Working papers series

WP ECON 09.15

Are Within-Groups `Abilities' Distribution Constant on Time?

Manuel Hidalgo (U. Pablo de Olavide)

JEL Classification numbers: J31

Keywords: ability distribution, residual wage inequality, education

Department of Economics

Are Within-Groups 'Abilities' Distribution Constant on Time?

Manuel Hidalgo-Pérez*

Universidad Pablo de Olavide

November 24, 2009

Abstract

The analysis of wage inequality has had a lot of tradition since the early nineties, in particular seeking an explanation for residual wage inequality defined as the inequality which is not explained by observables characteristics. However, some assumptions made in order to delve into the causes of wage inequality evolution are far from realistic, as the one assuming constancy in the distribution of non-observed characteristics within each cohort of workers employed (ability distribution). In this work, this latter hypothesis is relaxed. The main contribution is that this paper explicitly considers and values the change in ability distribution and its effects on total wage inequality. The conclusion is that non considering the changes in ability distribution may overvalued the effects of other possibles causes, as for example changes in prices paid to unobserved skills. I use Spanish data to evaluate this new approximation.

JEL: J31

Keywords: ability distribution, residual wage inequality, education

*Universidad Pablo de Olavide, ctra de Utrera, km 1, s/n, CP 41013. Teéfono: 0034 954 97 79 79, email: mhidper@upo.es.

1 Introduction

Wage inequality has changed in some countries since the seventies, especially the United States (Blackburn et al., 1990; Katz and Murphy, 1992; Levy and Murname, 1992). The causes for this increase in the United States remain unclear because the Mincerian equations used fail to explain at least half of the change when the explanatory variables are observable characteristics of workers, such as education, gender or experience (Mincer, 1974). The other half, which is called residual wage inequality, remains unexplained, and thus the analysis of variables contributing to such inequality is important (Juhn et al.1993).

Residual wage inequality corresponds to three factors: first, the existence of non observed characteristics, which many believe largely account for unmeasured productivity or 'ability' of workers (Chay and Lee, 2000)[1] ; second, the prices paid for these characteristics in the market; and third, measurement errors. Therefore, the literature provides three possible reasons why the residual wage dispersion can change: changes in 'ability' dispersion, changes in prices dispersion or changes in the variance of measurement errors.

No consensus has been reached on the role played by each of these causes. While Juhn et al (1993) consider that the price effect is the most relevant for United States, Lemieux (2006) argues that the change in the dispersion of 'abilities' is relevant enough in to explain the increase in residual wage dispersion. This argument is based on the fact that changes in the workforce composition change residual dispersion mechanically. For example, if the proportion of more highly educated workers increases at the expense of those with lower educational levels, the total dispersion is increased. This occurs because the former have greater dispersion of 'abilities' than the latter within their respective groups. As Juhn et al (1993) did not implicitly consider the composition effect, the role played by prices was overvalued according to

Lemieux, assuming that variance in measurement errors is constant over time. By removing the variable of the composition effect, Lemieux found said effect to be an important factor in explaining the increase of residual wage dispersion.

However, none of these authors have paid attention in the distribution of 'abilities' within cells narrowly defined by observed characteristics–what is called within-group ability distribution– and its effect on residual wage dispersion. Whereas Lemieux's composition effect assumes that residual wage dispersion depends on the weights of individual cells (composition), it also sustains that dispersion of 'abilities' within cells is constant over time.[2] But theoretically and empirically, the dispersion of abilities changes within each education- and experience-based group of workers that has been described. Through the use of several models, Galor and Moav (2000) and Gould, Moav and Weinberg (2001), show that the ability of different models within each educational group can increase within-group inequality. Especially, Gould, Moav and Weinberg's Model (2001) generates two inequality sources, including ability and the differential effect of technology in diverse educational groups. They show that wage inequality within educational groups may not only result from payment for ability but also from changes at its internal composition. Empirically, Gottschalk and Moffitt (1994) find that intra-group inequality changes in the 70s and 80s, and these changes are different by education groups. Hoxby et al. (1999) found that, among workers with a university education, the increase in wage inequality can be broken down into three factors: the increasingly diverse background of people who attend college, in the increase of returning to ability and thirdly the change in market structure of college education. Hoxby (1997a, b) finds that, in the U.S., differences between "abilitys" of students within each college fall, while the differences between students of different universities grow. This is what Hoxby calls intensive margin affecting wage inequality of workers with university degrees.

Then, the aim of this study is to incorporate into the debate of the explanation of residual wage dispersion the changes in the dispersion of abilities within each education- and experience-based cell. Specifically, this paper expands on Juhn et al. and Lemieux's analysis of changes in residual wage inequality. This extended analysis will cover four, instead of three, possible reasons for residual wage dispersion: changes in prices, measurement error, composition and the distribution of 'abilities' within education and experience cells. These four change factors will be called price effect, error effect, composition effect and distribution effect from now on. In implementing this extended analysis, we set forth several hypotheses. Let's suppose that each year "new" workers with different 'abilities' distributions are incorporated into the market, different from those who agreed in previous years, and referred to as "old" workers. Let's suppose that these differences between new and old workers continue to exist even if we compare them by education and experience cells. In this way, the within group dispersion of cells that the so called new workers' receive changes from year to year even if all cell weights and the prices paid for ability remain constant. This change will be more intense in cohorts with a high percentage of new workers. If, to the contrary, no new worker enters a particular cohort, the distribution of 'abilities' remains constant. Furthermore, when these workers begin to change cell values– for example, as their experience grows–they become "new" to the cell to which they have moved, once more changing within-group 'abilities' dispersion. In summary, changes in the dispersion of 'abilities' within each cohort are the result of the entry of new workers into the labor market and the subsequent spillover into the remaining cohorts.

To perform this breakdown, we need information on a set of workers, including when they entered the market and how long they have been working. To explore this approximation I use information for Spain, from the Spanish database called Continuous Sample on Working Lives (Muestra Continua de Vidas Laborales

in Spanish, and MCVL head on). In addition, this database allows to organize workers by the moment in which they begin to work in the market.

The results are as follow: Once Mincerian regressions are estimated, we observe a continuous increase in residual wage inequality for Spain.[3] Using Lemieux's decomposition, it can be observed that both price and composition effects are important. Nonetheless, the price effect gives rise to a paradox: while prices for non-observed abilitys increase residual inequality, prices for observed abilities (return to education) decreased wage inequality. By using the new decomposition, I found firstly that the within-group 'abilities' distribution increased wage inequality since the nineties, and secondly, that prices paid for 'abilities' decreased inequality. Thus the mystery is solved.

The rest of the study is arranged in the following manner: Section 2 explains the RWI decomposition method developed by Lemieux (2006). Section 3 proposes the extension of Lemieux's method when it is assumed that the distribution of the observed features is not constant within cohorts. In section 4 the MCVL, or the database used for the Spanish study, is described. In sections 5 and 6, Lemieux's decomposition and the new decomposition are applied to MCVL data in Spain. Section 6 shows the robustness of this analysis by adding the observed variables. Section 7 concludes.

2 Residual Wage Inequality: Skill Prices, Composition Effects, and Measurement Error. The "Lemieux" Specification

Many essays have been published regarding changes in wage inequality. Following an increase in wage inequality since the seventies in countries like the U.S. and United Kingdom, works have emerged that

attempt to explain these changes. The growth in residual wage inequality is of special interest, as it explains at least half the growth in wage inequality in the United States (Juhn et al., 1993).

The literature assumes that residual wage inequality changes because there is a change in skill ranges among workers, compensation for such abilities, or measurement error. According to Juhn et al (1993) the main reason for increasing residual wage inequality is the increase in payment for skills, or price effect. Thus, the authors assume that measurement error and the range in workers' skills remain constant over time. While the first assumption is explicit, the second is implied by the method and arguments (Lemieux, 2006). Lemieux criticizes this conclusión because it raises at least two questions. Firstly, given the correlation of ability with education, increase in intragroup inequality should coincide in time with an increase in inequality between groups. These groups are defined by workers' education level. While residual inequality began to increase in the early seventies, between-group inequality did not begin to do so until the close of the decade. Secondly, one would expect to note an increase in the wage gap based on gender or race, but such a trend was not observed.[4] Therefore, Lemieux (2006) extends the analysis beyond the price effect. According to Lemieux, the composition effect–changes in the weight of the different cells defined by education and experience values—explains much of the change in residual wage inequality. If we assume, for instance, that higher within-group educational level corrrelates with higher within-group inequality, an increase in the weight of the cells for workers with higher education will then increase, by composition, wage dispersion or inequality. To conclude, the price effect has also been important. Furthermore, when the effect of composition is discounted, the same price effect shows a correlation in time with the changing salaries paid for education and skills.

Lemieux's decomposition method for the residual wage inequality in workers is summarized below (Lemieux,

2006).

2.1 The Lemieux decomposition

To remove the effect of composition changes in residual wage inequality, Lemieux uses a Mincer type wage equation:

$$w_{it} = x_{it}b_t + \varepsilon_{it}, \tag{1}$$

where w_{it} is the natural logarithm of wages for individual i at time t; x_{it} is a vector of observed variables (education, age and gender); b_t is a vector with payments to the observed features; and ε_{it} is the error of regression and dispersion whose analysis underlies the study of residual wage dispersion. Both Juhn, Murphy and Pierce (1993) and Lemieux (2006) assume that the residual of the regression, ε_{it}, is the product of e_{it} abilities and return which are payed, p_t, and a measurement error v_{it}. Thus the regression residuals can be described through a model of components with the following specification:

$$\varepsilon_{it} = p_t e_{it} + v_{it}. \tag{2}$$

Assuming $E(e_{it}v_{it}) = 0$, the variance of the residual can be represented as

$$Var(\varepsilon_{it}) = p_t^2 Var(e_{it}) + Var(v_{it}). \tag{3}$$

The equation (3) shows that the variance of the residuals can be explained by changes in the prices paid to the abilities, p_t^2, by changes in the unconditional variance or dispersion thereof $Var(e_{it})$, or by changes in the dispersion of the error of measurement, $Var(v_{it})$. Assuming that the latter is constant in time, the dispersion of waste can only change because price and the dispersion of abilities do. This is the hypothesis

made by Lemieux (2006) contradicting the assumption JMP. They argue implicitly that $Var(e_{it})$ is constant in time (Lemieux, 2006), so by default, for JMP the evolution of residual wage dispersion in the United States from the seventies reproduces the trend of prices paid to the abilities.

The literature and evidence show that there is heterogeneity in wage dispersion between different cohorts defined by education, age and gender. In other words, the conditional variance of the residuals for education, age and gender is different for each cohort defined by these variables. For example, to higher education and older age, greater dispersion. Theory explains this evidence. For example, the work of Mincer (1974) and Chay and Lee (2000) shows that wage dispersion between groups is not the same if we define them by their education level. Also Mincer (1974) and Farber and Gibbons (1996) theoretically demonstrate using On-The-Job-Training and Learning models that the more experience, the greater wage dispersion. Therefore, changes in workforce composition in favor of groups with higher education and / or experience must have increased dispersion of residualss and wages mechanically.

The explanatory power of composition change on the behavior of the residual wage inequality is what Lemieux intends to prove. To do so it represents the dispersion of abilitiess in a time t as:

$$Var(e_{it}) = \sum_j \theta_{jt}\sigma^2_{jt}, \tag{4}$$

where σ^2_{jt} is the conditional variance of e_{it} skills to a group of j, where j represents the potential cohort groups for possible combinations of J education-age-gender values or $\sigma^2_{jt} = Var(e_{it}|x_{it} \in J)$ and θ_{jt} is the percentage of workers belonging to cohort j at time t.

According to (4) unless $\sigma_{jt} = \sigma_{kt}$, $\forall j \neq k$, any change in the weights θ_{jt} will change dispersion of e_{it}.

Following the Lemieux notation, conditioned variance or intra-group variance of cohort j residuals is

defined as:

$$V_{jt} = p_t^2 \sigma_{jt}^2,$$

so residual wage dispersion can be expressed as the sum of the conditional variances of the respective weighted residuals weight, plus the variance of measurement errors, ie:

$$Var(\varepsilon_{it}) = \sum_j \theta_{jt} V_{jt} + Var(v_{it}). \tag{5}$$

At this point Lemieux takes two assumptions. Firstly it assumes that the variance of measurement errors is constant over time.[5] Secondly, it assumes that the conditional variance of abilities for a cohort is constant over time. That is to say, $\sigma_{jt}^2 = \sigma_j^2$, $\forall t$. This implies that $V_{jt} = p_t^2 \sigma_j^2$, so that changes in the conditional variance of the residuals for the cohort j can only be due to changes in prices.

Given these hypothesis, we define counterfactual dispersion of the residualss at t moment with the weights at time $t-1$ as $Var(e_{it})^* = \sum_j \theta_{jt-1} V_{jt}$. Based on the change in the variance of the residualss from a time $t-1$ and a moment t can be expressed as:

$$\triangle Var(\varepsilon_{it}) = \sum_j (\theta_{jt} V_{jt} - \theta_{jt-1} V_{jt-1}).$$

Adding and substracting $Var(e_{it})^*$, we obtain:

$$\triangle Var(\varepsilon_{it}) = \sum_j \theta_{jt-1}(V_{jt} - V_{jt-1}) + \sum_j (\theta_{jt} - \theta_{jt-1}) V_{jt}. \tag{6}$$

Developing, the first term of this expression can be expressed as

$$\sum_j \theta_{jt-1}(V_{jt} - V_{jt-1}) = \sum_j \theta_{jt-1}(p_t^2 - p_{t-1}^2)\sigma_j^2,$$

and responding to changes in the prices paid to the abilities (the price effect). The second term is expressed as

$$\sum_j (\theta_{jt} - \theta_{jt-1})V_{jt} = \sum_j (\theta_{jt} - \theta_{jt-1})p_t^2\sigma_j^2,$$

and whose change explains the changes in the weights of the cohorts, ie the composition of the labor force (composition effect).

What equation (6) tells us is that the change in residual wage dispersion can be motivated by the price effect and the composition effect. According to Lemieux (2006) this second component has had a major role in the case of the United States so that, once removed, the change in residual wage inequality lies in time and manner with changes in other measures such as the college wage premium.

Then, using data from the MCVL for Spain in the period 1990-2006, the aforementioned exercise is going to be reproduced for the Spanish case. But first, we need to describe the database that is used as the manipulation of the same, which becomes necessary to achieve the objectives of this work.

3 Residual Wage Inequality and Non-Constant Ablity Distribution

Let's go back to the expression (6). This result is partly a consequence of assuming that V_{jt} and V_{jt-1} only change if prices paid to non-observed characteristics do. As $V_{jt} = p_t\sigma_j^2$ and $V_{jt-1} = p_{t-1}\sigma_j^2$, changes in time

only have their possible origin at p_t and p_{t-1}. This section shows how to relax this assumption, allowing the conditional variances of the cohorts j to be affected by changes in the distribution of those non-observed.

In section 2, it was assumed that the distribution of non-observable characteristics for each cohort j, σ_j^2 was constant over time. But now the variance for the cohort in one year will be assumed to be the same as in the preceding period but only for those workers who already belonged to labor market and the previous cohort at the time. By contrast, workers who enter job market or change their cohort during the year, will not do it necessarily with a distribution of non-observed features corresponding to the one the cohort had in $t-1$. The new workers will modify the distribution of non-observable in those cohorts they have access to, forcing it to change from year to year. This way constancy assumption is eliminated given that $\sigma_{jt}^2 \neq \sigma_{js}^2$, $\forall t \neq s$.

An example will help understand this assumption. Let's suppose year 0. In that year cohorts are defined according to observed features, qualification (C), age (E) and sex (S), and that we denote as the cohort j where $j \in \{1, ..., J\}$, being J the number of possible combinations of skills, age and sex. For this moment each cohort has its σ_{j0}^2 dispersion. If there were no entries of new workers into the market, and assuming that once in the market, the distribution of non-observable characteristics is held constant for each cohort regardless of who had access or not to them, the effect of changes in the dispersion of the non-observed characteristics will not only be motivated by the composition effect, as Lemieux assumes. Suppose, however, that in the year 1, new workers enter the labor market. They will enter the cohorts defined by their cualification, age and gender. What is now assumed is that these new workers will not have the same non-observed dispersion characteristics that those who belonged to the cohort they enter in the year 0. We assume the same for those workers who have access to new cohorts and belong to another cohort in the previous year. In this way, as

years pass, workers with new distributions for characteristics observed are accessing other cohorts to which they alter its previous distribution. This assumption does not negate the fact that among and workers not belonging to the market keep their dispersion while keeping cohort unchanged. In conclusion, the dispersion of each of the cohorts will change as new workers enter at that time.

Thus, at the time 1 of the cohort j, there are two types of workers. Those workers who were already in the labor market and in that cohort before that year and who maintain the same intra-group dispersion of the previous year, and those entering the job market or the cohort that year. In other words, if there were no entry of employees in or changes between cohorts there would not be changes in the dispersion of the same and any change in the dispersion of non-observable features would be produced by changes in composition, as Lemieux assumed (2006).

This course reveals how restrictive it was to assume that intra-group dispersion of observed characteristics observed are constant over time. Automatically it assumes that any worker can enter a cohort in which he/she shares stochastic structure with those who were already at this cohort. In turn, with the passing of years, it is assumed that the same cohort of workers are not affected in terms of intra-group distributions by major changes that have occurred in the rules, in society or customs that may affect the distribution of abilities.

What is the basis for assuming that new workers have different distributions in their non-observable characteristics? Let's think about the youngest workers to explain it easier. In this case, for example, changes in the education system significantly altering the distribution of qualities at schools will alter the distribution of abilities for those who are younger in the job market every year, the universalization of education, legislative changes... Therefore the distribution of non-observable in the lower age groups may change each year. It would be difficult to understand that distribution of cohorts with less age remained

constant over the decades or lustrums. However, it is easier to consider that the distribution of those who have been at the labor market for longer time can be kept more stable, since the entry of new workers in higher age groups may be minimal.

Thus, let's suppose that in moment for a specific education-age-sex j cohort, there are two different distributions for non-observed characteristics. The first is the one of those already in the labor market in the previous period and who belonged to this cohort and whose dispersion is defined by

$$V_{jt}(o) = p_t^2 \sigma_{jt}^2(o),$$

where o makes references to those workers in the period t who already at moment $t-1$ belonged to labor market and that cohort. Therefore, we assume that $\sigma_{jt}^2(o) = \sigma_{jt-1}^2 \ \forall t$.

The second one is that of workers who at the moment enter labor market or the aforementioned cohort from another one, and whose dispersion is

$$V_{jt}(n) = p_t^2 \sigma_{jt}^2(n),$$

And we assume that $\sigma_{jt}^2(n) \neq \sigma_{jt}^2(o) \ \forall j$ and for each moment in time t.

It is assumed that p_t is common both for the ones who enter for the first time and those already remaining at the market.

Therefore, the unconditional variance at the time of a cohort of workers j, V_{jt}, is the combination of two different variances: the workers that were already in the market before such moment and those entering for the first time. More specifically, if we assume that the difference in the distribution of abilities lies only at its dispersion, but not in its average, it can be shown that (see Annex A)

$$V_{jt} = \gamma_{jt}(n)V_{jt}(n) + (1 - \gamma_{jt}(n))V_{jt}(o), \tag{7}$$

where $\gamma_{jt}(n)$ is the weight represented by new workers over the total of workers at that moment t in the cohort j.

Rescuing from (6) the change that $\sum_j \theta_{js}(V_{jt} - V_{jt-1})$, according to Lemieux is affected to the price, this one can now be due to 2 main reasons. First, because price changes, as Lemieux suggests, or that distribution of non-observed characteristics change. From (6) and adding and subtracting $\sum_j \theta_{js}V_{jt}(o)$, we obtain that:

$$\begin{aligned} \sum_j \theta_{js}(V_{jt} - V_{t-1}) &= \sum_j \theta_{js}V_{jt} - \sum_j \theta_{js}V_{jt}(o) \\ &\quad + \sum_j \theta_{js}V_{jt}(o) - \sum_j \theta_{js}V_{jt-1}. \end{aligned} \tag{8}$$

Assuming that $\sigma^2_{jt}(o) = \sigma^2_{jt-1}$ $\forall t$ and (7), decomposition (8) can be represented as:

$$\sum_j \theta_{js}(V_{jt} - V_{jt-1}) = \sum_j \theta_{js}p_t^2\gamma_{jt}(n)(\sigma^2_{jt-1} - \sigma^2_{jt}(n)) + \sum_j \theta_{js}(p_t^2 - p_{t-1}^2)\sigma^2_{jt-1}. \tag{9}$$

The first part on the right of the expression (9) represents the change in the dispersion of residualss motivated by the change in the distribution of non-observed characteristics (distribution effect). The second part explains the change determined by changes in prices paid to such characteristics (price effect)

To obtain calculation of (8), availability of a panel with information over time for different workers is mandatory. Of course it is necessary to distinguish between those workers who at the time were already in the labor market with respect to those who enter at the moment and those changing their cohort. The

Continuous Sample on Working Lives (MCVL) provides this information. Given that it has knowledge regarding the working life of an employee from the moment when the worker enters for the first time the labor market, it allows not only to identify their working lives, but when he/she came to it, in which cohort, etc..This offers the possibility of estimating, at each given moment, the dispersion of wages for these workers and those who have entered the job market this year so that calculation shown in the expression (8) is feasible. Once you have identified the workers with respect to the time when they were entering the labor market, we can identify which part of the dispersion, not motivated by the composition effect, is due to price changes and changes in the distribution of non-observed characteristics.

4 The Continuous Sample on Working Lives

The Continuous Sample on Working Lives (MCVL in spanish) is a database derived from the records of the Social Security under the Ministry of Labor, incorporating in turn information from the Padrón Municipal Contínuo. It is being done annually since 2004. Samples from the year 2006 also includes tax information. It is a sample, because from the total number of people who, during the reference year, had some connection with the Social Security (contributions, pensions, etc. ...), and which in 2006 amounted to more than 20 million, it will select a number of them, not less than one million. In total one out of each four people. It is representative of the population included in the complete records of Social Security because the selection aims to reproduce the structure given by the set of records. It is continuous because the MCVL follows the evolution of selected individuals throughout their working life, both in labor relations and benefits received. That is, there are so many different records for each individual, so the MCVL is also a panel of data.

The MCVL offers a very broad set of variables, up to more than eighty different. These include variables

identifying the person in question, eg date of birth and, in its case, death, sex, education, nationality, province, where he joined for the first time, account code contribution, group contribution, the company contribution code ... Variables related to the job as the contribution scheme are also included, the register and leave dates, both real and effective, contract type and type of labor day, percentage of time worked compared to a full day, the group of contributions, etc ... In addition, for each of the working relations included for each individual, the basis for which the individual has contributed each month is provided. Out of that basis, salaries and unemployment benefits, if any, can be obtained. The MCVL completes the information with data from the employer or responsible for contributions such as economic activity, firm size, seniority as an employer, location and type of employer.

Given the enormous amount of information, it becomes necessary to select a sample set of data provided by MCVL conforming to the reference population for this study. MCVL has the information of members existing due to the existence of an employment relationship as well as those pensioners or other members who are simply listed to maintain certain rights to future receivables. For this analysis, only those appearing in the records because of an employment relationship are interesting for our purposes. Moreover, from these ones, those whose relations are not attached to the General System of Social Security, the special scheme for coal mining or the Household Workers, are eliminated. It excludes, for example, contributors belonging to the Agricultural System. Furthermore, workers belonging to certain groups of special contributions, such as under-18,[6] beneficiaries of war, self-employed, etc..have also been eliminated. Specifically the groups included are those between Group 1 and Group 10. Workers from the agricultural sector have not been included, given the particular case of this sector, and that would not have been removed to exclude the agricultural system. With regard to the employment relationship, some special cases as those contributors

who have some peculiarities that make them not be registered or those with a learning contract. Workers hired through temporary employment agencies were also eliminated. Finally, those workers with a lack of information that might be relevant for the subsequent analysis have also been eliminated.

These considerations made, we selected 200,000 members that met these requirements at some time during 1990 and 2006. From all of them, their complete set of relations in the period, as well as contributions which they have been bound to have been obtained. Specifically, from these 200,000 members, we take their working life meeting the requirements outlined in the previous paragraph. Of course, these relations can go from one, where the participant has not changed in any of its terms such a relationship since 1990, to hundreds of them when the member has a job profile that leads to changing it even several times a month, for instance, artists. A total of 1,478,735 relations, ie, 7.39 per member, is included.

From the totality of these relations, the contributions the affiliate has made have been obtained. Each affiliate has the obligation to contribute once a month due to laboral relation. Therefore, if an affiliate does not change working relationship in a whole year, he/she will have paid twelve months to the end of that year. In contrast, an affiliate who changed their employment relationship one or more times a month or several months in a year, will be over 12 annual contributions. Therefore, seeking not to increase too much information on the different contributions to all 200,000 members, only the price of October has been taken into account. Thus, if an affiliate was hired in October, left in unemployment in the same month and be employed again in October, he/she will appear on the show for the first and last year, but not so for the interval. This selection leads to provide a database that reaches 3.63 million.

Table 1 shows some descriptive statistics of the database. The average age of participants selected ranges from 34.4 and 35.9 years. The percentage of women increased from 32.4% in 1990 to 46.1% in 2006, with

values close to the national average and shown by other surveys, although they were slightly higher at the end of the period. For example, the Working Population Census estimated these percentages to the 31.4% and 40.53% respectively.

However, average study years show a downward bias with regard to the values observed for other works in Spain. Instituto Valenciano de Investigaciones Económicas (IVIE) calculates the average years of study in Spain between 1977 and 2007 period. For 1990 the value was 9.11 years of study while in 2006 this value amounted to 11.60. Therefore, it is clear that the sample does not capture this workers' feature properly. This is because such information comes from the municipality and corresponds to worker's response at a given point in time. Once this information is collected it is not changed so this variable may be a distant reality in many cases. For example, the weight of the college-educated workers in this sample is very low compared with other work surveys (García-Pérez, 2008). Because of that, the alternative has to be a cualification rather than education measurement. The qualification is measured by the contribution group to Social Security and can be a correct approach to the current worker's level of skills. Following García-Pérez (2008), the occupation groups 1 to 3 are considered high-level skills, group 4 to group 6 qualify as medium-high, groups 7 and 8 would be medium-low and finally groups 9 and 10 as low. According to this clasification, workers with high qualification. Finally, the last column shows the percentage of workers with full-time contracts. Clearly, an increase of workers with part-time contracts is observed, a result of recent laboral reforms.

As indicated, a problem with MCVL is that, technically, it provides no wages. Instead of these, MCVL offers the so-called contribution bases, which is the wage base on which to apply the tax rates for which each member contributed to the system of Social Security in Spain. In most cases this information is consistent with the worker's wages. However, the existence of limits both above and below the base rate provokes the

fact that there are workers whose salary information is censored. Usually, before this problem, the work done by chooses various options. The first one is eliminating these workers, given that, in the worst of cases, they do not go beyond a 10% of affiliates The second is to apply econometric methods in the estimates of wage equations. Here we propose to correct censorship using Tobit models. Thus censored wages are replaced for estimates that take into account the structure of the residualss. With this, we make good use of the information from all employees and minimize potential selection bias by eliminating those affiliated with maximum contributions.[7] Although there are no hours, as shown in Table 1, information is provided on the percentage of time worked over a full-time contract. Adjusting by that percentage, it is possible to homogenize the salaries of all full-time workers to make them comparable absolutely.

TABLE 1 HERE

Table 2 and Figure 1 show the first average wages estimated by the MCVL and other national surveys that provide information such as the Wage Structure Survey (ESS) and the Quarterly Wage Cost Survey (TCE), both from INE. As noted, although they do not show exactly the same values, they do converge to a significant degree. Figure 1 shows the growth rates of real wages estimated for the MCVL and the ECT, those derived from INE's National Accounting and those reflected in collective bargaining, collected from administrative sources in the Ministry of Labor and Immigration. In all three cases, there is a correlation between the wages of the Sample and the rest.

TABLE 2 HERE

FIGURE 1 HERE

Once data and wages are available it is interesting to replicate the performance of Lemieux (2006) on

decomposition of residual wage inequality evolution before explaining the extension of this method.

5 The Lemieux's Decomposition for the Spanish Residual Wage Inequality, 1990-2006

Figure 2 shows the evolution of the standard deviation of the residualss wage regression (1) when the explanatory variables of employees, qualifications, experience potential, potential experience squared and gender are included.[8] As shown, the timeline clearly shows that dispersion of residuals increased between 1991 and 1993, was moderated towards the end of the decade but did not stop growing, and doing it intensely until 2003. Finally, from then until 2006, it showed a continuing decline.

FIGURE 2 HERE

This evolution of Spanish residual wage inequality has already been analysed in other works and using other data sources. For example, using the Household Budget Survey for the years 1980 and 1990 Continuing Survey of Family Budgets for 1985, 1990, 1995 and 2000, Hidalgo (2008) obtains residualss from a Mincerian regression similar to (1). One result of this work is that the dispersion of the same grows since the late eighties. Pijoan-Mas and Sanchez-Marcos (2009) using data from the same source and comparing it to the Household Panel for Spain, found mixed evidence. With data from income of household heads of ECPF, they find that, while there is a clear decrease in the return to education, with increases in experience and stationary in sex, residual wage inequality shows a slight increase from 1991 to 1993, to be stabilize until the end of the decade. However, data from the Household Panel measuring hourly wages of workers, show a mixed trend in residualss, with continuing decline until 1998, while the rest shows a similar trend to that

obtained by another source. One explanation for this discrepancy lies in the different measures of income, total by household head in the ECPF and hourly wages in the Panel. In addition, the panel seems to fail in the determination of wages for workers with low wages (Pijoan-Mas and Sanchez-Marcos, 2009).

To verify that this evolution is not driven by the behavior of extreme and censored cases, percentiles 10, 50 and 90 have been estimated for the distribution of residualss and, this time, dispersion was measured as the distance between the aforementioned percentiles. Besides providing an extra measure for the analysis, these measures of dispersion can detect whether changes in the distribution of waste are being originated above or below the median.

Figure 3 shows the evolution of the distances between the percentiles 90 and 10 (solid line with triangles), between percentiles 50 and 10 (dot line with squares) and percentiles between 90 and 50 (solid line with large triangles). Again, these statistics show an increase in the dispersion, which is repeated on both sides of the median. However, it should be noted that, until 2002, the growth of residuals dispersion below the median was more intense. Moreover, from 2003, median residualss are the ones appearing to slow down and reduce dispersion.

FIGURE 3 HERE

Table 3 shows the standard deviation of the residualss of the equations and the wage gap between percentiles, all for only five years, 1990, 1994, 1998, 2002 and 2006. Column 1 shows standard deviation. The column "90-10" refers to the distance between percentiles 90 and 10, in the next column, the distance between the percentiles 50 and 10, while in the fourth and last column the distance between 90 and 50 percentiles. In rows 1 to 6 the values for the selected years are shown for each dispersion measurements while in rows 7 through 10 growth between these years is provided. In the last row of the table the growth

for the whole period is shown. According to data in the first column and, as shown in Figure 2, the standard deviation experiences continued growth during the 90s, changing to fall from 2002. The dispersion of wages in 2006 is the same as in 1994. The distance between 10 and 90 percentile for the whole period grew by 26 percentage points. This growth was more intense in 2002, the year in which the gap was 28 points higher than that shown in 1990. Between 2002 and 2006 the spread fell 2 points. While the bottom of the distribution of residualss (distance between 10 and 50) showed increased growth, the upper part of the distribution (percentiles 50 to 90) shows an increase much less intense and more uniform throughout the period.

TABLE 3 HERE

Of all these 26 points in the growth for the period, two thirds had its origin among residualss below the median. As can be seen in the last line of the second column, the distance between the 10 and 50 percentiles increased by 17 points, while the same line at the last column shows that among the residualss above the median, dispersion rose 9 points. This difference is even more intense before 2002. Until that year the dispersion below the median grew by 19 points, while above it, it grew by 8 points. This result is again similar to that obtained by Hidalgo (2008) with data until 2000. In conclusion, the data show that the dispersion of waste from a mincerian wage regression grows at least until 2002.

The table 4 contains the standard deviations for the years 1990 and 2006 and their growth in percent for cohorts of workers defined by cualification (high, medium-high, medium-low and low) by five age groups with ten years of interval each going from 16 to 65 and the two sexes. The aim is trying to find some change in the structure of within-group inequality conditioned to observed characteristics of workers.

TABLE 4 HERE

In principle, the most striking fact is that there are no significant changes when we condition residualss to observables. In all cases, with one exception (medium-high with 55 years or more) growth rates of within-group inequality are positive. However, in some cases, these rates are found not to be significantly different from zero.

As for differences between groups, we found some relevant issues. To begin with, growth in inequality appears to be stronger among younger and those with lower qualifications groups. While the inequality growth rate in this period was 12.3% for high and 19.3 for medium-high, the growth was 32.4 and 25.0% for medium-low and low respectively. Moreover, the growth of the younger cohort was 32.9%, higher than any other age group. Interestingly, we find that this rate declines with age, except the last group, which shows an upturn. It is interesting to find how these patterns are more heterogeneous in men and more homogeneous in women. It also appears that to higher age and higher qualifications we find increased dispersion, as is customary for this type of exercise. Finally, in cells corresponding to crossed characteristics of qualification and age, the pattern by which dispersion growth is greater the lower the value of variables, (less cualification and age), seems to be repeating itself, not without exceptions.

The table 4 shows patterns that can make us intuit some relevant ideas. The first is that inequality within groups is the heritage of all workers, with few exceptions. However, it appears that this increase in the waste dispersion is more intense for certain groups. These are the younger and less quaified. This change seems to be consistent with changes in the Spanish economy in this period. The crisis of the nineties and posterior recovery rose unemployment rate amongst the younger age groups with no qualification, as well as in older age groups for early retirement. However, job creation since the second half of the nineties was also intense at least in the first of the groups. Immigration, from the beginning of this decade, was very

important especially in groups with younger and less skilled groups than Spanish average (Economic Office of the President, 2006). Finally, the structure of economic growth has also affected these groups. All these reasons could be behind dispersion growth in the groups described.

However, the fact that this growth is common to all groups is important because it can be considered that there are other factors behind this pattern. One possibility is that we observe a change in the heterogeneity of workers within each group. It seems that these differences must fall more in the non-observed characteristics than in prices. If the answer were prices, to paraphrase Lemieux, we would find significant differences between men and women in terms of their growth in inequality within groups. Despite differences, these are not so important as to think that the prices paid to non-observed are causing this change. One possibility is that we observe a change in the heterogeneity of workers within each group. Another important idea is that, as inequality grows in all groups, composition effect will not be decisive in explaining the evolution of residual wage inequality.

Other reasons suggest that distribution may indeed be the reason, for example, when comparing the evolution of the RWI with other measures of inequality such as the skill wage premium. The figure 4 shows residuals dispersion jointly, again measured as the standard deviation of the residualss of the wage equation and represented in the second axis, and skill-wage premium of skilled workers, measured by the coefficient associated with the dummy "skilled worker" in these wage regressions. Therefore, it represents, in percentage, the wage gap for high-skilled workers facing a low-skilled worker. This series is measured from far left of the axis of the graph.

FIGURE 4 HERE

This figure shows that the evolution of both inequality measures have been completely the opposite. The skill wage premium for skilled workers with respect to the unskilled fell from 36% in 1990 to 25% in 2006. It is a similar result to that obtained in a multitude of works and for periods comparable to that used in this study, in which it is observed that remuneration of the qualifications has fallen considerably especially in Spain since the second half of the nineties (Pijoan-Mas and Sanchez-Marcos, 2009; Lacuesta and Izquierdo, 2007). Clearly this trend is completely different from that shown by the RWI. If we assume once again that there should be no obvious differences in the evolution of prices paid to non-observed and observed skills, the evolution of residual wage inequality should be driven by factors other than payment.

To demonstrate that composition effect can not explain the growth in inequality is to apply the method of Lemieux to MCVL data. Using the expression (5) and keeping the weights fixed in 1990, we can obtain a new measure of residuals dispersion without composition effect. This requires defining the cohort for which the variances are calculated, and apply weights fixed at the time, particularly where the definition of cohorts is broader than that shown in Table 2. For that, definition of cohorts have been broadened raising the figure to 80 -4 skill levels, ten age intervals, from 16-20 years to 61-65 and two by gender-and whose weights were held constant for 1990.

The figure 5 shows two lines. The first, with small triangles, and defined as the deviaiton of residualss, represents residual wage inequality previously calculated. The second line, with tables, la residual wage inequality without composition effect. The result shows that there was a clear composition effect given that both lines do not match. Moreover, the composition effect has clearly raised residual wage inequality. In case the structure of the labor force had not changed, residual wage inequality would have grown less than it really did. However, residual wage inequality, once it has been discounted from composition effect, keeps

showing a clear growth, which does not yet allow us to define the reasons for its different development from other inequality measures.

FIGURE 5 HERE

The figure 6 shows the same year but using the weights of 2006. As expected, the choice of the year to set the weights do not change previous conclusion.

So next section explains how to extend Lemiuex method if we assume that the distribution of non-observed characterisitics is not constant. There is a distribution effect in the terminology of this work. Moreover, it is shown that under certain feasible conditions it is possible to decompose change in the residual wage distribution between changes in the non-observed distribution characteristics and changes in their wages.

FIGURE 6 HERE

6 Accounting for Changes in Spanish Residual Wage Distribution

Once you have defined how to extend the decomposition of the variance between price changes, composition and distribution of non-observable characteristics, the exercise is executed for MCVL data and the period considered. The figure 7 shows residual wage inequality without composition effect, as proved in section 5, distribution effect growth and price effect growth. The latter is obtained by subtracting the growth from the first, the distribution effect growth ($\sum_j \theta_{js}(V_{jt} - V_{jt-1})$ -$\sum_j \theta_{js} p_t^2 \gamma_{jt}(n)(\sigma^2_{jt-1} - \sigma^2_{jt}(n))$) in (9). The dashed line, the one representing the effect on dispersion of residualss from the change in the distribution of non-observed characteristics, stays above zero. Therefore, the most important news is that the intra-group dispersion has grown during this period, although it seems that in recent years this growth has been

shrinking, exception made of 2000-2003 interval. Moreover, there seems to be no major changes in this trend, showing that the effect of changes in the qualities of the workforce has been steady but with no abruptness, which can be a priori adjusted to the idea that changes of this nature can not be overstated. The line corresponding to price change maintains its structure throughout time, although it increases the number of years in which rates are negative. While previously growth had been negative in the years 1995, 1998 and 2004 to 2006, eliminating the distribution effect years are extended with 1991, 1992, 1997, 1999 and 2003. That is to say, it goes from 5 years to 10 in which drops in the prices paid to non-observed characteristics are detected.

FIGURE 7 HERE

In the figure 8 three lines are shown. The so-called original residualss is residual wage inequality measured as the standard deviation of the residualss once composition effect has been removed, and that was calculated in section 5. If we remove to this series distribution effect growth, the result is the evolution of dispersion only due to the price paid to them. That is to say, dispersion is constructed of residualss that would have prevailed in the absence of composition effect and distribution effect. To construct this series, the value of the standard deviation for 1990 is used and from this, and using calculated growth rates, the series is generated. The series constructed only with prices have a very different profile from the original, as can be seen in figure 8. As can be seen, the difference is that this new series has a clear declining character. The difference between both series would be the evolution that dispersion would have experienced if only there had been distribution effect.

FIGURE 8 HERE

Finally, figure 9 shows the evolution of the dispersion caused by the prices paid to non-observed characteristics and the skill-wage premium already shown in section 5. What can be seen at the figure is that both series show a clearly decreasing profile, plus the second, and with converging patterns. For example, between 1990 and 1992, both series show an increasing behavior, while they don't exactly match for the years 1993-1995, although they do not offer a completely different behavior. Between 1995 and 1999 both show decreases in dispersion, therefore caused by both a reduction in wage differentials between workers with different qualifications, and, in the case where we assume that a great part of the non-observed characteristics of variability is ability, due to decreases in retribution to the latter. Between 2002 and last year, albeit with slight differences between the two series, there was again an increase in the wage gap, while from 2001-2002, depending on the series, we observed intense fall in wage differentials arising from remunerations.

FIGURE 9 HERE

This last figure reconciles the evidence that, a priori, the evolution of remuneration to the observed and non-observed abilities should be matched. In conclusion, if we assume, as it is usual in the literature, that much of the residual wage inequality is explained by ability, once it is controlled by the composition effect and the change in the distribution of this ability, changes in the wage gap are largely driven by changes in payments to the non-observed and observed skills. Furthermore, as expected, both changes are similar, and are really two versions of the same act.

7 Robust Analysis

7.1 Omitted Variables

A possible criticism to this exercise is that omitted variables are many, known and not included at the residuals directly assuming that they are non-observed. For example, in the regression (1), which is where everything is based on further analysis, it would be possible to include many more explanatory variables allowing to ensure that residualss obtained in this case are closer to the concept of skill.

Because of that, the exercise shown in sections 5 and 6 has been replicated to verify that the results are independent from the omission of variables. Specifically, sector, provincial, company size dummies have been included and also characteristics of the worker, type of contract and time, and other personal characteristics, as if he is married or is an immigrant.

In the figure 10, the dispersion growth caused by prices paid to non-observable characteristics, both the previously obtained (non-adjusted) and the one obtained by expanding the number of explanatory variables in wage regression. The figure shows an almost perfect match between the two series. The conclusion is that we can not affirm that the above results are motivated by the variation of the dispersion of omitted variables. Therefore, the conclusion reached at the previous section is maintained.

FIGURE 10 HERE

7.2 Test on the change in RWI change

To determine whether distribution effect is significant, the ideal would be to make a contrast of hypotheses about whether growth $\sum_j \theta_{js} p_t^2 \gamma_{jt}(n)(\sigma_{jt-1}^2 - \sigma_{jt}^2(n))$ is significantly different from zero. The quickest way of doing this is by way of a contrast of equality in variances. What is done is simply to check whether the variance caused by the distribution effect in the year is less than or equal to the variance arising in the year $t+1$. In case of rejecting the null hypothesis, we accept that distribution effect is significant and has therefore generated a growth in residual wage inequality during the period analyzed. If we accept the null hypothesis we are accepting the fact there has been no distribution and all that was attached to the price effect on the decomposition of Lemieux responds only to wage changes.

Take σ_t^2 as the variance in year t and σ_{t+1}^2 as the variance that was observed in year t +1 if they only had experienced changes due to effect distribution. Null hypothesis is

$$H_0 : \sigma_{t+1}^2 \leq \sigma_t^2$$

Versus the alternative one

$$H_1 : \sigma_{t+1}^2 > \sigma_t^2.$$

We reject null hypohesis if and only if

$$F_{ratio} = \frac{\sigma_{t+1}^2}{\sigma_t^2} > F_{1-\alpha},$$

Where α is the level of significance, and F is a F-Snedecor distribution.

As

$$\frac{\sigma_{t+1}^2}{\sigma_t^2} = 1 + \frac{\triangle\sigma_{t+1}^2}{\sigma_t^2},$$

the ratio $\frac{\sigma_{t+1}^2}{\sigma_t^2}$, it is easily calculable from (9). In table 5 test results are shown. As can be proven, null hypothesis is rejected for each year at almost 1%, so that $\sigma_{t+1}^2 > \sigma_t^2$ for period joint. That is to say, changes in abillity distributions has increased wage dispersion in Spain between 1990 and 2006.

TABLE 5 HERE

8 Conclusions

The aim of this work is to incorporate the explanation of residual wage dispersion the changes in the dispersion of abilities within each education-experience cell. Concretely, changes in residual wage dispersión was divided into four: changes in prices, composition, measurement error and the distribution of 'abilities' within education and experience cells. Previous to this paper, only the first three of them were considered. The decomposition is possible thanks to make an important difference between "new" and "older" workers. Knowing who is new in the labor market and who came into some years ago, it is possible to extract that part of residual wage inequality due to changes in abilities distribution.

Data form Continues Sample on Working Lives from Spanish Social Security Records are used to implement this new decomposition.

Once this decomposition has been executed, several facts are shown. First, the dispersion caused by the prices paid to the different non-observed characteristics has dropped, exactly the same comparing with

other skills princes, while the dispersion caused by their distribution has grown. This reconcile a puzzle for Spanish data, a decrease in the return to observable skills jointly an increase in residual wage inequality. Then I conclude that, for Spanish data, the within-group abilities' distribution is not constant on time, and it must be considered when we analyse wage inequality and its changes on time.

The last question is may be wondering why the increase of within-cells abilities dispersion could occur. Reviewing the literature some answers are proposed, such as universalization of education in Spain, more dispersion in school quality, or significant advance in the number of Universities of high diversity.

A Decomposition for σ^2_{jt}

The decomposition (9) can be obtained simply from the hypothesis proposed. The first states that the distributions of abilities among those workers already in the labor market in the previous period new workers only differ in their dispersion and not in her average. Thus, the variance of workers belonging to a cohort can be decomposed as the linear combination of the variances for both groups of workers in this cohort.

Variance σ^2_{jt} can be expressed as:

$$\sigma^2_{jt} = \frac{\sum_i (e^j_{it} - \overline{e}^j_t)^2}{N_{jt}},$$

where $e^j_{it} = \{e_{it} | x_{it} \in j\}$, the bar represents median and N_{jt} is the number of workers belonging to the group j at the moment t. If teh median is the same both for group o and group n, then:

$$\sigma^2_{jt} = \frac{\sum^o_i (e^j_{it}(o) - \overline{e}^j_{it})^2}{N(o)_{jt}} \frac{N(o)_{jt}}{N_{jt}} + \frac{\sum^n_i (e^j_{it}(n) - \overline{e}^j_{it})^2}{N(n)_{jt}} \frac{N(n)_{jt}}{N_{jt}},$$

where $e^j_{it}(o) \in \{e_{it} | x_{it} \in \{j, o\}\}$ and $e^j_{it}(n) \in \{e_{it} | x_{it} \in \{j, n\}\}$. Both $N(o)_{jt}$ and $N(n)_{jt}$ represent the number of workers already in the market and the new ones respectively. Changing the note we can express the variance σ^2_{jt} as

$$\sigma^2_{jt} = \gamma_{jt}(n)\sigma^2_{jt}(n) + (1 - \gamma_{jt}(n))\sigma^2_{jt}(o)$$

where $\gamma_j(n) = \frac{N(n)_j}{N_j}$. Multiplying both sides by p^2_t we obtain:

$$V_{jt} = \gamma_{jt}(n)V_{jt}(n) + (1 - \gamma_{jt}(n))V_{jt}(o)$$

Which is the expression (7).

B Selection and Information Processing from the MCVL

We take affiliation registers as the starting point. In each record information from each of the different relationships between the individual and Social Security is included. Each of these relationships include information on the date of commencement (real registers) and its end (real leaving), logically, this last exists for those relationships which occurred before the end of 2006. Thus, for each person, we will have as many records as changes its relationship with Social Security has had from initial register. At this point, homogenization of existing information is necessary with the objective marked by research. In this case, whose purpose is that of analyzing the evolution of wages and inequality in a period of time, it will be necessary to select a valid specific sample.

Not all relationships have been considered. Only those relationships originating a salary are interesting for this purpose given that this means paying the corresponding contributions. Therefore the unemployed and pensioners in all forms are not included in the sample. Of the remaining workers, all those labor relations not belonging to the general regime of social security, special arrangements for coal mining and domestic workers have also been removed. Furthermore, workers belonging to certain groups of special contributions, such as under-18, beneficiaries of war, self-employed, etc.. have also been eliminated. Specifically the groups included are those between Group 1 and Group 10. Workers from the agricultural sector have not been included, given the particular case of this sector, and that would not have been removed given that the agricultural system was not included. With regard to the employment relationship, some special cases, as those contributors who have some peculiarities that make them not being registered or those with a learning

contract have been eliminated. Workers hired through temporary employment agencies were also eliminated. Finally, those workers with a lack of information that might be relevant for the subsequent analysis have also been eliminated.

The MCVL for 2006 included 1,170,895 records of people who had some kind of relationship with Social Security this year, and who generated a total of 14,218,748 different affiliations for their entire working life Filtering according to preliminary information, we reduce the total membership of 874,655, which generated a total of 6,885,239 records of membership. Obviously this involves excessive handling of information, given that later on it will be interesting to distribute these relationships by years of contribution, which will multiply total information. Because of that, working with a random sample of them has been chosen. Of the 874,655 affiliates 200,000 records have been selected, less than a quarter.

These 200.000 selected members have generated a total of 1,478,735 affiliations registered for Social Security. 50% of these have generated less than 3 relationships or affiliations throughout his working life.. 75% percent reaches 8 affiliations. 90% has had a maximum of 16 affiliations. The average of affiliations by person raises in this selection to 7,39 per individual. .

Once we have the sample for affiliations, we contrast information with contribution file. Each affiliation can be extended for a determined period of time originating successive contributions throughout months or years. On the contrary, in just one year it is possible that a member has two or more records associated with changes in the contract, or any other reason which gave rise to its low or high post in the records of the SS. Because of that, contrasting the 1.478.735 different affiliations with contribution file we obtain different contribution registers. These 200.000 affiliates selected generated 7.39 different affiliations throughout their labor lives and 18.19 contribution registers for each one of these affiliations. Out of these registers, 68.19%

were affiliations generating contributions, and therefore they lasted at least one year, 10% changed at least once a year and 97% of the contributions were due to affiliation changes of not less than once a month.

For this case, those contributions for the selected affiliates previous to 1990 have been eliminated. This leaves us 3.391.166 contributions and 196.671 affiliates. We this information we start the necessary variable treatment to carry out our exercise.

From the contributions end information has only been taken for those corresponding to October of each year. The aim is to be able to compare the results of this work with those made with other data sources such as the Survey of Wage Structure. Of those workers who may appear with more than a different relationship for the month of October, for example, because of a change of contract for that month, one of them has been taken randomly. In this way, we have one observation per year and worker, of course, for all those years in which he/she was registered for the month of October. Once workers have been filtered, there are 2.563.059 different contributions.

C Adjusted wages

The salary information is inferred from the data of members' monthly contribution. In this case, the usual problem facing any analysis using MCVL is that higher wages are censored because of the existence of maximum contribution bases. There are also minimum contribution bases, which, however, concern us less given the existence of minimum legal wages that would support such censorship.. This fact can relevantly condition estimations done from wage equations with this sample.

Although the percentage of participants who show censorships in their contributions is not too high (10%), it will affect the estimates by the existence of a bias in the results of the same. To do this, in this work,

censored contributions biases are corrected. The idea is transferring distribution structure of those wages which are near to censorship, but not censored, to those that did were. For that, the following methodology base don Tobit estimation models

The (log of the) wage of a worker belonging to a group contribution g can be expressed as:

$$\begin{aligned} w_{ig} &= l_g \quad si \quad x_{ig}\beta_g + \varepsilon_{ig} \leq l_g \\ w_{ig} &= u_g \quad si \quad x_{ig}\beta_g + \varepsilon_{ig} \geq u_g \\ w_{ig} &= x_{ig}\beta_g + \varepsilon_{ig} \leq l_g \\ & \quad en \quad caso \quad contrario \quad a \quad los \quad anteriores \end{aligned} \tag{10}$$

where l_g and u_g are lower and upper limits of the contribution base for the group contribution g; x_{ig} it is a group of characteristics associated with worker, the job and where the company carries out its work , β_g the returns to each of the above characteristics and ε_{ig} the error term.

Therefore, the idea is to estimate the model (10) by a double-censored Tobit, properly defining the role of maximum-likelihood assuming normality in the error term. Once the model has been estimated and given the structure for given estimated error, we simulate the wage and contribution base for those workers whose original base was censored. For those other workers no simulation is carried out, unless they are used to estimate the stochastic structure of the errors that will be used to describe simulation.

Being s_g the standard error of the original wage series w_g, and defining $\widehat{u}_g$ as the estimated standard error and adjusted such that:

$$\widehat{u}_g = \frac{u_g - \widehat{w}_g}{s_g}$$

so that $\widehat{u}_g \sim N(0,1)$,we re-estimate wages for those who are censored by the expression:

$$\widehat{w}_{ig} = \widehat{w}_g + s_g \frac{\phi(\widehat{u}_g)}{1 - \Phi(\widehat{u}_g)} + s_g \phi^{-1} \left(\Phi(\widehat{u}_g) + \theta_i(1 - \Phi(\widehat{u}_g))\right)$$

Where $\theta_i \sim U(0,1)$ and ϕ is normal density function with zero average and variance 1 and ϕ^{-1} is its reverse. That is to say, given a probability value a, $\phi^{-1}(a)$ gives us a value in $\mathbb{R}$. The second term on the right $\widehat{w}_g$ corrects the estimate of bias derived from censorship. The third term on the right introduces randomness to the individual i and that is a function of the distribution of estimated errors with the information available for non-censored individuals.

In this way we correct the salaries of those employees whose basic minimum or maximum rate is equal to the limits set by law. This method ensures the maintenance of the structure due to stochastic information for the vast majority of workers.

References

Blackburn, Mc., D. Blomm & Richard B. Freeman (1990), The declining position of less-skilled americam males, *in* G.Burtless, ed., 'A future of lousy jobs?: the changing structure of US wages', Brookings Institution Press.

Chay, K.Y. & D.S. Lee (2000), 'Changes in relative wages in the 1980s returns to observed and unobserved skills and black–white wage differentials', *Journal of Econometrics* **99**(1), 1–38.

Farber, H.S. & R. Gibbons (1996), 'Learning and wage dynamics', *The Quarterly Journal of Economics* pp. 1007–1047.

Galor, O. & O. Moav (2000), 'Ability-Biased Technological Transition, Wage Inequality, and Economic Growth*', *Quarterly Journal of Economics* **115**(2), 469–497.

Garcia Perez, J.I. (2008), 'La muestra continua de vidas laborales: una guia de uso para el analisis de transiciones', *Revista de Economia Aplicada* (EXTRAORDINARIO)), 5–28.

Gottschalk, P. & R.A. Moffitt (1994), 'Welfare dependence: concepts, measures, and trends', *The American Economic Review* pp. 38–42.

Gould, E.D., O. Moav & B.A. Weinberg (2001), 'Precautionary demand for education, inequality, and technological progress', *Journal of Economic Growth* **6**(4), 285–315.

Hidalgo, M. (2008), Wage Inequality in Spain 1980-2000, Technical report, Universidad Pablo de Olavide.

Hoxby, Caroline M. (1997), How the changing market structure of u.s. higher education explains college tuition, NBER Working Papers 6323, National Bureau of Economic Research, Inc.

Hoxby, C.M., B.T. Long & L.E. Building (1999), 'Explaining rising income and wage inequality among the college educated', *NBER working paper* .

Izquierdo, Mario & Aitor Lacuesta (2006), Wage inequality in spain: recent developments, Banco de España Working Papers 0615, Banco de España.

Juhn, C., K.M. Murphy & B. Pierce (1993), 'Wage inequality and the rise in returns to skill', *Journal of Political Economy* **101**(3), 410–442.

Katz, L.F. & K.M. Murphy (1992), 'Changes in relative wages, 1963-1987: Supply and demand factors', *The Quarterly Journal of Economics* pp. 35–78.

Lemieux, T. (2006), 'Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?', *The American Economic Review* pp. 461–498.

Levy, F. & R.J. Murnane (1992), 'US earnings levels and earnings inequality: A review of recent trends and proposed explanations', *Journal of Economic Literature* pp. 1333–1381.

Mincer, J. (1974), *Schooling, experience, and earnings*, National Bureau of Economic Research New York.

Pijoan-Mas, J. & V. Sanchez-Marcos (2008), Spain is Different: Falling Trends of Inequality, Technical report, mimeo CEMFI.

Notes

[1] From this point on non-observed characteristics will be denoted as abilities.

[2] For example both Juhn, Murphy and Pierce (1993), Chay and Lee (2000) as Lemieux (2006), when analyzing the evolution of residual wage dispersion in the United States, assume that the distribution of each cohort is constant in time. Moreover, although Lemiuex (2006) finds in a footnote that this assumption can be forced, for example, because the younger workers could access the market with different ability distributions, that is convincing to assume its constancy.

[3] This result has already been found by other surveys using different data. For example, Hidalgo (2008) using data from the Household Budget Survey and the Continuous Household Budget Survey for the years 1980, 1985, 1990, 1995 and 2000, found

that the RWI grows from the late eighties, especially among those with lower wages. Pijoan and Sánchez (2009) find similar results with the same databases.

[4]Clearly, this is because it is considered that the skills learned are different for men and women workers and for white and black. School segregation, the effects peers and the less educated women explain these differences in skill endowments until recent decades. If the effect were only price, to greater endowment inequality, greater wage inequality.

[5]This assumption is logical, because if there is no clear methodological changes in compiling the information, we should not suppose changes in measurement errors.

[6]Which do not necessarily imply getting rid of under 18 year old individuals.

[7]This correction is explained in appendix B.

[8]The regression is the usual:

$$w_{it} = c + b_1 Q_{it,1} + b_2 Q_{it,2} + b_3 Q_{it,3} + b_4 \exp_{it} + b_5 \exp_{it}^2 + b_6 Sex_i + \varepsilon_{it}$$

where w_{it} is the log of the wages for individual i ijn year t; $Q_{it,1}$, $Q_{it,2}$ y $Q_{it,3}$, are dummies, each of them for high, medium and low qualified workers, *expe* is potencial experience valued as age minus the year of firts affiliation and lastly Sex is one when the worker is a woman..

Table 1. Descriptive statistics of the database

Año	number	average age	women	years of schooling	full time
1.990	130.959	34,5	32,4	8,5	95,7
1.991	140.781	34,6	33,6	8,6	95,0
1.992	144.419	35,2	33,6	8,6	94,5
1.993	149.875	35,9	32,8	8,7	94,2
1.994	154.240	35,9	33,5	8,8	93,3
1.995	160.425	35,7	34,6	8,9	92,0
1.996	170.247	35,4	35,9	9,0	90,3
1.997	186.616	35,2	37,2	9,1	88,4
1.998	208.443	34,6	39,6	9,2	85,5
1.999	230.269	34,4	40,9	9,2	84,0
2.000	239.727	34,5	41,5	9,3	83,2
2.001	248.822	34,7	42,4	9,2	82,1
2.002	274.286	35,1	42,7	9,2	81,7
2.003	273.195	35,2	44,0	9,2	80,1
2.004	287.382	35,4	44,6	9,1	79,3
2.005	307.390	35,6	45,5	9,1	78,1
2.006	330.671	35,8	46,1	9,0	77,0

Source: The Continuous Sample on Working Lives and own elaboration

Table 2: Average month wages, euros per worker

	Average wages		
	MCVL	EES	ECT
1990	766.4		
1991	825.2		
1992	891.0		
1993	952.3		
1994	1,013.6		
1995	1,051.4	1,124.3	
1996	1,080.4		
1997	1,110.6		
1998	1,139.7		
1999	1,169.1		
2000	1,210.0		1,290.0
2001	1,262.1		1,337.1
2002	1,310.2	1,389.3	1,387.8
2003	1,356.3		1,430.4
2004	1,396.8		1,461.5
2005	1,435.0		1,489.7
2006	1,494.0	1,438.5	1,549.4

Nota: MCVL, The Continuous Sample on Working Lives; EES, Structure Wage Survey; ECT, Quarterly Labor Cost SurveyEncuesta de Coste Salarial Trimestral

Source: Labor Ministry and Spanish National Statistic Institute and own elaboration

Table 3. Changes in inter-percentiles gap (log-points), 1990-2006

year	standard deviation	percentiles		
		90-10	50-10	90-50
1990	0.33	0.85	0.39	0.46
1994	0.37	0.95	0.47	0.49
1998	0.38	1.03	0.51	0.52
2002	0.42	1.13	0.59	0.54
2006	0.41	1.10	0.56	0.55
94-90	0.05	0.11	0.07	0.03
98-94	0.01	0.07	0.04	0.03
02-98	0.04	0.10	0.08	0.02
06-02	-0.01	-0.02	-0.03	0.01
06-90	0.08	0.26	0.17	0.09

Fuente: MCVL

Table 4. Within-Group Standard Deviation of Residuals and Growth, Qualification-Age-Gender Cells, 1990-2006

	Total				Men				Women			
	1990	2006	growth		1990	2006	growth		1990	2006	growth	
TOTAL	.329	.410	24.7	***	.309	.367	18.9	***	.323	.407	25.8	***
QUALIFICATION												
High	.414	.465	12.3	***	.435	.470	8.0	***	.360	.449	24.9	***
Medium-High	.353	.421	19.3	***	.360	.414	14.9	***	.340	.425	***25.1***	***
Medium-Low	.276	.365	***32.4***	***	.260	.340	***30.8***	***	.306	.389	***27.4***	***
Low	.308	.385	***25.0***	***	.284	.360	***26.5***	***	.342	.416	21.4	***
AGE												
[16,25)	.311	.414	***32.9***	***	.301	.395	***31.3***	***	.327	.441	***34.8***	***
[25,35)	.314	.395	***25.7***	***	.304	.372	22.4	***	.331	.416	***25.6***	***
[35,45)	.339	.405	19.5	***	.337	.385	14.1	***	.344	.429	24.6	***
[45,55)	.359	.408	13.7	***	.357	.395	10.5	***	.360	.424	17.8	***
[55,65]	.347	.426	22.9	***	.348	.440	***26.5***	***	.330	.396	20.0	***
QUALIFICATION x AGE												
High												
[16,25)	.422	.552	***30.8***	***	.421	.508	***20.8***	*	.389	.539	***38.8***	***
[25,35)	.391	.460	***17.4***	***	.406	.444	9.3	***	.362	.460	27.1	***
[35,45)	.418	.471	12.8	***	.444	.480	8.1	***	.348	.450	***29.4***	***
[45,55)	.439	.449	2.3		.452	.466	3.0	*	.369	.410	11.0	***
[55,65]	.377	.435	15.6	***	.386	.460	***19.2***	***	.318	.365	14.9	*
Medium-High												
[16,25)	.344	.455	***32.0***	***	.341	.421	***23.4***	***	.346	.465	***34.5***	***
[25,35)	.340	.425	***24.9***	***	.344	.417	***21.3***	***	.332	.425	***28.2***	***
[35,45)	.350	.412	17.7	***	.358	.403	12.6	***	.335	.419	25.1	***
[45,55)	.366	.399	9.3	***	.372	.400	7.5	***	.347	.395	13.8	***
[55,65]	.356	.373	4.8		.360	.392	8.7		.340	.335	-1.2	
Medium-Low												
[16,25)	.290	.370	27.7	***	.266	.341	28.2	***	.310	.389	25.5	***
[25,35)	.274	.363	32.6	***	.256	.339	***32.3***	***	.299	.386	***29.2***	***
[35,45)	.270	.361	***34.0***	***	.258	.335	29.7	***	.301	.399	***32.3***	***
[45,55)	.271	.357	31.9	***	.264	.335	27.3	***	.309	.394	27.5	***
[55,65]	.265	.370	***39.6***	***	.258	.367	***42.4***	***	.308	.374	21.6	***
Low												
[16,25)	.305	.411	***34.8***	***	.296	.386	***30.8***	***	.326	.460	***41.0***	***
[25,35)	.290	.362	24.6	***	.262	.339	29.5	***	.335	.395	***18.0***	***
[35,45)	.319	.372	16.8	***	.283	.348	22.8	***	.356	.395	10.9	***
[45,55)	.336	.388	15.5	***	.296	.361	21.9	***	.369	.407	10.2	***
[55,65]	.284	.395	***38.8***	***	.267	.398	***49.0***	***	.325	.391	20.2	

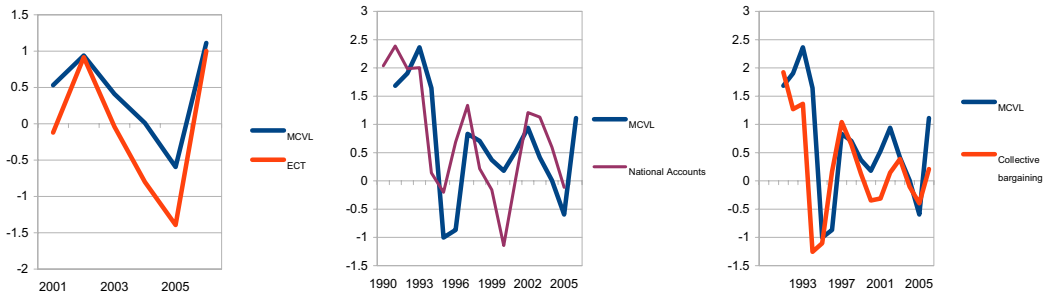
Note: * indicates that growth of standard deviation between 1990 and 2006 is positive and significant diferent from zero at 90-confidence level. ** indicates that the confidence is at 95%, and *** indicates that the confidence is at 99%. "Growth" are rates in %.

Table 5. Hypothesis testing of abilities' distribution inequality.

year	$\sigma^2_{t+1}/\sigma^2_t$	n_t	n_{t+1}	$F_{1-\alpha}$	p-value
1990	1.0000	139,959			
1991	1.0233	140,781	139,959	1.0105	0.014
1992	1.0260	144,419	140,781	1.0104	0.014
1993	1.0275	149,875	144,419	1.0103	0.013
1994	1.0172	154,240	149,875	1.0101	0.013
1995	1.0159	160,465	154,240	1.0099	0.013
1996	1.0161	170,247	160,465	1.0097	0.013
1997	1.0155	186,616	170,247	1.0093	0.012
1998	1.0147	208,443	186,616	1.0089	0.012
1999	1.0189	230,269	208,443	1.0084	0.011
2000	1.0087	239,727	230,269	1.0081	0.011
2001	1.0122	248,822	239,727	1.0080	0.010
2002	1.0158	274,286	248,822	1.0077	0.010
2003	1.0168	273,195	274,286	1.0075	0.010
2004	1.0127	287,382	273,195	1.0074	0.010
2005	1.0079	307,390	287,382	1.0072	0.009
2006	1.0061	330,671	307,390	1.0070	0.009

Note: n_t and n_{t+1} are the number of observations (individuals) for year t and t+1.
$F_{1-\alpha}$ is the value for the F-distribution at α level of significance.
P-value is the probability that the null hypothesis $\sigma^2_{t+1}<=\sigma^2_t$ must be accepted instead of $\sigma^2_{t+1}>\sigma^2_t$

Figure 1. Real growth rate of average wages.



Note: MCVL, The Continuous Sample on Working Lives; EES, Structure Wage Survey; ECT, Quarterly Labor Cost SurveyEncuesta de Coste Salarial Trimestral
Source: Labor Ministry and Spanish National Statistic Institute and own elaboration

Figure 1: Residuals standard deviation evolution

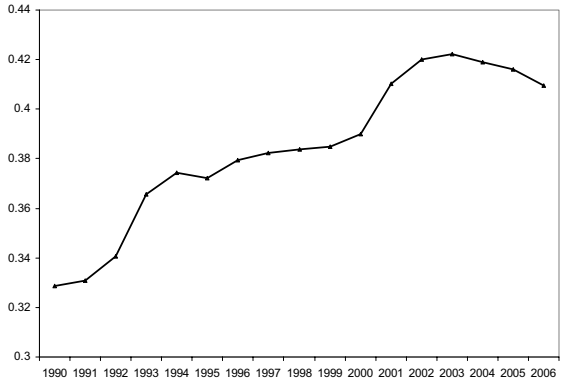


Figure 2:

Figure 3: Evolution of wage equation residuals percentiles gap.

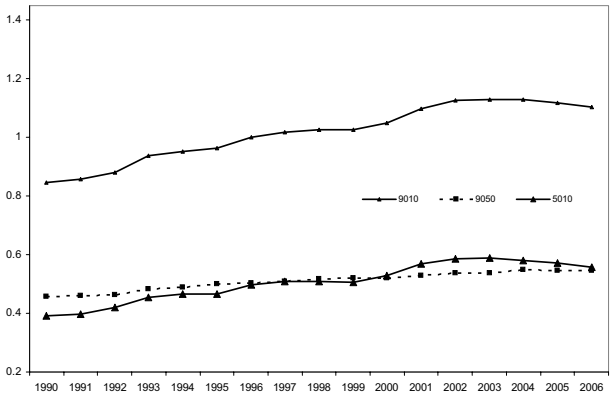


Figure 4: Skill wage premium and within-group wage inequality

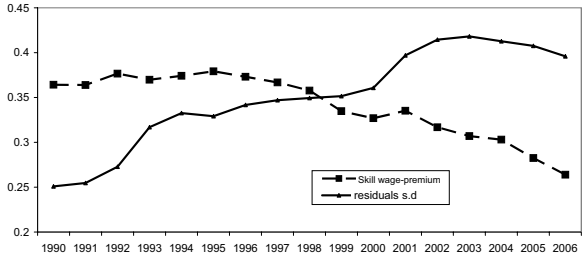


Figure 5: Composition effect on residual standard deviation evolution. Base year = 1990

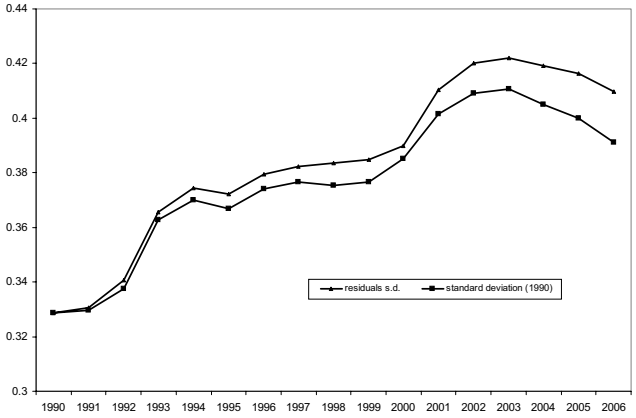


Figure 6: Composition effect on residual standard deviation evolution. Base year = 1990

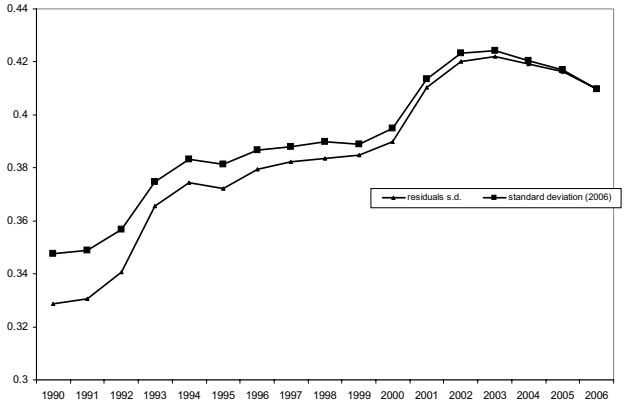


Figure 7: Residual wage inequality growth decomposition: changes in non-observed distribution, characteristics distribution and prices

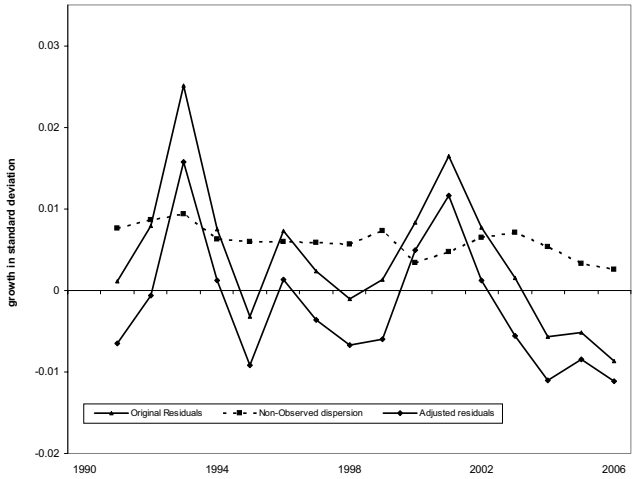


Figure 8: Evolution of residuals standard deviation, only with non observed prince efefct and with only non-observed distribution effect

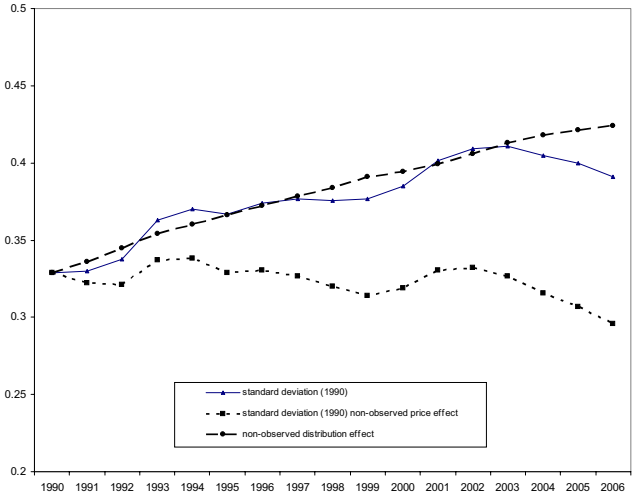


Figure 9: Skill wage-premium and non-observed price inequality change

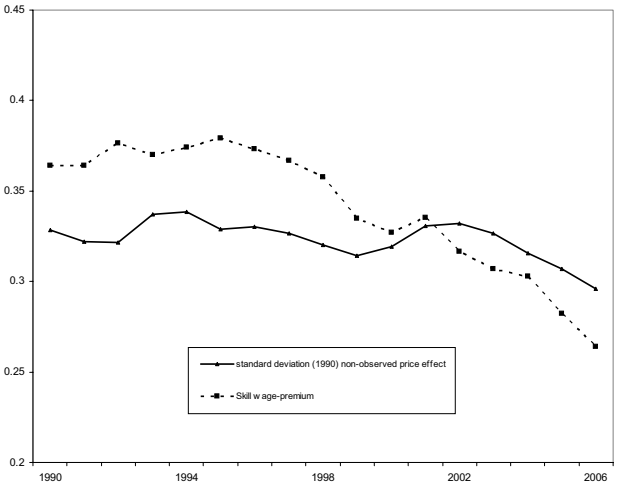


Figure 10: Comparisons of adjusted and non-adjusted increase in price dispersion for non observed characteristics