629***	0.1625***	0.1475***	0.1494***	0.1538***	0.1568***	0.1430***	0.1452***	0.1475***	0.1471***	0.1547***	0.1586***
Tagus Valley	(0.0177)	(0.0167)	(0.0145)	(0.0143)	(0.0162)	(0.0159)	(0.0156)	(0.0152)	(0.0148)	(0.0145)	(0.0150)	(0.0145)
Inland	-0.1131***	-0.1124***	-0.0892***	-0.0885***	-0.1027***	-0.0999***	-0.1086***	-0.1094***	-0.1051***	-0.1065***	-0.0923***	-0.0943***
	(0.0140)	(0.0139)	(0.0141)	(0.0140)	(0.0152)	(0.0150)	(0.0161)	(0.0160)	(0.0159)	(0.0158)	(0.0151)	(0.0150)
Algarve	0.0499**	0.0523**	0.0121	0.0159	0.0072	0.0093	0.0056	0.0071	0.0106	0.0131	0.0126	0.0144
	(0.0212)	(0.0211)	(0.0190)	(0.0189)	(0.0227)	(0.0225)	(0.0222)	(0.0221)	(0.0225)	(0.0224)	(0.0204)	(0.0203)
Observations	213,901	213,901	190,929	190,929	194,143	194,143	177,799	177,799	169,451	169,451	159,173	159,173
R-squared	0.535	0.535	0.558	0.558	0.597	0.596	0.587	0.587	0.589	0.589	0.607	0.607
RMSE	0.378	0.378	0.380	0.380	0.380	0.380	0.383	0.383	0.380	0.380	0.371	0.371
Hansen J stat.		13.29		8.354		7.846		7.342		9.132		7.505
p-value of HJ		0.652		0.870		0.854		0.834		0.610		0.677

Continued on next page

Table A.12 – continued from previous page

	1999		2000		2002		2003		2004	
	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$
Education	0.0614***	0.0612***	0.0560***	0.0553***	0.0580***	0.0574***	0.0639***	0.0645***	0.0670***	0.0680***
	(0.0039)	(0.0038)	(0.0042)	(0.0040)	(0.0048)	(0.0046)	(0.0051)	(0.0050)	(0.0053)	(0.0053)
Experience	0.0087	0.0141	0.0123	0.0207	-0.0048	-0.0008	-0.0413	-0.0174	-0.2016***	-0.1818***
	(0.0208)	(0.0200)	(0.0263)	(0.0248)	(0.0352)	(0.0332)	(0.0373)	(0.0362)	(0.0455)	(0.0445)
Experience sq	-0.0002	-0.0003	-0.0003	-0.0004	-0.0000	-0.0001	0.0005	0.0002	0.0025***	0.0022***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0005)	(0.0005)	(0.0006)	(0.0006)
Centre	-0.0272*	-0.0295**	-0.0203	-0.0201	-0.0022	-0.0023	-0.0014	0.0023	-0.0016	0.0009
	(0.0146)	(0.0145)	(0.0130)	(0.0128)	(0.0143)	(0.0142)	(0.0137)	(0.0137)	(0.0136)	(0.0136)
Lisbon and	0.1542***	0.1553***	0.1719***	0.1716***	0.1774***	0.1795***	0.1617***	0.1699***	0.1633***	0.1693***
Tagus Valley	(0.0154)	(0.0151)	(0.0168)	(0.0162)	(0.0159)	(0.0155)	(0.0156)	(0.0153)	(0.0152)	(0.0150)
Inland	-0.1026***	-0.1043***	-0.0908***	-0.0905***	-0.0885***	-0.0889***	-0.1004***	-0.0964***	-0.0951***	-0.0944***
	(0.0142)	(0.0141)	(0.0150)	(0.0149)	(0.0152)	(0.0150)	(0.0160)	(0.0158)	(0.0151)	(0.0150)
Algarve	0.0233	0.0244	0.0265	0.0270	0.0137	0.0157	0.0227	0.0295	0.0053	0.0077
	(0.0181)	(0.0180)	(0.0199)	(0.0199)	(0.0191)	(0.0189)	(0.0203)	(0.0202)	(0.0193)	(0.0193)
Observations	147,103	147,103	129,497	129,497	92,529	92,529	78,654	78,654	65,242	65,242
R-squared	0.600	0.600	0.615	0.615	0.582	0.582	0.567	0.567	0.566	0.565
RMSE	0.383	0.383	0.382	0.382	0.397	0.397	0.412	0.413	0.426	0.427
Hansen J statistic		10.02		11.98		4.645		13.04		6.401
p-value of HJ statistic		0.349		0.152		0.590		0.0230		0.171

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include a set of dummies to control for occupation and sector of activity. North is the omitted category in region.

Table A.13: IV estimates using q_1 and $q_1 \times year - of - birth$ as instruments, Men born 1950–1966

	1986		1987		1988		1989		1991		1992	
VARIABLES	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$
Education	0.0541***	0.0494***	0.0544***	0.0487***	0.0582***	0.0553***	0.0610***	0.0580***	0.0708***	0.0713***	0.0696***	0.0697***
	(0.0022)	(0.0021)	(0.0023)	(0.0020)	(0.0022)	(0.0020)	(0.0024)	(0.0023)	(0.0024)	(0.0023)	(0.0020)	(0.0020)
Experience	0.0910***	0.0728***	0.0903***	0.0697***	0.0857***	0.0736***	0.0735***	0.0563***	0.0453***	0.0442***	0.0330***	0.0322***
	(0.0055)	(0.0047)	(0.0058)	(0.0047)	(0.0057)	(0.0048)	(0.0060)	(0.0051)	(0.0069)	(0.0060)	(0.0057)	(0.0053)
Experience sq	-0.0030***	-0.0024***	-0.0030***	-0.0023***	-0.0026***	-0.0022***	-0.0022***	-0.0017***	-0.0011***	-0.0010***	-0.0008***	-0.0008***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0001)
Centre	-0.0209**	-0.0173**	-0.0233**	-0.0200**	-0.0139	-0.0123	0.0026	0.0035	-0.0024	-0.0017	0.0116	0.0102
	(0.0087)	(0.0085)	(0.0091)	(0.0090)	(0.0089)	(0.0088)	(0.0088)	(0.0087)	(0.0096)	(0.0093)	(0.0098)	(0.0097)
Lisbon and	0.1174***	0.1110***	0.1122***	0.1091***	0.1130***	0.1069***	0.1190***	0.1172***	0.1229***	0.1277***	0.1345***	0.1320***
Tagus Valley	(0.0117)	(0.0113)	(0.0119)	(0.0115)	(0.0111)	(0.0107)	(0.0105)	(0.0103)	(0.0136)	(0.0128)	(0.0134)	(0.0129)
Inland	-0.0536***	-0.0628***	-0.0720***	-0.0720***	-0.0879***	-0.0882***	-0.0777***	-0.0784***	-0.0840***	-0.0798***	-0.0848***	-0.0856***
	(0.0127)	(0.0125)	(0.0103)	(0.0103)	(0.0090)	(0.0090)	(0.0099)	(0.0099)	(0.0132)	(0.0129)	(0.0120)	(0.0119)
Algarve	0.0715***	0.0721***	0.0689***	0.0735***	0.1154***	0.1136***	0.1122***	0.1104***	0.0616***	0.0678***	0.0868***	0.0868***
	(0.0179)	(0.0178)	(0.0173)	(0.0171)	(0.0190)	(0.0188)	(0.0165)	(0.0164)	(0.0171)	(0.0169)	(0.0174)	(0.0173)
Observations	275,909	275,909	294,018	294,018	300,608	300,608	327,995	327,995	343,191	343,191	344,325	344,325
R-squared	0.525	0.526	0.527	0.527	0.507	0.508	0.484	0.486	0.489	0.489	0.486	0.486
RMSE	0.272	0.272	0.274	0.274	0.284	0.283	0.305	0.305	0.349	0.349	0.361	0.361
Hansen J stat.		59.62		53.32		46.49		46.85		29.70		14.51
p–value of HJ		6.5e-07		6.68e-06		8.17e-05		7.18e-05		0.020		0.561

Continued on next page

Table A.13 – continued from previous page

	1993		1994		1995		1996		1997		1998	
VARIABLES	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$
Education	0.0709***	0.0712***	0.0753***	0.0749***	0.0690***	0.0693***	0.0683***	0.0680***	0.0677***	0.0676***	0.0680***	0.0677***
	(0.0024)	(0.0023)	(0.0024)	(0.0023)	(0.0027)	(0.0026)	(0.0027)	(0.0026)	(0.0026)	(0.0025)	(0.0025)	(0.0024)
Experience	0.0264***	0.0267***	0.0238***	0.0233***	0.0116**	0.0141***	0.0102*	0.0115**	0.0067	0.0089*	-0.0008	0.0014
	(0.0063)	(0.0058)	(0.0061)	(0.0058)	(0.0051)	(0.0048)	(0.0052)	(0.0050)	(0.0051)	(0.0049)	(0.0048)	(0.0046)
Experience sq	-0.0006***	-0.0006***	-0.0003**	-0.0003**	-0.0001	-0.0001	-0.0001	-0.0001	0.0001	0.0000	0.0002*	0.0001
	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Centre	0.0066	0.0090	0.0004	0.0026	-0.0019	-0.0001	0.0024	0.0054	0.0146	0.0148	0.0057	0.0060
	(0.0104)	(0.0103)	(0.0098)	(0.0098)	(0.0102)	(0.0101)	(0.0102)	(0.0100)	(0.0098)	(0.0097)	(0.0096)	(0.0095)
Lisbon and	0.1425***	0.1442***	0.1325***	0.1367***	0.1375***	0.1383***	0.1344***	0.1390***	0.1423***	0.1419***	0.1390***	0.1386***
Tagus Valley	(0.0147)	(0.0138)	(0.0130)	(0.0126)	(0.0127)	(0.0124)	(0.0133)	(0.0128)	(0.0107)	(0.0105)	(0.0101)	(0.0100)
Inland	-0.0910***	-0.0888***	-0.0829***	-0.0825***	-0.0953***	-0.0926***	-0.0861***	-0.0833***	-0.0870***	-0.0879***	-0.0784***	-0.0776***
	(0.0134)	(0.0130)	(0.0124)	(0.0123)	(0.0121)	(0.0119)	(0.0148)	(0.0144)	(0.0128)	(0.0126)	(0.0118)	(0.0115)
Algarve	0.0875***	0.0887***	0.0398**	0.03446**	0.0485***	0.0520***	0.0563***	0.0603***	0.0559***	0.0565***	0.0460***	0.0469***
	(0.0178)	(0.0174)	(0.0160)	(0.0158)	(0.0175)	(0.0173)	(0.0165)	(0.0163)	(0.0153)	(0.0152)	(0.0147)	(0.0146)
Observations	339,556	339,556	341,560	341,560	375,529	375,529	368,409	368,409	383,575	383,575	381,686	381,686
R–squared	0.472	0.431	0.497	0.433	0.567	0.567	0.556	0.556	0.558	0.558	0.580	0.580
RMSE	0.378	0.392	0.384	0.408	0.376	0.376	0.382	0.382	0.378	0.378	0.371	0.371
Hansen J stat.		9.199		15.50		10.79		12.11		14.49		14.49
p–value of HJ		0.905		0.488		0.822		0.736		0.562		0.562

Continued on next page

Table A.13 – continued from previous page

	1999		2000		2002		2003		2004		2005	
VARIABLES	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$
Education	0.0690***	0.0686***	0.0660***	0.0665***	0.0617***	0.0615***	0.0662***	0.0656***	0.0740***	0.0725***	0.0760***	0.0742***
	(0.0025)	(0.0024)	(0.0028)	(0.0027)	(0.0027)	(0.0027)	(0.0024)	(0.0024)	(0.0027)	(0.0026)	(0.0025)	(0.0025)
Experience	-0.0117**	-0.0090**	-0.0109*	-0.0059	-0.0222***	-0.0218***	-0.0243***	-0.0246***	-0.0272***	-0.0271***	-0.0319***	-0.0321***
	(0.0051)	(0.0049)	(0.0065)	(0.0060)	(0.0061)	(0.0059)	(0.0059)	(0.0057)	(0.0064)	(0.0061)	(0.0059)	(0.0057)
Experience sq	0.0003***	0.0002***	0.0003***	0.0002*	0.0004***	0.0004***	0.0004***	0.0005***	0.0005***	0.0005***	0.0006***	0.0006***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Centre	0.0163	0.0159	0.0256**	0.0276***	0.0290***	0.0296***	0.0230**	0.0259***	0.0211*	0.0190*	0.0135	0.0125
	(0.0099)	(0.0098)	(0.0100)	(0.0099)	(0.0108)	(0.0105)	(0.0109)	(0.0106)	(0.0110)	(0.0107)	(0.0112)	(0.0109)
Lisbon and	0.1421***	0.1409***	0.1623***	0.1613***	0.1616***	0.1595***	0.1531***	0.1537***	0.1544***	0.1507***	0.1570***	0.1557***
Tagus Valley	(0.0104)	(0.0103)	(0.0119)	(0.0115)	(0.0115)	(0.0111)	(0.0117)	(0.0113)	(0.0114)	(0.0110)	(0.0121)	(0.0116)
Inland	-0.0832***	-0.0797***	-0.0768***	-0.0755***	-0.0756***	-0.0759***	-0.0811***	-0.0757***	-0.0844***	-0.0841***	-0.0799***	-0.0790***
	(0.0115)	(0.0113)	(0.0121)	(0.0119)	(0.0118)	(0.0114)	(0.0123)	(0.0118)	(0.0120)	(0.0117)	(0.0124)	(0.0121)
Algarve	0.0459***	0.0479***	0.0546***	0.0559***	0.0248*	0.0263*	0.0311**	0.0370**	0.0078	0.0069	-0.0086	-0.0054
	(0.0139)	(0.0139)	(0.0146)	(0.0144)	(0.0146)	(0.0141)	(0.0153)	(0.0149)	(0.0146)	(0.0144)	(0.0145)	(0.0141)
Observations	395,205	395,205	391,504	391,504	377,627	377,627	375,745	375,745	385,692	385,692	394,777	394,777
R-squared	0.583	0.583	0.599	0.599	0.587	0.587	0.569	0.569	0.562	0.563	0.549	0.550
RMSE	0.377	0.377	0.374	0.375	0.388	0.388	0.401	0.401	0.417	0.416	0.427	0.427
Hansen J stat.		24.17		27.23		18.49		29.59		20.29		22.04
p–value of HJ		0.086		0.039		0.296		0.020		0.208		0.142

Continued on next page

Table A.13 – continued from previous page

	2006		2007		2008		2009	
VARIABLES	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$
Education	0.0732***	0.0719***	0.0734***	0.0713***	0.0719***	0.0698***	0.0708***	0.0691***
	(0.0026)	(0.0025)	(0.0027)	(0.0026)	(0.0028)	(0.0027)	(0.0028)	(0.0028)
Experience	-0.0355***	-0.0352***	-0.0477***	-0.0474***	-0.0474***	-0.0457***	-0.0476***	-0.0464***
	(0.0062)	(0.0060)	(0.0064)	(0.0062)	(0.0065)	(0.0063)	(0.0066)	(0.0064)
Experience sq	0.0006***	0.0006***	0.0008***	0.0008***	0.0007***	0.0007***	0.0007***	0.0007***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Centre	0.0175*	0.0173*	0.0138	0.0146	0.0224**	0.0225**	0.0129	0.0106
	(0.0104)	(0.0102)	(0.0107)	(0.0105)	(0.0096)	(0.0095)	(0.0102)	(0.0101)
Lisbon and	0.1666***	0.1671***	0.1728***	0.1725***	0.1680***	0.1666***	0.1652***	0.1639***
Tagus Valley	(0.0109)	(0.0106)	(0.0119)	(0.0116)	(0.0105)	(0.0102)	(0.0115)	(0.0112)
Inland	-0.0789***	-0.0786***	-0.0814***	-0.0796***	-0.0730***	-0.0725***	-0.0739***	-0.0743***
	(0.0118)	(0.0114)	(0.0125)	(0.0122)	(0.0115)	(0.0113)	(0.0118)	(0.0116)
Algarve	0.0069	0.0081	0.0100	0.0137	0.0148	0.0171	0.0077	0.0079
	(0.0135)	(0.0133)	(0.0137)	(0.0135)	(0.0126)	(0.0125)	(0.0133)	(0.0132)
Observations	386,044	386,044	383,973	383,973	375,137	375,137	351,151	351,151
R-squared	0.546	0.547	0.534	0.536	0.530	0.532	0.540	0.541
RMSE	0.430	0.430	0.435	0.435	0.434	0.434	0.435	0.435
Hansen J statistic		18.11		25.22		26.36		32.17
p–value of HJ statistic		0.318		0.066		0.049		0.010

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include a set of dummies to control for occupation and sector of activity. North is the omitted category in region.

Table A.14: IV estimates using q_1 and $q_1 \times year - of - birth$ as instruments, Men born 1967–1980

	1992		**1993**		**1994**		**1995**		**1996**		**1997**	
VARIABLES	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$
Education	0.0496***	0.0490***	0.0506***	0.0505***	0.0492***	0.0488***	0.0546***	0.0544***	0.0530***	0.0533***	0.0574***	0.0574***
	(0.0025)	(0.0025)	(0.0027)	(0.0027)	(0.0025)	(0.0024)	(0.0024)	(0.0024)	(0.0025)	(0.0025)	(0.0023)	(0.0022)
Experience	0.0926***	0.0921***	0.0937***	0.0916***	0.0944***	0.0934***	0.0822***	0.0816***	0.0784***	0.0780***	0.0825***	0.0821***
	(0.0082)	(0.0081)	(0.0082)	(0.0089)	(0.0062)	(0.0061)	(0.0054)	(0.0053)	(0.0052)	(0.0052)	(0.0041)	(0.0041)
Experience sq	-0.0050***	-0.0051***	-0.0052***	-0.0050***	-0.0048***	-0.0048***	-0.0037***	-0.0037***	-0.0035***	-0.0035***	-0.0032***	-0.0032***
	(0.0006)	(0.0006)	(0.0005)	(0.0006)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.0002)
Centre	0.0373***	0.0380***	0.0320***	0.0326***	0.0175**	0.0168**	0.0267***	0.0265***	0.0198**	0.0197**	0.0322***	0.0314***
	(0.0070)	(0.0069)	(0.0078)	(0.0077)	(0.0082)	(0.0082)	(0.0087)	(0.0086)	(0.0094)	(0.0094)	(0.0076)	(0.0076)
Lisbon and	0.1080***	0.1104***	0.1203***	0.1217***	0.1156***	0.1144***	0.1149***	0.1164***	0.1196***	0.1204***	0.1161***	0.1154***
Tagus Valley	(0.0098)	(0.0097)	(0.0115)	(0.0113)	(0.0122)	(0.0119)	(0.0117)	(0.0115)	(0.0126)	(0.0124)	(0.0101)	(0.0101)
Inland	-0.0432***	-0.0414***	-0.0484***	-0.0477***	-0.0586***	-0.0587***	-0.0610***	-0.0605***	-0.0599***	-0.0594***	-0.0625***	-0.0630***
	(0.0076)	(0.0076)	(0.0099)	(0.0098)	(0.0092)	(0.0091)	(0.0091)	(0.0090)	(0.0105)	(0.0104)	(0.0085)	(0.0085)
Algarve	0.1088***	0.1110***	0.1151***	0.1165***	0.0670***	0.0660***	0.0555***	0.0571***	0.0670***	0.0674***	0.0644***	0.0630***
	(0.0149)	(0.0149)	(0.0150)	(0.0149)	(0.0143)	(0.0143)	(0.0147)	(0.0146)	(0.0146)	(0.0146)	(0.0131)	(0.0130)
Observations	102,047	102,047	120,693	120,693	150,832	150,832	194,573	194,573	213,515	213,515	263,581	263,581
R–squared	0.312	0.313	0.319	0.317	0.351	0.351	0.463	0.463	0.457	0.457	0.473	0.473
RMSE	0.312	0.312	0.342	0.342	0.348	0.348	0.323	0.323	0.335	0.335	0.327	0.327
Hansen J stat.		15.03		9.000		9.960		12.43		9.982		7.228
p–value of HJ		0.090		0.253		0.619		0.493		0.764		0.926

Continued on next page

Table A.14 – continued from previous page

	1998		1999		2000		2002		2003		2004	
VARIABLES	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$
Education	0.0597***	0.0596***	0.0625***	0.0623***	0.0667***	0.0661***	0.0693***	0.0687***	0.0689***	0.0684***	0.0716***	0.0712***
	(0.0022)	(0.0022)	(0.0022)	(0.0021)	(0.0021)	(0.0021)	(0.0022)	(0.0022)	(0.0022)	(0.0021)	(0.0022)	(0.0022)
Experience	0.0756***	0.0737***	0.0706***	0.0700***	0.0711***	0.0682***	0.0753***	0.0712***	0.0731***	0.0698***	0.0748***	0.0717***
	(0.0039)	(0.0039)	(0.0036)	(0.0036)	(0.0034)	(0.0032)	(0.0032)	(0.0031)	(0.0031)	(0.0030)	(0.0030)	(0.0030)
Experience sq	-0.0028***	-0.0027***	-0.0026***	-0.0026***	-0.0023***	-0.0022***	-0.0022***	-0.0021***	-0.0021***	-0.0020***	-0.0021***	-0.0020***
	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Centre	0.0182***	0.0194***	0.0243***	0.0255***	0.0353***	0.0360***	0.0356***	0.0365***	0.0292***	0.0305***	0.0268***	0.0274***
	(0.0066)	(0.0066)	(0.0063)	(0.0063)	(0.0065)	(0.0065)	(0.0064)	(0.0063)	(0.0070)	(0.0069)	(0.0073)	(0.0072)
Lisbon and	0.1096***	0.1058***	0.1078***	0.1093***	0.1244***	0.1251***	0.1271***	0.1263***	0.1276***	0.1292***	0.1261***	0.1267***
Tagus Valley	(0.0098)	(0.0096)	(0.0082)	(0.0082)	(0.0084)	(0.0082)	(0.0079)	(0.0078)	(0.0082)	(0.0081)	(0.0083)	(0.0082)
Inland	-0.0636***	-0.0623***	-0.0717***	-0.0697***	-0.0561***	-0.0555***	-0.0575***	-0.0573***	-0.0584***	-0.0582***	-0.0579***	-0.0562***
	(0.0089)	(0.0088)	(0.0085)	(0.0083)	(0.0086)	(0.0085)	(0.0075)	(0.0074)	(0.0086)	(0.0086)	(0.0090)	(0.0090)
Algarve	0.0634***	0.0602***	0.0510***	0.0517***	0.0596***	0.0603***	0.0518***	0.0522***	0.0549***	0.0553***	0.0443***	0.0437***
	(0.0125)	(0.0124)	(0.0118)	(0.0117)	(0.0110)	(0.0109)	(0.0103)	(0.0102)	(0.0108)	(0.0107)	(0.0123)	(0.0122)
Observations	283,475	283,475	322,209	322,209	347,462	347,462	385,309	385,309	399,331	399,331	419,834	419,834
R-squared	0.499	0.499	0.512	0.512	0.521	0.522	0.523	0.524	0.515	0.516	0.517	0.518
RMSE	0.323	0.323	0.329	0.329	0.332	0.332	0.348	0.347	0.357	0.357	0.368	0.368
Hansen J stat.		23.86		21.88		34.04		48.66		49.93		44.72
p–value of HJ		0.048		0.081		0.001		5.05e-06		3.06e-06		2.33e-05

Continued on next page

Table A.14 – continued from previous page

	2005		2006		2007		2008		2009	
VARIABLES	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$	q_1	q_1, $q_1 \times yb$
Education	0.0722***	0.0720***	0.0723***	0.0723***	0.0706***	0.0702***	0.0697***	0.0694***	0.0719***	0.0717***
	(0.0023)	(0.0023)	(0.0023)	(0.0022)	(0.0023)	(0.0023)	(0.0022)	(0.0022)	(0.0024)	(0.0024)
Experience	0.0764***	0.0734***	0.0712***	0.0712***	0.0686***	0.0678***	0.0641***	0.0636***	0.0612***	0.0608***
	(0.0033)	(0.0032)	(0.0031)	(0.0030)	(0.0031)	(0.0031)	(0.0030)	(0.0030)	(0.0032)	(0.0032)
Experience sq	-0.0022***	-0.0021***	-0.0018***	-0.0018***	-0.0018***	-0.0018***	-0.0016***	-0.0016***	-0.0014***	-0.0014***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Centre	0.0205***	0.0234***	0.0279***	0.0290***	0.0219***	0.0215***	0.0220***	0.0215***	0.0156*	0.0159**
	(0.0074)	(0.0073)	(0.0075)	(0.0075)	(0.0080)	(0.0079)	(0.0079)	(0.0079)	(0.0081)	(0.0080)
Lisbon and	0.1303***	0.1308***	0.1426***	0.1420***	0.1425***	0.1415***	0.1360***	0.1345***	0.1285***	0.1273***
Tagus Valley	(0.0084)	(0.0083)	(0.0082)	(0.0081)	(0.0092)	(0.0091)	(0.0095)	(0.0094)	(0.0100)	(0.0099)
Inland	-0.0598***	-0.0564***	-0.0612***	-0.0616***	-0.0696***	-0.0701***	-0.0688***	-0.0703***	-0.0686***	-0.0685***
	(0.0099)	(0.0098)	(0.0096)	(0.0096)	(0.0103)	(0.0102)	(0.0107)	(0.0106)	(0.0102)	(0.0102)
Algarve	0.0384***	0.0380***	0.0447***	0.0445***	0.0501***	0.0484***	0.0503***	0.0487***	0.0416***	0.0406***
	(0.0121)	(0.0120)	(0.0103)	(0.0102)	(0.0124)	(0.0123)	(0.0122)	(0.0122)	(0.0129)	(0.0128)
Observations	449,972	449,972	453,254	453,254	461,434	461,434	465,057	465,057	448,227	448,227
R–squared	0.509	0.509	0.506	0.506	0.500	0.500	0.499	0.499	0.504	0.504
RMSE	0.379	0.379	0.386	0.386	0.394	0.394	0.396	0.396	0.398	0.398
Hansen J statistic		35.93		15.63		22.40		13.11		7.132
p–value of HJ statistic		0.001		0.336		0.071		0.518		0.930

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include a set of dummies to control for occupation and sector of activity. North is the omitted category in region.

Table A.15: OLS and IV estimates using CSL6 as instrument

	1991		1992		1993		1994		1995		1996	
VARIABLES	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**
Education	0.0529***	0.0584***	0.0539***	0.0650***	0.0540***	0.0683***	0.0520***	0.0717***	0.0496***	0.0718***	0.0480***	0.0720***
	(0.0003)	(0.0028)	(0.0003)	(0.0026)	(0.0003)	(0.0028)	(0.0003)	(0.0023)	(0.0002)	(0.0025)	(0.0002)	(0.0024)
Experience	0.0462***	0.0455***	0.0460***	0.0450***	0.0454***	0.0445***	0.0439***	0.0437***	0.0446***	0.0462***	0.0435***	0.0470***
	(0.0006)	(0.0011)	(0.0006)	(0.0011)	(0.0006)	(0.0011)	(0.0006)	(0.0011)	(0.0005)	(0.0012)	(0.0005)	(0.0014)
Experience sq	-0.0012***	-0.0011***	-0.0012***	-0.0011***	-0.0011***	-0.0010***	-0.0010***	-0.0009***	-0.0010***	-0.0009***	-0.0009***	-0.0009***
	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Centre	0.0078***	0.0074	0.0210***	0.0202**	0.0176***	0.0163*	0.0094***	0.0077	0.0134***	0.0114	0.0124***	0.0112
	(0.0019)	(0.0078)	(0.0019)	(0.0083)	(0.0020)	(0.0091)	(0.0020)	(0.0085)	(0.0017)	(0.0088)	(0.0017)	(0.0092)
Lisbon and	0.1150***	0.1109***	0.1275***	0.1192***	0.1382***	0.1269***	0.1340***	0.1181***	0.1339***	0.1162***	0.1372***	0.1179***
Tagus Valley	(0.0016)	(0.0119)	(0.0016)	(0.0125)	(0.0016)	(0.0143)	(0.0016)	(0.0129)	(0.0014)	(0.0126)	(0.0014)	(0.0136)
Inland	-0.0675***	-0.0678***	-0.0677***	-0.0687***	-0.0711***	-0.0727***	-0.0726***	-0.0752***	-0.0785***	-0.0805***	-0.0709***	-0.0736***
	(0.0024)	(0.0123)	(0.0025)	(0.0102)	(0.0025)	(0.0121)	(0.0025)	(0.0106)	(0.0022)	(0.0104)	(0.0022)	(0.0128)
Algarve	0.0776***	0.0751***	0.0988***	0.0949***	0.1029***	0.0963***	0.0596***	0.0501***	0.0583***	0.0495***	0.0640***	0.0547***
	(0.0044)	(0.0154)	(0.0044)	(0.0155)	(0.0045)	(0.0164)	(0.0043)	(0.0147)	(0.0038)	(0.0156)	(0.0037)	(0.0149)
Observations	289,060	289,060	314,061	314,061	330,103	330,103	362,743	362,743	426,915	426,915	439,679	439,679
R–squared	0.470	0.469	0.466	0.463	0.452	0.447	0.468	0.459	0.553	0.544	0.537	0.526
RMSE	0.332	0.332	0.346	0.347	0.366	0.367	0.372	0.375	0.353	0.357	0.361	0.365

Continued on next page

Table A.15 – continued from previous page

	1997		1998		1999		2000		2002		2003	
VARIABLES	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**
Education	0.0469***	0.0675***	0.0488***	0.0727***	0.0505***	0.0753***	0.0400***	0.0730***	0.0495***	0.0716***	0.0510***	0.0734***
	(0.0002)	(0.0022)	(0.0002)	(0.0021)	(0.0002)	(0.0020)	(0.0002)	(0.0021)	(0.0002)	(0.0021)	(0.0002)	(0.0019)
Experience	0.0443***	0.0486***	0.0425***	0.0484***	0.0414***	0.0483***	0.0582***	0.0456***	0.0367***	0.0448***	0.0364***	0.0453***
	(0.0004)	(0.0013)	(0.0004)	(0.0014)	(0.0004)	(0.0014)	(0.0006)	(0.0017)	(0.0004)	(0.0016)	(0.0005)	(0.0017)
Experience sq	-0.0009***	-0.0010***	-0.0008***	-0.0009***	-0.0008***	-0.0009***	-0.0023***	-0.0008***	-0.0006***	-0.0007***	-0.0006***	-0.0007***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Centre	0.0279***	0.0256***	0.0146***	0.0135*	0.0240***	0.0220***	0.0391***	0.0346***	0.0392***	0.0374***	0.0348***	0.0318***
	(0.0016)	(0.0079)	(0.0016)	(0.0072)	(0.0015)	(0.0071)	(0.0015)	(0.0074)	(0.0016)	(0.0075)	(0.0016)	(0.0076)
Lisbon and	0.1426***	0.1241***	0.1355***	0.1154***	0.1382***	0.1182***	0.1605***	0.1391***	0.1582***	0.1417***	0.1561***	0.1393***
Tagus Valley	(0.0013)	(0.0101)	(0.0013)	(0.0094)	(0.0012)	(0.0087)	(0.0012)	(0.0096)	(0.0013)	(0.0085)	(0.0013)	(0.0088)
Inland	-0.0716***	-0.0745***	-0.0702***	-0.0721***	-0.0748***	-0.0770***	-0.0626***	-0.0664***	-0.0649***	-0.0661***	-0.0652***	-0.0677***
	(0.0020)	(0.0103)	(0.0020)	(0.0099)	(0.0019)	(0.0098)	(0.0019)	(0.0097)	(0.0019)	(0.0085)	(0.0020)	(0.0094)
Algarve	0.0652***	0.0568***	0.0584***	0.0501***	0.0557***	0.0460***	0.0695***	0.0558***	0.0479***	0.0392***	0.0551***	0.0436***
	(0.0034)	(0.0134)	(0.0033)	(0.0127)	(0.0032)	(0.0120)	(0.0031)	(0.0120)	(0.0030)	(0.0113)	(0.0030)	(0.0116)
Observations	492,174	492,174	504,141	504,141	540,648	540,648	554,018	554,018	566,629	566,629	574,309	574,309
R–squared	0.539	0.531	0.556	0.544	0.560	0.548	0.576	0.559	0.563	0.554	0.551	0.542
RMSE	0.353	0.357	0.348	0.353	0.354	0.359	0.351	0.358	0.369	0.372	0.379	0.383

Continued on next page

Table A.15 – continued from previous page

	2004		2005		2006		2007		2008		2009	
VARIABLES	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**	**OLS**	**IV**		
Education	0.0547***	0.0790***	0.0559***	0.0808***	0.0544***	0.0794***	0.0542***	0.0792***	0.0523***	0.0787***	0.0522***	0.0769***
	(0.0002)	(0.0019)	(0.0002)	(0.0021)	(0.0002)	(0.0019)	(0.0002)	(0.0020)	(0.0002)	(0.0019)	(0.0002)	(0.0020)
Experience	0.0372***	0.0474***	0.0362***	0.0470***	0.0327***	0.0439***	0.0304***	0.0421***	0.0261***	0.0386***	0.0241***	0.0360***
	(0.0005)	(0.0017)	(0.0005)	(0.0022)	(0.0005)	(0.0018)	(0.0006)	(0.0018)	(0.0006)	(0.0017)	(0.0006)	(0.0019)
Experience sq	-0.0006***	-0.0007***	-0.0005***	-0.0007***	-0.0004***	-0.0006***	-0.0004***	-0.0006***	-0.0003***	-0.0005***	-0.0003***	-0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Centre	0.0326***	0.0296***	0.0247***	0.0219***	0.0306***	0.0275***	0.0242***	0.0213**	0.0290***	0.0261***	0.0214***	0.0185**
	(0.0017)	(0.0079)	(0.0017)	(0.0082)	(0.0017)	(0.0080)	(0.0017)	(0.0085)	(0.0017)	(0.0080)	(0.0018)	(0.0083)
Lisbon and	0.1571***	0.1390***	0.1601***	0.1421***	0.1722***	0.1544***	0.1730***	0.1561***	0.1675***	0.1497***	0.1632***	0.1463***
Tagus Valley	(0.0013)	(0.0090)	(0.0013)	(0.0096)	(0.0013)	(0.0089)	(0.0013)	(0.0099)	(0.0013)	(0.0095)	(0.0014)	(0.0103)
Inland	-0.0647***	-0.0677***	-0.0643***	-0.0676***	-0.0642***	-0.0678***	-0.0719***	-0.0749***	-0.0682***	-0.0713***	-0.0691***	-0.0724***
	(0.0020)	(0.0096)	(0.0020)	(0.0105)	(0.0020)	(0.0102)	(0.0021)	(0.0110)	(0.0021)	(0.0109)	(0.0021)	(0.0108)
Algarve	0.0418***	0.0292**	0.0313***	0.0183	0.0429***	0.0298***	0.0463***	0.0335***	0.0489***	0.0352***	0.0410***	0.0287**
	(0.0030)	(0.0126)	(0.0030)	(0.0126)	(0.0030)	(0.0113)	(0.0030)	(0.0126)	(0.0030)	(0.0122)	(0.0032)	(0.0128)
Observations	594,757	594,757	622,017	622,017	618,488	618,488	622,599	622,599	619,828	619,828	593,114	593,114
R–squared	0.549	0.539	0.538	0.528	0.537	0.526	0.527	0.517	0.524	0.513	0.532	0.522
RMSE	0.392	0.396	0.403	0.408	0.407	0.412	0.415	0.420	0.415	0.420	0.416	0.421

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include a set of dummies to control for occupation and sector of activity. North is the omitted category in region.

Table A.16: First Stage of IV2SLS – Estimation using Average Education in Portugal and Average Education by Region

Year	**Av. Schooling** (country)	R^2	F–**test on excluded var.**	**Av. Schooling by region**	R^2	F–**test on excluded var.**
1986	-4.570***	0.4028	1253***	-1.071***	0.3589	227***
	(.1291)			(.0706)		
1987	-4.734***	0.4154	1405***	-1.142***	0.3696	296***
	(.1263)			(.0664)		
1988	-4.879***	0.4188	1509***	-1.139***	0.3691	254***
	(.1256)			(.0714)		
1989	-5.302***	0.4427	1873***	-1.273***	0.3861	366***
	(.1225)			(.0666)		
1991	-6.044***	0.5047	3307***	-1.490***	0.4349	422***
	(.1051)			(.0726)		
1992	-6.253***	0.5179	4458***	-1.612***	0.4435	488***
	(.0936)			(.0730)		
1993	-6.416***	0.5370	5143***	-1.732***	0.4597	666***
	(.0895)			(.0671)		
1994	-6.777***	0.5748	9616***	-2.136***	0.4914	800***
	(.0691)			(.0755)		
1995	-5.654***	0.6726	4923***	-2.023***	0.6192	712***
	(.0806)			(.0758)		
1996	-5.390***	0.6789	13935***	-2.163***	0.6279	1026***
	(.0457)			(.0675)		
1997	-5.159***	0.6849	21227***	-2.268***	0.6355	1387***
	(.0354)			(.0609)		
1998	-5.232***	0.6927	23223***	-2.589***	0.6439	2027***
	(.0343)			(.0575)		
1999	-5.217***	0.6997	24061***	-2.810***	0.6528	3198***
	(.0336)			(.0497)		
2000	-5.089***	0.7114	23867***	-2.934***	0.6688	3597***
	(.0329)			(.0489)		
2002	-5.226***	0.7130	25050***	-3.460***	0.6744	7132***
	(.0330)			(.0410)		
2003	-5.336***	0.7089	28167***	-3.675***	0.6699	8973***
	(.0318)			(.0387)		
2004	-5.401***	0.7072	31028***	-3.804***	0.6674	12799***
	(.0307)			(.0336)		
2005	-5.314***	0.7097	39275***	-3.789***	0.6709	18739***
	(.0268)			(.0277)		
2006	-5.197***	0.7089	44979***	-3.718***	0.6708	22034***
	(.0245)			(.0251)		
2007	-5.049***	0.7049	43264***	-3.579***	0.6666	17764***
	(.0243)			(.0269)		
2008	-4.894***	0.7035	49770***	-3.449***	0.6653	18252***
	(.0219)			(.0255)		
2009	-4.693***	0.6988	45383***	-3.294***	0.6608	15112***
	(.0220)			(.0268)		

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A.17: OLS and IV estimates using Average Education Country and Average Education by Region as instruments

VARIABLES	1986			1987			1988			1989		
	OLS	IV		OLS	IV		OLS	IV		OLS	IV	
		Country	Region		Country	Region		Country	Region		Country	Region
Education	0.0551***	0.0433***	0.1459***	0.0558***	0.0497***	0.1488***	0.0560***	0.0553***	0.1445***	0.0586***	0.0618***	0.1425***
	(0.0002)	(0.0020)	(0.0180)	(0.0002)	(0.0021)	(0.0140)	(0.0002)	(0.0018)	(0.0158)	(0.0002)	(0.0018)	(0.0133)
Experience	0.0328***	0.0316***	0.0420***	0.0330***	0.0324***	0.0418***	0.0324***	0.0323***	0.0412***	0.0331***	0.0334***	0.0415***
	(0.0002)	(0.0009)	(0.0023)	(0.0002)	(0.0008)	(0.0015)	(0.0002)	(0.0007)	(0.0019)	(0.0002)	(0.0006)	(0.0017)
Experience sq	-0.0005***	-0.0005***	-0.0006***	-0.0005***	-0.0005***	-0.0006***	-0.0005***	-0.0005***	-0.0006***	-0.0005***	-0.0005***	-0.0006***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Centre	-0.0367***	-0.0360***	-0.0418***	-0.0422***	-0.0418***	-0.0474***	-0.0308***	-0.0308***	-0.0358***	-0.0138***	-0.0139	-0.0164*
	(0.0014)	(0.0113)	(0.0099)	(0.0014)	(0.0109)	(0.0097)	(0.0014)	(0.0103)	(0.0093)	(0.0014)	(0.0098)	(0.0087)
Lisbon and	0.1357***	0.1416***	0.0904***	0.1318***	0.1348***	0.0862***	0.1271***	0.1275***	0.0815***	0.1364***	0.1346***	0.0911***
Tagus Valley	(0.0010)	(0.0155)	(0.0120)	(0.0010)	(0.0149)	(0.0106)	(0.0010)	(0.0133)	(0.0105)	(0.0011)	(0.0127)	(0.0102)
Inland	-0.0637***	-0.0633***	-0.0670***	-0.0815***	-0.0813***	-0.0851***	-0.0938***	-0.0937***	-0.0975***	-0.0816***	-0.0816***	-0.0829***
	(0.0019)	(0.0133)	(0.0113)	(0.0018)	(0.0117)	(0.0100)	(0.0018)	(0.0101)	(0.0086)	(0.0017)	(0.0101)	(0.0088)
Algarve	0.0546***	0.0589***	0.0214	0.0594***	0.0618***	0.0239	0.1079***	0.1082***	0.0705***	0.1021***	0.1009***	0.0714***
	(0.0035)	(0.0202)	(0.0161)	(0.0034)	(0.0196)	(0.0156)	(0.0033)	(0.0208)	(0.0185)	(0.0033)	(0.0170)	(0.0153)
Observations	561,384	561,384	561,384	580,967	580,967	580,967	588,812	588,812	588,812	626,711	626,711	626,711
R–squared	0.559	0.556	0.389	0.562	0.561	0.384	0.548	0.548	0.380	0.529	0.528	0.380
RMSE	0.304	0.305	0.358	0.304	0.304	0.361	0.305	0.305	0.358	0.323	0.323	0.370

Continued on next page

Table A.17 – continued from previous page

VARIABLES	1991			1992			1993			1994		
	OLS	IV		OLS	IV		OLS	IV		OLS	IV	
		Country	Region		Country	Region		Country	Region		Country	Region
Education	0.0602***	0.0658***	0.1147***	0.0600***	0.0673***	0.1137***	0.0597***	0.0681***	0.1058***	0.0577***	0.0664***	0.0918***
	(0.0002)	(0.0016)	(0.0110)	(0.0002)	(0.0015)	(0.0093)	(0.0002)	(0.0019)	(0.0084)	(0.0002)	(0.0015)	(0.0054)
Experience	0.0310***	0.0316***	0.0369***	0.0310***	0.0318***	0.0371***	0.0310***	0.0320***	0.0364***	0.0313***	0.0324***	0.0356***
	(0.0002)	(0.0006)	(0.0015)	(0.0002)	(0.0006)	(0.0013)	(0.0002)	(0.0006)	(0.0013)	(0.0002)	(0.0006)	(0.0010)
Experience sq	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Centre	-0.0140***	-0.0140	-0.0144	0.0018	0.0016	0.0005	-0.0025	-0.0029	-0.0046	-0.0075***	-0.0079	-0.0090
	(0.0014)	(0.0103)	(0.0095)	(0.0015)	(0.0093)	(0.0085)	(0.0015)	(0.0094)	(0.0086)	(0.0015)	(0.0091)	(0.0087)
Lisbon and	0.1341***	0.1310***	0.1042***	0.1458***	0.1416***	0.1145***	0.1501***	0.1450***	0.1218***	0.1405***	0.1349***	0.1185***
Tagus Valley	(0.0011)	(0.0140)	(0.0111)	(0.0011)	(0.0127)	(0.0106)	(0.0012)	(0.0133)	(0.0113)	(0.0011)	(0.0117)	(0.0109)
Inland	-0.0840***	-0.0838***	-0.0820***	-0.0784***	-0.0784***	-0.0783***	-0.0890***	-0.0892***	-0.0898***	-0.0795***	-0.0798***	-0.0807***
	(0.0018)	(0.0119)	(0.0109)	(0.0018)	(0.0106)	(0.0095)	(0.0019)	(0.0113)	(0.0104)	(0.0019)	(0.0104)	(0.0098)
Algarve	0.0634***	0.0614***	0.0441***	0.0789***	0.0766***	0.0613***	0.0814***	0.0782***	0.0635***	0.0402***	0.0367***	0.0266*
	(0.0032)	(0.0176)	(0.0158)	(0.0032)	(0.0172)	(0.0157)	(0.0034)	(0.0159)	(0.0151)	(0.0032)	(0.0142)	(0.0136)
Observations	683,741	683,741	683,741	694,774	694,774	694,774	690,591	690,591	690,591	699,919	699,919	699,919
R–squared	0.543	0.543	0.488	0.535	0.534	0.482	0.512	0.511	0.474	0.527	0.526	0.506
RMSE	0.352	0.352	0.372	0.361	0.361	0.381	0.375	0.376	0.390	0.379	0.379	0.387

Continued on next page

Table A.17 – continued from previous page

VARIABLES	1995			1996			1997			1998		
	OLS	IV		OLS	IV		OLS	IV		OLS	IV	
		Country	Region		Country	Region		Country	Region		Country	Region
Education	0.0522***	0.0613***	0.0860***	0.0500***	0.0598***	0.0772***	0.0477***	0.0570***	0.0745***	0.0487***	0.0591***	0.0767***
	(0.0002)	(0.0018)	(0.0057)	(0.0002)	(0.0017)	(0.0053)	(0.0002)	(0.0014)	(0.0043)	(0.0002)	(0.0014)	(0.0036)
Experince	0.0324***	0.0336***	0.0369***	0.0314***	0.0328***	0.0353***	0.0320***	0.0334***	0.0360***	0.0314***	0.0330***	0.0359***
	(0.0002)	(0.0005)	(0.0009)	(0.0002)	(0.0005)	(0.0010)	(0.0002)	(0.0006)	(0.0008)	(0.0001)	(0.0006)	(0.0008)
Experience sq	-0.0005***	-0.0005***	-0.0005***	-0.0004***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0005***	-0.0004***	-0.0004***	-0.0005***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Centre	-0.0059***	-0.0062	-0.0073	-0.0048***	-0.0050	-0.0053	0.0115***	0.0107	0.0091	0.0025**	0.0023	0.0019
	(0.0014)	(0.0094)	(0.0090)	(0.0014)	(0.0095)	(0.0092)	(0.0013)	(0.0088)	(0.0084)	(0.0013)	(0.0084)	(0.0081)
Lisbon and	0.1439***	0.1380***	0.1219***	0.1411***	0.1346***	0.1231***	0.1459***	0.1391***	0.1261***	0.1426***	0.1355***	0.1233***
Tagus Valley	(0.0011)	(0.0118)	(0.0107)	(0.0011)	(0.0124)	(0.0117)	(0.0010)	(0.0102)	(0.0091)	(0.0010)	(0.0098)	(0.0085)
Inland	-0.0878***	-0.0880***	-0.0885***	-0.0824***	-0.0828***	-0.0834***	-0.0810***	-0.0816***	-0.0828***	-0.0752***	-0.0753***	-0.0754***
	(0.0017)	(0.0104)	(0.0098)	(0.0018)	(0.0122)	(0.0118)	(0.0016)	(0.0104)	(0.0101)	(0.0016)	(0.0099)	(0.0096)
Algarve	0.0402***	0.0372**	0.0290*	0.0503***	0.0471***	0.0415***	0.0516***	0.0484***	0.0422***	0.0477***	0.0446***	0.0392***
	(0.0030)	(0.0156)	(0.0149)	(0.0030)	(0.0150)	(0.0146)	(0.0028)	(0.0140)	(0.0135)	(0.0027)	(0.0131)	(0.0127)
Observations	782,266	782,266	782,266	778,301	778,301	778,301	836,295	836,295	836,295	846,319	846,319	846,319
R–squared	0.588	0.587	0.570	0.572	0.570	0.560	0.573	0.572	0.561	0.588	0.586	0.574
RMSE	0.368	0.368	0.375	0.372	0.373	0.377	0.365	0.366	0.370	0.358	0.359	0.364

Continued on next page

Table A.17 – continued from previous page

VARIABLES	1999			2000			2002			2003		
	OLS	IV		OLS	IV		OLS	IV		OLS	IV	
		Country	Region		Country	Region		Country	Region		Country	Region
Education	0.0504***	0.0632***	0.0780***	0.0474***	0.0614***	0.0754***	0.0471***	0.0640***	0.0707***	0.0492***	0.0671***	0.0701***
	(0.0002)	(0.0013)	(0.0032)	(0.0002)	(0.0013)	(0.0036)	(0.0002)	(0.0014)	(0.0027)	(0.0002)	(0.0013)	(0.0023)
Experience	0.0309***	0.0329***	0.0353***	0.0304***	0.0327***	0.0349***	0.0301***	0.0327***	0.0337***	0.0309***	0.0335***	0.0340***
	(0.0001)	(0.0005)	(0.0007)	(0.0001)	(0.0005)	(0.0007)	(0.0001)	(0.0005)	(0.0006)	(0.0002)	(0.0005)	(0.0005)
Experience	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Centre	0.0108***	0.0102	0.0094	0.0229***	0.0217***	0.0204***	0.0280***	0.0267***	0.0261***	0.0233***	0.0211***	0.0208***
	(0.0013)	(0.0082)	(0.0079)	(0.0012)	(0.0080)	(0.0076)	(0.0012)	(0.0080)	(0.0078)	(0.0013)	(0.0081)	(0.0080)
Lisbon and	0.1427***	0.1338***	0.1236***	0.1613***	0.1517***	0.1420***	0.1601***	0.1484***	0.1437***	0.1546***	0.1419***	0.1398***
Tagus Valley	(0.0010)	(0.0095)	(0.0083)	(0.0010)	(0.0101)	(0.0088)	(0.0010)	(0.0089)	(0.0084)	(0.0010)	(0.0091)	(0.0086)
Inland	-0.0807***	-0.0810***	-0.0813***	-0.0682***	-0.0687***	-0.0693***	-0.0666***	-0.0671***	-0.0673***	-0.0675***	-0.0690***	-0.0693***
	(0.0016)	(0.0094)	(0.0092)	(0.0015)	(0.0098)	(0.0095)	(0.0015)	(0.0087)	(0.0085)	(0.0015)	(0.0093)	(0.0092)
Algarve	0.0478***	0.0431***	0.0377***	0.0592***	0.0533***	0.0473***	0.0444***	0.0377***	0.0350***	0.0523***	0.0434***	0.0420***
	(0.0026)	(0.0121)	(0.0117)	(0.0025)	(0.0121)	(0.0117)	(0.0023)	(0.0113)	(0.0110)	(0.0023)	(0.0117)	(0.0115)
Observations	891,526	891,526	891,526	903,222	903,222	903,222	913,885	913,885	913,885	927,908	927,908	927,908
R–squared	0.586	0.583	0.572	0.596	0.593	0.583	0.579	0.574	0.569	0.565	0.560	0.557
RMSE	0.363	0.364	0.369	0.360	0.361	0.366	0.369	0.372	0.374	0.378	0.381	0.382

Continued on next page

Table A.17 – continued from previous page

VARIABLES	2004			2005			2006			2007		
	OLS	IV		OLS	IV		OLS	IV		OLS	IV	
		Country	Region		Country	Region		Country	Region		Country	Region
Education	0.0529***	0.0748***	0.0761***	0.0539***	0.0797***	0.0776***	0.0526***	0.0808***	0.0756***	0.0521***	0.0859***	0.0780***
	(0.0002)	(0.0015)	(0.0019)	(0.0002)	(0.0018)	(0.0018)	(0.0002)	(0.0017)	(0.0018)	(0.0002)	(0.0019)	(0.0019)
Experience	0.0330***	0.0361***	0.0363***	0.0340***	0.0375***	0.0372***	0.0337***	0.0375***	0.0368***	0.0331***	0.0374***	0.0364***
	(0.0002)	(0.0005)	(0.0005)	(0.0001)	(0.0007)	(0.0006)	(0.0001)	(0.0005)	(0.0005)	(0.0001)	(0.0006)	(0.0005)
Experience sq	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Centre	0.0207***	0.0179**	0.0178**	0.0145***	0.0115	0.0117	0.0209***	0.0172**	0.0179**	0.0162***	0.0120	0.0129
	(0.0013)	(0.0082)	(0.0081)	(0.0013)	(0.0080)	(0.0080)	(0.0013)	(0.0077)	(0.0077)	(0.0013)	(0.0079)	(0.0080)
Lisbon and	0.1534***	0.1377***	0.1368***	0.1549***	0.1368***	0.1382***	0.1641***	0.1446***	0.1482***	0.1646***	0.1423***	0.1475***
Tagus Valley	(0.0010)	(0.0089)	(0.0086)	(0.0010)	(0.0090)	(0.0089)	(0.0010)	(0.0083)	(0.0084)	(0.0010)	(0.0090)	(0.0092)
Inland	-0.0678***	-0.0699***	-0.0700***	-0.0657***	-0.0683***	-0.0681***	-0.0647***	-0.0683***	-0.0676***	-0.0702***	-0.0739***	-0.0731***
	(0.0016)	(0.0092)	(0.0091)	(0.0015)	(0.0094)	(0.0095)	(0.0015)	(0.0091)	(0.0092)	(0.0015)	(0.0096)	(0.0097)
Algarve	0.0375***	0.0267**	0.0261**	0.0276***	0.0147	0.0157	0.0391***	0.0255**	0.0280***	0.0451***	0.0291***	0.0328***
	(0.0023)	(0.0118)	(0.0117)	(0.0023)	(0.0114)	(0.0114)	(0.0023)	(0.0102)	(0.0102)	(0.0022)	(0.0109)	(0.0110)
Observations	967,227	967,227	967,227	1,027,128	1,027,128	1,027,128	1,030,386	1,030,386	1,030,386	1,053,358	1,053,358	1,053,358
R–squared	0.562	0.554	0.553	0.550	0.538	0.540	0.545	0.531	0.535	0.532	0.511	0.519
RMSE	0.389	0.393	0.393	0.396	0.401	0.401	0.398	0.404	0.402	0.402	0.411	0.407

Continued on next page

Table A.17 – continued from previous page

VARIABLES	2008			2009		
	OLS	IV		OLS	IV	
		Country	Region		Country	Region
Education	0.0507***	0.0871***	0.0785***	0.0510***	0.0911***	0.0783***
	(0.0002)	(0.0020)	(0.0021)	(0.0002)	(0.0023)	(0.0027)
Experience	0.0320***	0.0364***	0.0354***	0.0323***	0.0369***	0.0354***
	(0.0001)	(0.0006)	(0.0005)	(0.0002)	(0.0006)	(0.0005)
Experience sq	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Centre	0.0198***	0.0154**	0.0164**	0.0114***	0.0062	0.0078
	(0.0013)	(0.0073)	(0.0074)	(0.0013)	(0.0075)	(0.0077)
Lisbon and	0.1571***	0.1338***	0.1393***	0.1486***	0.1230***	0.1311***
Tagus Valley	(0.0010)	(0.0084)	(0.0088)	(0.0010)	(0.0090)	(0.0095)
Inland	-0.0670***	-0.0708***	-0.0699***	-0.0664***	-0.0717***	-0.0700***
	(0.0015)	(0.0094)	(0.0094)	(0.0016)	(0.0091)	(0.0092)
Algarve	0.0482***	0.0307***	0.0349***	0.0393***	0.0212**	0.0270**
	(0.0022)	(0.0104)	(0.0106)	(0.0023)	(0.0108)	(0.0110)
Observations	1,063,182	1,063,182	1,063,182	1,020,794	1,020,794	1,020,794
R–squared	0.527	0.502	0.512	0.534	0.504	0.520
RMSE	0.400	0.410	0.406	0.399	0.411	0.404

Source: Computations of the author based on *Quadros de Pessoal*, 1986–2009.
Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include a set of dummies to control for occupation and sector of activity. North is the omitted category in region.

Appendix B

Appendices to Chapter 4

B.1 Human Capital Index

B.1.1 Individual Skill Index

The skill index proposed by Portela (2001) takes into account different dimensions of the productivity of the worker, namely schooling, experience in the labour market and unobservable ability. This index for individual i is given by:

$$S_i = mschool \times \left(0.5 + \frac{e^{\frac{school_i - mschool}{sschool}}}{1 + e^{\frac{school_i - mschool}{sschool}}}\right) \times \left(0.5 + \frac{e^{\frac{exper_i - mexper_{|school_i}}{sexper_{|school_i}}}}{1 + e^{\frac{exper_i - mexper_{|school_i}}{sexper_{|school_i}}}}\right) \times$$

$$\times \left(0.5 + \frac{e^{\frac{effect_i - meffect_{|school_i, exper_i}}{seffect_{|school_i, exper_i}}}}{1 + e^{\frac{effect_i - meffect_{|school_i, exper_i}}{seffect_{|school_i, exper_i}}}}\right)$$

where $school_i$ is the schooling level of worker i, evaluated as years of schooling, $mschool$ represents the average of school level in the population and $sschool$ denotes the standard deviation of schooling; $exper_i$ is the experience of the worker i, $mexper_{|school_i}$ and $sexper_{|school_i}$ represent the average and the standard deviation of experience within school level $school_i$, respectively; $effect_i$ corresponds to the individual's specific effect, $meffect_{|school_i, exper_i}$ and $seffect_{|school_i, exper_i}$ are, respectively, the average and the standard deviation of the effects for the individuals with the same schooling and experience. These individual's specific effects are estimated as the worker's specific effect within a wage regression, where one uses the fixed effects estimation (Portela, 2001). As this approach allows to control for the effects of omitted variables, which, to some extent, caracterize individuals' heterogeneity, but are constant over time, the time-invariant variables are not considered as explanatory variables being captured by the unobserved

effects of the regression.

B.1.2 Average of the skill index of individuals within a county *vs.* the skill index of a county

Considering that n_r represents the number of workers of a given county and n the number of workers in the economy. The average of the skill index of the individuals of the county is given by:

$$\overline{SI} = \frac{\sum_{i=1}^{n_r} S_i}{n_r} = \frac{\sum_{i=1}^{n_r} \left[mschool \left(0.5 + \frac{e^{\frac{school_i - mschool}{sschool}}}{1+e^{\frac{school_i - mschool}{sschool}}} \right) \right]}{n_r}$$

$$\overline{SI} = \frac{mschool}{n_r} \sum_{i=1}^{n_r} \left(0.5 + \frac{e^{\frac{school_i - mschool}{sschool}}}{1 + e^{\frac{school_i - mschool}{sschool}}} \right)$$

$$\overline{SI} = \frac{mschool}{n_r} \left(0.5 n_r + \sum_{i=1}^{n_r} \frac{e^{\frac{school_i - mschool}{sschool}}}{1 + e^{\frac{school_i - mschool}{sschool}}} \right)$$

$$\overline{SI} = 0.5 \times mschool + \sum_{i=1}^{n_r} \left(\frac{e^{\frac{school_i - mschool}{sschool}}}{1 + e^{\frac{school_i - mschool}{sschool}}} \right) \tag{B.1}$$

The skill index of a given county with n_r workers is given by:

$$SI_r = mschool \left(0.5 + \frac{e^{\frac{school_r - mschool}{sschool}}}{1 + e^{\frac{school_r - mschool}{sschool}}} \right)$$

$$SI_r = mschool \left(0.5 + \frac{e^{\frac{\frac{\sum_{i=1}^{n_r} school_i}{n_r} - mschool}{sschool}}}{1 + e^{\frac{\frac{\sum_{i=1}^{n_r} school_i}{n_r} - mschool}{sschool}}} \right)$$

$$SI_r = 0.5 \times mschool + mschool \times \left(\frac{e^{\frac{\frac{\sum_{i=1}^{n_r} school_i}{n_r} - mschool}{sschool}}}{1 + e^{\frac{\frac{\sum_{i=1}^{n_r} school_i}{n_r} - mschool}{sschool}}} \right) \tag{B.2}$$

Comparing the two equations (B.1) and (B.2), one concludes that the skill index of a given county and the average of the individual skill index of the n workers of a county are different. In Chapter 4 the skill index of a county is used as a proxy for the county's aggregate human capital.

B.2 Fixed–effects estimation: collinearity between "experience" and time dummies within worker's spell

Consider the following wage equation:

$$y_{it} = x'_{it}\beta + \lambda_t + \varepsilon_{it}, \qquad i = 1, \ldots, N, \qquad and \qquad t = 1, \ldots, T$$

where y_{it} represents log real hourly wage of individual i in time t and x_{it} is individual i's experience at time t in the labour market; experience variable is an individual specific time trend.

In the current discussion one wants to show that, if a set of time variables is included, the fixed-effects estimation is not feasible. However, β can be estimated by OLS. First, one demonstrates the possibility of obtaining an estimate for the parameter β by OLS. Second, we show the impossibility of estimating β by fixed–effects. One assumes a simplified sample in order to facilitate the discussion and arithmetic operations.

Sample data

i	**t**	**y**	**Experience**
1	1	y_{11}	1
1	2	y_{12}	2
1	3	y_{13}	3
1	4	y_{14}	4
2	1	y_{21}	4
2	2	y_{22}	5
2	3	y_{23}	6
2	4	y_{24}	7

Considering only one individual ($i = 1$), the model can be rewriten as follows (the intercept is not included):

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \beta + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \end{bmatrix}$$

Assuming that the basic hypotheses of OLS are valid, the OLS estimator is given by: $\widehat{\beta} = (X'X)^{-1}X'y$, it follows that:

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

where

$$\mathbf{X'X} = \begin{bmatrix} 2 & 2 & 3 & 4 \\ 2 & 8 & 6 & 8 \\ 3 & 6 & 18 & 12 \\ 4 & 8 & 12 & 32 \end{bmatrix}$$

and

$$(\mathbf{X'X})^{-\mathbf{1}} = \begin{bmatrix} \frac{4}{5} & -\frac{1}{10} & -\frac{1}{15} & -\frac{1}{20} \\ -\frac{1}{10} & \frac{1}{5} & -\frac{1}{30} & -\frac{1}{40} \\ -\frac{1}{15} & -\frac{1}{30} & \frac{4}{45} & -\frac{1}{60} \\ -\frac{1}{20} & -\frac{1}{40} & -\frac{1}{60} & \frac{1}{20} \end{bmatrix}$$

The matrix $(X'X)^{-1}$ exist, thus, it is guaranteed that the four columns of the matrix X are linearly independent. Thus, $\widehat{\beta}$ computation by OLS is feasible.

Considering now two individuals ($i = 1, 2$), one has:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 3 & 0 \\ 4 & 0 \\ 0 & 4 \\ 0 & 5 \\ 0 & 6 \\ 0 & 7 \end{bmatrix} \beta + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \end{bmatrix}$$

thus,

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 2 & 0 & 0 \\ 3 & 0 & 0 & 0 & 3 & 0 \\ 4 & 0 & 0 & 0 & 0 & 4 \\ 0 & 4 & 1 & 0 & 0 & 0 \\ 0 & 5 & 0 & 2 & 0 & 0 \\ 0 & 6 & 0 & 0 & 3 & 0 \\ 0 & 7 & 0 & 0 & 0 & 4 \end{bmatrix}$$

where

$$\mathbf{X'X} = \begin{bmatrix} 30 & 0 & 1 & 4 & 9 & 16 \\ 0 & 126 & 4 & 10 & 18 & 28 \\ 1 & 4 & 2 & 0 & 0 & 0 \\ 4 & 10 & 0 & 8 & 0 & 0 \\ 9 & 18 & 0 & 0 & 18 & 0 \\ 16 & 28 & 0 & 0 & 0 & 32 \end{bmatrix}$$

and

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} \frac{7}{5} & -\frac{2}{3} & -\frac{61}{30} & -\frac{23}{15} & -\frac{41}{30} & -\frac{77}{60} \\ \frac{2}{3} & \frac{1}{3} & -1 & -\frac{3}{4} & -\frac{2}{3} & -\frac{5}{8} \\ -\frac{61}{30} & -1 & \frac{211}{60} & \frac{34}{15} & \frac{121}{60} & \frac{227}{120} \\ -\frac{23}{15} & -\frac{3}{4} & \frac{34}{15} & \frac{439}{240} & \frac{91}{60} & \frac{683}{480} \\ -\frac{41}{30} & -\frac{2}{3} & \frac{121}{60} & \frac{91}{60} & \frac{253}{180} & \frac{19}{15} \\ -\frac{77}{60} & -\frac{5}{8} & \frac{227}{120} & \frac{683}{480} & \frac{19}{15} & \frac{1171}{960} \end{bmatrix}$$

As before, all columns of the matrix X are linearly independent, the matrix $(X'X)^{-1}$ exists, and, consequently, the absence of multicollinearity between the experience variable and the time dummies is guaranted, so $\widehat{\beta}$ computation by OLS is feasible. The inclusion of a set of time dummy variables does not prevent the estimation of the experience variable coefficient.

Now, consider the fixed-effects model where both an individual–specific effect and a time–specific effect are included. The model becomes

$$y_{it} = x_{it}\beta + \alpha_i + \lambda_t + \varepsilon it \qquad i = 1, \ldots, N, \qquad and \qquad t = 1, \ldots, T$$

which can be written as

$$Y = \beta X + D_N\alpha + D_T\lambda + \upsilon$$

Considering $i = 1, 2$ and $t = 1, 2, 3, 4$, one has:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{23} \\ y_{24} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 3 & 0 \\ 4 & 0 \\ 0 & 4 \\ 0 & 5 \\ 0 & 6 \\ 0 & 7 \end{bmatrix} \beta + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \end{bmatrix}$$

The model considers only $T-1$ dummy variables, since one of the time effects must be dropped to avoid perfect collinearity – the group effects and time effects both sum to one Greene (2012).

The *within* transformation is given by:

$$y_{it} - \overline{y}_{i.} - \overline{y}_{.t} + \overline{y}_{..} = \beta(x_{it} - \overline{x}_{i.} - \overline{x}_{.t} + \overline{x}_{..}) + (\varepsilon_{it} - \overline{\varepsilon}_{i.} - \overline{\varepsilon}_{.t} + \overline{\varepsilon}_{..}) \tag{B.3}$$

where,

$$\overline{x}_{i.} = \frac{1}{T}\sum_{t=2}^{T} x_{it} \qquad \overline{x}_{.t} = \frac{1}{N}\sum_{i=1}^{N} x_{it} \qquad \overline{y}_{i.} = \frac{1}{T}\sum_{t=2}^{T} y_{it} \qquad \overline{y}_{.t} = \frac{1}{N}\sum_{i=1}^{N} y_{it}$$

$$\overline{x}_{..} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T} x_{it} \qquad \overline{y}_{..} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T} y_{it}$$

The model may be rewritten as follows:

$$\widetilde{y}_{it} = \widetilde{x}_{it}\beta + \widetilde{\varepsilon}_{it}$$

This is a regression model in deviations from individual and time means, and this called a within transformation. The fixed–effects estimator for β is obtained by pooled OLS from this transformed model B.4, defined as

$$\widetilde{\beta} = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{y}$$

therefore,

$$\widetilde{\mathbf{X}} = \begin{bmatrix} \frac{11}{12} & \frac{11}{12} & \frac{11}{12} & \frac{11}{12} \\ \frac{5}{3} & \frac{5}{3} & \frac{5}{3} & \frac{5}{3} \end{bmatrix}$$

and

$$\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} = \begin{bmatrix} \frac{521}{144} & \frac{521}{144} & \frac{521}{144} & \frac{521}{144} \\ \frac{521}{144} & \frac{521}{144} & \frac{521}{144} & \frac{521}{144} \\ \frac{521}{144} & \frac{521}{144} & \frac{521}{144} & \frac{521}{144} \\ \frac{521}{144} & \frac{521}{144} & \frac{521}{144} & \frac{521}{144} \end{bmatrix}$$

One verifies that the determinant of the matrix $(\widetilde{X}'\widetilde{X})$ is equal to zero, then its inverse does not exist and therefore $(\widetilde{X}'\widetilde{X})$ is a singular matrix. Consequently, it can be stated that there is perfect multicollinearity between the experience variable and the time dummies. In other words, if the individual works every year, *experience* increases one unit each year, thus the experience variable increases in the same amount for all individuals in the sample. The effect of increasing one year of experience cannot be distinguished from aggregated temporal effects (Wooldridge, 2010; Greene, 2012).

B.3 Description of variables included in the empirical models

Table B.1: Description of variables

Variable	Description	Source
Individual level variables		
Real hourly wage	Real hourly wage (Euro) (wages are adjusted using the IPC)	QP
Education	Years of Education calculated by imputing the nominal number of years required to complete the level of schooling	QP
Female	Dummy variable: 1 if female; 0 otherwise	QP
Age	Years of age	QP
Tenure	Years of tenure (years of experience in the current firm), which is equal to the date of the questionnaire minus date of admission	QP
Experience	Years of experience which is calculated by: $exper = age - years\,of\,education - 6$ if $(years\,of\,education + 6) \geq 16$ $exper = age - 16$ if $(years\,of\,education + 6) < 16$	QP
Tenure	Years of tenure (years of experience in the current firm), which is equal to the date of the questionnaire minus date of admission	QP
Occupation	Occupation level inside the firm is a categorical variable 1 = Executive and managerial 2 = Intermediate managerial and executive 3 = Low managerial 4 = Technicians highly specialized 5 = Sales, administrative and precision production 6 = Administrative support, and production 7 = Unskilled 8 = Apprentice 9 = Other & unknown	QP
Firm level variables		
Average schooling	Average education of the workforce in the firm	QP
Share of most qualified workers in the firm	Percentage of workers with 12 years of schooling or more	QP
Skill index in the firm	Adjusted skill index within firm (see B.1 for more details)	QP
Size	Number of workers in the firm	QP
Real sales (1000 Euros)	Real sales (1000 Euros) (sales adjusted using the GDP deflator)	QP

Continued on next page

Table B.1 – continued from previous page

County level variables		
Population	Number of inhabitants in the county	Sales Index 2013
Average schooling	Average education of the workforce in the county	QP
Share of most qualified workers in the county	Percentage of workers with 12 years of schooling or more	QP
County skill index	Adjusted skill index within county (see B.1 for more details)	QP
Attractiveness index	Marktest Index which is considered a overall Index of Development of the county and it is computed as follows: $Attractiveness\,Index = \frac{PopulationIndex}{AreaIndex} \times 100 - 100$, where Attractiveness Index, Population and Area Index are expressed per thousand and positive values of the Attractiveness Index indicate that the county attracts population and negative values indicate the opposite	Sales Index 2013
Housing price	Ratio between the price of evaluation of accommodation per square meter and Lisbon, by county	INE
Unemployment rate	The proxy for unemployment rates is computed using information of the working population (2001 and 2011 Census) and the number of unemployed people registered at Portuguese Employment Centres (IEFP)	INE IEFP
Average size	Average firm size in the county	QP

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting – Sales Index 2013, INE and, *Instituto de Emprego e Formação Profissional–Ministério do Trabalho, Solidariedade e Segurança Social*, 2002–2009.

Table B.2: Descriptive statistics in 2009

Variable	Obs.	Mean	Std. Dev.	Min	Max
Individual–level variables					
Real hourly wage (Euro)	1,882,207	6.225	6.181	1.984	1352
Female	1,882,207	0.442			
Education	1,882,207	9.471	3.966	0	16
$\leq$ 1st cycle		0.168			
2nd cycle		0.183			
3rd cycle		0.254			
Secondary		0.251			
Tertiary		0.163			
Age	1,882,207	38.99	10.59	16	64
Tenure	1,882,207	8.182	8.372	0.083	52.58
Experience	1,881,647	21.54	11.23	0	48
Occupation	1,882,170				
Executive/managerial		0.045			
Intermediate managerial/executive		0.075			
Low managerial		0.128			
Technicians highly specialized		0.164			
Sales, administ. and precision prod.		0.173			
Administ. support and production		0.012			
Unskilled		0.194			
Apprentice		0.102			
Unknown		0.108			
Firm–level variables					
Average schooling	249,868	8.844	3.156	0	16
Share of most qualified workers	249,868	34.56	40.63	0	100
Firm Skill Index	108,983	8.618	3.136	2.510	20.09
Size	249,868	11.22	103.40	1	19,967
Real sales (1000 Euros)	224,460	1,285,199	1.72e+07	5000	2.11e+09
County–level variables					
Population	278	18,140	28,352	846	223,449
Average schooling	278	8.425	0.728	6.879	11.10
Share of most qualified workers	278	29.82	7.806	12.31	60.32
County Skill Index	278	7.565	1.333	5.062	12.83
Attractiveness Index (log form)	278	6.987	0.3283		
Housing price (ratio)	278	46.63	11.25	32.49	100
Unemployment rate	278	5.126	1.763	1.901	12.77
Average size of firms	278	7.516	3.729	1.455	23.42

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting – Sales Index 2013, INE and, *Instituto de Emprego e Formação Profissional – Ministério do Trabalho, Solidariedade e Segurança Social*, 2009.

B.4 Additional tables of results

Table B.3: Additional results of estimation for pooled OLS

VARIABLES	(1)	(2)	(3)	(4)	(5)
Education	0.0124***	0.0325***	0.0261***	0.0449***	0.0452***
	(0.0029)	(0.0011)	(0.0018)	(0.0005)	(0.0005)
Average education	-0.0028			0.0302***	
(county)	(0.0043)			(0.0035)	
Education×Aver. education	0.0037***				
(county)	(0.0003)				
Share of HQ workers		-0.0002			
(county)		(0.0004)			
Education×Share of HQ		0.0004***			
workers (county)		(0.0000)			
Skill index (county)			-0.0032		
			(0.0025)		
Education×skill index			0.0022***		
(county)			(0.0002)		
Log variation coefficient				0.1497***	
				(0.0267)	
Hirschman–Herfindahl					-0.00002***
index					(0.0000)
Female	-0.1551***	-0.1553***	-0.1547***	-0.1520***	-0.1515***
	(0.0032)	(0.0032)	(0.0032)	(0.0033)	(0.0033)
Experience	0.0203***	0.0203***	0.0202***	0.0204***	0.0206***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Experience sq	-0.0003***	-0.0003***	-0.0003***	-0.0003***	-0.0003***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Tenure	0.0141***	0.0141***	0.0141***	0.0141***	0.0141***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Tenure sq	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Log size	-0.0270***	-0.0268***	-0.0270***	-0.0280***	-0.0284***
	(0.0045)	(0.0045)	(0.0045)	(0.0045)	(0.0045)
Log sales	0.0710***	0.0709***	0.0710***	0.0719***	0.0724***
	(0.0031)	(0.0031)	(0.0031)	(0.0030)	(0.0031)
Log aver. size	0.0072*	0.0073*	0.0078*	0.0101***	0.0098**
	(0.0042)	(0.0042)	(0.0042)	(0.0042)	(0.0042)
Log population	0.0032	0.0041*	0.0035	0.0016	0.0015
	(0.0022)	(0.0022)	(0.0022)	(0.0022)	(0.0023)
Log attractiveness	-0.0322***	-0.0368***	-0.0328***	-0.0260**	-0.0177***
index	(0.0046)	(0.0048)	(0.0047)	(0.0047)	(0.0045)
Housing price	0.0014***	0.0014***	0.0014***	0.0017***	0.0024***
(ratio)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Unemployment rate	-0.0023***	-0.0023**	-0.0026***	-0.0046***	-0.0048***
	(0.0009)	(0.0009)	(0.0009)	(0.0009)	(0.0009)
Observations	12,562,901	12,562,901	12,562,557	12,562,901	12,562,901
R–squared	0.634	0.634	0.634	0.632	0.633
RMSE	0.332	0.332	0.332	0.333	0.333

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting – Sales Index 2013, INE and, *Instituto de Emprego e Formação Profissional–Ministério do Trabalho, Solidariedade e Segurança Social*, 2002–2009.

Note: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. The dependent variable is log real hourly wage. All regressions include a set of dummies to control for occupation, sector of activity and year.

Table B.4: Fixed–effects estimation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All workers			**Less qualified workers**			**Most qualified workers**		
Education	0.0127***	0.0127***	0.0132***	0.0079***	0.0079***	0.0082***	0.0509***	0.0509***	0.0516***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0003)	(0.0003)
Average education (county)	0.0105***			0.0074***			0.0109***		
	(0.0001)			(0.0001)			(0.0003)		
Share of HQ workers (county)		0.0011***			0.0008***			0.0011***	
		(0.0000)			(0.0000)			(0.0000)	
Skill index (county)			0.0057***			0.0040***			0.0070***
			(0.0001)			(0.0001)			(0.0001)
Observations	12,584,353	12,584,353	12,584,009	11,135,257	11,135,257	11,134,955	4,137,302	4,137,302	4,137,260
Number of workers	3,341,727	3,341,727	3,341,674	2,970,715	2,970,715	2,970,715	1,127,529	1,127,529	1,127,525
Number of firms	393,674	393,674	393,672	383,818	383,818	383,818	201,548	201,548	201,548
σ_α	0.4384	0.4386	0.4403	0.3623	0.3623	0.3637	0.4747	0.4743	0.4739
σ_ϵ	0.1473	0.1472	0.1473	0.1409	0.1409	0.1409	0.1664	0.1664	0.1664
ρ	0.8986	0.8987	0.8994	0.8687	0.8687	0.8695	0.8906	0.8904	0.8903
F–test	17.42	17.39	17.37	16.02	16.00	15.98	18.09	18.06	18.04
Prob > F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting – Sales Index 2013, INE and, *Instituto de Emprego e Formação Profissional–Ministério do Trabalho, Solidariedade e Segurança Social*, 2002–2009.

Notes: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1; F–test is a statistical test of the hypothesis that all $\alpha_i = 0$. The dependent variable is log real hourly wage. All regressions include individual controls, firm controls and a set of dummies to control for occupation and sector of activity.

Table B.5: Additional results of fixed–effects: Individuals who did not changed the educational level

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All workers			**Less qualified workers**			**Most qualified workers**		
Av. education (county)	0.0056***			0.0056***			0.0101***		
	(0.0002)			(0.0002)			(0.0003)		
Share of HQ (county)		0.0007***			0.0007***			0.0011***	
		(0.0000)			(0.0000)			(0.0000)	
Skill index (county)			0.0040***			0.0035***			0.0069***
			(0.0001)			(0.0001)			(0.0002)
Observations	10,926,530	10,926,530	10,926,240	9,669,703	9,669,703	9,669,446	3,638,781	3,638,781	3,638,748
Number of workers	3,286,360	3,286,360	2,975,360	2,628,317	2,628,317	2,628,269	968,683	968,683	968,679
Number of firms	433,468	433,468	372,964	362,552	362,552	362,548	184,730	184,730	184,730
σ_α	0.4862	0.4850	0.4872	0.3788	0.3783	0.3800	0.5279	0.5276	0.5277
σ_ε	0.1424	0.1424	0.1424	0.1367	0.1367	0.1367	0.1646	0.1646	0.1646
ρ	0.9210	0.9207	0.9213	0.8847	0.8845	0.8854	0.9114	0.9113	0.9114
F–test	20.80	20.78	20.83	17.97	17.96	17.98	21.71	21.68	21.69
Prob > F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting – Sales Index 2013, INE and, *Instituto de Emprego e Formação Profissional–Ministério do Trabalho, Solidariedade e Segurança Social*, 2002–2009.
Notes: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1; F–test is a statistical test of the hypothesis that all $\alpha_i = 0$. The dependent variable is log real hourly wage. All regressions include individual controls, firm controls and a set of dummies to control for occupation and sector of activity.

Table B.6: IV–FE results: All workers, less qualified and most qualified workers

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	***Measure of Human Capital: Average Education of the county***					
	All workers		Less qualified workers		Most qualified workers	
	Polyt. Inst.	Share 1900	Polyt. Inst.	Share 1900	Polyt. Inst	Share 1900
Education	0.0321***	0.0308***	0.0171***	0.0164***	0.0569***	0.0535***
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0004)	(0.0004)
Av. education	-0.0051***	0.0034***	-0.0037***	0.0024***	-0.0049***	0.0107***
(county)	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0006)	(0.0004)
Observations	6,609,603	6,609,603	5,768,666	5,768,666	2,291,400	2,291,400
	Measure of Human Capital: Share of HQ workers in the county					
	All workers		Less qualified workers		Most qualified workers	
	Polyt. Inst.	Share 1900	Polyt. Inst.	Share 1900	Polyt. Inst	Share 1900
Education	0.0319***	0.0309***	0.0170***	0.0165***	0.0567***	0.0540***
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0004)	(0.0004)
Share of HQ	-0.0005***	0.0003***	-0.0003***	0.0002***	-0.0004***	0.0009***
(county)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0000)
Observations	6,609,603	6,609,603	5,768,666	5,768,666	2,291,400	2,291,400
	Measure of Human Capital: Skill Index of county					
	All workers		Less qualified workers		Most qualified workers	
	Polyt. Inst.	Share 1900	Polyt. Inst.	Share 1900	Polyt. Inst	Share 1900
Education	0.0317***	0.0310***	0.0169***	0.0166***	0.0564***	0.0545***
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0004)	(0.0003)
Skill index	-0.0027***	0.0018***	-0.0020***	0.0013***	-0.0027***	0.0060***
(county)	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0003)	(0.0002)
Observations	6,609,409	6,609,409	5,768,495	5,768,495	2,291,377	2,291,377

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting – Sales Index 2013, INE and, *Instituto de Emprego e Formação Profissional–Ministério do Trabalho, Solidariedade e Segurança Social*, 2002–2009.

Notes: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1; F–test is a statistical test of the hypothesis that all $\alpha_i = 0$. The dependent variable is log real hourly wage. All regressions include individual controls, firm controls and a set of dummies to control for occupation and sector of activity.

Table B.7: Fixed–effects estimation including aggregated human capital at firm level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All workers			**Less qualified workers**			**Most qualified workers**		
Education	0.0101***	0.0112***	0.0117***	0.0044***	0.0055***	0.0062***	0.0484***	0.0502***	0.0505***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0003)	(0.0003)
Av. education (county)	0.0084***			0.0056***			0.0066***		
	(0.0001)			(0.0001)			(0.0003)		
Av. education (firm)	0.0128***			0.0130***			0.0149***		
	(0.0001)			(0.0001)			(0.0001)		
Share of HQ (county)		0.0009***			0.0006***			0.0009***	
		(0.0000)			(0.0000)			(0.0000)	
Share of HQ (firm)		0.0010***			0.0010***			0.0010***	
		(0.0000)			(0.0000)			(0.0000)	
Skill index (county)			0.0050***			0.0036***			0.0054***
			(0.0001)			(0.0001)			(0.0001)
Skill index (firm)			0.0106***			0.0097***			0.0099***
			(0.0001)			(0.0001)			(0.0001)
Observations	12,091,700	12,091,700	11,397,056	10,444,174	10,444,174	9,890,349	4,283,954	4,283,954	4,010,356
Number of workers	2,921,773	2,921,773	2,795,927	2,575,643	2,575,643	2,474,471	1,000,658	1,000,658	947,295
Number of firms	140,108	140,108	132,250	139,301	139,301	129,876	97,746	97,746	67,473
σ_α	0.4543	0.4554	0.4620	0.3748	0.3758	0.3794	0.4761	0.4777	0.4823
σ_ϵ	0.1451	0.1451	0.1430	0.1392	0.1392	0.1379	0.1602	0.1602	0.1574
ρ	0.9074	0.9078	0.9126	0.8787	0.8793	0.8833	0.8984	0.8989	0.9036
F–test	18.33	18.35	18.54	16.91	16.94	17.06	19.44	19.58	19.91
Prob > F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting – Sales Index 2013, INE and, *Instituto de Emprego e Formação Profissional–Ministério do Trabalho, Solidariedade e Segurança Social*, 2002–2009.

Notes: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1; F–test is a statistical test of the hypothesis that all $\alpha_i = 0$. The dependent variable is log real hourly wage. All regressions include individual controls, firm controls and a set of dummies to control for occupation and sector of activity..

Table B.8: Fixed–effects estimation including aggregated human capital at inter–regional level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All workers			**Less qualified workers**			**Most qualified workers**		
Education	0.0127***	0.0122***	0.0120***	0.0079***	0.0077***	0.0073***	0.0510***	0.0495***	0.0504***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.0003)	(0.0004)
Av. education (county)	0.0104***			0.0073***			0.0109***		
	(0.0019)			(0.0001)			(0.0002)		
Neighbouring education	0.0003***			-0.0006***			-0.0007***		
	(0.0001)			(0.0001)			(0.0002)		
Share of HQ (county)		0.0011***			0.0008***			0.0012***	
		(0.000)			(0.0000)			(0.0000)	
Share HQ (neighbours)		0.0018***			0.0010***			0.0013***	
		(0.000)			(0.0000)			(0.0001)	
Skill index (county)			0.0057***			0.0041***			0.0077***
			(0.0001)			(0.0001)			(0.0002)
Skill index (neighbours)			0.0013***			0.0002*			0.0008***
			(0.0001)			(0.0001)			(0.0003)
Observations	12,584,376	12,584,376	10,954,331	12,381,330	12,381,330	9,746,562	4,137,327	4,137,327	3,490,440
Number of workers	3,341,727	3,341,727	3,197,091	2,970,715	2,970,715	2,855,241	1,127,529	1,127,529	1,036,473
Number of firms	393,674	393,674	372,072	383,818	383,818	363,555	201,548	201,548	184,818
σ_α	0.4513	0.4355	0.4394	0.3625	0.3606	0.3636	0.4747	0.4758	0.4824
σ_ϵ	0.1491	0.1472	0.1432	0.1409	0.1409	0.1374	0.1664	0.1663	0.1621
ρ	0.9015	0.8975	0.9040	0.8687	0.8676	0.8750	0.8906	0.8911	0.8986
F–test	18.84	17.39	16.83	16.02	16.00	15.50	18.08	18.04	17.63
Prob > F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Source: Computations of the author based on *Quadros de Pessoal*, Marktest Consulting – Sales Index 2013, INE and, *Instituto de Emprego e Formação Profissional–Ministério do Trabalho, Solidariedade e Segurança Social*, 2002–2009.

Notes: Robust standard errors in parentheses, clustered at the firm level. Significance levels: *** p<0.01, ** p<0.05, * p<0.1; F–test is a statistical test of the hypothesis that all $\alpha_i = 0$. The dependent variable is log of real hourly wage. All regressions include individual controls, firm controls and neighbouring controls, and a set of dummies to control for occupation and sector of activity.

Appendix C

Appendices to Chapter 5

C.1 Description of variables included in the empirical models

Table C.1: Description of variables

Variable	Description
Student–level variables	
Mathematics	Score on the Mathematics A exam (12^{th} grade)
Portuguese	Score on the Portuguese exam (12^{th} grade)
Maths 9^{th}	Score on the Maths exam (9^{th} grade)
Portuguese 9^{th}	Score on the Portuguese exam (9^{th} grade)
Female	Dummy variable: 0 for male and 1 for female
Age	Years of student's age
Portuguese student	Dummy variable: 0 for non Portuguese student and 1 for Portuguese student
Internet	Dummy variable: 1 if student has internet at home and 0 otherwise
Beneficiary social support	Dummy variable: 1 if student has social support and 0 otherwise
Parent/legal–guardian	Person responsible for the school education of the student 1 = father; 2 = mother; 3 = own; 4 = other
Parent/legal–guardian education	Education of the responsible for the school education 1 = Tertiary; 2 = Secondary education; 3 = 3^{rd} cycle; 4 = 2^{nd} cycle; 5 = 1^{st} cycle or less
Parent/guardian employment status	Employment status of the responsible for the school education 1 = Worker for others; 2 = Self-employed; 3 = Unemployed; 4 = student; 5 = domestic/retired; 6 = other

Continued on next page

Table C.1 – continued from previous page

Class–level variables	
Class size	Number of students per class
% economically disadvantaged	Share of students in the class receiving social support
% of girls	Share of female students in the class
% of immigrants	Share of immigrant students in the class
% less educated parent/legal–guardian	Share of parents/legal-guardians with secondary education or less in the class
School–level variables	
School size	Number of students per school
% economically disadvantaged	Share of students in the school receiving social support
% of girls	percentage of female students in the school
% of immigrants	Share of immigrant students in the school
% less educated parent/legal–guardian	Share of parents/legal-guardians with secondary education or less in the school
Expenditure–student ratio	Ratio between teaching expenditure and the number of students in school

Source: Created by the author based on *MISI* and *JNE Statistics*, 2010-2012.

C.2 Additional tables of estimations

Table C.2.1: OLS regressions (Year of examination: 2010)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Dependent variable: Log of Mathematics scores					**Dependent variable: Log of Portuguese scores**				
Log Maths 9	0.6047***	0.5887***	0.5942***	0.5843***	0.5946***					
	(0.0180)	(0.0178)	(0.0179)	(0.0177)	(0.0180)					
Log Portuguese 9						0.6552***	0.6526***	0.6531***	0.6516***	0.6529***
						(0.0132)	(0.0132)	(0.0133)	(0.0132)	(0.0132)
Female	0.0884***	0.0913***	0.0986***	0.0989***	0.0987***	0.0637***	0.0638***	0.0591***	0.0594***	0.0590***
	(0.0101)	(0.0101)	(0.0110)	(0.0109)	(0.0109)	(0.0044)	(0.0044)	(0.0048)	(0.0048)	(0.0048)
Age	-0.2027***	-0.1910***	-0.2055***	-0.1910***	-0.2058***	-0.0990***	-0.0988***	-0.1007***	-0.0987***	-0.0999***
	(0.0231)	(0.0226)	(0.0230)	(0.0226)	(0.0230)	(0.0062)	(0.0062)	(0.0062)	(0.0062)	(0.0062)
Beneficiary S.S.	-0.0312**	-0.0228	-0.0179	-0.0177	-0.0178	-0.0110**	-0.0073	-0.0073	-0.0072	-0.0073
	(0.0145)	(0.0146)	(0.0149)	(0.0148)	(0.0149)	(0.0053)	(0.0053)	(0.0055)	(0.0054)	(0.0055)
Internet	0.0195*	0.0342**	0.0139	0.0335**	0.0141	0.0150***	0.0205***	0.0128***	0.0203***	0.0117**
	(0.0108)	(0.0144)	(0.0109)	(0.0144)	(0.0112)	(0.0044)	(0.0054)	(0.0044)	(0.0053)	(0.0045)
Portuguese student	0.0041	-0.0061	0.0091	-0.0104	0.0088	0.0367	0.0520**	0.0441	0.0543**	0.0470*
	(0.0681)	(0.0679)	(0.0680)	(0.0670)	(0.0680)	(0.0272)	(0.0254)	(0.0272)	(0.0254)	(0.0272)
Class size			0.0251**	0.0311***	0.0243**			0.0064*	0.0035	0.0073**
			(0.0098)	(0.0117)	(0.0099)			(0.0034)	(0.0038)	(0.0034)
Class size sq			-0.0006***	-0.0007***	-0.0006***			-0.0002**	-0.0001	-0.0002***
			(0.0002)	(0.0002)	(0.0002)			(0.0001)	(0.0001)	(0.0001)
% of economically			-0.0011**	-0.0007	-0.0011**			-0.0003*	-0.0001	-0.0002
disadvantaged (class)			(0.0004)	(0.0007)	(0.0005)			(0.0002)	(0.0002)	(0.0002)
% of female (class)			-0.0006**	-0.0005	-0.0007**			0.0004***	0.0005***	0.0004***
			(0.0003)	(0.0003)	(0.0003)			(0.0001)	(0.0001)	(0.0001)
% less educated parents/			-0.0009**	-0.0014***	-0.0008*			-0.0007***	-0.0008***	-0.0007***
guardians (class)			(0.0004)	(0.0005)	(0.0004)			(0.0002)	(0.0002)	(0.0002)
School controls	No	No	No	No	Yes	No	No	No	No	Yes
School fixed–effects	No	Yes	No	Yes	No	No	Yes	No	Yes	No
Observations	7,800	7,800	7,800	7,800	7,800	12,773	12,773	12,773	12,773	12,773
R–squared	0.312	0.367	0.316	0.369	0.317	0.307	0.353	0.310	0.354	0.311
RMSE	0.429	0.429	0.429	0.429	0.429	0.231	0.231	0.231	0.231	0.231

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2010.
Note: Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. All regressions include a set of dummies to control for parents'/legal–guardians' education and for district/region.

Table C.2.2: OLS regressions (Year of examination: 2011)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Dependent variable: Log of Mathematics scores					**Dependent variable: Log of Portuguese scores**				
Log Maths 9	0.7060***	0.6666***	0.6884***	0.6599***	0.6866***					
	(0.0307)	(0.0304)	(0.0305)	(0.0302)	(0.0306)					
Log Portuguese 9						0.7250***	0.7065***	0.7164***	0.6995***	0.7161***
						(0.0175)	(0.0173)	(0.0176)	(0.0174)	(0.0176)
Female	0.0575***	0.0607***	0.0748***	0.0754***	0.0748***	0.0762***	0.0760***	0.0726***	0.0717***	0.0725***
	(0.0122)	(0.0124)	(0.0128)	(0.0130)	(0.0128)	(0.0057)	(0.0057)	(0.0061)	(0.0061)	(0.0061)
Age	0.0223	0.0065	0.0175	0.0079	0.0163	-0.0751***	-0.0777***	-0.0757***	-0.0764***	-0.0756***
	(0.0201)	(0.0206)	(0.0200)	(0.0205)	(0.0200)	(0.0063)	(0.0065)	(0.0064)	(0.0065)	(0.0064)
Beneficiary S.S.	-0.0455***	-0.0396**	-0.0371**	-0.0352**	-0.0378**	-0.0071	-0.0044	-0.0051	-0.0031	-0.0047
	(0.0173)	(0.0177)	(0.0176)	(0.0177)	(0.0175)	(0.0070)	(0.0071)	(0.0073)	(0.0072)	(0.0073)
Internet	0.0214	0.0281	0.0125	0.0265	0.0109	-0.0024	0.0068	-0.0043	0.0062	-0.0030
	(0.0179)	(0.0199)	(0.0178)	(0.0198)	(0.0178)	(0.0070)	(0.0078)	(0.0071)	(0.0078)	(0.0071)
Portuguese student	0.1059	0.1116	0.0931	0.1091	0.0881	-0.0355	-0.0192	-0.0422	-0.0194	-0.0416
	(0.0871)	(0.0787)	(0.0851)	(0.0770)	(0.0849)	(0.0278)	(0.0289)	(0.0275)	(0.0283)	(0.0278)
Class size			0.0188	0.0136	0.0179			-0.0001	0.0023	0.0008
			(0.0116)	(0.0140)	(0.0117)			(0.0041)	(0.0046)	(0.0042)
Class size sq			-0.0004*	-0.0003	-0.0004*			0.0000	-0.0000	-0.0000
			(0.0002)	(0.0003)	(0.0002)			(0.0001)	(0.0001)	(0.0001)
% of economically			-0.0000	-0.0007	0.0005			-0.0001	-0.0003	-0.0001
disadvantaged (class)			(0.0006)	(0.0009)	(0.0007)			(0.0002)	(0.0003)	(0.0003)
% of female (class)			-0.0011***	-0.0011***	-0.0012***			0.0005***	0.0006***	0.0006***
			(0.0004)	(0.0004)	(0.0004)			(0.0002)	(0.0002)	(0.0002)
% less educated parents/			-0.0027***	-0.0028***	-0.0026***			-0.0010***	-0.0010***	-0.0011***
guardians (class)			(0.0005)	(0.0007)	(0.0006)			(0.0002)	(0.0003)	(0.0003)
School controls	No	No	No	No	Yes					
School fixed–effects	No	Yes	No	Yes	No	No	Yes	No	Yes	No
Observations	6,109	6,109	6,109	6,109	6,109	9,550	9,550	9,550	9,550	9,550
R–squared	0.284	0.345	0.291	0.349	0.292	0.323	0.371	0.326	0.373	0.327
RMSE	0.467	0.467	0.467	0.467	0.467	0.262	0.262	0.262	0.262	0.262

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2011.
Note: Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. All regressions include a set of dummies to control for parents'/legal–guardians' education and for district/region.

Table C.2.3: OLS regressions (Year of examination: 2012)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Dependent variable: Log of Mathematics scores					**Dependent variable: Log of Portuguese scores**				
Log Maths 9	1.0430***	1.0153***	1.0230***	1.0000***	1.0228***					
	(0.0334)	(0.0337)	(0.0334)	(0.0336)	(0.0334)					
Log Portuguese 9						0.5946***	0.5869***	0.5944***	0.5863***	0.5940***
						(0.0113)	(0.0113)	(0.0113)	(0.0114)	(0.0113)
Female	0.0679***	0.0683***	0.0828***	0.0817***	0.0840***	0.0509***	0.0512***	0.0448***	0.0445***	0.0446***
	(0.0107)	(0.0108)	(0.0115)	(0.0115)	(0.0115)	(0.0044)	(0.0044)	(0.0047)	(0.0047)	(0.0048)
Age	-0.1029***	-0.1086***	-0.1026***	-0.1048***	-0.1022***	-0.1017***	-0.1026***	-0.1015***	-0.1029***	-0.1014***
	(0.0149)	(0.0153)	(0.0148)	(0.0152)	(0.0148)	(0.0052)	(0.0053)	(0.0052)	(0.0053)	(0.0052)
Beneficiary S.S.	-0.0615***	-0.0541***	-0.0544***	-0.0495***	-0.0556***	-0.0073	-0.0069	-0.0072	-0.0056	-0.0073
	(0.0155)	(0.0155)	(0.0159)	(0.0156)	(0.0159)	(0.0057)	(0.0058)	(0.0059)	(0.0059)	(0.0059)
Internet	0.0088	0.0307*	0.0001	0.0274	0.0021	-0.0060	0.0029	-0.0053	0.0029	-0.0060
	(0.0151)	(0.0183)	(0.0151)	(0.0183)	(0.0153)	(0.0058)	(0.0070)	(0.0058)	(0.0069)	(0.0059)
Portuguese student	-0.0086	-0.0116	-0.0170	-0.0137	-0.0112	-0.0351	-0.0417	-0.0420	-0.0454*	-0.0396
	(0.0749)	(0.0803)	(0.0757)	(0.0796)	(0.0756)	(0.0262)	(0.0262)	(0.0267)	(0.0264)	(0.0265)
Class size			0.0329***	0.0371***	0.0304***			0.0107***	0.0134***	0.0112***
			(0.0091)	(0.0118)	(0.0092)			(0.0029)	(0.0033)	(0.0029)
Class size sq			-0.0007***	-0.0008***	-0.0007***			-0.0002***	-0.0003***	-0.0002***
			(0.0002)	(0.0002)	(0.0002)			(0.0001)	(0.0001)	(0.0001)
% of economically			-0.0004	-0.0016*	-0.0008			-0.0000	-0.0003	-0.0005*
disadvantaged (class)			(0.0005)	(0.0008)	(0.0007)			(0.0002)	(0.0003)	(0.0002)
% of female (class)			-0.0008**	-0.0004	-0.0007**			0.0005***	0.0006***	0.0005***
			(0.0003)	(0.0003)	(0.0003)			(0.0001)	(0.0001)	(0.0001)
% less educated parents/			-0.0028***	-0.0043***	-0.0034***			0.0002	0.0003	0.0004
guardians (class)			(0.0004)	(0.0007)	(0.0005)			(0.0002)	(0.0003)	(0.0002)
School controls	No	No	No	No	Yes					
School fixed–effects	No	Yes	No	Yes	No	No	Yes	No	Yes	No
Observations	8,454	8,454	8,454	8,454	8,454	12,583	12,583	12,583	12,583	12,583
R–squared	0.272	0.327	0.280	0.335	0.282	0.340	0.380	0.342	0.382	0.343
RMSE	0.479	0.479	0.479	0.479	0.479	0.235	0.235	0.235	0.235	0.235

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2012.
Note: Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. All regressions include a set of dummies to control for parents'/legal–guardians' education and for district/region.

Table C.3.1: Random effects estimations (Year of examination: 2010)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: Mathematics scores			**Dependent variable: Portuguese scores**		
Log Maths 9	0.5983***	0.5910***	0.5911***			
	(0.0128)	(0.0128)	(0.0128)			
Log Portuguese 9				0.6531***	0.6523***	0.6523***
				(0.0117)	(0.0118)	(0.0118)
Female	0.0895***	0.0987***	0.0987***	0.0638***	0.0594***	0.0593***
	(0.0098)	(0.0105)	(0.0105)	(0.0042)	(0.0045)	(0.0045)
Age	-0.1996***	-0.2005***	-0.2003***	-0.0988***	-0.0993***	-0.0990***
	(0.0172)	(0.0172)	(0.0172)	(0.0054)	(0.0055)	(0.0055)
Beneficiary S.S.	-0.0280**	-0.0184	-0.0182	-0.0081	-0.0074	-0.0073
	(0.0130)	(0.0134)	(0.0134)	(0.0052)	(0.0053)	(0.0053)
Internet	0.0286**	0.0251**	0.0256**	0.0195***	0.0187***	0.0187***
	(0.0118)	(0.0118)	(0.0120)	(0.0051)	(0.0051)	(0.0051)
Portuguese student	-0.0013	-0.0032	-0.0034	0.0476**	0.0504**	0.0516**
	(0.0525)	(0.0529)	(0.0529)	(0.0216)	(0.0218)	(0.0218)
Parent/legal-guardian						
father	0.0062	0.0061	0.0062	0.0101**	0.0098*	0.0099*
	(0.0119)	(0.0119)	(0.0119)	(0.0051)	(0.0051)	(0.0051)
own	0.0328	0.0313	0.0303	-0.0006	0.0006	-0.0016
	(0.0864)	(0.0864)	(0.0866)	(0.0315)	(0.0315)	(0.0315)
other	0.0036	0.0079	0.0084	0.0036	0.0049	0.0050
	(0.0372)	(0.0372)	(0.0374)	(0.0150)	(0.0150)	(0.0149)
Parent/legal-guardian education						
tertiary	0.0846***	0.0745***	0.0745***	0.0572***	0.0530***	0.0530***
	(0.0157)	(0.0160)	(0.0160)	(0.0067)	(0.0068)	(0.0068)
secondary	0.0363**	0.0342**	0.0342**	0.0235***	0.0227***	0.0232***
	(0.0156)	(0.0156)	(0.0156)	(0.0064)	(0.0064)	(0.0064)
2^{nd} *cycle*	-0.0362**	-0.0328*	-0.0328*	-0.0061	-0.0057	-0.0057
	(0.0171)	(0.0171)	(0.0171)	(0.0069)	(0.0069)	(0.0069)
$\leq 1^{st}$ *cycle*	-0.0694***	-0.0657***	-0.0659***	-0.0082	-0.0080	-0.0080
	(0.0186)	(0.0186)	(0.0186)	(0.0074)	(0.0074)	(0.0074)
Parent/guardian employm. status						
self-employed	0.0248	0.0256	0.0256	0.0016	0.0021	0.0020
	(0.0156)	(0.0156)	(0.0156)	(0.0066)	(0.0066)	(0.0066)
unemployed	-0.0293	-0.0279	-0.0276	0.0043	0.0039	0.0041
	(0.0250)	(0.0249)	(0.0249)	(0.0097)	(0.0097)	(0.0097)
student	-0.1226	-0.1209	-0.1217	-0.0157	-0.0157	-0.0156
	(0.0814)	(0.0812)	(0.0813)	(0.0295)	(0.0295)	(0.0295)
domestic/retired	0.0411**	0.0424***	0.0424***	-0.0075	-0.0069	-0.0068
	(0.0160)	(0.0160)	(0.0160)	(0.0066)	(0.0066)	(0.0066)
other	0.0082	0.0105	0.0101	-0.0141	-0.0136	-0.0135
	(0.0519)	(0.0518)	(0.0518)	(0.0223)	(0.0223)	(0.0223)
Class size		0.0271***	0.0266***		0.0048	0.0050
		(0.0095)	(0.0096)		(0.0037)	(0.0037)
Class size sq		-0.0006***	-0.0006***		-0.0001	-0.0001
		(0.0002)	(0.0002)		(0.0001)	(0.0001)

Continued on next page

C.3.1 – continued from previous page

% of economically disadvantaged (class)		-0.0010**	-0.0010*		-0.0001	-0.0001
		(0.0005)	(0.0005)		(0.0002)	(0.0002)
% of female (class)		-0.0006*	-0.0006*		0.0004***	0.0004***
		(0.0003)	(0.0003)		(0.0001)	(0.0001)
% of immigrants (class)		-0.0009	-0.0012		0.0006	0.0010*
		(0.0014)	(0.0015)		(0.0006)	(0.0006)
% less educated parents/ guardians (class)		-0.0010**	-0.0010**		-0.0007***	-0.0007***
		(0.0005)	(0.0005)		(0.0002)	(0.0002)
School size			(0.0000)			(0.0000)
% of economically disadvantaged (school)			0.0001			-0.0005
			(0.0009)			(0.0004)
% of female (school)			0.0014			0.0009
			(0.0020)			(0.0011)
% of immigrants (school)			0.0011			-0.0032**
			(0.0027)			(0.0013)
% less educated parents/guardians (school)			-0.0000			0.0006
			(0.0011)			(0.0006)
Expenditure/student			-0.0000			-0.0000
			(0.0000)			(0.0000)
Observations	7,800	7,800	7,800	12,773	12,773	12,773
Number of schools	282	282	282	284	284	284
RMSE	0.421	0.421	0.421	0.227	0.227	0.227

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2010.
Note: Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Mother is the omitted category in parent/legal–guardian, 3^{rd} cycle is the omitted category in parent/legal–guardian education and employee is the omitted category in parent/legal–guardian employment status.

Table C.3.2: Random effects estimations (Year of examination: 2011)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: Log of Mathematics scores			**Dependent variable: Log of Portuguese scores**		
Log Maths 9	0.6931***	0.6798***	0.6783***			
	(0.0201)	(0.0201)	(0.0202)			
Log Portuguese 9				0.7152***	0.7080***	0.7078***
				(0.0151)	(0.0152)	(0.0152)
Female	0.0583***	0.0746***	0.0747***	0.0764***	0.0723***	0.0722***
	(0.0121)	(0.0128)	(0.0128)	(0.0055)	(0.0059)	(0.0059)
Age	0.0164	0.0136	0.0128	-0.0766***	-0.0763***	-0.0762***
	(0.0163)	(0.0162)	(0.0163)	(0.0057)	(0.0057)	(0.0057)
Beneficiary S.S.	-0.0431***	-0.0360**	-0.0364**	-0.0061	-0.0042	-0.0040
	(0.0163)	(0.0169)	(0.0169)	(0.0069)	(0.0071)	(0.0071)
Internet	0.0239	0.0176	0.0170	0.0017	0.0006	0.0012
	(0.0167)	(0.0167)	(0.0167)	(0.0074)	(0.0074)	(0.0074)
Portuguese student	0.1070	0.0966	0.0949	-0.0289	-0.0328	-0.0324
	(0.0816)	(0.0816)	(0.0817)	(0.0367)	(0.0368)	(0.0368)
Parent/legal-guardian						
father	0.0170	0.0166	0.0160	0.0003	0.0005	0.0008
	(0.0153)	(0.0152)	(0.0152)	(0.0069)	(0.0069)	(0.0069)
own	0.0035	0.0155	0.0176	-0.0486	-0.0455	-0.0460
	(0.0896)	(0.0894)	(0.0895)	(0.0386)	(0.0386)	(0.0386)
other	-0.0075	0.0074	0.0070	-0.0388*	-0.0349*	-0.0346*
	(0.0470)	(0.0470)	(0.0470)	(0.0200)	(0.0200)	(0.0200)
Parent/legal-guardian education						
tertiary	0.1533***	0.1263***	0.1259***	0.0741***	0.0659***	0.0660***
	(0.0192)	(0.0196)	(0.0197)	(0.0087)	(0.0090)	(0.0090)
secondary	0.0396**	0.0346*	0.0338*	0.0283***	0.0274***	0.0277***
	(0.0183)	(0.0182)	(0.0182)	(0.0081)	(0.0081)	(0.0081)
2^{nd} *cycle*	-0.0019	0.0045	0.0044	-0.0066	-0.0048	-0.0048
	(0.0213)	(0.0213)	(0.0213)	(0.0090)	(0.0090)	(0.0090)
$\leq 1^{st}$ *cycle*	-0.0353	-0.0315	-0.0320	-0.0221**	-0.0208**	-0.0206**
	(0.0229)	(0.0229)	(0.0229)	(0.0097)	(0.0097)	(0.0097)
Parent/guardian employ. status						
self-employed	-0.0154	-0.0214	-0.0207	0.0083	0.0081	0.0078
	(0.0203)	(0.0202)	(0.0202)	(0.0090)	(0.0090)	(0.0090)
unemployed	-0.0539*	-0.0547*	-0.0551*	-0.0146	-0.0155	-0.0153
	(0.0296)	(0.0295)	(0.0295)	(0.0122)	(0.0122)	(0.0122)
student	0.0750	0.0661	0.0656	0.0346	0.0328	0.0319
	(0.0867)	(0.0864)	(0.0864)	(0.0373)	(0.0372)	(0.0373)
domestic/retired	-0.0055	-0.0072	-0.0065	0.0183**	0.0180**	0.0177**
	(0.0211)	(0.0210)	(0.0210)	(0.0090)	(0.0090)	(0.0090)
other	0.0557	0.0465	0.0460	0.0411	0.0383	0.0384
	(0.0847)	(0.0844)	(0.0845)	(0.0373)	(0.0373)	(0.0373)
Class size		0.0193*	0.0185*		0.0012	0.0016
		(0.0111)	(0.0112)		(0.0044)	(0.0044)
Class size sq		-0.0004*	-0.0004*		-0.0000	-0.0000
		(0.0002)	(0.0002)		(0.0001)	(0.0001)

Continued on next page

C.3.2 – continued from previous page

% of economically disadvantaged (class)		-0.0002	0.0001		-0.0002	-0.0003
		(0.0006)	(0.0007)		(0.0003)	(0.0003)
% of female (class)		-0.0011***	-0.0012***		0.0006***	0.0006***
		(0.0004)	(0.0004)		(0.0002)	(0.0002)
% of immigrants (class)		-0.0009	-0.0010		-0.0010	-0.0009
		(0.0016)	(0.0017)		(0.0006)	(0.0007)
% less educated parents/ guardians (class)		-0.0026***	-0.0025***		-0.0009***	-0.0010***
		(0.0006)	(0.0006)		(0.0003)	(0.0003)
School size			0.0000			-0.0000*
			(0.0000)			(0.0000)
% of economically disadvantaged (school)			-0.0004			-0.0001
			(0.0009)			(0.0005)
% of female (school)			0.0000			-0.0006
			(0.0021)			(0.0011)
% of immigrants (school)			0.0009			-0.0008
			(0.0027)			(0.0013)
% less educated parents/guardians (school)			-0.0003			0.0003
			(0.0012)			(0.0006)
Expenditure/student			0.0000			-0.0000
			(0.0000)			(0.0000)
Observations	6,109	6,109	6,109	9,550	9,550	9,550
Number of schools	281	281	281	283	283	283
RMSE	0.461	0.461	0.461	0.257	0.257	0.257

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2011.
Note: Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Mother is the omitted category in parent/legal–guardian, 3^{rd} cycle is the omitted category in parent/legal–guardian education and employee is the omitted category in parent/legal–guardian employment status.

Table C.3.3: Random effects estimations (Year of examination: 2012)

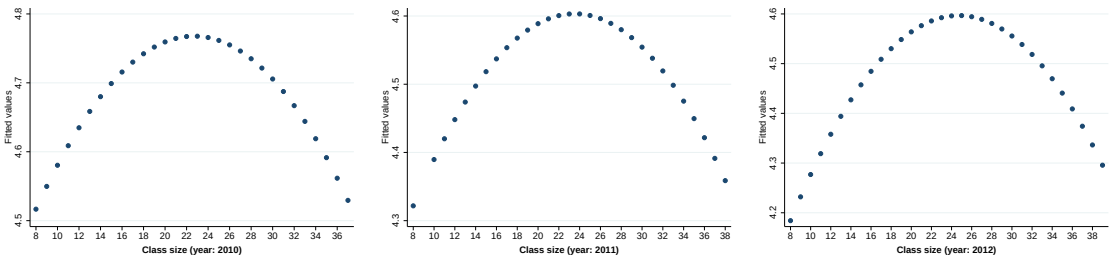
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: Mathematics scores			**Dependent variable: Portuguese scores**		
Log Maths 9	1.0277***	1.0110***	1.0113***			
	(0.0248)	(0.0247)	(0.0247)			
Log Portuguese 9				0.5895***	0.5892***	0.5891***
				(0.0098)	(0.0099)	(0.0099)
Female	0.0689***	0.0816***	0.0824***	0.0511***	0.0445***	0.0444***
	(0.0105)	(0.0111)	(0.0111)	(0.0043)	(0.0046)	(0.0046)
Age	-0.1058***	-0.1044***	-0.1036***	-0.1023***	-0.1024***	-0.1023***
	(0.0130)	(0.0129)	(0.0129)	(0.0045)	(0.0045)	(0.0045)
Beneficiary S.S.	-0.0575***	-0.0522***	-0.0527***	-0.0077	-0.0065	-0.0065
	(0.0144)	(0.0146)	(0.0146)	(0.0055)	(0.0056)	(0.0056)
Internet	0.0187	0.0131	0.0147	-0.0011	-0.0006	-0.0012
	(0.0159)	(0.0159)	(0.0159)	(0.0063)	(0.0063)	(0.0063)
Portuguese student	-0.0065	-0.0109	-0.0077	-0.0390	-0.0446	-0.0437
	(0.0736)	(0.0735)	(0.0734)	(0.0286)	(0.0287)	(0.0287)
Parent/legal-guardian						
father	-0.0063	-0.0081	-0.0084	0.0106*	0.0110**	0.0114**
	(0.0135)	(0.0134)	(0.0134)	(0.0054)	(0.0054)	(0.0054)
own	0.1096	0.1241*	0.1302*	-0.0218	-0.0187	-0.0191
	(0.0745)	(0.0742)	(0.0742)	(0.0300)	(0.0300)	(0.0300)
other	-0.0561	-0.0457	-0.0452	0.0139	0.0131	0.0141
	(0.0396)	(0.0395)	(0.0395)	(0.0150)	(0.0150)	(0.0150)
Parent/legal-guardian education						
tertiary	0.1460***	0.1188***	0.1191***	0.0667***	0.0694***	0.0694***
	(0.0166)	(0.0169)	(0.0169)	(0.0067)	(0.0069)	(0.0069)
secondary	0.0368**	0.0289*	0.0285*	0.0082	0.0094	0.0092
	(0.0158)	(0.0158)	(0.0158)	(0.0062)	(0.0063)	(0.0063)
2^{nd} *cycle*	-0.0014	0.0017	0.0011	-0.0108	-0.0113	-0.0118
	(0.0185)	(0.0184)	(0.0184)	(0.0072)	(0.0072)	(0.0072)
$\leq 1^{st}$ *cycle*	0.0067	0.0111	0.0106	-0.0086	-0.0080	-0.0083
	(0.0205)	(0.0205)	(0.0205)	(0.0079)	(0.0079)	(0.0079)
Parent/guardian employ. status						
self-employed	-0.0053	-0.0047	-0.0044	-0.0070	-0.0067	-0.0072
	(0.0172)	(0.0171)	(0.0171)	(0.0070)	(0.0070)	(0.0070)
unemployed	-0.0234	-0.0217	-0.0219	0.0033	0.0041	0.0040
	(0.0237)	(0.0236)	(0.0236)	(0.0090)	(0.0090)	(0.0090)
student	-0.0695	-0.0777	-0.0800	-0.0010	-0.0050	-0.0055
	(0.0737)	(0.0734)	(0.0734)	(0.0294)	(0.0294)	(0.0294)
domestic/retired	-0.0136	-0.0143	-0.0141	-0.0044	-0.0043	-0.0052
	(0.0186)	(0.0185)	(0.0185)	(0.0072)	(0.0072)	(0.0072)
other	-0.0071	-0.0031	-0.0035	-0.0076	-0.0068	-0.0063
	(0.0586)	(0.0584)	(0.0583)	(0.0238)	(0.0238)	(0.0238)
Class size		0.0324***	0.0315***		0.0120***	0.0123***
		(0.0095)	(0.0096)		(0.0030)	(0.0030)
Class size sq		-0.0007***	-0.0007***		-0.0002***	-0.0002***
		(0.0002)	(0.0002)		(0.0001)	(0.0001)

Continued on next page

C.3.3 – continued from previous page

% of economically disadvantaged (school)		-0.0007	-0.0012*		-0.0002	-0.0004*
		(0.0006)	(0.0007)		(0.0002)	(0.0002)
% of female (class)		-0.0006*	-0.0005		0.0005***	0.0005***
		(0.0003)	(0.0003)		(0.0001)	(0.0001)
% of immigrants (class)		-0.0001	0.0001		-0.0006	-0.0003
		(0.0017)	(0.0017)		(0.0006)	(0.0006)
% less educated parents/ guardians (class)		-0.0033***	-0.0038***		0.0003	0.0004
		(0.0005)	(0.0006)		(0.0002)	(0.0002)
School size			(0.0000)			(0.0000)
SESescola1			0.0010			0.0006
			(0.0010)			(0.0004)
% of female (school)			-0.0008			0.0004
			(0.0023)			(0.0009)
% of immigrants (school)			-0.0015			-0.0018
			(0.0032)			(0.0013)
% less educated parents/guardians (school)			0.0021*			-0.0006
			(0.0012)			(0.0005)
Expenditure/student			0.0000			-0.0000
			(0.0001)			(0.0000)
Observations	8,454	8,454	8,454	12,583	12,583	12,583
Number of schools	320	320	320	322	322	322
RMSE	0.470	0.470	0.470	0.231	0.231	0.231

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2012.
Note: Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1. Mother is the omitted category in parent/legal–guardian, 3^{rd} cycle is the omitted category in parent/legal–guardian education and employee is the omitted category in parent/legal–guardian employment status.



Source: Created by the author based on *MISI* and *JNE Statistics*, 2010–2012.

Figure C.1: Fitted values for the class size (in Mathematics) with a 2^{nd} degree polynomial model, in 2010, 2011 and 2012

C.3 DEA efficiency scores

Table C.4: Efficiency scores of public schools: DEA results

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Output: Mathematics scores							**Output: Portuguese scores**						
Year of examination: 2010													
Rank	School	Efficiency score	School	Efficiency score	School	Efficiency score	Rank	School	Efficiency score	School	Efficiency score	School	Efficiency score
1	41	1	105	1	105	1	1	250	1	250	1	86	1
2	105	1	2	1	314	1	2	81	1	46	1	242	1
3	81	1	41	1	146	1	3	46	0.9992	105	1	323	1
4	305	0.9963	184	1	242	1	4	323	0.9404	86	1	105	1
5	242	0.9547	81	1	305	1	5	59	0.9289	59	1	146	1
6	51	0.9343	305	0.9963	41	1	6	140	0.9271	184	1	108	1
7	172	0.9267	55	0.9703	2	1	7	261	0.9220	81	1	46	1
8	33	0.9195	242	0.9547	55	1	8	169	0.9132	333	0.9598	250	1
9	97	0.9126	51	0.9343	323	1	9	65	0.9118	323	0.9420	48	1
10	48	0.9081	274	0.9325	184	1	10	247	0.9097	74	0.9395	81	1
11	251	0.8988	172	0.9267	81	1	11	105	0.9061	64	0.9370	59	1
12	221	0.8973	141	0.9205	51	0.9641	12	321	0.9043	247	0.9285	184	1
13	146	0.8949	33	0.9195	48	0.9498	13	108	0.9041	140	0.9276	65	0.9860
14	272	0.8947	97	0.9126	241	0.9455	14	221	0.9020	261	0.9250	314	0.9747
15	295	0.8922	48	0.9081	274	0.9363	15	204	0.8962	271	0.9202	247	0.9643
16	346	0.8814	272	0.9050	97	0.9335	16	64	0.8944	150	0.9158	64	0.9642
17	195	0.8790	251	0.8988	172	0.9315	17	346	0.8889	108	0.9143	333	0.9598
18	241	0.8759	221	0.8973	33	0.9273	18	113	0.8836	169	0.9135	51	0.9594
19	274	0.8748	146	0.8949	141	0.9176	19	86	0.8833	65	0.9120	74	0.9406
20	128	0.8743	295	0.8922	177	0.9113	20	153	0.8829	151	0.9098	238	0.9396

Continued on next page

Table C.4 – continued from previous page

Rank	School	Efficiency score	School	Efficiency score	School	Efficiency score	Rank	School	Efficiency score	School	Efficiency score	School	Efficiency score
262	302	0.5679	292	0.6016	192	0.6071	264	281	0.7078	251	0.7205	244	0.7302
263	286	0.5667	302	0.5981	292	0.6016	265	136	0.7068	248	0.7159	281	0.7296
264	254	0.5502	348	0.5935	281	0.6014	266	5	0.7042	60	0.7156	327	0.7274
265	317	0.5473	207	0.5913	302	0.5981	267	285	0.7030	147	0.7145	248	0.7272
266	6	0.5458	303	0.5838	348	0.5935	268	18	0.7019	203	0.7133	303	0.7260
267	304	0.5449	6	0.5808	207	0.5913	269	110	0.7006	290	0.7085	60	0.7167
268	343	0.5440	281	0.5788	303	0.5838	270	201	0.7004	281	0.7082	203	0.7133
269	153	0.5422	153	0.5721	304	0.5811	271	184	0.6966	136	0.7072	290	0.7085
270	147	0.5384	304	0.5714	6	0.5808	272	249	0.6939	285	0.7031	136	0.7072
271	235	0.5344	317	0.5679	147	0.5714	273	199	0.6909	201	0.7009	201	0.7063
272	348	0.5332	152	0.5631	317	0.5696	274	171	0.6873	199	0.6911	199	0.6911
273	303	0.5285	147	0.5541	152	0.5681	275	240	0.6845	171	0.6873	171	0.6877
274	278	0.5117	278	0.5513	254	0.5609	276	304	0.6791	69	0.6857	69	0.6857
275	152	0.5091	254	0.5502	198	0.5536	277	8	0.6790	304	0.6847	304	0.6847
276	5	0.5041	343	0.5440	278	0.5513	278	69	0.6734	8	0.6796	8	0.6796
277	244	0.4951	198	0.5431	343	0.5440	279	245	0.6702	245	0.6704	245	0.6749
278	198	0.4946	18	0.5401	18	0.5401	280	152	0.6656	152	0.6697	152	0.6697
279	113	0.4917	244	0.5187	244	0.5258	281	307	0.6590	307	0.6603	159	0.6620
280	240	0.4730	240	0.5078	240	0.5136	282	159	0.6317	159	0.6451	307	0.6603
281	18	0.4602	113	0.4917	113	0.4917	283	31	0.6220	182	0.6331	182	0.6331
282	301	0.4201	301	0.4756	301	0.4811	284	182	0.5985	31	0.6222	31	0.6318

Continued on next page

Table C.4 – continued from previous page

Year of examination: 2011													
Rank	School	Efficiency score	School	Efficiency score	School	Efficiency score	Rank	School	Efficiency score	School	Efficiency score	School	Efficiency score
1	46	1	46	1	46	1	1	346	1	86	1	48	1
2	139	1	74	1	242	1	2	280	1	346	1	86	1
3	244	1	139	1	271	1	3	39	0.9910	104	1	316	1
4	241	1	90	1	251	1	4	332	0.9278	280	1	74	1
5	251	1	101	1	74	1	5	251	0.9204	332	1	44	1
6	346	0.9914	251	1	54	1	6	90	0.9060	90	1	271	1
7	238	0.9631	244	1	44	1	7	48	0.8972	39	0.9910	346	1
8	3	0.9339	3	1	90	1	8	104	0.8926	4	0.9803	39	1
9	193	0.9302	241	1	139	1	9	86	0.8831	251	0.9544	242	1
10	325	0.9249	346	0.9914	101	1	10	4	0.8756	9	0.9365	251	1
11	101	0.9138	325	0.9801	241	1	11	305	0.8686	48	0.9327	104	1
12	182	0.9040	238	0.9631	238	1	12	158	0.8671	170	0.9219	280	1
13	169	0.8994	182	0.9381	256	1	13	125	0.8658	321	0.9216	332	1
14	83	0.8940	193	0.9316	244	1	14	28	0.8604	125	0.9148	90	1
15	232	0.8891	232	0.9279	3	1	15	321	0.8603	327	0.9110	256	0.9839
16	172	0.8775	9	0.9081	346	0.9973	16	193	0.8588	202	0.9077	4	0.9803
17	78	0.8717	141	0.9052	325	0.9801	17	170	0.8489	88	0.9075	99	0.9600
18	103	0.8685	169	0.8994	264	0.9652	18	88	0.8413	179	0.9073	112	0.9367
19	74	0.8674	83	0.8916	182	0.9381	19	136	0.8342	13	0.9069	9	0.9365
20	187	0.8647	26	0.8795	169	0.9332	20	108	0.8283	193	0.9023	156	0.9308

Continued on next page

Table C.4 – continued from previous page

Rank	School	Efficiency score	School	Efficiency score	School	Efficiency score	Rank	School	Efficiency score	School	Efficiency score	School	Efficiency score
259	323	0.5554	303	0.5698	41	0.5747	261	63	0.6321	83	0.6672	343	0.6715
260	53	0.5540	136	0.5624	235	0.5722	262	245	0.6301	63	0.6663	303	0.6686
261	333	0.5538	6	0.5606	303	0.5698	263	167	0.6299	345	0.6632	167	0.6677
262	181	0.5485	41	0.5601	6	0.5606	264	350	0.6289	80	0.6630	83	0.6672
263	180	0.5472	323	0.5554	323	0.5554	265	210	0.6271	146	0.6623	63	0.6663
264	6	0.5459	53	0.5547	53	0.5547	266	102	0.6256	317	0.6586	285	0.6645
265	303	0.5458	180	0.5472	288	0.5530	267	171	0.6254	211	0.6573	80	0.6630
266	77	0.5327	181	0.5467	77	0.5487	268	232	0.6238	285	0.6528	317	0.6586
267	288	0.5279	77	0.5402	180	0.5472	269	314	0.6218	350	0.6521	345	0.6576
268	317	0.5275	288	0.5364	181	0.5467	270	3	0.6210	245	0.6494	211	0.6573
269	339	0.5173	317	0.5264	317	0.5342	271	83	0.6177	76	0.6431	350	0.6521
270	343	0.5142	10	0.5219	45	0.5246	272	141	0.6142	300	0.6428	245	0.6494
271	20	0.5137	339	0.5173	10	0.5219	273	303	0.6124	144	0.6374	300	0.6428
272	45	0.5116	343	0.5142	339	0.5211	274	144	0.6107	240	0.6355	144	0.6374
273	306	0.5028	45	0.5116	343	0.5142	275	40	0.6065	210	0.6338	240	0.6355
274	10	0.5012	306	0.5028	306	0.5087	276	317	0.6049	171	0.6254	210	0.6338
275	140	0.4697	140	0.4725	268	0.4950	277	340	0.6004	340	0.6237	171	0.6254
276	29	0.4563	29	0.4549	140	0.4725	278	240	0.5809	101	0.6222	340	0.6237
277	240	0.4338	268	0.4403	29	0.4549	279	101	0.5772	40	0.6196	101	0.6222
278	1	0.4310	240	0.4330	240	0.4540	280	113	0.5755	113	0.6149	40	0.6196
279	268	0.4281	1	0.4305	1	0.4305	281	54	0.5599	54	0.5986	113	0.6149

Continued on next page

Table C.4 – continued from previous page

Year of examination: 2012													
Rank	School	Efficiency score	School	Efficiency score	School	Efficiency score	Rank	School	Efficiency score	School	Efficiency score	School	Efficiency score
1	217	1	217	1	132	1	1	193	10.000	279	1	279	1
2	136	1	136	1	134	1	2	261	10.000	193	1	48	1
3	301	0.9527	275	1	217	1	3	125	10.000	125	1	132	1
4	125	0.9337	301	0.9527	64	1	4	279	10.000	261	1	188	1
5	64	0.8898	125	0.9337	242	1	5	111	0.9713	175	1	193	1
6	241	0.8735	64	0.8898	346	1	6	50	0.9431	315	1	216	1
7	346	0.8554	269	0.8749	316	1	7	335	0.9431	2	1	242	1
8	168	0.8494	241	0.8735	216	1	8	236	0.9414	66	1	50	1
9	242	0.8459	346	0.8554	275	1	9	2	0.9284	26	0.9977	251	1
10	161	0.8434	168	0.8494	136	1	10	224	0.9284	105	0.9976	111	1
11	48	0.8107	3	0.8461	7	0.9969	11	13	0.9196	111	0.9966	125	1
12	183	0.8080	242	0.8459	237	0.9926	12	25	0.9190	277	0.9959	335	1
13	28	0.8077	161	0.8434	48	0.9859	13	251	0.9189	25	0.9953	336	1
14	305	0.8036	14	0.8424	76	0.9770	14	328	0.9174	209	0.9935	261	1
15	3	0.8018	315	0.8375	301	0.9551	15	241	0.9163	236	0.9900	209	1
16	129	0.7977	74	0.8352	241	0.9519	16	64	0.9157	9	0.9880	74	1
17	251	0.7972	53	0.8259	256	0.9511	17	15	0.9147	74	0.9840	175	1
18	238	0.7952	263	0.8187	251	0.9455	18	321	0.9142	19	0.9794	315	1
19	146	0.7942	310	0.8130	74	0.9375	19	255	0.9098	317	0.9766	2	1
20	254	0.7901	48	0.8107	125	0.9337	20	4	0.9082	20	0.9759	105	1

Continued on next page

Table C.4 – continued from previous page

Rank	School	Efficiency score	School	Efficiency score	School	Efficiency score	Rank	School	Efficiency score	School	Efficiency score	School	Efficiency score
297	226	0.4813	337	0.4970	80	0.5119	299	348	0.7315	211	0.7364	214	0.7470
298	252	0.4797	329	0.4898	139	0.5051	300	210	0.7303	350	0.7362	350	0.7446
299	347	0.4782	226	0.4811	276	0.5043	301	107	0.7273	339	0.7349	107	0.7440
300	329	0.4759	252	0.4797	120	0.5037	302	182	0.7261	186	0.7334	113	0.7402
301	229	0.4725	347	0.4782	327	0.5024	303	343	0.7160	23	0.7316	210	0.7369
302	123	0.4706	229	0.4725	337	0.4970	304	280	0.7151	348	0.7315	211	0.7364
303	250	0.4699	123	0.4706	250	0.4910	305	79	0.7141	210	0.7303	23	0.7343
304	114	0.4564	250	0.4699	329	0.4898	306	32	0.7130	107	0.7273	348	0.7315
305	348	0.4557	114	0.4564	226	0.4811	307	205	0.7094	280	0.7166	60	0.7208
306	352	0.4542	348	0.4557	123	0.4747	308	60	0.7086	8	0.7161	62	0.7197
307	35	0.4538	35	0.4538	35	0.4729	309	222	0.7046	343	0.7160	39	0.7177
308	147	0.4516	147	0.4516	229	0.4725	310	62	0.7037	79	0.7147	343	0.7163
309	215	0.4470	215	0.4470	114	0.4564	311	352	0.6957	32	0.7130	8	0.7161
310	303	0.4022	303	0.4022	348	0.4557	312	39	0.6952	60	0.7086	79	0.7147
311	188	0.3991	188	0.3991	69	0.4364	313	214	0.6927	62	0.7037	280	0.7144
312	69	0.3965	69	0.3965	303	0.4022	314	8	0.6894	39	0.6952	182	0.7108
313	100	0.3479	100	0.3636	100	0.3830	315	309	0.6790	309	0.6790	309	0.6907
314	266	0.3143	266	0.3143	266	0.3143	316	136	0.6644	136	0.6780	136	0.6780
315	338	0.3119	338	0.3119	338	0.3096	317	81	0.6548	226	0.6557	226	0.6557
316	208	0.2768	208	0.2768	208	0.2768	318	226	0.6147	81	0.6548	81	0.6452
317	81	0.1575	81	0.1575	81	0.1881	319	59	0.5666	59	0.5670	59	0.5825

Source: Computations of the author based on *MISI* and *JNE Statistics*, 2010–2012.

Appendix D

Appendices to Chapter 6

D.1 Description of variables included in the empirical models

Table D.1: Description of variables

Variable	Description
Student–level variables	
12th national exam score	Score on the Mathematics exam and Portuguese exam
9th national exam score	Score on the Mathematics and Portuguese language of 9th exam
Female student	Dummy variable: 0 for male and 1 for female
Age	Years of student's age
Internet	Dummy variable: 1 if student has internet at home and 0 otherwise
Beneficiary social support	Dummy variable: 1 if student has social support and 0 otherwise
Teacher–level variables	
Female teacher	Dummy variable: 0 for male teacher and 1 for female teacher
Advanced degree	Dummy variable: 0 if teacher's academic education is *bacharelato* or bachelor degree and 1 if teacher's academic education is postgraduate studies, master degree or PhD
Experience	Years of experience which is calculated by the ratio between the total number of days of service and 365 days
Commuting	Dummy variable: 1 if if the residence county is different from the county in which the school where the teacher works is located and 0 otherwise
Class size	Number of students per class

Source: Created by the author based on *MISI* and *JNE Statistics*, 2010–2012.