34133

POLICY RESEARCH REPORT ON
GENDER AND DEVELOPMENT
Working Paper Series No. 20

Wage Distribution in Chile: Does Gender Matter? A Quantile Regression Approach

Claudio Montenegro

This paper analyzes gender differentials in the returns to education, the returns to experience, and gender wage differentials in the Chilean case. It uses the standard "Mincerian" wage equation and estimates it separately by gender using the quantile regression method. Then, using the Oaxaca decomposition, we breakdown the total wage gap into an explained term (due to differences in endowments) and into a residual (or unexplained term). Chilean nation-wide data for the years 1990, 1992, 1994, 1996, and 1998 are used in the estimations. The results show systematic differences in the returns to education and to experience by gender along the conditional wage distribution. The paper also finds clear evidence that the unexplained wage differential is higher in the upper quantiles of the conditional wage distribution. The results are remarkably stable and consistent across years.

December 2001
The World Bank
Development Research Group/
Poverty Reduction and Economic Management Network

Wage Distribution in Chile: Does Gender Matter? A Quantile Regression Approach

Claudio E. Montenegro [*]

The World Bank
Development Research Group

Keywords: Chile; returns to education; quantile regression, wage differentials, labor markets.

JEL classification: J24, J31

Abstract.

This paper analyzes gender differentials in the returns to education, the returns to experience, and gender wage differentials in the Chilean case. It uses the standard "Mincerian" wage equation and estimates it separately by gender using the quantile regression method. Then, using the Oaxaca decomposition, we breakdown the total wage gap into an explained term (due to differences in endowments) and into a residual (or unexplained term). Chilean nation-wide data for the years 1990, 1992, 1994, 1996 and 1998 are used in the estimations. The results show systematic differences in the returns to education and to experience by gender along the conditional wage distribution. The paper also finds clear evidence that the unexplained wage differential is higher in the upper quantiles of the conditional wage distribution. The results are remarkably stable and consistent across years.

[*] Responsibility for the contents of this paper is entirely that of its author and should not be attributed to The World Bank, its Executive Directors or the countries they represent. I thank without implicating Maria Borda, Wendy Cunningham, Indermit Gill, Elizabeth King, Julian Lampietti, William Maloney, Andrew Mason, Marcelo Olarreaga, Lucia Pardo, Martin Rama, Luz Saavedra, Claudia Sepulveda, Isidro Soloaga, and the participants of the *XVII Latin American Meeting of the Econometric Society (Cancun, Mexico, August 1999)* for encouraging comments and help. Needless to say, any remaining errors are my own.

I. INTRODUCTION

Unexplained wage gaps occur when two people with equal abilities and skills are doing similar jobs but are treated differently by the employer. Although this difference in treatment may take many forms (wages, job assignments, promotions, or any other type of retribution), in this study we deal exclusively with wage gaps. These wage gaps in the labor market have various consequences. Tzannatos (1994) showed that wage and employment discrimination result in lower earnings and in efficiency loss (lower production). Therefore, policies aimed at preventing discrimination in the labor markets have two obvious benefits: an increase in income of the discriminated group (usually more prone to poverty than the dominant group) and an increase in total product.

Some studies have suggested that the gender wage gap may not be constant along the wage distribution. The standard methodology for estimating wage gaps uses regression equation estimates to decompose the gender wage gap into two components: a gender difference in endowments and a residual component (Oaxaca, 1973). Therefore, the procedure measures the wage gap at the mean of the wage distribution. New evidence suggests that the gender wage gap may be higher in the upper part than in the lower part of the wage distribution. For instance, Meng and Miller (1996) using data for Chinese rural industrial sector found that "the gender wage gap for staff is higher than for workers" (pp. 138); Kuhn (1987) reports that women at the higher top of the income distribution are more likely to report being discriminated against; also Garcia, Hernandez and Lopez-Nicolas (2001) using similar techniques to the ones used in this paper, found that "the wage gaps increases with the pay scale" (pp. 165).

To explore the idea that wage gaps are bigger in the upper part than in the lower part of the wage distribution we use figures 1-(a) and 1-(b). These figures show the hourly nominal average female wage relative to the hourly nominal average male wage by years of education and by years of experience, in the case of Chile, using five independent surveys. Although these figures are an inexact representation of the true earning gaps (because interactions are not considered and because the samples may be biased samples of their populations), they clearly suggest bigger wage gaps for women with more education and/or with more experience (and hence, with higher incomes). The purpose of using five years instead of just one, is to give some sense of robustness to our estimates. The exact definition of the sample considered for these calculations is given in the data section and it is kept constant throughout the paper.

Chile is a very interesting case to analyze. In spite of being a developing country, Chile ranks 38th among all the countries in the world when considering the human development indicators (UNDP, 2000), and has a very open and competitive economy with a relatively well-developed labor market. Regardless of this, only a few studies have addressed the issue of gender wage gaps in Chile. Paredes (1982) and Paredes and Riveros (1994) looked at the gender wage gap, but their sample includes only the metropolitan area of Santiago, excluding about two thirds of the country implying that agricultural, mining, maritime activities, among others, were excluded from the sample. In another study (Gill, 1982), the author uses country-wide data, but is unable to properly define an hourly wage measure of income because the survey used in his estimates did not include the number of hours worked. In this paper we use nationwide surveys that permit us to control also for the number of hours worked.

The aims of this study are: (i) to complement and extend the gender wage gap analysis in the Chilean case; (ii) to analyze the gender differentials in returns to education, returns to experience and the gender wage gap using the Mincer equation and the Quantile Regression Method (QRM); and (iii) to analyze the stability of these results. The use of the QRM permits fitting hyperplanes through out the conditional wage distribution and is ideal for characterizing the entire wage distribution. It is especially useful when the effect on the dependent variable of the covariates differ for different conditional quantiles of the wage distribution. The method has the potential of generating different responses in the dependent variable at different quantiles. These different responses may be interpreted as differences in the response of the dependent variable to changes in the regressors at various points in the conditional distribution of the dependent variable. The study does that simultaneously for men and women using comparable samples and variable definitions which permits us to make comparisons not only across gender but also across years.

The results clearly show that the returns to education, the returns to experience and the gender wage differential are not constant along the wage distribution. In particular, we show that in the lower part of the wage distribution women have higher returns to education than men, but similar returns to experience. In the upper part of the wage distribution, women have similar returns to education but lower returns to experience than men do. When using the Oaxaca decomposition these results imply that about 10% of the gender wage gap in the lower part of the distribution and around 40% in the upper part of the wage distribution cannot be explained by differences in education and experience. It is possible that at least a portion of this wage gap can be attributed to discrimination. The analysis also shows that the results are remarkably similar for the five different years used in the study.

The rest of the paper is organized as follows: section two presents the basic model used, it also discusses gender wage differentials and how they can be included in that model; section three presents the QRM and the basic data used; section four presents and discusses the results; section five presents the main conclusions.

II. THE MODEL

The standard model used to analyze earning differentials is based on the human capital earnings function developed by Mincer (1974) that has the form:

$$\ln(Y_i) = \varphi(X_i) + u_i \qquad (1)$$

where $\ln(Y_i)$ is the natural log of earnings or wages for individual i, X_i is a vector that usually includes a measure of schooling or educational attainment, a measure of the stock of experience, and some other factors that may affect earnings such as occupation, training, race, gender, abilities, marital status, number of children, seniority in actual job, hours of work, health, region, firm size, etc.; and u_i is a random i.i.d. disturbance term that reflects unobserved characteristics. Note that in equation (1) nothing is said about the functional form of the equation. The empirical estimation usually has the form (see for instance, Willis, 1986; Polachek and Siebert, 1993):

$$\ln Y_i = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + u_i \qquad (2)$$

where $\ln(Y_i)$ is the natural log of the hourly wage, S_i is the years of schooling, E_i is the level of experience (proxied, for data reasons, by age minus years of schooling minus 6), E_i^2 is the square of the level of experience (included to account for the commonly observed effect of a declining age-earning profile for a given level of experience).

It is important to recall that equation (2) is based on some restrictive assumptions. It assumes that individuals are of equal abilities and face equal opportunities (i.e., it assumes perfect capital and labor markets, which allows us to take earnings as a proxy for marginal productivity). It also ignores direct costs of schooling and overlooks earnings while attending school. Moreover, it assumes a constant return per year of schooling. A closer look at equation (2) also shows us that the parameter for years of schooling is an estimate of the impact of schooling on wages rather than an internal rate of return on investment. If it were an internal rate of return it would be a private one, since this specification ignores any subsidization of schooling and omits any positive or negative externalities to schooling.

Equation (2) also omits a potentially very relevant variable: ability. Ability is likely to be positively correlated with schooling, so omitting ability measures from the regression equation will bias the estimated returns to schooling upward. However, ability is difficult to conceptualize and measure, and there is no consensus as to whether it is significant enough to differentiate earnings. For these reasons and because the survey data do not include any variable that could conceivably be used as a proxy of ability, this problem is ignored in our estimations. Another problem associated with the estimation of equation (2) is that we have to proxy experience by its potential term: age minus years of education minus six. This is a poor proxy. Furthermore, potential experience is an even poorer proxy for women than for men especially in the case of women who drop out of the labor force to raise children. Therefore, women's potential experience overstates true work experience relative to men's and so it is not surprising to find that women appear underpaid for comparable experience. An additional problem is that equation (2) assumes that education is assigned randomly across the population. In reality education is endogenous and the estimation of the relationship between earnings and education may be biased upward or downward depending on the way individuals make their education choices. Like in the case of abilities, there is a lack of adequate instruments in the sample at hand so we were not able to correct the problem of endogeneity of education and it is ignored in our estimations.

III. METHODOLOGY AND DATA

The traditional method used to estimate the Mincerian equation has been Ordinary Least Squares (OLS). The limitations of this method are that (i) it characterizes the wage distribution only at the mean of the distribution, and (ii) it is not robust to the presence of outliers. Such disadvantages have induced the use of the median regression methods (also known as least absolute deviation -LAD- estimators) where the objective is to estimate the median of the dependent variable. The median regression method fits the regression hyperplane that minimizes the sum of the absolute residuals rather than the sum of the squared residuals. Putting this in a slightly different but equivalent way, it adjusts a regression hyperplane that leaves 50% of the errors "above" the hyperplane and 50% "below" the hyperplane. In the same vein these positive and negative errors are weighted differently, such that the weights are one giving origin to the statistical method QRM (for a more detailed description of the QRM, see methodological appendix). This technique allows us to estimate conditional quantile functions (among them the conditional median function) and obtain statistical inference about the parameters estimated. The purpose of the classical least squares estimation is to determine the conditional mean of a random variable Y, $E(Y|X)$ given some explanatory variables X, usually under some assumptions about the functional form of $E(Y|X)$, for instance, linearity. The QRM enables us to pose such a question at any quantile of the conditional distribution. Let us remember that a real-valued random variable Y is fully characterized by its distribution function $F(y) = P(Y \le y)$. Given $F(y)$, we can for any $\theta \in (0,1)$ define θ-th quantile of Y given by $Q_Y(\theta) = \inf\{y \in \Re \mid F(y) \ge \theta\}$. The quantile function, $Q_Y(\theta)$ as a function of θ, completely describes the distribution of the random variable Y. Hence, the estimation of conditional quantile functions allows us to obtain a more complete picture about the dependence of the conditional distribution of Y on X. In other words, this means that we have the possibility to investigate the influence of explanatory variables on the *shape* of the distribution, as well as on the conditional impact of the regressors on the dependent variables at different “layers” of the distribution.

The QRM just discussed is not exempt from criticism. One criticism is related to the arbitrariness in the election of the proportion θ. This criticism is not relevant to our study. Our main purpose is to characterize the whole distribution of the wage structure and so we use *seventeen* different conditional quantiles. Another problem associated with the QRM is that the final solution may not be unique. But uniqueness can always be achieved by selecting an appropriate design or by using an arbitrary rule to select from any set of multiple solutions (this problem is exactly similar to selecting the median among a sample that has an even number of observations).

Data

This paper uses micro data sets of the Caracterización Socioeconómica Nacional (CASEN), for the years of 1990, 1992, 1994, 1996 and 1998. The CASEN is a nationally and regionally representative household survey carried out by MIDEPLAN, through the Department of Economics at the Universidad de Chile, with the dual objectives of generating a reliable portrait of socioeconomic conditions across the country, and of monitoring the incidence and effectiveness of the government's social programs and expenditures. The working samples in this study contain workers who worked at least 35 hours per week and were not self-employed. The sample includes only "empleados" (white-collar workers) and "obreros" (blue-collar workers). Self-employed workers, domestic servants, and military personnel were omitted from the sample. Self-employed workers were omitted because the data did not allow the separation of income into returns to labor and returns to capital. Domestic servants were omitted because their recorded earnings could misrepresent their labor income, which for live-in domestic servants includes room and board, both difficult to value. Military personnel were omitted because their salaries do not correspond to a market productivity criterion. The unemployed and people who work in voluntary services were also excluded. The same type of survey and variable definitions are used throughout the period in order to have year by year absolutely comparable results. The dependent variable is defined as the log of the hourly wage. It is important to stress that all analyses carried out in this paper always use the same sample definition.

Before proceeding with the formal analysis, an examination performed using Q-Q plots (not shown) revealed the presence of outliers in the dependent variable of our equation (the natural log of the hourly wage), but as stated before, this is not of a concern since the technique we will use is robust to the presence of outliers. The same analysis did not show any outliers in the independent variables.

Table 1 shows for each year the basic statistics of the variables in the sample. The table reveals some very interesting facts. First, the overall wage gap between men and women is far smaller than the observed wage gap in the United States and other developed countries (Gunderson, 1989). The wage ratio (defined as $(Y^f/Y^m)*100$) is 82%, 87%, 88%, 93% and 96% for the years 1990, 1992, 1994, 1996 and 1998. Although this seems to suggest that the wage gap has been eliminated, a closer look shows that women are more educated than men, on average. On that basis one might expect them to earn, more (and not less!) than men. The education ratio (defined in a similar way) is 117%, 117%, 119%, 116% and 116% for the same years. Given that the women who work are the best educated women, these education ratios overestimate women's *overall* population means; for the entire working age population -i.e. for all people over 14 years old- there is no significant difference in the education mean by gender. In contrast, Table 1 shows that women have an "experience gap" relative to men. The experience ratio (again defined in a similar way) is 81%, 80%, 80%, 82% and 82% for the same years.

Table 1 also reveals that women constitute about one-quarter of the sample, but that proportion increased in the last two years of the sample. The proportion of women in the working force is 25%, 25%, 24%, 27% and 29% for the years 1990, 1992, 1994, 1996 and 1998. Given that education is one of the most important factors affecting wages, we take a closer look at the distribution of education by sex.

Figures 2-(a) and 2-(b) show the marginal distribution of educational attainment for men and women, respectively, for each year in the sample. Both graphs show the percentage of people in each category of education as a proportion of the total gender category. These figures highlight the differences of educational attainment between men and women. Working women are highly educated with thirteen years plus of education while working men have significantly lower levels of education attainment. The old Chilean system of education had a first cycle of six years, and a second of also six years. The new system has a first cycle of eight years and a second of four. Therefore, high percentages of educational attainment are found at six, eight and twelve years of education. It is interesting to note that a significantly higher proportion of women compared to men tend to complete secondary school.

Given that the distributions for educational attainment vary significantly, it is interesting to observe the proportion of males in the labor force, by the number of years of education and by range of years of experience. Figures 3-(a) and 3-(b) present the proportion of males in the labor force by educational attainment and years of experience, respectively. Figure 3-(a) shows a high disproportion of men versus women in the labor force by years of education. Males represent slightly more than 80% of the working labor force that have eleven years or less of education. This proportion falls approximately to 67% for twelve years of education, to around 55% for thirteen to seventeen years and again increases to 72% for eighteen and nineteen years of education. This analysis indicates that education is a very important factor concerning women's decision to participate in the labor force. This impact is even more remarkable for women who have completed secondary school or who have some tertiary schooling. Figure 3-(b) also shows a high disproportion between the genders by range of years of experience: it reveals that women participating in labor market have less experience than their male counterparts. Women represent approximately 40% of the working labor force with 0 to 4 years of experience, approximately 30% of the working labor force with 5 to 9 years of experience, and around only 25% of the working labor force with 10 years or more of experience.

IV. EMPIRICAL ANALYSIS

There is abundant literature on the issue of sample selectivity bias in wage equations estimation using only data for working people (Gronau, 1974; Heckman, 1979; Killingsworth, 1983; Manski, 1989; Schultz, 1993). The bias results from the fact that the workers in the labor force are those who obtained wage offers higher than their reservation wages. Workers who receive offers lower than their reservations wages do not participate in the observed labor force, therefore, their actual wages are unobserved. This problem is accentuated for the female's wage equation estimation since it is assumed that their reservation wages are higher than male's due to female's presumed higher productivity in home activities. Heckman's (1979) procedure is the most widely used procedure to account for sample selectivity when estimating wage equations. However, this traditional selectivity correction procedure cannot be used when using the QRM (see Buchinsky 2001). Moreover, it is not exempt from criticism even when used to correct the OLS estimates. Manski (1989) argues that the procedure lacks robustness and is sensitive to identification. Puhani (2000) recommends a "case by case" use of the Heckman selectivity correction. Heckman (1979) himself warns against the use of the procedure with inadequately specified selection model. Due to these considerations and the lack of good instruments to represent a labor market decision mechanism in our sample we decided to present our results without sample selectivity correction. This is not an unusual decision (see, for instance, Wooden, 1999; Liu, Meng, and Zhang 2000; Newell and Reilly, 2001).

We estimate the usual version of the Mincerian equation that includes years of schooling, years of experience, and the square of years of experience. The equation was estimated using the QRM for different values of θ, for each of the five years in the sample, and separately for men and women. Also the equation was estimated using OLS for a familiar point of comparison. We present the parameters estimated using the QRM and the OLS methods for each θ, year, and gender in Annex Table I and depicted in figures 4, 5 and 6. The graphs include the 95% confidence intervals for men and women. To facilitate comparison with the traditional methodology the table (but not the figures) includes the OLS estimates.

Figure 4 illustrates wage returns to education for men and women. In general, the returns to education are higher for women than for men in the lower quantiles of the distribution. The results also show that those differences are statistically significant at the 95% confidence level (i.e. the two bands showing the 95% confidence interval do not intersect). On the contrary, in the upper part of the distribution men tend to have a higher return to education than women. In other words, the two bands intersect each other. Since this intersection is around the θ=0.50 this implies that simple OLS would probably not show any statistically significant difference in the returns to education for men and women. The statistically significant difference is at the tails, especially at the lower tail, of the conditional distribution of wages.

Figure 5 shows the returns to experience for men and women. The figure shows that the pattern of return to experience by quantiles evaluated for both groups at a level of 17 years of experience is very similar for the five years considered in our sample. Nevertheless, they present a different pattern for men and women. Although they are also almost equal for men and women in the lower quantiles, in the upper quantiles men definitively have a higher return than women. At around θ=.50 the return to experience for men becomes statistically significant higher than the return to experience for women.

Finally, figure 6 shows the estimated constant in all the models estimated. Women have a statistically significant lower constant than men in the lower part of conditional distribution of wages, but this difference tends to decrease as we moved upward.

Our analysis has shown that there are systematic differences in the returns to education and experience by gender. Moreover, the patterns are remarkably stable from year to year. The next step is to compare the predicted wages for men and women and try to identify systematic differences in their wages along the conditional distribution of wages. Two different sources that could explain the total difference in the predicted wages by gender are: differences in the endowments for each group or differences in the returns to those endowments. To measure and disentangle both effects we use the Oaxaca decomposition (Oaxaca, 1973).

Oaxaca Decomposition

The decomposition is based on the basic human capital earnings function. Let's denote the male group by *m* and the female group and by *f*. According to equation A.3 (see appendix) the θ^{th} quantile regression predictions of the hourly earnings in each sex group can be expressed as:

$$\hat{Y}_{\theta}^{m} = X^{m}\hat{\beta}_{\theta}^{m} \qquad (3)$$

$$\hat{Y}_{\theta}^{f} = X^{f}\hat{\beta}_{\theta}^{f} \qquad (4)$$

where $\hat{Y}_{\theta}^{i}$, $(i = m, f)$, indicates the quantile regression prediction of the θ^{th} percentile of Y given X. Therefore, the θ^{th} percentile value of Y predicted for each group (at the mean values of the covariates) is given by:

$$\hat{Y}_{\theta,\overline{X}}^{m} = \overline{X}^{m}\hat{\beta}_{\theta}^{m} \qquad (5)$$

$$\hat{Y}_{\theta,\overline{X}}^{f} = \overline{X}^{f}\hat{\beta}_{\theta}^{f} \qquad (6)$$

where $\hat{Y}_{\theta,\overline{X}}^{i}$, $(i = m, f)$, indicates the predicted value of the θ^{th} percentile of Y given the mean values of the covariates. The purpose of the Oaxaca decomposition is to breakdown the total wage gap between the two groups, $\hat{Y}_{\theta,\overline{X}}^{m} - \hat{Y}_{\theta,\overline{X}}^{f}$, into differences of observable characteristics and a residual or an unexplained part. This is done by assuming that the "discriminated" group is paid the wages of the other group. This is equivalent to assuming that men and women are perfect substitute factors of production. In other words, a man and a woman with the same characteristics of education and years of experience performing the same job should earn the same. In absence of unexplained differences the wage of the female group would be given by:

$$\hat{Y}^{f*}_{\theta,\overline{X}} = \overline{X}^{f}\,\hat{\beta}^{m}_{\theta} \qquad (7)$$

We can now decompose the total wage gap into two components as follows:

$$\hat{Y}^{m}_{\theta,\overline{X}} - \hat{Y}^{f}_{\theta,\overline{X}} = (\hat{Y}^{m}_{\theta,\overline{X}} - \hat{Y}^{f*}_{\theta,\overline{X}}) + (\hat{Y}^{f*}_{\theta,\overline{X}} - \hat{Y}^{f}_{\theta,\overline{X}}) \qquad (8)$$

In the OLS case, by restriction of the method, the adjusted hyperplane goes through the mean of all the variables. This is not the case when using LAD estimators, and so, the strictly rigorous equation (8) should include a residual. This residual is arbitrary and depends on the values at which we decide to evaluate the functions (see Garcia, Hernandez and Lopez-Nicolas, 2001).

Substituting equation (11), (12) and (13) into (14) and rearranging terms we get:

$$\hat{Y}^{m}_{\theta,\overline{X}} - \hat{Y}^{f}_{\theta,\overline{X}} = \hat{\beta}^{m}_{\theta}(\overline{X}^{m} - \overline{X}^{f}) + (\hat{\beta}^{m}_{\theta} - \hat{\beta}^{f}_{\theta})\,\overline{X}^{f} \qquad (9)$$

And normalizing by the predicted women wage in the respective quantile we get:

$$\frac{\hat{Y}^{m}_{\theta,\overline{X}} - \hat{Y}^{f}_{\theta,\overline{X}}}{\hat{Y}^{f}_{\theta,\overline{X}}} = \frac{\hat{\beta}^{m}_{\theta}(\overline{X}^{m} - \overline{X}^{f})}{\hat{Y}^{f}_{\theta,\overline{X}}} + \frac{(\hat{\beta}^{m}_{\theta} - \hat{\beta}^{f}_{\theta})\,\overline{X}^{f}}{\hat{Y}^{f}_{\theta,\overline{X}}} \qquad (10)$$

The first term on the right hand side corresponds to differences in wages in each quantile due to differences in the endowments (education and years of experience) of the two groups. The second term on the right hand side reflects differences in return to the endowment, unequal pay for the same endowment. This term is the unexplained wage differential that corresponds to the Becker's index of discrimination.

The Oaxaca decomposition is not exempt from criticism. Since the unexplained wage differential is measured as a residual, it is not clear if the decomposition will over or under estimate the residual. This will depend on whether the omitted variables are positively or negatively correlated with productivity and on the distribution of the omitted variables across both groups. Knowing the relationship between productivity and the omitted factor, and also the distribution of the omitted factor among the two groups, we could, in principle, establish the direction of the bias. A second criticism is that the evaluation of the unexplained wage gap is made at the levels of the "non-discriminated" group. This assumes that, in the case of "no-discrimination", the prevailing wage would be the "no-discrimination" wage. This can be easily corrected assuming that the final wage is a weighted average of both group wages. The weights being the relative size of each group. A third critique is that the residual calculation of the unexplained wage becomes meaningless in the presence of measurement error of the independent variables. Although this is a very important criticism, it is not relevant in our study. We are more interested in comparing how the unexplained gender wage gap differs in different conditional quantiles of the wage distribution given the sample definition, the variable definition and the variable measurement than in the absolute level of the gender wage gap itself.

Oaxaca Calculations

As stated in equation (10), we estimated the decomposition at the mean of the vector of predictors. We used the same parameters depicted in figures 4, 5 and 6 and that are also tabulated in Annex Table I. The calculations are presented in Table 2. The first five columns under the heading of Male Wages in Table 2 present the predicted wage for males at each quantile, for each year of the analysis. The next five columns under the heading or Female Wages present the predicted wages for females at each quantile, for each year of the analysis. The last five columns present the wages that women would receive if they were paid male's return. In terms of equations (5), (6) and (7), each group of five column represent, respectively, $\hat{Y}^{m}_{\theta,\overline{X}}$, $\hat{Y}^{f}_{\theta,\overline{X}}$, and $\hat{Y}^{f*}_{\theta,\overline{X}}$.

The lower part of Table 2 shows the Oaxaca decomposition. The first five columns present the total wage differentials as a percentage of the female wage in the respective quantile, $\left(\hat{Y}^{m}_{\theta,\overline{X}} - \hat{Y}^{f}_{\theta,\overline{X}} \right) \Big/ \hat{Y}^{f}_{\theta,\overline{X}}$, for each year and quantile. The next five columns present the difference due to endowments as a percentage of the female wage in the respective quantile, $\left(\hat{Y}^{m}_{\theta,\overline{X}} - \hat{Y}^{f*}_{\theta,\overline{X}} \right) \Big/ \hat{Y}^{f}_{\theta,\overline{X}}$, for each year and quantile. Finally, the last five columns present the gender wage gap due to differences in returns as a percentage of the female wage in the respective quantile, $\left(\hat{Y}^{f*}_{\theta,\overline{X}} - \hat{Y}^{f}_{\theta,\overline{X}} \right) \Big/ \hat{Y}^{f}_{\theta,\overline{X}}$, for each year and quantile. Note that Table 2 also presents the same calculations using the standard method of OLS.

Several points should be noted about Table 2. First, the predicted male wage for the 90th quantile is approximately 4.8 times the predicted male wage for the 10th quantile. In the case of women the difference is approximately 4.1 times. When we look at female's wages with male's pay we see that the predicted female wage for the 90th quantile is approximately 5.0 times the predicted female wage for the 10th quantile. This suggests that female wages are far more compressed than male wages. Note that these differences in wages are due only to differences in returns to the endowments along the conditional distribution of wages. All the calculations assume that the individuals have the same endowments in each year. The stability of these values across years should also be noted. A second important point to note is the symmetric distribution of errors indicated by the closeness of the OLS wage prediction, the mean wage prediction, to the median prediction.

We now turn our attention to the standardized Oaxaca decomposition presented in the lower part of Table 2 and figure 7. As we mentioned in section III, the total gender wage gap in Chile is low and it has been steadily decreasing. This result is confirmed by our measure of the standardized Oaxaca total wage gap as a percentage of females wages. The total wage gap in predicted wages is 12%, 11%, 4%, 5% and 3% for years 1990, 1992, 1994, 1996 and 1998, respectively at the median θ=0.50. Using the OLS method to calculate the total wage gap give us 11%, 11%, 6%, 5% and 2%. Figure 7-(a) presents the Oaxaca measures of the total wage gap, for each year and quantile. Two things are worth noting. First, the patterns of the total gap are very similar for the five different years considered in our sample. Second, the results clearly indicate that the total gap is greater in the upper quantiles than in the lower quantiles.

The standardized Oaxaca median estimates for the differences due to endowments are -12%, -13%, -17%, -18% and –17% for years 1990, 1992, 1994, 1996 and 1998, respectively. These measures are negative because the average female education is higher than the male average education. To understand this we take a second look at the first term on the right hand side of equation (10). Given that $\hat{\beta}_{\theta}^{m}$ and $\hat{Y}_{\theta,\bar{X}}^{f}$ are positive terms, the negative sign clearly comes from the term $(\bar{X}^{m}-\bar{X}^{f})$. Female education is higher than male education, on average. When this negative difference is multiplied by higher male returns it gives even lower values of the wage gap due to endowments. The same estimates when using the OLS method are -14%, -14%, -18%, -18% and -18%. It is evident from these estimates that there is a relatively high unexplained wage differential when the differences in endowments are taking into account. This reflects our previous statement that women should earn more (and not less) than men because they have more education than men, on average. It should also be noted that in this case the mean and the median estimates are relatively similar, as shown in Table 2. Figure 7-(b) shows that the wage gap due to difference in endowments generally increases as we move from the lower quantiles to the upper quantiles of the conditional distribution of wages. Again, two things are worth noting here. First, the patterns of the gap are very similar for the five different years considered in our sample. Second, the results suggest that the gap is greater in the upper quantiles than in the lower quantiles.

The standardized Oaxaca median estimates of the unexplained wage gap are 24%, 24%, 21%, 22% and 19% for years 1990, 1992, 1994, 1996 and 1998, respectively. The same estimates when using the OLS method are 25%, 25%, 24%, 23% and 20%. These estimates suggest that there is a relatively high proportion of unexplained gender wage gap when the differences in endowments are taken into account. Note that in the case of the unexplained gender wage gap the mean and the median estimates are relatively similar. Figure 7-(c) shows that the wage residual gap increases as we move from the lower quantiles to the upper quantiles of the conditional distribution of wages. In the lower part of the conditional distribution of wages the unexplained residual is about 10% of the average predicted female wage. This percentage linearly increases up to about 40% in the upper part of the conditional distribution of wages. Once again, two things are worth noting here. First, the patterns of the unexplained wage gaps are very similar for the five years in the sample, not only in terms of the levels but also in terms of the patterns found across quantiles. Second, the results clearly show that the gap is far greater in the upper quantiles than in the lower quantiles.

V. CONCLUSIONS

The principal objective of this paper is to analyze gender differentials in the returns to education, the returns to experience and the wage differentials along the conditional wage distribution. Our results show that the returns to education are significantly different for women and men by quantiles. Four points should be noted here. First, for both men and women private returns to education tend to increase as we move from the lower to the upper part of the conditional wage distribution. Second, women have higher returns to education than men in the lower quantiles of the distribution and similar returns in the upper quantiles of the wage distribution. Third, the returns to education at the median, θ=0.5, produce very similar results for men and women, implying that an OLS mean estimate will not be able to detect gender differences. Finally, the gender differences in the rates of return to education are stable and systematically decreasing when we move from the lower quantiles to the upper quantiles of the conditional distribution of wages. Our results for returns to years of experience show that in the lower quantiles men and women have similar rates of returns whereas in the upper quantiles men tend to have higher rates of return. Finally, our analysis shows that the unexplained wage gap is not constant along the conditional wage distribution. In particular, we show that the unexplained wage gap steadily increases from 10% to 40% as we move from the lower part to the upper part along the conditional wage distribution. These findings are consistent with those of Garcia, Hernandez and Lopez-Nicolas (2001) who also report that the residual or unexplained part is greater in the upper part of the conditional wage distribution.

REFERENCES

Buchinsky, Moshe. 1994. 'Changes in the U. S. Wage Structure 1963-1987: Application of Quantile Regression', *Econometrica*, 62(2):405-458.

Buchinsky, Moshe. 1998. 'Recent Advances in Quantile Regressions Models: A Practical Guideline for Empirical Research.' *Journal of Human Resources* 33(1):88-126.

Buchinsky, Moshe. 2001. 'Quantile Regression with Sample Selection: Estimating Women's Return to Education in the U.S.' *Empirical Economics* 26:87-113.

Deaton, A. 1997. *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*, The World Bank, Washington, DC.

Even, William. 1990. 'Sex Discrimination in Labor Markets: The Role of Statistical Evidence: Comment' *American Economic Review* 80:287-289.

Garcia, Jaume, Pedro J. Hernandez, and Angel Lopez-Nicolas. 2001. 'How wide is the gap? An Investigation of Gender Wage Differences using Quantile Regression.' *Empirical Economics* 23:149-167.

Gill, Indermit. 1992. 'Is there Sex Discrimination in Chile? Evidence from the CASEN Survey', in G. Psacharopoulos and Z. Tzannatos (eds.) *Case Studies on Women's Employment and Pay in Latin America*, World Bank Regional and Sectoral Studies. The World Bank, Washington, DC.

Gill, Indermit and Claudio E. Montenegro. Forthcoming. 'Responding to Earning Differentials in Chile'. In '*Crafting Labor Policy: Techniques and Lessons from Latin America*,' Indermit Gill, ~~and~~Claudio E. Montenegro, and Dorte Dormeland (eds.). The World Bank, Washington, DC.

Gronau, Reuben. 1974. 'The Effect of Children on the Housewife's Value of Time' in T. W. Schultz (ed.) *Economics of the Family*. Chicago: University of Chicago Press.

Gunderson, Morley. 1989. 'Male-Female Wage Differentials and Policy Responses.' *Journal of Economic Literature* 27(1):46-72.

Heckman, James. 1979. 'Sample Selection Bias as Specification Error.' *Econometrica*, 47(1):153-161.

Killingsworth, M. R. 1983, *Labor Supply*. Cambridge: Cambridge University Press.

Koenker, Roger and Gilbert Basset Jr. 1978. 'Regression Quantiles.' *Econometrica* 46:33-50.

Koenker, Roger and Gilbert Basset Jr.. 1982. 'Test of Linear Hypothesis and l_1-Estimation.' *Econometrica* 50: 1577-1583.

Kuhn, P. 1987. 'Sex Discrimination in Labor Markets: The Role of Statistical Evidence.' *American Economic Review* 77:567-583.

Kuhn, P. 1990. 'Sex Discrimination in Labor Markets: The Role of Statistical Evidence: Reply' *American Economic Review* 80:290-297.

Liu, Pak-Wai, Xin Meng, and Junsen Zhang. 2000. 'Sectoral Gender Wage Differentials and Discrimination in the Transitional Chinese Economy.' *Journal of Population Economics* Vol. 13: 331-352.

Manski, C. 1989. 'Anatomy of the Selection Problem.' *Journal of Human Resources* 24(3):343-360.

Meng, Xin and Paul Miller. 1995. 'Occupational Segregation and its Impact on Gender Wage Discrimination in China's Rural Industrial Sector.' *Oxford Economic Papers* 47(1):136-55.

Mincer, Jacob. 1974. *Schooling, Experience and Earnings*. New York: The National Bureau of Economic Research.

Montenegro, Claudio E. 1998. 'The Structure of Wages in Chile 1960-1996: An Application of Quantile Regression.' *Estudios de Economia* 25(1):71-98.

Newell, Andrew and Barry Reilly. 2001. 'The Gender Pay Gap in the Transition from Communism: Some Empirical Evidence.' Institute for the Study of Labor (IZA) discussion paper No. 268, March 2001.

Oaxaca, Ronald. 1973. 'Male-Female Differentials in Urban Labor Markets'. *International Economic Review* 14:693-709.

Paredes, Ricardo. 1982. 'Diferencias de ingreso entre hombres y mujeres en el Gran Santiago 1969 y 1981.' *Estudios de Economia* 18:99-121.

Paredes, Ricardo. and Riveros, L. 1994. 'Gender Wage Gaps in Chile. A Long Term View: 1958-1990.' *Estudios de Economia*.

Polachek, S. and Siebert, W. S. 1993. *The Economics of Earnings*. Cambridge; New York and Melbourne: Cambridge University Press.

Powell, James. 1984. 'Least Absolute Deviation Estimation of the Censored Regression Model.' *Journal of Econometrics* 25:303-325.

Puhani, Patrick A. 2000. 'The Heckman Correction for Sample Selection and its Critique.' *Journal of Economic Surveys* 14(1):53-68.

Schultz, T. Paul. 1993. 'Returns to Women's Education.' In *Women's Education in Developing Countries: Barriers, Benefits, and Policies*.' Baltimore and London: Johns Hopkins University Press for the World Bank.

Tzannatos, Zafiris 1994. 'The Cost of Ethnic Discrimination to Society and Minorities: A Note on the Methodology and results from Guatemala' (mimeo, The World Bank).

United Nations Development Program. 2000. *Human Development Report 2000: Human Rights and Human Development*. New York: United Nations.

Willis, Robert J. 1986. 'Wage Determinants, A Survey and Reinterpretation of Human Capital Earnings Functions.' In Orley Ashenfelter and Richard Layard (eds.) *The Handbook of Labor Economics*, Vol. 1, Amsterdam: North Holland-Elsevier Science Publishers; pp. 525-602.

Wooden, Mark. 1999. 'Gender Pay Equity and Comparable Worth in Australia: A Reassessment.' *Australian Economic Review*, Vol. 32(2):157-171.

Table 1: Basic Statistics of the Sample															
	1990			1992			1994			1996			1998		
	wage	educ	exp	wage	educ	exp	wage	educ	exp	wage	educ	exp	wage	educ	exp
Total Sample															
N	21838	21838	21838	30919	30919	30919	34935	34935	34935	28942	28942	28942	41653	41653	41653
Mean	408.2	10.9	17.8	575.2	10.7	18.4	749.1	11.0	18.7	1111	11.2	18.5	1291	11.3	18.9
Std.	576.6	4.0	12.7	808.7	4.0	13.2	932.1	4.0	13.0	1685	4.0	12.8	2254	4.0	13.0
Med.	239.7	12.0	15.0	340.6	12.0	16.0	476.2	12.0	16.0	671.3	12.0	16.0	784.4	12.0	17.0
Skew.	8.7	-0.4	0.9	6.1	-0.4	0.9	5.4	-0.5	0.8	11.8	-0.4	0.8	35.8	-0.4	0.8
Kurt	141.9	2.9	3.2	62.2	2.9	3.2	53.5	3.0	3.1	291.6	3.0	3.3	2682	3.1	0.3
Men															
N	16331	16331	16331	23313	23313	23313	26381	26381	26381	20992	20992	20992	29943	29943	29943
Mean	432.2	10.4	18.9	598.5	10.2	19.5	776.8	10.4	19.9	1137	10.7	19.6	1308	10.7	20.1
Std.	650.7	4.1	13.1	866.3	4.1	13.5	1012	4.1	13.4	1808	4.1	13.1	2527	4.1	13.4
Med.	246.3	11.0	16.0	344.7	11.0	17.0	471.2	11.0	17.0	651.0	12.0	17.0	770.4	12.0	18.0
Skew.	8.3	-0.3	0.8	5.9	-0.2	0.8	5.2	-0.4	0.7	12.0	-0.3	0.8	37.2	-0.3	0.8
Kurt	122.6	2.8	3.1	58.1	2.8	3.0	49.0	2.8	3.0	297.2	2.9	3.1	2538	3.0	3.1
Women															
N	5507	5507	5507	7606	7606	7606	8554	8554	8554	7950	7950	7950	12230	12230	12230
Mean	353.0	12.1	15.3	518.1	11.9	15.6	683.2	12.4	15.9	1055	12.4	16.1	1257	12.4	16.5
Std.	339.8	3.6	11.5	644.3	3.6	11.8	699.0	3.4	11.6	1385	3.5	11.6	1575	3.5	11.8
Med.	239.7	12.0	12.0	323.6	12.0	13.0	485.1	12.0	14.0	714.3	12.0	14.0	784.4	12.0	14.0
Skew.	3.8	-0.6	0.9	6.3	-0.7	1.0	5.4	-0.9	0.8	9.4	-0.8	0.8	7.9	-0.7	0.8
Kurt	29.1	3.4	3.3	64.6	3.6	3.7	53.3	4.1	3.4	157.5	4.1	3.4	157.8	3.8	3.2

Note: Wages are in hourly nominal values.

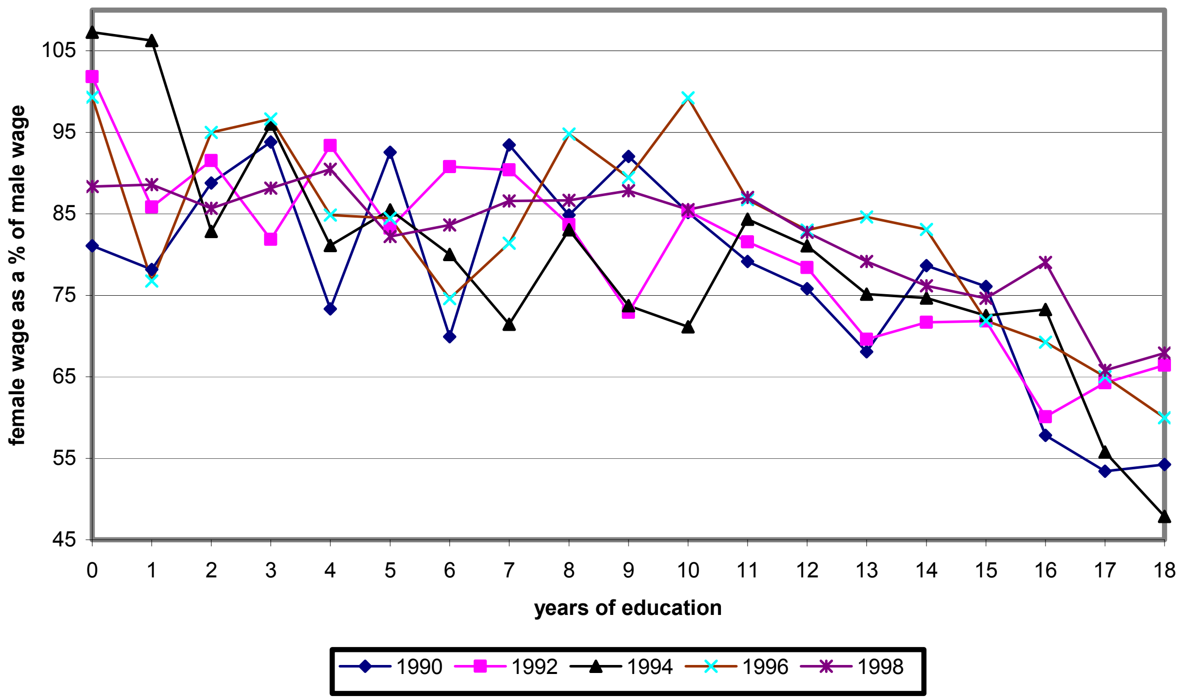
Table 2: Oaxaca Decomposition by Quantiles

	Male Wages					Female Wages					Female Wages at Men's β's				
	1990	1992	1994	1996	1998	1990	1992	1994	1996	1998	1990	1992	1994	1996	1998
θ=0.10	148	200	270	354	429	144	194	278	360	440	160	217	302	404	476
θ=0.15	166	227	303	410	487	162	220	310	415	491	181	247	341	468	545
θ=0.20	183	254	336	459	540	176	240	342	460	540	200	277	380	527	611
θ=0.25	198	276	368	508	592	191	259	371	497	587	216	302	419	587	672
θ=0.30	215	302	399	554	642	201	280	403	541	637	237	333	456	643	736
θ=0.35	233	325	432	604	692	215	303	433	580	684	258	358	497	702	797
θ=0.40	252	354	464	652	745	231	323	458	625	739	279	391	534	759	864
θ=0.45	270	382	502	705	800	248	344	490	677	791	300	423	579	822	928
θ=0.50	293	408	540	757	868	263	368	518	724	846	326	455	627	885	1010
θ=0.55	318	444	585	818	936	279	396	551	784	906	355	495	681	956	1092
θ=0.60	345	481	635	890	1011	299	426	589	830	963	386	537	740	1043	1183
θ=0.65	382	525	697	972	1091	326	457	631	903	1035	429	589	817	1140	1281
θ=0.70	421	577	764	1067	1190	354	491	673	983	1126	474	650	898	1257	1398
θ=0.75	471	636	847	1180	1317	388	545	737	1070	1232	532	714	999	1395	1554
θ=0.80	521	716	947	1328	1458	435	613	815	1171	1354	593	803	1122	1574	1721
θ=0.85	598	835	1085	1528	1673	492	701	908	1297	1554	684	942	1290	1807	1991
θ=0.90	723	990	1272	1842	1998	610	831	1052	1566	1885	831	1129	1524	2174	2401
ols	309	428	564	784	899	279	385	533	747	883	349	483	662	921	1056

	Total Wage Gap as % of Female Wages in Quantile					Gap Due to Gender Differences in Endowments as % of Female Wages in Quantile					Gap Due to Gender Differences in Returns to Endowments as % of Female Wages in Quantile				
	1990	1992	1994	1996	1998	1990	1992	1994	1996	1998	1990	1992	1994	1996	1998
θ=0.10	2	3	-3	-2	-2	-9	-9	-12	-14	-10	11	12	9	12	8
θ=0.15	2	3	-2	-1	-1	-9	-9	-12	-14	-12	11	12	10	13	11
θ=0.20	4	6	-2	-0	-0	-10	-10	-13	-15	-13	14	15	11	15	13
θ=0.25	4	6	-1	2	1	-10	-10	-14	-16	-14	13	17	13	18	15
θ=0.30	7	8	-1	2	1	-11	-11	-14	-17	-15	18	19	13	19	15
θ=0.35	9	7	-0	4	1	-12	-11	-15	-17	-15	20	18	15	21	16
θ=0.40	9	10	1	4	1	-12	-12	-15	-17	-16	20	21	17	21	17
θ=0.45	9	11	2	4	1	-12	-12	-16	-17	-16	21	23	18	21	17
θ=0.50	12	11	4	5	3	-12	-13	-17	-18	-17	24	24	21	22	19
θ=0.55	14	12	6	4	3	-13	-13	-17	-18	-17	27	25	24	22	21
θ=0.60	15	13	8	7	5	-14	-13	-18	-18	-18	29	26	26	26	23
θ=0.65	17	15	10	8	5	-14	-14	-19	-19	-18	32	29	29	26	24
θ=0.70	19	18	13	9	6	-15	-15	-20	-19	-18	34	32	33	28	24
θ=0.75	21	17	15	10	7	-16	-14	-21	-20	-19	37	31	36	30	26
θ=0.80	20	17	16	13	8	-17	-14	-21	-21	-19	36	31	38	34	27
θ=0.85	21	19	20	18	8	-17	-15	-23	-22	-20	39	34	42	39	28
θ=0.90	18	19	21	18	6	-18	-17	-24	-21	-21	36	36	45	39	27
ols	11	11	6	5	2	-14	-14	-18	-18	-18	25	25	24	23	20

Note: Wages are in hourly nominal values.

Figure 1. Female to Male Average Wage
(a) By years of Education



(b) By Years of Experience

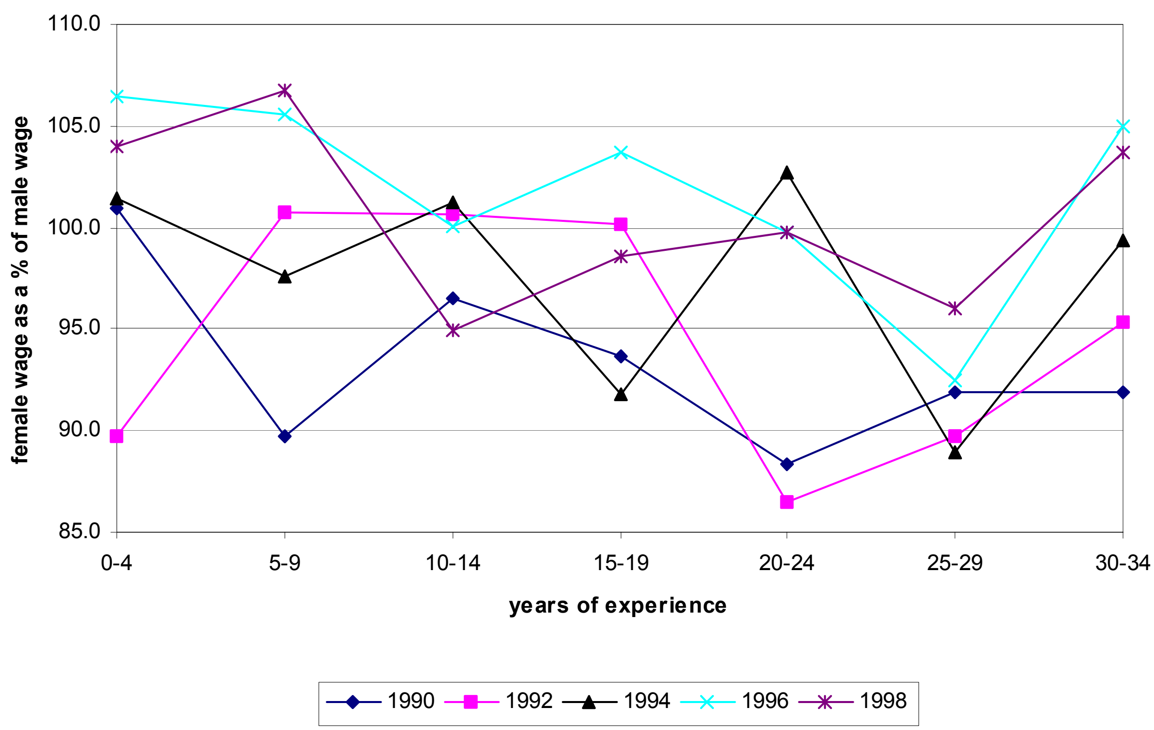
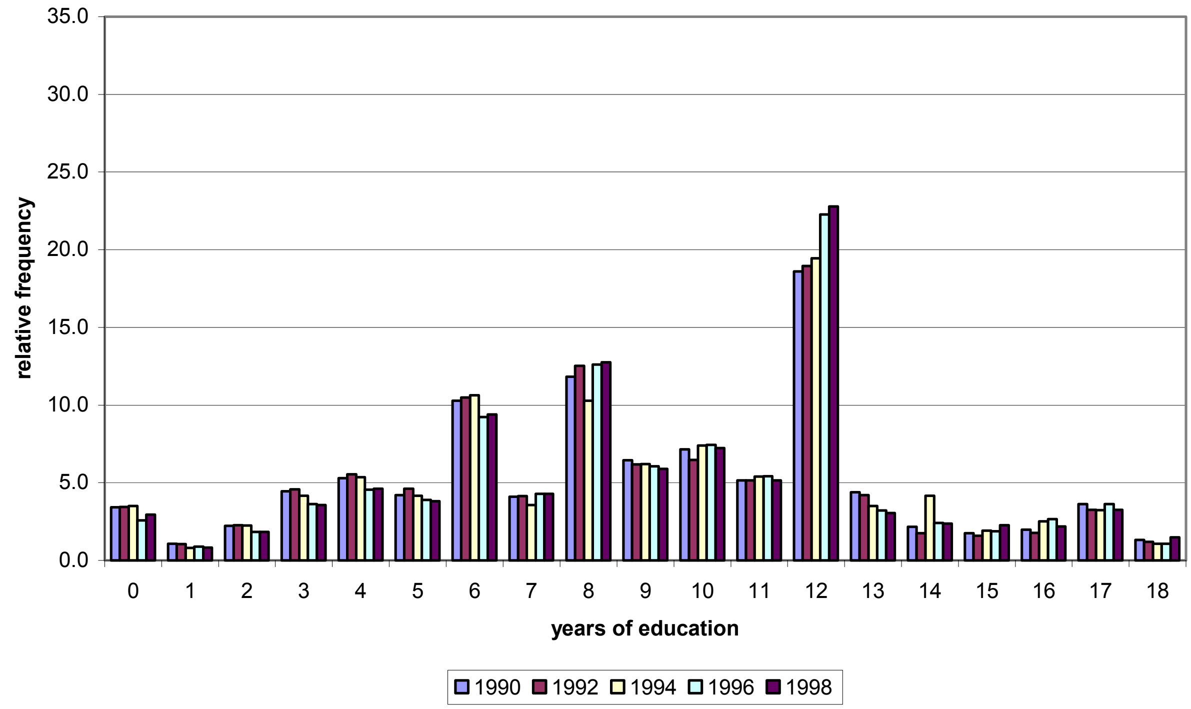


FIGURE 2. UNIVARIATE DISTRIBUTION OF EDUCATION

(a) Men



(b) Women

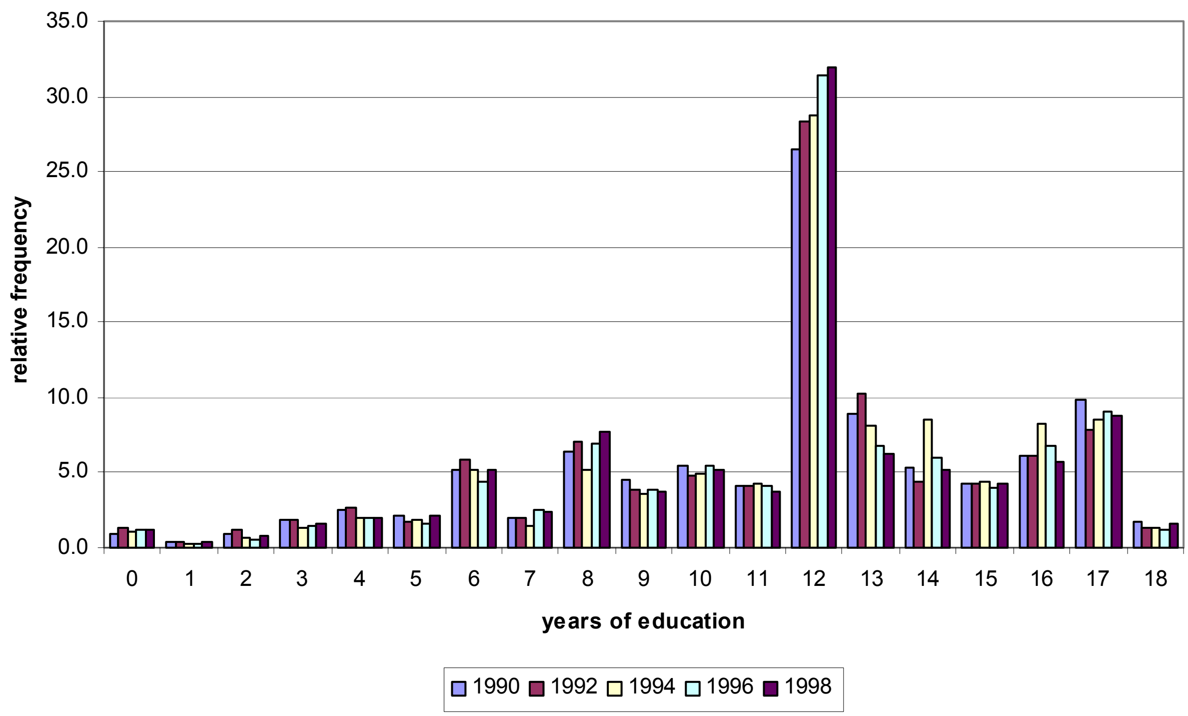
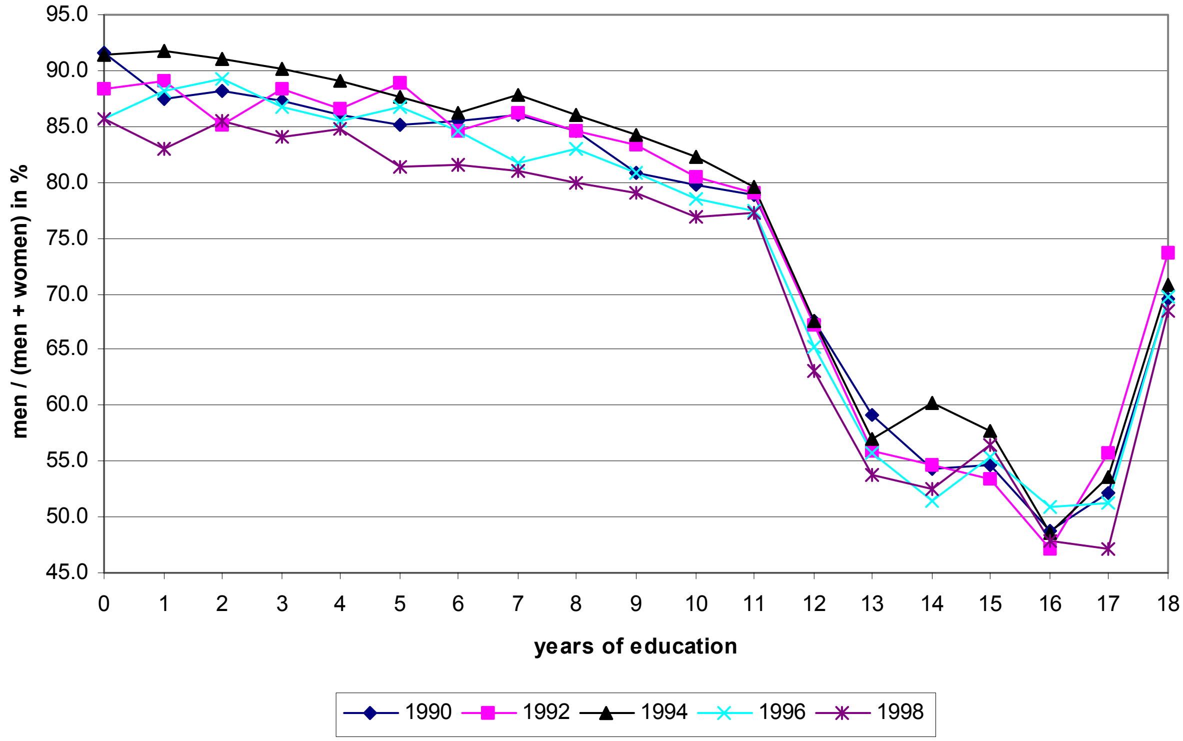


Figure 3. Proportion of Men

(a) By Years of Education



(b) By Years of Experience

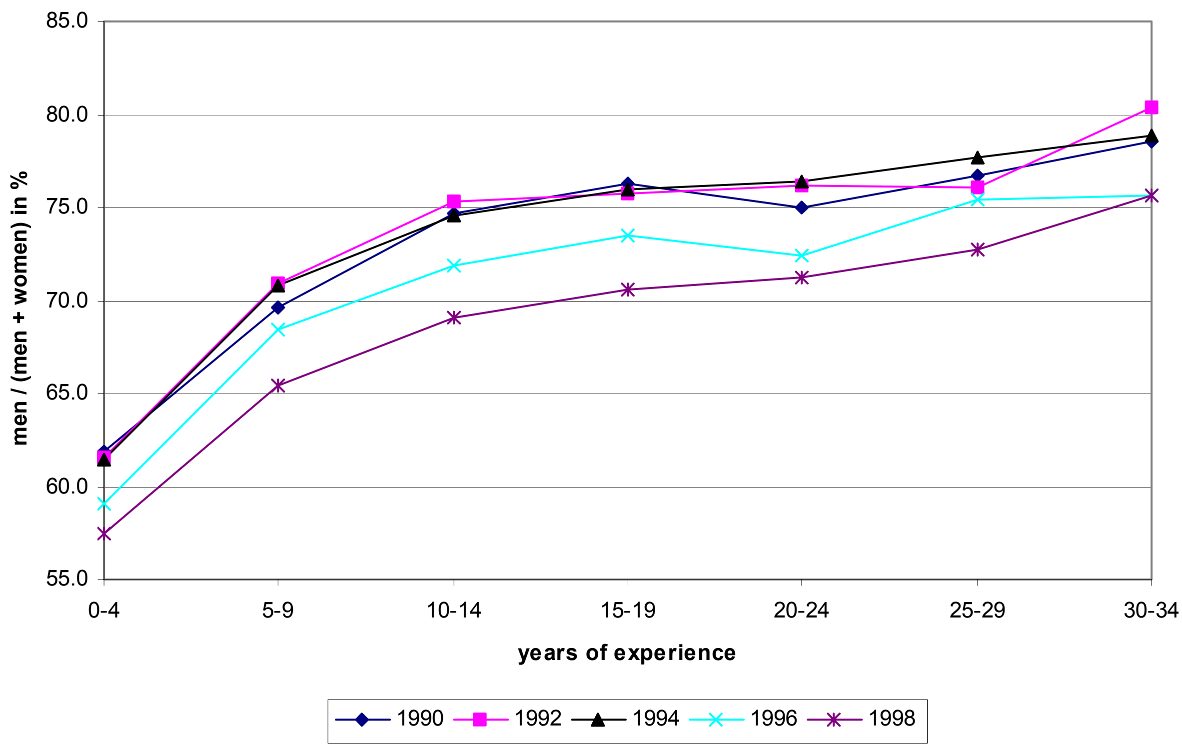
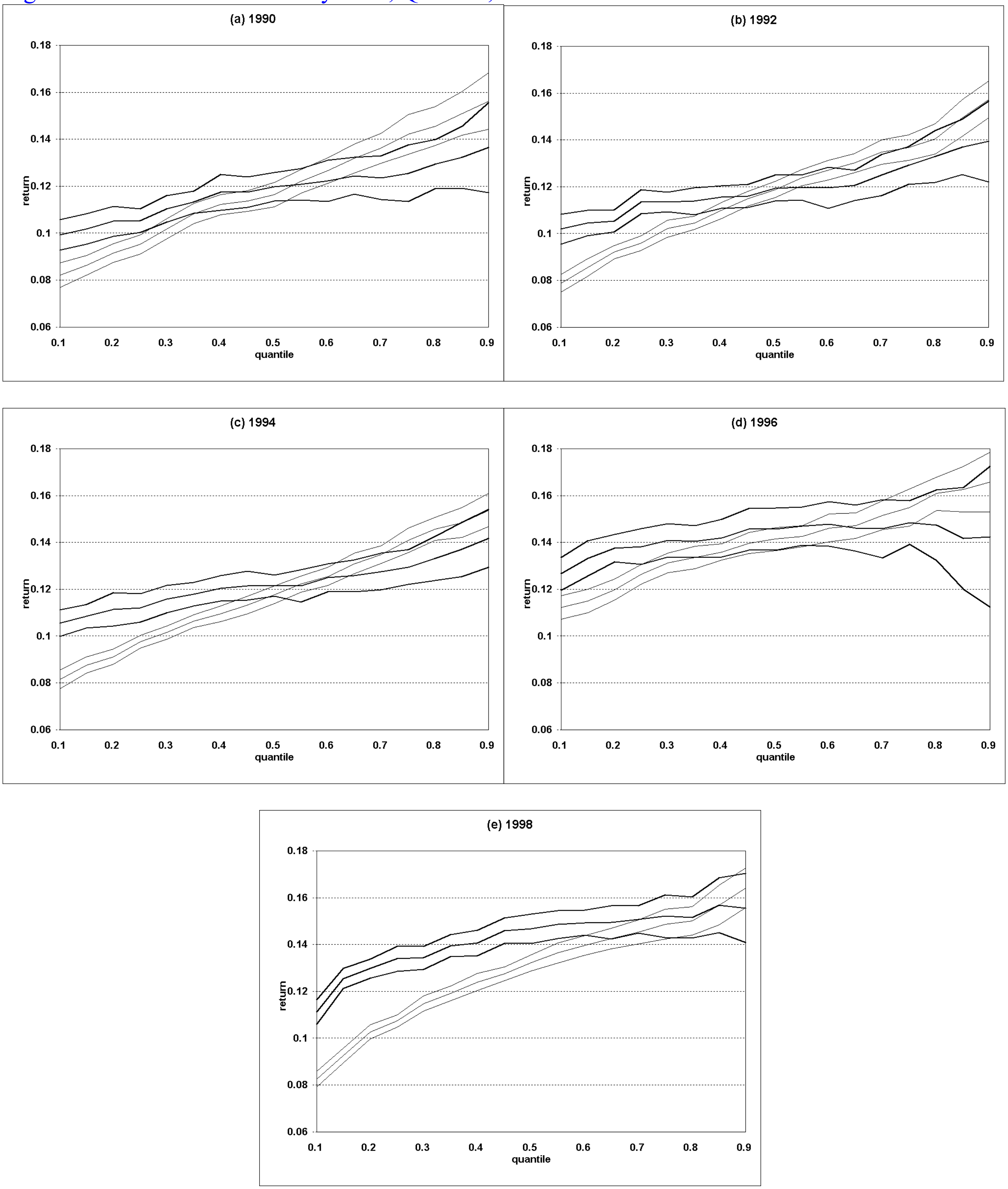
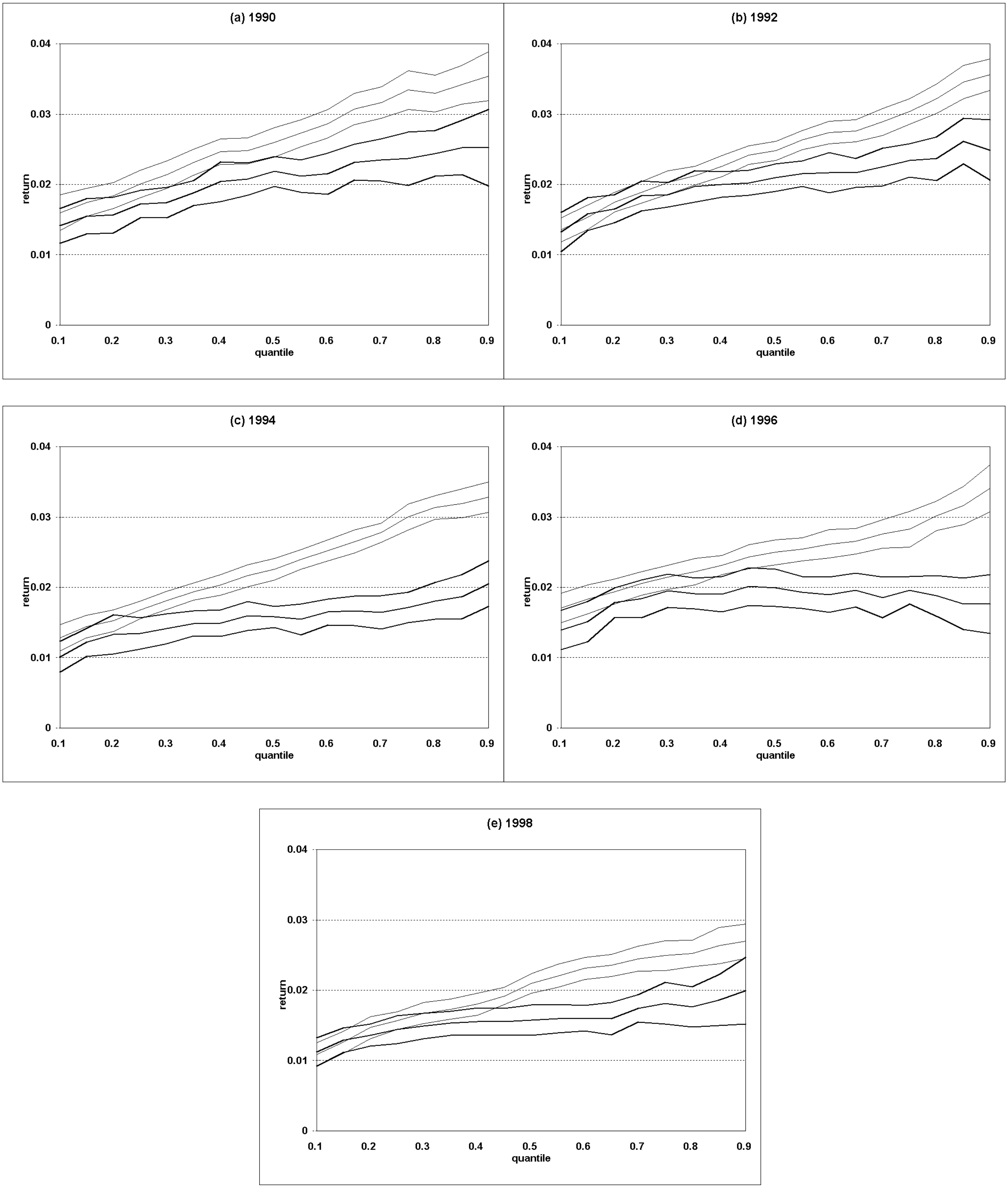


Figure 4. Return to Education by Year, Quantile, and Gender



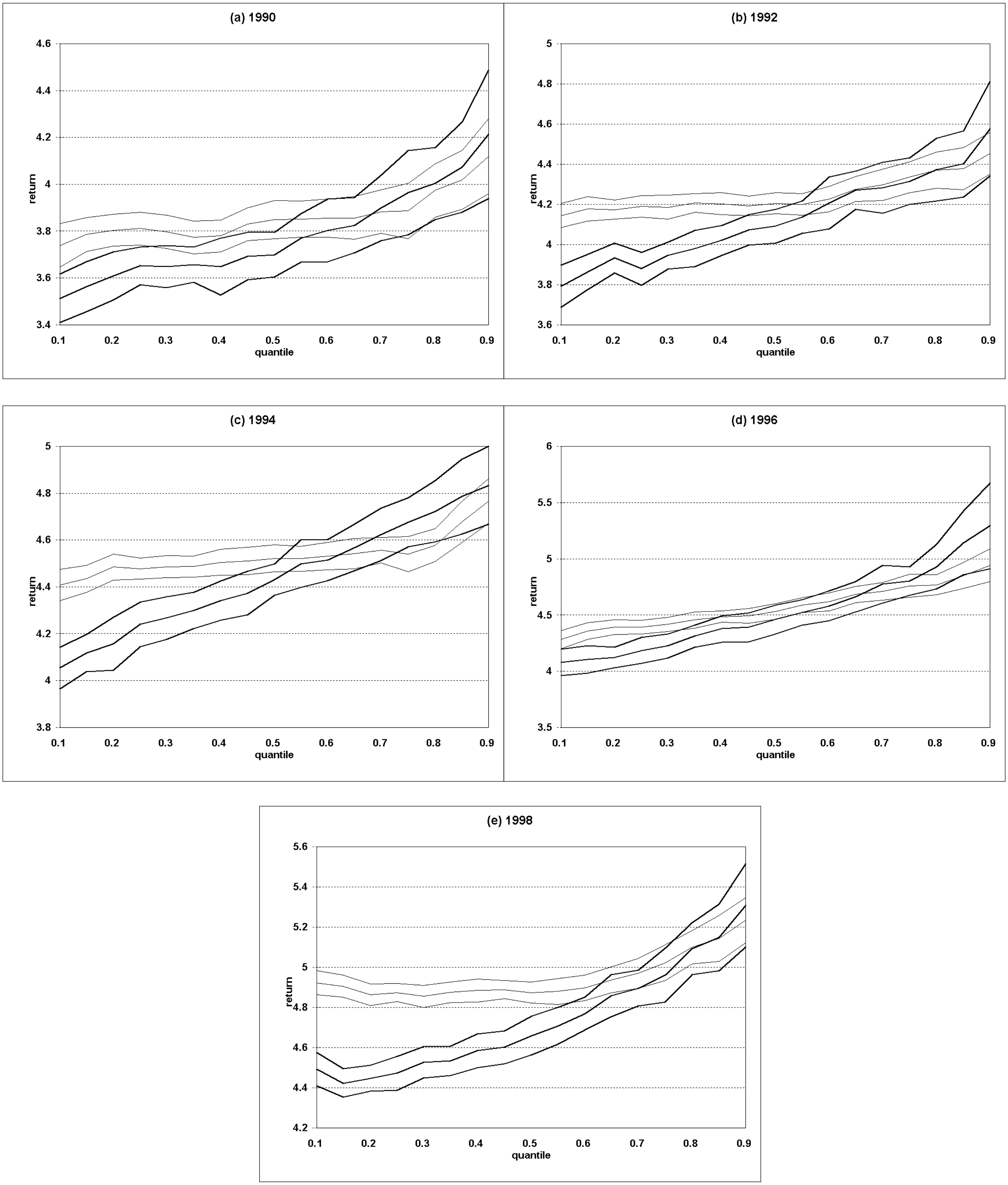
Note: thin lines represent men and thick lines represent women.

FIGURE 5. RETURN TO EXPERIENCE BY YEAR, QUANTILE, AND GENDER



Note: thin lines represent men and thick lines represent women.

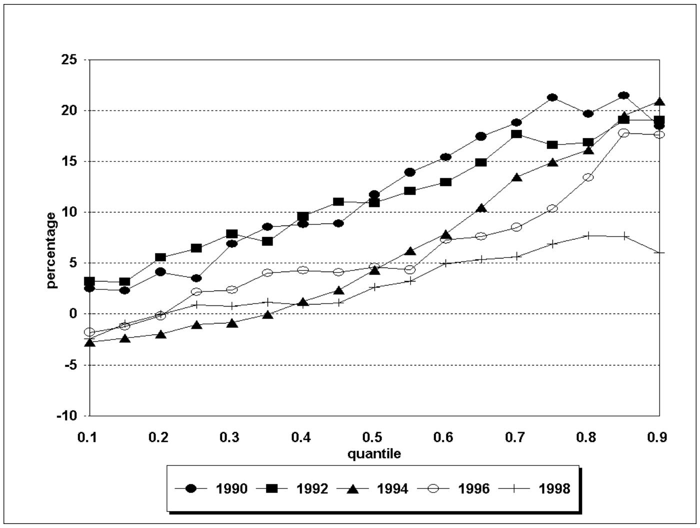
FIGURE 6. CONSTANT BY YEAR, QUANTILE, AND GENDER



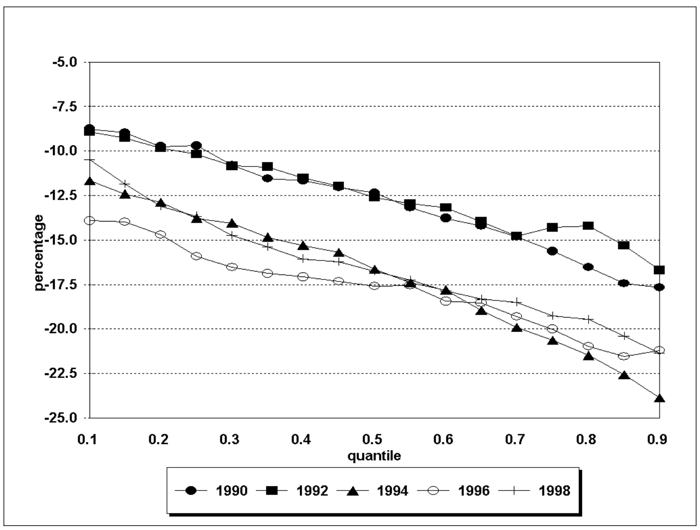
Note: thin lines represent men and thick lines represent women.

Figure 7. Gaps As a Proportion of Females Wages

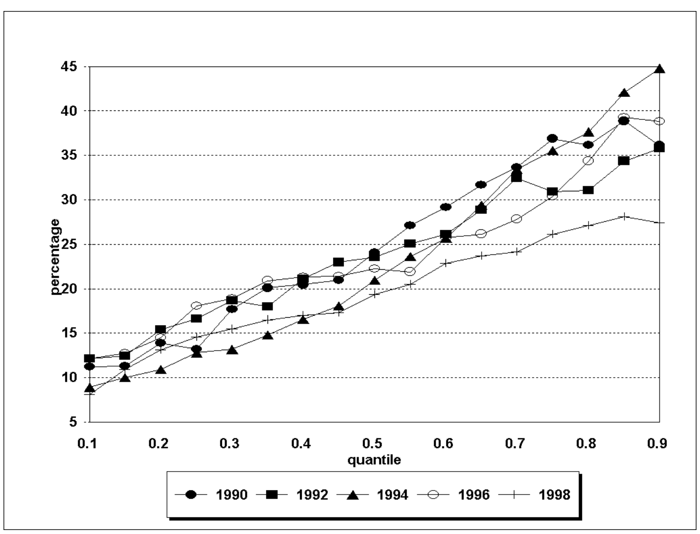
(a) Total Wage Gap



(b) Gap due to Differences in Endowments



(c) Residual Gap



Methodological Appendix: The Quantile Regression Approach

The standard econometric model used to estimate equation (1) is OLS, which implies the minimization of

$$\psi_{OLS} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - x_i'\beta)^2 \qquad \text{(A.1)}$$

The method is based on the following assumptions:

(i) $y_i = \beta' x_i + u_i \quad \forall\, i$, i = 1, 2, ... n, or, alternatively, Y = Xβ + U (true model)

(ii) $E(u_i|x_i) = 0 \;\; \forall \;\; i$

(iii) X is a non-stochastic matrix n *x* k with rank k.

(iv) U ~ N(0,σ I). Note that this implies that $E(u_i) = 0$ and that E(U'U)=σI.

The previous formulation and assumptions imply that $E(y_i|x_i) = \beta' x_i$. In other words, the minimization implies that the fitted curve is the prediction of the *mean* of y_i (the dependent variable) given the values of a certain vector of independent variables x_i. When disturbances are non-normal or when there is a relatively large proportion of outliers the traditional mean regression estimation method is not robust. This has led to the study of an alternative method that does not minimize the sum of the squared errors but minimizes instead the sum of the absolute value of the errors. The Least Absolute Deviation (LAD) technique implies the minimization of:

$$\psi_{LAD} = \sum_{i=1}^{n}|y_i - \hat{y}_i| = \sum_{i=1}^{n}|y_i - x_i'\beta| = \sum_{i=1}^{n}(y_i - x_i'\beta)\;\; sign(y_i - x_i'\beta) \qquad \text{(A.2)}$$

where sign(R) is 1 if R is non-negative and -1 if R is negative. In the case where there is just one constant and no covariates in equation (A.2), the LAD estimator becomes the median of y and so it is the value that leaves 50% of the observations above the median value and the remaining 50% below that value. The intuition for this result (Deaton, 1997) comes from thinking about the first order conditions that are satisfied by the parameters that minimize Eq. (4), which are: $\sum_{i=1}^{n} x_{ij}\, sign(y_i - x_i'\beta) = 0$, ∀j, j=1, 2, ...

k. When there is only a constant in the model this equation becomes $\sum_{i=1}^{n} sign(y_i - \beta_0) = 0$. This implies that the constant should be chosen so that there are equal **number** of observations on either side of it, and so β_0 is the median value of y. The estimation of β when there are covariates is a linear programming problem that can be solved using linear programming techniques as it is shown in Koenker and Bassett (1978, 1982). The solution does not have an explicit form.

The obvious natural extension of equation (A.2) is to fit different curves that would leave different percentages of the observations above and below the fitted curve. This can be achieved by weighting differently positive and negative errors in equation (A.2). This is the origin of the quantile regression method (QRM). A very common misconception about the QRM is that is equivalent to adjusting different curves to different sub-samples, where the sub-samples are being determined by the dependent variable. This is not the case. In each QRM estimation the same number of observations is used: the total sample. Another common misconception about the QRM is to think about the returns to education or experience in the upper vs. the lower quantiles as if this were capturing differences between rich and poor people. This is not the case. The distribution is of the residuals, after compensating for education and experience. In other words, those individuals who are high in the distribution are those who are earning above what would be expected given their level of human capital, which may include rich and poor people. Using θ and (1-θ) to weight positive and negative errors in equation (A.2) we proceed to present the QRM in a more formal way.

Let y_i (i=1,.....,n) be the wage of individual i and let x_i be a known vector of covariates. Let us assume that the θth quantile of the conditional distribution of y_i given x_i is linear, that is,

$$y_i = x_i'\beta_\theta + u_{\theta i} \quad with \quad Q_\theta(y_i | x_i) = x_i'\beta_\theta \quad i = 1, 2, ..., n \qquad \text{(A.3)}$$

where x'_i is a k x 1 vector of covariates with $x_{i1}=1$, and β_θ for all i, and is an unknown k *x* 1 parameter vector whose estimation, for different values of θ [0 < θ < 1] is in our interest. The term $Q_\theta(y_i|x_i)$ denotes the conditional quantile of y_i given x_i. In this specification $u_{\theta i}$ is defined by $u_{\theta i} = y_i - x_i'\beta_\theta$, from where it follows that $Q_\theta(u_{\theta i}|x_i)=0$. Therefore, the θth quantile regression based on a sample (y_i,x_i) i=1, 2, ..., n, is a vector β_θ that minimizes:

$$\psi_{QRM,\theta} \quad = \quad \theta \sum_{i|y_i \geq x_i'\beta_\theta}^{n} |y_i - x_i'\beta_\theta| \; + (1-\theta) \sum_{i|y_i < x_i'\beta_\theta}^{n} |y_i - x_i'\beta_\theta| \qquad \text{(A.4)}$$

The QRM technique is particularly useful for our objectives not only because it is a robust method, but also because it permits us to compute several regression curves corresponding to different conditional quantiles in the distribution. It robustness comes from the fact that the QRM method fits hyperplanes among the observations so that a certain proportion θ of the observations will be below of the hyperplane and the rest above it. What is important in this type of estimation, is the number of errors that are positive or negative, but not their magnitude. Put it in another way, this means that the vector $\hat{\beta}_\theta$ that minimizes equation (A.4) is invariant to the presence of outliers. The method has the potential of generating different solutions (i.e. distinct β's) at different quantiles. In theory we can have as many different conditional derivatives (conditional quantile estimates) $\partial Q_\theta(y_i|x_i)/\partial x_j$ as we want. In reality (given that the any data set is finite) only a finite number of distinct β can be estimated. These different "responses" may be interpreted as differences in the response of the dependent variable to changes in the regressors at various points in the conditional distribution of the dependent variable. In a regression analysis, where the errors are i.i.d., the conditional quantiles become a set of parallel hyperplanes and thus, we would expect the estimated coefficients be the same for different quantiles (with the obvious exception of β_0). In this case, not much information is lost just using simple OLS. But even in this set up the QRM technique is useful to characterize the distribution (the bigger the difference among the β_0's for two different θ's, the "flatter" the distribution).

Inference can also be performed in the QRM. It can be shown that the first order conditions implied by Eq. 6 can be used along with the Generalized Method of Moments to establish consistence and asymptotic normality of $\hat{\beta}_0$ (Koenker and Basset, 1978 and 1982; Powell, 1984; Buchinsky, 1998). If

$\lim[(X'X)/n] \rightarrow V$ with $X_{i1}=1$ for all i, and the error distribution, F, has strictly positive density at the θth quantile [i.e., $f(F^{-1}(\theta))>0$], then $\hat{\beta}_\theta$ is asymptotically normal, i.e.,

$$\sqrt{n}(\hat{\beta}_\theta - \beta_\theta) \xrightarrow{d} N(0, g^2(\theta, F)V^{-1}) \qquad \text{(A.5)}$$

where $\beta_\theta=\beta +(F^{-1}(\theta), 0,....., 0)'$ and $g^2(\theta,F)=\theta(1-\theta)/f^2(F^{-1}(\theta))$. From this, it follows that a natural statistical test for H_o: $M\beta=m$ is:

$$\lambda = \hat{g}^{-2}(M\hat{\beta}_\theta - m)'(M(X'X)^{-1}M')(M\hat{\beta}_\theta - m) \qquad \text{(A.6)}$$

where $\hat{g}$ is a consistent estimator of g. The accuracy with which the parameters of the vector β_θ are estimated depends on the magnitude of the density for each quantile. This means that quantiles in the upper or lower tale of the distribution, where the density is low, are more difficult to estimate and the corresponding test have less power.

The estimation of the variance-covariance matrix is not an easy task. Koenker and Basset (1982) derive an asymptotic formula that provides with an estimate for the variance-covariance matrix. However, when the residuals are heteroscedastic, the formula appears to grossly underestimate the true standard errors. In order to have a more reliable estimate of the variance-covariance matrix we used the bootstrapping method to calculate standard errors in our empirical tests.

Annex Table I: Parameters Estimated

		MEN					WOMEN				
quantile	**variable**	**1990**	**1992**	**1994**	**1996**	**1998**	**1990**	**1992**	**1994**	**1996**	**1998**
OLS	**schooling**	0.1284	0.1281	0.1247	0.1450	0.1365	0.1202	0.1245	0.1231	0.1397	0.1457
0.10	**schooling**	0.0821	0.0788	0.0815	0.1123	0.0825	0.0993	0.1020	0.1056	0.1266	0.1112
0.15	**schooling**	0.0863	0.0856	0.0877	0.1150	0.0926	0.1019	0.1046	0.1085	0.1332	0.1255
0.20	**schooling**	0.0916	0.0921	0.0912	0.1199	0.1027	0.1051	0.1055	0.1114	0.1376	0.1298
0.25	**schooling**	0.0952	0.0959	0.0977	0.1265	0.1074	0.1054	0.1136	0.1120	0.1383	0.1339
0.30	**schooling**	0.1019	0.1022	0.1016	0.1312	0.1149	0.1104	0.1135	0.1158	0.1409	0.1344
0.35	**schooling**	0.1084	0.1046	0.1065	0.1337	0.1191	0.1132	0.1138	0.1179	0.1405	0.1395
0.40	**schooling**	0.1122	0.1099	0.1096	0.1360	0.1240	0.1174	0.1156	0.1204	0.1419	0.1407
0.45	**schooling**	0.1137	0.1150	0.1133	0.1397	0.1275	0.1175	0.1161	0.1215	0.1457	0.1459
0.50	**schooling**	0.1164	0.1188	0.1177	0.1415	0.1323	0.1198	0.1195	0.1217	0.1458	0.1468
0.55	**schooling**	0.1221	0.1239	0.1223	0.1427	0.1364	0.1209	0.1197	0.1216	0.1469	0.1486
0.60	**schooling**	0.1266	0.1271	0.1255	0.1463	0.1397	0.1224	0.1196	0.1249	0.1479	0.1494
0.65	**schooling**	0.1318	0.1302	0.1310	0.1472	0.1426	0.1244	0.1206	0.1258	0.1462	0.1495
0.70	**schooling**	0.1362	0.1348	0.1348	0.1516	0.1454	0.1237	0.1250	0.1276	0.1459	0.1507
0.75	**schooling**	0.1421	0.1367	0.1409	0.1550	0.1487	0.1255	0.1291	0.1295	0.1486	0.1522
0.80	**schooling**	0.1456	0.1405	0.1458	0.1609	0.1501	0.1294	0.1329	0.1332	0.1474	0.1517
0.85	**schooling**	0.1511	0.1495	0.1486	0.1628	0.1569	0.1323	0.1370	0.1371	0.1417	0.1568
0.90	**schooling**	0.1562	0.1573	0.1538	0.1658	0.1642	0.1364	0.1394	0.1418	0.1425	0.1557
OLS	**experience**	0.0428	0.0388	0.0360	0.0357	0.0272	0.0322	0.0283	0.0236	0.0204	0.0160
0.10	**experience**	0.0280	0.0240	0.0232	0.0234	0.0154	0.0191	0.0199	0.0222	0.0154	0.0147
0.15	**experience**	0.0294	0.0240	0.0240	0.0263	0.0171	0.0225	0.0206	0.0217	0.0185	0.0135
0.20	**experience**	0.0307	0.0271	0.0248	0.0282	0.0190	0.0214	0.0207	0.0235	0.0196	0.0149
0.25	**experience**	0.0327	0.0290	0.0263	0.0301	0.0209	0.0247	0.0229	0.0220	0.0197	0.0149
0.30	**experience**	0.0344	0.0309	0.0284	0.0306	0.0221	0.0237	0.0244	0.0224	0.0203	0.0170
0.35	**experience**	0.0368	0.0327	0.0305	0.0316	0.0224	0.0256	0.0282	0.0234	0.0193	0.0167
0.40	**experience**	0.0384	0.0349	0.0314	0.0334	0.0230	0.0281	0.0281	0.0216	0.0184	0.0177
0.45	**experience**	0.0389	0.0372	0.0336	0.0353	0.0247	0.0303	0.0289	0.0233	0.0196	0.0164
0.50	**experience**	0.0416	0.0382	0.0344	0.0365	0.0278	0.0323	0.0288	0.0227	0.0193	0.0158
0.55	**experience**	0.0423	0.0406	0.0365	0.0366	0.0292	0.0298	0.0316	0.0223	0.0201	0.0153
0.60	**experience**	0.0442	0.0415	0.0389	0.0375	0.0308	0.0316	0.0319	0.0226	0.0191	0.0142
0.65	**experience**	0.0473	0.0421	0.0402	0.0386	0.0309	0.0351	0.0315	0.0232	0.0216	0.0121
0.70	**experience**	0.0488	0.0436	0.0431	0.0397	0.0323	0.0373	0.0313	0.0215	0.0194	0.0142
0.75	**experience**	0.0523	0.0461	0.0467	0.0405	0.0333	0.0375	0.0335	0.0229	0.0212	0.0146
0.80	**experience**	0.0503	0.0490	0.0477	0.0432	0.0338	0.0401	0.0348	0.0229	0.0194	0.0114
0.85	**experience**	0.0529	0.0524	0.0479	0.0457	0.0346	0.0416	0.0396	0.0218	0.0158	0.0123
0.90	**experience**	0.0557	0.0532	0.0487	0.0503	0.0347	0.0448	0.0367	0.0255	0.0193	0.0169
OLS	**exp squared**	-0.0005	-0.0004	-0.0004	-0.0003	-0.0002	-0.0004	-0.0002	-0.0002	-0.0001	0.0000
0.10	**exp squared**	-0.0004	-0.0003	-0.0003	-0.0002	-0.0001	-0.0002	-0.0002	-0.0004	-0.0000	-0.0001
0.15	**exp squared**	-0.0004	-0.0003	-0.0003	-0.0002	-0.0001	-0.0002	-0.0001	-0.0003	-0.0001	-0.0000
0.20	**exp squared**	-0.0004	-0.0003	-0.0003	-0.0003	-0.0001	-0.0002	-0.0001	-0.0003	-0.0001	-0.0000
0.25	**exp squared**	-0.0004	-0.0003	-0.0003	-0.0003	-0.0002	-0.0002	-0.0001	-0.0003	-0.0000	-0.0000
0.30	**exp squared**	-0.0004	-0.0003	-0.0003	-0.0003	-0.0002	-0.0002	-0.0002	-0.0002	-0.0000	-0.0001
0.35	**exp squared**	-0.0004	-0.0003	-0.0003	-0.0003	-0.0002	-0.0002	-0.0003	-0.0003	0.0000	-0.0000
0.40	**exp squared**	-0.0004	-0.0004	-0.0003	-0.0003	-0.0002	-0.0002	-0.0002	-0.0002	0.0000	-0.0001
0.45	**exp squared**	-0.0004	-0.0004	-0.0004	-0.0003	-0.0002	-0.0003	-0.0003	-0.0002	0.0000	-0.0000
0.50	**exp squared**	-0.0005	-0.0004	-0.0004	-0.0003	-0.0002	-0.0003	-0.0002	-0.0002	0.0000	0.0000
0.55	**exp squared**	-0.0004	-0.0004	-0.0004	-0.0003	-0.0002	-0.0003	-0.0003	-0.0002	-0.0000	0.0000
0.60	**exp squared**	-0.0005	-0.0004	-0.0004	-0.0003	-0.0002	-0.0003	-0.0003	-0.0002	-0.0000	0.0001
0.65	**exp squared**	-0.0005	-0.0004	-0.0004	-0.0004	-0.0002	-0.0004	-0.0003	-0.0002	-0.0001	0.0001
0.70	**exp squared**	-0.0005	-0.0004	-0.0005	-0.0004	-0.0002	-0.0004	-0.0003	-0.0002	-0.0000	0.0001
0.75	**exp squared**	-0.0006	-0.0005	-0.0005	-0.0004	-0.0003	-0.0004	-0.0003	-0.0002	-0.0001	0.0001
0.80	**exp squared**	-0.0005	-0.0005	-0.0005	-0.0004	-0.0003	-0.0005	-0.0003	-0.0001	-0.0000	0.0002
0.85	**exp squared**	-0.0006	-0.0005	-0.0005	-0.0004	-0.0002	-0.0005	-0.0004	-0.0001	0.0001	0.0002
0.90	**exp squared**	-0.0006	-0.0005	-0.0005	-0.0005	-0.0002	-0.0006	-0.0004	-0.0002	-0.0001	0.0001
OLS	**constant**	3.7543	4.1453	4.4719	4.5368	4.8753	3.7643	4.0868	4.4366	4.5740	4.7135
0.10	**constant**	3.7393	4.1452	4.4085	4.2817	4.9230	3.5139	3.7921	4.0540	4.0802	4.4931
0.15	**constant**	3.7856	4.1789	4.4347	4.3596	4.9054	3.5625	3.8625	4.1169	4.1054	4.4230
0.20	**constant**	3.8043	4.1737	4.4855	4.3930	4.8632	3.6083	3.9328	4.1570	4.1226	4.4473
0.25	**constant**	3.8114	4.1903	4.4778	4.3935	4.8745	3.6515	3.8793	4.2409	4.1858	4.4721
0.30	**constant**	3.7975	4.1858	4.4863	4.4176	4.8555	3.6486	3.9455	4.2677	4.2248	4.5273
0.35	**constant**	3.7738	4.2070	4.4870	4.4570	4.8755	3.6572	3.9812	4.2996	4.3117	4.5338
0.40	**constant**	3.7794	4.2033	4.5054	4.4856	4.8847	3.6486	4.0209	4.3421	4.3769	4.5844
0.45	**constant**	3.8303	4.1923	4.5104	4.4946	4.8885	3.6932	4.0723	4.3731	4.3919	4.6013
0.50	**constant**	3.8497	4.2061	4.5221	4.5310	4.8738	3.6995	4.0921	4.4310	4.4614	4.6598
0.55	**constant**	3.8507	4.1998	4.5208	4.5889	4.8799	3.7707	4.1365	4.4995	4.5240	4.7083
0.60	**constant**	3.8552	4.2275	4.5322	4.6187	4.8978	3.8044	4.2076	4.5149	4.5819	4.7695
0.65	**constant**	3.8564	4.2761	4.5429	4.6835	4.9372	3.8247	4.2700	4.5676	4.6617	4.8591
0.70	**constant**	3.8831	4.2971	4.5567	4.7098	4.9702	3.8993	4.2836	4.6249	4.7752	4.8962
0.75	**constant**	3.8863	4.3358	4.5406	4.7601	5.0232	3.9651	4.3152	4.6755	4.8037	4.9612
0.80	**constant**	3.9738	4.3716	4.5788	4.7694	5.0998	4.0040	4.3735	4.7228	4.9306	5.0928
0.85	**constant**	4.0196	4.3787	4.6777	4.8515	5.1442	4.0738	4.4015	4.7866	5.1428	5.1496
0.90	**constant**	4.1205	4.4545	4.7664	4.9432	5.2354	4.2131	4.5773	4.8332	5.2922	5.3080