LUND UNIVERSITY
School of Economics and Management

Master's Programme in Economic Development and Growth

Exploring the Occupational Mismatch in Spain (1990s-2013)

by

Akshaya Sharma

akshaya.sharma.629@lu.se

[This research aims to give a thorough look at the incidence, and evolution of occupational mismatch in the Spanish labor market from the 1990s till 2013. This is done by examining which parts of the labor market have been exhibiting this feature, and what the odds are of being mismatched, given individual characteristics. Two, large micro-data surveys were used; the Wage Structure Survey (WSS) and the Economically Active Population Survey (EAPS). Additionally, information from the WSS was used to analyze the effect of over/undereducation on earnings. The results indicate that the Spanish labor market suffers from mismatch in its lowest and highest educated individuals (polarization), and that overeducation appears to be more of a permanent than transient phenomenon.]

Key words: occupational mismatch, OCR, overeducation, Great Recession, Spain

EKHM51
Master's Thesis (15 credits ECTS)
June 2015
Supervisor: Anders Nilsson
Examiner
Word Count:12,378

Table of Contents

1 Introduction 2
1.1 Research Problem 2
1.1.1 Relevance of Occupational Mismatch 2
1.1.2 Interest in Spanish Labor Market 3
1.2 Aim and Scope 4
1.3 Outline of the Thesis 7
2 Relevant Changes in the Labor Market and Theory 8
2.1 Relevant history and changes within the Spanish Labor Market 8
2.2 Signalling and Screening 11
2.2 Occupational Mismatch Theory 14
2.1.2 Measures 14
2.1.3 Drawbacks of Measures 15
2.1.3 Consensus on findings in Occupational Mismatch 16
2.2 The Role of Binary Outcomes in Occupational Mismatch 18
3 Data 20
4 Methods 23
5 Emperical Analysis 29
6 Conclusion 45
References 47
Appendix A 51
Appendix B 52
Appendix C 54
Appendix B 61

Introduction

The increase in the educational level of populations has been quite notable amongst several nations, especially amongst *Organization for Economic Co-operation and Development* (OECD) members during the past decades. It has been found that 65% of the population aged 25 to 34 years in 1992 had attained upper secondary education, while this number rose to 83% in 2014 (OECD, 2015). Tertiary education has also seen a high increase in graduates. While economic policies encourage educational expansion in order to capture the private and social benefits of higher worker productivity, or more equitable resource distribution; inefficient allocation of this newly acquired human capital can cause a mismatch between the supply and demand of labor. (Known as occupational or skill mismatch) Given the context of the 2008 recession where unemployment rates rose in several countries, while at the same time firms are experiencing changes in technological requirements, the aforementioned misallocation could be quite expected. It is thus relevant to study the incidence and evolution of this mismatch, and determine how it is affecting the labor market, while human capital investment continues to expand further.

1.1 Research Problem

As previously implied, it must be primarily addressed that simply providing many individuals with higher skills does not ensure an efficient use of resources, nor optimal levels of productivity for firms (OECD, 2015). In order for rising enrolment levels to have the desired effect, the appropriate quantity of demand for high skilled labor (opportunities in the labor market) must be present to match the supply, and efficiently allocate the skills being acquired from the educational process.

1.1.1 Relevance of Occupational Mismatch

There are several factors that could cause an occupational-skill mismatch, which itself could occur in two forms. A common event is that the supply of highly educated workers may grow faster than their demand, resulting in a mismatch between the skills of the worker and the ones required in their occupation. This form of skill mismatch could be referred to as

‘overeducation’. The same way, ‘undereducation’ could be a problem caused by the same forces working in the reverse direction, such as an increase in school/university dropout rates.

The presence of a mismatch between workers’ skills and their occupations has been shown to have negative effects on individual workers, their respective firms, and consequently the economy. This is especially true if the mismatch happens to lean more towards being a permanent phenomenon than a transient one. (Tsang & Levin (1985); García & Malo, 1996). An economy which has a large presence of ‘overeducation’, as found in literature, has a negative impact on wages, in the way that overeducated workers earn less than their colleagues that are ‘appropriately matched’ for their occupation, with the same amount of human capital investment. (Budría & Moro-Egido, 2008; Alba-Ramírez; and 1993García-Montalvo & Peiró 2009) Consequently, a decrease in wage returns to education, for a specific level/degree of education suggests a problem of overeducation (Rumberger, 1981) Eventually, firms have negative influences on their productivity and competitiveness from overeducated workers. (Tsang, Russell, Rumberger, & Levin (1991). In regards to the other end of the spectrum, the literature has also seen mismatch from undereducation causing workers to receive lesser wages than those that are appropriately educated for the job.

Thus, the problem that will be studied here is whether there occur such mismatches between one’s education, and the education required for their occupation. It will further be analyzed which aspects of the Spanish labor force are more severely affected by a mismatch, and what effect it has on an individual’s wages. Lastly, an attempt will be made to look at this with a pre and post recessionary lens; and give the reader an idea of how certain variables affect the possibility of being misallocated in the labor market after 2008.

1.1.2 Interest in Spanish Labor Market

The analysis in this investigation focuses on the Spanish labor market, with the context of the 2008 recession in mind. This is due to several reasons that make it a unique case study, especially in comparison to its OECD counterparts. Spain is a country that has witnessed large changes in its educational system during the past three decades. It has risen from less than 7 million students in the educational system in 1970, to almost 11 million in 2015 (Ministerio de Educación, 2016). When compared to other nations, 35% of adults had obtained tertiary education in 2015, while the average for the OECD was 33% and 32% for

the EU-21. These figures show a rising educated population, and relatively higher proportion of tertiary educated individuals. Higher educated populations in a country can at many times be one of the main factors in development strategies (Minguez, 2013). However, the information thus far does not indicate whether the labor market is able to allocate these individuals with tertiary education into suitably matched occupations for their skills.

Thus, when looking the supply side of the labor market, Congregado et. al (2014) states that Spain was creating more jobs than most of its counterparts prior to the 2008 recession, but was later losing more jobs than most of its counterparts in the post-recessionary period. Its unemployment rate was 17.9% in 2009, the highest in the Euro area. Even for a crisis, such a fluctuation from one end of the spectrum to another suggests that its labor market allocation mechanisms are not the most efficient. Features of the demand and supply side of the labor market as the ones mentioned so far, suggest that the incidence of mismatch could be interesting to look at.

When studying unemployment and mismatch, it is also essential to focus on the youth that is currently entering the labor market, or has recently entered it. Youth employment rates in Spain have been comparatively the lowest in the EU region. Given this, it is also known that during the post crisis period, enrolment rates in Spain have were noted to rise. This was especially for the age group of 20-24, much higher than its OECD counterparts (Dolado, 2013). However, there is a large share of youth that is not receiving high levels of education. 35% is the reported rate, as compared to the OECD average of 20% for the number of individuals who do not complete their upper secondary level of education (Minguez, 2013). At the same time, there is a high proportion of youth that are "not in employment, education or training" (NEET), which can potentially signify structural problems faced during the transition from education to the labor force The NEET rates in Spain were one of the highest in the European Union even prior to the Great Recession. Moreno, (2013) states that relatively high rates of individuals not completing secondary education are not new to Spain. The share of NEET slightly declined post 2008, but as mentioned, is still one of the highest in the region. Thus, when looking at the educational levels of the population (especially the youth), it appears to be quite polarized. This is highly relevant to the discussion as it can be considered an indicator of how well the education system is able to enhance the employability of the youth, and prevent an occupational mismatch if they enter the labor market (OECD,

2015). Given the context of the Great Recession, it is curious whether such a polarization if present, has been even worsened by 2015.

Spain has other aspects related to the misallocation of human capital that could be interesting to include in the analysis. Congregado (2014) has also shown this in studies involving gender differences in the Spanish labor market. He shows that female participation rates in the labor force rose steadily from the 1970s into this decade. On the other hand, male participation rate did not look the same; it declined until around the mid 1990s. However at the same time, it must be noted that the female participation rate is still lower in Spain than in most EU nations. Prieto and Rodriguez (2000) find that females' labor participation is dependent on the husband's labor status (if married), along with potential declines in fertility. The current literature does not provide much information on this, and provokes more questions to the discussion of mismatch in Spain.

The incidence of temporary unemployment in Spain is another factor that has been historically linked to occupational mismatch, and to wage skill premium inequality. Florentino (2010) finds that a higher incidence of temporary contracts in the labor force (especially youth), creates a more instable environment for them, while transitioning from education into the labor market. He states that the incidence of temporary employment has increased amongst the population aged 15 to 24 years, and that it is related to youth unemployment and dropping out of the labor force. 46% of jobs for 20 to 29 year olds have temporary contracts. This could have several implications in skill mismatch. Literature such a Toharia (1993) find that temporary employment also pays lower wage returns for the same job in Spain, than on a permanent contract. Davia (2002) also finds a significant wage gap between the two. Moreno (2013) adds that the youth at times do not have a choice but simply have to enter this instable work environment, with high wage differentials. This provokes the question of whether temporary unemployment is indeed related in Spain to the issue of mismatch. As can be seen, Spain appears to have certain paradoxical conditions concerning its labor supply and demand that provoke discussion over the allocation of its human capital.

1.2 Aim and Scope

Thus, the following investigation will report on the incidence of occupational-skill mismatch in Spain, and its post-recessionary outlook. It will contribute an analysis using ORU regressions on the effect on secondary and tertiary wage premiums, along with any potential differences between genders. Further on, the influence of these changes will be expanded upon in terms of educational levels attained, and factors such as temporary employment. Finally, an analysis will be done to see the odds of being occupationaly mismatched in the Spanish labor market, given certain conditions. The investigation will be provoked by these operational questions:

1) How has the incidence of skill mismatch (both overeducation and undereducation) changed in the past 20 years? (Including pre and post recessionary differences)

2) Where in the population is it more likely to occur? (educational levels, age groups, occupational sectors)

3) What is the relation between wages and occupational mismatch? Is this relevant to educational levels attained?

4) Are there gender differences in occupational mismatch, and different effects on wages?

5) Is the feature of being employed on a temporary contract relevant to occupational mismatch?

This way, the investigation aims to provide a much more detailed picture than present in the current literature about where (within the Spanish labor market) the occupational mismatch has been occurring, and what its recent evolution has been after the Great Recession. The mixture of regression analysis relating mismatch to wages, and the change in probabilities of being mismatched over time attempt to give the reader a very deep insight into this issue.

1.3 Outline of the Thesis

The rest of this investigation is presented as follows; section 2 will describe the history and conditions in the Spanish labor market that are relevant to occupational mismatch, and what changes have defined its characteristics today. It will also discuss the importance of signaling and screening processes in the Spanish labor market, which are associated with occupational mismatch. Finally, binary outcome models will be discussed in terms of their relevance to this analysis and how they can enhance the mismatch literature. Section 3 will describe the data sources used, including their strengths and shortcomings. Section 4 will describe the exact methodology being used, including the limitations of its usage in this study. Later, in Section 5, the empirical analysis conducted will visualized, after which it will be discussed. Finally, section 6 will connect the findings with the aims and objectives of this investigation, concluding with the implications of such research and how it may be implemented in the future.

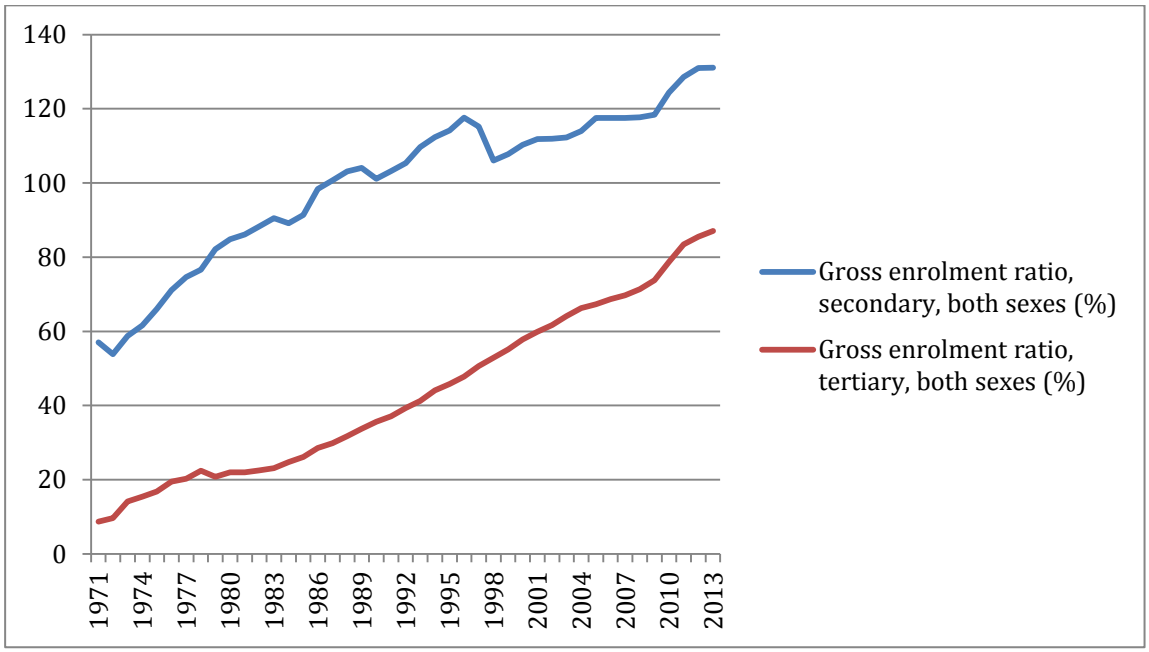
2 Relevant Changes in the Labor Market and Theory

This section aims to provide a brief understanding of the history of the Spanish labor market, and the changes that occurred from the 1970s, which shape its current features. It also describes how signaling and screening processes interact with each other, and potentially aggravate situations of mismatch. Lastly, literature on occupational mismatch theory and choice models are presented in order to indicate the motivations for choosing the methodology in this investigation, including its benefits and drawbacks.

2.1 Relevant History and Changes within the Spanish Labor Market

It is key to first address the discussion of why Spain has the labor market trends that it does today, and why it might have problems allocating human capital efficiently in the labor market. The first, is to understand the rise of educational enrolment. There are several reasons throughout the past decades for the rise in the education enrolment, out of which a few have been persistent. A number of these underlying factors have been outlined in Schofer (2005) as global trends in economic development, democratization, scientization, and decentralized systems, which began in the 1950s. These trends were seen to influence, and finally take place in Spain during the 1970s. Figure 1 below portrays the evolution of enrolment, regardless of age as a percentage of the population of official secondary/tertiary education age.

Figure 1. Gross enrolment ratios, secondary and tertiary (%)



Source: World Bank Databank

The percentage thus can surpass 100% due to early or late school entrances or even due to repeating a level. However, the objective is to simply show the change over time in the enrolment ratios. One can see that both ratios have been steadily rising without fail from 1971 (previous data unavailable), and both see a small jump shortly after 2008.

The event that occurred in Spain during the 1970s, which had the biggest influence on its economic and educational outlook, (previously and indirectly has been referred to twice) was the end of its totalitarian dictatorship in 1975. This marked a shift from authoritarian rule, towards a democratically elected government, along with the help of the titular monarch. The new political environment also saw the end of highly centralized forms of governing, which has been found in the literature to be correlated with higher education enrolment rates (Schofer, 2005). Another global trend at the time that is correlated with higher educational enrolment rates is the process of 'scientization'. This refers to the wider acceptance of scientific research and thinking in order to solve daily problems, as well as at the firm and state level. It is not a surprise that this would make the idea of university a more popular one, contributing to a greater portion of society pursuing human capital investment. Consequently, the purpose of education evolved from one that simply created candidates to maintain a nation's elite class, towards one where human capital investment was a public benefit that would empower both individuals and society. The difference can be visualized in the fact that

university centers were previously only in large cities (where the wealthy lived) such as Madrid, Barcelona, Salamanca (Vila and Mora, 1998). Following this, in the period of democratization, university centers increased from 15 in 1976 to 23 in 1988 ((Gutiérrez and Ortega, 2007) Starting from this period onwards, students were less geographically and financially limited from pursuing tertiary education.

The end of the dictatorship was the most significant, but only one of the three major events that influenced the Spanish economy from that period onwards. The two other changes were Spain's entrance to the European Union in 1986 (Bentolila and Jimeno, 2003), and its deregulation process in order to enter the European Economic and Monetary Union (EMEU) in 1994. (Verheul et al., 2006). Labor reforms in relation to these three events caused changes in the form of more gender equality, rising unemployment, and employment legislation affecting the types of contracts being offered in the market. (Congregado et al., 2011)

Thus began a change from a state with autarky-styled policies (such as price settings on basic goods) towards one of free market policies with fewer regulations. This caused the first set of labor reforms to take place in the late 1970s, which influenced positive effects, and lacked in other areas. (Carrera et al., 2001). The first result was a much higher female labor force participation rate as compared to before. There was still however, very poor allocation of labor among firms. This was simply because of strict labor market regulations that did not allow enough autonomy for firms in the making of contracts yet, in the context of a newly competetive market. (García-Serrano et al., 1998; Toharia, 2003; Bentolila and Jimeno, 2003; Dolado and Jimeno, 1997).

The poor allocation of labor caused high unemployment rates through to the mid-1980s, which is when the Spanish government gradually started introducing further labor reforms. These essentially aimed to lower the unemployment rates with policies such as making it easier to form temporary employment contracts. The 1980 Workers' Statue made open-ended contracts as the default form of contract, as opposed to the permanent ones. (Davia, 2002). This allowed many individuals to reenter the labor force, and also for employers to manage their enterprises with greater ease. The details of such reforms can be seen in Bover et al. (2002), García-Pérez (1997), Toharia (2003), and Congregado and García-Pérez (2002).

Spain's entrance to the European Union in the 1980s seemed to have net benefit for its labor market. However, the prerequisites to this involved objectives of economic openness

and competitiveness that required the government to restructure its firms and sectors, involving an even further set of reforms. However, with the help of EU funding, Spain was able to succeed in actually becoming more open/competitive, and saw its unemployment rate drop in the early 1990s. Notwithstanding, it rose by 1994 again, influenced by the European recession, shortly after which, the national currency was abandoned in favor of the Euro. With the signing of the Maastricht Treaty came additional labor reforms such as unemployment cost reductions, direct tax and social security fee reductions. This enabled higher economic growth in Spain, lower unemployment, and contributed to the fact that Spain created more jobs than its European counterparts until the Great Recession (Congregado et al., 2011). Following the signing of the treaty, temporary employment was put through another set of reforms in order to reduce employee turnover, but ended in the opposite result (Toharia and Malu, 1999). Also, the share of temporary workers in the labor force stayed at a constant from 1990s onwards at approximately 30%. This is the highest amongst all of Spain's European counterparts. It has been suggested that product market conditions of the Spanish labor market present from the 1990s are part of the reason for no show in the decline in the share of temporary workers. Florentino (2010) states that this has an effect (negative) on the returns to work experience and tenure, when evaluating wage skill premiums. Wage setting institutions seem to favor temporary contracts as a result, and this lowers wages of workers recently joining the labor force. This depression of wages, coupled with unemployment rates also seem to exaggerate the signaling young workers have to do in order to pass screening and filtering processes, which will be discussed in the following section.

2.2 Signalling and Screening

Signalling and screening have been mentioned in this study but not defined as of yet. The origin of this is in classical human capital theory. Becker, (1962) states that individuals who have higher levels of general human capital are more efficient, and are usually more productive than the latter. However, screening theory does not view education with the same viewpoint. Screening theory simply sees education as a signalling device, which does not add to one's productivity, and one that individuals make use of when trying to obtain a job position. Employers therefore implement screening processes and filters to identify those potential employees with the desired abilities. (Weiss, 1995). This difference is easily

reflected in the tete-a-tete in a labor market with highly educated individuals where employers raise requirements to obtain information about individual innate abilities, while potential employees try to signal this with even higher educational attainment. Thus, (as is described in segmentation theory) it is not hard to see that signalling can quite easily lead to overeducation, and possibly exaggerate an already present occupational mismatch (Malo, 2004). Looking from the point of screening theory, additional screening could be seen as necessary in the context of an environment where enrolment rates keep steadily rising, and many individuals are signaling ”high abilities”. Even without such a context, it is usually viewed during recruitment that unfavorable employee characteristics can be associated with relatively low levels of education, which leads to higher educational levels (especially when compared to their cohort) becoming a key hiring criterion (Schofer 2005). Employers for higher qualified jobs thus tend to search for workers on the outer (educated) end of the educational spectrum, where one's ability may be implied by the years they have studied.

A standard example of this filtering and screening process for abilities is the requirement of university education. It differentiates those who were able to enter the tertiary educational system. While it is considered a filter in some studies, others look at it as a double filter, such as in the case of Arrow (1973). Here, the first filter involves obtaining admission to the university or institution, and the second filter is successfully completing the degree or program there. There could potentially be even more filters, especially in the first part, where the process of admission involves certification exams. This is the case in some universities in Spain, but many do not require them.

When concerning occupational mismatch and wages, it has been found in the literature that filtering and screening such as with universities has an effect on workers' earnings Weiss, (1995). It would not be strange to then relate changes in private returns with screening processes that become more complex and widespread due to the constantly increasing number of educated individuals. From the point of view of screening theory, this could be relevant in a situation where an individual successfully and falsely signals 'desirable' abilities with their educational attainment. This employee would be considered in theory to be more productive than in reality. The implications of this are that the individual will receive higher private returns to their education than someone who actually possesses that ability. (and works at a job that requires that level of skills/ability) At the same time, the benefits to society would diminish. Such scenarios are market inefficiencies.

Several authors have attempted to analyze private returns in order to see if screening and filtering produce efficient allocations of human capital. The skill premium for education in Spain and other European countries has been termed at around ’10 percent’ in Psacharopoulos, (2004). However, studies made on twins regarding the measurement error of ability recalculate that value at around 8 percent. It was thought that ability would have a greater role in the reduction by accounting for measurement error, but was indicated otherwise. Maybe this should caution employers from using educational attainment by itself as an indicator of potential employees’ abilities. Regarding the returns to society that get affected by the process of signaling, the literature does not provide much information or consensus on a given number in the past decades. This is merely because a suitable methodology of estimating it has not been formed yet. There are many problems starting with its measurement, and identifying returns themselves. The lack of information on changes in social benefits could thus be considered a missing part of the picture when estimating wage skill premiums in the labor market and analyzing a relation between its fluctuations and the negative effects on society from occupational mismatch. When discussing the rise of wage skill premiums, another relevant point amongst literature such as Psacharopoulos (2004) is the false signaling of abilities with the time spent earning certifications. It has be studied that individuals at times may pursue degrees for a given number of years while in reality, they could be earned sooner. The difference is known as ‘actual years of education versus efficient years of education. Miller (1995) states that signaling theory favors those workers that finished their education faster, as it could imply a greater level of ability in the respective field. However, human capital theory on the other hand seems to favor merely a higher value to form as part of the model, indicating a more positive effect on earnings. The extra time is seen to augment productivity and possibly gain a more thorough understanding of the skill/trade/field. This suggests caution to employers when interpreting an additional year of education as a positive signal of a potential employers abilities.

The question remains on how to properly allocate labor in accordance with educational requirements, and how appropriate decisions made in the labor market today based on years of education are. Some studies attempt at evaluating the results of screening processes and whether relations between wages and education show workers in “appropriately matched” occupations. Riley (1979) shows that these models were relatively correct for those occupations where screening is regarded as highly important. However, it was also concluded that the wage skill premiums were higher in occupations that did not highly regard or use

screening. This shows that while screening could be important and useful for some occupations, analyzing schooling decisions is another issue that is more complex. According to Psacharopoulos (1994), the relation between ability and years of education also rises at a lower rate in screened occupations than occupations that use little screening. Nevertheless, both imply that education, while not being a perfect measure for ability is the most cost effective one for employers as of date. It would be expensive to implement the next alternative, on-the-job filtering, due to the replacement costs of even a few inappropriately matched employees. However, increasing enrolment rates, as pointed by Arrow (1973), mean that inevitably, more and more individuals will successfully 'deceive' the screening and filtering processes.

As long as there are forces that continue exaggerating the processes of signaling and screening, potential overeducation and undereducation can occur, causing occupational-skill mismatches in the labor market. There is no consensus on cheaper or better alternatives to these processes in hiring practices for several occupations, as well as a lack of quantitative analysis on how social returns are actually being affected. The following section thus looks at how measures of overeducation and undereducation can be related to the aforementioned mismatch, given the changes in the labor market over the past years.

2.3 Occupational Mismatch Theory

As will be seen in this section, there are several ways of attempting to measure a mismatch between one's occupation and skills. However, the baseline remains that whenever an individual works in an occupation that requires a different level of skills than they have obtained through educational attainment, an occupational mismatch is said to be present.

2.3.1 Measures

When looking at the literature, there is no single accepted measure of overeducation or occupational mismatch. Five different types of measures are usually made use of, as stated by Groot et al. (2000) and Hartog (2000). Hartog (2000) provides a good discussion of these different measures, and shows that conclusions obtained from most measures do not differ by

a huge amount. However, there are a few some results such as by Groot et al. (2000) which suggest otherwise. In this case, the difference was 10 to 42%. All forms of measurement can be seen to have their drawbacks, and most authors select a methodology based on data availability.

The aforementioned measures can be divided amongst subjective and objective ones, all in which the relation between educational level and occupation is visualized. Subjective measures can be of two types. In both, the mismatch is calculated using information workers provide about themselves, their occupation, and whether they consider themselves underqualified/appropriately qualified/overqualified for the position. However, one of the subjective measures also cross-checks this information with their actual investment in education, and their skills.

On the other hand, objective measures are analyzed using data collected in relation to the jobs their characteristics, level of difficulty, and the training/experience actually needed to perform them (Murillo, 2008). Among them, one common methodology has been to estimate the mean value (of educational investment) of the distribution in each occupation, and categorize those individuals that are one standard deviation away from that value as undereducated or overeducated. This has been highly popularized by Verdugo and Verdugo (1989). Another measure, is to use the modal value as a reference for the appropriate years of education required for a given occupation, and calculate the mismatch as such. This measure was pioneered by Duncan and Hoffman (1981) and made more popular for use by Kiker et al. (1997). A third, but not widely used method due to data availability is to use systematic, expert job analyst evaluations of requirements for occupations, and the skills and abilities that workers in these occupations possess.

2.3.2 Drawbacks of Measures

There are drawbacks to all the measures mentioned. When looking at the subjective measures, their strengths lie in the fact that they provide up to date information at the individual level. However, a respondent to a survey is free to exaggerate as much as they want in regards to their qualifications and those required in their job. They might even simply read their firm’s manual on hiring regulations. The variation of answers may not be systematic at all.

Regarding objective methods using the mean value of educational investment as a reference, much literature such as Bauer (2002) criticize this since there is the reasoning behind the decision of using one standard deviation is quite arbitrary. Also, it is an estimation that will be sensitive to technological changes in the labor market and changes in workplace organization. It could potentially lead to incorrect conclusions when concerning undereducation. (Kiker, 1997) A slightly better alternative seems to be the one using the modal reference for education investment. With this methodology, the analysis and conclusions will be less sensitive to technological changes in the labor market and outliers. (Verdugo and Verdugo, 1989)) However, this definition focuses on the assignment and allocation of workers in reality, and the allocation is endogenous. It's strengths lie in using it to analyze market assignment, and not to analyze shifts in labor demand. If possible, it is recommended by Hartog (2000) to use data and evaluations by job analyst experts in estimating occupational mismatch. However, the work required to carry out such evaluations over time is usually quite expensive, and few countries have data that could be used to make a proper analysis over many years (Hartog, 2000) This is an issue since a lacking part of the picture in such analyses is usually that job requirements are not well defined in theory. As stated by (Groot et al., 2000), one could assume that different jobs ask for different training levels from potential employees. If one considers this to be true, then an employee's productivity would not depend mostly on their years of education, but there would also be a greater weight than usual on the characteristics of the job. This is backed by screening and competition theory. In such a scenario, wage skill premiums would show an even greater dependence on a skill match between the employee and the occupation needs. There are fields such as psychology that have substantial information on this, but have not been passed over to the economics field as of yet. Many aspects of specific job requirements such as abilities and aptitudes are not included when simply looking at the number of years invested in education.

2.3.3 Consensus on findings in occupational mismatch

The findings generally agreed upon in the occupational mismatch literature so far consist of the following:

1) The wage skill premium for appropriate years of schooling are higher than usually higher than a worker's actual investment in human capital. This is the result of comparing ORU regressions to Mincerian wage equations.

2) The wage skill premiums for studying more years than needed is usually positive, but not as high as the premium for studying the required amount for the job. It can range from 50% to 67% of the returns on required years of education.
3) The returns to studying fewer years than required for the occupation are usually negative. However, undereducation does not have as much of a negative effect on earnings than studying more years than required has a postitive effect. In several studies, the coefficient of undereducation remains almost constant.

These have been found in have noticed the following: (Duncan and Hoffman, (1981); Alba-Ramirez, (1993); Hartog and Tsang, (1987); Kiker and Santos, (1991); Hartog and Oosterbeek, (1988); Sicherman (1991); and Groot & Maassen van den Brink, (1995).

While some of the results of measuring occupational mismatch have seen a consensus, it must be commented that many results do not reach a consensus on explaining its incidence. This is both in recent and past literature. Some recent findings have been that overeducation could possibly be explained by lack of human capital assets including ability, experience, or on-the-job training. Groot et al., (2000) confirmed that employees that had higher than the required amount of education rank lower on such factors than those workers who have been 'appropriately matched' in their occupation. They noticed that several workers who are not appropriately matched have had events such as child rearing and other activities that have been related to this. Classical human capital theory by Becker, (1964) and Mincer, (1974) state that occupational mismatch is a temporary phenomenon and that marginal productivity is formed by education, experience and on-the-job training. Thus according to this, would be no relation between one's education being above the appropriate level for the job and their earnings.

Nevertheless, authors such as Duncan (1981) do see a relation between the two and have attempted at analyzing the relation between earnings and mismatch. Psacharapoulos (2004) investigates whether private returns to education have been rising or falling over a longer period of time. His conclusions are that private returns have been rising until recent years. Few studies have looked at the effect of the mismatch in the context of the 2008 recession. In the context of Spain, several authors studied the private returns to tertiary education and found a decreasing trend over a longer period (Vila & Mora, 1998). It appears in the literature

thus far that the evolution of returns to education is an informative path to analyze if one is not able to look at the various job characteristics that differ in each occupational category.

2.3.4 The role of binary outcome methodologies in occupational mismatch

This section simply seeks to briefly address the use of models such as binomial/multinomial Probit and Logit in the occupational mismatch literature. There is not a huge literature of binary outcome models related to occupational mismatch, but substantial enough to make some comments relevant to this investigation. Relevant examples are Kalfa et al., (2013), who studies the probability of mismatch when entering a first job in the labor market. Boll et al. (2016) does a dynamic mixed multinomial logit and probit analysis to study overeducation in East Germany, and Joona et al., (2014) do a study concerning overeducation and wage effects for immigrants to Sweden. The usefulness of binary outcomes in all such studies is to be able to conclude with the odds or probability that explanatory variables being observed can cause an influential outcome (unemployment, mismatch, overeducation)

Models used in the investigation by Kalfa et al., (2013) check how explanatory variables such as country of origin, having had a previous job, temporary contract, and years spent in the country determine the probability of occupational mismatch, and selection into unemployment. The findings showed that job experience and having a temporary contract were the strongest predictors of receiving a 0 in terms of not having occupational mismatch as the outcome.

The analysis using probit models for seeing how explanatory variables related to job characteristics such as part-time/full time, household variables related to marriage and children, and parental characteristics played a role in predicting occupational mismatch. The most significantly related covariates were the ones related to personal biography, and work-related variables such as working part-time. Overeducation however, was discovered to be a variable that could differ regionally and was state-dependant. Some limitations of this studies, including others is not being able to include earnings in the model. Meanwhile, the example of analyzing immigrants' overeducation in Sweden uses explanatory variables such as

previous overeducation, age, number of children, years of schooling, and years present in Sweden in a dynamic, probit model.

The nature of occupational mismatch itself may not be so binary in reality (there is a wide spectrum of mismatch amounts) but an analysis such as one of these give quantities to the influence explanatory variables can have on mismatch, in the absence of other useful data, such as wages.

3 Data: Description and Drawbacks

The data in this investigation comes from two main sources. The first occupational concerning occupational information and wages, is the Spanish Wage Structure Survey (WSS). This comes in four waves in 1995, 2002, 2006 and 2010. This was produced by National Institute of Statistics (INE) in Spain with the aim of providing more information on wage levels, at a micro-data level, along with other variables that could have an influence on wages. Workers can be classified by sex, type of working days, economic activity and other, similar categories. The information is obtained through the Social Security system and the State Tax administration Agency.

The data used is gathered in firms, and are classified into one of the following sectors: business services, scientific professionals, support technicians and professionals, administrative employees, personal and protection services, skilled workers in the agricultural sector, skilled workers in art, manufacturing and construction (except for machinery), machinery, and unskilled workers. These are taken from the CNO-1994 and CNO-2011 occupatoinal codes in Spain. Education, health, and other social activities are not included. There is a great potential to do regional analysis with the data, but is not part of the aim of this investigation.

In order to make the data more streamlined across its different waves, firms with less than 10 workers were eliminated. Also any additional, minor occupational categories that might have been added in the recent waves were not included. Related to variables used for estimating wages, gross monthly wages were used. (from the month of October; which is representative of the whole year) Since Spain, used the peseta as a form of currency in 1999, this was converted into 2002 euros. Information on educational investment was not provided in the form of years completed. This was modified into years completed of education, based on the highest number of years needed in order to complete schooling in a respective title or degree. Experience was not provided, so a commonly used proxy in the literature of taking the difference of 6 years and years of schooling from the individual's age was implemented.

The second data source used was the Economically Active Population Survey (EAPS) by the INE. This is a quarterly survey, having been conducted through the past decades in Spain and focuses on families. The main purpose is to be able to get data on the

labor workforce, and whether individuals are currently in employment. It has been especially used in past investigations to study problems related to unemployment. The survey covers around 65,000 families which makes it a rich data source too. It is used in the investigation as a major complement to the Wage Structure Survey as it covers many more years, and still contains relevant information about worker's occupations, and related job/demographic characteristics. The third quarter was used in the selected years, in order to be consistent with the Wage Structure Survey. The data from this survey was also used in a similar manner to assign years of education to various individuals and their highest title/qualification. 1999 had more detailed information on the level of schooling obtained but was streamlined to the level of detail as the other years. Data was organized so that each individuals occupation would categorized in the same manner as in the Wage Structure Survey. The changes made to the dataset are similar to changes made in similar studies. Depending on if one focuses on wage inequality, or immigrant wages, it could be restricted more, but is not relevant to this study.

These data sources are the main ones used to analyze the evolution of labor market forces, in Spain. The only other survey which is used by authors in this topic is the European Household Labor Force Survey, which is not accessible for this study. There are some smaller surveys such as the Survey on Transition from Education and Training to Labor Market Insertion (2014), but it does not provide many details that are not already available in the Wage Structure Survey or the Economically Active Population Survey.

Finally, the main limitation of these datasets is the fact that the WSS does not contain micro-data on wages after 2010, and that neither of the two surveys can be used to construct a usable panel. Using a panel would allow the analysis to work with the problem of unobserved heterogeneity of the individuals. Occupational mismatch being corelated with innate abilities of the individuals could be a potential problem. However, the panel created could only contain around 200 individuals and was not considered sufficient for a rigorous analysis. One problem that several datasets have in investigations such as these is that certain occupational categories in the dataset are quite small. This could potentially cause unreliable estimates, and comparisons in the results, but is not an issue in this investigation. There are a substantial amount of observations for the nine occupational categories in both datasets.

Hence, both micro-data surveys provide some of the most representative information on the working population in Spain. Their information is quite relevant to the variables being investigated, in relation to wages, occupational characterisitics, and details

related to educational attainment. There is also little question about its reliability as they are highly reputed and competent sources used for most investigations in Spain.

4 Methods

4.1 Mincerian wage equation

The first part of the methodology is to observe private returns on education using the traditional Mincerian wage equations. As shown in the model below, wages are explained by years of schooling, experience and its square, and a term representing random error such as individual characteristics that have an influence on wages

$$\ln(w_i) = \alpha + \beta S_i + \gamma_1 E_i + \gamma_2 E^2 i + u_i \qquad (1)$$

This model simply allows the reader to get a look at how private returns have been evolving until just past the recessionary period. Rising returns in the context of the Spanish economy have been seen as slight indications of overeducation as per the aforementioned literature (Murillo, 2012). However, all educational levels are considered to have an equal effect in the latter. This is questionable, given the range of different levels of education that one could have attained as their highest. One way to discern this in a more realistic manner is to treat the various educational levels as dummy variables: first level of secondary education (SEC1), second level of secondary education (SEC2), middle grade vocational training (mvt), upper grade vocational training (uvt), a diplomature (DIP, one year less than a B.A. education), and then a university education equal to a B.A. or higher (B.A.). Educational titles in Spain are visualized and better described in Figure 15 in Appendix A.

Model (2) is described as follows:

$$\ln(w_i) = \alpha + \beta_1 SEC1_i + \beta_2 SEC2_i + \beta_3 MVT_i + \beta_4 UVT_i + \beta_5 DIP_i + \beta_6 B.A._i + \gamma_1 E_i + \gamma_2 E2_i + u_i \qquad (2)$$

After observing effects of additional years of education in each title, model (3) is a continuation of the analysis that attempts at estimating marginal benefits of obtaining the next highest educational title. This would give more information about the labor market forces of

signaling versus screening, and how successful increased signaling might be in the Spanish labor market. There have been similar calculations performed by Raymond (2009) in the past.

$$R_B = \frac{(\beta_B - \beta_A)}{(S_B - S_A)} \qquad (3)$$

where R_B is the wage skill premium associated with attaining an educational level of B as opposed to attaining an educational level of A. S signifies the years of education required in order to obtain the title, and β from each educational level signifies the wage skill premium. The intuition behind this is that an individual gets a net benefit as he/she incurs an opportunity cost when pursuing the next academic qualification, in terms of wages, but then may recover this through the higher level of wages that they receive after obtaining this qualification.

4.2 ORU Regressions: Estimating the effect of mismatch on earnings

Following these basic studies of the evolution of private returns on education, the latter will be modified in order to relate it to occupational mismatch. One successful method by (Duncan and Hoffman, 1981) has been to use the modal reference for overeducation in order to modify the traditional, Mincerian wage equation.

This would appear as such:

$$\ln(wi) = \alpha + \beta_o So_i + \beta_a Sa_i + \beta_u Su_i + \gamma_1 E_i + \gamma_1 E^2 i + ui \qquad (4)$$

where β_o is the coefficient of overeducation if the individual has studied more years than their occupation's modal value, β_a is the coefficient of *appropriate* years of education, β_u is the coefficient related to years of undereducation in relation to the modal value, and the following terms signify experience and the error term.

As mentioned before, a strength of this estimation is that it adjusts for technological changes in occupations and is not strongly affected by outliers. When comparing ORU regressions to straightforward Mincer equations, the benefits of this measure are its focus on demand side variables. This is usually not as prevalent in traditional earnings functions, unless one includes dummies for industry or alike.

However, the shortcomings of Mincer equations and ORU regressions in general must be addressed. As is widely known, Mincer equations are liable to bias when estimated in OLS regressions. This bias is present in several factors, such as measurement error of ability, omitted variable bias, and the endogeneity of the years of schooling. Thus, in the ORU regressions, required education will most likely be measured with substantial error. Using an instrumental variable that measures motivation or abilities would be quite useful here, in order to relate it to studying too many or too few years for one's occupation but there are not many cases of this in the literature.

In terms of interpreting the results of such regressions, As mentioned by Hartog (2000) ORU regressions serve as a good way of looking at the market assignment of human capital. They do not provide any direct information on how shifts in supply or demand change coefficients of earnings functions. However, they do provide some information on inequality in earnings. One of the recommendations in the literature for fully utilizing the potential of ORU regressions is to be able to look at changes over greater periods of times, and observe the evolution of overeducation and undereducation measures. The problem of measurement error would need to be resolved with the use of instrumental variables. More research needs to be done in order to connect it to structural models of the labor market in order for its interpretation to be more powerful. Lastly, a potential problem in most of the studies using ORU regressions is that they mostly make use of cross-sectional data. The heterogeneity of the individuals could thus be missing. One method of solving this is to use panel estimation techniques. However, this again depends on the data available to the researcher.

4.3 Skill mismatch using EAPS

Thus, in order to give a bit more of a timeline to the picture of occupational mismatch, the percentage of mismatched candidates will be calculated for the same years, and for 2013 in

the EAPS survey. It will also be specified in detail here, which educational levels, and types of people are affected by this by creating dummies for the respective specifications. This can be compared as a check against mismatch results from the WSS, except for any interaction between the mismatch of workers and their wages, since the EAPS does not have any information on wages.

4.4 Logit Model

4.4.1 Empirical Strategy

As previously stated, one of the objectives in the analysis is to run a binary outcome model in order to visualize which parts of the labor market have a higher probability of being inefficiently allocated, and how these evolve over the past two decades. In order to do this, the model to be tested must be correctly specified. Problems with such models often include misspecification of the dependant variable, and a constant that is not significant. Also, given the nature of probit and logit models, one cannot simply insert all the characteristics they would like to investigate for and consequently test the model. After running a criterion test (Akaike, Bayesian Information) it appeared that a logit model would be a more appropriate fit for the data. (less flattened ends of the distribution) Robustness checks for this were also done, such as running a probit function for the same system, and comparing the results. Goodness-of-fit tests, along with checking the area under the ROC curve were implemented in order to verify the logit model as being an appropriate one. (These can be found to have been in order with the data results, as can be seen in the Appendix) Thus, based on the given literature, a model was formed such as the one in (5) will be tested using Logit.

$$Y_i = \alpha_0 + \beta_1 X_i + \delta sec1_1 + \delta mvt_1 + \delta ter_1 + \delta female_1 + \delta temp_1 + \delta otherjob_1 + \delta tenuremonths_1 + \delta twentothryfour_1 + \delta thirtyfivetosixfour_1 + \varepsilon \quad (5)$$

In this model, *sec1* relates to lower secondary education, *mvt* relates to lower grade vocational training, *ter* refers to tertiary education (diplomature from WSS, and regular university degrees), female refers to being a female, *temp* to having a temporary contract, *otherjob*

relates to having another job, *tenuremonths* refers to how many months they have worked in the same job, *twetothryfour* refers to the age group 20-34 years old, *thirtyfivetosixfour* refers to the age group 25-64 years old, and ε is a residual term. The link test was conducted in order to verify that mismatch could be an appropriate dependant variable. The results described it to be so in each case. It was taken care that no dichotomous errors were committed, as male was not included as a dummy for gender, and neither were all education, job related, and age dummy variables.

The explanatory variables are mostly dummy variables, with the value of 1 if an individual contains that characteristic, and 0 otherwise. Tenure is the only variable that is continous. Further on, the marginal effects of the logit model will be presented, where the effect of all the explanatory variables can be seen to have on the odds that the dependent variable, *mismatch,* can have a value of 1. (Signifiying that an individual on average would have an occupational mismatch, given those conditions) Models such as this have been tested in Kalfa (2013).

5 Empirical Analysis

5.1 Results

The first result presented below is from the basic Mincerian wage equation, as previously discussed.

Table 1. Evolution in the wage skill premium in Spain (1995-2010)

	1995	s.e.	2002	s.e.	2006	s.e.	2010	s.e.
Constant	-0.012	-1.44	0.278	14.85	0.643	86.99	0.758	88.68
Years of Education	0.094	224.92	0.105	115.04	0.079	239.36	0.087	229.56
Experience	0.059	117.5	0.040	33.91	0.040	83.78	0.035	65.88
Experience^2	-0.001	-61.28	0.000	-11.98	0.000	-35.32	0.000	-24.88
R squared	0.307		0.305		0.241		0.243	
Standard Deviation	0.539		0.581		0.593		0.558	
F-Statistic	25180.2		5005.1		23322.2		20575.8	
Observations	170584		34188		220403		192066	

One can see fluctuation in the returns on education (From a 9.5% to 10.5% wage return on an additional year of education) but definitely lowering by 2006. (7.9%) Later in 2010, it appears the wage returns are rising again. (8.7%) This would almost make it seem like a U-shape is in form when looking at the evolution of wages from the 1990s through to the recession to today. However, it would be too bold to claim that simply from four data points in time. It seems that the value of studying an extra year changed momentum at some point between 2006 and 2010; it is possible that in the presence of overeducation and screening/signaling, that this would be a plausible results. However, further analysis must be conducted in order to make such a statement. Experience on the other hand, lowered its effect on wages from a bit below 6% in 1995 to a bit below 4% in 2002, staying around level before dropping a bit further to 3.52% in 2010.

Following this, Table 2 is presented, where the different levels of schooling no longer have an equal influence on wages.

Table 2. Evolution in the wage skill premium for each educational level (1995-2010)

	1995	s.e.	2002	s.e.	2006	s.e.	2010	s.e.
university graduate	1.210	0.01	1.177	0.01	1.018	0.00	1.028	0.00
diplomature	0.897	0.01	0.899	0.01	0.780	0.00	0.761	0.01
sec 2	0.581	0.00	0.504	0.01	0.437	0.00	0.403	0.00
uvt	0.642	0.01	0.655	0.01	0.601	0.01	0.583	0.01
mvt	0.454	0.01	0.435	0.02	0.441	0.01	0.387	0.01
sec 1	0.147	0.00	0.157	0.01	0.160	0.00	0.136	0.00
R^2	0.353		0.336		0.264		0.277	
Standard Deviation	0.521		0.567		0.584		0.545	
F-Statistic	11612.3		2164.5		9866.0		9183.8	
Observations	170584.0		34188.0		220403.0		192066.0	

Thus, the wage skill premiums respective to an additional year within each educational title obtained can be seen. Generally and as expected, the returns to an additional year within each educational title keeps are of a larger amount with a higher title. However, it can be observed that undertaking an additional year of upper grade vocational training provides a slightly higher return than an additional year of upper secondary education. Several educational titles have decreasing private returns over the past fifteen years, some having a temporary increase in 2006, two years before the Great Recession. Previous literature such as Murillo (2013) have found in the past that returns were declining more in the lowest levels of education. The opposite trend has been seen here except for the returns of the very highest university qualifications. The declines in returns appear to be highest when increasing in educational level from lower secondary levels towards the diplomature qualification. This is argued by

authors to be an increase in the number of graduates in general, without an increase in the supply of high skilled jobs. Vila and Mora (1998), Budría and Moro-Egido (2008)

However, one does not obtain any information here about whether it is worth the opportunity cost of not working temporarily, and obtaining a higher degree. This is now presented in Table 3.

Table 3. Marginal benefits of getting a higher educational title (1995-2010)

	1995	2002	2006	2010
university graduate vs.				
diplomature	0.156	0.139	0.119	0.134
diplomature vs. sec2	0.317	0.395	0.343	0.358
sec 2 vs. sec1	0.217	0.174	-0.028	0.134
uvt vs. mvt	0.094	0.110	0.080	0.098
mvt vs. sec1	0.153	0.139	0.140	0.126

As can be seen, the highest marginal benefits on wages of obtaining the next highest educational title lie within obtaining a diplomature degree as opposed to having upper secondary education as the highest academic qualification. The benefits seem to following a U-shape over time. Hypothetically, it might make sense that the marginal benefits would rise towards 2010 in the context of high signaling/screening, and a recessionary period where higher qualifications than previously may be needed in order to obtain a position requiring a given level of skills. Working in a positions after having obtained upper secondary education as opposed to lower secondary education decreases at first, even showing in 2006 that it an individual with lower secondary education would be receiving higher wage returns for the same position (negative return on overeducation) , but rises again after the Great Recession by 2010. It appears possible that due to market allocation changes in human capital, the previous and unnecessary overeducation, is maybe closer to the required years of education for certain job positions for most educational titles. Thus all the comparisons seem to have a pseudo- U-shaped trend, with the trough being between 2006 and 2010, where marginal wage returns for the next highest title seem to be rising again, except for when comparing obtaining medium grade vocational training as opposed to lower secondary education. The trend seems to be a decreasing one even after 2006. There is not a straightforward, consecutively positive trend of

marginal benefits with the next higher title, but it appears that obtaining higher titles do provide individuals with greater marginal benefits in wages.

As aforementioned, literature such as Murillo (2013) suggest that declines in wage skill premiums could be related to increases in overeducation. This seems to have happened, followed by an increase after 2013. This could maybe be a decrease in the amount of individuals allocated as 'overeducated' in the labor market. To explore this further, ORU regressions are conducted on the data from the Wage Structure Survey.

Table 4. ORU regression (1995-2010)

	1995	s.e.	2002	s.e.	2006	s.e.	2010	s.e.
Constant	0.7773	(93.4)	1.3845	(-71.71)	1.5207	(200.4)	1.5646	(176.1)
Undereducation	-0.067	(-72.21)	-0.0252	(-9.17)	0.0119	(15.55)	-0.0106	(-11.4)
Appropriate years	0.0398	(94.42)	0.0344	(38.04)	0.0208	(61.16)	0.0284	(83.39)
Overeducation	0.0712	(80.12)	0.0936	(28.39)	0.0507	(46.41)	0.1119	(84.82)
Experience	0.0617	(109.6)	0.0347	(25.6)	0.039	(73.92)	0.0416	(69.82)
Experience^2	-0.0008	(-84.01)	-0.0004	(-15.64)	-0.0005	(-48.98)	-0.0005	(-48.07)
R^2	0.1496		0.0908		0.0815		0.0919	
Standard Deviation	0.4284		0.66412		0.65239		0.61011	
F-Statistic	6002.02		682.64		3909.29		4018.46	
Observations	170584		34188		220403		192066	

Source: WSS, own elaboration

Note: For the purpose of clarity, no (*)s have been added; all the coefficients are significant at the 1% level

The results can be better summarized in the following table, which is expressed in percentages.

Table 5. Summary of returns to schooling, by educational mismatch (%)

	1995	2002	2006	2010
β_u	-6.7%	-2.5%	1.2%	-1.1%
β_a	4.0%	3.4%	2.1%	2.8%
β_o	7.1%	9.4%	5.1%	11.2%
β_o-β_a	3.1%	5.9%	3.0%	8.4%
β_u+β_a	-2.7%	0.9%	3.3%	1.8%

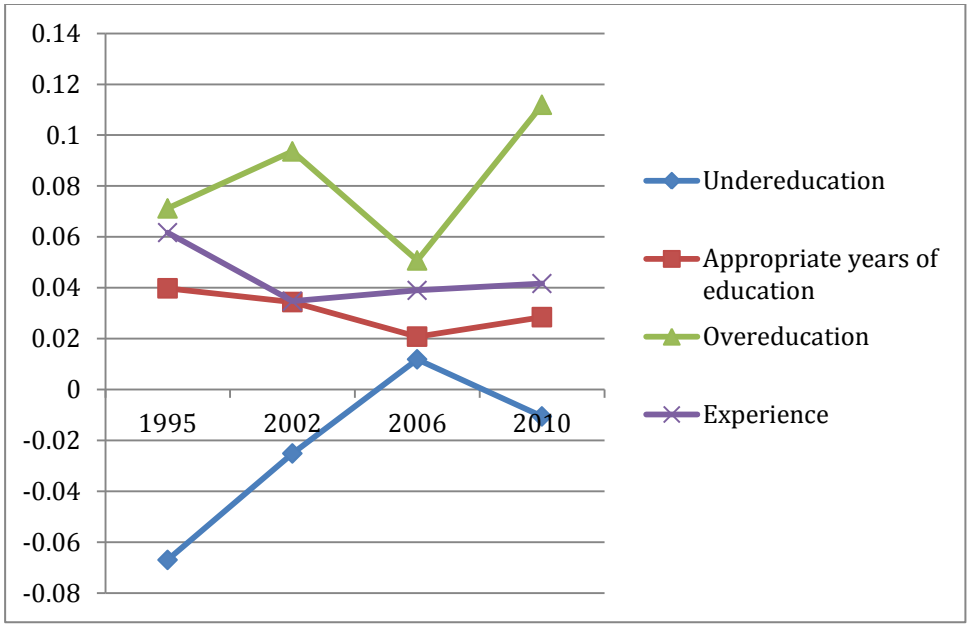
Source: WSS, own elaboration

The first three rows display the influence of an additional year of undereducation, appropriate education, and overeducation on wages, respectively. The returns to an additional yar of undereducation, which are mostly negative in the Spanish labor market, seem to have been steadily increasing (becoming even positive in 2006) but then begin to fall between 2006 and 2010. One could hypothesize that being undereducated in a recessionary economy would see lowering benefits in terms of wages. Having an additional year of appropriate education has greater benefits than an additional year of undereducation (obvious) but not as much as an additional year of overeducation. (This is slightly different from some results in the literature where overeducation has lower wage returns than appropriate years of education. However, both of these types of wage returns are lower in 2006 than in 2002, but rise again by 2010.

The term (β_o-β_a) signify the loss of income of a worker who has an extra year of overeducation compared with another who has achieved a match between his job and education with the same educational level. As implied above, The wage premium for additional overeducation is always higher than the premium for additional required education. The difference between the two has fluctuations, but definitely increases over time, signifying some allocation problems in the Spanish labor market. Secondly, the term (β_u+β_a) signifies the additional wages received by a worker who has an extra year of undereducation compared to someone who works in a well-matched job, but with the same educational level. The results show that this difference is negative in 1995, but turns positive, and stays positive throughout the next few years. The worker who has the same education but works in an appropriately matched job is 'worse off' in 1995. However, the difference moves in favor of the undereducated person in the post-recessionary period, in comparison to 2006. Finally, the penalty for an additional year of undereducation always maintains itself to be smaller in size than the wage premium benefit that is obtained with an additional year of overeducation. This

is maintained throughout the four years observed in the 15 years total. This can be visualized more easily in Figure 2 below.

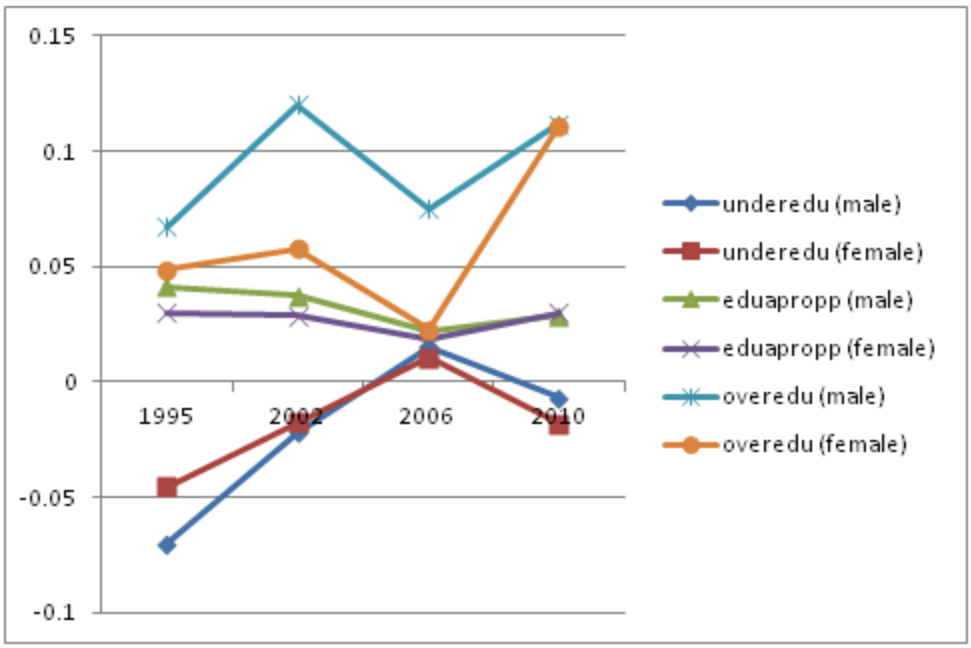
Figure 2. Effects of occupational match/mismatch on wages (1995-2010)



Source: WSS, own elaboration

The results in general suggest that the mismatch for overeducation benefits overeducated workers in the Spanish labor market more than appropriately educated workers. However, undereducated workers who were previously receiving a smaller penalty, are now receiving a higher penalty in 2010 again. It would be interesting to see these effects by gender and other characteristics, which will be presented next

Figure 3. Incidence of occupational match/mismatch by gender (1995-2010)

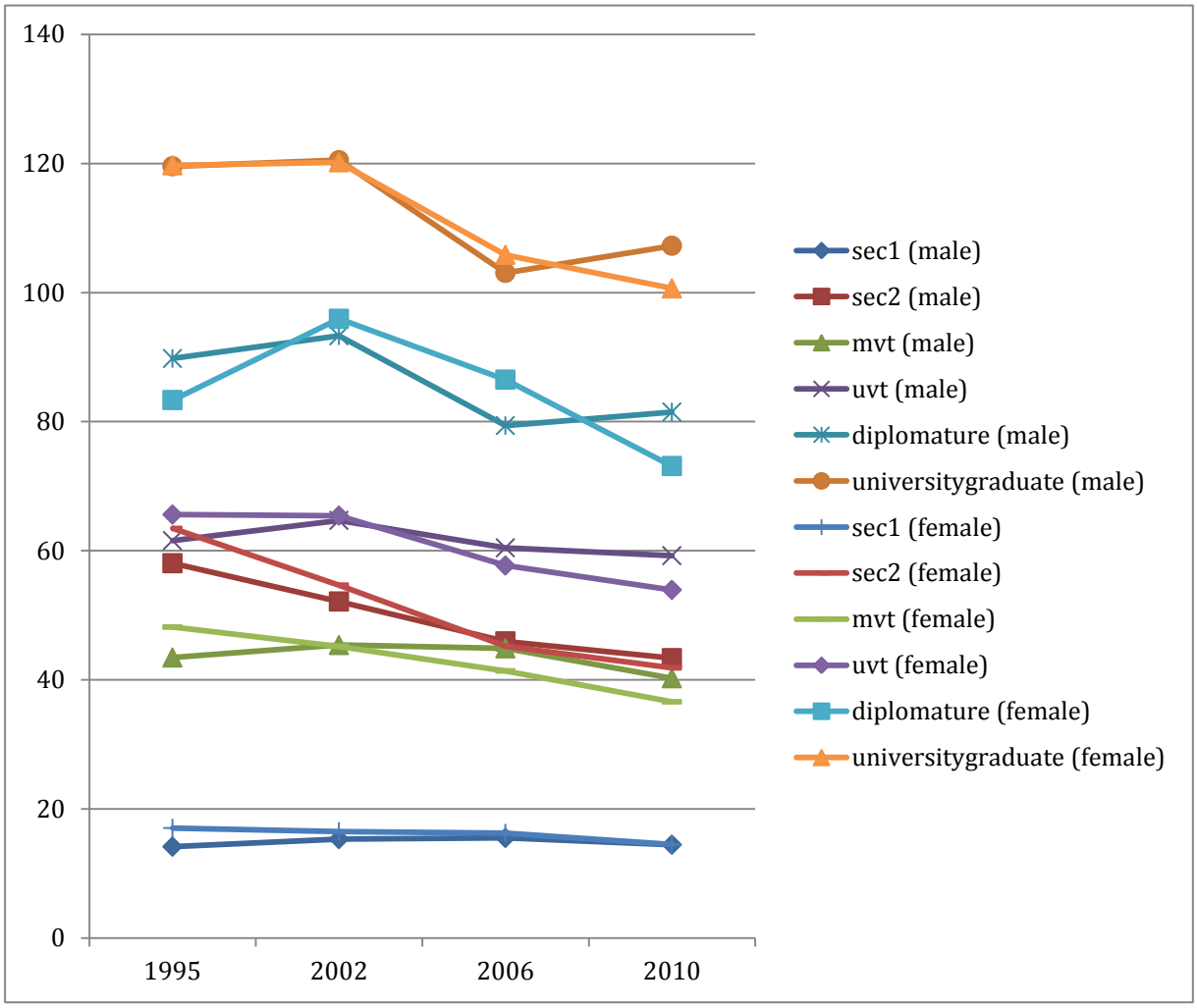


Source: WSS, own elaboration

Above can be seen the wage skill premiums on additional years of undereducation, appropriate education, and overeducation. The premiums for additional years of undereducation and appropriate education are slightly different for men and women in 2006, but converge towards 2010. The penalty for additional undereducation was slightly higher for males, but they also received a slightly higher wage premium for an additional year of appropriate education. When it comes to overeducation, both males and females saw the difference widen over time, with males receiving higher premiums, but by 2010 receive the same level. These results are quite interesting as there is literature that suggests gender wage inequality in Spain is decreasing over time.

However, this is assuming that all levels of education have equal gender differences in Spain. In order to verify the reality, dummies have been incorporated by educational level and gender in the model, and visualized in Figure 4. (Tables representing these changes can be found in Appendix C.

Figure 4. Effect of educational attainment on wages, by gender (1995-2010)

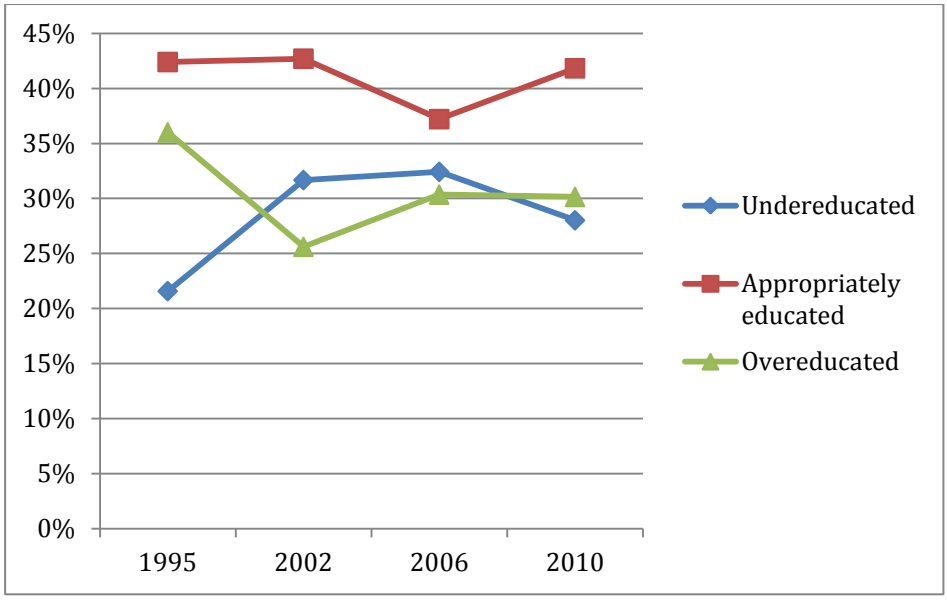


Source: WSS, own elaboration

Although the difference is not pronounced at an extremely high level, it can be seen that there are small difference between males and females in certain educational levels. It could be said that the difference seems to rise slightly as does the educational level. The more skilled a male/female is, the more pronounced the difference in the wage premiums. In vocational training, it seems females who used to receive higher skill premiums for an additional year of training, have switched places with males. The biggest gap is aroud 9% in diplomature education, where the female who used to receive a higher premium in 2006 now receives a lower one in 2010. With regards to a higher university title, females who used to receive the same level of premiums as males, receive a premium in 2010 of approximately 7% less. Thus, it seems that gender differences are not pronounced in jobs for lower educated individuals, but there is a noticeable difference in higher skilled jobs.

In order to summarize the mismatch information available in the WSS, Figure 5 below show the percentage of mismatched individuals among the Spanish labor force. This shows the evolution of the latter from 1995 to 2010.

Figure 5. Evolution of mismatch shares in Spanish labor force from WSS



Source: WSS, own elaboration

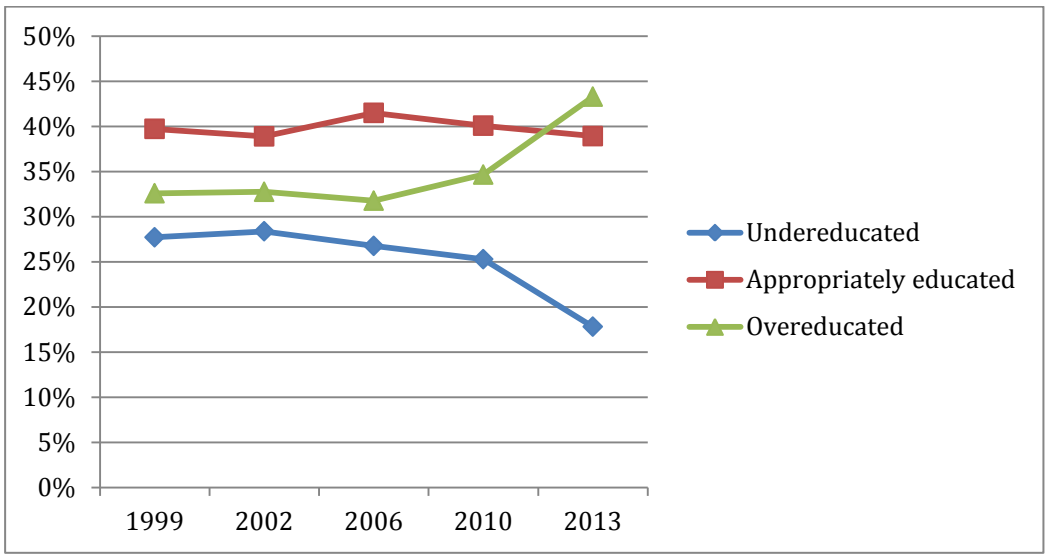
Thus far, some effects of occupational mismatch have been seen on the wage premiums, and additionaly observed through separate genders and educational levels. The results seem to suggest that towards 2006, screening versus signaling effects were favoring overeducation less, and undereducation more in terms of wage premiums. However, there appears to be the start of a U-shape, since the trends move in the opposite direction between 2006 and 2010. The behavior of labor market allocation seems to have changed during the time of the recession. Nevertheless, wage skill premiums have still been falling, and slightly more so for highly educated females. The differences in the wage skill premiums of overeducation, undereducation and appropriate years of education appear to have converged more over time for both genders though, probably in the lower levels of education. This suggests maybe more labor market allocation problems at higher education levels, where the marginal benefits of

obtaining education are relatively higher than at the lower levels. However, the observation for the post-recessionary period is only one year (2010), which can be argued to be very limited in terms of analyzing post-recessionary mismatch in Spain. Therefore, the Economically Active Population Survey, which has information past 2010, but not much information on wages, will be used to complement this investigation.

Analysis from the EAPS

Data was observed in the EAPS in three years prior to and three years following the Great Recession. This provides two more observations over the course of five more years in the Spanish labor market. However, it must be noted that there is a slight difference in the information available on educational levels in the EAPS. The incidence of undereducation, overeducation, and appropriate years of education is presented below in Figure 6.

Figure 6. Occupational Mismatch using the EAPS (1999-2013)



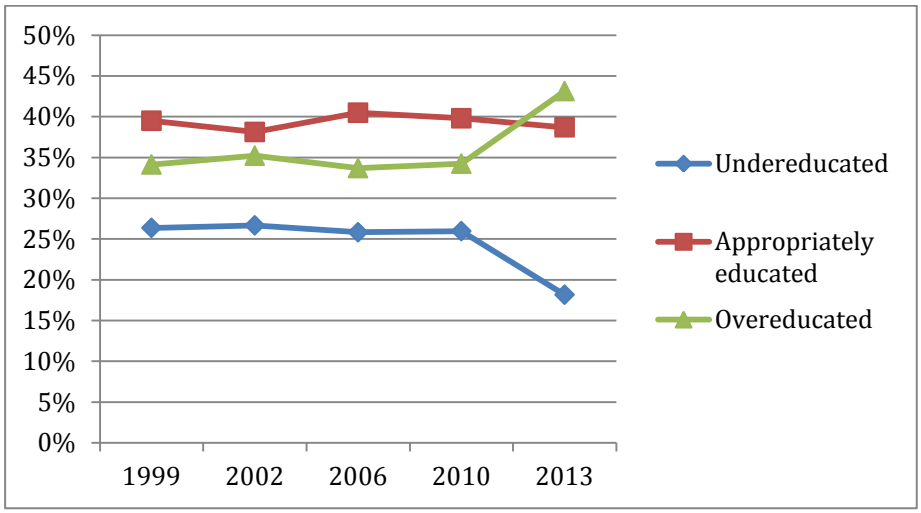
Source: EAPS, own elaboration

These results can be considered as a robustness check for the previous results in Figures 3 and 4. However, the major difference in the survey to be noted is that the WSS was a survey undertaken in Spanish firms, while the EAPS was undertaken in households. The observations from 1999 to 2010 seem to follow the same trend. From 2010 onwards however, the shares of

overeducation and undereducation seem to follow opposite trends. This could suggest some changes that took effect in the Spanish labor market from 2012 onwards, such as differences in technological aspects of available jobs, or the labor reforms mentioned in section 2. In order to obtain more details about these changes, it can be observed whether this differs by gender or educational level in the labor market.

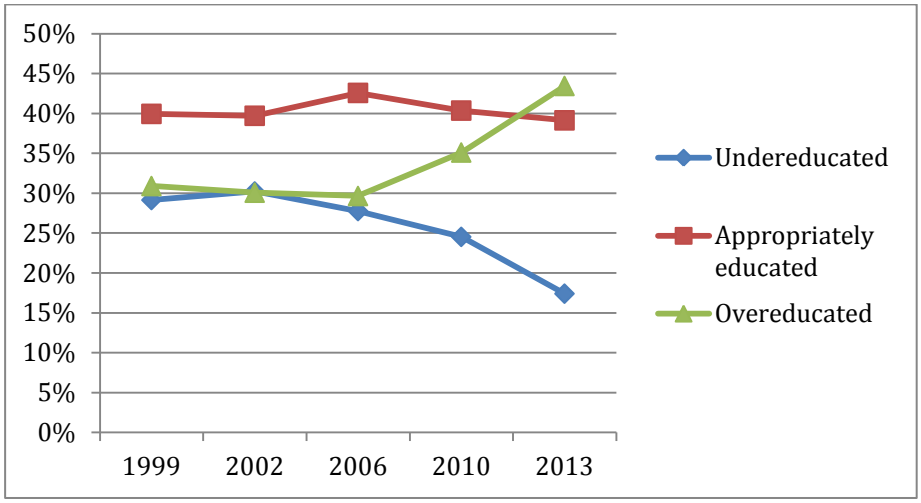
The figures below present the shares of mismatch by gender and educational levels in the EAPS.

Figure 7. Share of Mismatch in Female Population



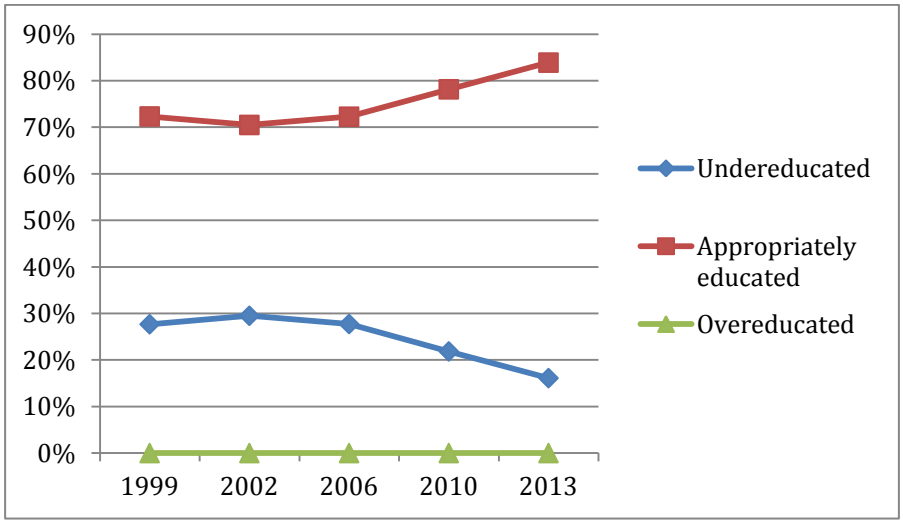
Source: EAPS, own elaboration

Figure 8. Share of Mismatch in Male Population



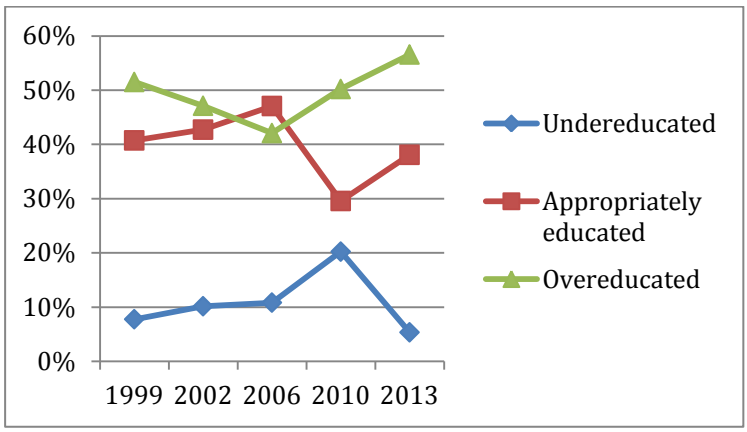
Source: EAPS, own elaboration

Figure 9. Share of Mismatch in individuals with Primary educational attainment (highest)



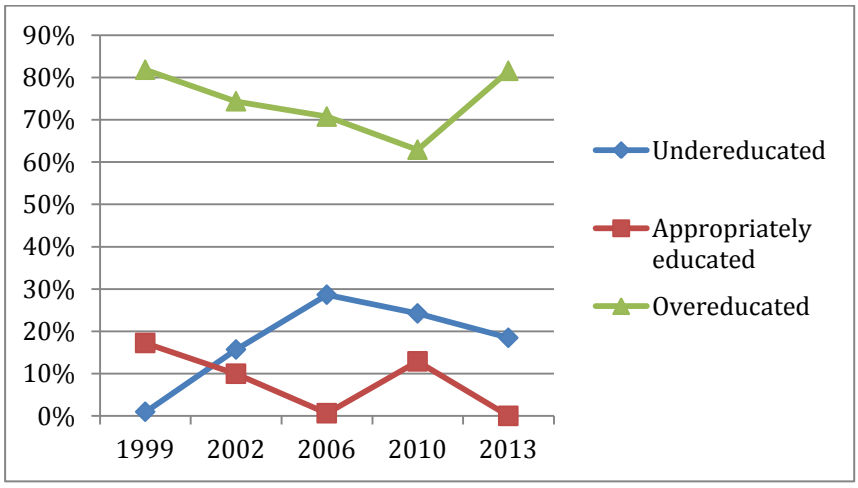
Source: EAPS, own elaboration

Figure 10. Share of Mismatch in individuals with lower secondary educational attainment



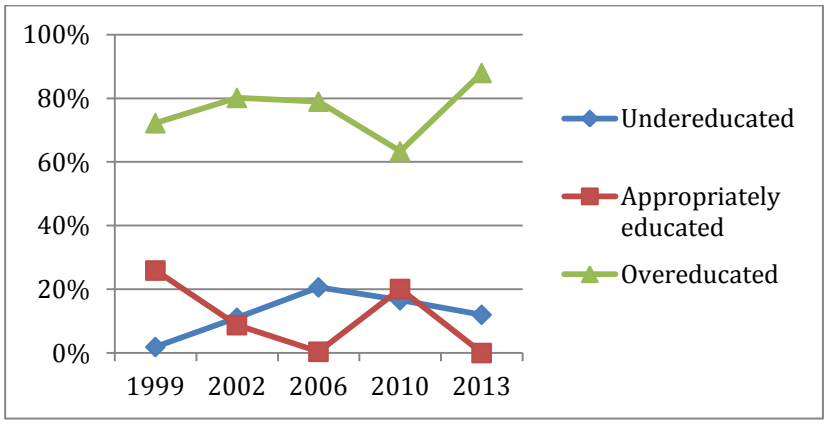
Source: EAPS, own elaboration

Figure 11. Share of Mismatch in individuals with upper secondary educational attainment (highest)



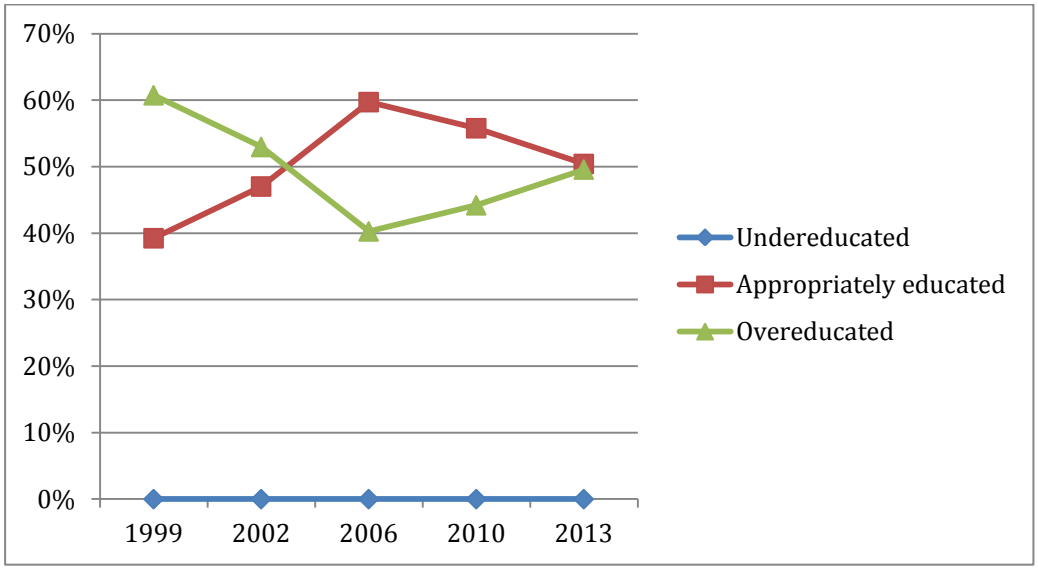
Source: EAPS, own elaboration

Figure 12. Share of Mismatch in individuals with upper vocational training (highest)



Source: EAPS, own elaboration

Figure 13. Share of Mismatch in individuals with Tertiary/Higher Education (highest)



Source: EAPS, own elaboration

In terms of gender, there is not much of a pronounced difference in mismatch shares if observing on average all educational levels. There are a slightly higher amount of undereducated females than overeducated females, while the share of both for males is equal. When separating the analysis simply by educational attainment, there are as expected, some major differences. Individuals with primary education were mostly appropriately educated for their occupations, with a small portion of undereducated workers. However, also after the

benchmark of 2013, the shares of undereducated and overeducated workers made a drastic change, almost all primary educated (maximum attainment) individuals qualify as undereducated for their occupations. When observing individuals that obtained a highest of lower secondary education, a similar process seems to occur but with the share of overeducated and appropriately educated workers. The share of appropriately educated workers rises in 2013, while the share of overeducated workers sees a sharp decline. The share of undereducated and overeducated individuals rose in the recessionary period, showing a polarizing effect in the allocation of these individuals, but changed to the latter in 2013.

With regards to upper secondary educated and upper vocationally trained individuals, similar trends occur. The trend is an inverted –U falling till around 2010. Following this, the incidence of overeducation rises again. Meanwhile, the incidence of appropriate educated workers seems to follow a trend in direct opposite relation with this. It rises when the incidence of overeducation falls, and vice-versa. A small difference is that there is a slightly higher incidence for undereducation among upper secondary educated individuals as opposed to upper vocationally trained individuals. However, they are low for both demographics. It could make sense in the labor market since individuals with upper secondary education tend to apply for a wider range of occupations than upper vocationally trained individuals, who are trained for specific positions in the labor market.

Tertiary educated individuals see an exchange between overeducation and appropriately educated allocation of their skills through the observed time period. The incidence of overeducation rises through the recessionary period while the incidence of appropriate education (which is higher) falls. they both reach a 50-50 level in 2013. However, it should be noted that 50% of tertiary educated individuals working in jobs that do not require those skills is quite a high number. The rising rate of overeducation was slowed a bit by the recession, but it appears that it has started rising at the same rate as it did before the recession. This could possibly be due to allocation related to labor market reform or technological changes in Spain.

As was previously described, a logit model was used to evaluate the effect of the variables of educational levels and other job characteristics on a binary outcome (1) of having an occupational mismatch, or (0) not.

Table 6 . Marginal Effects influencing Occupational Mismatch (1999-2000)

	1999	s.e.	2002	s.e.	2006	s.e.	2010	s.e.
sec1	-0.5323**	(0.0024)	-0.5824**	(0.0026)	-0.6681**	(0.0102)	-0.4053**	(0.0045)
Mvt	-0.2051**	(0.0059)	-0.0596**	(0.0075)	-0.0055	(0.0199)	-0.1844**	(0.0070)
Ter	-0.3303**	(0.0041)	-0.3981**	(0.0033)	-0.6057**	(0.0102)	-0.5330**	(0.0035)
Female	-0.0451**	(0.0039)	-0.0411**	(0.0035)	-0.0324**	(0.0029)	-0.0239**	(0.0038)
Temp	0.0173**	(0.0045)	0.0124**	(0.0041)	-0.0107**	(0.0034)	-0.0190**	(0.0049)
Otherjob	-0.0576**	(0.0136)	0.0196**	(0.0025)	0.0064**	(0.0020)	0.0100**	(0.0026)
tenuremonths	-0.0001**	(0.0000)	-0.0001**	(0.0000)	-0.0001**	(0.0000)	-0.0001**	(0.0000)
twentothrtyfour	-0.0274**	(0.0070)	-0.0083*	(0.0074)	0.0200**	(0.0071)	-0.0546**	(0.0091)
thirtyfivetosixfour	0.0055	(0.0067)	-0.0146	(0.0070)	0.0027	(0.0067)	-0.0453**	(0.0081)

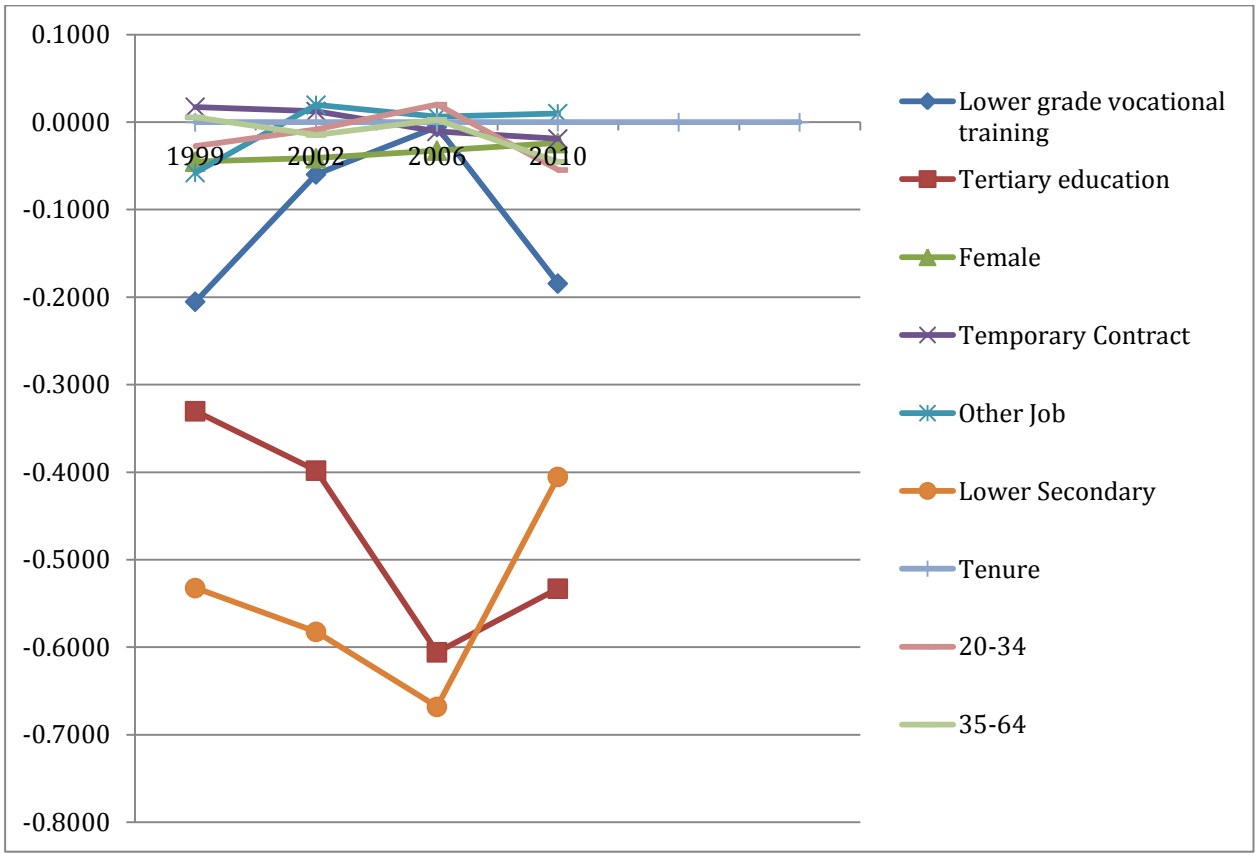
Source: EAPS, own elaboration

** significant at 5% level

***significant at 1% level

This is also presented in a more visual display, in Figure 14.

Figure 14. Marginal Effects influencing Occupational Mismatch (1999-2000)



Source: EAPS, own elaboration

The marginal effects from the logit model show that most of the explanatory variables do not predict log odds of an individual having an occupational mismatch (1) on average. The variables that do show this over time are the dummy variables representing having a temporary contract, having another job, and being in the age group of 20-34 years old. This can be seen with their positive coefficients; however, the only one showing this in 2010 is the variable of having a second job. (A small 1 % log odds of the individual having an occupational mismatch). None of the log odds ever rise above 2%, showing that in the way that the model is currently specified, other variables would probably predict better odds on average of a worker experiencing occupational mismatch. Omitted variable bias is an obvious limitation of this model but given the dataset used, and the variables at hand, this is the closest model that could be tested, that involved the same variables as previously analyzed in ORU regressions.

6 Conclusion

6.1 Summary

Thus, rigourous analysis has been conducted in the form of wage equations, ORU regressions, calculations of mismatch shares, and a binary outcome model to better examine which variables/personal characteristics are most associated with it. The results show that there is indeed a polarization (educationally) in the Spanish labor force. (It must be noted that those with tertiary education are not the most highly mismatched however, the most highly mismatched are those with secondary education. Nevertheless, the polarized educational system seems to be hold as a general trend) When looking at the analysis during the 2008 period of the recession, higher educated workers find themselves to be more 'appropriately matched' while lower educated workers seem to suffer negative returns to their undereducation. Nevertheless, it seems that some of these effects are starting to subside by 2013, which is 5 years post the bubble burst. The results also imply that overeducation in Spain is not a transient phenomenon; it can be seen that on average, and for educational levels of secondary and upwards, that the incidence has been increasing, just at a slightly lower rate in the years surrounding the recession. This suggests towards some permanent structural issues that are still present in the Spanish labor market, possibly some of them even since the 1970s. Issues that were mentioned, but not delved into deeply, such as the incidence of temporary contracts could be quite related to this. The Spanish government, as previously described has leveled out reforms several times, but does not always succeed in their objectives. The effects on earnings seems to be mostly positive for those highly educated workers who are 'overqualified' for their jobs, but not so much for the others during the years just after the recession. This suggests that there could be further polarization taking place in the Spanish labor market and that signaling and screening could possibly increase the incidence of overeducation in the years past 2013. This as mentioned, is validated by segmentation theory in the literature.

6.2 Review

The aims of this investigation were to provide more information on what has been happening with Spanish occupational mismatch, and who exactly it affects in the labor force. The research aims related to finding the relation between wages, gender, and temporal have been fulfilled. (with limitations of course) This research angle was slightly broad and can be used as a motivation for future investigations to delve deeper into specific issues: relations to wages, incidence of temporary employment, and even exploring which sectors are more affected. The Spanish government and educational institutions could attempt to provide more direct information about the requirements in occupations, which would also save lost resources through the excessive signaling. Binary outcomes would be quite interesting if a better model could be developed and tested with the available data. As a result, a little more light has been shed on the grey term of occupational mismatch, and hopefully, the reader has been able to visualize the relationship between its associated forces in the labor market.

References

Alba-Ramírez, A. (1993). Mismatch in the Spanish labour market: Overeducation? *The Journal of Human Resources, 28*, 259–278.

Arrow, Kenneth, 'Higher Education as a Filter', *Journal of Public Economics*, 2: 1973.

Bauer, T. K. (2002). Educational mismatch and wages: A panel analysis. *Economics of Education Review, 21*(3), 221-229. doi:10.1016/S0272-7757(01)00004-8

Bentolila and Jimeno, 2003 S. Bentolila, J.F. Jimeno. Spanish unemployment: the end of the wild ride? FEDEA Documento de Trabajo 2003–10, April, 2003 (2003)

Boll, C., Leppin, J.S. and Schömann, K., 2016. Who is overeducated and why? Probit and dynamic mixed multinomial logit analyses of vertical mismatch in East and West Germany. *Education Economics*, pp.1-24.

Bover et al., 2002 O. Bover, M. Arellano, S. Bentolila Unemployment duration, benefit duration and the business cycle Economic Journal, 112 (2002), pp. 223–265

Budría, S., & Moro-Egido, A. (2008). Education, educational mismatch and wage inequality: Evidence for Spain. *Economics of Education Review, 27*(3), 332–341.

Carrera, N., Carmona, S., & Gutiérrez, I. (2008). Human capital, age and job stability: Evidence from spanish certified auditors (1976-1988). *Accounting and Business Research, 38*(4), 295-312. doi:10.1080/00014788.2008.9663343

Congregado and García-Pérez, 2002 E. Congregado, J.I. García-Pérez. Análisis económico de la reciente reforma del sistema de prestaciones por desempleo Cuadernos de Información Económica, 169 (2002), pp. 105–116

Congregado, E., Golpe, A. A., & van Stel, A. (2011). Exploring the big jump in the spanish unemployment rate: Evidence on an 'added-worker' effect. *Economic Modelling,28*(3), 1099-1105. doi:10.1016/j.econmod.2010.11.018

Congregado, E.; Carmona, M.; Golpe, A. and van Stel, A. (2014): "Unemployment, Gender and Labor Force Participation in Spain: Future Trends in the Labor Market", Romanian Journal of Economic Forecasting, 2014 (1), 53-66.

Davia Rodríguez, Ma. Angeles (María Angeles), Hernanz Martín, V. & Fundación de Estudios de Economía Aplicada (Spain) 2002, *Temporary employment and segmentation in the spanish labour market: an empirical analysis through the study of wage differentials*, FEDEA, Madrid.

Dolado and Jimeno, 1997 J.J. Dolado, J.F. Jimeno The causes of Spanish unemployment: a structural VAR approach European Economic Review, 41 (1997), pp. 1281–1307

Dolado, J. J., Jansen, M., Felgueroso, F., Fuentes, A., & Wölfl, A. (2013). Youth labour market performance in spain and its determinants: A MICRO-LEVEL PERSPECTIVE.*OECD Economic Department Working Papers*, (1039), 0_1.

Duncan, G., & Hoffman, S. D. (1981). The incidence and wage effects of overeducation. *Economics of Education Review, 1*(1), 75–86.

Florentino Felgueroso & Manuel Hidalgo & SergiJiménez Martín, 2010. "Explaining the fall of the skill wage premium in Spain," Working Papers 2010-19, FEDEA.

García-Serrano, C., & Malo, M. A. (1996). Desajuste educativo y movilidad laboral en Espa˜na. *Revista de Economía Aplicada*, *11*(4), 105–131.

García-Serrano et al., 1998 C. García-Serrano, J.F. Jimeno, C. García-Serrano, J.F. Jimeno Labour reallocation, job tenure, labour flows and labour market institutions: evidence from Spain FEDEA Documento de Trabajo 98-07 (1998)

García-Montalvo, J., & Peiró, J. M. (2009). *Análisis de la sobrecualificación y la flexibilidad laboral*. Fundación Bancaja.

Groot, W., & Maassen van den Brink, Henriëtte. (2000). Overeducation in the labor market: A meta-analysis.*Economics of Education Review, 19*(2), 149-158. doi:10.1016/S0272-7757(99)00057-6

Gutiérrez, I. and Ortega, J. (2007). 'Higher education in the business fields: the case of Spain'. In Durand, T. and Dameron, S. (Eds.), The future of business schools, Hampshire: Palgrave. In press.

Hartog, J. (2000). Over-education and earnings: Where are we, where should we go? *Economics of Education Review*, *19*, 131–147.

Joona, P.A., Gupta, N.D. & Wadensjö, E. 2014, "Overeducation among immigrants in Sweden: incidence, wage effects and state dependence", *IZA Journal of Migration*, vol. 3, no. 1, pp. 1-23.

Kalfa, Eleni & Piracha, Matloob, 2013. "Immigrants' Educational Mismatch and the Penalty of Over-Education," IZA Discussion Papers 7721, Institute for the Study of Labor (IZA).

Kiker, B. F., Santos, M. C., & Mendes de Oliviera, M. (1997). Overeducation and undereducation: Evidence for Portugal. *Economics of Education Review*, *16*(2), 111–125.

Malo, M. A., & Garcia-Serrano, C. (2004). Overeducation, undereducation and expected promotions. *Revista Universitaria De Ciencias Del Trabajo,* 145-160. Retrieved June 1, 2016, from http://uvadoc.uva.es/handle/10324/11283

Malo, M. A., & Garcia-Serrano, C. (2004). Overeducation, undereducation and expected promotions. *Revista Universitaria De Ciencias Del Trabajo,* 145-160. Retrieved June 1, 2016, from http://uvadoc.uva.es/handle/10324/11283

Mínguez, Almudena Moreno, and Enrique Crespo Ballesteros. (2013)."School to Work Transitions among Young People in Spain from European Perspective."

Mincer, Jacob (1974), Chapter 4 in *Schooling, Experience, and Earnings*. Columbia University Press, New York.

Miller, Paul, Mulvey, Charles & Martin, Nick (1995), 'What do Twins Studies Reveal about the Economic Returns to Education? A Comparison of Australian and US Findings.' American Economic Review, 85: 586-99.

Ministerio de Educación. (2016). Las cifras de la educación en España. Curso 2013-2014 (Edición 2016). Madrid.

Murillo, I. P., Rahona-Lopez, M., & Salinas-Jimenez, M. d. M. (2012;2011;). Effects of educational mismatch on private returns to education: An analysis of the spanish case (1995-2006). *Journal of Policy Modeling, 34*(5), 646-659. doi:10.1016/j.jpolmod.2011.07.012

OECD (2015), Education at a Glance 2015: OECD Indicators, OECD Publishing. http://dx.doi.org/10.1787/eag-2015-en

Prieto, J. and Rodríguez, C., 2000. Participation of married women in the labour marker and the added worker effect in Europe, *IRISS Working Paper Series*, No. 2000-12.

Psacharopoulos, George & Harry Anthony Patrinos (2004), 'Human Capital and Rates of return, Introduction + Methodology', in Johnes, G & Johnes J (eds.), International Handbook of the Economics of Education, Edward Elgar Publishing.

Raymond, J. L., Roig, J. L., & Gómez, L. M. (2009). Rendimientos de la educación en Espa˜na y movilidad intergeneracional. *Papeles de Economía Espanola*, *119*, 188–205.

Riley, J. (1979b), 'Testing the educational screening hypothesis', *Journal of Political Economy*, 87 (5), S227–S252.

Rumberger, R. W. (1981). The rising incidence of overeducation in the US labour market. *Economics of Education Review*,*1*, 293–314.

Schofer, Evan & John W. Meyer (2005), 'The worldwide Expansion of Higher Education in the Twentieth Century', *American Sociological Review*, 70: 898-920.

Stiglitz, Joseph E (1975), 'The Theory of Screening, Education, and the Distribution of Income', *American Economic Review*, 65: 283-300.

Toharia, 2003 L. Toharia. El mercado de trabajo en España, 1978–2003 ICE, 811 (2003), pp. 203–220

Toharia, L., Malo, M. (1999) The Spanish Experiment: pros and cons of flexibility at the margin. In: Andersen E. (ed.) Why deregulate labour market? Oxford University

Tsang, M. C., & Levin, H. M. (1985). The economics of overeducation. *Economics of Education Review*, *4*(2), 93–104.

Tsang, L. C., Russell, W., Rumberger, W., & Levin, H. M. (1991). The impact of surplus schooling on worker productivity.*Industrial Relations*, *30*(2), 209–228.

Verheul et al., 2006 I. Verheul, A. van Stel, R. Thurik, D. Urbano The relationship between business ownership and unemployment in Spain: a matter of quantity or quality? Estudios de Economía Aplicada, 24 (2006), pp. 435–457

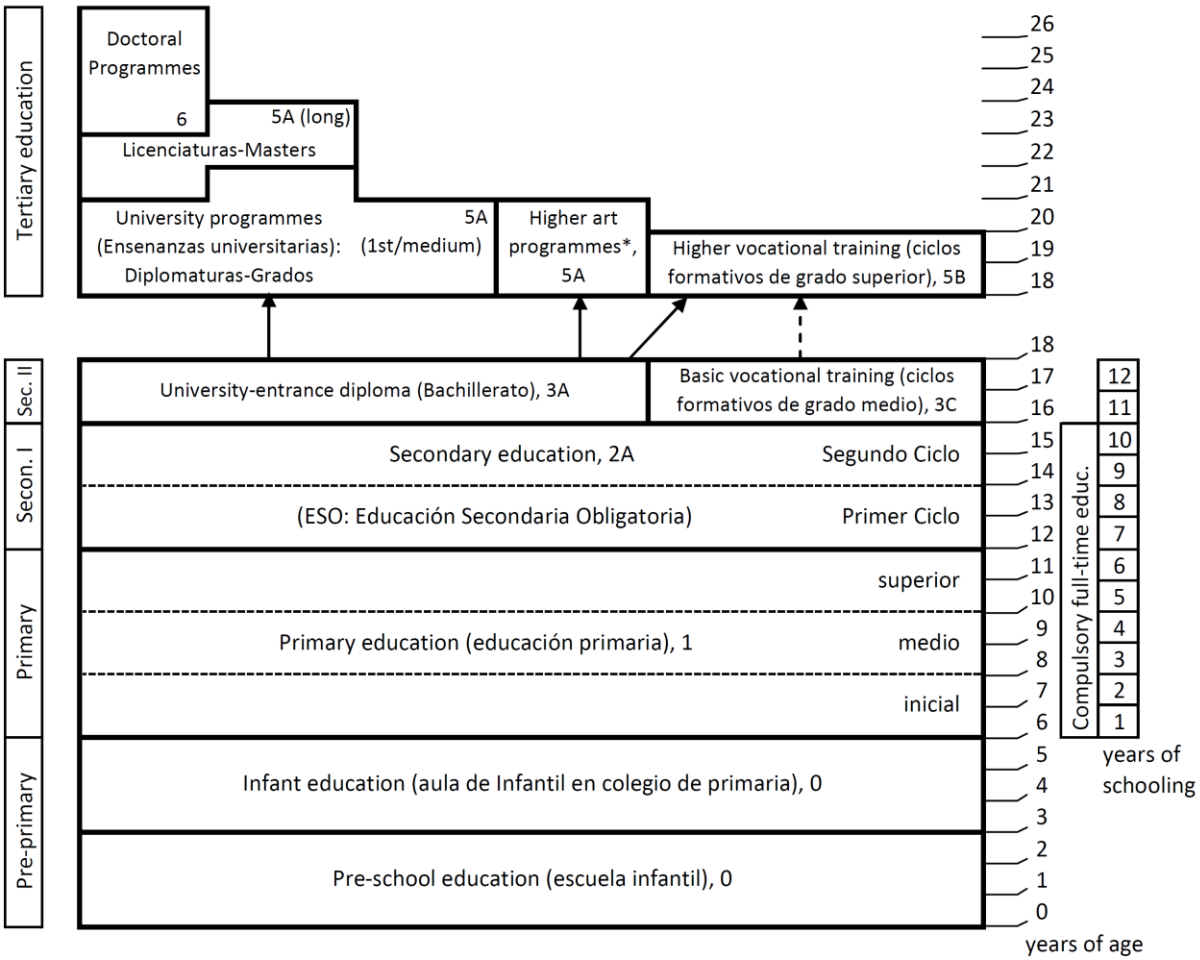
Verdugo, R. R., & Verdugo, N. T. (1989). The impact of surplus schooling on earnings. Some additional findings. *Journal of Human Resources, 24*(4), 629–643.

Vila, L.E. and Mora, J.G. (1998). 'Changing returns in education in Spain during the 1980s'. Economics of Education Review, 17(2): 173-178.

Weiss, Andrew (1995), 'Human Capital vs Signalling Explanations of Wages.' Journal of Economic Perspectives, 9: 133-54.

Appendix A

Figure 15. Spanish Educational System



The educational attainment variables used in this investigation can be visualized here. Primary education can be seen as being completed after 6 (compulsory) years. Lower secondary education can be seen as being completed after the tenth year. Upper secondary, which can also be interchanged, for medium grade vocational training both are completed after the twelfth year. Students can then choose from various programs if they wish to study further. Upper grade vocational training consists of two additional year of study. A diplomature, which is almost equivalent to a Bachelor's degree takes three years of study following upper secondary education. Following this are Master's and doctoral programs that are all included in the same variable for tertiary education in the analyses.

Appendix B

Table 7 . Occupational-Skill mismatch (percentages) from EAPS (1999-2013)

Average	Undereducated	Appropriately educated	Overeducated
1999	28%	40%	33%
2002	28%	39%	33%
2006	27%	41%	32%
2010	25%	40%	35%
2013	18%	39%	43%
Female			
1999	26%	39%	34%
2002	27%	38%	35%
2006	26%	40%	34%
2010	26%	40%	34%
2013	18%	39%	43%
Male			
1999	29%	40%	31%
2002	30%	40%	30%
2006	28%	43%	30%
2010	25%	40%	35%
2013	17%	39%	43%
Primary Education			
1999	28%	72%	0%
2002	29%	71%	0%
2006	28%	72%	0%
2010	22%	78%	0%
2013	16%	84%	0%

Source: EAPS, own elaboration

Table 7 . Occupational-Skill mismatch (percentages) from EAPS (1999-2013), continued

Lower Secondary Education	Undereducated	Appropriately educated	Overeducated
1999	8%	41%	52%
2002	10%	43%	47%
2006	11%	47%	42%
2010	20%	30%	50%
2013	5%	38%	57%
Upper Secondary Education			
1999	1%	17%	82%
2002	16%	10%	74%
2006	29%	1%	71%
2010	24%	13%	63%
2013	18%	0%	82%
Upper Grade Vocational Training			
1999	2%	26%	72%
2002	11%	9%	80%
2006	21%	0%	79%
2010	17%	20%	63%
2013	12%	0%	88%
Tertiary Education			
1999	0%	39%	61%
2002	0%	47%	53%
2006	0%	60%	40%
2010	0%	56%	44%
2013	0%	50%	50%

Source: EAPS, own elaboration

Appendix C

Figure 16. Link test for logit model (1999)

```
Iteration 0:   log likelihood = -36445.725
Iteration 1:   log likelihood = -27341.795
Iteration 2:   log likelihood = -27299.582
Iteration 3:   log likelihood = -27299.455
Iteration 4:   log likelihood = -27299.455

Logistic regression                             Number of obs   =      52589
                                                LR chi2(2)      =   18292.54
                                                Prob > chi2     =     0.0000
Log likelihood = -27299.455                     Pseudo R2       =     0.2510
```

mismatch	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_hat	.9821015	.0086932	112.97	0.000	.9650632	.9991398
_hatsq	.1013581	.0089878	11.28	0.000	.0837423	.118974
_cons	-.1547762	.0173607	-8.92	0.000	-.1888026	-.1207498

Figure 17. Goodness of fit test for logit model (1999)

Logistic model for mismatch, goodness-of-fit test

```
      number of observations =     52589
number of covariate patterns =      7814
          Pearson chi2(7804) =  11679.79
                Prob > chi2 =         0.0000
```

Figure 18. Logit -Area under ROC curve(1999)

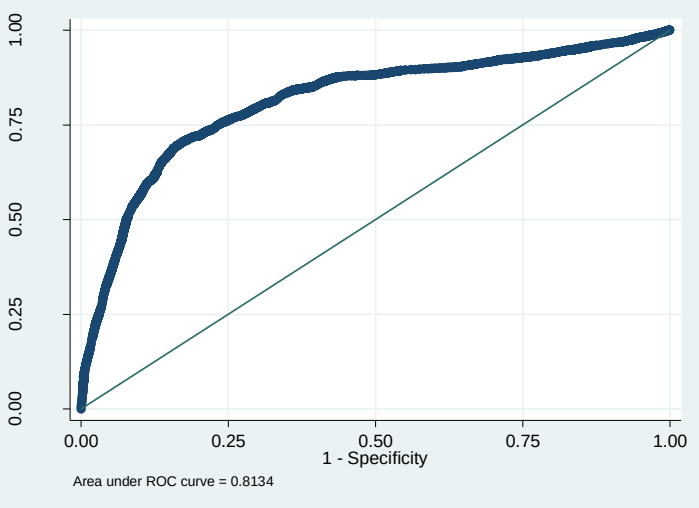


Figure 19. Robustness - comparison of marginal effects to probit model (1999)

```
Average marginal effects                        Number of obs    =      52589
Model VCE    : Robust

Expression   : Pr(mismatch), predict()
dy/dx w.r.t. : sec1 mvt ter female temp otherjob tenuremonths twentothrtyfour thirtyfivetosixfour
```

	Delta-method dy/dx	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sec1	-.547962	.0024283	-225.65	0.000	-.5527214	-.5432025
mvt	-.2080002	.0061271	-33.95	0.000	-.2200091	-.1959913
ter	-.3443647	.0042819	-80.42	0.000	-.352757	-.3359723
female	-.0423164	.0038219	-11.07	0.000	-.0498072	-.0348256
temp	.0182431	.0045058	4.05	0.000	.0094118	.0270743
otherjob	-.0574591	.0138598	-4.15	0.000	-.0846238	-.0302945
tenuremonths	-.0001009	.0000206	-4.90	0.000	-.0001412	-.0000605
twentothrtyfour	-.03043	.0072347	-4.21	0.000	-.0446097	-.0162502
thirtyfivetosixfour	.0059951	.0069819	0.86	0.391	-.0076892	.0196793

Figure 20. Robustness – Information Criterion of logit model (1999)

```
Akaike's information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	52589	-36445.72	-27363.05	10	54746.1	54834.81

Note: N=Obs used in calculating BIC; see **[R] BIC note**

Figure 21. Robustness – AIC, BIC Criteria of probit model (1999)

```
Akaike's information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	52589	-36445.72	-27367.68	10	54755.35	54844.06

Note: N=Obs used in calculating BIC; see **[R] BIC note**

Figure 22. Link test for logit model (2002)

```
Iteration 0:   log likelihood = -36469.355
Iteration 1:   log likelihood = -24333.284
Iteration 2:   log likelihood = -24216.776
Iteration 3:   log likelihood = -24216.122
Iteration 4:   log likelihood = -24216.122

Logistic regression                             Number of obs   =      52637
                                                LR chi2(2)      =   24506.47
                                                Prob > chi2     =     0.0000
Log likelihood = -24216.122                     Pseudo R2       =     0.3360
```

mismatch	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_hat	.968896	.0089082	108.76	0.000	.9514362	.9863558
_hatsq	.0520965	.0070726	7.37	0.000	.0382344	.0659587
_cons	-.1141989	.0193838	-5.89	0.000	-.1521904	-.0762075

Figure 23. Goodness of fit test for logit model (2002)

Logistic model for mismatch, goodness-of-fit test

```
       number of observations =     52637
number of covariate patterns =      8132
          Pearson chi2(8122) =     11636.53
                Prob > chi2 =         0.0000
```

Figure 24. Logit -Area under ROC curve(2002)

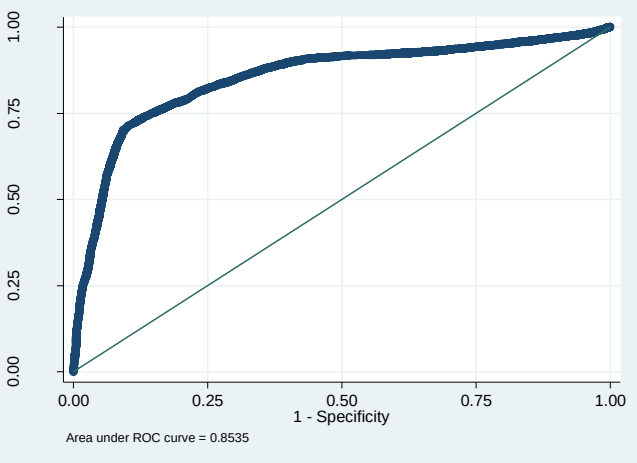


Figure 25. Robustness - comparison of marginal effects to probit model (2002)

```
Average marginal effects                        Number of obs   =      52637
Model VCE    : Robust

Expression   : Pr(mismatch), predict()
dy/dx w.r.t. : secl mvt ter female temp otherjob tenuremonths twentothrtyfour thirtyfivetosixfour
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
secl	-.5997192	.002359	-254.22	0.000	-.6043428	-.5950956
mvt	-.0540486	.0071386	-7.57	0.000	-.06804	-.0400571
ter	-.414213	.0032537	-127.31	0.000	-.4205901	-.4078359
female	-.0409617	.0034866	-11.75	0.000	-.0477953	-.034128
temp	.0122344	.0041265	2.96	0.003	.0041465	.0203222
otherjob	.0192529	.0025725	7.48	0.000	.0142108	.0242949
tenuremonths	-.0001295	.0000186	-6.98	0.000	-.0001659	-.0000931
twentothrtyfour	-.0134751	.0073917	-1.82	0.068	-.0279626	.0010125
thirtyfivetosixfour	-.0146895	.007027	-2.09	0.037	-.0284621	-.0009169

Figure 26. Robustness – Information Criterion of logit model (2002)

```
Akaike's information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	52637	-36469.35	-24243.08	10	48506.15	48594.87

Note: N=Obs used in calculating BIC; see **[R] BIC note**

Figure 27. Robustness – AIC, BIC Criteria of probit model (2002)

```
Akaike's information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	52637	-36469.35	-24261.7	10	48543.4	48632.11

Note: N=Obs used in calculating BIC; see **[R] BIC note**

Figure 28. Link test for logit model (2006)

```
Iteration 0:   log likelihood =  -34954.85
Iteration 1:   log likelihood = -16596.356
Iteration 2:   log likelihood = -15911.569
Iteration 3:   log likelihood = -15825.767
Iteration 4:   log likelihood = -15823.494
Iteration 5:   log likelihood = -15823.493

Logistic regression                             Number of obs   =      50583
                                                LR chi2(2)      =   38262.71
                                                Prob > chi2     =     0.0000
Log likelihood = -15823.493                     Pseudo R2       =     0.5473
```

mismatch	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_hat	1.008383	.0228973	44.04	0.000	.9635055	1.053261
_hatsq	-.0030053	.006618	-0.45	0.650	-.0159764	.0099658
_cons	.0196121	.0488674	0.40	0.688	-.0761662	.1153904

Figure 29. Area under ROC curve (2006)

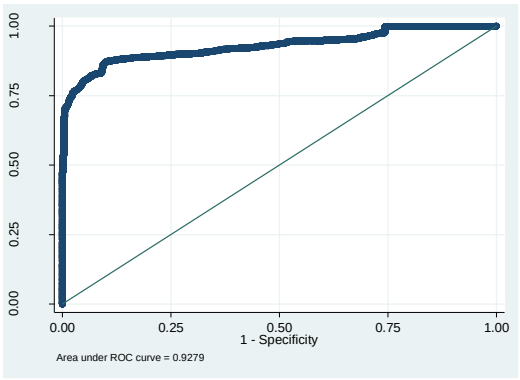


Figure 30. Goodness of fit test for logit model (2006)

Logistic model for mismatch, goodness-of-fit test

```
      number of observations =     50583
number of covariate patterns =      9214
          Pearson chi2(9204) =     14951.91
                 Prob > chi2 =         0.0000
```

Figure 31. Robustness - comparison of marginal effects to probit model (2006)

mismatch	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
prim1	0	(omitted)				
sec1	-3.606085	.04647	-77.60	0.000	-3.697164	-3.515005
mvt	.0858633	.084874	1.01	0.312	-.0804868	.2522133
ter	-2.776629	.0422361	-65.74	0.000	-2.85941	-2.693848
female	-.1686053	.0202179	-8.34	0.000	-.2082317	-.1289789
temp	-.0311096	.0213055	-1.46	0.144	-.0728677	.0106485
otherjob	-.0017391	.0124863	-0.14	0.889	-.0262119	.0227336
tenuremonths	.0005301	.0001011	5.24	0.000	.0003319	.0007283
sixteentotwenty	-.1340056	.0846687	-1.58	0.113	-.2999532	.0319419
twentothrtyfour	-.1081783	.0644235	-1.68	0.093	-.2344461	.0180895
thirtyfivetosixfour	-.1206626	.0612658	-1.97	0.049	-.2407414	-.0005838
indfoodtextwood	-.5718044	.143316	-3.99	0.000	-.8526987	-.2909102
extractind	-1.784729	.1390306	-12.84	0.000	-2.057224	-1.512234
machequiptrans	-.0480411	.1336272	-0.36	0.719	-.3099456	.2138633
construct	.5402443	.1338452	4.04	0.000	.2779125	.8025761
hotel	.7499426	.1325405	5.66	0.000	.4901679	1.009717
transport	.2282238	.1511001	1.51	0.131	-.0679269	.5243746
finanestat	.412401	.1323628	3.12	0.002	.1529748	.6718272
paedu	.3013593	.1331829	2.26	0.024	.0403256	.562393
otherservices	.1595403	.1337551	1.19	0.233	-.1026149	.4216956
_cons	2.278252	.1755741	12.98	0.000	1.934133	2.622371

Figure 32. Robustness – Information Criterion of logit model (2006)

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	50583	-34954.85	-15823.6	10	31667.19	31755.51

Note: N=Obs used in calculating BIC; see **[R] BIC note**

Figure 33. Robustness – AIC, BIC Criteria of probit model (2006)

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	50583	-34954.85	-15850.49	10	31720.97	31809.28

Figure 34. Link test for logit model (2010)

```
Iteration 0:   log likelihood = -36687.514
Iteration 1:   log likelihood = -29483.495
Iteration 2:   log likelihood = -29474.215
Iteration 3:   log likelihood = -29474.206
Iteration 4:   log likelihood = -29474.206

Logistic regression                               Number of obs   =      53185
                                                  LR chi2(2)      =   14426.62
                                                  Prob > chi2     =     0.0000
Log likelihood = -29474.206                       Pseudo R2       =     0.1966
```

mismatch	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_hat	.9954272	.0104329	95.41	0.000	.9749792	1.015875
_hatsq	.0123261	.0116612	1.06	0.291	-.0105294	.0351815
_cons	-.0158262	.0182061	-0.87	0.385	-.0515096	.0198572

Figure 35. Goodness of fit test for logit model (2010)

Logistic model for mismatch, goodness-of-fit test

```
      number of observations =     53185
 number of covariate patterns =      9280
          Pearson chi2(9270) =     12436.65
                 Prob > chi2 =         0.0000
```

Figure 36. Logit -Area under ROC curve(2010)

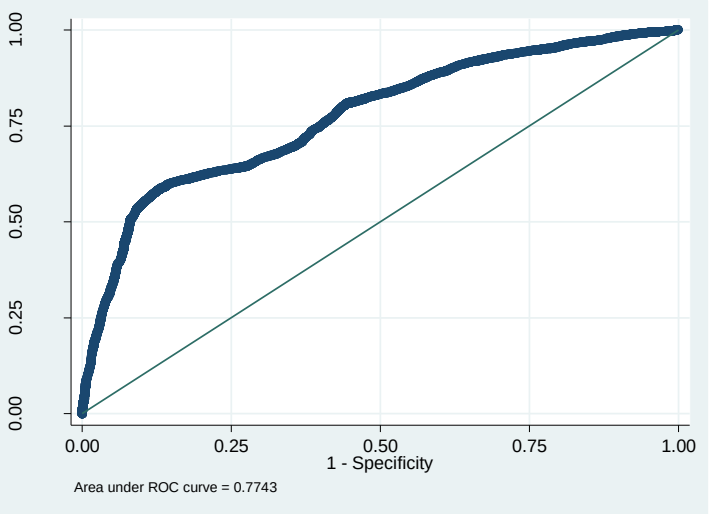


Figure 37. Robustness – Information Criterion of logit model (2010)

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	53185	-36687.51	-29474.76	10	58969.53	59058.34

Note: N=Obs used in calculating BIC; see **[R] BIC note**

Figure 38. Robustness – AIC, BIC Criteria of probit model (2010)

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	53185	-36687.51	-29465.22	10	58950.43	59039.25

Note: N=Obs used in calculating BIC; see **[R] BIC note**

Figure 39. Robustness - comparison of marginal effects to probit model (2010)

Average marginal effects Number of obs = 53185
Model VCE : Robust

Expression : Pr(mismatch), predict()
dy/dx w.r.t. : sec1 mvt ter female temp otherjob tenuremonths twentothrtyfour thirtyfivetosixfour

	Delta-method dy/dx	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sec1	-.4142112	.0044537	-93.00	0.000	-.4229403	-.4054821
mvt	-.1813189	.0069305	-26.16	0.000	-.1949025	-.1677354
ter	-.5442481	.0034392	-158.25	0.000	-.5509887	-.5375074
female	-.0280152	.0037776	-7.42	0.000	-.0354191	-.0206114
temp	-.0173118	.0048992	-3.53	0.000	-.0269141	-.0077095
otherjob	.0100739	.0026092	3.86	0.000	.00496	.0151879
tenuremonths	-.0001192	.0000178	-6.69	0.000	-.0001542	-.0000843
twentothrtyfour	-.05906	.0091298	-6.47	0.000	-.076954	-.0411659
thirtyfivetosixfour	-.0468559	.0082268	-5.70	0.000	-.0629801	-.0307316

Appendix D

Table 8. Effect of Occupational match/mismatch on wages for males (1995-2010)

Males	1995	s.e.	2002	s.e.	2006	s.e.	2010	s.e.
Undereducation	-0.070	0.001	-0.022	0.003	0.015	0.001	-0.007	0.001
Appropriate years	0.041	0.000	0.037	0.001	0.022	0.000	0.029	0.000
Overeducation	0.067	0.001	0.120	0.004	0.075	0.001	0.112	0.002
Experience	0.066	0.001	0.037	0.002	0.043	0.001	0.044	0.001
Experience^2	-0.001	0.000	0.000	0.000	0.000	0.000	-0.001	0.000
Constant	0.789	0.010	1.363	0.025	1.479	0.010	1.612	0.012
R^2	0.15		0.11		0.11		0.11	
Standard Deviation	0.58		0.65		0.63		0.61	
F-Statistic	4730.24		566.24		3552.14		2788.81	
Observations	131534		22049		139429		116654	

Table 9. Effect of Occupational match/mismatch on wages for females (1995-2010)

Females	1995	s.e.	2002	s.e.	2006	s.e.	2010	s.e.
Undereducation	-0.046	0.002	-0.018	0.005	0.011	0.001	-0.018	0.001
Appropriate years	0.030	0.001	0.029	0.001	0.019	0.001	0.030	0.001
Overeducation	0.049	0.002	0.058	0.005	0.023	0.002	0.111	0.002
Experience	0.045	0.001	0.031	0.002	0.032	0.001	0.036	0.001
Experience^2	-0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Constant	0.949	0.016	1.445	0.030	1.595	0.012	1.541	0.013
R^2	0.08		0.07		0.05		0.08	
Standard Deviation	0.61		0.66		0.66		0.58	
F-Statistic	675.80		157.01		832.52		1361.42	
Observations	39050		12139		80974		75412	

Table 10. Effect of educational attainment on wages, for females (1995-2010)

Female	1995	s.e.	2002	s.e.	2006	s.e.	2010	s.e.
sec1	0.170	0.008	0.165	0.017	0.163	0.007	0.145	0.007
sec2	0.635	0.010	0.547	0.019	0.452	0.008	0.419	0.008
Mvt	0.482	0.014	0.452	0.024	0.414	0.010	0.366	0.009
Uvt	0.656	0.012	0.654	0.023	0.577	0.009	0.539	0.009
diplomature	0.833	0.014	0.959	0.019	0.865	0.008	0.731	0.009
universitygraduate	1.197	0.013	1.202	0.019	1.058	0.008	1.006	0.008
experience	0.063	0.001	0.050	0.002	0.047	0.001	0.042	0.001
experiencesqr	-0.001	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000
_cons	0.464	0.015	0.775	0.025	0.903	0.011	1.176	0.011
R squared	0.29		0.36		0.28		0.27	
Standard Deviation	0.54		0.54		0.58		0.52	
F-Statistic	1982.93		1982.93		3483.57		3435.59	
Observations	39050		39050.0		80974.0		75412.0	

Table 11. Effect of educational attainment on wages, for males (1995-2010)

Male	1995	s.e.	2002	s.e.	2006	s.e.	2010	s.e.
sec1	0.142	0.004	0.153	0.011	0.155	0.004	0.145	0.005
sec2	0.581	0.005	0.521	0.015	0.460	0.006	0.434	0.006
Mvt	0.435	0.007	0.454	0.018	0.449	0.007	0.403	0.007
Uvt	0.616	0.006	0.647	0.015	0.605	0.006	0.592	0.006
diplomature	0.898	0.007	0.933	0.016	0.794	0.007	0.815	0.007
universitygraduate	1.196	0.006	1.205	0.014	1.030	0.006	1.072	0.006
experience	0.066	0.001	0.046	0.001	0.049	0.001	0.048	0.001
experiencesqr	-0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
_cons	0.655	0.008	1.056	0.020	1.117	0.008	1.297	0.010
R squared	0.36		0.36		0.29		0.31	
Standard Deviation	0.50		0.55		0.56		0.54	
F-Statistic	9446.4		1538.5		7087.0		6442.1	
Observations	131534.0		22049.0		139429.0		116654.0	