INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY

CAMPUS MONTERREY

ESCUELA DE GRADUADOS EN ADMINISTRACIÓN
PÚBLICA Y POLÍTICA PÚBLICA

TECNOLÓGICO
DE MONTERREY®

"QUALITY, PRODUCTIVITY AND DISCRIMINATION EFFECTS IN MEXICAN LABOUR REMUNERATIONS"

TESINA

PRESENTADA COMO REQUISITO PARCIAL PARA OBTENER EL GRADO ACADEMICO DE:

MAESTRO EN ECONOMÍA Y POLÍTICA PÚBLICA
MEK 2009

POR:

ELIUD DIAZ ROMO

MONTERREY, N.L. DICIEMBRE DE 2010

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY

CAMPUS MONTERREY

ESCUELA DE GRADUADOS EN ADMINISTRACION
PUBLICA Y POLITICA PUBLICA

Los miembros del comité de tesina recomendamos que el presente proyecto de tesina presentado por el Lic. Eliud Díaz Romo sea aceptado como requisito parcial para obtener el grado académico de:

Maestro en Economía y Política Pública

Comité de Tesina:

Dr. Raymundo C. Rodríguez Guajardo
Asesor

Dr. José de Jesús Salazar Cantú
Sinodal

Dr. Lester B. García Olvera
Sinodal

Dedication

This thesis is dedicated to my parents for their endless love, support and encouragement

Acknowledgments

First and foremost, I would like to thank my supervisor, Prof. Raymundo C. Rodriguez Guajardo, for his guidance and support, and especially for his confidence in me.

I would also like to express my gratitude to Prof. Jose de Jesus Salazar Cantu for his insightful comments. They were more helpful than he may think.

Finally, I would like to thank Odette Melvin for her assistance in editing the English version of this paper.

Quality, Productivity and Discrimination Effects in Mexican Labour Remunerations

By Eliud Diaz Romo

Abstract

This investigation analyzes the definition and measurement of worker quality, productivity, and discrimination in labour market remunerations. The methodology proposed in this paper considers Mincerian equations for human capital accumulation (Mincer, 1970) and Ronald Oaxaca (1973) decomposition of wage differentials to formally derive the productivity and discrimination effects of labour remuneration differentials. Based on this specification, increases in remunerations depend on changes in the quality of human capital, discrimination effects and productivity differentials. Econometric regressions are estimated for Mexico using data from the National Surveys of Occupation and Employment (Encuesta Nacionales de Ocupación y Empleo - ENOE) of 2005, 2007 and 2009. Results indicate that there are remuneration differentials across time that do not necessarily depend on improvements in the quality of human capital and are better explained by productivity shifts. Additionally, it is statistically proven that gender discrimination effects are significant towards explaining wage differentials and do not lose magnitude across time.

*Keywords: discrimination, human capital, productivity, quality, wage differentials.

Efectos de la Calidad, Productividad y Discriminación en las Remuneraciones Salariales de México

Por Eliud Diaz Romo

Resumen

Esta investigación analiza la definición y medición de la calidad del trabajador, productividad y discriminación en las remuneraciones del mercado laboral. Considerando las ecuaciones Mincerianas de capital humano (Mincer, 1970) y el método de descomposición de las diferencias salariales de Ronald Oaxaca (1973) se derivan formalmente los efectos de productividad y discriminación en las diferencias salariales. La metodología propuesta supone que los aumentos salariales dependen de cambios en la calidad del capital humano, efectos de discriminación y diferencias de productividad. Las especificaciones econométricas formuladas son estimadas para México usando información de las Encuestas Nacionales de Ocupación y Empleo (ENOE) 2005, 2007 y 2009. Los resultados sugieren que en México existen aumentos salariales que no dependen necesariamente de diferencias en la calidad de los trabajadores y son mejor explicados por aumentos de productividad. Por otro lado, se demuestra estadísticamente que en las diferencias salariales existen efectos derivados de discriminación por género y que estos efectos no prierden magnitud a medida que pasan los años.

*Palabras clave: calidad, capital humano, diferencias salariales, discriminación, productividad

Table of contents

1. Introduction

The measurement of labour quality, productivity or discrimination and their effects on labour remunerations are difficult to estimate. Few methodologies have been proposed that try to integrate labour quality, productivity and discrimination concepts and even fewer empirical works have been conducted. The next section of the document endeavours to review the principal literature analyzing quality (Nicholson, 1967 and; Hicks and Johnson, 1968), essential papers studying human capital (Mincer, 1970, 1974) and other investigations of decomposing methods (Oaxaca, 1973; Blinder, 1973 and; Nicole Fortin, Thomas Lemieux and Sergio Firpo, 2010)

Various limitations to measuring quality include the interpretation of available data, the lack of measurement variables, the existence of few formal studies and the complexity of generalizing findings. It is difficult to avoid subjectivity in definitions and variable specifications. The acceptance of a definition of quality in academics, management or politics depends on opinions, given that it is difficult to test that one definition is correct. In this investigation, worker quality will be defined in terms of human capital. A more qualified worker will have abilities derived from years of education, experience and other variables that improve a worker's capacities or abilities.

Improvements in human capital lead to increases in productivity. Productivity is considered to be the relationship between output of production and the resources used to generate it. More specifically, labour productivity will be the amount of goods and services produced per worker hour. It is generally accepted that a more qualified worker will produce more products and services than a less qualified one in a given time frame.

Real wage increases can be explained in terms of better employee performance (quality and productivity), however wage increases can differ among men, women, pregnant women, indigenous persons, etc. with similar characteristics and equivalent performance because of discrimination. Not many authors have tried to combine the concept of labour quality with productivity and discrimination effects in workers' earnings. Ronald Oaxaca (1973) developed a decomposition methodology to separate discrimination and human capital components of workers' remunerations using Mincerian specifications for human capital (Mincer, 1970). This decomposition procedure has been widely accepted by academics and it has been used to analyze wage differences across countries, industries and groups.

Based on Oaxaca (1973) decomposition technique, a theoretical and empirical specification is formulated in this paper that decomposes labour remuneration increases into three components: worker quality in terms of human capital, productivity, and discrimination. The main purpose of the investigation is to offer a time series analysis of the evolution of productivity, quality of human capital and discrimination in labour remunerations.

The decomposition of wage differentials will be useful in creating governmental and private policies that encourage increases in human capital, provide incentives to boost productivity and avoid discrimination differentials. Particularly in Mexico this methodology could be beneficial for the Department of Labour and Social Forecasting (Secretaría de Trabajo y Previsión Social - STPS), the state department with the responsibility of supervising labour market outlook, promoting productivity in labour, encouraging training to improve labour productivity and carrying out investigations to develop strategies to increase employment in Mexico.

Quarterly data from Mexico, taken from National Survey's of Ocupation and Employment (Encuesta Nacional de Ocupación y Empleo -ENOE)[1], is analyzed. Different estimations compare labour information from the same quarter of different years, adjusted to 2009 prices. To prevent periodical abnormal differentials, the periods analyzed are the third quarter of 2005, the third quarter of 2007 and the same quarter of 2009.

Estimated decompositions resulting from econometric regressions suggest that improvements in salaries across time can be explain in term of increases in quality of human capital and increases of minimum wage, as the intercept represents remuneration paid to an individual without any education or experience. In terms of discrimination it is found that these effects are significant and statistically constant over time.

The second section of this document reviews relevant investigations that try to incorporate quality into different economic models, other papers that explain productivity in human capital and selected investigations that study the impacts of discrimination in wage

[1] ENOE is collected since 2005 for National Institute of Statistics and Geography (Instituto Nacional de Estadística y Geografía - INEGI) accounting with nationally representative information. ENOE 2005 has information for 421, 715 individuals, whereas ENOE 2007 and 2009 have information for 418, 327 and 402, 919, respectively.

differentials. In the third section, economic theory, labour market concepts and discrimination effects are connected obtaining the quality, productivity and discrimination components in a human capital model. The fourth part covers methodology aspects and description of the proposed human capital specification that includes impacts of discrimination and productivity in labour remunerations. In the fifth section, the proposed equation is used to estimate a numerical example for Mexico using the Ordinary Least Squares estimation method, correcting for heteroskedasticity and verifying the inexistence of multicollinearity. Results, analysis, conclusions and limitations are presented in the last part of the document.

2. Literature review

What is quality? Quality does not have a specific definition unless it is related to a determined area of study. In business, quality is a perceptual variable interpreted as the superiority of something. Different people in diverse places will attribute different levels of quality to the same good. In economics, quality is the perception of the level to which a product satisfies a customer's expected benefits or utility.

Originally in economic science economists did not pay attention to quality measurement due to the fact that subjective perceptions are frequently present in quality analysis. It was not until the second part of the twentieth century that economists started explaining the concept of quality in economics. Nicholson (1967) formalized the problem; he argued that changes in quality and the introduction of new commodities were possibly the most difficult problem faced by compilers of index numbers. Well-accepted solutions were not available and different views had been expressed until Nicholson decided to determine the first fundamental principles to tackle the problem. According to Nicholson the principles used to study quality are the following:

1 Measurement of quality depends on new substitutes (commodities) and the variation of production quality through time, for example the season of the year in which the product is produced

differentials. In the third section, economic theory, labour market concepts and discrimination effects are connected obtaining the quality, productivity and discrimination components in a human capital model. The fourth part covers methodology aspects and description of the proposed human capital specification that includes impacts of discrimination and productivity in labour remunerations. In the fifth section, the proposed equation is used to estimate a numerical example for Mexico using the Ordinary Least Squares estimation method, correcting for heteroskedasticity and verifying the inexistence of multicollinearity. Results, analysis, conclusions and limitations are presented in the last part of the document.

2. Literature review

What is quality? Quality does not have a specific definition unless it is related to a determined area of study. In business, quality is a perceptual variable interpreted as the superiority of something. Different people in diverse places will attribute different levels of quality to the same good. In economics, quality is the perception of the level to which a product satisfies a customer's expected benefits or utility.

Originally in economic science economists did not pay attention to quality measurement due to the fact that subjective perceptions are frequently present in quality analysis. It was not until the second part of the twentieth century that economists started explaining the concept of quality in economics. Nicholson (1967) formalized the problem; he argued that changes in quality and the introduction of new commodities were possibly the most difficult problem faced by compilers of index numbers. Well-accepted solutions were not available and different views had been expressed until Nicholson decided to determine the first fundamental principles to tackle the problem. According to Nicholson the principles used to study quality are the following:

1 Measurement of quality depends on new substitutes (commodities) and the variation of production quality through time, for example the season of the year in which the product is produced

2 The market defines prices but quality depends on the expected marginal utility and quantity supplied.

3 It is assumed that quality and prices are positively related.

4 An individual is willing to accept and can differentiate the grade of quality offered and the cost of a productive factor.

5 An individual has notion between the value of the goods and the period that he or she is interested in. The individual may identify the value of a productivity factor according to its capabilities and performance.

6 There is a range of prices at which some commodities are preferred over others with a slightly upper level of price according to the quality offered. Some labour quality factors are preferred more than others, even though they demand a higher rate of wage, which does not compensate his quality.

7 Productivity factors have no transitivity in a specific period of time. Commodities do not have the same attributes nor do they provide the same marginal utility at a given time, so it is difficult to compare the marginal contributions (utility).

8 Temporary, personal, or seasonal (mid-term) interests ("fashions") affect the measurement of quality.

9 There is a tendency on demand. The more budget income an individual has, the more they will be willing to pay for a higher quality product.

10 Another factor that affects a person's demand is the liquid currency and therefore the only product to buy is short term quality (poor quality products).

11 Seasonal characteristics and attributes may increase the quality.[2]

These principles can be used to analyze the labour market, where goods will be hours of work (employment) and price will be the wage rate per hour of work. In this manner, measurement of quality will depend on the presence of new employees, variation of workers quality through time, expected productivity, the assumption that quality and wages are positively related, temporary cultural and economic tendencies and characteristics or attributes of workers.

---

[2] Nicholson (1967) principle 12 is not included. This principle reviews empirical, not theoretical, examples.

The problem of quality estimation persisted. One of the first formal specifications of quality and quantity components of income elasticity was proposed by Hicks and Johnson (1968). They identified that increases in food expenditure can be devoted to increasing the quantity of food consumed and increasing the quality of the diet. Knowing that expenditures for food in high per capita income countries were excessively larger than expenditures in low per capita income countries the authors present a simplified model to determine the income elasticities in terms of quantity and quality components.

Development of theoretical models boomed. Aggregating a person's preferences is one of the most contested topics in determining quality effects. Cox and Wohlgenant (1986) estimated aggregated cross-sectional demand functions to obtain quality effects on supply related to variation in prices. They found that differences in parameters resulting from the failure to adjust cross-sectional prices for quality changes are small in disaggregated, homogeneous commodities. Nelson (1991) proposes alternative measures of demand when goods are heterogeneous, derived from restrictions on quality variation, consumer preferences and relative prices. Most recently Babcock and Carriquiry (2005) developed a micro-foundations model of a firm that needs to manage the quality of a heterogeneous good in the presence of different consumer tastes and quality expectations.

Following Nicholson (1967), according to principle 4, employers tend to differentiate workers based on productivity factors. These productivity factors must be defined if quality is to be estimated. In labour economics, worker quality is explained in terms of human capital. Mincer (1970) proposed a human capital model explaining earnings in terms of education, experience and other worker characteristic variables. Four year later Mincer (1974) used data from the 1950 and 1960 USA Censuses to estimate wage changes in terms of variations in education and training among workers. Specifications developed by Mincer are now called Mincerian equations and are used in labour economics for estimating econometric human capital models.

Hanushek, Eric A. and Dongwook Kim (1995) recognize the importance of human capital for Growth Economics but they maintain that empirical investigations do not clarify the size or magnitude of human capital, nor do they policy implications. They develop a direct measure of quality in human capital measured by cognitive skills in mathematics and science obtained from international test scores across countries. The estimated growth effects of increases in labour force quality are significant for the precise specification of the regressions. The significance of quality presents a dilemma in policy making, because simple investments to improve cognitive skills seem ineffective. Five years later Eric A. Hanushek and Dennis Kimko (2000) considered two ways to measure quality: direct measures of cognitive skills (like Hanushek in 2005) and measures of schooling inputs (books, teachers, technological resources). They carry out different estimations trying to develop and use a consistent set of cognitive test measures of quality. However different estimates of micro-productivity's effect introduce uncertainty about the magnitude of quality's impact on economic growth.

The measurement problem continued. The concept of quality has a practical interpretation as the usefulness or superiority of something. Quality is a perceptual and to some extent a subjective concept that may be understood differently by different individuals. Firms may focus on a worker's quality by looking for better productivity, by ensuring goods are produced correctly or by innovating and differentiating their products. Productivity is defined as a relationship between the output of production and the resources used to obtain it. Productivity is an indicator of efficiency that relates quantity of input used in production process and the output produced. Labour productivity is generally measured as output production in a given amount of time.

The most important challenge of this paper is the integration of quality, productivity and discrimination effects into labour capital theory. Giannias (1998) presents a theory for labour quality that includes a methodical process to assure that wage differentials can be explain in terms of variations in labour quality. Dimitrios Giannias derives a demand for labour quality obtaining and calculating the equilibrium wage equation. His method yields consistent estimates of the equilibrium demand for labour quality equation. The author offers an application on the market for seminars, exploring wage and quality characteristics of trainers in Nova Scotia. Finally Giannias presents equations for the labour quality index and wage equilibrium.

What about discrimination differentials in labour remunerations? An explanation for wage differentials must consider changes in three components: productivity, labour capital (quality) and discrimination. A diverse literature addresses how one can decompose wage differentials into variations into productive endowments and discrimination. A common econometric procedure for measuring discrimination effects on wage differentials was developed by Ronald Oaxaca (1973). Basically he indicates two different groups, an advantaged group and a disadvantaged one, and supposes that, in the absence of discrimination, the estimated effects of labour capital variables on remunerations are identical for each group. Variations in intercept and any other coefficients are explained by differentials in labour capital variables and discrimination.

Blinder (1973) provides analysis of wage differentials between white and black men and between white males and females using regressions of white males, white females and black males. The author describes the same decomposition method as that used by Oaxaca (1973) to analyze wage differentials. When the estimations are combined he finds that the major component of wage differentials are attributable to diverse types of discrimination.

Nicole Fortin, Thomas Lemieux and Sergio Firpo (2010) present an overview of decomposition methods developed since Oaxaca and Blinder. They focus on different distributional statistics besides the mean such as variance, the Gini coefficient or quantiles, as opposed to Oaxaca and Blinder, who only studied the mean. They discuss assumptions required for decomposing equations and other estimation methods proposed. Finally the authors explain decomposition in structural models and include extensions for dealing with self-selection, endogeneity and panel data.

Using Oaxaca (1973) decomposition methodology, the next section describes an aggregated specification that decomposes productivity, human quality and discrimination components in wage differentials.

3. Theoretical framework

Modern labour economics was developed with empirical foundations in human capital theory. The fundamental conceptual framework used in this investigation is based on Mincerian equations (Mincer 1970, 1974). Mincer presents an earnings function that summarizes the various categories of human capital investments as determinants of wages. These categories are analysed in life cycle intervals to capture the impact of human capital investments on paid salaries. The individual wage function is as follows:

econometric procedure for measuring discrimination effects on wage differentials was developed by Ronald Oaxaca (1973). Basically he indicates two different groups, an advantaged group and a disadvantaged one, and supposes that, in the absence of discrimination, the estimated effects of labour capital variables on remunerations are identical for each group. Variations in intercept and any other coefficients are explained by differentials in labour capital variables and discrimination.

Blinder (1973) provides analysis of wage differentials between white and black men and between white males and females using regressions of white males, white females and black males. The author describes the same decomposition method as that used by Oaxaca (1973) to analyze wage differentials. When the estimations are combined he finds that the major component of wage differentials are attributable to diverse types of discrimination.

Nicole Fortin, Thomas Lemieux and Sergio Firpo (2010) present an overview of decomposition methods developed since Oaxaca and Blinder. They focus on different distributional statistics besides the mean such as variance, the Gini coefficient or quantiles, as opposed to Oaxaca and Blinder, who only studied the mean. They discuss assumptions required for decomposing equations and other estimation methods proposed. Finally the authors explain decomposition in structural models and include extensions for dealing with self-selection, endogeneity and panel data.

Using Oaxaca (1973) decomposition methodology, the next section describes an aggregated specification that decomposes productivity, human quality and discrimination components in wage differentials.

3. Theoretical framework

Modern labour economics was developed with empirical foundations in human capital theory. The fundamental conceptual framework used in this investigation is based on Mincerian equations (Mincer 1970, 1974). Mincer presents an earnings function that summarizes the various categories of human capital investments as determinants of wages. These categories are analysed in life cycle intervals to capture the impact of human capital investments on paid salaries. The individual wage function is as follows:

$$w_t = w_{t-1} + rC_{t-1} \tag{1}$$

where w_t is the wage paid in period t, C_{t-1} is the amount of money paid for net investment in period t-1 and r is the rate of return on this particular investment. For simplicity Mincer assumes that r is the same for each period.

$$w_t = w_0 + r\sum_{j=0}^{t-1} C_j \tag{2}$$

where w_0 is the initial earning capacity, or the base salary an individual would receive who has not invested in education, experience or any other ability.

Individual investments are not easily observed, except for schooling or training programs. Even for this case, data of the money spent is not as readily available as data of years of educational attainment. For this reason it is preferred to express explanatory variables in terms of time rather than in money spent. To better understand this, the ratio of investment (K_t) expenditure is shown:

$$K_t = \frac{C_t}{w_t} \tag{3}$$

K_t is the proportion of wages spent on investment in a given period t. If the cost of investment represents only time costs, then K_t is the fraction of the period spent on investment activities.

Expressing equation (3) for period t-1 and substituting in equation (1):

$$C_{t-1} = K_{t-1}w_{t-1}$$

$$w_t = w_{t-1} + rK_{t-1}w_{t-1} = w_{t-1}(1 + rK_{t-1})$$

and for two years of school:

$$w_t = w_{t-2}(1 + rK_{t-2})(1 + rK_{t-1})$$

subsequently by recursion starting in the period where no investments are made:

$$w_t = w_0(1 + rK_0)(1 + rK_1)\cdots(1 + rK_{t-1})$$

$$w_t = w_0 \prod_{j=0}^{t-1} (1 + rK_j) \quad (4)$$

If rK_j is a relatively small number, a logarithmic approximation is appropriate for simplifying the equation[3], and:

$$\ln w_t = \ln w_0 + r\sum_{j=0}^{t-1} K_j \quad (5)$$

Equation (5) can be extended to include other investments in human capital like pre-school care, medical care, health activities, experience, and so forth. Finally Mincer (1974) estimated two specifications that relate individual earnings to accumulated human capital at each point in a person's working life:

$$\ln w_t = \delta_0 + \delta_1 s + \delta_2 t + \delta_3 t^2 + v \quad (6)$$

$$\ln w_t = \delta_0 + \delta_1 s_t + \delta_2 t_t + \delta_3 t_t^2 + \delta_4 \ln H_t + u_t \quad (7)$$

where $\delta_0 = \ln w_0$, $\delta_1 = r$, s represents years of school, t is individual experience during a person's working life, t^2 is included to allow for diminishing marginal returns to experience[4] and H is Actual amount of working during the year. δ_0, δ_1, δ_2 and δ_4 are expected to be positive. The initial salary should be positive (δ_0) and more education or experience produces better qualified workers, δ_1 and δ_2 are anticipated to be positive. δ_3 is expected to be negative, in the sense that working productivity diminishes when an individual's working life is close to an end.

Described specifications that explain the natural logarithm of wages in terms of investments in human capital (education, experience, etc) are called Mincerian equations

---

[3] Assuming r is relatively small and that $K_j < 1$

$$w_t = w_0 \prod_{j=0}^{t-1} (1 + rK_j) \approx w_0 \ell^{\, r \sum_{j=0}^{t-1} K_j}$$

[4] Marginal impact of experience = $\partial \ln w_t / \partial t = \delta_2 + 2\delta_3 t$

and are commonly used in labour economics for analyzing human capital components of labour productivity. Mincerian linear specifications may be estimated using Ordinary Least Squares (OLS). The next part of the document describes Oaxaca's (1973) extensions of Mincerian specifications. Oaxaca decomposition will be used later in the investigation.

3.1 Decomposed theoretical specification for discrimination

Using Oaxaca's (1973) methodology it is possible to decompose wage differences into two components, discrimination and adjustment from the means. For the analysis presented in this investigation the adjustment from means component will be represented as the worker's quality effect. The simple matrix models for distinguishing discrimination are:

$$\ln w_t^* = X_t^* \alpha^* + \varepsilon_t^*$$

$$\ln w_{*t} = X_{t*} \alpha_* + \varepsilon_{t*}$$

where $\ln w_t^* = X_t^* \alpha^* + e_t^*$ is the model for the advantaged group and $\ln w_{*t} = X_{t*} \alpha_* + e_{t*}$ is the model for the disadvantaged group. $\ln w_t^*$ and $\ln w_{t*}$ are t rows and 1 column vectors that contain the natural logarithm of individual wages. X_t^* and X_{t*} are matrices for all human capital variables explaining earnings, with t rows and k columns; t represents the number of observations and k represents the number of variables. X_t^* and X_{t*} may include years of school, experience, marital status, etc. α^* and α_* are k rows by 1 column matrices that are the confined coefficients (marginal effects) of the dependent variables.

Due to Ordinary Least Squares (OLS) properties the fitted regression crosses through the means and the difference in means is represented by the following:

$$\overline{\ln w_t^*} - \overline{\ln w_{t*}} = \overline{X_t^*} a^* - \overline{X_{t*}} a_*$$

The difference in coefficients is: $a^* - a_* = \Delta a$. Alternative decompositions are:

$$\overline{\ln w_t^*} - \overline{\ln w_{t*}} = \overline{X_t^*} (\Delta a + a_*) + \overline{X_{t*}} a_*$$

$$\overline{\ln w_t^*} - \overline{\ln w_{t*}} = \overline{X_t^*} \Delta a + \left(\overline{X_t^*} - \overline{X_{t*}} \right) a_* \qquad (8)$$

or

$$\overline{\ln w_t^*} - \overline{\ln w_{*t}} = \overline{X_t^*} a^* - X_{*t} \left(a^* - \Delta a \right)$$

$$\overline{\ln w_t^*} - \overline{\ln w_{*t}} = \overline{X_{*t}} \Delta a + \left(\overline{X_t^*} - \overline{X_{*t}} \right) a^* \qquad (8')$$

where wage difference is explained in terms of discrimination effects and average differences, in that order. This decomposition method is analyzed in the following pages to expand the analysis of wage differences.

4. Methodological framework

In this investigation, worker quality will be defined in terms of human capital. A better qualified worker will have abilities derived from the number or years of education, experience and other variables that improve the worker's capacities ore abilities. There is not sufficient information on school quality, trainings received, acquired capability tools or any other form of human capital investment for all individuals. It will be assumed that school quality is the same for all individuals and quality will depend solely on years of attainment. Oaxaca (1973) methodology can be used to observe wage differences over time.

4.1 Time decomposition

Using the same methodology (Oaxaca 1973) for differences in wages across time, it is possible to decompose two components, productivity and adjustment from means. The decomposition starts with the matrix equations that explain wages in the present period and the final time period

$$\ln w_t = X_t \beta^* + e_t^*$$

$$\ln w_{t-1} = X_{t-1} \beta_* + e_{*t}$$

For OLS properties the fitted regression crosses through the means.

$$\overline{\ln w_t} = \overline{X}_t b^*$$

$$\overline{\ln w_{t-1}} = \overline{X}_{t-1} b_*$$

The difference in means in one period of time is:

$$\overline{\ln w_t} - \overline{\ln w_{t-1}} = \overline{X}_t b^* - \overline{X}_{t-1} b_*$$

The difference in coefficient vectors across time is: $b^* - b_* = \Delta b$. This expression gives the possibility of two alternative decompositions:

$$b^* = \Delta b - b_*$$

where wage difference is explained in terms of discrimination effects and average differences, in that order. This decomposition method is analyzed in the following pages to expand the analysis of wage differences.

4. Methodological framework

In this investigation, worker quality will be defined in terms of human capital. A better qualified worker will have abilities derived from the number or years of education, experience and other variables that improve the worker's capacities ore abilities. There is not sufficient information on school quality, trainings received, acquired capability tools or any other form of human capital investment for all individuals. It will be assumed that school quality is the same for all individuals and quality will depend solely on years of attainment. Oaxaca (1973) methodology can be used to observe wage differences over time.

4.1 Time decomposition

Using the same methodology (Oaxaca 1973) for differences in wages across time, it is possible to decompose two components, productivity and adjustment from means. The decomposition starts with the matrix equations that explain wages in the present period and the final time period

$$\ln w_t = X_t \beta^* + e_t^*$$

$$\ln w_{t-1} = X_{t-1} \beta_* + e_{*t}$$

For OLS properties the fitted regression crosses through the means.

$$\overline{\ln w_t} = \overline{X}_t b^*$$

$$\overline{\ln w_{t-1}} = \overline{X}_{t-1} b_*$$

The difference in means in one period of time is:

$$\overline{\ln w_t} - \overline{\ln w_{t-1}} = \overline{X}_t b^* - \overline{X}_{t-1} b_*$$

The difference in coefficient vectors across time is: $b^* - b_* = \Delta b$. This expression gives the possibility of two alternative decompositions:

$$b^* = \Delta b - b_*$$

substituting

$$\overline{\ln w_t} - \overline{\ln w_{t-1}} = \overline{X_t}(\Delta b + b_*) - \overline{X_{t-1}} b_*$$

$$\overline{\ln w_t} - \overline{\ln w_{t-1}} = \overline{X_t}\Delta b + (\overline{X_t} - \overline{X_{t-1}}) b_* \qquad (9)$$

or

$$b_* = b^* - \Delta b$$

substituting

$$\overline{\ln w_t} - \overline{\ln w_{t-1}} = \overline{X_t} b^* - \overline{X_{t-1}}(b^* - \Delta b)$$

$$\overline{\ln w_t} - \overline{\ln w_{t-1}} = \overline{X_{t-1}}\Delta b + (\overline{X_t} - \overline{X_{t-1}}) b^* \qquad (9')$$

By selecting comparable groups considering geographical distribution, industry sectors, discrimination and human capital variables, the difference in log wages is decomposed into two components. In (2) and (2') the first element represents the difference in log salaries due to the difference in estimated coefficients; Oaxaca (1973) captured the discrimination effect with this component. Differentiating salaries over time yields the productivity effect. It symbolizes the worker's capability to obtain better wages independently of improvements in human capital. The second element represents differences in log wages due to differences in average human capital variables and represents the effect of changes in worker quality.

The quality and productivity components obtained could be biased because they could be missing the evolution of discrimination effects over time. Comparison and integration of wage discrimination and productivity evolutions is needed to justify the statistical and empirical relevance of the decomposed quality, quantity and discrimination components.

4.2 Quality, productivity and discrimination in wage differentials

Integrating productivity and discrimination components of wage differentials requires analyzing wage differentials between the advantaged group in period *t* and disadvantaged group in period *t-1*. To examine this, four basic specifications are used, assuming that OLS regressions cross through the means:

1. $\overline{\ln w_t^*} = \overline{X_t^*}\alpha_1$

2. $\overline{\ln w_{t-1}^*} = \overline{X_{t-1}^*}\alpha_2$

3. $\overline{\ln w_{t*}} = \overline{X_{t*}}\alpha_3$

4. $\overline{\ln w_{t-1*}} = \overline{X_{t-1*}}\alpha_4$

where 1 corresponds to the human capital model for the advantaged group in period *t*, 2 corresponds to the advantaged group in period *t-1*, 3 symbolizes the disadvantaged group in period *t* and 4 represents the disadvantaged group in period *t-1*

There are eight possible combinations that decompose discrimination, quality and productivity components. Two possible decompositions are explained here. Using 1 and 2, productivity and quality components are obtained.

$$\overline{\ln w_t^*} - \overline{\ln w_{t-1}^*} = \overline{X_t^*}\alpha_1 - \overline{X_{t-1}^*}\alpha_2$$

$$\Delta\alpha_{12} = \alpha_1 - \alpha_2$$

$$\overline{\ln w_t^*} - \overline{\ln w_{t-1}^*} = \overline{X_t^*}\Delta\alpha_{12} + \left(\overline{X_t^*} - \overline{X_{t-1}^*}\right)\alpha_2 \qquad (10)$$

or

$$\overline{\ln w_t^*} - \overline{\ln w_{t-1}^*} = \overline{X_{t-1}^*}\Delta\alpha_{12} + \left(\overline{X_t^*} - \overline{X_{t-1}^*}\right)\alpha_1 \qquad (10')$$

where the first component ($\overline{X_t^*}\Delta\alpha_{12}$ or $\overline{X_{t-1}^*}\Delta\alpha_{12}$) will be the productivity component, which is the wage differential explained by increases of an advantaged worker's capabilities in one period of time. The second component ($\left(\overline{X_t^*} - \overline{X_{t-1}^*}\right)\alpha_2$ or $\left(\overline{X_t^*} - \overline{X_{t-1}^*}\right)\alpha_1$) will be the advantaged worker quality differential explained by differences in the arithmetic mean of the independent variables.

Using 2 and 4 a the second specification is constructed, decomposing the wage differentials of the advantaged and disadvantaged groups in period *t-1*.

$$\overline{\ln w_{t-1}^*} - \overline{\ln w_{t-1*}} = \overline{X_{t-1}^*}\alpha_2 - \overline{X_{t-1*}}\alpha_4$$

$$\Delta\alpha_{24} = \alpha_2 - \alpha_4$$

$$\overline{\ln w_{t-1}^*} - \overline{\ln w_{t-1*}} = \overline{X_{t-1}^*}\Delta\alpha_{24} + \left(\overline{X_{t-1}^*} - \overline{X_{t-1*}}\right)\alpha_4 \qquad (11)$$

or

$$\overline{\ln w_{t-1}^*} - \overline{\ln w_{t-1*}} = \overline{\overline{X_{t-1*}}}\Delta\alpha_{24} + \left(\overline{X_{t-1}^*} - \overline{X_{t-1*}}\right)\alpha_2 \qquad (11')$$

where the first elements ($\overline{X^*_{t-1}}\Delta\alpha_{24}$ and $\overline{X_{t-1*}}\Delta\alpha_{24}$) correspond to the discrimination impact in wage differentials and the second elements ($\left(\overline{X^*_{t-1}} - \overline{X_{t-1*}}\right)\alpha_4 , \left(\overline{X^*_{t-1}} - \overline{X_{t-1*}}\right)\alpha$) correspond to the difference from means of human capital variables.

Now an integrated model with quality, quantity and discrimination components is obtained by adding (10)-(11), (10)-(11'), (10')-(11) and (10')-(11'). Examining the first combination:

(10) + (11)

$$\left(\overline{\ln w^*_t} - \overline{\ln w^*_{t-1}}\right) + \left(\overline{\ln w^*_{t-1}} - \overline{\ln w_{t-1*}}\right) = \overline{X^*_t}\Delta\alpha_{12} + \left(\overline{X^*_t} - \overline{X^*_{t-1}}\right)\alpha_2 + \overline{X^*_{t-1}}\Delta\alpha_{24} + \left(\overline{X^*_{t-1}} - \overline{X_{t-1*}}\right)\alpha_4$$

$$\overline{\ln w^*_t} - \overline{\ln w_{t-1*}} = \overline{X^*_t}\Delta\alpha_{12} + \left(\overline{X^*_t} - \overline{X^*_{t-1}}\right)\alpha_2 + \left(\overline{X^*_{t-1}} - \overline{X_{t-1*}}\right)\alpha_4 + \overline{X^*_{t-1}}\Delta\alpha_{24} \quad (12)$$

(10') + (11')

$$\overline{\ln w^*_t} - \overline{\ln w^*_{t-1}} + \overline{\ln w^*_{t-1}} - \overline{\ln w_{t-1*}} = \overline{X^*_{t-1}}\Delta\alpha_{12} + \left(\overline{X^*_t} - \overline{X^*_{t-1}}\right)\alpha_1 + \overline{X_{t-1*}}\Delta\alpha_{24} + \left(\overline{X^*_{t-1}} - \overline{X_{t-1*}}\right)\alpha_2$$

$$\overline{\ln w^*_t} - \overline{\ln w_{t-1*}} = \overline{X^*_{t-1}}\Delta\alpha_{12} + \left(\overline{X^*_t} - \overline{X^*_{t-1}}\right)\alpha_1 + \left(\overline{X^*_{t-1}} - \overline{X_{t-1*}}\right)\alpha_2 + \overline{X_{t-1*}}\Delta\alpha_{24} \quad (12')$$

where wage differentials of advantaged and disadvantaged groups from one period to the next one can be interpreted as the addition of productivity increases of the advantaged group, differences in means of the advantaged group in two periods, differences from means of the advantaged and disadvantaged groups in *t-1* and the discrimination component in period *t-1*.

The same process could be repeated using 1 with 3 and then 3 with 4.[5] This process will produce another four different specifications that decompose wage differentials into productivity, quality and discrimination components. In this investigation it is assumed that all specifications result in the same decomposition results. Nevertheless, two alternative specifications are estimated, (12) and (12'), to contrast regressions results. The next section will estimate proposed equation for the Mexican labour market.

[5] The decomposition process is repeated using 1 with 3 and then 3 with 4 in Appendix A1

5. Estimated specifications for México

The proposed decomposition for productivity, quality and discrimination effects in wage differentials represents a useful tool for analyzing labour market behavior in Mexico, for supporting labour policies intended to encourage human capital development and for trying to avoid any kind of differentials explained by discrimination. The Department of Labour and Welfare (Secretaría de Trabajo y Previsión Social)

5.1 Description of information

The described models will be estimated for Mexico using the "Encuesta Nacional de Ocupación y Empleo" (ENOE) for three periods considered to be economically stable and two years apart, starting with the third quarter of 2005, followed by the same quarter of 2007 and finishing with the third quarter of 2009.

Before estimating the models, it is necessary to select comparable groups for analysis. ENOE offers representative information for Mexico, although each year the different individuals are surveyed. In this paper it is assumed that observations from ENOE are comparable across years, under this hypothesis estimated Mincerian econometric regressions are comparable across time.

5.2 Empirical analysis

Basic Mincerian equations to be estimated for Mexico are specified as follows using Mincer's (1974) conceptual framework:

$$\ln(w_{it}) = \alpha + \beta_1 S_{it} + \beta_2 S_{it}^2 + \beta_3 E_{it} + \beta_4 E_{it}{}^2 + u_{it} \qquad (13)$$

The next table shows variable definitions.

Table 1. Variable definitions

Variable	Description
w_i	Real Mexican pesos paid per hour of work to individual *i* in period *t,* based on 2009 prices
E_i	Experience. Calculated as the age of individual *i* minus years of study minus 6, which is assumed to be the average of age for starting elementary school. It is expected to have a positive relationship between experience and perceived wage.
E_i^2	Squared Experience. It is expected to have a negative relationship with the dependent variable as a result of diminishing marginal returns to experience.[6]
S_i	Education. Total years of school of individual *i*. It is expected to have a positive relationship with the dependent variable.
S_i^2	Squared education of individual *i*. The coefficient of this explanatory variable is expected to be negative for decreasing marginal returns of education.[7]

The worker's experience variable is not available in ENOE. Nonetheless, this variable is needed for estimating equation (13). Following Mincer (1970) and Oaxaca (1973), in this investigation the experience variable is created by taking the age of each individual minus his or her years of education minus 6. This definition of the experience variable requires three important assumptions. First, all individuals begin education (not including preschool education or/and nursery school) at 6 years old which is at present the minimum age required in Mexico to be accepted in elementary school. Second, years of school and a worker's experience are strictly independent, any individual studying cannot work and vice

[6] Marginal Impact of Experience in ln(wage): $\partial \ln w_{it} / \partial E_{it} = \beta_3 + 2\beta_4 E_{it}$

[7] Marginal Impact of School in ln(wage): $\partial \ln w_{it} / \partial S_{it} = \beta_1 + 2\beta_2 S_{it}$

versa. Finally, activities besides studying and working and not considered, thus leisure is not present.

Table 1.1 presents the means of the variables used in the equation (13); paid wage per hour, years of work experience and years of schooling completed. The wage and human capital investments (school and experience) gain value across time for all groups. The constant increment of the wage paid per hour of work could be explained by increases in years of experience and years of education.

In all periods men are better paid than women. Men have better work payments, although this difference seems to diminish over time. In 2005 the difference in male-female wage per hour is of $1.85, in 2007 it diminishes to $1.4 and in 2009 it is $1.29. In all the periods women have more years of education than men and men have more experience than women. Using this information, the male female wage differentials could be explained statistically if years of experience are better paid than years of education, otherwise the wage differentials would be explained by gender discrimination.

Table 1.1 Mean of variables

Variables	2005			2007			2009		
	All	Men	Women	All	Men	Women	All	Men	Women
Wage per hour	17.78	18.48	16.63	20.49	21.03	19.63	20.93	21.43	20.14
Years of experience	21.93	22.35	21.20	21.97	22.43	21.20	22.42	22.70	21.96
Yeas of school	8.97	8.80	9.26	9.24	9.02	9.59	9.25	9.04	9.59

In specification (13) variables for discrimination differentials (gender, marital status, race, etc) are not included and specification (13) is not decomposed to discern productivity or quality components. This equation must be estimated in different groups; advantaged and disadvantaged groups (men and women), in periods t and $t-1$ (2005, 2007 and 2009). All regressions are estimated using OLS, ensuring the estimated equations cross through the variables' means, to be consistent with Oaxaca decomposition methodology.

Table 2.1 shows the estimated regressions for all available data in each ENOE survey, and Table 2.2 gives the regression coefficients for men and women, excluding non-paid workers and atypical information[8]. For the estimated regressions only those individuals with paid work are considered, these are observations with wage per hour larger than zero.

In all regressions the p-value of the F statistic rejects the null hypothesis that all coefficients, except the intercept, are equal to zero. Adjusted R^2 represents the goodness of fit adjusted for the number of explanatory terms in a model and moves between 0 and 1. A larger R^2 means a better goodness of fit. All estimated regressions have a reasonably low R^2, limiting their capacity to predict salaries in the future. Nevertheless, all coefficients are significant, validating the regression's capacity to represent the relationship between the explanation variables and the natural logarithm of wages.

Table2.1 Estimated regressions of equation (13)[9]

Equation (13)	Coefficients		
Variable	2005	2007	2009
Constant	1.9405***	2.1247***	2.2207***
Experience	0.0306***	0.0286***	0.0268***
Experience^2	-0.0004***	-0.0004***	-0.0004***
Education	0.0232***	0.0217***	0.0109***
Education^2	0.0036***	0.0034***	0.0039***
Adjusted R^2	0.26	0.25	0.23
Observations	132180	132055	121115
Prob(F statistic)	0.000	0.000	0.000

*, **, *** Significant at a 90%, 95% and 99% confidence level, respectively

[8] Atypical information refers to variables suspicions of sampling errors, particularly individuals with more years of education than their age (633 observations, 271 in 2005, 186 in 2007 and 176 in 2009), or variables that does not meet fundamental assumptions, specifically individuals whit age smaller than their years of education plus 6 (1111 observations, 398 in 2005, 21 in 2007 and 292 in 2009).

Table2.2 Estimated regressions of equation (13) by gender

Equation (13)	Coefficients					
	Men			Women		
Variable	2005	2007	2009	2005	2007	2009
Constant	2.0339***	2.1903***	2.2912***	1.7527***	1.9839***	2.0588***
Experience	0.0308***	0.0288***	0.0273***	0.0309***	0.0290***	0.0269***
$Experience^2$	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0004***
Education	0.0207***	0.0225***	0.0113***	0.0282***	0.0218***	0.0132
$Education^2$	0.0035***	0.0032***	0.0036***	0.0040***	0.0039***	0.0043***
Adjusted R^2	0.25	0.24	0.22	0.29	0.27	0.26
Observations	83488	82298	74577	48692	49757	46538
Prob(F)	0.000	0.000	0.000	0.000	0.000	0.000

*, **, *** Significant at a 90%, 95% and 99% confidence level, respectively

The problem of heteroskedasticity is present in all regressions[10]. Heteroskedasticity exists when the estimated variance is not constant for all observations. This indicates that the basic assumption of constant variance in OLS is not satisfied. In order to deal with this problem all coefficient standard errors are transformed to robust standard errors based on a list of explanatory variables of equation-level scores and a covariance matrix.[11] Robust standard errors offer confidence and permit tests of any coefficient hypothesis.

Based on the nature of the explanatory variables, and the way they were created, they are supposed to be independent. However, in order to confirm that there is no multicollinearity[12], for all variables the variance inflation factor (VIF) is calculated[13]. Low values of the VIF statistic in the estimated regression suggest that multicollinearity is not suspected.[14]

---

[9] See general statistics of regressions in Appendix C.

[10] The White test proves the existence of homoskedasticity. See theoretical framework in Appendix A2. The test results are reported for each regression in Appendix C.

[11] The robust standard errors consist of a transformation of variance-covariance matrix in terms of multiple variables for multiple-equation models. In all regressions presented, robust coefficient standard errors replace the original covariance matrix with heteroskedasticity problems. See Appendix A3 for explanation of the variance-covariance matrix with robust standard errors.

[12] In econometrics multicollinearity refers to the situation in which strong correlation exists between explanatory variables of a model. This condition disturbs basic assumptions of OLS that no relationship exists between explanatory variables.

[13] VIF quantifies the dimension of multicollinearity using auxiliary OLS regressions between explanatory variables. It provides an Index to measure multicollinearity. The bigger it is because collinearity. See Appendix A4 for theoretical details.

[14] See Appendix C where VIF statistics are below 17 for all coefficients

The calculated intercept is positive, as expected; this parameter represents the natural logarithm of the wage received if an individual does not have any experience or education. Women have smaller constant than men. These results expose the wage differentials explained by discrimination. Men would be better paid if the education and the work experience were not considered. The intercept becomes bigger through time, each period it is bigger than the last. These shifts could be consequence of productivity rises across time.

The sign and magnitude of the estimated intercept differs among investigations depending on the survey employed, the period analyzed and the unit in which wages are measured (annual data, USA dollars per hour, USA dollars per month, etc.). The most common coefficients are significant and positive (e.g. Oaxaca, 1973; Blinder, 1973 and; Mincer, 1974).

The estimated parameter accompanying experience is positive and represents the proportional changes in salaries due to changes in years of experience. In addition, squared years of experience affects wages negatively. This means that the positive returns to experience get smaller as experience increases; this process is explained by diminishing marginal returns to experience. For example in 2005 (Table 2.1), if an individual improve his/her years of work experience from 20 to 21 years, then the expected remunerations will increase 4.74% (3.06-2x0.04x21). There is not an observable difference between the work experience effects of men and women, but it is evident that the effect is diminishing through time. The payment for each year of work experience could be worse through time.

In the literature it is typical to find the signs of the coefficients of experience to be positive and squared experience to be negative. The parameters magnitude and sign in Tables 2.1 and 2.2 are comparable with Blinder (1973, 1974), that uses data from Michigan Survey Reasearch Center's "Panel Study of Income Dymamics", and Oaxaca (1973), that employs 1967 Survey of Economic Opportunity conduced for the Bureau of the Census. The same signs but bigger coefficients' absolute values are found for Mincer (1974) using data from the 1950 and 1960 USA Censuses

The coefficients accompanying education have the expected sign but, contrasting what is expected, the coefficients of squared education are positive. For these results it could be assumed increasing marginal returns to education. This means that additional years of education are always better paid. These effects could be present because average education of Mexican employees (represented in ENOE) is lower than 10 and biased to the left[15]. This means that the expected coefficient signs could be found if the Mexican average years of education were higher to reach a level.

Squared schooling is not commonly used as explanatory variable. In the papers reviewed previously in this investigation (Mincer, 1974; Oaxaca, 1973 and; Blinder, 1973) only Oaxaca (1973) uses schooling squared, and finds the same signs as in this paper, which are positive coefficients accompanying education and squared education. In other paper, Oaxaca, Regan and Burghardt (2007) estimated Mincerian equations including squared years of education using data from the National Longitudinal Survey of Youth 1979 (NLSY79[16]).. Contrary to estimated regressions of this paper, they found larger coefficients and decreasing marginal returns to education on the natural logarithm of wages. Obtaining the same results for Mexico should be possible with sequential information of the same individuals across time.

A human capital representation for Mexico is estimated for Villarreal and Mehta (2003) using data from National Survey of Homes' Income and Expenditure (Encuesta Nacional de Ingreso y Gasto de los Hogares – ENIGH) obtaining comparable estimations, in magnitude and signs, as those found in the regressions presented here. The authors found decreasing marginal returns to experience; this means a positive coefficient accompanies experience and a negative coefficient accompanies squared work experience. In the specification proposed for Villarreal and Mehta the impacts of education on wages are separated, attempting to obtain different diploma effects (kindergarden, elementary school, middle school, etc). Significant diploma effects are only observed upon graduation from primary school and returns to college are larger for young workers. These returns to

---

[15] Years of school's means below 10 and positive skewnesses, that represents the compilation of years of education bellow the mean, are presented in Appendix B.

[16] National Longitudinal Survey of Youth 1979 (NLSY79) is a United States nationally representative sample of 12,686 individuals who were 14 to 22 years in 1979 and are periodically interviewed until today

education are greater than those presented in table 2.1 and 2.2. Nevertheless, Villarreal and Mehta's specification has suspected multicollinearity problems[17]

Table 2.3 shows the maximum achievable wage for each estimated representation according to econometric regressions, assuming individuals begin school at 6 years old and the maximum years of education are 24[18]. The maximization of wages comes when work experience is around of 35 years and the individuals obtain a PhD degree. In the maximum achievable wage the salary paid per hour is better for women than men, although women are older in that point. From regressions this is explained because girls are better paid for each year of education and the estimated experience return peaks earlier for men.

Work experience is better paid in the first years but after 4 years of school, or even less years depending on the estimated regression analyzed in Tables 2.1 and 2.2, the marginal returns received for each year of education are greater than the marginal returns received for experience. Keeping in mind that women have more education than men (Table 1.1) it is easy to conclude that male-female wage differentials favouring men are not explained by differentials in human capital investments. The difference in coefficients is statistically analyzed subsequently in this investigation, when decomposed discrimination, quality and productivity effects are analyzed.

---

[17] The proposed equation of Villarreal and Mehta (2003) is:

$$\ln(w_{it}) = \alpha_0 + \beta_Y Y + \alpha_P P + \beta_J P(Y-7) + \alpha_J J + \beta_H J(Y-10) + \alpha_H H + \beta_C H(Y-13) + \alpha_C C + \beta_G C(Y-17) + \varepsilon_1 E + \varepsilon_2 E^2$$

were Y is the number of years of education completed, E is the number of years or work experience, and P, J, H and C are indicator variables that take a value of 1 if an individual has completed primary school, junior-high, high school and college respectively, zero otherwise. The multicollinearity problem is present in dichotomous variables (P, J, H and C) where the effect of a diploma does not exclude effects of diplomas from before. For example, if an individual has completed college then C=1, but also P, J and H are equal to 1, the coefficient beside C will not represent the effect of accumulated education until college because it will be biased by the effects of P, J and H. In an appropriate specification, when C=1 then P, J and H should equal zero.

[18] In ENOE of 2005, 2007 and 2009 the maximum years of study are 24 (see Appendix B). In this paper, according to Mexican school programs, it will be assumed 6 years of primary school, 3 year of secondary school, 3 years of High School, 5

Table 2.3 Estimated maximum wage

	Year	Years of Experience	Yeas of school	Years Old	Maximum Achievable Wage ($/h)
All	2005	35.5	24	65.5	170.4
	2007	34.7	24	64.7	168.2
	2009	34.8	24	64.8	175.1
Men	2005	34.6	24	64.6	161.0
	2007	33.9	24	63.9	158.3
	2009	33.6	24	63.6	161.0
Women	2005	37.4	24	67.4	197.9
	2007	36.4	24	66.4	194.4
	2009	37.8	24	67.8	209.8

5.3 Productivity effects

This part of the document proves statistically whether wage differentials across time are attributable to differences in the quality of human capital and differences in the returns to schooling and/or experience. Table 3.1 shows that 13.6% of the increases in wages from 2005 to 2007 are explained by more years of study and that the other 86.4% can be attached to individuals' improvements in productivity. Although wage shifts were lower from 2007 to 2009, 14.3% of wage differentials are attributable to increases in experience among Mexicans, augmenting the quality of human capital. The payments for years of education and years of experience diminish across time; the shifts attributable to productivity are explained by constant differences, this means that minimum wage increases across time.

years of professional School, 2 years of master degree and 5 years of PhD making a total of 24 years of possible school if an individual does not repeat years of school or degrees of education.

Table 3.1 Productivity effects[19]

Equation (9)				
$\overline{\ln w_t} - \overline{\ln w_{t-1}}$	Variable	Difference for Quality	Difference for Productivity	Total Wage Difference because of Variables
2007-2005	Constant		0.184***	0.184
	Experience	0.001	-0.044***	-0.044
	Experience²	0.000	0.013	0
	Education	0.006***	-0.014	0.006
	Education²	0.016***	-0.021	0.016
	Total Wage Difference by Effect	**0.022 (13.6%)**	**0.14 (86.4%)**	**0.162**
2009-2007	Constant		0.096***	0.096
	Experience	0.013***	-0.041***	-0.028
	Experience²	-0.010***	0.020**	0.01
	Education	0.000	-0.100***	-0.1
	Education²	-0.001	0.043	0.043
	Total Wage Difference by Effect	**0.003 (14.3%)**	**0.018 (85.7%)**	**0.021**

Total differential in the natural logarithm of wages, by effect or variable, are calculated considering only significant effects with at least 90% of confidence level
Wage difference in percentages attributable to each effect is in parenthesis
*, **, *** Significant at a 90%, 95% and 99% confidence level, respectively

5.4 Discrimination effects

Table 3.2 summarizes the male-female wage differentials for 2005, 2007 and 2009. Analyzing the quality in human capital indicates that wage differentials should favour women. As it has been discussed before women have higher levels of education and men have higher levels of experience and the marginal returns for education are greater than those of experience. Nevertheless, men are better paid because discrimination effects are more than two times larger that quality effects. It is interesting to pay attention to differences attributable to squared variables, which all favour women. This means that marginal returns for experience peaks at a higher level for men and marginal returns for education are much larger for women when they accumulate this human capital investment.

[19] See more statistics and productivity effects for women and men in Appendix D

Table 3.2 Discrimination rffects

Equation (8)				
$\overline{\ln w_t^*} - \overline{\ln w_{t*}}$	Variable	Difference for Quality	Difference for Discrimination	Total Wage Difference because of Variables
	Constant		0.28***	0.28
	Experience	0.035***	-0.002	0.035
	Experience²	-0.027***	-0.024*	-0.027
	Education	-0.013***	-0.066*	-0.013
	Education²	-0.035***	-0.045**	-0.08
2005	Total Wage Difference by Effect	-0.04 (21.1%)	0.15 (78.9%)	0.11
	Constant		0.206***	0.206
	Experience	0.036***	-0.005	0.036
	Experience²	-0.031***	-0.020	-0.031
	Education	-0.012***	0.006	-0.012
	Education²	-0.043***	-0.069***	-0.112
2007	Total Wage Difference by Effect	-0.05 (26.7%)	0.137 (73.3%)	0.087
	Constant		0.232***	0.232
	Experience	0.020***	0.009	0.02
	Experience²	-0.018***	-0.039***	-0.057
	Education	-0.007***	-0.017	-0.007
2009	Education²	-0.048***	-0.071***	-0.119
	Total Wage Difference by Effect	-0.053 (30.3%)	0.122 (69.7%)	0.069

Total differential in the natural logarithm of wages, by effect or variable, are calculated considering only significant effects with at least 90% of confidence level
Wage difference in percentages attributable to each effect is in parenthesis
*, **, *** Significant at a 90%, 95% and 99% confidence level, respectively

Effects of discrimination in wage differentials have not clearly diminished across time. Table 3.3 shows that estimated intercept in 2007 is lower than in 2005. Nonetheless there is no statistical difference in discrimination effects between periods.

3.3 Probing different discrimination effects across time

Equation (8)		Ho: No difference in wages caused by discrimination Ha: There are difference in variables caused by discrimination
$\overline{\ln w_t^*} - \overline{\ln w_{t*}}$	Variable	$Ho: \overline{X_t^*}\Delta a = \overline{X_{t-1}^*}\Delta a$ $Ha: \overline{X_t^*}\Delta a \neq \overline{X_{t-1}^*}\Delta a$
2005-2007	Constant	Ho rejected**
	Experience	Ho not rejected
	Experience²	Ho not rejected
	Education	Ho not rejected
	Education²	Ho not rejected
	Total Wage Difference for Discrimination	Ho not rejected
2007-2009	Constant	Ho not rejected
	Experience	Ho not rejected
	Experience²	Ho not rejected
	Education	Ho not rejected
	Education²	Ho not rejected
	Total Wage Difference for Discrimination	Ho not rejected

*, **, *** Significant at a 90%, 95% and 99% confidence level, respectively

5.5 Effects of discrimination and productivity

Using equation (12) it is possible to combine different results from regressions to analyze wage differentials by quality, discrimination and productivity effects. Table 3.5 shows differences in salaries between men in 2007 (2009) and women in 2005 (2007) explained by discrimination, productivity, quality in human capital by gender and quality in human capital across time.

Discrimination explains the greater part of wage differentials between men of 2005 and women of 2007. The second greater component of wage differential for the same periods is the difference attributable to productivity across time. Both effects favour men. In Table 3.5 the wage differentials for men in 2009 and women in 2007 are also better explained for discrimination effects but, the second greater component is the quality in human capital component that favour women. The better paid remunerations for each year of education favour women, even so, the discrimination, productivity and quality in human capital effects across time are positive and total wage differential favour men

Table 3.4 Effects of quality, discrimination and productivity in 2007.

Equation (12) Men in 2007 vs Women in 2005					
Variable	Difference for Discrimination	Difference for Quality , by gender	Difference for Quality across time	Productivity	Wage Difference
Constant	0.206***			0.156***	0.362
Experience	-0.005	0.035***	0.002	-0.046***	-0.011
$Experience^2$	-0.020	-0.027***	-0.003	0.015	-0.027
Education	0.006	-0.013***	0.005***	0.017	-0.008
$Education^2$	-0.069***	-0.035***	0.013***	-0.030*	-0.091
Total Wage Difference by Effect	**0.137 (44.9%)**	**-0.04 (13.1%)**	**0.018 (5.9%)**	**0.08 (36.1%)**	**0.195**

Total differential in the natural logarithm of wages, by effect or variable, are calculated considering only significant effects with at least 90% of confidence level
Wage difference in percentages attributable to each effect is in parenthesis
*, **, *** Significant at a 90%, 95% and 99% confidence level, respectively

Table 3.5 Effects of quality, discrimination and productivity in 2009.

Equation (12) Men in 2009 vs Women in 2007					
Variable	Difference for Discrimination	Difference for Quality by gender	Difference for Quality across time	Productivity	Wage Difference
Constant	0.232***			0.101***	0.333
Experience	0.009	0.036***	0.008***	-0.033***	0.011
$Experience^2$	-0.039***	-0.031***	-0.006***	0.014	-0.076
Education	-0.017	-0.012***	0.000	-0.101***	-0.113
$Education^2$	-0.071***	-0.043***	-0.001	0.038**	-0.076
Total Wage Difference by Effect	**0.122 (68.2%)**	**-0.05 (27.9%)**	**0.002 (1.1%)**	**0.005 (2.8%)**	**0.079**

Total differential in the natural logarithm of wages, by effect or variable, are calculated considering only significant effects with at least 90% of confidence level
Wage difference in percentages attributable to each effect is in parenthesis
*, **, *** Significant at a 90%, 95% and 99% confidence level, respectively

6. Findings and conclusions

In this paper wage discrimination and productivity effects considering gender and time are computed for the first quarter of 2005 to the third quarter of 2009 using ENOE. Results presented prove the existence of discrimination and productivity effects in wage differentials. In general, according to the significance of the estimated effects it can be concluded that:

Quality

- There is a positive and significant effect across time of the effects of education and experience on wage differentials.
-Women are rewarded by their higher average levels of education. Although, men have more years of experience those remuneration suffer decreasing marginal returns. On the other hand, women have more years of school and the remunerations for this variable have increasing marginal returns.

Productivity

- In all regressions (total, men and women) the intercept increases across time. This increase indicates that there are wage increases between the periods that are independent of schooling and work experience.
- Returns to years of education or work experience decrease across time.

Discrimination

- The estimated intercepts indicate that there are wage differentials that are independent of experience or education that favour men.
- Payments for years of schooling favour women. Although men receive better remunerations for each year of experience than women the remunerations for women's education are much larger.

In Mexico, the proposed methodology used in this paper can be employed by STPS, or other political organization, to encourage labour market policies that favour human capital creation, better paid jobs and the elimination of gender discrimination. There are three points that labour policies encouraging human capital accumulation in Mexico should consider:

1. The wage differentials explained by quality in human capital must be protected and encouraged. Although the quality effect is significant and positive for all estimations presented here the size of its effects is relatively small.
2. Wage differentials across time must favour improvements in education and experience, not just improvements in the minimal wages.
3. Occupational barriers for woman must be removed in addition to putting incentives in place for those women to encourage their education and take advantage of greater returns for each year of study than men.

ENOE offers extensive socio-demographic information for more than 402,000 individuals. One interesting extension would be the analysis of the seven industry sectors organized by Mexican federal entities: 1) Construction; 2) Manufacturing; 3) Commerce; 4) Services; 5) Other; 6) Agriculture; 7) Not specified. In this extension, the purpose of analysis could be proving that industry sectors are comparable across time and between federal entities. Another extension would be the aggregation of different explanatory variables that better explain the Mexican labour market or a particular production sector.

References

Becker, G. (1971), "The economics of discrimination," The University of Chicago Press

Blinder, Alan S. (1973), "Wage Discrimination: Reduced Form and Structural Estimates," Journal of Human Resources 8:436.455.

Blinder, Alan S. (1974), "On Dogmatism in Human Capital Theory," Princeton University, Department of Economics, Industrial Relations Section.

Cox, Thomas and Michael Wohlgenant (1986), "Prices and Quality Effects on Cross-sectional Demand Analysis," *American Journal of Agricultural Economics* 68: 908-919

Dimitrios A Giannias (1998), "Labour Quality and Wage Differentials. Scottish Journal of Political Economy," 45: 188-97

points that labour policies encouraging human capital accumulation in Mexico should consider:

1. The wage differentials explained by quality in human capital must be protected and encouraged. Although the quality effect is significant and positive for all estimations presented here the size of its effects is relatively small.
2. Wage differentials across time must favour improvements in education and experience, not just improvements in the minimal wages.
3. Occupational barriers for woman must be removed in addition to putting incentives in place for those women to encourage their education and take advantage of greater returns for each year of study than men.

ENOE offers extensive socio-demographic information for more than 402,000 individuals. One interesting extension would be the analysis of the seven industry sectors organized by Mexican federal entities: 1) Construction; 2) Manufacturing; 3) Commerce; 4) Services; 5) Other; 6) Agriculture; 7) Not specified. In this extension, the purpose of analysis could be proving that industry sectors are comparable across time and between federal entities. Another extension would be the aggregation of different explanatory variables that better explain the Mexican labour market or a particular production sector.

References

Becker, G. (1971), "The economics of discrimination," The University of Chicago Press

Blinder, Alan S. (1973), "Wage Discrimination: Reduced Form and Structural Estimates," Journal of Human Resources 8:436.455.

Blinder, Alan S. (1974), "On Dogmatism in Human Capital Theory," Princeton University, Department of Economics, Industrial Relations Section.

Cox, Thomas and Michael Wohlgenant (1986), "Prices and Quality Effects on Cross-sectional Demand Analysis," *American Journal of Agricultural Economics* 68: 908-919

Dimitrios A Giannias (1998), "Labour Quality and Wage Differentials. Scottish Journal of Political Economy," 45: 188-97

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo (June 2010), “Decomposition Methods in Economics,” National Boreau of Economic Research

Hanushek, Eric A. and Dennis Kimko (2000), "Schooling, Labor Force Quality, and the Growth of Nations," American Economic Review, 90(5): 1184-1208

Hanushek, Eric A. and Dongwook Kim (1995), " Schooling, Labor Force Quality, and Economic Growth ," National Bureau of Economic Research

Hicks, W. and S. R. Johnson (1968), “Quantity and Quality Components for Income Elasticities of Demand for Food,” *American Journal of Agricultural Economics* 50: 1512-1517.

Mincer, Jacob (1970), “The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach,” Journal of Economic Literature, American Economic Association 8: 1-26

Mincer, Jacob (1974), “Progress in Human Capital Analysis of the Distribution of Earnings,” National Bureau of Economic Research

Nelson, Julie (1991), “Quality variation and Quality Aggregation in Consumer Demand for Food,” American Journal of Agricultural Economics 73: 1204-1212

Nicholson, J.L. (1967), “The Measurement of Quality Changes,” The Economic Journal 77: 152-530

Oaxaca, Ronald L. (1973), “Male-Female Wage Differentials in Urban Labor Markets,” International Economic Review 14: 693-709

Oaxaca, Ronald L., Tracy L. Regan and Galen Burghardt (2007), “A Human Capital Model of the effects of Ability and Family Background on Optimal Schooling Levels,” Economic Inquiry 45: 721-738

Rogers, W. H. (1993). “Regression standard errors in clustered samples,” Stata Technical Bulletin 13: 19–23.

Villarreal, Hector J. and Aashish Mehta (2003), “Returns to Schooling, Institutions and Heterogeneous Diploma Effects: An Expanded Mincerian Framework applied to Mexico,” University of Wisconsin, Agricultural and Applied Economics

White, Halbert (1980), “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity,” Econometrica 48: 817-838

Appendix

Appendix A. Algebra and theoretical framework

Appendix A1) Wage differential decomposition

Using one again

1. $\overline{\ln w_t^*} = \overline{X_t^*}\alpha_1$
2. $\overline{\ln w_{t-1}^*} = \overline{X_{t-1}^*}\alpha_2$
3. $\overline{\ln w_{t*}} = \overline{X_{t*}}\alpha_3$
4. $\overline{\ln w_{t-1*}} = \overline{X_{t-1*}}\alpha_4$

using 1 with 3 to differentiate wages between advantaged and disadvantaged group in period t

$$\overline{\ln w_t^*} - \overline{\ln w_{t*}} = \overline{X_t^*}\alpha_1 - \overline{X_{t*}}\alpha_3$$

$$\Delta\alpha_{13} = \alpha_1 - \alpha_3$$

$$\overline{\ln w_t^*} - \overline{\ln w_{t*}} = \overline{X_t^*}\Delta\alpha_{13} + \left(\overline{X_t^*} - \overline{X_{t*}}\right)\alpha_3 \qquad \text{(a)}$$

or

$$\overline{\ln w_t^*} - \overline{\ln w_{t*}} = \overline{X_{t*}}\Delta\alpha_{13} + \left(\overline{X_t^*} - \overline{X_{t*}}\right)\alpha_1 \qquad \text{(a')}$$

and using 3 with 4 for differentiating wages between disadvantaged group across time

$$\overline{\ln w_{t*}} - \overline{\ln w_{t-1*}} = \overline{X_{t*}}\alpha_3 - \overline{X_{t-1*}}\alpha_4$$

$$\Delta\alpha_{34} = \alpha_3 - \alpha_4$$

$$\overline{\ln w_{t*}} - \overline{\ln w_{t-1*}} = \overline{X_{t*}}\Delta\alpha_{34} + \left(\overline{X_{t*}} - \overline{X_{t-1*}}\right)\alpha_4 \qquad \text{(b)}$$

or

$$\overline{\ln w_{t*}} - \overline{\ln w_{t-1*}} = \overline{X_{t-1*}}\Delta\alpha_{34} + \left(\overline{X_{t*}} - \overline{X_{t-1*}}\right)\alpha_3 \qquad \text{(b')}$$

An integrated model with quality, quantity and discrimination components is obtained by adding (a)+(b), (a')+(b'), (a')+(b), (a)+(b'). Examining the first combination:

(a) + (b)

$$\overline{\ln w_t^*} - \overline{\ln w_{t-1*}} = \overline{X_t^*}\Delta\alpha_{13} + \left(\overline{X_t^*} - \overline{X_{t*}}\right)\alpha_3 + \overline{X_{t*}}\Delta\alpha_{34} + \left(\overline{X_{t*}} - \overline{X_{t-1*}}\right)\alpha_4$$

where wage differentials of advantaged and disadvantaged from one period to the next one can be interpreted as a discrimination component in period *t*, difference in means of advantaged and disadvantaged group in period *t*, the productivity component of disadvantage group and difference from means of disadvantage group between periods, in that order.

Appendix A2) White test

The White test is a statistical test proposed by Halbert White (1980) that proves if the variance of regression residuals is constant (homoskedasticity). He makes use of the squared residuals from an OLS regression in terms of explained variables, the squared explained variables and the cross-terms of those independent variables, to estimate an auxiliary regression. This auxiliary specification for equation (13) is

$$\hat{u}_{it} = \theta_0 + \theta_1 S_{it} + \theta_2 S_{it}^2 + \theta_3 E_{it} + \theta_4 E_{it}^{\ 2} + \upsilon_{it}$$

where $\hat{u}_{it}$ are the estimated residuals for variable i in period t and it is calculated for all the estimations presented.

Looking for homoscedasticity the null and alternative hypothesis are:

Ho: $\theta_1 = \theta_2 = \theta_3 = \theta_4 = 0$	homoscedasticity
Ha: at least one is different	heteroscedasticity

The test statistic is the product of the R^2 of the estimated auxiliary regression and the number of observations employed ($T*R^2$). Ho is rejected and heteroskedasticity problem is present if the test statistic is bigger than a Chi-square statistic with 4 degrees of freedom

Appendix A3) Robust standard errors

In OLS the matrix estimator for the coefficients of the regression model $Y = X\beta + \ell$, assuming $\ell \sim N(0, \sigma^2)$, is

$$\hat{\beta} = (X'X)^{-1} X'Y$$

where β is a k (number of coefficients in a regression) x 1 vector, X is a T (number of observations) x k matrix and y is a T x 1 vector. When the estimated residuals are heteroskedastic they have different variances and does not meet OLS typical assumptions. In this case, the coefficient standard errors will be biased and therefore regression analysis will be biased. Biased standard errors lead to biased inference and the parameters' hypothesis testing does not have dough confidence. The robust standard errors allow for confidence and are used against heteroskedasticity. All the regressions were estimated and corrected using the statistical software Stata. This program uses the robust estimate of variance developed by Rogers (1993). He takes into account the White (1980) method to derive the heteroskedasticity-consistent covariance estimator. In this procedure everything is considered conditional on X, then $(X'X)^{-1}$ is considered a constant matrix. Consequently, the coefficients variance is

$$Var(\hat{\beta}) = Var\left((X'X)^{-1}X'Y\right) = (X'X)^{-1}Var\left((X'Y)\right)(X'X)^{-1} \qquad \text{(c)}$$

Nonetheless, the matrix dimension of $X'Y$ is k x 1. Due to the fact that X is considered conditional, it can be treated as a constant, consequently the variance for the first column of X is;

$$Var(X_1'Y) = x_{11}^2 Var(y_1) + x_{21}^2 Var(y_2) + \ldots + x_{T1}^2 Var(y_T)$$

assuming that y_t are independent. Taking the estimated squared residuals $(\hat{\ell}^2)$ as estimate for $Var(y_t)$ Rogers (1993) estimate the off-diagonal terms of variance-covariance matrix of $X'Y$ and establishing that

$$\hat{V}ar(X'Y) = \sum_{t=1}^{T} \hat{\ell}_t^2 x_t' x_t$$

where x_t is a 1 x k vector and $x_t'x_t$ is a matrix of k x k.

substituting in (c)

$$Var(\hat{\beta}) = (X'X)^{-1} \sum_{t=1}^{T} \hat{\ell}_t^2 x_t' x_t \, (X'X)^{-1} \qquad \text{(d)}$$

This is the robust standard errors estimator for OLS computed coefficients.

Appendix A4) Variance Inflation Factor

The VIF (Variance inflation Factor) index attempts to estimate the severity of multicollinearity in OLS regressions. It offers a measure of how much the estimated regression coefficient variance is influenced because collinearity between explanatory variables. Following equation (13) the VIF statistic for each independent variable is estimated from the auxiliary regression models:

$$S_{it} = \eta_0 + \eta_1 S_{it}^2 + \eta_2 E_{it} + \eta_3 E_{it}^{\ 2} + \mu_{1it} \qquad \text{(e)}$$

$$S_{it}^2 = \phi_0 + \phi_1 S_{it} + \phi_2 E_{it} + \phi_3 E_{it}^{\ 2} + \mu_{2it} \qquad \text{(f)}$$

$$E_{it} = \delta_0 + \delta_1 S_{it} + \delta_2 S_{it}^2 + \delta_3 E_{it}^{\ 2} + \mu_{3it} \qquad \text{(g)}$$

$$E_{it}^{\ 2} = \lambda_0 + \lambda_1 S + \lambda_2 S_{it}^2 + \lambda_3 E_{itit} + \mu_{4it} \qquad \text{(h)}$$

The VIF estimators for (e),(f), (g) and (h) are calculated using the determination coefficient of each regression. In alphabetical order: $VIF_s=1/(1-R_{s2})$; $VIF_{s^2}=1/(1-R_{s2}^{\ 2})$; $VIF_E=1/(1-R_E^2)$ and; $VIF_{E2}=1/(1-R_{E2}^2)$. Multicollinearity is reflected in the size of the VIF value. The bigger the VIF index is, more multicollinearity suspicious is present.

Appendix B. Descriptive statistics

Appendix B1) Descriptive statistics of variables in ENOE 2005

Individuals	Variable	Mean	Observations	Standard deviation	Maximum	Minimum	Skewness	Kurtosis
All	ln(wage/hour)	2.877864	132180	.8246283	10.86085	-2.873869	-.1317145	5.187191
	Experience	21.92918	132180	15.56439	92	0	.8009949	3.257186
	Experience^2	723.1375	132180	929.1716	8464	0	2.242253	9.308102
	Yeas of school	8.969587	132180	4.59799	24	0	.0765853	2.507707
	(Yeas of school)^2	101.5948	132180	88.00353	576	0	1.068094	3.298431
Men	ln(wage/hour)	2.916773	83488	.8151556	10.86085	-2.873869	-.1553523	5.563531
	Experience	22.35188	83488	15.72263	92	0	.7901836	3.192595
	Experience^2	746.8049	83488	950.0432	8464	0	2.174321	8.829531
	Yeas of school	8.802403	83488	4.562434	24	0	.1469213	2.569189
	(Yeas of school)^2	98.29785	83488	87.30767	576	0	1.162653	3.547853
Women	ln(wage/hour)	2.811151	48692	.8364287	7.751725	-2.798358	-.0837295	4.658
	Experience	21.20441	48692	15.2622	92	0	.816278	3.367416
	Experience^2	682.5569	48692	890.7968	8464	0	2.366672	10.25508
	Yeas of school	9.256243	48692	4.644385	23	0	-.0437106	2.44665
	(Yeas of school)^2	107.2479	48692	88.90083	529	0	.9174021	2.94815

Appendix B2) Descriptive statistics of variables in ENOE 2007

Individuals	Variable	Mean	Observations	Maximum	Minimum	Standard deviation	Skewness	Kurtosis
All	ln(wage/hour)	3.01999	132055	8.088228	-1.795105	.8039172	-.1323129	5.287168
	Experience	21.96686	132055	91	0	15.53781	.7541782	3.165335
	Experience²	723.9646	132055	8281	0	916.9388	2.189893	8.975825
	Yeas of school	9.236303	132055	24	0	4.55551	.0310387	2.528537
	(Yeas of school)²	106.0618	132055	576	0	88.59057	1.006494	3.167191
Men	ln(wage/hour)	3.045999	82298	7.751725	-4.143955	.8000572	-.2146205	5.727546
	Experience	22.43208	82298	91	0	15.80194	.7591297	3.142161
	Experience²	752.8962	82298	8281	0	951.2758	2.148767	8.622235
	Yeas of school	9.022771	82298	24	0	4.523022	.1059887	2.582334
	(Yeas of school)²	101.8679	82298	576	0	87.62456	1.110274	3.438944
Women	ln(wage/hour)	2.976971	49757	8.088228	-1.795105	.8084349	.0021394	4.652243
	Experience	21.1974	49757	91	0	15.05944	.7324466	3.164214
	Experience²	676.1118	49757	8281	0	854.9947	2.234161	9.463854
	Yeas of school	9.589485	49757	23	0	4.587022	-.0940677	2.494315
	(Yeas of school)²	112.9986	49757	529	0	89.73732	.8493586	2.81381

Appendix B3) Descriptive statistics of variables in ENOE 2009

Individuals	Variable	Mean	Observations	Maximum	Minimum	Standard deviation	Skewness	Kurtosis
All	ln(wage/hour)	3.041004	121115	9.883006	-2.488192	.7987146	-.1546154	5.338482
	Experience	22.41728	121115	92	0	15.67448	.7180989	3.087172
	Experience^2	748.2218	121115	8464	0	931.5741	2.13465	8.685012
	Yeas of school	9.24949	121115	24	0	4.507804	.0196701	2.56553
	$(\text{Yeas of school})^2$	105.8732	121115	576	0	87.56107	1.007888	3.17978
Men	ln(wage/hour)	3.064836	74577	8.638494	-2.844901	.7908782	-.1845634	5.650895
	Experience	22.7049	74577	92	0	15.86582	.7301305	3.068376
	Experience^2	767.2331	74577	8464	0	956.5338	2.094551	8.300274
	Yeas of school	9.038819	74577	24	0	4.458088	.0877726	2.630146
	$(\text{Yeas of school})^2$	101.5745	74577	576	0	86.13996	1.1111	3.456759
Women	ln(wage/hour)	3.002813	46538	9.883006	-2.488192	.8096616	-.1040841	4.904821
	Experience	21.95638	46538	92	0	15.35183	.6915543	3.099386
	Experience^2	717.7563	46538	8464	0	889.2809	2.193236	9.346074
	Yeas of school	9.58709	46538	24	0	4.566173	-.0922704	2.511405
	$(\text{Yeas of school})^2$	112.7618	46538	576	0	89.36228	.856182	2.832044

Appendix C Econometric estimations

Appendix C11) Estimation of Mincerian equation in ENOE 2005

All (2005)	Coefficient	VIF	Robust Std. Error	t statistic	Probability
Constant	1.940468		.0107387	180.70	0.000
Experience	.0305732	15.48	.0004172	73.28	0.000
$Experience^2$	-.0004302	13.53	8.08e-06	-53.26	0.000
Education	.0231898	9.60	.0018779	12.35	0.000
$Education^2$	.0036423	9.24	.0000888	41.01	0.000
R^2	0.2568				
Adjusted R^2	0.2568				
Observations	132180				
F statistic	10652.09				
Prob(F)	0.0000				
White (obs*R^2)	2838.84				
Prob(White)	0.0000				

Appendix C12) Estimation of Mincerian equation for men in ENOE 2005

Men (2005)	Coefficient	VIF	Robust Std. Error	t statistic	Probability
Constant	2.033856		.0132568	153.42	0.000
Experience	.0308237	15.10	.0005244	58.78	0.000
$Experience^2$	-.0004451	13.30	9.97e-06	-44.64	0.000
Education	.0206856	9.85	.002337	8.85	0.000
$Education^2$	.0035021	9.47	.0001116	31.37	0.000
R^2	0.2473				
Adjusted R^2	0.2473				
Observations	83488				
F statistic	6129.26				
Prob(F)	0.0000				
White (obs*R^2)	1867.60				
Prob(White)	0.0000				

Appendix C13) Estimation of Mincerian equation for women in ENOE 2005

Women (2005)	Coefficient	VIF	Robust Std. Error	t statistic	Probability
Constant	1.752697		.0181103	96.78	0.000
Experience	.0308966	16.13	.0006894	44.81	0.000
Experience²	-.0004136	13.93	.0000138	-29.89	0.000
Education	.0281687	9.25	.0031125	9.05	0.000
Education²	.0039615	8.96	.000145	27.32	0.000
R^2	0.2868				
Adjusted R^2	0.2868				
Observations	48692				
F statistic	4836.49				
Prob(F)	0.0000				
White (obs*R^2)	1036.08				
Prob(White)	0.0000				

Appendix C21) Estimation of Mincerian equation in ENOE 2007

All (2007)	Coefficient	VIF	Robust Std. Error	t statistic	Probability
Constant	2.124713		.0109194	194.58	0.000
Experience	.0285927	15.95	.0004063	70.38	0.000
Experience²	-.0004121	14.00	7.95e-06	-51.82	0.000
Education	.0216605	9.62	.0019045	11.37	0.000
Education²	.0034457	9.16	.000089	38.70	0.000
R^2	0.2461				
Adjusted R^2	0.2460				
Observations	132055				
F statistic	9714.79				
Prob(F)	0.0000				
White (obs*R^2)	2625.57				
Prob(White)	0.0000				

Appendix C22) Estimation of Mincerian equation for men in ENOE 2007

Men (2007)	Coefficient	VIF	Robust Std. Error	t statistic	Probability
Constant	2.190275		.0136365	160.62	0.000
Experience	.0287842	15.59	.0005088	56.57	0.000
Experience^2	-.0004247	13.75	9.74e-06	-43.58	0.000
Education	.0225481	9.78	.0024066	9.37	0.000
Education^2	.0032035	9.31	.0001138	28.14	0.000
R^2	0.2386				
Adjusted R^2	0.2385				
Observations	82298				
F statistic	5532.31				
Prob(F)	0.0000				
White (obs*R^2)	2027.70				
Prob(White)	0.0000				

Appendix C23) Estimation of Mincerian equation for women in ENOE 2007

Women (2007)	Coefficient	VIF	Robust Std. Error	t statistic	Probability
Constant	1.983905		.0181842	109.10	0.000
Experience	.0290093	16.54	.0006785	42.76	0.000
Experience^2	-.0003988	14.41	.0000139	-28.77	0.000
Education	.0218339	9.52	.0031065	7.03	0.000
Education^2	.0038796	9.05	.0001426	27.21	0.000
R^2	0.2702				
Adjusted R^2	0.2701				
Observations	49757				
F statistic	4456.35				
Prob(F)	0.0000				
White (obs*R^2)	672.34				
Prob(White)	0.0000				

Appendix C31) Estimation of Mincerian equation in ENOE 2009

All (2009)	Coefficient	VIF	Robust Std. Error	t statistic	Probability
Constant	2.220668		.0115976	191.48	0.000
Experience	.0267719	15.75	.0004256	62.91	0.000
Experience^2	-.0003852	13.84	8.21e-06	-46.91	0.000
Education	.0108894	9.78	.002017	5.40	0.000
Education^2	.0038505	9.30	.0000942	40.86	0.000
R^2	0.2282				
Adjusted R^2	0.2282				
Observations	121115				
F statistic	8247.99				
Prob(F)	0.0000				
White (obs*R^2)	2186.73				
Prob(White)	0.0000				

Appendix C32) Estimation of Mincerian equation for men in ENOE 2009

Men (2009)	Coefficient	VIF	Robust Std. Error	t statistic	Probability
Constant	2.291201		.0143717	159.42	0.000
Experience	.0273105	15.25	.00054	50.57	0.000
Experience^2	-.000407	13.50	.0000102	-39.75	0.000
Education	.0113464	9.95	.0025299	4.48	0.000
Education^2	.0035759	9.46	.0001202	29.76	0.000
R^2	0.2173				
Adjusted R^2	0.2173				
Observations	74577				
F statistic	4521.78				
Prob(F)	0.0000				
White (obs*R^2)	1639.31				
Prob(White)	0.0000				

Appendix C33) Estimation of Mincerian equation for women in ENOE 2009

Women (2009)	Coefficient	VIF	Robust Std. Error	t statistic	Probability
Constant	2.058805		.019626	104.90	0.000
Experience	.0269107	16.56	.0006921	38.88	0.000
Experience^2	-.0003558	14.39	.0000138	-25.81	0.000
Education	.0132072	9.61	.0033329	3.96	0.000
Education^2	.0042738	9.17	.0001521	28.11	0.000
R^2	0.2572				
Adjusted R^2	0.2572				
Observations	46538				
F statistic	3962.71				
Prob(F)	0.0000				
White (obs*R^2)	634.96				
Prob(White)	0.0000				

Appendix D. Quality, productivity and discrimination effects

Appendix D1) Quality and productivity effects of Equation (9).

Equation (9)			Productivity			Quality		
Time	Gender	Variable	Value	s.e.	t stat	Value	s.e.	t stat
2005-2007	All	Constant	0.184***	0.015	12.030			
		Experience	-0.044***	0.013	-3.401	0.001	0.002	0.623
		Experience²	0.013	0.008	1.597	0.000	0.002	-0.230
		Education	-0.014	0.025	-0.572	0.006***	0.000	14.978
		Education²	-0.021	0.013	-1.564	0.016***	0.001	13.003
	Men	Constant	0.156***	0.019	8.225			
		Experience	-0.046***	0.016	-2.791	0.002	0.002	1.036
		Experience²	0.015	0.010	1.464	-0.003	0.002	-1.304
		Education	0.017	0.030	0.555	0.005***	0.000	9.876
		Education²	-0.030*	0.016	-1.873	0.013***	0.002	8.309
	Women	Constant	0.231***	0.026	9.009			
		Experience	-0.040*	0.021	-1.951	0.000	0.003	-0.073
		Experience²	0.010	0.013	0.756	0.003	0.002	1.158
		Education	-0.061	0.042	-1.441	0.009***	0.001	11.325
		Education²	-0.009	0.023	-0.403	0.023***	0.002	10.101
2007-2009	All	Constant	0.096***	0.016	6.024			
		Experience	-0.041***	0.013	-3.094	0.013***	0.002	7.253
		Experience²	0.020**	0.009	2.354	-0.010***	0.002	-6.594
		Education	-0.100***	0.026	-3.883	0.000	0.000	0.732
		Education²	0.043***	0.014	3.124	-0.001	0.001	-0.538
	Men	Constant	0.101***	0.020	5.094			
		Experience	-0.033***	0.017	-1.986	0.008***	0.002	3.408
		Experience²	0.014	0.011	1.255	-0.006***	0.002	-2.972
		Education	-0.101***	0.032	-3.208	0.000	0.001	0.707
		Education²	0.038**	0.017	2.250	-0.001	0.001	-0.668
	Women	Constant	0.075***	0.027	2.799			
		Experience	-0.046***	0.021	-2.165	0.022***	0.003	7.737
		Experience²	0.031**	0.014	2.195	-0.017***	0.002	-7.398
		Education	-0.083*	0.044	-1.893	0.000	0.001	-0.081
		Education²	0.044*	0.024	1.891	-0.001	0.002	-0.410

*, **, *** Significant at a 90%, 95% and 99% confidence level, respectively

Appendix D2) Quality and discrimination effects Equation (8)

Equation (8)		Discrimination			Quality		
Time	Variable	Value	s.e.	t stat	Value	s.e.	t stat
2005	Constant	0.28***	0.022	12.527			
2005	Experience	-0.002	0.019	-0.084	0.035***	0.003	13.039
2005	$Experience^2$	-0.024*	0.013	-1.850	-0.027***	0.002	-12.340
2005	Education	-0.066*	0.034	-1.923	-0.013***	0.001	-17.248
2005	$Education^2$	-0.045**	0.018	-2.511	-0.035***	0.002	-17.772
2007	Constant	0.206***	0.023	9.079			
2007	Experience	-0.005	0.019	-0.265	0.036***	0.003	14.170
2007	$Experience^2$	-0.020	0.013	-1.526	-0.031***	0.002	-15.150
2007	Education	0.006	0.035	0.182	-0.012	0.001	-21.870
2007	$Education^2$	-0.069***	0.019	-3.706	-0.043***	0.002	-22.036
2009	Constant	0.232***	0.024	9.554			
2009	Experience	0.009	0.020	0.455	0.020***	0.002	8.148
2009	$Experience^2$	-0.039***	0.013	-2.984	-0.018***	0.002	-9.146
2009	Education	-0.017	0.038	-0.445	-0.007***	0.000	-20.511
2009	$Education^2$	-0.071***	0.020	-3.600	-0.048***	0.002	-21.487

*, **, *** Significant at a 90%, 95% and 99% confidence level, respectively

Appendix D3) Quality and productivity effects of Equation (9').

Equation (9')			Productivity			Quality		
Time	Gender	Variable	Productivity	se	t stat	Quality	se	t stat
2005-2007	All	Constant	0.184***	0.015	12.030			
		Experience	-0.043***	0.013	-3.395	0.001	0.002	0.582
		Experience²	0.013	0.008	1.595	0.000	0.002	-0.221
		Education	-0.014	0.025	-0.555	0.006***	0.000	13.990
		Education²	-0.020	0.013	-1.498	0.015***	0.001	12.301
	Men	Constant	0.156***	0.019	8.225	0.000		
		Experience	-0.046***	0.016	-2.781	0.002	0.002	0.967
		Experience²	0.015	0.010	1.452	-0.003	0.002	-1.245
		Education	0.016	0.030	0.542	0.005***	0.000	10.765
		Education²	-0.029	0.016	-1.808	0.011***	0.002	7.601
	Women	Constant	0.231***	0.026	9.009	0.000		
		Experience	-0.040**	0.021	-1.952	0.000	0.003	-0.068
		Experience²	0.010	0.013	0.763	0.003	0.002	1.116
		Education	-0.059	0.042	-1.390	0.007***	0.001	8.778
		Education²	-0.009	0.023	-0.382	0.022***	0.002	9.892
2007-2009	All	Constant	0.096***	0.016	6.024	0.000		
		Experience	-0.040***	0.013	-3.032	0.012***	0.002	6.791
		Experience²	0.019**	0.009	2.277	-0.009***	0.002	-6.164
		Education	-0.099***	0.026	-3.877	0.000	0.000	0.368
		Education²	0.043***	0.014	3.129	-0.001	0.001	-0.602
	Men	Constant	0.101***	0.020	5.094	0.000		
		Experience	-0.033**	0.017	-1.962	0.007***	0.002	3.233
		Experience²	0.013	0.011	1.232	-0.006***	0.002	-2.849
		Education	-0.101***	0.032	-3.202	0.000	0.001	0.356
		Education²	0.038**	0.017	2.256	-0.001	0.001	-0.746
	Women	Constant	0.075***	0.027	2.799	0.000		
		Experience	-0.044**	0.021	-2.090	0.020***	0.003	7.178
		Experience²	0.029**	0.014	2.068	-0.015***	0.002	-6.601
		Education	-0.083*	0.044	-1.894	0.000	0.001	-0.049
		Education²	0.045*	0.024	1.895	-0.001	0.002	-0.452

*, **, *** Significant at a 90%, 95% and 99% confidence level, respectively

Appendix D4) Quality and discrimination effects Equation (8')

Equation (8')		Discrimination			Quality		
Time	Variable	Value	s.e.	t stat	Value	s.e.	t stat
2005	Constant	0.281***	0.022	12.527			
2005	Experience	-0.002	0.019	-0.080	0.035***	0.003	13.008
2005	Experience²	-0.022*	0.013	-1.691	-0.029***	0.002	-13.280
2005	Education	-0.069**	0.034	-2.022	-0.009***	0.001	-12.666
2005	Education²	-0.049***	0.018	-2.739	-0.031***	0.002	-15.711
2007	Constant	0.206***	0.023	9.079			
2007	Experience	-0.005	0.019	-0.251	0.036***	0.003	14.060
2007	Experience²	-0.018	0.013	-1.370	-0.033***	0.002	-16.134
2007	Education	0.007	0.035	0.193	-0.013***	0.001	-22.586
2007	Education²	-0.076***	0.019	-4.111	-0.036***	0.002	-18.196
2009	Constant	0.232***	0.024	9.554			
2009	Experience	0.009	0.020	0.440	0.020***	0.002	8.269
2009	Experience²	-0.037***	0.013	-2.791	-0.020***	0.002	-10.463
2009	Education	-0.018	0.038	-0.472	-0.006***	0.000	-17.621
2009	Education²	-0.079***	0.020	-3.996	-0.040***	0.002	-17.978

*, **, *** Significant at a 90%, 95% and 99% confidence level, respectively