Essays on Wage Inequality and Human Capital in Spain

Manuel A. Hidalgo Pérez [1] [2]

April 2008

[1]Universitat Pompeu Fabra. Graduate Program of Economics, Finance and Management (GPEFM)

[2]Address: Ctra de Utrera, s/n, km 1, CP 41013. Sevilla. Telephone: 0034.954.97.79.79, E-mail: mhidper@upo.es.

Contents

List of Figures

List of Tables

Acknowledgements

First of all, I would like to thank my advisor Antonio Ciccone for his guidance, patience and suggestions in all stages of writing my thesis. I am especially indebted to him for his support and for encouraging me all this time. I would also like to thank Juan Carlos Conesa, José Ignacio García, Walter García-Fontes, Teresa García-Milá, José García Montalvo, Paula González, Sergi Jiménez, José Luis Pinto, Nicolás Porteiro, Jesús Rodríguez López, Sara de la Rica, Carmen Mª Rubio, José Luis Torres and Carlos Usabiaga and all the members of the VI Workshop in International Economics in Malaga and the III Jornadas de Economia de Granada the helpful of their comments. I would also like to thank my colleagues and friends who have provided comments, discussion, encouragement and technical advice on various aspects of this project: Judit Montoriol, Claudia Canals, Emilia García-Appendini, Raúl González, Josep Mestres, Josepa Miquel, Thomas Rangel, José Luis Fillat and Gustavo Solorzano. I would like to thank to my family and Manuel Rubio. Finally I also specially thank to Francisco Ferraro and José Mª O'Kean without whom this thesis does not exist.

A Carmen

Chapter 1

Wage Inequality in Spain, 1980-2000

1.1 Introduction

Wage inequality has grown substantially since the 1970s in many countries. In the United States, for example, the 90/10 percentile ratio of male hourly earnings grew by 23.4 log points between 1979 and 2003 (Autor, Katz, and Kearney 2005a). This ratio increased by about 20 log points in the UK between the early 1980s and late 1990s, with similar increases taking place in Germany between the early 1980s and the mid-1990s. In Canada during the 1990s, the ratio increased by 15 log points. These figures contrast with slight increases or even declines in other countries such as the Netherlands, Sweden or Belgium, as summarized by Acemoglu (2003).

This paper aims to provide a detailed analysis of wage inequality trends in Spain between 1980 and 2000s, providing a level of detail similar to that found in analyses for other countries. Specifically, it will examine the trends in overall wage inequality and decompose these trends into three components, corresponding to three types of inequality: that which responds to changes in the wage premium paid for education and experience (between-group inequality, BI); that which responds to the observed skill distribution (composition effects); and that which responds to changes in wage dispersion among workers with the same levels of experience and education (residual wage inequality, RWI).

The literature on this subject mentions a number of approaches to the decomposition of changes in overall wage inequality into between-group, composition effects, and residual wage inequality. Juhn, Murphy, and Pierce (1993)'s seminal study extended the "mean" Oaxaca-Blinder procedure to include the decomposition of distributions. In applying this approach, Juhn, Murphy and Pierce (1993) first estimated yearly Mincerian wage equations to obtain returns for education and experience, as well as a measure of the residual wage distribution (wage dispersion that cannot be explained by differences in education, experience or other observable worker characteristics). They then used the results to simulate counterfactual densities (the wage distributions that would prevail if some parameters, such as the return to education, were changed but the rest remained constant). Lemieux's (2002, 2006) decomposition of changes into overall wage inequality extends the kernel procedure of DiNardo, Fortin, and Lemieux (1996). An advantage of his approach is that it allows residual wage inequality to be a function of workers' observable characteristics. Finally, Autor, Katz, and Kearney (2005a) extend Machado and Mata's (2005) quantile decomposition technique. We have chosen to rely on the latter approach, because it incorporates the work of both Juhn, Murphy and Pierce and that of DiNardo, Fortin and Lemieux.

Our analysis uses data from the Household Budgets Surveys for 1980-81 and 1990-91 and the Continuous Household Budgets Surveys for 1985-86, 1990-91, 1995-96 and 2000-01. These are the only data sources that contain all of the information needed for this study and that cover the entire period of interest to us here. By contrast, other data sources either cover shorter time periods or lack key information on individual workers.

We find that overall wage inequality in Spain, measured as the difference between log wages at the 90th and 10th distribution percentiles, increased only very slightly between 1980 and 2000. More specifically, we find a decrease in wage inequality during the 1980s, compensated for by an increase during the 1990s. Within each of these decades, wage inequality initially increased and then decreased during the later years of the decade. Given these results, it is interesting to note that GDP grew below trend during the early 1980s and the 1990s and above trend during the final years of both decades. Hence, overall wage inequality behaved countercyclically within each of the two decades.[1] For the 1980s, our findings are on a par with those of earlier studies, published during the 1980s. Arellano, Bentolila, and Bover (2001), for example, used data from the Social Security archives to show that wage dispersion rose during the first half of the 1980s, while Abadie (1997) demonstrated that wage inequality fell slightly throughout the 1980s on the basis of data from the Household Budget Survey for 1980-81 and 1990-91.

We also decompose changes in wage inequality into between-group inequality (the distance between distributions conditional to one observable characteristic), residual inequality (inequality within conditional distributions), and composition effects (changes in wage inequality that arise due as the result of a shift in labor force composition). Between-group inequality behaves much like overall inequality. This is true for the entire 20-year period, as well as for and within each of the two decades encompassed by that period. The composition effect contributes positively to overall wage inequality until the late 1990s. Residual inequality, on the other hand, only begins to increase during the early 1990s.

One advantage of our decomposition method is that it allows us to exam-

[1]For a discussion of the link between business cycle and wage inequality in the US and the U.K. see Dimelis and Alexandra (1999).

ine wage inequality trends at different wage distribution points. In particular, we look at wage gaps situated in the lower part of the distribution, between the 50th and the 10th percentiles (where the gap is 50/10), and in the upper part of the distribution between the 90th and the 50th percentiles (where the gap is 90/50). Regarding overall inequality, we find that the 90/50 wage gap behaves qualitatively like the 90/10 gap. The 50/10 wage gap, on the other hand, has been increasing since the second half of the 1980s.

A close look at these figures reveal that between-group inequality evolves similarly in the upper and the lower part of the wage distribution; thus, both the 50/10 and the 90/50 gaps behave qualitatively like the 90/10 gap with respect to between-group inequality. At a quantitative level, however, the changes are more marked in the upper half of the distribution. By contrast, residual inequality behaves rather differently when we look at the upper half instead of the lower half of the distribution. While the 50/10 gap has been increasing since the second half of the 1980s, there is no clear pattern to be found with respect to the 90/50 gap. Hence, in the upper part of the distribution the rising overall wage inequality shown is mirrored by rising between-group inequality, but not by residual inequality. In the lower part of the distribution, on the other hand, wage inequality behaves much like residual inequality but not like between-group inequality. The different patterns of wage inequality above and below the median, and the behavioral differences that can be observed between the different components of our analysis, suggest that there was no unique driving force behind inequality trends in Spain during the period under study.[2]

The rest of the paper is organized as follows: Section 1.2 reviews the related literature, Section 1.3 describes our data, Section 1.4 gives a preliminary survey of wage inequality between 1980 and 2000, Section 1.5 explains

[2]Autor, Katz, and Kearney (2005a) reach the same conclusion for the US and Abadie (1997) for Spain in the 1980s.

our decomposition method, Section 1.6 discusses our main empirical results, and Section 1.7 concludes.

1.2 Related Literature

Wage inequality and the decomposition of overall wage inequality into between-group inequality (BI, hereafter), composition effects, and residual wage inequality (RWI) has generated a large body of research to date. Most of the literature uses one of three methods: the full variance accounting method presented by Juhn, Murphy and Pierce (1993); the semi-parametric procedure of DiNardo, Fortin, and Lemieux (1996); or the Mata and Machado (2005) approach as extended by Autor, Katz, and Kearney (2005a).

The full variance accounting decomposition put forth by Juhn, Murphy and Pierce (1993) (JMP) begins with the supposition that wages can be characterized by the canonical Mincer equation

$$w_{it} = X_{it}\beta_t + u_{it}, \tag{1.1}$$

where w_{it} denotes the wage logs at time t for individual i, X_{it} is a specific set of individuals and environmental characteristics that may potentially affect wages, β_t is a set of returns or prices that set the value of these characteristics, and u_{it} is a compendium of non-observable characteristics that may potentially affect individual wages.

Within this framework, wage inequality changes may come from three sources: changes in the distribution of observable characteristics (changes in the distribution of X_{it}); changes in the prices of observable skills (changes in β_t); and changes in the wage distribution accounted for by unobservable characteristics (u_{it}). With this structure, JMP estimate the counterfactual densities that would prevail if any subset of these three components were held to be constant, finding that residual inequality accounted for a great deal of the overall increase in U.S. wage inequality between 1964 and 1988. They also demonstrate that nearly all of the rising inequality explained by observables is due to changes in returns on observable skills (rather than to changes in the distribution of observable skills). While they find that residual wage inequality began to increase during the late 1960s, they also find that increasing returns on skills have contributed to a higher overall wage inequality only since the early 1980s. Their favorite explanation for these wage inequality trends is that the demand for both observable and unobservable skills has been increasing, and that this demand has translated into higher skill prices.[3]

A shortcoming of this approach is that it does not fully account for the links between observable characteristics and residual wage dispersion, as pointed out by Lemieux (2006).[4] For example, hourly wage dispersion in the U.S. is typically found to be greater for college graduates than for less educated workers (Autor, Katz and Kearny, 2005a, 2005b). As a result, a rise in the number of college degree-holding workers may cause overall wage inequality to increase as the result of a "mechanical" composition effect. Such composition effects working through residual wage dispersions must be accounted for, since they might easily be confused with the effects of changing prices on unobservable skills (Lemieux, 2006 and Autor, Katz and Kearny, 2005a, 2005b). Lemieux's (2002,2006) extension of the approach developed by DiNardo, Fortin and Lemieux (1996) was designed to resolve this problem. Lemieux modelled overall residual wage dispersion as a weighted average of wage dispersions by skill group. Under this model, changes in the distribution of observable skills will also change the balance of wage dispersion by skill group, thereby mechanically changing the overall wage dispersion. Lemieux

[3]An important open question is why BI and RWI started rising at different times.

[4]In principle, the approach of JMP should be able to deal with this issue. But Lemieux (2006) argues that it does not do so in practice.

(2006), who uses this approach in order to examine wage inequality trends in the U.S., finds that composition effects play a greater role and changes in RWI play a smaller role than they do in JMP (composition effect now encompasses all changes in wage inequality caused by changes in skill composition, including those caused by re-weighted residual wage dispersions by skill; RWI are changes in wage inequality that are unaccounted for by composition effects and BI). He also argues that a great many of the changes in RWI can be explained by institutional factors (rather than by changes in skill prices), such as the declining real minimum wage.

Our empirical work will rely on the latest decomposition method appearing in the literature on this subject, Autor, Katz and Kearny's (2005a) (AKK) extension of the Machado and Mata (2005) (MM) quantile decomposition approach, which is explained in detail in Section 5. This method corrects the shortcomings of the original full distribution accounting method developed by Juhn, Murphy, and Pierce (1993), and nests the approach proposed by DiNardo, Fortin and Lemieux (1996) and Lemieux (2002,2006). As applied to U.S. data (Autor, Katz and Kearny, 2005a, 2005b), the AKK approach confirms the importance of the composition effects emphasized by Lemieux (2006) for the lower part of the wage distribution. But in the upper part of the wage distribution, rising wage inequality is found to be almost entirely explained by rising prices for observable skills and by greater RWI (the two main driving forces behind wage inequality emphasized by JMP).

Wage inequality in Spain has been studied by Abadie (1997), Arellano, Bentolila, and Bover (2001) and Izquierdo and Lacuesta (2007). Abadie's analysis examines wage inequality trends during the 1980s using quantile regressions. He documents a fall in the return to education during this period, which mostly affects the lower part of the distribution for younger workers and the upper part for elderly workers. Our approach differs from his in that it allows for a detailed characterization of composition effects as well as for wage inequality trends within and between groups. Arellano, Bentolila, and Bover use a large Social Security data sample to examine wage inequality trends for the 1980-1987 period. Their analysis focuses on the behavior of returns to skill and experience both over time and across sectors. Izquierdo and Lacuesta (2007) use non-parametric techniques to analyze Spanish wage inequality between 1995 and 2002 using the Wage Structure Survey. They show that changes in the return to education and tenure decreased inequality, while changes in composition increased the inequality. Our approach differs from theirs with respect to period of analysis, method, and the special attention paid in our study to within-group inequality.

1.3 Data

Let us now review the data sets that can be used to analyze wage inequality trends in Spain.

Household surveys. All of the information necessary for our analysis can be found in the "Encuesta de Presupuestos Familiares" or EPF (Household Budget Survey) for 1980-81 and 1990-91 and its newer counterpart, the quarterly Encuesta Continua de Presupuestos Familiares or ECPF (Continuous Household Budget Survey), available from 1985 to 2005 data sets. While they provide useful data on wages, education, age, gender, such information is only available for heads of families. Nevertheless, the alternatives (some of which are explained below) present even greater drawbacks, making these surveys the main source of information for the study of wage inequality trends (Oliver, Raymond, Roig, and Barceinas, 1999). Appendix A gives more details on these surveys.

Other data sources. There are two other of wage surveys in Spain, the Wage Structure Survey and the (Quarterly) Labour Cost Survey, both compiled by the Spanish National Institute of Statistics (*Instituto Nacional de Estadistica* or INE). The first of these provides information for about two thousand industrial and service workers for the years 1995, 2002 and, recently, 2006. While the individual information in this survey is very detailed, the time span covered is too short to be useful here. The (Quarterly) Labour Cost Survey (previously called the Survey of Wages in Industry and Services) surveys wages for the 1980-2000 period but provides no information on education levels.[5]

The Spanish Social Security records provide another source of information on wages. The large sample size and the length of the period covered make this an appealing data set for those analyzing changes in wage inequality over time, (see Arellano, Bentolila and Bover, 2001). Recently, the "Muestra Contínua de Vidas Laborales" from the Social Security records further develops the information used by these authors, covering a panel of workers and an expanded number of years and providing more data on individual worker characteristics. Also, for 2004 and 2006 this new survey has information on fiscal variables. The drawback of using these records is that, with the exception of 2004 and 2006, wages are not directly stated but have to be inferred from social security contributions.

[5]Moreover, sample selection is at the firm level, which implies that the sample may not be representative of Spanish workers.

1.4 Descriptive Statistics and Wage Inequality

Our analysis of wage inequality trends in Spain is based on the data for heads of household who work full-time provided in the EPF surveys for 1980-81 and 1990-91, and in the ECPF for 1985-1986, 1990-1991, 1995-1996 and 2000-2001. The reasons for this selection are described in Appendix A. One drawback of this selection is the smaller female participation as compared with other exercises. For comparative purposes, some of the consequences of this lower participation will be pointed out when we present our results.

Some important characteristics of these data are reported in Table 1.1. It can be seen that individual yearly earnings, in 2000 pesetas, rose from 2,172,996 in 1980 to 2,513,765 in 2000. This corresponds to a real growth rate of about 15.6% over 20 years. The wage dispersion, measured as the standard deviation of wages, also grew between 1980 and 2000. Within the 20-year period, real average wages grew unevenly. Between 1980 and 1985 real wages increased only minimal; however, they increased substantially between 1985 and 1990. Real wage growth during the 1990s was more even, however. Other Spanish wage statistics for the same period yield similar results (Survey of Wages in Industry and Services or National Accounts).

Table 1.1 also shows that the average age of workers oscillated between 42 and 43 years during the sample period. Average years of schooling, on the other hand, rose by almost 50 percent between 1980 and 2000 (from 8.63 years to 12.45 years). While schooling rose in nearly every country in the world during this period, this rise was much more pronounced in Spain than it was in most other countries (Acemoglu 2003). The percentage of female heads of household who work full-time has also grown since the EPF80/81.

Some trends in wage inequality can be analyzed without a full decomposition approach. Table 1.2 plots the difference between log wages at the 90th

Table 1.1: EPF-80/81-90/91 and ECPF 85/86-90/91-95/96-00/01 main characteristics

	EPF		ECPF			
	80/81	90/91	85/86	90/91	95/96	00/01
number	7,027	8,193	2,683	1,965	1,696	2,057
av. wage (c.p. 2000)	2,172,996	2,282,235	2,165,996	2,350,950	2,430,859	2,513,765
sd. dv.	252,818	259,729	233,101	270,562	277,327	310,028
wage growth rate		0.49		1.65	0.67	0.67
age	42.02	42.93	41.58	41.83	41.99	43.60
schooling	8.63	9.28	10.49	11.02	11.69	12.45
women (%)	6.06	9.37	5.14	6.36	9.35	11.74

Notes: The row *av. wage (c.p. 2000)* shows average wages for each surveys at 2000 constant pesetas. *Sd. Dv.* are standard deviations. *Wage growth rate* gives the average growth rate per year between the year corresponding to the column the data and the previous one. The rows *age* and *schooling* present the average age and years of schooling for each survey. Lastly, *women* are the women participation in each sample.

and 10th percentiles of the distribution (the 90/10 log wag gap), between log wages at the 90th and 50th percentiles (the 90/50 log wag gap), and between log wages at the 50th and 10th percentiles (the 50/10 log wag gap). It can be seen that between 1980 and 1990, the 90/10 log wage gap fell by about 15 log points (average real wages at the 90th percentile rose by only 1.8% per year during this period, while at the 10th percentile they registered an increase of 3.3% per year). Between 1985 and 2000, however, 90/10 wage inequality grew by 36 points. Inequality trends for the 1980s differ qualitatively in the lower and upper halves of the distribution. Looking at the lower tail, it can be seen that the 50/10 log wage gap fell by 29 points. By contrast, the 90/50 gap increased by 14 points, indicating greater wage inequality in the upper half of the distribution. During the 1990s, the rise in inequality was concentrated entirely in the lower tail, with a sharp 55 log-point rise rise at the 50/10 gap. Inequality in the upper tail decreased somewhat during this period.

Analyzing different subperiods and parts of the wage distribution yields further interesting results. First, the lower half seems to present increasing inequality between 1985 and 1990. The distance in log wages between the

10th and 50th percentiles falls between 1980 and 1990 but remains constant between 1985 and 1990. This gap also increases between 1990 and 1995, and between 1995 and 2000. However, the upper half shows a different trend for the entire twenty-year period. For each decade, the 90/50 log wage gap increased during the early years when GDP growth was low, but fell during the later years, when GDP grew more rapidly.[6] Wage inequality above the median thus seems to follow a countercyclical pattern. This result is important enough to merit further attention in the following sections.

Wage inequality trends may be driven by changes in between-group or within-group wage dispersion, or by changes in the composition of the labor force. As a preliminary means of isolating changes in between-group inequality due to changes in education premia, we estimate Mincerian wage regressions. Table 1.3 shows different measures of the return to education, all obtained using Mincerian wage equations. All premia/penalties are obtained using Mincerian wage regressions that include age, age squared, and gender as additional explanatory variables.[7] In particular, this Table shows education wage premia earned by college-educated workers relative to those with only a secondary education ("college"), and also the wage penalty for primary-schooled workers relative secondary-schooled ones ("primary"). The results show that wage differentials between primary- and secondary-schooled workers narrowed continuously between 1980 and 2000. This tendency should

[6]Spanish real GDP growth was 1.5%, between 1980 and 1985, 4.5% between 1985 and 1990, 1.3% between 1990 and 1995, and 3.9% between 1995 and 2000.

[7]These results are derived from a standard Mincer equation that takes the following form:

$$w_{it} = \alpha_t + \beta_{1t} p_{it} + \beta_{2t} c_{it} + \gamma_{1t} age_{it} + \gamma_{2t} age^2_{it} + \rho_t sex_{it} + e_{it}$$

where w_{it} is the log of individual wage, (p_t, c_t) is the vector of educational dummies, $(age_{it}, age^2_{it}, sex_{it})$ is the vector of other wage-influencing variables and $(\alpha_t, \beta_{1t}, \beta_{2t}, \gamma_{1t}, \gamma_{2t}, \rho_t)$ are the returns on wages for these variables.

Table 1.2: Wage Inequality Change 1980-2000

	EPF		ECPF			
	1980/81	1990/91	1985/86	1990/91	1995/96	2000/01
90th-10th	0,926	0,911	0,966	0,960	1,003	1,002
90th-50th	0,520	0,534	0,543	0,537	0,541	0,526
50th-10th	0,406	0,377	0,422	0,422	0,462	0,477

Note: Each row represent distances in logs between the wage distribution percentiles represented at the first column.

Table 1.3: Between-group and residual inequality. Mincerian Wage regressions

	EPF		ECPF			
	1980/81	1990/91	1985/86	1990/91	1995/96	2000/01
primary	-0.311	-0.282	-0.286	-0.223	-0.212	-0.202
	(0.008)	(0.008)	(0.014)	(0.015)	(0.018)	(0.018)
college	0.266	0.283	0.308	0.2961	0.3526	0.316
	(0.011)	(0.011)	(0.021)	(0.023)	(0.024)	(0.02)
age	0.037	0.048	0.055	0.059	0.064	0.035
	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)	(0.007)
res 90-10	0.778	0.714	0.773	0.778	0.816	0.828
res 90-50	0.425	0.394	0.417	0.404	0.419	0.420
res 50-10	0.354	0.319	0.356	0.375	0.398	0.407
R^2	0.29	0.35	0.32	0.280	0.29	0.220
years of sch.	0.063	0.060	0.079	0.064	0.075	0.070
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)

Note: Each three first rows are the standard Mincerian regression coefficients for education (represented by two dummies, the first if the worker has only primary or less education and second if the worker has at least college education) and age (square of age and gender are not showed). The following three rows are the distance in the 90, 50 and 10 residuals percentiles which evaluates changes in non-observed, residual or within-group inequality. R^2 show the goodness of fit, and years of schooling is the return to education if we use average years of schooling instead of the previous dummies in the Mincerian regression.

have reduced wage inequality. By contrast, the college/high school wage premium evidenced a more uneven trend. It grew between 1980 and 1985 and also between 1990 and 1995, but fell during the remaining periods. Hence, the college-high school premium seems to follow a counter-cyclical pattern.

The bottom row of Table 1.3 shows another standard statistic of the return to schooling, the average return to an additional year of schooling.[8] This statistic shows a fall in the return to schooling during the 1980s and a modest increase from 1990 onwards.

Age/experience contributed to increasing wage inequality between 1980 and 1995. The return to experience shows clearly the increase in the inter-group wage gap defined by this characteristic. During the 1980s there was a sharp increase of about 30%. This trend remains roughly similar, but slightly lower, before breaking in 1995. The last five years see a contraction in BI inequality given by age/experience.

Table 1.3 also shows statistics for residual wage inequality. For now, we simply evaluate changes in the RWI using the distribution of OLS residuals. The analysis of RWI is important because it explains about two thirds of the overall inequality in our data (this number is a quite standard finding using the Mincerian approach). The rows in Table 1.3 reveal a contrast before and after 1990. The 1980s' trend was negative, despite the increase in the second half of that decade. This was only possible thanks to the considerable decrease in residual wage inequality that took place during the first half of the decade. Thus, the increase in wage inequality since 1985 appears to be due to residual inequality.

[8] In this case, instead of introducing dummies to capture the school effect on wages, here we used years of schooling. Then, the estimated mincerian equation is defined by

$$w_{it} = \alpha_t + \beta_t s_{it} + \gamma_{1t} age_{it} + \gamma_{2t} age_{it}^2 + \rho_t sex_{it} + e_{it}$$

where, in this case s_{it} gives the years of schooling of individual i.

Figure 1.1: Wage inequality, college and primary/high-school wage premium and residual inequality

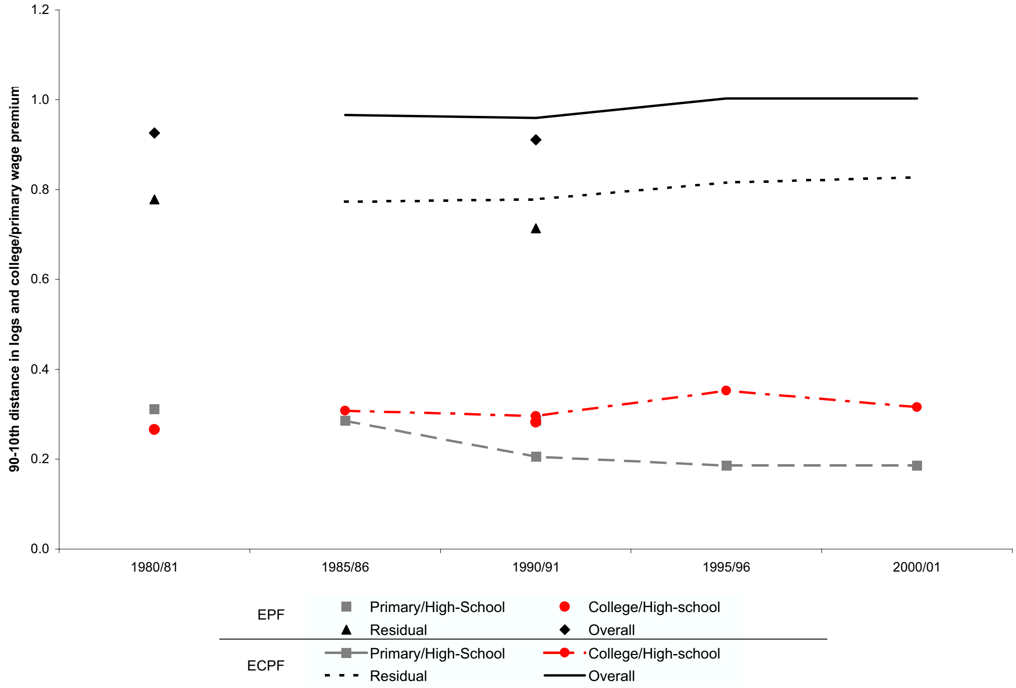


Figure 1.1 summarizes the trends in 90/10, the college wage premium, the primary wage penalty, and residual inequality (the data come from Table 1.3). Differences in wage inequality trends between decades are evident. The graph shows clearly that wage inequality fell during the 1980s, widened between 1990 and 1995, and remained constant until 2000. The college premium rose until the last period (1995-2000). Changes in the primary wage penalty tended to reduce inequality up to the year 2000. Thus, up to the year 1995, education reduced inequality for less educated workers - those in the lower tail of the wage distribution- and increased inequality in the upper tail.

Overall, residual inequality appears to have run parallel to wage inequality, except during the period between 1985 and 1990 when residual inequality increased while total inequality fell. During 1985-1990, changes in wage inequality seem to mirror changes in BI.

Our analysis so far does not allow us to assess the role of changes in labor force composition for changes in wage inequality. As a simple, preliminary way to examine this issue, we estimated Table 1.3 again, but holding the labor force composition constant at its 1980 value. To do so, workers in each survey year were classified by level of education (primary, high college and college), age (one-year intervals from 20 to 65 years), and gender, which yielded 276 cells. We then generated counterfactual samples for each survey year by combining average worker characteristics in each cell with 1980 sample weights. Estimating Table 1.3 using these counterfactual samples yields the results in Table 1.4. It can be seen that differences with Table 1.3 are very small, which suggest that labor force composition did not play a major role in Spanish wage inequality trends.

The main limitation of our preliminary analysis (based on OLS estimation) is that it does not allow for differences in the evolution of prices at different parts of the wage distribution. Similarly, it does not show how changes in the labor force composition affect wages at different parts of the

Table 1.4: Between-group and residual inequality. Minceerian Wage regressions (1980 constant weights)

	EPF		ECPF			
	1980/81	1990/91	1985/86	1990/91	1995/96	2000/01
Primary	-0.311 (0.008)	-0.261 (0.012)	-0.271 (0.017)	-0.206 (0.020)	-0.186 (0.025)	-0.186 (0.018)
College	0.266 (0.011)	0.295 (0.015)	0.338 (0.032)	0.308 (0.030)	0.357 (0.034)	0.295 (0.02)
Age	0.037 (0.002)	0.051 (0.003)	0.056 (0.005)	0.063 (0.005)	0.072 (0.011)	0.039 (0.009)
Res 90-10	0.778	0.713	0.777	0.782	0.819	0.842
Res 90-50	0.425	0.391	0.423	0.406	0.428	0.424
Res 50-10	0.354	0.322	0.354	0.376	0.390	0.417
R^2	0.29	0.35	0.32	0.280	0.29	0.220

Note: Each three first rows are the standard Mincerian regression coefficients for education (represented by two dummies, the first if the worker has only primary or less education and second if the worker has at least college education) and age (square of age and gender are not showed). The following three rows are the distance in the 90, 50 and 10 residuals percentiles which evaluates changes in non-observed, residual or within-group inequality. R^2 show the goodness of fit, and finally years of schooling is the return to education if we use average years of schooling instead of the previous dummies in the Mincerian regression. Constant weights implies that each regression are made using 1980 population structure for observable variables.

wage distribution. Moreover, our approach may have overestimated the degree to which residual inequality contributes to changes in wage inequality by not fully accounting for links between observable characteristics and residual wage dispersion, as emphasized by Lemieux (2006). The following section explains a quantile regression technique that can be used to resolve these issues.

1.5 Methodology

The decomposition technique proposed by Machado and Mata (2005,MM) and extended by Autor, Katz and Kearny (2005a,AKK) uses quantile regressions to decompose wage distributions into "price" and "quantity" components. These components are then used to assess the importance of changes in prices and quantities in explaining wage inequality trends using counterfactual analysis. Our exposition of the MM and AKK methodology follows AKK.

Let $Q_\theta(w_t|x_t)$ for $\theta \in (0,1)$ be log wages (w_t) at the θ^{th} quantile of the distribution of wages given the vector of k covariates x_t for year t. Assume that the conditional quantiles can be represented as a linear function of covariates, or more formally, that there are $k \times 1$ vectors of quantile regression coefficients $\beta_t(\theta)$ such that

$$Q_\theta(w_t|x_t) = x_t'\beta_t(\theta). \tag{1.2}$$

If $Q_\theta(w_t|x_t)$ in (1.2) is specified correctly, $x_t'\beta_t(\theta)$ provides a full characterization of the distribution of wages given the covariates x_t. The distribution of w_t given x_t can therefore be obtained by (i) repeatedly drawing θ from a uniform distribution on the open interval $(0,1)$; (ii) obtaining the price vector $\beta_t(\theta)$ corresponding to θ; (iii) calculating $x_t'\beta_t(\theta)$. In general, the specification in (2) must be thought of as an approximation. How good the approximation turns out to be is easy to check and depends on the particular application. For the case of Spain, we will show that the resulting approximation is quite accurate.

As is well known, for a given θ, the vector $\beta_t(\theta)$ can be estimated by solving the following minimization problem,

$$\min_\beta \quad n^{-1}\sum_{i=1}^{n} \rho_\theta(w_{it} - x_{it}'\beta_t); \tag{1.3}$$

where the ρ_θ is a "check function" (Koenker and Bassett 1978),

$$\rho_\theta(u) \equiv \begin{cases} \theta u & \text{for } u \geq 0 \\ (\theta - 1)u & \text{for } u < 0; \end{cases}$$

where in this case $u = w_{it} - x_{it}'\beta_t$. This method estimates $\beta_t(\theta)$ consistently under conditions similar to those required for the asymptotic consistency of OLS estimation.

Once $\beta_t(\theta)$ has been estimated for $\theta_i, i = 1, ...m$ with a large value for m spread over the open unit interval, the distribution of wages given x_t is obtained as $\{\hat{w}_{it} = x_{it}'\hat{\beta}_t(\theta_i)\}_{i=1}^m$, where hats denotes estimated values.

This describes the simulation of the wage data for any given x, but does not provide the marginal density of w. The marginal density also depends on the distribution of the covariates, which we will denote by $g(x)$. The marginal density of w is obtained by (i) repeatedly drawing rows of data from $g(x)$, x_i; (ii) drawing corresponding θ_i from a uniform (0,1) distribution; (iii) obtaining wages as $\hat{w}_i \equiv x_i'\hat{\beta}(\theta_i)$.

The MM conditional quantile decomposition procedure has two important properties. First, like the Oaxaca-Blinder OLS procedure (Oaxaca, 1973; Blinder, 1973), it separates the observed wage distribution into price and quantity components. But while the Oaxaca-Blinder procedure only characterizes the central tendency of wages (between-group wage differences), the MM approach characterizes both the central tendency of wages and their dispersion (linked to residual wage inequality). This is a key issue if research is aimed at decomposing wage inequality into composition effects, between inequality, and residual inequality. Second, under the assumption that aggregate quantities of skills in the labor market do not affect skill prices (a strong but convenient assumption), the conditional quantile model can be used to simulate how changes in the labor force composition or skill prices affect the distribution of wages. For example, to see what wages would have prevailed with the labor force composition of period t, $g_t(x)$, and labor market prices of period s, $\beta_s(\theta)$, one simply simulates wages using $g_t(x)$ and $\beta_s(\theta)$.

AKK extend the MM approach to a counterfactual analysis of residual wage inequality. Their approach uses the skill price vector at the 50th percentile, $\beta(0.5)$, to characterize changes in between-group inequality. Hence, like the OLS price vector in the Oaxaca-Blinder decomposition, $\beta(0.5)$ is used to estimate the central tendency of the data conditional on x. Within-group inequality is quantified using the difference between the estimated coefficient vector $\beta(\theta)$ and the median coefficient vector $\beta(0.5)$,[9]

$$\beta^w(\theta) \equiv \beta(\theta) - \beta^b \equiv \beta(\theta) - \beta(0.5). \tag{1.4}$$

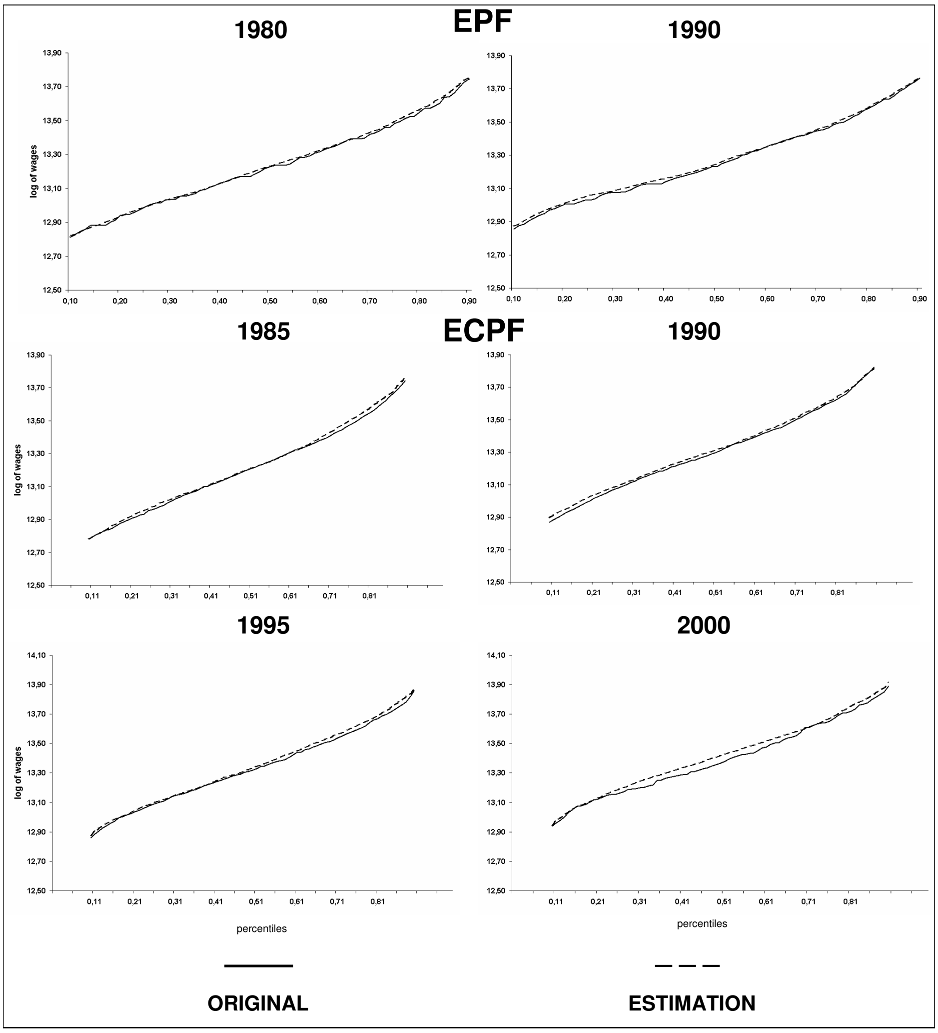
Hence, there are now three ingredients in each wage simulation exercise: (i) estimated "within"coefficient vectors $\hat{\beta}^w(\theta)$ for a large number of θs spread over the open unit interval; (ii) an estimated "between"coefficient vector $\hat{\beta}^b \equiv \hat{\beta}(0.5)$; (iii) and the distribution of covariates, $g(x)$. AKK perform counterfactual analysis by changing one of these three elements at a time. This allows them to assess the contribution of residual wage inequality (changes in the wage distribution when changing $\hat{\beta}^w(\theta)$ only), between group inequality (changes in the wage distribution when changing $\hat{\beta}^b$ only), and labor force composition effects (changes in the wage distribution when changing $g(x)$ only).

1.6 Results

In this section, we perform the following four-step exercise in order to decompose wage inequality change. First, four thousand θ were selected for each year from a uniform distribution $U(0, 1)$. Next, the quantile regressions for each percentile were estimated, the counterfactual exercise described in the previous section was implemented and, lastly the comparison are made.

[9]See Appendix B for a prove.

Figure 1.2: Original versus QR Estimated Wage Distributions' Percentiles.



Solid lines represent original wage distributions percentiles. Dotted lines show the simulated wage distribution percentiles using the Machado and Mata (2005) algorithm.

Before analyzing the results, we have to check whether the method used the conditional quantiles with accuracy. Figure 1.2 describes both the original as well as the simulated quantiles. Since the simulated percentiles match the originals, the accuracy of this procedure for Spanish data is almost perfect, with the exception of the year 2000. Even for this year, the error seems be restricted to levels (constant at about 3%); thus, it vanishes when we compare the evolution of distances between percentiles. So, to the exent that we are only concerned with log changes, the magnitude of the error is negligible. With the exception of this case, the average error ranges from 0.8% (for the 1989 data) and 1.5% for the 1995 data. Thus, we conclude that the MM algorithm is a suitable tool for decomposing changes in wage distributions.

Overall Wage Inequality

Table 1.5 shows the log change in 10th-50th, 50th-90th and 10th-90th percentile distances, with the last of these representing what we call overall wage inequality. The first result shown in the top panel of the table refers to the fall in inequality during the eighties and late nineties and the sharp rise in inequality during the early nineties. This result is very similar to that found in section 1.4. For the whole period, the result is a slight increase. Despite the lack of observations for the 1980-1985 period, inequality might have grown during the first half of the eighties despite the over fall for that decade. This intuition is driven by the 4.7 percent fall in wage inequality between 1980 and 1990 shown by the EPF, when the ECPF shows it to have contracted by a higher rate, 6.7 log points, during the second half of the period under study.[10]

[10] These results are complementary to those found by other Spanish studies, for different periods and using different data: Arellano, Bentolila, and Bover (2001) using Social Security Records for 1981-1987, Abadie (1997) for the eighties as a whole and Izquierdo and Lacuesta (2007) since 1995.

Table 1.5: 100 $\times$ log changes in Overall, Between Groups and Residual

	EPF 80-90	ECPF 85-00		
		85-90	90-95	95-00
overall				
10-50	-4.6	-1.3	5.1	1.4
50-90	-0.1	-5.3	2.6	-4.4
10-90	-4.7	-6.7	7.7	-2.9
prices				
10-50	-4.4	-3.2	3.7	1.9
50-90	-0.7	-7.2	0.2	-1.8
10-90	-5.1	-10.3	3.9	0.1
quantity				
10-50	-0.2	1.8	1.4	-0.5
50-90	0.5	1.9	2.4	-2.6
10-90	0.3	3.7	3.9	-3.1
between				
10-50	-0.8	-1.0	0.5	-2.1
50-90	-1.8	-3.5	1.6	-2.6
10-90	-2.6	-4.4	2.1	-4.7
residual				
10-50	-3.6	-2.2	3.2	4.0
50-90	1.1	-3.7	-1.4	0.7
10-90	-2.5	-5.9	1.8	4.8

Note: Each value represents changes in log percentiles distances. *Overall* defines simulated wages distributions using MM algorithm. *Prices* represents counterfactuals distributions when only returns to skills (education and experience) change, *quantity* when only the distribution of attributes change, *between* represents changes between counterfactual densities when only princes changes when all of the workers has the same returns to skills no matter the location they are within the distribution of wages, and *residual* shows the inequality evolution using counterfactuals densities when only changes in returns distributions are evaluated.

Examining the two halves of the density instead of the entire distribution gives rise to a familiar question. As we pointed out earlier, the two halves do not always share the same trend; thus, total inequality depends on which half prevails. While the upper tail increased during the 1980-1985 and 1990-1995 periods and decreased during the 1985-1990 and 1995-2000 ones, the lower half was highly negative during the eighties (especially between 1980 and 1985) and positive during the nineties. Thus, the causes driving total wage inequality seem to differ above and below the median.[11]

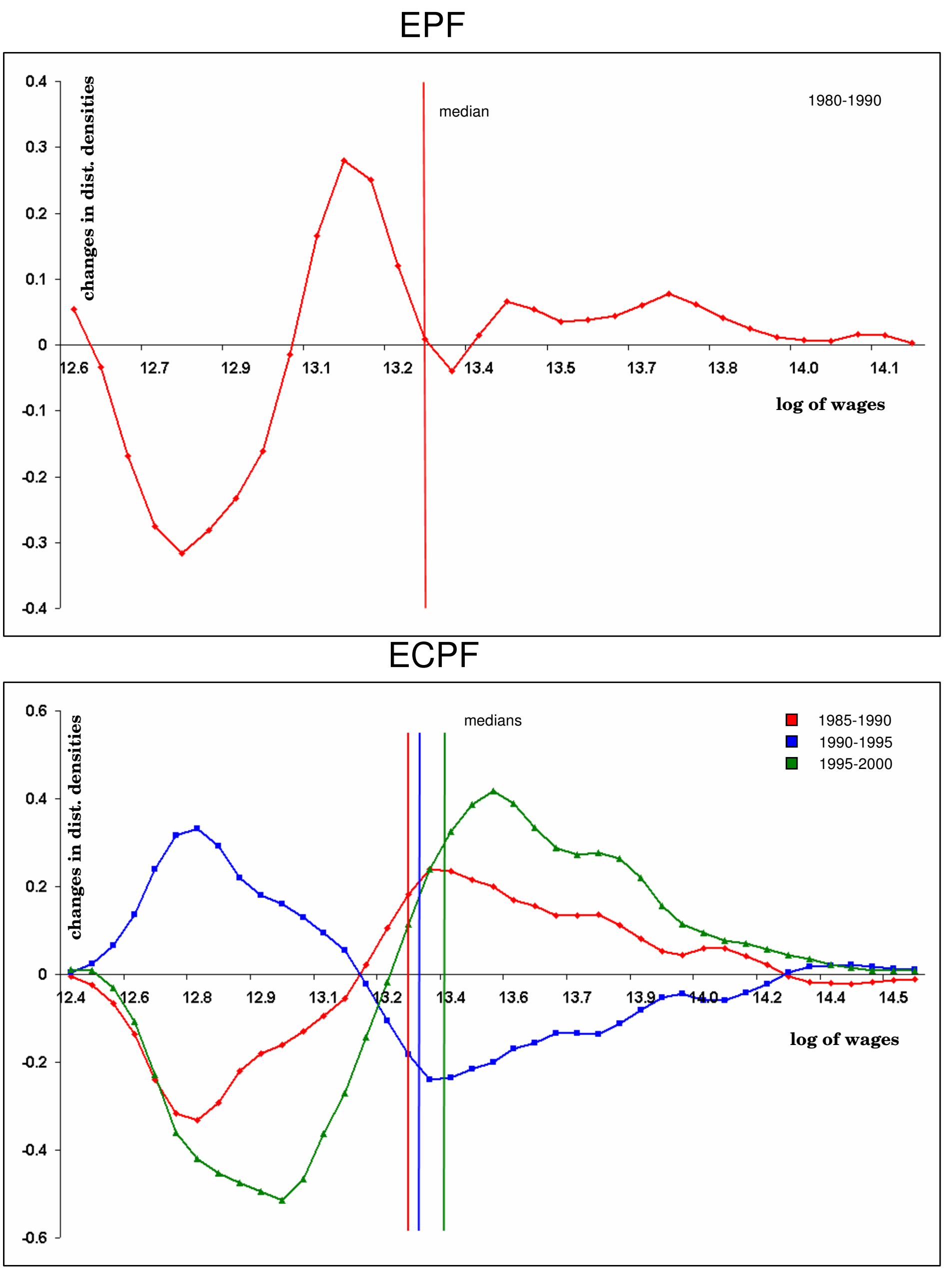
To help clarify the dynamic under discussion, Figure 1.3 presents changes in wage density distributions. Each line states the increase or decrease in the density of a given wage between two specific years. For example, in the EPF line, the density of a wage equal to 12.6 increased by about 0.05 between 1980 and 1990.[12] The vertical lines represent median levels during the first years of our comparison. Thus, between 1980 and 1990, this figure represents an accumulation of density near but below the median. Between 1985 and 1990, we observe a new low wage concentration beside an increase of density around the median. This explains the results of Table 1.5. First, the inequality drop of about -6.7% can explained by the huge concentration of density around the median instead of around the tails. However, the drop in density around the upper half can be explained by the median increase, which is not followed by the higher percentiles. Between 1990 and 1995 a dispersion increase can be observed, especially around the lower part. Finally, the pattern between 1995 y 2000 mirrors that to be found for the 1980s.

In view of this information, our first impression is that inequality adopts a countercyclical pattern. This is a well known issue that finds its parallel in inequality analysis for other countries. For example, Dimelis and Alexandra (1999) find that different US and UK inequality measures show a negative

[11] This is also AKK's main finding for the United States.

[12] The density are estimated using the Epanechnikov kernel procedure using simulated wages (Epanechnikov 1969). The width is the one that would minimize the mean integrated squared error if the data were Gaussian.

Figure 1.3: Changes in Wage Distributions



correlation with the GDP business cycle (measured as the difference between GDP and its long-term trend). But they also find a positive correlation for some countries, such as Greece, and mixed results for others, such as Italy. This initial information therefore shows that the Spanish case is closer to that of the US and the UK.

In summary, during the period under analysis wage inequality showed an uneven rather than a steady pattern, presenting differences above and below the median. Furthermore, Spanish wage inequality shows a counter-cyclical pattern shared with 50th-90th inequality, while 10th-50th inequality has increased since 1985. This is our first result.

Composition and Prices

To explore the possible causes of this dynamic, we must isolate the effects of both price and composition on wage inequality. At a first glance, it is not clear which of these two effects predominates. During the 1980s, for example, the 5.1 log point decrease in 90th-10th inequality is caused by prices, while quantities play a minor offsetting role (0.3 log points). The predominance of prices is also evident during the second half of the 1980s. Nevertheless, during the 1990s prices and composition jointly increase the 90th-10th inequality by 3.9 log points until 1995, while composition plays the main role thereafter, decreasing inequality by 3.1 log points while the price trend remains steady. Both effects are important, however. This result is coherent with a major change in the labor force composition in Spain during this period, and especially after 1985: the increased participation of both highly educated workers and women.[13] This change affects 10-50th and 50-90th inequality to the same degree. For all of these reasons, therefore, the

[13]As it can be sawn in Table 1.1, the increase in average years of schooling that stems from these surveys was around 0.65 in eighties, while during 1990-1995 it increased 0.67 and 1995-2000 by 0.76.

composition effect in wage inequality change must not be negligible.

Despite the compositional factor, the pace marked by total wage inequality repeats observable price movements until the last period, when the shift in inequality was almost zero. Thus, our data shows that prices reduced inequality during the 1980s and increased it during the 1990. Nevertheless, the reduction during the 1980s might have been concentrated in the second half of that period. Except for the 1985-1990 period, the effect of these changes mainly shows up in the lower half, which governs the total wage inequality trend. Despite the important role played by composition effect, the overall evolution of wage inequality mirrors price inequality. Thus, inequality caused by price changes shows an increased trend for the 10th-50th percentile distance and a counterciclycal one for the 50th-90th one.

These results are coherent with those of previous studies (e.g., Abadíe, 1997; Arellano, Bentolila and Bover, 2001 and Izquierdo and Lacuesta, 2007) but not with those of Izquierdo and Lacuesta (2007) with regard to the composition effect. There are two reasons for this discrepancy. First the period of study analyzed by the latter (1995-2000) differs from our own. Second, we are restricted to using data from heads of family. In the EPF and ECPF surveys, a woman is classified as the head of her family when no other male family member is employed. Thus, although our data reflects the fact that ten percent of the workers considered for this exercise were women, our analysis shows a quite small level of female participation and also a gap due to market participation. Thus, our results differ because our data –unlike that of Izquierdo and Lacuesta (2007)– does not completely reflect shifts in female participation taking place during the period under study, nor does it show the dramatic recent rise in the latter. As they argue, women's participation in the job market was the main driving force behind wage inequality during this period, and our study does not account for this trend.

Our second result, therefore, is that both price and labor composition played an important role in the evolution of Spanish wage inequality during our study period. Furthermore, while composition effect is important, overall wage inequality changes mirrored the changes in price inequality during this period. Thus, the 10th-50th distance when only prices changes increased after 1985, and the 50th-90th distance shows a countercyclycal pattern.

Between and within group inequality

Next, price inequality is decomposed between the BI and RWI change in inequality. Note that BI shows inequality due to the distance between the conditionals distributions, while RWI represents the dispersion within the conditional distributions.

As shown in the lower panel of Table 1.5 BI captures a symmetric trend on both sides. Here again, the trend suggests that prices reduced inequality during the late 1980s and 1990s but increased it during the early years of both decades. Nevertheless, analysis of RWI reveals a completely different picture, with the most salient feature being the increasing trend. From a fall during the early 1980s to a rise during the late 1990s, the trend is clearly one of steady growth. In RWI there is no symmetry, with most of the change clustering in the lower half with the exception of the 1985-1990 period. Our third result, therefore, is the difference between the inequality evolution at BI and RWI. Clearly, these two different forces jointly draw the picture described previously regarding a price-driven change in inequality. How can we explain the BI and RWI trends? An initial guess is that each of these trends must respond to different causal factors.

Education and Experience

Due to the existence of previous literature that investigates the effects of a return to education and experience in Spanish wage inequality, one further step is to distinguish the BI for both prices. Table 1.6 presents the counterfactual densities obtained when only one price changes. For example, education became less dispersed throughout our study period, with the exception of the years between 1990 and 1995. Again, this is a countercyclical trend, but one located between the 50th-90th percentiles, so that almost all of the decrease can be found in the upper half. The reason for this asymmetry might be the massive incorporation of more highly skilled workers into the labour market during this period. Bearing in mind the results discussed in section 1.4, between 1980 and 2000 the fall in the high school wage premium relative to other premiums, followed by a fall in the college wage premium, reduced inequality when only the price of education is accounted. Nevertheless, the lower tail shows a different movement where growth is prevalent. Again, weaker institutions might be the reason for this trend. The period 1980-1985 is of special interest. In this case, comparison of the results for the decade as a whole and for the 1985-1990 period would reveal an increase in BI in education. This result is similar to that found by Arellano, Bentolila, and Bover (2001) (for the first half of the 1980s) and to that found by Abadie (1997) (for the 1980s) and Izquierdo and Lacuesta (2007) (for the second half of 1990s). Experience presents a countercyclical and completely asymmetric trend on both sides of median. The greater weight of less experienced workers within the lower tail of the wage distribution and the greater weight of more experienced workers within the upper tail would explain the different effects. Again, these results mirror those obtained in other Spanish studies for the same periods.

The range of variation not only between BI, RWI and composition effects, but also within each simulated counterfactual density, suggests that changes in wage inequality were caused by more than one factor. The hidden factors

Table 1.6: 100 × log changes in Education and Experience Prices and Quantity Inequality

	EPF 80-90	ECPF 85-00		
Prices		85-90	90-95	95-00
Education				
10-50	0.0	1.7	-1.7	1.7
50-90	-2.5	-5.4	2.9	-4.2
10-90	-2.5	-3.7	1.2	-2.5
Experience				
10-50	1.3	-2.7	2.2	-3.8
50-90	-1.5	1.9	-1.3	1.6
10-90	-0.2	-0.7	0.9	-2.1

underlying changes in earnings dispersions differ for different wage distribution areas, and in the intensity of their effects. Nevertheless, it is necessary to tackle this result with some possible explanations that could be developed in future research.

Some tentative explanations

Note that our results for BI once again support the idea of countercyclical inequality. In years of low GDP growth, the distance between the returns paid to skilled workers increases, whereas in years of high GDP growth the distance between conditional dispersions contracts. A number of theoretical explanations for this result can be given. For example, despite that the fact the skill bias technical change can be seen to result from unobservable change in the production function (Katz and Murphy, 1992), more recently Krusell, Ohanian, Rios-Rull, and Violante (2000) associate technological change with capital equipment increase. The important assumption is that of capital-skill complementarity, which means that the elasticity of substitution between capital and unskilled labor is higher than the elasticity of substitution between capital and skilled labor. Thus, in the middle of both decades, both output and the use of inputs declined. Assuming that capital is a quasi-fixed production factor and that capital-skill complementarity prevails, the demand for unskilled labor reacts more than the demand for skilled labor. When aggregate demand increases, the opposite effect occurs. Thus, it follows that the relative demand for skills is countercyclical due to a relative supply curve of skilled labor with a positive slope, and that the relative wage and employment of skilled labor are also countercyclical.

Supporting this explanation, Hidalgo, O'Kean and Rodríguez (2008) have found strong capital-skill complementarity evidence for Spain between 1980 and 2004, which could explain the Spanish countercyclical property that we have been discussing. While this countercyclicality has also been found for other countries,[14] for the United States there is only evidence of countercyclical difference in wages earned by skilled and unskilled workers until 1984. The trend turns procyclical in 1984, a shift that has been explained by the concomittant reduction in the degree of capital-skill complementarity. (Castro and Coen-Pirani 2005).[15] Nevertheless, since other explanations have been suggested,[16] this issue awaits a more in-depth analysis.

The evolution of RWI can be explained by institutional factors, for a number of reasons. First, this feature has tended to be considered as an institutional effect (i.e. DiNardo, Fortin and Lemieux, 1996, Card and Lemi-

[14]For example Denmark (Skaksen and Sorensen 2005).

[15]The cyclicality of the skill premium has been largely ignored in the literature since Reder (1955), who was the first who formally investigated the movements of wage differentials between skilled and unskilled workers over the business cycle. Since then, Keane and Prasad (1993), Young (2003) or Lindquist (2004) has analyzed this issue.

[16]For example, firm-specific human capital of skilled labor (see e.g., Becker, 1964), and hiring costs that are higher for skilled labor than unskilled labor (see e.g., Bentolila and Bertola, 1990) or implicit contracts literature (see e.g., Pourporides, 2007).

uex, 2001 and Autor, Katz and Kearny, 2005a and 2005b). Second, it fits quite well with Spanish evidence regarding labor markets reforms. For example, Peraita (2003) argues that the Spanish labor market was one of the most rigid in the industrialized world during the early 1980s in particular, but market labor reforms in 1984, 1992, 1993 and especially in 1994 induced a flexibilization process. Third, labour institution reforms since 1984 and the fact that Spain's growth structure since 1996 has been skewed towards low-skilled labour sectors would explain the wide dispersion found, especially, in the lower half of the distribution. For example, the inequality decrease described in the lower half of RWI during the early 1980s could be explained by the increase in the bottom wage percentile due to the appearance of a new institutional wage-setting framework. Peraita (2003) explains, by contrast, that Spanish labour market reforms should increase wage inequality among lower-paid workers, especially from 1994 onwards. To summarize, it seems that the possible increase in relative supply due to unemployment and in skill bias technical change due to capital-skill complementarity (BI) and, more particularly, to weaker labour institutions, (RWI), may jointly explain the increase in inequality during the early 1990s and thus explain the overall increase in wage inequality during this period. These factors would also explain the lack of movement in the upper tail: the lesser impact of unemployment effects in these cohorts and the limited labour reforms which had no effect on indefinite labour contract privileges might explain this pattern, in which the labour market appears to dominate the wage inequality trend.

There is room for alternative explanations, however. For example, there may have been an increase in unobservable heterogeneity, which would have increased RWI. A natural explanation might be linked to the generalization of education and the increase in number of scholarships granted since the 1980s, which pushed a greater number of students with non-observable skills into higher education. This scenario might have increased the non-observable skills within each group of workers, especially the more educated ones.

This last hypothesis can be easily tested using counterfactual densities to simulate the densities of groups of workers with, say, 16 years and zero years of schooling, and exactly the same experience, gender and other characteristics. Under this hypothesis it should be expected that the conditional RWI of the more educated workers would mirror the overall RWI trend, especially in the lower tail. People with less skills and more education would be located in the lower tail of the conditional "more educated" distribution. Thus its dispersion must increase through time. Then, once the "more and less educated" conditional counterfactual distributions are simulated, it is possible to calculate the percentile distances. The results for the 1985-2000 period are shown in table 1.7.

Table 1.7: 100 × log changes in percentiles distances for 16 and 0 years of schooling counterfactual densities

	85-90	90-95	95-00
16 years sch.			
10-50	5.8	-5.3	12.9
50-90	-18.8	6.8	-9.6
10-90	-13.0	1.5	3.3
0 years sch.			
10-50	-0.1	6.9	13.8
50-90	-6.9	-4.6	-7.6
10-90	-7.0	2.4	6.2

In these cases, the less educated workers' RWI appears to be closer to overall RWI than to that of more educated workers. In this last case, the 10th-

50th distance grows between 1985 and1990 and falls between 1990 and 1995, while the 10th-50th distance in the less educated conditional distribution follows the same trend as the total RWI lower tail dispersion. Thus, the increase in heterogeneity among more educated workers due to easier access to education is not enough to explain the lower tail trend of its conditional residual wage distribution and the resultant increase in total RWI. Rather, institutions seem to represent a key driving force behind the change in the residual inequality for lower wages.

1.7 Conclusions

This paper attempts to decompose the change in Spanish wage inequality into three components, using a counterfactual analysis based on the quantile regressions developed by Machado and Mata (2005) and extended by Autor, Katz and Kearney (2005a and 2005b): changes in between- and within-group inequality and changes in labor composition.

The results for the 1980-2000 period are fourfold. First, because the Spanish wage behaves countercyclically, the inequality change is almost zero but positive. Second, both price and labor composition play important roles in the evolution of Spanish wage inequality; however, while composition effect is important, the changes in overall wage inequality mirror the changes in price inequality. Third, the between-group inequality, which measures the distances in conditional to observable characteristics in wage distributions, follows a countercyclical trend, while the residual wage inequality, which measures the dispersion within the conditional distributions and that caused by non-observable variables, increased since 1985 onwards. Also, an important result is that while the between-group inequality shows a symmetric pattern above and below the median, the residual inequality tells a different story for each half. Forth and finally, the inequality found for education and experience mirrors previous results for Spain and, again, behaves countercyclically.

There are several possible explanations of these results. While for between-group inequality changes in the supply and demand of different worker cohorts may explain price changes, and therefore changes in wage distribution, institutions may be behind the increased dispersion in the lower tail of the residual wage inequality. For example, countercyclical changes in relative demand for skilled (more educated) workers could be explained by skill-biased technological change, and especially by the degree of complementarity between this factor and capital. However, there are other possible explanations for this behavior, such as changes in Spanish labour institutions in favor of collective bargaining improved between 1980 and 1985, and market reforms since the early 1990s. Nevertheless, these intuitions must be explored in greater depth.

Chapter 2

A Demand-Supply Analysis of the Spanish Education Wage Premium in the 1980s and 1990s

2.1 Introduction

Recent decades have seen substantial heterogeneity in the evolution of the education wage premium, both across countries and over time (Katz, Loveman, and Blanchflower 1995, Gottschalk and Smeeding 1997, Gottschalk and Joyce 1998, Acemoglu 2003). A natural starting point for the analysis of these differences is the demand-supply framework (D&S). The purpose of the D&S framework is to examine whether the evolution of the education wage premium can be approximated by supply-driven movements along a labor demand curve with a stable slope, plus shifts in labor demand. The results have been quite encouraging in a variety of contexts. Katz and Murphy (1992), for example, conclude that the education wage premium in the U.S. between

1963 and 1987 can be explained by steady, secular shifts in the demand for educated workers combined with observed changes in relative supply. Katz, Loveman, and Blanchflower (1995) show that the D&S framework is also useful for understanding the evolution of the wage premium in four OECD countries (the U.S., U.K., Japan, and France). Card, Kramarz, and Lemieux (1999) incorporate wage-setting institutions in a D&S framework and show that this helps to explain relative wage trends among less-skilled workers in the U.S., Canada, and France in the 1980s. Acemoglu (2003) finds that the D&S with steady, secular shifts in the demand for educated workers can account for the differences in the evolution of wage inequality between Finland and Norway.

While the Spanish education wage premium has been studied quite intensively,[1] the literature has not yet explored whether its evolution in time may fit within the D&S framework. The goal of this study is to ascertain whether the D&S framework can help explain the evolution of the education wage premium in Spain during the two decades between 1980 and 2000. Our main finding is that the evolution of the premium during those two decades can be well approximated by combining the observed changes in labor supply with steady growth in the demand for education over the 1980-200 period. Interestingly, our estimates of the slope of the Spanish demand curve for education, and education demand growth, are quite similar to U.S. estimates.

One of the key elements of the D&S framework is the slope of the demand curve for education (which, in the standard D&S framework, is the inverse of the elasticity of substitution between more and less educated workers). The main difficulty faced when estimating this slope is that education supply and the education wage premium are determined simultaneously by demand and supply. Estimation therefore requires solving the standard identification

[1]See Abadie (1997), Arellano, Bentolila, and Bover (2001), Torres (2002) and Martinez-Ros (2001) for example.

problem (see Hamermesh, 1993, for a summary of this problem in the context of labor demand estimation). The empirical literature on the demand curve for education stretches back to the 1970s. Johnson (1970) estimates the elasticity of substitution between more and less educated workers to be 1.34 for a cross section of U.S. states in 1960. Ciccone and Peri (2005), using a panel of US states for the 1950-1990 period, and employing Acemoglu and Angrist's (2001) state-time-dependent child labor and compulsory school attendance laws as instruments for changes in the supply of education, find an elasticity of substitution of about 1.5. Angrist (1995) finds an elasticity of substitution of about 2 for data on Palestinian workers in the West Bank and the Gaza Strip during the 1980s; he uses the number of local higher-education institutions as an instrument for education supply. Fallon and Layard (1975), using cross-country data and employing income per capita as their instrument for education supply, obtain an estimate of 1.49 for the elasticity of substitution between more and less educated workers. Caselli and Coleman (2000) apply a D&S framework with endogenous technology to cross-country data and obtain an elasticity of substitution of 1.31. Katz and Murphy (1992) derive from U.S. time-series data for the 1963-1987 period an elasticity of substitution of about 1.4.

Because the national time-series data for Spain is insufficient to allow a valid estimate of the elasticity of substitution, we use the approach developed by Katz and Murphy to estimate the elasticity of substitution in a panel of Spanish regions, employing the beginning-of-period population structure as our instrument for education supply. The resulting estimate of the elasticity of substitution between more and less educated workers in Spain is close to the estimates reported by Katz and Murphy and Ciccone and Peri for the United States.

Our estimate of the slope of the Spanish demand curve for education for the 1980-2000 period allows us to examine the degree to which the D&S

framework can be used to explain the evolution of the Spanish education wage premium during that period. Our chief empirical finding is that the evolution of the education wage premium as predicted by the framework fits quite closely with its actual evolution. For example, our estimates imply a fall in the relative wage of more educated workers during the 1980s of 0.6% and an increase of 1.1% in the 1990s, which comes close to the actual 0.7% drop of relative wages in the 1980s and the 1.4% increase in the 1990s. Interestingly, we find similar annual growth rates for education demand in Spain (labor demand shifts) during the 1980s and the 1990s (2.7% and 3.1%, respectively). These estimates come close to estimates for the United States: for example, Katz and Murphy (1992) estimate the relative U.S. demand shifts to be about 3.3% per year, while Acemoglu (2002) reports an increase of about 2.5% annually.

One explanation for cross-country differences in the evolution of wage inequality, especially in Europe, is that wage-setting institutions differ by country (Acemoglu, 2003, Card, Kramarz, and Lemieux, 1999, Abraham and Houseman, 1993...). Arguably, the most relevant institution for the Spanish case is collective wage bargaining; however, taking this into account does not affect our conclusion that the D&S framework is able to capture the evolution of the Spanish education wage premium.

The rest of the paper is structured as follow. Section 2.2 explains the data used; Section 2.3 explains the measurement of relative wages and education supply; Section 2.4 presents the estimation and decomposition results; Section 2.5 evaluates the possible effects of collective bargaining on relative demand estimates; and Section 2.6 concludes.

2.2 The Demand and Supply Framework

According to the demand and supply framework, the wage of more relative to less educated workers (the education wage premium) is determined by education demand and supply. The simplest model of relative demand is based on the constant elasticity of substitution (CES) firm-level production function (see, for example, Katz and Murphy, 1992). The model assumes that firms f have access to the following production function:

$$Y = [A_f L^\rho + B_f H^\rho]^{\frac{1}{\rho}} \tag{2.1}$$

where Y is output, H is the input of more educated (skilled) workers, and L the input of less educated (unskilled) workers. A_f and B_f denote the levels of factor-augmenting technology to which firms have access. It is straightforward to show that the production function parameter ρ determines the elasticity of substitution between factors σ. In particular, $\sigma = 1/(1-\rho)$, which implies that $\rho \leq 1$ is necessary for the isoquants to be convex and the education demand curve to be well-defined ($\rho = 1$ corresponds to the case where the two types of labor are perfect substitutes, while $\rho \to -\infty$ implies that there is no substitutability at all between more and less educated workers).

Firms are assumed to take wages in the labor market as given when making their hiring decisions. Firms' demand for education, the demand for more relative to less educated workers H/L, can be obtained from their first-order conditions for profit-maximization as

$$\left(\frac{H}{L}\right)_D = \left(\frac{B_f}{A_f}\right)^\sigma \left(\frac{w^H}{w^L}\right)^{-\sigma},$$

where we have used that $\sigma = 1/(1-\rho)$.

The D&S framework can be applied to the regional level by assuming that firms in region i have levels of factor-augmenting technology $A_f = A_i$

and $B_f = B_i$. A region's equilibrium education wage premium can now be determined by equating education demand with education supply $(H/L)_{Si}$in region i and solving for the relative wage for educated workers,

$$\left(\frac{w^H}{w^L}\right)_i = \left(\frac{B}{A}\right)_i \left(\frac{H}{L}\right)_{Si}^{-\frac{1}{\sigma}}.$$

Taking logs on both sides yields

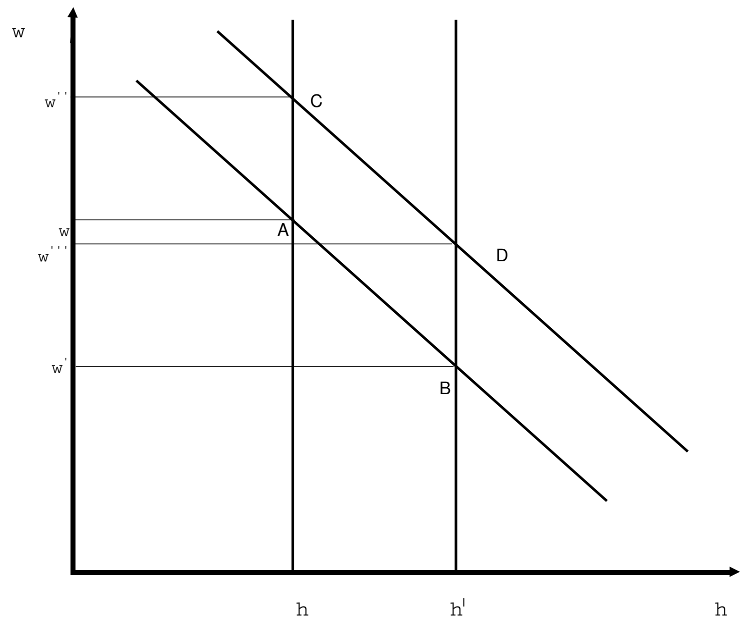
$$\omega_i = b_i - \frac{1}{\sigma} h_{Si}; \qquad (2.2)$$

where $\omega_i = Ln(w^H/w^L)$, $b = Ln(B/A)$ and $h = Ln(H/L)$. Taking differences over time (denoted by Δ) yields

$$\Delta\omega_{it} = \Delta b_{it} - \frac{1}{\sigma}\Delta h_{iSt}. \qquad (2.3)$$

Hence, log changes in the education wage premium, $\Delta\omega_{it}$, are equal to shifts in education demand, Δb_{it}, plus supply-driven movements along the education demand curve, $-\frac{1}{\sigma}\Delta h_{iSt}$. The strength of the effect of supply changes on the wage premium depends the slope of the inverse education demand curve, $1/\sigma$, which is equal to the inverse of the elasticity of substitution between more and less educated workers. When the elasticity of substitution is high, supply changes will have small effects on the education wage premium (the inverse demand curve is flat). As the elasticity of substitution between more and less educated workers falls, the sensitivity of the education wage premium to changes in education supply increases. Figure 1 provides a graphic illustration of the relative wage effects of demand shifts and supply-driven movements along the demand curve. An increase in the relative supply, from h to $h^{'}$, moves the equilibrium point along the downward-sloping inverse demand curve (A to B) and reduces the education wage premium. An increase in the relative demand for educated workers moves the equilibrium point to C and increases the education wage premium. When there are demand and supply shifts, the equilibrium rests at point D; in this case the behavior of the education wage premium depends on which shift prevails.

Figure 2.1: The Relative Demand for Education



The key feature of (2.3) from our point of view is that, once the elasticity of substitution between more and less educated workers has been estimated, it can be used to determine how supply and demand affect the evolution of the education wage premium. In order to resolve the standard simultaneous-equation identification problem, estimating the elasticity of substitution requires a valid instrument for shifts in the regional education supply.

2.3 Data and Measurement

2.3.1 Data

Individual Data

The wage data for this study comes from the Household Budget Survey (EPF) and Continuous Household Budget Survey (ECPF),[2] both of which cover a wide range of individual characteristics, such as education, age, region, annual earnings, type of employment contract, etc. The EPF is available for 1974, 1980-81 and 1990-1991; the ECPF is available for every quarter since 1985. Although there are some differences, the information in the EPF since 1980 is quite similar to that in the ECPF. As the focus here is on long-term trends, we will use data corresponding to 1980-1981, 1990-1991 and 2000-2001. The 1974 survey is not considered because it provides no usable education data. Other sources of wage data either lack needed individual data only or cover only short periods. An additional advantage of the EPF is that the methodology used to compile this data has remained relatively stable since 1980.

We will focus on heads of households aged 20 to 65 who work full time and are not self-employed.[3] Our schooling data refers to the highest of four possible degrees attained upon the subject's completion of primary school, lower secondary school, upper secondary school, or university. We use this information to impute years of schooling to each of the individuals in our sample.[4]

[2]Dating from 1985, but with a change in methodology in 1997 and processed by the Spanish National Institute of Statistics (INE)

[3]In 1990-91 and 2000-01 some characteristics are included for the head of family only.

[4]We follow former Spanish studies to impute these years, especially Vila and Mora (1998).

Aggregate Supply Data

Our data on the supply of schooling comes from Instituto Valenciano de Investigaciones Económicas (IVIE) (Mas, Pérez, Uriel, Serrano, and Soler, 2002). The schooling data refers to workers in 17 (of 19) Spanish regions from 1964 to 2001. The categories are:

1. No schooling or primary schooling only (less than eight years of schooling)
2. Lower level secondary schooling (between eight and twelve years of schooling)
3. Upper secondary education (thirteen years of schooling)
4. College degree (sixteen to eighteen years of schooling)

Less educated workers are defined as belonging to either the first or the second group. That is, workers with fewer than 12 years of schooling are considered to be less educated, while those with an upper secondary or university educations are defined as more educated. These definitions are quite similar to others in analysis for the U.S., U.K. and other countries.[5]

Instruments

Our approach is uses the beginning-of-period population structure as an instrument for analyzing regional changes in schooling supply. The needed population data comes from the 1981 and 1991 Spanish Population Censuses provided by the National Statistics Institute.

[5]Acemoglu (2002) defines this classification for the US. However, he considers this a simplification in a context in which there is a continuum of imperfectly substitutable skills.

2.3.2 Measurement and Descriptive Stats

Education Wage Premium

Wages depend not only on schooling but on many other individual characteristics. To isolate the role of schooling, two approaches might be used. One could use all of the available individual characteristics to build a narrow definition of worker cohorts, then calculate the education wage premium as the wage of one narrowly-defined cohort relative to another, less-educated cohort that is very similar to the first in all other dimensions. However, this strategy requires many observations. We therefore focus on a second strategy based on Mincer wage regressions (Mincer 1974). Specifically, using j for individuals, t for years, and i for regions, we estimate

$$ln(w_{it}^j) = \alpha_{it} + \beta_{it} S_t^j + \gamma_{it}^1 E_t^j + \gamma_{it}^2 (E_t^j)^2 + \mu_{it} X_t^j + \varepsilon_t^j. \tag{2.4}$$

The left-hand side is the log of individual wages and the right-hand side contains a list of explanatory variables: years of schooling (S_t^j), years of experience (E_t^j), and other k variables (represented by the $k \times 1$ vector X_t^j) such as marital status, employment sector, gender, etc. As usual, experience is calculated as age minus years of schooling minus six. The key parameter is β_{it}, that is, the percentage increase in wages (the return) from one year of schooling in any given region/year. Once we have estimated this return,we obtain the log education premium of workers with S^H years of schooling relative to workers with S^L years of schooling in region i for year t by multiplying the difference in years of schooling by the estimated return to schooling ($\hat{\beta}_{it}$). The method used to estimate (2.4) is ordinary least squares.

A standard concern with Mincerian wage regressions estimated using ordinary least squares is that schooling can be correlated with unobservable characteristics (e.g., ability) that may also affect wages. While some Spanish studies have sought to address these concerns using instrumental variables, none of them use EPF or ECPF data, since these surveys do not provide suitable instruments. Nevertheless, there two reasons to believe this concern should not affect our analysis. First, many studies have shown that the bias is quite small (Card 1999). Moreover, since our study focuses on the the evolution of the education wage premium, our analysis will not be affected by the bias as long at the latter remains approximately constant in time. Another issue is that our estimating equation implies that the return to an additional year of schooling is independent of the level of schooling. In principle, we could relax this assumption by estimating the return to schooling only for those who have attained certain levels of education (degrees). But we do not have sufficient data to follow this approach for some of the smaller Spanish regions.

Table 2.1 contains our return-to-schooling estimates (or β) for Spain as a whole, obtained using the data on individuals available in our surveys.[6] Here, it can be seen that the number of individuals returning to school fell during the 1980s and increased slightly during the 1990s, a pattern that is in keeping with the results found for other countries (for example, see (see for some examples Gottschalk and Smeeding 1997, Freeman and Katz 1995, Acemoglu 2003). There are similar, previous findings for Spain; for example, Abadie (1997) finds that Spanish wage inequality fell during 1980s, partly due to a decrease in the return to education.[7]

[6] Some of the values of the estimated regional coefficients are not shown.

[7] Other works point in a different direction however, maybe because of the use of different surveys to compare trends in the return to education. Barceinas, Oliver, Raymond, and Roig (2000) estimate the return to education using EPF for 1980 and ECPF for 1985 - 1996. They found that return to education increased during this period, except in between 1985 and 1991 when it fell. Their estimate of the return to education is 5.9% for 1980 and 7.0% for 1990. The estimates obtained in this paper are 6.4% for 1980, 6.0% for 1990, and 7.0% in 2000. The major differences in the 1990 figures are due to the use of ECPF data for that year.

Table 2.1: Spanish Returns to Education. 1980/81, 1990/91 and 2000/01

	1980-81	1990-91	2000-01
β_t	0.064	0.060	0.070
	(0.001)	(0.001)	(0.003)

Note: Estimations for 1980/81 and 1990/91 are based on EPF and 2000/01 on ECPF. β_t represents the average return to education for Spain. Data in parenthesis represent standard deviations. The re- turns are estimated using Mincer equations and OLS. The sample is limited to heads of family and non self-employed workers.

Relative Supply

We first aggregate workers with only primary schooling and workers with lower-level secondary schooling using

$$L_{it} = L_{it}^1 + a_{it}^L L_{it}^2,$$

where a_{it}^L is the efficiency of workers with lower-level secondary schooling relative to workers with primary schooling. This efficiency parameter is obtained as the education premium in region i and year t of workers with no more than a secondary-school education relative workers with no more than a primary-school education. Here, the supply of more educated workers is obtained by aggregating workers with an upper-secondary education and college educated workers using

$$H_{it} = H_{it}^1 + a_{it}^H H_{it}^2,$$

where a_{it}^H is obtained obtained as the education premium in region i and year t of workers with no more than an upper-secondary education relative workers with a university degree. The log supply of education can now be

Table 2.2: Relative supply and logs relative wage in Spain.

	1980-81	1990-91	2000-01
w_t	0.51	0.43	0.49
h_t	0.057	0.096	0.121

Note: w_t denotes the Spanish average education wage premium for year t while h_t represents relative supply for better-educated workers.

obtained as

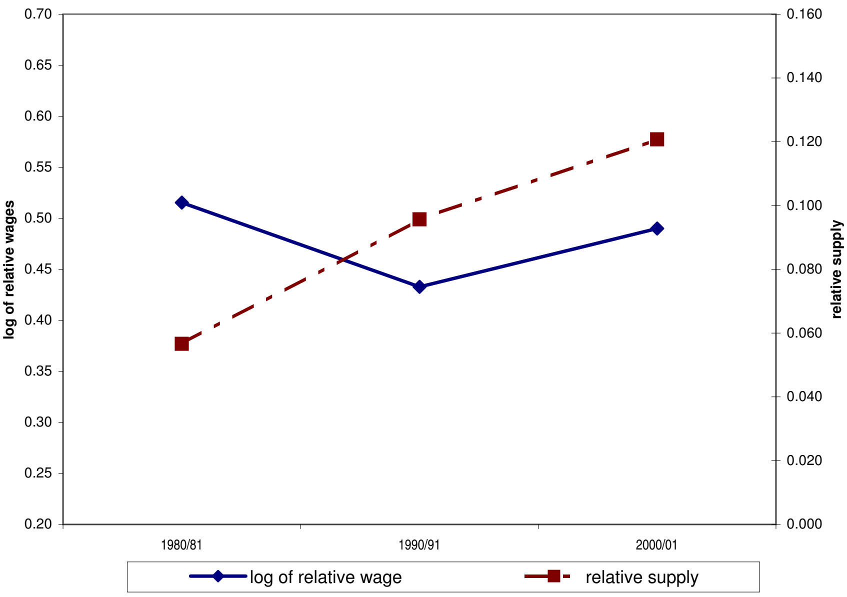
$$h_{i,t} = ln\frac{H_{it}}{L_{it}}.$$

Descriptive Statistics

Figure 2.2 and Table 2.2 contain information on the education wage premium and the relative supply of schooling for the 1980s and 1990s. It can be seen that the education wage premium fell between 1980 and 1990 (from 0.51 to 0.43) and increased between 1990 and 2000 (from 0.43 to 0.49). The implied annual growth rates are equal to -0.8% during the 1980s and 0.6% during the 1990s. The (log) relative supply of schooling, on the other hand, increased from 0.057 to 0.096 during the 1980s and from 0.096 to 0.121 during the 1990s. The implied annual growth rates were 5.2% during the 1980s and 2.3% during the 1990s.

Whether the pattern in Figure 2.2 and Table 2.2 is sensitive to the way education groups are aggregated is an important issue. As a robustness check, therefore, we classify workers with lower-secondary schooling in the higher education group and then repeat the analysis using this new classification.

Figure 2.2: Education Wage Premium and Relative Supply of Skills in Spain. 1980-2000. (More educated workers have previous to college or college education)



The results are shown in Figure 2.3 and Table 2.3.

Qualitatively, the evolution of the education wage premium and relative supply of schooling for this new classification is very similar to the one we obtained earlier.

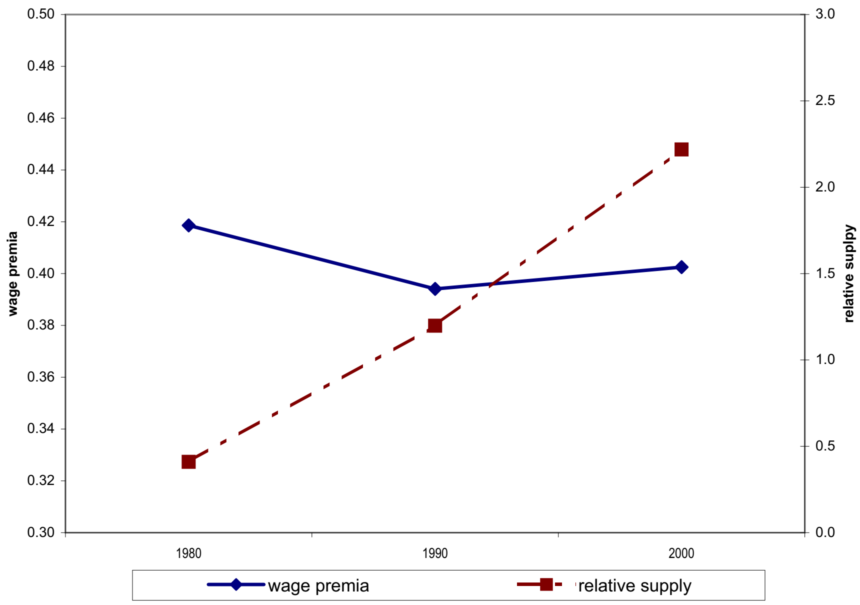
2.4 Estimation and Results

To gauge the extent to which the evolution of the education wage premium can be explained by the demand-supply framework, we estimate

$$\Delta\omega_{it} = \Delta b_t - \frac{1}{\sigma}\Delta h_{iSt} + (\Delta b_{it} - \Delta b_t). \tag{2.5}$$

where $\Delta\omega_{it}$ is the change in the education wage premium, Δb_t the national shift in education demand, $-(1/\sigma)\Delta h_{iSt}$ captures supply-driven movements

Figure 2.3: Education Wage Premium and Relative Supply of Skills in Spain. 1980-2000. (More educated workers have secondary or higher education)



along regional education demand curves, and $\Delta b_{it} - \Delta b_t$ are regional shocks to labor demand.

Changes in the regional supply of educated workers are likely to be positively correlated to shifts in regional labor demand, which implies that the inverse elasticity of substitution between more- and less-educated workers cannot be estimated by applying ordinary least squares estimation to (2.5). This positive correlation may be the result of worker migration to regions with rapidly rising wages, or it may reflect the fact that individuals living in regions where education is highly paid may decide to remain in school longer. Thus, it is necessary to find instruments for changes in the supply of education. Since the beginning-of-period population structure should be unaffected by shocks to regional labor demand, we will use the regional population structure in 1980 as an instrument for changes in the education supply

Table 2.3: Relative supply and logs relative wage in Spain (II)

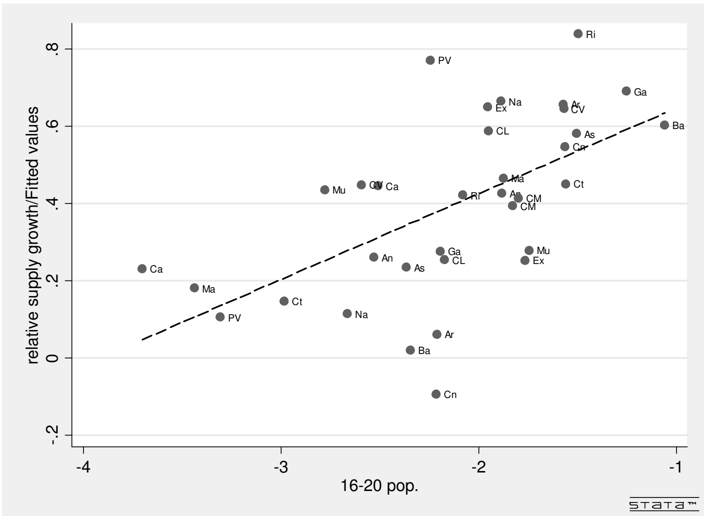
	1980-81	1990-91	2000-01
w_t	0.42	0.39	0.40
h_t	0.41	1.20	2.22

Note: w_t denotes the Spanish average education wage premium for year t while h_t represents relative supply for better-educated workers.

during the 1980s and the population structure in 1990 as an instrument for changes in the education supply during the 1990s. Since changes in legal schooling requirements and better educational opportunities have increased the educational levels of younger people throughout Spain between 1980 and 2000, relative to that of the generation that retired during those same two decades, changes in the regional supply of educated workers should also correlate to changes in the supply of education. Hence, the average level of schooling should have increased more rapidly in regions where there were more young people in 1980, relative to other regions. Figure 2.4 shows the relationship between the population share of people between the ages of 16 and 20 at the start of the period and the variation in relative supply over the course of the following decade. Predictably, it suggests a positive link between these two variables. The beginning-of-period population share of 16-to-20-year-old is the instrument used to estimate (2.5). Hence, our identifying assumption is that this population share affects the change in the education premium over the course of the following decade only through its effects on the relative supply of education.

Table 2.4, which presents the first stage regression, shows that the 16-to-20-year-old population share has a highly statistically significant positive

Figure 2.4: Relation between instruments and relative supply growth



Source: Spanish Demographic Censuses (INE) and this chapter's estimate with Human Capital Series (IVIE).

effect on the growth of education supply. Interestingly, these results become even stronger once we introduce region-fixed effects to control for region-specific trends in the relative supply of education.

Table 2.5 contains the second stage results. The first column of results is the baseline specification. The remaining columns give the results of various robustness checks using other variables such as physical capital stock per worker for all sectors, physical capital per worker in intensive information and communication technology (ICT) sectors, and employment level.[8] Our data on physical capital per worker comes from BD-Mores, published by the Spanish Ministry of the Economy (specifically, the Department of the Economy). The ideas of these robustness checks using capital is that (some

[8]ICT (Information and Communication Technology) sectors are those with a high technological content (Mas, Pérez, and Uriel 2006).

Table 2.4: First Stage Regression

Dependent Variable	Changes in log of relative supply.			
	I		II	
16-20 years old	0.229	***	0.428	***
	(0.069)		(0.106)	
Constant	0.877	***	1.242	***
	(0.131)		(0.194)	
fixed effects	no		yes	
R^2	0.307^1		0.601	
F-statistic	8.32		11.31	

Note: dependent variable are log changes in regional relative supply due to extra education. The regressor (the 16-20 age group variable) represents the regional share in total population of people aged between 16 and 20. Column I shows results exclusive of fixed effects, while column II shows results inclusive of fixed effects. Data in parenthesis are standard errors.
*** means significance at 1%.
(1) In I regression is adjusted R^2

types of) capital may be complementary to educated workers and therefore affect the education wage premium. Total employment is included to test for aggregate scale effects.

The results indicate that -1/σ is between -0.58 and -0.65, meaning that the education demand curve is downward sloping. Moreover, our estimates are statistically different from zero at the 1% level. The implied elasticity of substitution between more- and less.educated workers is between 1.5 and 2. This value is very similar to that found elsewhere. Johnson (1970), for example, estimates the elasticity of substitution between more- and less-educated workers to be 1.34 for a cross-section of U.S. states in 1960. Fallon and Layard (1975) find an elasticity of substitution between less- and more-educated workers of 1.49, using cross-country data. Angrist (1995) reports an elasticity of substitution of about 2, and Caselli and Coleman (2000) estimate the elasticity of substitution between more- and less-educated workers to be approximately 1.3. Katz and Murphy (1992), using U.S. time-series data for the 1963-1987 period, report an inelasticity of about 1.4 for substitution between more- and less-educated workers. Using different estimation methods, Ciccone and Peri (2005) argue that the long-term elasticity of substitution in the U.S. between 1950 and 1990s was between 1 and 2.

By how much did the demand for education increase in the 1980s and 1990s according to our estimates? The first column shows a demand shift of 0.2715 during the 1980s, which represents an annual increase of about 2.8%. Since the difference between the pace of the education demand shifts during the 1990s relative to the 1980s is not statistically different from zero at any conventional level, we cannot reject the hypothesis that labor demand increased by the same amount during the 1990s as during the 1980s. Interestingly, our estimated increase in education demand for Spain is very similar to that estimated by Katz and Murphy (1992), who report a value of 3.3%.

Summarizing, we find that education demand grew during the 1980s at

Table 2.5: Relative Demand Estimation

Dependent variable: log changes in regional education wage premium.

	I	II	III	IV
Δb_{80}	0,275***	0,245**	0,281***	0,024
	(0.084)	(0.114)	(0.082)	(0.291)
$\Delta b_{90} - \Delta b_{80}$	0.066	0.064	0.077	0.063
	(0.140)	(0.139)	(0.141)	(0.132)
$-\frac{1}{\sigma}$	-0,654***	-0,596**	-0,654***	-0,582***
	(0.199)	(0.251)	(0.207)	(0.176)
Δk	-	0,116	-	-
	-	(0.291)	-	-
Δk_{ict}	-	-	-0,050	-
	-	-	(0.294)	-
employ.	-	-	-	0,035
	-	-	-	(0.044)
Adj. R^2	0,48	0,49	0,48	0,49
n	34	34	34	34

Note: The dependent variable is changes in logs of regional relative wages due to extra education. $\Delta b_{80}t$ estimates changes in the(inverse) relative demand intercept in eighties. $\Delta b_{90} - \Delta b_{80}$ estimates differences in changes in the(inverse) relative demand intercept in nineties relative to eighties. $-\frac{1}{\sigma}$ estimates the coefficient associates to changes in logs of regional relative supply due to extra education or the (inverse) relative demand slope. Δk are changes in logs of regional physical capital per workers and Δk_{ict} are changes in logs of regional physical capital per workers in ICT sectors. Employment responds to log of regional employment at the start of the period. All estimations are performed controlling for regional and time effects.
*** implies significance at 1%, ** at 5% and * at 10%.
Values in parenthesis are standard deviations.

Table 2.6: Decomposition of Relative Wage Changes

	wages	supply	demand	error
1980/81 – 1990/91	-0,7	-3,4	2,8	0,0
1990/91 – 2000/01	1,4	-1,7	2,8	0,4
Average	0,4	-2,6	2,8	0,2

Note: *Supply* represents relative wage growth rates if the only change is in relative supply, or changes in relative wages along the relative demand curve. Wages are the values for relative wages derived from section 2.3.2. *Demand* represents the growth rates given by the common constant in (2.5).

a rate roughly similar to that for 1990s, and that this increase in education demand approximates that found for the United States during the same period. Moreover, the elasticity of substitution between more- and less-educated workers found by us is also quite similar to that estimated for the United States.

We are now ready to decompose the change in the education wage premium into the part attributable to changes in demand and the part attributable to changes in supply. Table 2.6 shows this exercise. Our results show that with no shift in education demand, the education wage premium would have fallen by 2.6% per annum between 1980 and 2000. By decade, this decrease would have been much stronger during the 1980s (3.4%) than during the 1990s (1.7%). In the presence of education demand shifts only, the education wage premium would have increased by 2.8% during both the 1980s and the 1990s. The last column shows that the change in the education wage premium predicted by the demand-supply model comes close to the change actually observed in the premium.

2.5 Labor Institutions

Let us now examine whether our conclusions above are robust to the influence of wage-setting institutions. These institutions have changed in almost all countries during recent decades. In Spain, a new system of labor regulation was introduced during the 1980s, [9] the most important feature of which was centralized collective bargaining (CB).[10] The latter involved wages being negotiated between unions and employer associations, as it did in other European Countries such as Germany.[11] CB agreements set a wage floor, and while only 18% of workers are paid at the negotiated rate, they tend to be the least-paid of all workers; see (Dolado, Felgueroso, and Jimeno 1997) (hereafter DFJ).[12] Hence, education premia could in principle be affected by CB wage floors; more importantly from our perspective, the evolution of the education wage premium in Spain may have been affected by rising CB wage floors leading to wage compression (while there is a minimum wage in Spain, this wage falls below the wage floor set by CB and is therefore not regarded as binding). One way to check whether trends in CB wage floors did indeed raise the wages of least-educated workers, relative to the market-clearing wage level, is to examine whether the unemployment rate among less-educated workers increased more rapidly than it did among other education-classified worker groups. Table 2.7, which lists the percentage change in unemployment rates by worker education category, shows no

[9] Ley del Estatuto de los Trabajadores (1980).

[10] Almost 50% of negotiations take place at sector-province level; 26.6% are at sectoral negotiations at the national level.

[11] Despite very low union affiliation, almost 80% of workers are covered by some collective bargaining agreement as negotiations are binding for most non-union workers.

[12] Dolado, Felgueroso, and Jimeno (1997) also explain that the real value of the minimum wage does not play any role in wage determination. They show that the lowest wages are determined by the CB wage floors, and that there is no link between the minimum wage and CB wage floors.

Table 2.7: Growth in Unemployment Rate by Education Groups in Spain

	1980-85	1985-90	1990-95	1995-00
none or primary only	77.3	-20.1	51.6	-34.2
lower-secondary	63.3	-35.4	28.8	-45.5
upper-secondary	56.4	-27.5	47.2	-34.0
tertiary	75.4	-27.4	31.1	-43.7

Note: data from Human Capital Series compiled by IVIE. Spain

marked differences between less- versus more-educated workers as far as unemployment trends are concerned. For example, between 1980 and 1985 (a time of rising overall unemployment), the rates for workers with higher versus lower levels of education exhibited similar trends. Between 1995 and 2000, a period characterized by falling unemployment, both the primary and upper-secondary education groups registered the same decline. In summary, over the course of the past twenty years a similar trend has characterized unemployment among workers in all education categories.

Another way to check whether our conclusions are driven by CB is to re-estimate our demand-supply model after excluding workers for whom the floors of CB agreements are likely to be binding. DFJ argue that bargained wages earned by the most CB-influenced workers fell below 125% of the minimum wage in the early 1980s and below 140% of the minimum wage in 1990. Because their study ends in 1996, we have to approximate their criterion for the year 2000. Our basic assumption is that the most CB-influenced workers earned less than the average wage reported for a Spanish "peones" (unskilled workers) in 2002 Wage Structure Survey, if we take into account CPI inflation between 2000 and 2002. Table 2.8 shows the results of our re-estimated model, after eliminating workers whose wages fell below the specified cutoffs. Here, the slope and the intercept are almost identical to those obtained earlier. Our conclusions regarding the evolution of the Spanish education wage premium therefore continue to hold.

2.6 Conclusions

The main aim of this paper was to examine the extent to which a demand-supply model may be used to explain the evolution of the education wage premium during the 1980s and the 1990s. Our key finding was that the evolution of relative supply represented the main driving force behind the changes in this premium, since the demand for education rose at a similar pace during the two decades under study. We also found that the increase in the demand for education during our study period–about 2.8% per year–roughly approximated that estimated for other countries, including the United States. Another important finding was that the elasticity of substitution between more- and less-educated workers in Spain was about 1.5, and thus almost identical to that obtained in previous cross-country and cross-regional studies. Our study therefore suggests that the trend in the Spanish wage premium can be explained quite well by market forces.

Table 2.8: Relative Demand Estimation exclusive of institutional effects

	I	II	III	IV
Δb_{80}	0.267*** (0.085)	0.237** (0.115)	0.271*** (0.083)	-2.800 (4.180)
$\Delta b_{90} - \Delta b_{80}$	0.063 (0.141)	0.063 (0.140)	0.074 (0.143)	0.064 (0.131)
$-\frac{1}{\sigma}$	-0.632*** (0.200)	-0.574** (0.252)	-0.631*** (0.208)	-0.535*** (0.198)
Δk	- -	0,112 (0.252)	- -	- -
Δk_{ict}	- -	- -	-0,038 (0.208)	- -
employ.	- -	- -	- -	0,488 (0.198)
Adj. R^2	0,47	0,49	0,49	0,49
n	34	34	34	34

Note: The dependent variable is changes in logs of changes in logs of regional relative wages to due extra education. $\Delta b_{80}t$ estimates changes in the(inverse) relative demand intercept in eighties. $\Delta b_{90} - \Delta b_{80}$ estimates differences in changes in the(inverse) relative demand intercept in nineties relative to eighties. $-\frac{1}{\sigma}$ estimates the coefficient associates to changes in logs of regional relative supply due to extra education or the (inverse) relative demand slope. Δk are changes in logs of regional physical capital per worker and Δk_{ict} are changes in logs of regional physical capital per worker in ICT sectors. Employment represents the log of regional employment at the start of the period. All estimations are performed controlling for regional and time effects. Institutional effects are assumed to be removed by deleting workers with wages below 25%, 40% and 60% of the minimum wage for 1980/81, 1990/81 and 2000/01. Institutions are assume to be eliminated deleting those workers with wages lower than 25%, 40% and 60% of minimum wage for 1980/81, 1990/81 and 2000/01.

*** implies significance at 1%, ** at 5% and * at 10%.

Values in parenthesis are standard deviations.

Chapter 3

Estimating Human Capital Externalities: The Case of Spanish Regions

3.1 Introduction

Verifying the existence of human capital externalities, which are defined as the difference between private and social marginal returns to education within a city, region or country, is of crucial importance because of different reasons. First, because the existence of externalities may justify public subsidies to education and to immigration of highly qualified workers. Second, because the empirical identification of human capital externalities can help in explaining income differences across countries, according to recent contributions to the development literature (for instance Lucas, 1988). Nevertheless, the identification of human capital externalities at the aggregate level is far from being convincing, despite the relevant existing theoretical and empirical work (see for some examples Lucas 1988, Azariadis and Drazen 1990, Benabou 1996, Black and Henderson 1999, Acemoglu and Angrist 2001, Rudd 2000, Moretti 2004a, Moretti 2004b, Ciccone and Peri 2006).

For the Spanish case, there are several estimations on human capital returns, but almost all of them have focused on private returns (among others Alba and Segundo (1995); Barceinas, Oliver, Raymond, and Roig (2000); Raymond (2002); De la Fuente (2003) and De la Fuente, Domenech, and Jimeno (2003)). In Spain the study of externalities from education could be of particular importance since the percentage of workers having attained secondary education has increased dramatically in the last decades. In 1980 the percentage of workers having attained high school or higher degrees was slightly more than 20%, while in 2001 this percentage was larger than 70%. If there are any externalities from education the effects should be sizable, and therefore Spain could be an interesting case for this issue. Nevertheless, there are very few estimations of education externalities in Spain. One of them is Alcalá and Hernández (2005), who use data coming from a wage survey to check for possible externalities at the firm and sector level. They find that both the average level of education of workers as well as the percentage of workers with college degrees increase the average wage within firms. Our paper, instead, contributes to this literature by trying to identify education externalities at the aggregate level using both individual information to obtain average wages at the regional level, and regional data to obtain empirical evidence on the effects of the intensity of human capital at the regional level.

This paper uses the main methodologies that have been proposed to identify human capital externalities. One methodology uses Mincerian equations including a variable related to the general level of human capital (for example Rudd 2000, Acemoglu and Angrist 2001, Moretti 2004a). This methodology identifies human capital externalities by looking at the marginal effect on individual wages of the average human capital of the city, region or country where the individual lives. The basic idea is that the individual wage is determined not only by individual characteristics but also by the average human capital level of the region. A variable that measures the average human capital level is included in the traditional Mincerian regression. We call this the Mincerian approach (M). This methodology may have some problems in identifying human capital externalities because of various problems. First, it implicitly assumes that there is perfect substitutability between the different groups of workers defined by different human capital levels. Imperfect substitutability between workers with different skills, though, would imply a relative labor demand with negative slope, and the M approach implies a positive effect of the aggregate human capital variable even if the social return is smaller than the private return. The intuition is simple: an increase in the number of skilled workers in a region relative to the rest of regions would imply an increase of the wages of unskilled workers simply because of the substitution effect, even if there are no externalities. The empirical evidence for the United States (Katz and Murphy, 1992; Ciccone and Peri, 2006), as well as for other countries (Angrist 1995) and particularly for Spain, in this thesis, chapter 2 shows that the elasticity of substitution is quite different from one, and therefore the M approach may fail in identifying human capital externalities. It is necessary therefore to complement the M approach with other methodologies that take into account the possibility of imperfect substitutability of skilled and unskilled workers. For instance Ciccone and Peri (2006) propose the Constant Composition approach (CC) which is valid for any value of the elasticity of substitution. According to this approach, and as they show theoretically, the externalities are defined as the marginal effect of the intensity of human capital on average aggregate wages, holding constant the composition of the labor force.

We are going to use data from the Spanish autonomous regions (Comunidades Autónomas), even despite the fact there is more disaggrated data available. Individual data at a more disaggregated level is not available for a sufficient time span to apply the different methodologies. We choose to work on a more aggregated level and get a longer time span (1980-2000) that covers a period where there were important changes in the average education level of the labor force.

One of the main problems that arises when estimating human capital externalities is that a regression of aggregate average wages against an aggregate measure of human capital is very likely to imply a problem of endogeneity in all possible specifications. Human capital is not distributed randomly at the regional level. Cities and regions with higher productivity, and therefore with higher average aggregate wages, will attract more skilled workers simply because of higher standards of living or a larger supply of amenities. Firms may also choose to locate in cities and regions where there is a higher average level of human capital in order to reduce search costs or trying to appropriate externalities generated by a more specialized labor market. Therefore it is reasonable to assume that these two variables are both endogenous. As a consequence, the direction of causality between wages and aggregate human capital level has to be identified using appropriate instruments. Furthermore the use of different proxies for the average human capital intensity at the regional level may imply important measurement errors which may exacerbate the correlation of the explanatory variables and the random terms of our equations. We are going to use as instruments the population structure in terms of the weight of young and old groups prior to 1980. This is predetermined for the period that we consider, but the change in human capital may be correlated with the weight of young and old people in the population at the beginning of the sample period.

Using different methodologies we find obviously quite different results, but overall our results seem to confirm the existence of important education externalities of human capital at the regional level in Spain. As expected, the largest effects can be found using the M approach, since this methodology adds the substitution effects to the externalities effects, if any, as we are going to show later. Using the CC approach, which controls for substitution effects, we still find a positive effect. The overall conclusion is that the human capital externalities account for slightly more that one third of the increase in salaries during the period 1980-2000.

The rest of the paper is organized as follows. Section 3.2 summarizes the relevant literature, while section 3.3 reviews the two main empirical methodologies used in this paper. Section 3.4 presents the data sources used. The main results and its interpretation can be found in section 3.5 and finally section 3.6 concludes.

3.2 Related Literature

Theoretically, the externalities of human capital are identified as the difference between the social and private marginal product of skilled workers. Typically, human capital is associated with education. The estimation of education returns, after decades of empirical work, has arrived to a consensus both in Spain and for other countries. In most cases, the return for an extra year of education is estimated to be an increase between 5 and 12% of average wages. For the case of the social return, instead, a consensus has still not be reached. From an empirical point of view.[1] there are several attempts to estimate the difference between social and private returns (J.E. (1993), Acemoglu and Angrist (2001), Rudd (2000), Moretti (2004a) and Moretti (2004b)) In the first contribution, Rauch (1993) uses a Mincerian equation where he includes an aggregate measure of qualification. The intuition behind this approach is that in the determination of wages apart from individual characteristics, such as education, experience, sex, productive sector, etc... it is also important to consider also the average education level of

[1] For a theoretical survey see Moretti, 2003

the city, region or country where the individual lives. The idea is that skilled workers can increase the productivity of the rest of workers because they increase the productivity of productive activities (productivity spillovers). For this reason a variable is included in the Mincerian equation to proxy the average education level (the weight of skilled workers over the total) of the workers for the location that is considered. For this estimation, Rauch uses data for wages and human capital from the U.S. Census for individuals of 237 cities in 1980. His results show that the marginal effect on average wages of an extra year of schooling is statistically significant and estimated to be between 3 and 5 per cent.

There are two weaknesses in this approach. First, unobservable characteristics of both individuals and cities that could be correlated with the explanatory variable could bias the results, since this omitted variables would be included in the residuals. The same effect would be present if there were measurement errors for the aggregate and individual human capital variables. Moretti (1998) tries to control for possible endogeneity of average years of schooling and measurement errors. As instruments, he uses the age structure of cities as well as an indicator variable for the presence of a land-grant college in the city. Using the share of college students as a proxy for the aggregate level of human capital he finds that an increase in 1% in this share raises average wages between 0.6 and 1.2 % above individual returns. Using also instrumental variable methods, Conley, Flyer, and Tsiang (2003) find significant externalities of education for Malaysia. Acemoglu and Angrist (2001) use changes in legislation related to compulsory basic education at the state level as instruments for the aggregate level of education. They also use instruments to determine the possible endogeneity of the individual level of education (quarter of birth). In this case, and despite the fact that their OLS results are similar to the results by Rauch and Moretti, the estimated coefficient is smaller and in a lot of cases not significantly different from 0.

The second weakness of Rauch's approach is that he identifies externalities under the assumption that workers with different human capital endowments are perfect substitutes in production. Perfect substitutability simplifies identification because it implies that changes in the relative supply of human capital do not affect the relative wages of the different human capital groups, if we hold total factor productivity constant. Consequently, all the effects that human capital supply changes have on workers with a given level of human capital have to come through total factor productivity and can be therefore interpreted as externalities. Nevertheless, the assumption of perfect substitutability of workers may bias the estimation of externalities. When workers from different schooling groups are imperfect substitutes in production, externalities can be wrongly identified (Ciccone and Peri 2006). This is so because the increase in human capital supply could increase average wages even without externalities. In any case, Moretti (2004a) shows that the possible bias caused by the assumption of perfect substitutability between different groups of workers when estimating externalities is not the most important one. He finds externalities analyzing social returns of education for different schooling groups. The idea is that if there is imperfect substitution, changes in the wages of low skilled workers, due to a larger share of high skilled workers, would be given by the sum of the effects of relative supply changes and by externalities. Both effects would be clearly positive, and therefore the total effect would be also positive. Nevertheless the change in the wages of high skilled workers would be the sum of two contradictory effects. If the total effect were positive, the effects of spillovers would clearly be larger than the effects of relative supply changes. Moretti finds that workers with high-school or high-school dropouts have a average wage increase between 1.6 and 1.9% as the results of an increase of a 1% of the college share in the city. Furthermore, wages of college graduates increases 0.4% as an effect of the increase in the college share.

Ciccone and Peri (2006) find that the bias which is incurred because workers of different skills are assumed to be substitutes can be very large (70% of estimated externalities). For this reason they propose an alternative methodology. According to these authors, the externalities can be estimated through the marginal effect of human capital on average wages holding constant the composition of the labor market. This method would be valid for all degrees of substitutability among workers. Another advantage of this approach with respect to previous methodologies is that human capital externalities can be estimated without an estimation of the individual return to education and experience, which may be problematic since individual education may be endogenous and measured with error (Acemoglu and Angrist 2001). This approach is based on the estimation of aggregate social returns and therefore does not depend on parametric assumptions on the relation of education and experience with individual human capital levels or wages. Using this methodology, Ciccone and Peri do not find any statistically significant externalities for the American cities and states between 1970 and 1990.

In the next section we describe these methodologies in detail.

3.3 Empirical Approaches to Human Capital Externalities

According to the literature review on human capital externalities, we can find two main approaches to identify them empirically. The first one is based on an analysis based on Mincerian equations. We call this the Mincerian approach (M approach). The second approach is the one proposed by Ciccone and Peri (2006), which has been termed the Constant Composition approach (CC approach).

3.3.1 The Mincerian Approach

As mentioned earlier, since the work by Rauch (1993) there have been many attempts to estimate social returns to education using a typical Mincerian equation, as in Mincer (1974). The main idea is to introduce an additional variable proxying the average endowment of human capital at the city, regional or country level. Let's look at a simple case including only two types of workers (qualified and non-qualified). In this case, a simple version of the wage equation can be written as:

$$\log(w_{rt}) = \theta h_{rt} + \alpha_{rt} + bD_{it}, \tag{3.1}$$

where w_{irt} is the wage of worker i in region r and year t, h_{rt} represents the value of human capital in the region and D_{it} is a dummy which is equal to 1 for qualified workers and 0 otherwise. In this case, θ can be interpreted as the social marginal return of the regional human capital on individual wages, showing therefore the value of externalities.

In order to estimate θ a two-stage procedure can be used. In the first stage, a classical Mincerian regression is estimated on the log of individual wages against individual characteristics which are supposed to affect wage determination: in our case we use education, experience, sex and sectoral and regional dummies. The goal of this first stage is to estimate average regional wages clean of individual characteristics, including private returns to education:

$$\log(w_{irt}) = \alpha_{rt} + \gamma s_{irt} + \sum_{k=0}^{K} \beta_t^k z_{irt}^k + u_{irt} \tag{3.2}$$

where w_{irt} is the wage of individual i at region r during year t, s_{irt} is average years of schooling, z_{irt} are the K individual characteristics that we want to control, and u_{irt} captures the effect of unobservable variables and estimating errors. The constant terms α_{rt} correspond to average wages for each region and year.

Since there may be unobservable variables included in the error term, which may be correlated with the schooling variable, the model may present endogeneity problems. These omitted unobservable variables, for instance innate ability, can cause estimation biases and have to be dealt by using appropriate instruments. In the previous section we commented on some of the instruments used in the literature. For the Spanish case it is not clear which instruments to use, since individual data are not available.[2]

In the second stage, in order to obtain the marginal effect of the human capital of region r over average wages, we compute the first difference in constant terms for each region ($\Delta\hat{\alpha}_{rt} = \hat{\alpha}_{rt} - \hat{\alpha}_{r(t-1)}$) and we regress these differences in average wages on the change in our human capital indicator:

$$\Delta\hat{\alpha}_{rt} = controls + \theta\Delta\hat{h}_{rt} + v_{rt} \tag{3.3}$$

where $\Delta\hat{h}_{rt}$ represents the change in human capital endowment in region r between $t-1$ and t. The value of human capital externalities according to the M approach is represented by θ. Some controls are also introduced corresponding to variables which affect changes in average wages which are not related with changes in the endowment of average human capital in the region. Finally, estimation of (3.3) has to be performed using instrumental variables due to the possible endogeneity of average wages and human capital stocks.

As instruments we use the demographic structure in 1981 and 1991. The absence of changes in legislation related to education across regions, as well as the scarcity of statistical information, seriously restricts our choice of instruments. Thus, we assume that the demographic structure of a region at the beginning of our sample period, determines how much the population with a

[2]There are several examples of the estimation of private returns to education without controlling for possible endogeneity, for instance Alba and Segundo (1995), Oliver, Raymond, Roig, and Barceinas (1999) and Vila and Mora (1998). Nevertheless there are some examples of the use of instruments, for instance Arrazola et. al. (2001).

given educational level can grow. If we assume that education is subsidized and that, a priori, everybody can access up to college, it can be assumed that those regions with more young people will experience a higher increase in the proportion of high skilled workers. In this case we can safely assume that the initial demographic structure of regions is independent of the wage level for future years, and therefore we can use this demographic structure to proxy regional changes in our human capital indicator. Summing up, the identifying assumption is that the demographic structure of the Spanish provinces in 1980 and 1990 is independent of the technological change at the province level for 1981-1991, as well as measurement errors and omitted variables.

But as stated earlier, the main problem with this methodology is that, even if the externalities were non-existent, the estimated coefficient would be positive and significant in the case that workers with different schooling levels were not perfect substitutes. To verify this, we turn to the simplest specification (3.1). We sum up this expression for workers and compute the average wage. This way we obtain $(1-h)\log(w_L)+h\log(w_H) = (\theta+b)h+a$ and therefore $\theta = \partial((1-h)\log(w_L)+h\log(w_H)/\partial h - b$. Since $b \equiv \log(w_H) - \log(w_L)$, we have that:

$$\frac{\partial((1-h)\log(w_L)+h\log(w_H))}{\partial h} - (\log(w_H) - \log(w_L)) = \theta^M \tag{3.4}$$

being h the weight of skilled workers over the total. If we maintain the weight between both groups constant to a period 0, expression (3.4) is equivalent to:

$$\theta^M = (1-h_0)\frac{w'_L}{w_L} + h_0\frac{w'_H}{w_H} \tag{3.5}$$

where $w'_L = \partial w_L/\partial h$ and $w'_H = \partial w_H/\partial h$. In the case of no externalities this method predicts that $\theta^M = 0$ if and only if $w'_L = w'_H = 0$, that is, if both groups of workers are perfect substitutes. Therefore, the M approach can incorrectly identify externalities in the case where there are none if the groups are not perfect substitutes.[3] It is therefore necessary to use other estimation methods which are valid also for the case of imperfect substitutability.

3.3.2 The Constant Composition Approach

The theoretical basis for the CC approach proposed by Ciccone and Peri (2006) is summarized as follows: under general conditions, the value of the externalities of human capital is equal to the average weighted effect that human capital has on wages, which in turn is equal to the marginal effect of human capital over average wages holding constant the composition of the labor force.

Let us suppose the we have, as in the previous section, two types of workers, skilled (H) and unskilled (L). If high skilled workers generate externalities, the social marginal product, $\partial Y/\partial H$, will be larger than the difference of the wages of high skilled workers compared to the rest, $w_H - w_L$. Therefore it can be said that

$$\frac{\partial Y}{\partial H} = EXT + (w_H - w_L) \tag{3.6}$$

Assuming Constant Returns to Scale in the production of Y, total product will be equal to the weighted sum of total earnings, that is $Y = hw_H + (1-h)w_L$, with h representing the weight of skilled workers over the total. Differentiating with respect to the relative supply of high skilled workers $\partial Y/\partial h = (w_H - w_L) + (1-h)\partial w_L/\partial h + h\partial w_H/\partial h$ and equating this last expression with the one in (3.6), defining $\theta^{CC} = EXT/Y$ as the proportion

[3] Furthermore, Ciccone and Peri (2006) evaluate the possible bias in the estimation of θ^M as

$$\theta^M = \frac{1}{\rho}\left(\frac{w_H - w_L}{w}\right)$$

where ρ is the elasticity of substitution between both groups of workers and w is average wage, a CES production function is used and there are no externalities. This implies that the M approach offers an estimation which is positively biased by 70%, according to the evaluation of these authors.

of externalities over total product and $\beta = Hw_H/Y$ as the share in total earnings of high-skilled workers, we obtain

$$\theta^{CC} = \frac{EXT}{Y} = (1-\beta)\frac{\partial w_L/\partial h}{w_L} + \beta\frac{\partial w_H/\partial h}{w_H} = \frac{\partial \log((1-h_0)w_L + h_0 w_H)}{\partial h} \tag{3.7}$$

where h_0 has been held constant for period 0. Therefore externalities can be identified as the log-change in the average wage holding skill-composition constant.

Therefore to apply this estimation methodology we need again to perform a two-stage procedure. In the first stage it is necessary to obtain a measure of weighted average wages holding the composition constant for the different periods. For this purpose average wages are computed for each qualification group and they are used to obtain weighted average wages using the weights for one of the years included in the period. To implement this we estimate average wages by schooling levels once we eliminate wage differences unrelated to education. Call w_{isrt} the wage of individual i with schooling level s in region r at period t, and z_{isrt} the characteristics of this individual that we want to clean out from our measure of average wages. We construct a measure of adjusted average wages of workers with schooling level s in region r and period t as the estimated constant in the following regression:

$$\log w_{isrt} = \sum_{q=0}^{R} \alpha_{qst} D(r=q) + \sum_{k=0}^{K} \beta_{st} z_{isrt} + u_{isrt} \tag{3.8}$$

where, as in equation (3.2), R is the total number of regions, K are the individual characteristics that we want to control, $D(r = q)$ is a dummy variable which is equal to 1 if $r = q$ and 0 otherwise, and u_{isrt} is a residual. This equation provides estimations for $\hat{\alpha}_{qst}$, the average wage of workers with qualification s in region r at moment t, adjusted by characteristics z.

Next, we use these adjusted wages by qualification, region and time to construct an average adjusted wage holding composition constant:

$$\log \hat{w}_{rt}^F = \log \sum_{s=1}^{S} l_{srT} \alpha_{srt} \tag{3.9}$$

Notice that the proportions l_{srT} correspond to the base year T. The average wages computed in (3.9) allow us to evaluate the increase in adjusted average wages holding constant the composition of the labor force, $\log(\hat{w}_{rt}) - \log(\hat{w}_{rT})$. Finally, we estimate the intensity of externalities by an empirical formulation of a discrete version of (3.7):

$$\log(\hat{w}_{rt}^F) - \log(\hat{w}_{rT}^F) = controls + \theta(\log(h_{rt}) - \log(h_{rT})) + u_t \tag{3.10}$$

Controls include the change in total employment in the regions to take into account scale effects as well as changes in physical capital, since an increase in this factor may also increase average wages. Since we are working with growth rates, permanent changes in wages at the regional level do not affect our results. For instance, if firms in the service sector of Madrid or Catalonia pay wages which are 30% larger than the firms in Andalusia due to higher living costs, then these differences in wages will not affect our results as long as they are constant over the period considered. Generally, shocks that increase average wages in all regions equally (such as the national inflation rate) do not affect the results since they simply get absorbed by the constant of our regressions. For the estimation of equation (3.10) it is necessary to take into account that the proportion of high skilled workers and the total number of workers in each region, used as a control, may be endogenous and measured with error. Furthermore there may be measurement errors in both the change in the measure of human capital intensity as well as the increase in the labor force. Measurement error refers in this case not only to the usual problem of how variables are measured or proxied, but also to an additional problem: the schooling systems of the different regions show important differences in quality which we are not taking into consideration, and, therefore, the schooling measures should have to be adjusted for these differences in quality in the educational systems. Unfortunately, quality adjusted educational data is not available.

We need again to find instruments to deal with the possible endogeneity of average wages and the human capital stock of a city, region or country. The instrumental variable methodology can also help in correcting the biases implied by the use of indicators that have measurement errors as well as not taking into account production prices of each region. In this last case, the problem can be interpreted as an omitted variable problem. This is because our model states that the change in real average wages is a function of the change in our labor force quality measure, $w/p = f(h)$. If we had changes in production prices in each province between 1981 and 2001, we could estimate directly this last equation. Since there is no data on this, we estimate $w = f(h)$. Therefore the percentual change in production prices would appear as an omitted variable in the right hand side of the equation. As long as this variable is not correlated with the instruments that we are using, instrumental variable estimation will correct the potential biases. Consequently, we are going to use the same instruments as the ones defined in the second-stage estimation of the M approach.

An important issue in this method is that the estimation of externalities depends on the weights chosen to compute the constant composition average wages. Ciccone and Peri (2006) show that if the production function is concave with respect to high skilled workers, that is, if it has marginal returns to scale with respect to high skilled workers net of externalities, then the values of the estimated externalities that we obtain choosing the weights of the initial and final period constitute a lower and an upper bound of the true externalities (not necessarily respectively). The true value of externalities lies in an intermediate unknown point between the two bounds. For this reason, in this work we are going to present two complementary estimations.

The first one uses a base year equal to the initial year of each period that we study (1980 and 1990), while the second one uses the central year of each period.

3.4 Data and Instruments

In order to implement the empirical approach presented in the previous section several data sources have been used. Individual data (used for individual and average wages) come from the Survey of Family Budgets[4] corresponding to 1981 and 1991 (EPF-81 and EPF-91) and the Continuous Survey of Family Surveys [5] corresponding to 2000/01 (ECPF-00-01). Regional data for human capital regional endowments were provided by the Human Capital Project of the Instituto Valenciano de Investigaciones Económicas (IVIE). Regional physical capital data has also been used (BDMores database from Instituto Nacional de Estadística) for robustness analysis. Finally, as instruments we have used the age composition of the population on the basis of the Population Census for the years 1981 and 1991.[6] The regional level that we use is the autonomous regions (Comunidades Autónomas, 17 excluding Ceuta and Melilla).

3.4.1 Individual Data

Individual data come from the Survey of Family Budgets (1980-81 and 1990-91) and the Continuous Survey of Family Budgets (2000 and 2001). These surveys give information on earnings for each member of family units. To better proxy wage earnings, we restrict the sample to main earners between 16 and 65 with dependent earnings and working more than 15 hours. We

[4] *Encuesta de Presupuestos Familiares, Instituto Nacional de Estadística,*

[5] *Encuesta Continua de Presupuestos Familiares, Instituto Nacional de Estadística.*

[6] *Censo de la población, Instituto Nacional de Estadística.*

also eliminate individuals with earnings below the minimum wage. We end up with 7,027 individuals for 1980-81, 8,193 for 1990-91 and 2,057 for years 2000 and 2001.

Individual data are necessary to implement equations (3.3) and (3.9). The variables constructed are the following:

- *Wage earnings:* Earnings from personal work excluding self–employment.

- *Years of schooling:* This is a crucial variable for the estimation in the M approach. It represents the average level of human capital for each worker and is used to estimate private returns to education. It is constructed by assigning to each schooling level the minimum amount of years needed to complete a degree. If the education system changes across the different surveys (EPF-81, ECP-91 and ECPF-00-01) a common denominator has to be found and education levels have to be aggregated into wider schooling groups. Table 3.1 shows the different schooling groups that we define and the schooling levels included for each of the surveys. As shown by the table, EPF-81 defines 8 schooling levels, EPF-91 defines 10 and ECPF-00-01 defines 7. This way, assigning to each level the corresponding years of schooling, we can construct our variable of average years of schooling. In those groups where there are 2 o more schooling levels, average schooling is computed as a weighted average according to the weight that each group has in the group total.

  The results confirm that between 1980 and 2000 average years of schooling increased, according to this methodology, from 9 to above 11, with the largest increase occurring during the 1990s. These results are very similar to the results of other authors, for instance Olivier, Ramos, and Raymond (2001).

Table 3.1: Schooling groups. Individual Data

	epf 80/81	epf 90/91	ecpf 00/01
Illetary-Primary	1-3	1-4	1-2
Low Secondary	4	5 & 7	3
Acad. & Voc. Upper Sec.	5-6	6 & 8	4-5
Higher Educ. (Short cycle)	7	9	6
Higher Educ. (Long cycle)	8	10	7

To estimate equation (3.9) instead of average years of schooling we need different schooling groups. We use the following variable:

- *Education:* This is a categorical variable with three schooling levels, the same levels that we used before: primary education or less, secondary education (lower and upper) and college (higher education, short and long cycle). We construct dummies using this categorical variable. Consequently we estimate equation (3.8) 9 times (for each of the cross sections 1981, 1991 and 2001 and for each of the schooling levels).

The controls that we use in equation (3.8) are the following:

- *Sex*

- *Experience:* We construct this variable as the difference between age and years of schooling (according to the Education variable) minus 6. We construct a categorical variable grouping experience in three levels: 20 years or less, between 20 and 30, between 30 and 40, and finally 40 or more. Dummies are derived from this categorical variable.

- *Sector:* Categorical variable with two categories: rural and non-rural worker.

- *Married:* Married or non-married worker.

3.4.2 Regional Data

The Human Capital project provides data for workers with different schooling levels for the years 1980-81, 1990-91 and 2000-01. The BDMores database provides data for the capital stock and employment, used for robustness checks.

We aggregate the schooling levels in three groups, that correspond to the groups that we can define for the CC approach. Once we define these groups, we compute the proportions l_{srt} used in equation (3.9):

- *Proportion of workers by education level.* With this variable we construct average wages used in equation (3.9) and compute the growth in adjusted average wages holding the composition constant. These two variables are used as our dependent variable in the second stage estimation.

Additionally, we construct the following variables at the regional level:

- Fraction of workers with high school or more. This variable is constructed with the Human Capital project of IVIE. With this variable we construct the variables that we use in the second stage, both for the M and CC approach. We are going to use also workers with high school and higher degrees for our estimations of the different production functions:

  LED: The variation in human capital intensity as defined by workers with high qualification at the regional level.

These are our measures of the intensity of high qualification at the regional level. Other variables that we use are:

- *Total workforce at the regional level.* This data comes from the Active Population Survey ($LEMP$).

- *Stock of physical capital.* Data from the BDMores database (K).

We complement these data with the Census of Population for 1981. We construct age group proportions: from 0 to 4, 5 to 9, and so on (18 groups). The last group gathers the population older than 85.

Table 3.2: Average wages in Spanish regions. 1980-2001

	1981	1991	2001
Andalucía	620.771	1.374.273	2.006.496
Aragón	726.164	1.552.504	2.404.515
Asturias	770.411	1.615.257	2.505.721
Baleares	670.676	1.460.103	2.191.425
Canarias	644.695	1.387.754	2.020.738
Cantabria	711.886	1.596.316	2.242.165
Castilla y León	690.036	1.567.936	2.155.668
Castilla-La Mancha	626.355	1.419.734	2.010.670
Cataluña	779.218	1.709.741	2.393.477
Comunidad Valenciana	675.078	1.389.134	2.207.454
Extremadura	541.652	1.320.384	1.760.094
Galicia	683.013	1.452.852	2.153.270
Madrid	815.915	1.657.278	2.468.719
Murcia	597.704	1.388.218	1.942.676
Navarra	801.067	1.777.856	2.683.833
País Vasco	803.115	1.758.933	2.381.716
La Rioja	697.790	1.531.849	2.183.584
España	694.516	1.514.939	2.232.096

3.4.3 Instruments

The endogeneity of explanatory variables may bias our results and generate inconsistent estimations. At the aggregate level, endogeneity may arise due to skilled workers interregional mobility. As it has been correctly stated by previous works, the geographic distribution of skilled workers does not seem

Table 3.3: **Wage per schooling. 1980-2001**

		1981	1991	2001
primary or less	mean	612.622	1.266.508	1.932.901
	stdv	278.190	584.631	652.061
secondary	mean	821.892	1.721.526	2.395.398
	stdv	324.575	1.182.154	886.831
college	mean	1.061.663	2.323.486	2.842.428
	stdv	371.898	938.236	1.027.295
Total	mean	694.516	1.514.939	2.232.096
	stdv	330.034	878.318	884.291

to be random. The usual explanation is that higher salaries in a region can imply high skilled workers migration to that region from other worse-paying regions and therefore a higher participation of high skilled workers in the receiving region. Another factor that may work in the same direction is that a region with average higher income will have better amenities, with a higher attraction of high skilled workers. Finally, firms may decide where to locate basing their decision on the characteristics of the city or region, especially wages, land cost or other factors related to the cost structure faced by the firm (Moretti 2003). Due to all these reasons there may be simultaneity between the average level of wages and the average educational level in the regions and cities, causing endogeneity problems. This may be exacerbated by measurement errors in some of the key variables used.

All these reasons call for the use of appropriate instruments for some of the problematic variables. In our case, the instruments used in a two-stage least square estimation procedure are related to the population structure of the regions. The instruments chosen are the proportions that different age groups represent over the total population. The data come from the Census of Population published by the Spanish National Institute of Statistics for

1980-81 and 1990-91. The variation in the supply of high skilled workers within a decade, assuming that access to education is free and identical in all regions (which is true for the case of Spain), among other things depends on the percentage of population in different age groups that may be acquiring skills. The main assumption is that the higher the percentage of young people, the higher the growth of skilled workers in the future. In other words, we assume that a region with a higher endowment of young people in 1980 may have had a larger increase in the number of high skilled workers in the next decade than another one with a smaller participation of young people in its demographic structure. Nevertheless it is also possible that the higher the proportion of workers older than 50, the larger the renewal of low skilled workers in the labor market, since older workers have a lower average education level than young workers. Considering the high correlation between the different age groups at the regional level (see table 4), it is possible that the sign of the coefficients estimated in a regression of these indicators and the variation in the percentage of of high skilled workers cannot be predicted a priori. But in any case the demographic structure at the beginning of the decade can be safely assumed to be independent of the growth in wages paid in the subsequent decade, and that is why we consider these indicators as appropriate instruments.

To verify this, we show the regressions of our main human capital intensity measure (LED) against the demographic structure age proportions in table 3.4. We show regressions for all age groups and for only young, intermediate and old groups. If we include all age groups, we find that these are jointly significant, but some of the intermediate groups appear as insignificant, which is a consequence of the collinearity between the different age groups. Excluding these intermediate groups increases the R^2, but some coefficients show counterintuitive signs. We prefer therefore to use all age groups in our regressions.

Table 3.4: First Stage Regression (II)

Dependent variable:. log changes in LED.		
Age Interval	**Coef.**	**Std.**
0-19	8.92	1.20
45-70	9.78	2.19
constant	-2.91	0.57
adjusted-R^2		0.69
F		36.78
Prob>F		0.00
n		34

3.5 Estimation and Results

We first discuss the results for the M approach, followed by the CC approach results.

3.5.1 The Mincerian Approach

In Table 3.5 we show the estimations of the Mincer Regressions according to equation (3.2). The variables included in these regressions are average years of schooling (s), experience, experience squared, a sex dummy (1=woman), a dummy for married (1=married), an agricultural worker dummy (1=agricultural worker)[7] and lastly regional fixed effects (which are the "$\hat{\alpha}$'s" in the equation), in order to capture the effects on individual wages of regional

[7]This dummy has been interacted with a regional dummy, since agriculture structures are quite different across regions, affecting in different ways wages paid in agriculture. In the table, in order to save space, we show only weighted average values.

characteristics not included in the previous variables. The results show that the fixed effects are highly significant. All other variables are also highly significant, but less so for more recent years. This may be due to limitations in the survey used for our data, which for more recent years include fewer individuals with relevant information for our estimations, especially for some regions, reducing considerably the degrees of freedom in our estimations and making them more imprecise. In any case, the results seem strong enough for all periods to warrant the interpretations below. As in previous studies, schooling shows a positive and significant impact on earnings at the individual level. Each additional year of schooling increases wage earnings by 6.4% in 1981, by 6.5% in 1991 and by 7.9% in 2001. The rest of the coefficients show reasonable signs and magnitude.

Once we have estimated the fixed effects we can estimate equation (3.3). In Table 3.6 we present the results for this second stage. In columns (1) and (2) we show the estimations considering 1981, 1991 and 2000 and the growth in these three years (two decades), while in columns (3) and (4) we show only the results for the global growth between 1980 and 2000. For the two first sets of results we have included a dummy to capture changes in trends between decades ($TIME$). For both cases we show OLS results (columns 1 and 3) and two-stage least squares (2SLS) using the instruments discussed in the previous section (columns 2 and 4). The value estimated for externalities in column 1 is 0.239 while for 2SLS it increases to 0.333. In the first case the estimated coefficients for externalities are significant at the 5% and in the second at the 1% level. Furthermore, overidentifying restrictions cannot be rejected at the usual significance levels. Columns 3 and 4 also reveal significant coefficients for externalities but with slightly lower values. These results show that the M approach is able to find significant educational externalities in the Spanish regions for the period between 1980 and 2000. Given these estimations, and considering that our human capital

Table 3.5: Mincer Regressions

	1981			1991			2001		
Aragón	0.183	0.021	***	0.204	0.023	***	0.124	0.030	***
Asturias	0.214	0.025	***	0.219	0.035	***	0.143	0.029	***
Baleares	0.108	0.033	***	0.163	0.033	***	0.119	0.036	***
Canarias	0.098	0.024	***	0.046	0.026	*	-0.026	0.032	
Cantabria	0.148	0.028	***	0.178	0.038	***	0.031	0.038	
Castilla y León	0.128	0.015	***	0.153	0.017	***	0.020	0.026	
Castilla-La Mancha	0.090	0.019	***	0.134	0.021	***	0.045	0.030	*
Cataluña	0.265	0.015	***	0.225	0.020	***	0.132	0.021	***
Comunidad Valenciana	0.163	0.018	***	0.122	0.020	***	0.104	0.022	***
Extremadura	-0.068	0.025	***	0.019	0.029		-0.082	0.033	***
Galicia	0.092	0.019	***	0.104	0.021	***	0.018	0.027	
Madrid	0.252	0.018	***	0.182	0.024	***	0.119	0.022	***
Murcia	0.049	0.034		0.098	0.033	***	-0.024	0.032	
Navarra	0.310	0.034	***	0.310	0.035	***	0.231	0.036	***
País Vasco	0.294	0.019	***	0.270	0.020	***	0.118	0.027	***
La Rioja	0.175	0.035	***	0.066	0.036	*	0.050	0.036	
s	0.064	0.002	***	0.069	0.002	***	0.079	0.003	***
e	0.032	0.002	***	0.034	0.002	***	0.026	0.004	***
e^2	-0.001	0.000	***	-0.001	0.000	***	0.000	0.000	***
sex	-0.296	0.020	***	-0.257	0.015	***	-0.287	0.024	***
marry	0.362	0.027	***	0.090	0.026	***	0.034	0.033	
agriculture	-0.450	0.014	***	-0.435	0.020	***	-0.212	0.029	***
constant	12.067	0.048	***	13.072	0.047	***	13.608	0.048	***
n		11402			8838			3751	
adj-R^2		0.315			0.310			0.278	
F		240.320			181.820			66.910	

Notes: s represents years of schooling. e and e^2 are experience and experience to the square. Agriculture is the average for the total regional effects of this sector

Data in parenthesis are standard deviations.

*,**,*** denotes estimates that are significantly different from 0 at the 10%, 5% and 1% levels.

intensity measure has increased in average 2.6% between 1980 and 2000, the estimated effect on the growth of wages due to externalities would be 0.80%

yearly considering the estimation in column 2 or 0.70% if we use results in column 4. If we consider an actual approximate growth of real wages of 0,86% since 1980,[8] our interpretation would be that almost of wages growth is due to educational externalities.

Table 3.6: Mincerian Approach Estimation

Dep. variable: change in regional fixed effects.								
	81/91/00				81/00			
	(1)		(2)		(3)		(4)	
Led	0.239	**	0.333	***	0.268	***	0.27	***
	(0.134)		(0.155)		(0.061)		(0.061)	
Time	-0.4913	***	-0.468	***	-		-	
	(0.036)		(0.040)		-		-	
Constant	0.881	***	0.833	***	-0.282	***	-0.283	***
	(0.069)		(0.080)		(0.056)		(0.055)	
Adj. R^2	0.97		0.97		0.52		0.52	
n	34		34		17		17	
P-value of overidentifying restrictions			0.21				0.26	

Note: Data in parenthesis are standard deviations.
Led. Growth rate of college vs total workers ratio.
*,**,*** denotes estimates that are significantly different from 0 at the 10%, 5% and 1% levels.

3.5.2 The Constant Composition Approach

As in the previous section, the estimation of externalities using the CC approach is done in two stages. In the first stage we compute average wages for each group by schooling level and year by means of regressions that clean

[8]This approximate growth of wages can be obtained for instance by dividing total real worker earnings by the total number of workers using National Accounting data.

wages of other effects other than education. These "cleaned regressions" (estimation of equation (3.8)) are shown in Table 3.7. Dummies for each region and year are included, as well as the ones included in the previous Mincer regressions. The variables included are highly significant as well as most of the regional and year dummies.

Table 3.7: Clean Regressions

	year	**Regression 1 Primary or without studies**		**Regression 2 Secondary**		**Regression 3 college**	
Andalucia	90	0.677	***	0.781	***	0.789	***
	00	1.376	***	1.351	***	1.129	***
Aragón	80	0.150	***	0.136	***	0.013	
	90	0.914	***	0.902	***	0.699	***
	00	1.501	***	1.403	***	1.277	***
Asturias	80	0.194	***	0.089	*	0.104	*
	90	0.945	***	0.824	***	0.799	***
	00	1.581	***	1.356	***	1.283	***
Baleares	80	0.039		-0.004		0.322	***
	90	0.848	***	0.889	***	0.566	***
	00	1.513	***	1.466	***	1.142	***
Canarias	80	0.026		0.083	*	0.039	
	90	0.689	***	0.814	***	0.753	***
	00	1.373	***	1.206	***	1.228	***
Cantabria	80	0.149	***	0.034		-0.159	**
	90	0.899	***	0.722	***	0.962	***
	00	1.478	***	1.293	***	1.149	***
Castilla y León	80	0.107	***	0.027		-0.076	*
	90	0.849	***	0.829	***	0.796	***
	00	1.405	***	1.293	***	1.193	***
Castilla-La Mancha	80	0.028		0.073	*	-0.088	*
	90	0.796	***	0.834	***	0.798	***
	00	1.418	***	1.320	***	1.216	***
Cataluña	80	0.229	***	0.180	***	0.089	**
	90	0.925	***	0.884	***	0.949	***
	00	1.495	***	1.510	***	1.238	***

Continued on next page

Table 3.7 –continued from previous page

	year	Regression 1 Primary or without studies		Regression 2 Secondary		Regression 3 college	
Comunidad Valenciana	80	0.122	***	0.077	*	0.076	
	90	0.817	***	0.749	***	0.699	***
	00	1.463	***	1.476	***	1.257	***
Extremadura	80	-0.199	***	0.090		-0.073	
	90	0.672	***	0.692	***	0.713	***
	00	1.314	***	1.229	***	1.166	***
Galicia	80	0.018		0.093	**	0.024	
	90	0.786	***	0.799	***	0.782	***
	00	1.367	***	1.346	***	1.180	***
Madrid	80	0.219	***	0.168	***	0.102	**
	90	0.892	***	0.873	***	0.782	***
	00	1.513	***	1.391	***	1.270	***
Murcia	80	0.058	*	-0.055		-0.151	*
	90	0.747	***	0.762	***	0.852	***
	00	1.362	***	1.281	***	1.093	***
Navarra	80	0.291	***	0.226	***	-0.099	
	90	1.041	***	0.962	***	0.900	***
	00	1.649	***	1.545	***	1.286	***
País Vasco	80	0.266	***	0.199	***	0.176	***
	90	0.992	***	0.875	***	0.884	***
	00	1.556	***	1.416	***	1.140	***
La Rioja	80	0.165	***	0.082		-0.119	
	90	0.817	***	0.596	***	0.777	***
	00	1.413	***	1.379	***	1.247	***
sex		-0.282	***	-0.173	***	-0.131	***
agriculture		-0.411	***	-0.330	***	0.026	
married		0.175	***	0.245	***	0.233	***
$expe < 20$		0.225	***	0.220	***	0.135	***
$30 < expe < 40$		0.279	***	0.273	***	0.182	***
$expe > 40$		0.334	***	0.321	***	0.155	***
Constant		12.715	***	13.005	***	13.457	***
Adjusted R-square		0,60		0,52		0,57	
n		15912		4468		3254	

*, **, *** Denotes estimates that are significantly different from 0 at the 10%, 5% or 1% levels.

Average estimated wages for each schooling level and region (the fixed effects of the "cleaned regressions") are used to compute weighted average wages for regions holding the composition of the labor market constant for the base year, according to equation (3.10).[9] For each region we have weighted average wages corresponding to the three cross-sections (1980,1990,2000). From these the difference in logarithms of weighted average wages (1980-1990 and 1990-2000) are computed and used as the dependent variable in a regression on our measure of the intensity of regional human capital. The results for this second stage are reported in Table 3.8. For this case results are reported considering the three crossections or looking at the difference between 1990 and 2000. As controls we use total employment (as a scale factor) and physical capital, only for the regressions that consider all three cross-sections, since otherwise total employment is highly collinear with our human capital intensity measure. Finally, for each case we present OLS and IV results, where we use the same instruments as in the first stage.

Our results show significant human capital externalities which are robust to the different methods used as well as the period definition, but the controls are not significant. For the IV estimations the overidentifying hypothesis is rejected. Estimated externalities range from 0.143 to 0.262. The coefficients are still significant at a 5% level if we consider the period as whole or at

[9]That is, for region r, schooling group s and years $t0$, $t1$ and $t2$, average wages by schooling group, year and region are $w_{sr,t0}$, $w_{sr,t1}$ and $w_{sr,t2}$. If the chosen base year is in central period, we have that the growth of average wages holding constant the composition of the labor market for each region r is defined as

$$\Delta w^*_{r,t1} = \sum_{s=1}^{3} l_{01,sr}(w_{r,t1} - w_{r,t0})$$

$$\Delta w^*_{r,t2} = \sum_{s=1}^{3} l_{12,sr}(w_{r,t1} - w_{r,t0})$$

where $l_{01,sr} = (l_{0,sr} + l_{1,sr})/2$ and $l_{12,sr} = (l_{1,sr} + l_{2,sr})/2$.

a 10% level if we use all three cross-sections, except for some of the OLS estimations. The estimated coefficients are smaller than for the M approach (0.264 and 0.174). If we assume a value of 0.174 for the externalities coefficient, the growth in wages that can be attributed to externalities would be approximately 0.45% annually, which represents a half of the total growth in real wages for the period, and 50% less than the estimation using the M approach.

3.6 Conclusions

Identifying human capital externalities is crucial both theoretically and empirically, both to give empirical support to growth models that consider human capital as a basic input in the production function and to analyze the relevance of private and public policy towards education.

Despite this relevance the empirical evidence is scarce, because of the lack of individual data to perform this type of analysis. In this paper Spanish data is used to identify human capital externalities between 1980 and 2000. The Spanish economy is an interesting case because of the considerable educational changes experimented by the workforce during this period.

Our results show clear evidence on the existence of externalities of human capital in Spain. Using the two most common used methodologies, which we call the Mincerian Approach and the Constant Composition Approach, we find that the social return to education is sizable. Our results show that during this specific period, in which there was an important increase in the endowment of human capital, approximately one third of the average annual growth in wages can be attributed to the externalities generated by the increment in the human capital stock.

Nevertheless there are still important limitations in this analysis and further research ahead to confirm these results. Particularly important is the availability of more detailed individual data and longer periods.

Table 3.8: Constant Composition Approach estimation

Dependent variable: log changes in weighted average wages (weights are for year 1990 for 1981-2001 estimation and middle year for 1981-1991-2001 estimation).

	1981-2001				1981-1991-2001			
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Led	0.143**	0.174**	0.162**	0.180**	0.154*	0.264*	0.158	0.262*
	(0.057)	(0.071)	(0.072)	(0.085)	(0.080)	(0.145)	(0.102)	(0.150)
Employment	-	-	0.090	0.030	-	-	0.011	-0.013
	-	-	(0.178)	(0.170)	-	-	(0.138)	(0.129)
Capital	-	-	0.012	0.004	-	-	0.057	0.054
	-	-	(0.113)	(0.115)	-	-	(0.159)	(0.157)
Time	-	-	-	-	-0.085**	-0.040	-0.085*	-0.042
	-	-	-	-	(0.040)	(0.062)	(0.045)	(0.064)
Constant	1.136***	1.104***	1.105***	1.094***	0.601***	0.521***	0.590***	0.518***
	(0.060)	(0.075)	(0.103)	(0.113)	(0.065)	(0.105)	(0.089)	(0.121)
Adj. R^2	0.24	0.24	0.15	0.13	0.67	0.65	0.65	0.66
n	17	17	17	17	34	34	34	34

Data in parenthesis are standard deviations.
Led. Log change of college vs total workers ratio.
Employment: Log change of regional employment.
Capital: Log changes of regional physical capital.
Time: dummy variable which is zero for the first decade change and one for the second.
*, **, *** Denotes estimates that are significantly different from 0 at the 10%, 5% or 1% levels.

Appendix A

The EPF 80-81/90-91 and ECPFs 1985-2004

Although these sources do not provide homogeneous wage statistic series, they give important information that is relevant to this kind of analysis, for which they have become the main current source of statistics. However, a number of problems need to be considered.

First, the EPFs from 1980-81 and 1990-91 provide a broad spectrum of data for about 20,000 families. However, the quarterly surveys use a smaller sample size, since their main objective is to offer a short-term analysis of consumption, rather than consumption structure. At any rate, since 1997 the sample has doubled. This problem may be resolved by using two-year samples to improve sample size, since the ECPFs poll the same family for six quarters, changing one sixth of this sample each quarter.

Second, wages are not immediately determined by recorded earnings data. The main problem is that there is no information on hours worked or similar criteria unless the head of family has worked for more than thirteen hours during the reference week. The only solution to this problem is to use only those workers who reported working more than 13 hours and to assume that they worked full time. This assumption will no doubt introduce some measurement errors.

Third, despite the vast amount of information available, complete information is only available for the head of family.[1] Therefore this paper, like other Spanish studies (Abadie 1997), works explicitly with this selection.

The fourth problem arises from the use of two similar but somewhat different sources, the EPF and the ECPF. Key differences between these sources center on the amount of information, number of characteristics reported, the richness of the classifications within each characteristic (for example, with regard to education) and others. Any contrast between these two sources must therefore derive from the effect of these differences.

The last problem stems from the heterogeneity of the definitions and the classifications of variables used. For example, different educational level classifications appear each year, partly because current surveys have modified their definitions over the years and partly because the Spanish legal definition of education changed during the period under study.[2] Table A.1 shows the different educational groups. To solve this problem, we used years of schooling instead of educational level, because the former is homogeneous across all years and classifications. The years imputed are the same as those given in Vila and Mora (1998).

To conclude, despite the limitations of working with these surveys, they possess many redeeming features which make them our best choice for data on the distribution of Spanish wage inequality between 1980 and 2000. To summarize, two main criteria may be cited in support of their use: first, the

[1]For instance, the ECPFs records contain information about education only for this group

[2]In the early nineties the education law changed from an earlier one passed in 1970, which introduced compulsory schooling in Spain up to the age of fourteen. In 1990, the LOGSE extended compulsory schooling to the age of sixteen and changed the grades as the 1970 law had.

Table A.1: Schooling groups. Individual Data

	epf 80/81	epf 90/91	ecpf 85/95	ecpf 00/01
Illetary-Primary	1-3	1-4	1-4	1-2
Low Secondary	4	5 & 7	4	3
Acad. & Voc. Upper Sec.	5-6	6 & 8	5	4-5
Higher Educ. (Short cycle)	7	9	6	6
Higher Educ. (Long cycle)	8	10	7	7

lack of no better alternatives, second, their usefulness as a basis for comparative analysis. Once the surveys were selected for each year, they were then refined. Only data for heads-of-family working over thirteen hours per week were considered, after eliminating all self-employed wage-earners. Because ECPF surveys are restricted to household income data, this study focused exclusively on households where the head of family is the only worker. Second, groups of workers with unrepresentative characteristics were eliminated. For instance, education and experience cohorts were defined by five-year segments, and cohorts with fewer than fifteen records were deleted. Third, it was assumed that wages reported as below the legal minimum were either earned by part-time workers or individuals who had been unemployed for at least one year, or represented erroneous replies. Thus, the records were modified when annual reported wages fell below the legal minimum for that year. In this case, a zero was imposed for censored analysis. Table A.2 shows the minimum wage value in 1980 constant prices, which will be the censor value and the percentage of records modified to zero. Fourth, calendar effects are taken into account in the ECPF information. The quarterly nature of these surveys implies that the wages reported might be influenced by the quarter in which they are given. To eliminate this effect, families with wages for all six quarters were taken first, and the calendar effect was analyzed in these cases. Then a wage-level factor was obtained for each quarter. Once these factors were obtained, all the workers' wages were "deflated". Finally, the average wage was taken and multiplied by four, to give the yearly wage for all of the workers, regardless of which quarter they had worked. In this case, the six quarter samples were combined to increase (double) the ECPF sample size, although the same result could probably be obtained using all four quarters in the calendar context.

Table A.2: Censored data

	EPF		ECPF		
	1,980	1,990	1,985	1,990	1,995
Censored wage	313,148	287,565	292,280	287,565	279,055
Percentage (%)	9.44	3.16	8.18	3.83	4.47

Appendix B

Heteroscedasticity and Quantile Regression

This Appendix describes the expression (1.4) that captures within-group inequality. The key is that quantile prices incorporate the heteroscedasticity as they change through the percentile estimates that they produce.

For the case in hand, different coefficients were estimated for different $\tau \in (0, 1)$, which differ from each other whenever conditional wage dispersion depends on the covariate values. In others words, it has been proved that quantile regressions serve to analyze the heteroscedasticity in errors (Koenker and Bassett 1982). In this case, the quantile coefficients show changes in the percentiles. More specifically, suppose that wage equation

$$w_{it} = \beta_t x_{it} + e_{it}$$

is a so-called location-scale model, where $e_{it} = \sigma(x_t)\varepsilon_{it}$, $\sigma(x_t)$ is some function of x_t and ε_{it} is a normal iid error term with continuous and positive density $f(\varepsilon_t)$ and a distribution given by $F(\varepsilon_t)$. In this case, the conditional quantile for year t is given by

$$Q_\tau(w_t|x_t) = x_t'\beta_t + \sigma(x_t)F_{\varepsilon_t}^{-1}(\tau).$$

Suppose, for the sake of simplicity, that $\sigma(x_t) = x_t$. Then for the case of (1.2) the vector of coefficients estimated is given by

$$\beta_t(\tau) = \beta_t + F_{\varepsilon_t}^{-1}(\tau).$$

Here, $F_{\varepsilon_t}^{-1}(\tau)$ is the inverse of the cumulative distribution function of ε and represents the value of the percentile τ. A test of heteroscedasticity would therefore be to check the null hypothesis $\beta_t(\tau) = \beta_t(\theta)$ for $\tau \neq \theta$. This does not imply that estimations are not consistent. If $\sigma(x_t)$ cannot be fitted to a linear expression, the intuition will be the same, but $\beta_t(\tau)$ will require a more complex expression.

To clarify matters, let us suppose that the quantile model is given by (1.2). Then, it is easy to see that

$$\beta_t(\tau) = \beta_t + \Psi(x_t; F_{\varepsilon_t}^{-1}(\tau));$$

which is a generalization of (B). In this case, the within-group coefficients are

$$\beta_t^w(\tau) = \beta_t(\tau) - \beta_t(0.5) = \Psi(x_t; F_{\varepsilon_t}^{-1}(\tau)) - \Psi(x_t; F_{\varepsilon_t}^{-1}(0.5)).$$

This expression implies that x_t induces the heteroscedasticity in the $\beta_t^w(\tau)$ estimation. But that is another story, because the within-group estimated counterfactual density changes $Q_\tau(\beta_1^b, \beta_1^w(\tau); g(x;0)) - Q_\tau(\beta_1^b, \beta_0^w(\tau); g(x;0)))$ use a constant weight, $g(x;0)$, applied to each conditional density. So, residual counterfactual densities use the same weighted rule over time, and composition effects are correctly removed from RWI.

Bibliography

ABADIE, A. (1997): "Changes in Spanish labor income structure during the 1980's: a quantile regression aproach," *Investigaciones Economicas*, 21(2), 253–272.

ABRAHAM, K. G., AND S. N. HOUSEMAN (1993): "Earnings Inequality in Germany," Staff Working Papers 94-24, W.E. Upjohn Institute for Employment Research.

ACEMOGLU, D. (2002): "Technical Change, Inequality, and the Labor Market," *Journal of Economic Literature*, 40(1), 7–72.

——— (2003): "Cross-Country Inequality Trends," *Economic Journal*, 113(485), 121–149.

ACEMOGLU, D., AND J. ANGRIST (2001): "How Large are the Social Returns to Education: Evidence from Compulsory Schooling Laws," in *NBER Macroeconomic Annual 2000*, ed. by B. Bernanke, and K. Rogoff, pp. 9–59. Cambridge, MA: The MIT Press.

ALBA, R., AND M. S. SEGUNDO (1995): "The Return to Education in Spain," *Economics of Education Review*, 14(2).

ALCALÁ, F., AND P. J. HERNÁNDEZ (2005): "Las externalidades del capital Humano en las Empresas Españolas," *Revista de Economía Aplicada*, XIV, 61–83.

ANGRIST, J. (1995): "The Economic Returns to Schooling in the West Bank and Gaza Strip," *American Economic Review*, 85(5), 1065–1087.

ARELLANO, M., S. BENTOLILA, AND O. BOVER (2001): "The Distribution of Earnings in Spain During the 1980s: The Effects of Skill, Unemployment and Union Power," CEPR Discussion Papers 2770, C.E.P.R. Discussion Papers.

ARRAZOLA, M., J. D. HEVIA, M. RISUEÑO, AND J. F. SANZ (2003): "Returns to education in Spain: Some evidence on the endogeneity of schooling," *Education Economics*, 11(3), 293–304.

AUTOR, D. H., L. F. KATZ, AND M. S. KEARNEY (2005a): "Rising Wage Inequality: The Role of Composition and Prices," Harvard Institute of Economic Research Working Papers 2096, Harvard - Institute of Economic Research.

——— (2005b): "Trends in U.S. Wage Inequality: Re-Assessing the Revisionists," NBER Working Papers 11627, National Bureau of Economic Research, Inc.

AZARIADIS, C., AND A. DRAZEN (1990): "Threshold Externalities in Economic Development," *The Quarterly Journal of Economics*, 105(2), 501–26.

BARCEINAS, F., J. OLIVER, J. RAYMOND, AND J. ROIG (2000): "Los Rendimientos de la Educación en España," *Papeles de Economía Española*, (86).

BECKER, G. (1964): *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education.* Columbia University Press, New York.

BENABOU, R. (1996): "Equity and Efficiency in Human Capital Investment: The Local Connection," *Review of Economic Studies*, 63(2), 237–64.

BENTOLILA, S., AND G. BERTOLA (1990): "Firing Costs and Labour Demand: How Bad Is Eurosclerosis?," *Review of Economic Studies*, 57(3), 381–402.

BLACK, D., AND V. HENDERSON (1999): "A Theory of Urban Growth," *Journal of Political Economy*, 107(2), 252–284.

BLINDER, A. (1973): "Wage Discrimination: Reduced Forms and Structural Estimation," *Journal of Human Resources*, 8(4), 436–455.

CARD, D. (1999): "The causal effect of education on earnings," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card.

CARD, D., F. KRAMARZ, AND T. LEMIEUX (1999): "Changes in The Relative Strustures of Wages and Employment: A Comparision of the United States, Canada and France," *The Canadian Journal of Economics*, 32(4), 843–877.

CARD, D., AND T. LEMIEUX (2001): "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis," *Quarterly Journal of Economics*, 2(116), 705–746.

CASELLI, F., AND W. J. COLEMAN (2000): "The World Technology Frontier," Working Paper 7904, National Bureau of Economic Research.

CASTRO, R., AND D. COEN-PIRANI (2005): "Why Have Aggregate Skilled Hours Become So Cyclical Since the Mid-1980's?," Cahiers de recherche 2005-19, Universite de Montreal, Departement de sciences economiques.

CICCONE, A., AND G. PERI (2005): "Long-Run Substitutability Between More and Less Educated Workers: Evidence from U.S. States, 1950-1990," *The Review of Economics and Statistics*, 87(4), 652–663.

——— (2006): “Identifying Human-Capital Externalities: Theory with Applications,” *Review of Economic Studies*, 73(2), 381–412.

Conley, T., F. Flyer, and G. Tsiang (2003): “Spillovers from Local Market Human Capital and the Spatial Distribution of Productivity in Malaysia,” *Advances in Economic Analysis & Policy*, 3(1), 1229–1229.

De la Fuente, A. (2003): “Human Capital in a Global Knowledge-based Economy, part II: assesment at the EU Country Level,” European Commission Report, Employment and Social Affairs.

De la Fuente, A., R. Domenech, and J. F. Jimeno (2003): “Human Capital as a Factor of Growth and Employment at the Regional Level. The Case of Spain,” UFAE and IAE Working Papers 610.04.

Dimelis, S., and L. Alexandra (1999): “Inequality and Business Cycles in the U.S. and European Union Countries,” *International Advances in Economic Research*, 5(3), 321–338.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996): “Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach,” *Econometrica*, 64(5), 1001–1044.

Dolado, J., F. Felgueroso, and J. Jimeno (1997): “The Effects of Minimum Bargained Wages on Earnings: Evidence from Spain,” *European Economic Review*, 41(4-6).

Epanechnikov, V. A. (1969): “Non-Parametric Estimation of a Multivariate Probability Density,” *Theory of Probability and its Applications*, 14(1), 153–158.

Fallon, P., and P. Layard (1975): “Capital-Skill Complementarity, Income Distribution and Output Accounting,” *Journal of Political Economy*, 83(2).

Freeman, R., and L. F. Katz (1995): *Differences and Changes in Wage Structures*chap. Introduction and Summary. University of Chicago Press, Chicago, freeman, richard and lawernce f. katz edn.

Gottschalk, P., and M. Joyce (1998): "Cross-National Differences In The Rise In Earnings Inequality: Market And Institutional Factors," *The Review of Economics and Statistics*, 80(4), 489–502.

Gottschalk, P., and T. M. Smeeding (1997): "Cross-National Comparasions of Earnings and Income Inequality," *Journal of Economic Literature*, 35(2), 633–687.

Hamermesh, D. (1993): *Labor Demand*. Princetown University Press.

Hidalgo, M. A., J. O'kean, and J. Rodríguez (2008): "Demanda de Trabajo y Tecnologías de la Información: Evidencia para España," Mimeograph. Universidad Pablo de Olavide.

Izquierdo, M., and A. Lacuesta (2007): "Wage inequality in Spain - recent developments," Working Paper Series 781, European Central Bank.

J.E., R. (1993): "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities," *Journal of Urban Economics*, 34(3), 380–400.

Johnson, G. E. (1970): "The Demand for Labor by Educational Category," *Southern Economic Journal*, (37), 190–203.

Juhn, C., K. M. Murphy, and B. Pierce (1993): "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy*, 101(3), 410–442.

Katz, L., G. Loveman, and D. Blanchflower (1995): "A Comparison of Changes in the Structure of Wages in Four OECD Countries," in *Differences and Changes in Wage Structures*, ed. by R. Freeman, and L. F. Katz. University of Chicago Press and NBER, Chicago.

Katz, L. F., and K. M. Murphy (1992): "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, 107(1), 35–78.

Keane, M., and E. Prasad (1993): "Skills Levels and the Cyclical Variability of Employment, Hours and Wages," Discussion Paper 4, IMF Staff Papers 40.

Koenker, R. W., and J. Bassett, Gilbert (1978): "Regression Quantiles," *Econometrica*, 46(1), 33–50.

——— (1982): "Robust Tests for Heteroscedasticity Based on Regression Quantiles," *Econometrica*, 50(1), 43–61.

Krusell, P., L. Ohanian, V. Rios-Rull, and G. Violante (2000): "Capital Skill Complementary and Inequality: A Macroeconomic Analysis," *Econometrica*, 68(5), 1029–53.

Lemieux, T. (2002): "Decomposing changes in wage distributions: a unified approach," *Canadian Journal of Economics*, 35(4), 646–688.

Lemieux, T. (2006): "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?," *American Economic Review*, 96(3), 461–498.

Lindquist, M. J. (2004): "Capital-Skill Complementarity and Inequality Over the Business Cycle," *Review of Economic Dynamics*, 7(3), 519–540.

LUCAS, R. J. (1988): “On the mechanics of economic development,” *Journal of Monetary Economics*, 22(1), 3–42.

MARTINEZ-ROS, E. (2001): “Wages and Innovations in Spanish Manufacturing Firms,” *Applied Economics*, 33(1), 81–89.

MAS, M., F. PÉREZ, AND E. URIEL (2006): *El stock y los servicios del capital en españa y su distribución territorial (1964-2003)*. Fundación BBVA.

MAS, M., F. PÉREZ, E. URIEL, L. SERRANO, AND A. SOLER (2002): “Metodología para la Estimación de las Series de Capital Humano. 1964-2001,” .

MATA, J., AND J. A. F. MACHADO (2005): “Counterfactual decomposition of changes in wage distributions using quantile regression,” *Journal of Applied Econometrics*, 20(4), 445–465.

MINCER, J. (1974): “Scholling, Experience and Earnings,” Discussion paper, National Bureau of Research, New York.

MORETTI, E. (1998): “Social returns to education and human capital externalities: evidence from cities,” Working Paper 9, Center for Labor Economics, UC Berkeley.

——— (2003): “Human Capital Externalities in Cities,” NBER Working Papers 9641, National Bureau of Economic Research, Inc.

——— (2004a): “Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data,” *Journal of Econometrics*, 121(1-2), 175–212.

——— (2004b): “Workers’ Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions,” *American Economic Review*, 94(3), 656–690.

Oaxaca, R. (1973): "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 14(3), 693–709.

Oliver, J., J. Raymond, J. Roig, and F. Barceinas (1999): "Returns to Human Capital in Spain: A Survey of the Evidence," in *Returns to human capital in Europe: A literature review*, chap. 13, pp. 280–298. R. Asplund and P.Pereira, Helsinki.

Olivier, J., X. Ramos, and J. Raymond (2001): "Capital humano y desigualdad en España 1985-1996," *Papeles de Economía Española*, (88), 240–256.

Peraita, C. (2003): "The Importance of Institutions and Skill-Based Technical Change in Wage Decompression: Evidence for Spain," V Jornadas de Economía Laboral, Reus.

Pourpourides, P. M. (2007): "Implicit Contracts, Wages and Wage Inequality over the Business Cycle," Cardiff Economics Working Papers E2007/19, Cardiff University, Cardiff Business School, Economics Section.

Raymond, J. (2002): "Convergencia real de las regiones españolas y Capital Humano," *Papeles de Economía Española*, 93, 109–121.

Reder, M. (1955): "The Theory of Occupational Wage Di®erentials," *American Economic Review*, 45(5), 833–852.

Rudd, J. (2000): "Empirical evidence on human capital spillovers," Finance and Economics Discussion Series 2000-46, Board of Governors of the Federal Reserve System (U.S.).

Skaksen, J., and A. Sorensen (2005): "Capital-Skill Complementarity and Rigid Relative Wages: Inference from the Business Cycle," *Contributions to Macroeconomics*, 5(1), 1145–1145.

TORRES, V. X. T. (2002): “Dispersión salarial y cambio tecnológico en la industria española,” *Investigaciones Economicas*, 26(3), 551–571.

VILA, L., AND J. MORA (1998): “Changing Returns to Education in Spain During the 1980s,” *Economics of Education Review*, 17(2), 173–178.

YOUNG, R. (2003): “The Wage Premium: A Puzzle,” Florida State University.