DT
DECON

DOCUMENTO DE TRABAJO

N° 442

HAS THE GENDER WAGE GAP BEEN REDUCED DURING THE 'PERUVIAN GROWTH MIRACLE?' A DISTRIBUTIONAL APPROACH

Juan Manuel del Pozo Segura

DEPARTAMENTO DE ECONOMÍA
ECONOMÍA
ET LUX IN TENEBRIS LUCET
MCMXVII
100 años PUCP

DOCUMENTO DE TRABAJO N° 442

HAS THE GENDER WAGE GAP BEEN REDUCED DURING THE 'PERUVIAN GROWTH MIRACLE?' A DISTRIBUTIONAL APPROACH

Juan Manuel del Pozo Segura

Julio, 2017

DEPARTAMENTO
DE **ECONOMÍA**

DOCUMENTO DE TRABAJO 442
http://files.pucp.edu.pe/departamento/economia/DDD442.pdf

Has the Gender Wage Gap been Reduced during the
'Peruvian Growth Miracle?' A Distributional Approach
Documento de Trabajo 442

Editado e Impreso:
© Departamento de Economía – Pontificia Universidad Católica del Perú,
Av. Universitaria 1801, Lima 32 – Perú.
Teléfono: (51-1) 626-2000 anexos 4950 - 4951
econo@pucp.edu.pe
http://departamento.pucp.edu.pe/economia/publicaciones/documentos-de-trabajo/

Encargado de la Serie: Jorge Rojas Rojas
Departamento de Economía – Pontificia Universidad Católica del Perú,
jorge.rojas@pucp.edu.pe
Primera edición – Julio, 2017.
Tiraje: 50 ejemplares

Hecho el Depósito Legal en la Biblioteca Nacional del Perú Nº 2017-09549.
ISSN 2079-8466 (Impresa)
ISSN 2079-8474 (En línea)

Se terminó de imprimir en julio de 2017.

Has the gender wage gap been reduced during the 'Peruvian Growth Miracle'? A distributional approach

Juan Manuel del Pozo Segura *[1,2]

[1]*Departamento de Economía, Pontificia Universidad Católica del Perú*
[2]*The University of Sussex*

Abstract

Between 2004 and 2014 the Peruvian economy experienced a noticeable growth which surpassed most of Latin American countries during that period, leading some to quote this episode as the *Peruvian Growth Miracle*. Yet, growth of wages would not have been accompanied by an equally marked reduction in wage differentials between men and women despite government efforts to address this issue. Consequently, this study analyzes and decomposes the gender wage gap in Peru for 2004 and 2014 using the Machado and Mata (2005) decomposition method correcting for sample selection bias in the context of quantile regression (Albrecht et al. 2009). This allows to decompose the differential in terms of the endowment and treatment effect at each point of the income distribution instead of, as has been customary in previous studies for Peru, only at the average of the distribution. Using data from the National Household Survey, we find that unconditional and conditional gaps, which favour men, have deepened between 2004 and 2014 at every point of the distribution, while there is evidence of *sticky floors* and *glass ceilings* in both years. Decompositions consistently reveal that, for both years, discrimination against women is the most important factor behind gender gaps at each percentile even though the effect of endowments plays in favor of those. All in all, this raise doubts about the aggregate effectiveness of pro-equity policies applied in recent years.

*E-mail jmdelpozo@pucp.pe. El presente estudio constituye la tesis de postgrado del autor como estudiante del programa MSc Economics en The University of Sussex, bajo la supervisión de Peter Dolton. El autor agradece a Peter Dolton por su apoyo e interés en el desarrollo de este estudio, a Barry Reilly por sus comentarios y a los asistentes al Viernes Económico de la PUCP del 11 de Noviembre del 2016 y del Congreso 2017 de la Asociación Peruana de Economía por sus valiosos comentarios.

Resumen

Entre el 2004 y 2014 la economía peruana experimentó un crecimiento notable en relación al de otros países Latinoamericanos, al punto que algunos acuñaron el término *El Milagro Peruano* para referirse a este episodio. Sin embargo, el crecimiento de salarios no habría estado acompañado por una reducción igualmente notable de los diferenciales de salarios entre hombres y mujeres pese a los esfuerzos del gobierno para abordar esta problemática. Consecuentemente, el presente estudio analiza y descompone la brecha salarial de género en el Perú mediante el método de descomposición de Machado and Mata (2005) corrigiendo por sesgo de selección muestral en el contexto de regresión cuantílica (Albrecht et al. 2009). Éste posibilita descomponer el diferencial en términos del efecto dotaciones y tratamiento en cada punto de la distribución de ingresos y no, como se ha hecho en los estudios previos para el Perú, sólo para el promedio de aquella. Usando datos de la Encuesta Nacional de Hogares, los resultados señalan que las brechas incondicionales y condicionales, que favorecen a los hombres, han crecido entre 2004 y 2014 en cada punto de la distribución a la vez que se encuentra evidencia de *pisos pegajosos* y *techos de cristal* en ambos años. Las descomposiciones consistentemente revelan que, para ambos años, la discriminación contra la mujer es el factor más importante detrás de las brechas de género en cada percentil a pesar de que el efecto de dotaciones favorece a aquellas. Estos resultados generan dudas sobre la eficacia agregada de políticas pro-equidad aplicadas en los últimos años.

JEL classification: C01, J08, J16, O12
Keywords: Inequality, Distributional Decomposition, Gender wage gap, Quantile regressions, Peru

1 Introduction

In the last 16 years, Peru experienced a noteworthy economic growth: according to the WB's World Development Indicators data, real per capita GDP growth was around 4.5%, a rate remarkably higher than the Latin American average (1.8%). This lively performance was not only confined to the mere statistical field since, due to the increase of public expenditure and deepening of the decentralization process, there was a noticeable advance in different areas of human welfare such as education and health services (Beteta and Del Pozo 2016), infrastructure development in terms of improved connectivity (Webb 2013) and water and sanitation (World Bank 2010), among other features of social development (for a brief survey, see PCM 2013 and INEI 2014b). This led many to coin this episode as the *Peruvian Growth Miracle* (Ross and Peschiera 2015).

Yet, this enthusiasm has faced a strong critique: fast growth of wages and productivity in labour market (World Bank 2015; Távara et al. 2014) has not been accompanied by an equally rapid reduction in inequality. On the one hand, according to data from Peru's Household National Survey, Gini coefficient for personal labour wages fell from 0.52 to 0.48 between 2004 and 2014 (implying an annual rate of change lower than 1%). On the other hand, men earn systematically higher wages than women after controlling for differences in qualifications and occupations (MTPE 2014; OIT and PNUD 2009; CEPAL et al. 2013). Considering that higher inequality is not only an ethical concern but also a functional matter, provided it tends to reproduce in time and to reduce prospects for future aggregate growth and effectiveness of poverty alleviation policies (Adams 2003; Persson and Tabellini 1994; Deininger and Squire 1998), understanding its causes and determinants contributes to attain a sustainable development path.

Among all the dimensions that term 'labour market inequality' encompasses, we focus here on the gender aspect and, more specifically, on the gender wage gap. The interest on this area lies in the fact that there are still noticeable wage differences among males and females despite that real wages of the population are higher now than fifteen years ago (INEI 2014a). In fact, this disparity has been a recurrent concern not only for the Peruvian government, since gender-related issues constitute one of the most referenced aspects in the legal standards established in the last decade (MTPE 2010), but also for international agencies, as stated in the Millennium Development Goals (2000) which contemplates labour-related gender equality as a key objective (CEPAL et al. 2013; OIT and PNUD 2009).

Consequently, several studies analysed the extent of the gender wage gap in the Peruvian labour market and, chiefly, what portion of this can be attributed to a discrimination factor against women (i.e. not explained by observed characteristics such as age, education, area of residence, etc.). Despite of the advancement these researches provide in the understanding of gender inequality, they feature, at least, two important limitations. Firstly, some studies (e.g. Garavito 2011; Coppola and Calvo-Gonzalez 2011) aim to synthesize the gap by a gender dummy variable on a Mincer equation which is estimated by Ordinary Least Squares (OLS). The fact that these estimates reflect average wage gaps do not allow to address gender inequality occurring at the top and at the bottom end of the wage distribution such as *glass ceilings* (defined as the limit women have on their earning prospects such that after a point their wages fall behind men's) and *sticky floors* (defined as the a tendency of women to be confined to poorly paid jobs compared to men). Secondly, those studies which carry out decompositions of the wage gap in that part attributable to differences in characteristics and that part attributed to discrimination (Castillo 2011; Montes 2007; Yamada et al. 2013 as well as Atal et al. 2009 for Latin American countries) apply exclusively the Oaxaca-Blinder (Oaxaca 1973) decomposition to analyse differences in mean wages. Given the strong right-skew of Peru's wage distribution, this (underlying) homogeneity assumption is not realistic and, hence, differences of *only* a measure of central tendency offers a partial and imprecise explanation of the gender gap. This led to some to emphasize the need for a more complete approach in order to analyse gaps beyond a unique point and to consider, instead, the whole distribution (Jaramillo et al. 2007; CIES 2011).

Having this in mind, this study applies the Machado and Mata 2005 decomposition (MM) method which, in the same spirit as the typical Oaxaca-Blinder (OB) decomposition, separates the gender wage gap in two parts: that which arises because females and males have different observable characteristics after they receive the same treatment and that part which arises because one group is more favourably treated despite having the same individual characteristics, being the latter part typically associated with discrimination. However, compared to alternative approaches, MM has two notable advantages. On the one hand, relative to the OB approach, MM does not only focus only on differences between males and females wages in terms of their averages but, instead, on differences in wages at *any* given point of their corresponding distributions, revealing a more detailed picture of the gender inequality. Accordingly, we will be able to assess, for example, if gender inequality can be more attributable to differences in credentials for those individuals located at the first quartile compared to those at the median of the distribution, or if the gap is more attributable to differences in returns to their endowments for those at the first decile compared to those at the 95th percentile. On the other hand, relative to the distributional approach stated by DiNardo et al. (1996) and Firpo et al. (2007), MM can be complemented with a method for correcting for sample selection using Albrecht et al. (2009) method. This selection is an inherent characteristic of samples for labour markets outcomes and, if not accounted, can lead to inconsistent estimates.

This study advances the understanding of Peru's male-female wage differentials in several ways. On the first place, we apply quantile regression which, unlike OLS, provides marginal effects of covariates for any point of the wages distribution, enabling us to explore phenomena such as glass ceilings and sticky floors. Furthermore, we account for the possibility of sample selection, which represents an improvement over studies which take a similar distributional approach for Latin American countries such as Salardi (2012) for Brazil and Pacheco (2013) for Nicaragua. Their decompositions are carried out under the assumption that selection does not have an important effect on the estimates, which cannot be assumed a priori. On the second place, application of MM decomposition, which implies construction of a counterfactual *distribution* (based on the parameters estimated by quantile regression), allows to decompose the gender gap on labour wages at *any* point of the distribution and not, as studies for Peru have done so far, at *only* the mean of the distribution. On the third place, we apply this for 2004 and 2014. This lapse of time, comprised by two points in time ten years apart, is of special interest provided the exceptional economic growth (around 4.5% annual in terms of real GDP per capita) experienced as well as different policies put in place by the government as an effort to reduce gender wage gaps. Therefore, comparison of the results for these two years allows us to assess how this favourable macroeconomic environment is correlated with labour market inequality. Likewise, application of MM decomposition to a country where only classical OB decomposition has been considered so far will provide more rigorous policy recommendations about gender disparities.

The study is organized as follows. Section 2 presents a brief background on the Peruvian labour-related gender disparities in the last decade and highlights the different government policies to offset labour market-related gender inequalities. Section 3 describes the analytical framework applied: quantile regression estimation implementing sample selection correction as well as the MM method, assessing its strengths and weaknesses compared to alternative methods, and provides a brief review of key studies which applied this approach. Section 4 describes the dataset and shows some descriptives that characterize the data and changes accrued between 2004 and 2014. Section 5 presents, on the first place, the results of quantile regression analysis in order to understand the sources of variation of wages at different points of its distribution between and within genders for the two years chosen and, on the second place, the results of the MM decomposition. Section 6 discusses the results and points out caveats and areas for future research.

2 Background: gender gaps in Peru's labour market

In developing countries, female population still faces limited opportunities in educational, social and economic aspects compared to men, which results in the persistence of gender differences in labour market. Latin America is a case in point: gaps in participation and employment rates between males and females have narrowed slowly over time but still favour men. Between 2007 and 2012, the gap in participation rates narrowed from 32 percentage points to 27 and the gap in employment rates fell from 30 percentage points to 26; in both cases the change was mainly due to the increase of females rates[1] (MTPE 2014). However, there is still a strong persistence of gender differences in aspects such as decent labour, occupational segregation and incomes. Considering the first, 5.2% and 10% of employed males and females are, respectively, considered hours-related subemployed[2]. The second aspect, occupational segregation, records a relatively high level and is decreasing at a rather slow pace. Duncan Index amounted 0.373 in 2000 and 0.366 in 2010; hence, it would take 559 years to attain an equal distribution of males and females in all economic sectors[3]. The third aspect, the lower incomes of females (despite that they have a higher schooling level than males in the urban areas), arises as a consequence of the former two and can be explained by the tendency of females to engage in low-productivity and informal jobs. Actually, by 2006, more than the half of female workers in the region are located in informal jobs (50.7% in the case of females and 40.5% for males) (OIT and PNUD 2009). Moreover, a large fraction of females in Latin America are unpaid family workers (more than one third of those over 15 years in 2010) and the time invested in household and caring activities is uneven (males work 45.3 hours per week and females, 37.8) (CEPAL et al. 2013).

Peru's labour market is not outside the realm of these inequalities. Female participation rate not only remains as the highest in Latin America (66% in 2013) but also experienced a notable growth in the last forty years, going from 30% in 1970 to 40% in 1985 and 50% in 1996 (INEI 2014a; MTPE 2006) [4]. The widespread expansion of employment during 2007-2012, due to the high economic growth experienced, meant a reduction in the unemployment rates: it went from 5.3% to 4.4% for males and from 5.3% to 3.2% for females during that period (MTPE 2012). However, the increased employment of females did not guarantee, by itself, better labour conditions for these. On the one hand, 70% of working males are adequately employed by 2012 while only 61% of females are considered as such. Indeed, this 9 percentage points differential is higher than what was found in 2007 (49.8% for males and 43.1% for females). On the

[1]Between 2000 and 2010, participation rates for females went from 49,2% to 52,6% while males' fell from 80,8% to 79,6%. Employment rates, between 2002-2012, increased from 45% to 49% for females and remained around 75% for males.

[2]According to CEPAL et al. (2013), adequately employed includes those who work at least more than the full working day (35 hours per week in Peru) and earn more than the minimum wage; income-related subemployed, those who work the full working day but earn less than the minimum wage and hours-related subemployed, those who involuntarily work less than the full working day.

[3]The Duncan Index compares the males-females relation and goes from 0 (males and females have equal employment distributions in a given sector) to 1 (only females or males work in a given sector). A value of 0.37 means that 37% females in labour market should shift to a sector where they are subrepresented in order to achieve an equal sectoral distribution.

[4]Explanations focusing on the demand side state that this is due to changes in industrial composition (mainly in non traditional exports, textiles, apparel and agroindustral) and labour markets flexibilization. Explanations from the supply side suggest that improvements in education and occupational training, decreasing the total fertility rate (from 4.3 children in 1986 to 2.9 in 2000) and changes in intrahousehold decision-making are behind these changes (Jaramillo et al. 2007; MTPE 2012, 2006).

other hand, income-related subemployment differentials have widened: while around 40% of both males and females were subemployed in 2007, prevalence of subemployment in females was higher (28.9% versus 23.4%) by 2012 [5]. An explanation for this lies in the engagement of women in informal activities, which provides them with wages lower than the legal minimum. Defining informal workers as those who are not affiliated to a pension system, prevalence of formal labour in men remains higher than in females: by 2012, 39% of the latter are affiliated to such system while only 25% are covered (MTPE 2014).

Some further evidence is worthwhile to consider. On average, out of the hours devoted on a weekly basis for economic activities (75 hours in the case of males and 66 for females), Pruvian males allocate the most part to paid work (50 hours per week) while females allocate more time to unpaid-household activities related to to family caring and household tasks (36 hours) (OIT and PNUD 2009; INEI 2014b). Also, the prevalence of women is lower as dependent (wage-earner) workers and higher as self-employed, being these latter characterized for a lower productivity and higher income volatility. As 2011, on average, 52 women out of 100 fall within these categories while in males the prevalence is 39% (INEI 2014b). Likewise, there is a tendency of females to locate in small and medium-sized enterprises, being these characterized by poorer working conditions (longer working hours, physical and legal unprotection, higher degree of risk exposure, etc.) (MTPE 2012). All in all, women are over-represented in low-income employment.

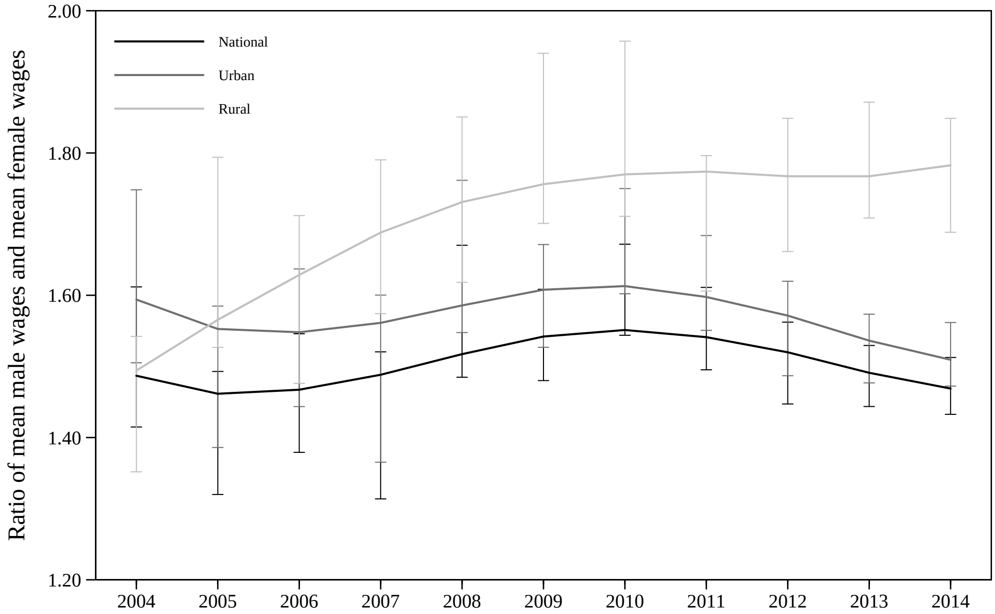
One way to synthesize these disparities is to analyse evolution of wage differentials. Focusing in the period under analysis, 2004-2014, the mean wages of men have been consistently higher than the mean wage of females, for the series depicts values statistically higher than one (Figure 1). This coincides with what is reported for other Latin American countries by CEPAL et al. (2013)[6]. Even though the ratio for the whole country has oscillated between a narrow interval (around 1.4 and 1.5), a different picture arises if we consider urban and rural areas separately. In the former case, the series show a similar fluctuation to the one found at national level but on a higher level (around 1.5 and 1.6); in the latter case, the ratio shows a sustained increase since it went from 1.5 on 2004 to around 1.8 in 2014.

Comparison of the wage distributions for the initial and ending years shows that the above-mentioned general stability of the ratio has occurred despite the right-shift of the wage distributions (Figure 2, upper panel). Put differently, the general increase of wages has not changed the fact that men earn a higher (observed) wage. More specifically, the

[5] Males also face a lower prevalence of hours-related subemployment; nevertheless, unlike the income-related subemployment, the gap has decreased between 2007 (6.4% for males versus 10.9% for females) and 2012 (3.5% for males versus 5.4% for females).

[6] Actually, this disparity holds despite that women have more years of education than males; however, for the most part they follow careers with lower returns (education, humanities, social services) in order to make it compatible with her family life MTPE (2012).

Figure 1 – Evolution of the male to female wage ratios, 2004 - 2014



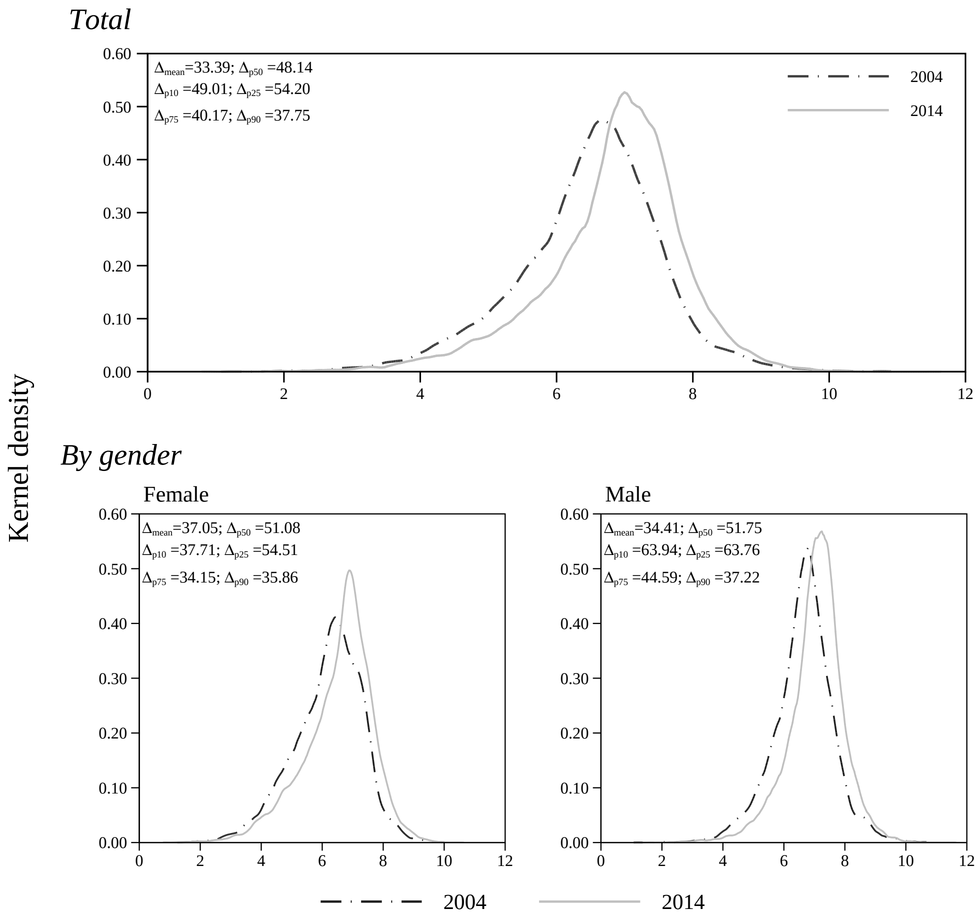
Note: Wages measured in constant 2014 Soles. Sample include individuals between 18 and 65 years. Observations weighted by expansion factor and VCE corrected according to survey's complex sample design. Local linear smooth plots shown. Vertical lines correspond to the confidence intervals at 95% of significance.
Elaborated by the author based on INEI − National Household Survey (2004−2014)

monthly average wage between 2004 and 2014 has increased in 33% (from S/. 1,095 to S/. 1,460) while the median has showed a higher growth, around 50% (increasing from S/. 721 to S/. 1,068, see Table A1 in Appendix A). Note that this shift corresponds not only to central tendency indicators but also to the lower and upper tails. While the 10th percentile increased near 49% (from S/. 161 to S/. 293), the 90th percentile increased 38% (S/. 2,074 to S/. 2,857). In the end, the Gini coefficient experienced a statistically significant reduction (at 95% of confidence), from 0.52 to 0.47.

Separation of the sample in men and females shows that the distribution changed on a similar way in both cases (Figure 2, bottom panel). Firstly, both distributions shifted to the right and the mean and percentile increased on very similar proportions (around 35% and 51%, respectively). Secondly, the top part of those distributions changed around 35%. Notwithstanding, there is a relevant difference: the bottom 10% increased in 38% in the case of females but it increased noticeably higher, around 64%, in the case of males. Indeed, the first quartile in the latter increased on an (almost) equal proportion as the first decile, whereas in the case of females the increase for this point of the distribution was higher than the bottom decile, around 55%. Accordingly, the (observed) Gini coefficient reduced (from 0.51 to 0.49) although not statistically significant, while for the case of males the observed reduction is more appreciable and significant (0.51 to 0.44) (Table A1).

Some large scale central-government-level efforts have been undertaken in the last

Figure 2 – Changes in the log-wages distribution, 2004 and 2014



Note: The variations shown are percentages calculated over the distribution of (observed) wages measured in constant 2014 Soles. Sample include individuals between 18 and 65 years; observations weighted by the expansion factor.
Elaborated by the author based on INEI − National Household Survey (2004−2014)

years to ameliorate these results. A remarkable example of these is the National Employment Policies (2012), a group of directions comprising actions of several government agencies which intends to promote employment, employability and entrepreneurship of females. Attention to labour-related gender issues was also manifested in the National Agreement (2002), a set of policies focused on improving working conditions for males and females, and two National Plans of Equality of Opportunities (2000-2005 and 2006-2010), aimed to assure decent work for women through specific instruments such as equitable labour legislation, programs to strengthen productive capacities and business management. In the end, these latter two resulted in the Law of Equality of Opportunities between Females and Males (Law N° 28983) in 2007 which, for the first time, gave prominence to the role of women as entrepreneurs and workers (rather than, as usually occurs, reproductive agents) by incorporating measures related to access to employment, training, promotion and working conditions (MTPE 2015).

Two additional legal norms stand out. Firstly, the Supreme Decree № 027-2007-PCM, signed in 2007, defined a set of mandatory compliance guidelines within all government institutions, among which lies a policy of non-discrimination as well as technical norms for the formulation and management of gender-equality-related public policies. Secondly, the Law 294098, signed in 2009 and devised similar as in other countries Latin America, seeks to strengthen the equitable distribution of housework tasks in case of the birth of a new child by providing both parents the right of paid parental leave (MTPE 210). Despite that the number of permitted days is uneven (4 days for the father and 3 months for the mother) and the scope of its implementation is unclear, this is the first effort to provide a relief in women's childcare burden, provided that their labour market participation is strongly influenced by decisions at household level (ILO 2012).

Although this preliminary evidence would lead us to conclude that the gender wage gap have remained steady between 2004 and 2014 despite the increase in wages during that period, it does not allow to take into consideration differences in relevant characteristics correlated with wages (such as age, education, area of residence, etc.). These are necessary to net out in order to provide a compelling estimate of the wage gap. Also, evaluation of the ratios themselves cannot lead us to state if there is a prevalence of glass ceiling or sticky floors against women for they only consider the mean wages; let alone the wages distribution depicted since they do not provide information on marginal returns of different characteristics. Furthermore, it is necessary to dig deeper on the observed labour wages differences between men and females. In other words, it is essential to assess if the gap is due to the fact that characteristics of males and females are notably different or to the fact that labour market returns are very different through the whole distribution. If this latter is the case, then the above mentioned government-induced policies might have not fully accomplish their objective of fostering a lower inter-gender discrimination. Consequently, the next section presents the analytical framework which will be applied in the research in order to get across these concerns.

3 Analytical framework

This section presents the methods which will be applied in this research. First, we show the quantile regression approach and how it can be extended in order to account for sample selection; second, we present the MM decomposition and undertake a brief comparison of this with other similar approaches, showing its strengths and weaknesses. Finally, we briefly review and assess key studies which apply this decomposition to gender gaps in the labour market.

3.1 Quantile regression with sample selection correction

The basic regression analysis in this study corresponds to quantile regression. This approach, as developed by Koenker and Bassett (1978), presents two compelling attributes that makes it more attractive and flexible method than the well known OLS approach. On the first place, it conveys a more complete picture of the effects of the covariates along the conditional distribution of variable of interest. While OLS provides a unique vector of estimates which represents only mean effect of the explanatory variables (*ceteris paribus*) on a response variable, quantile regression provides a vector of estimates which represents the effect of the explanatory variables over a given percentile of the conditional distribution of the continuous dependent variable. Put differently, while the former approach assumes homogeneity of the conditional distribution (such that regardless of the location of the observation, the marginal effect of a covariate is the same), the latter relaxes this assumption and provides a collection of vectors for each point of the distribution (say, the first decile, third quartile, the top decile). For our purposes, this heterogeneity assumption makes possible to uncover the presence of glass ceilings and sticky floors with respect to the conditional wage distribution and gender.

On the second place, quantile regression allows for heteroskedastic errors, since it relies on a more general dependence of the conditional distribution of the dependent variable on the covariates (instead of just the variance of the conditional mean). This is of special interest in the current application since functions estimated from survey data are typically not homoskedastic even when the variance of individual behaviour is constant within strata, since heterogeneity between clusters generates heteroskedasticity in the overall function. In this case, OLS is inefficient and the usual formulas for standard errors are incorrect. In the context of quantile regression, if the change of variance of the error is linked to the value of the covariants (with the distribution of residuals changing its dispersion as the values of the independent variables becomes larger), then quantile regressions for the percentiles (other than the median) will no longer be parallel to the regression line, but will diverge for larger values of the covariates (Deaton 1997).

Formally, we can define the θth quantile of the conditional distribution of log wage (w) given the vector of covariates $\mathbf{x}_i$ as

$$Q_\theta (w|\mathbf{x}) = \mathbf{x}_i^{'}\boldsymbol{\beta}(\theta) \text{ for } \theta \in (0,1) \tag{1}$$

where $\boldsymbol{\beta}(\theta)$ denotes a vector of parameters for the given quantile θ. This vector can be estimated by solving the linear programming problem given by

$$\min_{\boldsymbol{\beta}(\theta)} \left[\sum_{w_i > \mathbf{x}_i \boldsymbol{\beta}(\theta)} (\theta) \left| w_i - \mathbf{x}_i^{'}\boldsymbol{\beta} \right| + \sum_{w_i < \mathbf{x}_i \boldsymbol{\beta}(\theta)} (1-\theta) \left| w_i - \mathbf{x}_i^{'}\boldsymbol{\beta} \right| \right]$$

which corresponds to the minimization of an asymmetric loss function (giving differing weights to positive and negative residuals). However, in the case of median regression

($\theta = 0.5$), the expression inside brackets collapses to $\sum_{i=1}^{N}\left|w_i - \mathbf{x}_i^{'}\boldsymbol{\beta}\right|$ such that, at the optimum, the number of positive and negative residuals must be the same (Koenker and Hallock, 2001). If the equation 1 is correctly specified, the conditional quantile process provides a full characterization of the conditional distribution of wages given $\mathbf{x}$ in the same way as quantiles characterize a marginal distribution (Autor et al. 2005; Machado and Mata 2005). Indeed, under weak regularity conditions, the estimated conditional quantile function is a consistent estimator of the population quantile function (Bassett and Koenker 1986)[7].

The model estimated in this study is traced back to the seminal paper by Gronau (1973), who stated that selectivity bias induced by the job search process affects participation ratios between males and females given that the latter face a the lower wage-offer distribution and a higher value of time in absence of market opportunities. This implies that, in the context of quantile regression, the conditional quantiles depend also on a term of an unknown form which cannot be approached using the traditional parametric correction for sample selection. Consequently, we resort to a non-parametric method of sample selection correction originally proposed by Buchinsky (1998, 2001) which, in turn, mimics the correction on the mean regression case devised by Heckman (1979). At this point, it is useful to present the two equations of the wage model as presented in Buchinsky (1998).

On the first place, the outcome equation (wage offer) is given by

$$w_i^{*} = g^{*}\left(\mathbf{x}_i, \boldsymbol{\beta}\right) + u_i \tag{2}$$

where $\mathbf{x}_i$ is a $l \times 1$ vector of labour market characteristics for individual i, $\boldsymbol{\beta}$ is a vector of parameters and g^{*} is a general function which, for simplicity, is assumed to be $g^{*}\left(\mathbf{x}_i, \boldsymbol{\beta}\right) \equiv \mathbf{x}_i^{'}\boldsymbol{\beta}$. Further, we assume that $Median\left(u|\mathbf{x}\right) = 0$.

On the second place, the participation equation (reservation wage) is given by

$$w_i^{R} = g^{R}\left(\mathbf{z}_i, \boldsymbol{\alpha}\right) + v_i \tag{3}$$

where z_i is a vector of individual characteristics for individual i which impact his decision to work, $\boldsymbol{\alpha}$ is a vector of parameters and g^{R} is a function which, again, is assumed to be $g^{R}\left(\mathbf{z}_i, \boldsymbol{\alpha}\right) \equiv \mathbf{z}_i^{'}\boldsymbol{\alpha}$. Similar to the error term in the wage offer equation, it is assumed that

$$E\left(v|\mathbf{z}\right) = Median\left(v|\mathbf{z}\right) = 0$$

Note that, unlike the typical formulation of the error in the context of mean regression, we assume not only that the mean equals zero but also the median. In order to identify the parameters, it must be the case that $\mathbf{x}$ must be a subset of $\mathbf{z}$.

[7] While the conditional quantile functions are non-decreasing on the interval $(0,1)$ when the covariates are evaluated at sample mean (Bassett and Koenker 1982), this property need not hold for other values of the covariates and lack of monotonicity can ensue. However, given the consistency of the estimated conditional quantile function, it must necessarily be the case that, for any two values θ and $\theta^{'} > \theta$, the empirical quantile functions satisfy $\hat{Q}_{\theta}\left(w|\mathbf{x}\right) < \hat{Q}_{\theta'}\left(w|\mathbf{x}\right)$ for a sufficiently large sample size.

Given the quantile regression context, we can rewrite the previous set of equations as

$$w_i^* = \mathbf{x}_i^{'}\beta(\theta) + u_i \tag{4a}$$
$$w_i^R = \mathbf{z}_i^{'}\alpha(\theta) + v_i \tag{4b}$$

for $\theta \in (0,1)$. Focusing only on the wage offer, it is *assumed* that the conditional quantile of w^* conditional on $\mathbf{x}$ satisfies $Q_\theta(w|\mathbf{x}) = \mathbf{x}^{'}\beta(\theta)$ and hence

$$Q_\theta(u|\mathbf{x}) = 0$$

However, the wage offer is observed only if the individual accepts to work which, in turn, occurs if it exceeds the reservation wage, i.e. if $w_i^* > w_i^R \Leftrightarrow w_i^* - w_i^R > 0$. Then, we can express the observed wage as

$$w = I\left(w^* - w^R > 0\right)w^* = I\left(w^* - w^R > 0\right)\left(\mathbf{x}^{'}\beta(\theta) + u\right)$$

where $I(.)$ is the usual indicator function. Consequently, in the presence of this selection mechanism, the conditional quantile of the observed wage is given by

$$Q_\theta(w|\mathbf{x}) = Q_\theta(w^*|\mathbf{x}, I(.) = 1) = Q_\theta\left(\mathbf{x}^{'}\beta(\theta) + u|\mathbf{x}, I(.) = 1\right) = \mathbf{x}^{'}\beta(\theta) + Q_\theta(u|\mathbf{x}, I(.) = 1)$$

and so, in general, we *cannot* assume that $Q_\theta(u|\mathbf{x}, I(.) = 1) = 0$. Nevertheless, we can circumvent this problem by noting that in the latter expression we can define

$$Q_\theta(u|\mathbf{x}, I(.) = 1) \equiv h(\theta)(\mathbf{z}, \gamma) + \varepsilon$$

If $h(\theta)(\mathbf{z},\gamma)$ is only a function of an index $g_0 = g(\mathbf{z},\gamma)$, then $h(\theta)(\mathbf{z},\gamma) = h(\theta)(g_0)$. Further, under simplifying assumptions [8], the observed wage equation can be written as

$$w = \mathbf{x}^{'}\beta(\theta) + \left[h(\theta)\left(\mathbf{z}^{'}\gamma\right) + \varepsilon\right] \tag{5}$$

where, by definition, $Q_\theta(\varepsilon|\mathbf{x}, I(.) = 1) = 0$. The problem for correcting sample selection implies estimation of the function $\left[h(\theta)\left(\mathbf{z}^{'}\gamma\right) + \varepsilon\right]$ which is unknown[9].

[8] Buchinsky (1998, 2001) states two assumptions, continuity ($w = (u, v)$ has a continuous density) and dependence of w and $\mathbf{z}$ ($f_w(.|\mathbf{z}) = f_w(.|g(\mathbf{z}, \alpha_0))$ where $\alpha_0 = \tilde{\beta}_0 - \gamma_0$ and $\tilde{\beta}_0$ equals $\beta_0$with entries of zeros added in places where the variables in $\mathbf{z}$ do not appear in $\mathbf{x}$) to ensure that equation $Q_\theta(u|\mathbf{x}, I(.) = 1) \equiv h(\theta)(\mathbf{z},\gamma) + \varepsilon$ holds and that $h(.)$ is a continuous and increasing function of g. Even though Albrecht et al. (2009) point out that these assumptions could be deemed as controversial (since it seems difficult to specify a data generating process that conforms exactly to these two assumption), the objective is to allow for a selection effect that varies across quantiles and, so, $h(\theta)(\mathbf{z},\gamma)$ is an approximation to attain this aim.

[9] According to Albrecht et al. (2009), if we could regress the reservation wage on the observables, that would give a consistent estimate of γ. However, we only observe whether the difference between the market wage and the reservation wage is positive, i.e. if $I\left(w^* - w^R > 0\right)$ equals 0 or 1.

In the same vein as Newey (1988), we can estimate the single index term $h(\theta)\left(\mathbf{z}'\gamma\right)$ by a non parametric approximation on a two-step method. The first step consists on the estimation of the selection parameter $\hat{\gamma}$ by a Semi-parametric Least Squares (SLS) estimator suggested by Ichimura (1993) which makes no assumption about the particular parametric distribution of the selection equation error term. However, for this study, we choose to choose a probit estimator because of the high computational requirements that implies adopting the SLS estimation.

The second step consists on the estimation of the parameters of the wage offer equation $\beta(\theta)$ including as a right hand side variable the estimate of $h(\theta)\left(\mathbf{z}'\gamma\right)$, $\widehat{h}(\theta)\left(\mathbf{z}'\gamma\right)$, which controls for sample selection at the θth quantile and plays the same role as the Inverse Mills Ratio in the Heckman (1979) procedure. Given the assumptions of continuity and dependence of w and $\mathbf{z}$, $h(\theta)\left(\mathbf{z}'\gamma\right)$ can be approximated arbitrarily close by

$$\widehat{h}(\theta)\left(\mathbf{z}'\gamma\right) \equiv \widehat{\delta}'(\theta)\mathbf{P}_S\left(\mathbf{z}'\hat{\gamma}\right) \tag{6}$$

where $\mathbf{P}_S\left(\mathbf{z}'\hat{\gamma}\right) = \left(P_{S1}\left(\mathbf{z}'\hat{\gamma}\right), P_{S2}\left(\mathbf{z}'\hat{\gamma}\right), \ldots, P_{SS}\left(\mathbf{z}'\hat{\gamma}\right)\right)'$ is a polynomial vector of order S. For appropriate values of $\mathbf{P}_S$, $\widehat{h}(\theta)\left(\mathbf{z}'\gamma\right) \to h(\theta)\left(\mathbf{z}'\gamma\right)$ as the number of terms goes to infinity. Then, the vector of interest, $\hat{\beta}(\theta)$, is obtained from a quantile regression of y on x

$$Q_\theta(w|\mathbf{x},\mathbf{z}) = \mathbf{x}'\hat{\beta}(\theta) + \hat{\delta}'(\theta)\mathbf{P}_S\left(\mathbf{z}'\hat{\gamma}\right) \text{ for } \theta \in (0,1) \tag{7}$$

Three observations are in order. First, the parameters $\hat{\gamma}$ need to be estimated only once since the probit estimates, unlike those of quantile regression, result on a unique set for the sample. Second, we can only estimate $\hat{\beta}(\theta)$using observations of those who actually work, i.e. those where $I\left(w^* - w^R > 0\right) = 1$. Third, the last expression does not define the form of $\mathbf{P}_S\left(\mathbf{z}_i'\hat{\gamma}\right)$ and so several power series can be considered[10]. Buchinsky (2001) and Albrecht et al. (2009) consider the particular expression

$$\widehat{\delta}'(\theta)\mathbf{P}_S\left(\mathbf{z}'\hat{\gamma}\right) = \delta_0(\theta) + \delta_1(\theta)\lambda\left(\mathbf{z}'\hat{\gamma}\right) + \delta_2(\theta)\lambda\left(\mathbf{z}'\hat{\gamma}\right)^2 + \ldots$$

where $\lambda(.) = \frac{\phi(.)}{\Phi(.)}$ the Inverse Mills Ratio with $\phi(.)$ and $\Phi(.)$ the PDF and CDF, respectively, of a standard normal variable. The number of terms of the polynomials to be

[10] Albrecht et al. (2009) states that any function of $\mathbf{z}'\gamma$ can be used. Buchinsky (1998) takes three alternative expressions to the one used here. Let $\Phi(.)$ be the cumulative distribution function of a standard normal variable, $f(.)$ a non-parametric estimate of the probability density function of ε and $\hat{F}(.)$ a non-parametric estimate of the cumulative density function of ε; he considers, $\mathbf{P}_{Sj}\left(\mathbf{z}_i'\hat{\gamma}\right) = \left(1 - \Phi\left(\hat{\mu} + \hat{\sigma}\left(\mathbf{z}_i'\hat{\gamma}\right)\right)\right)^{j-1}$, $\mathbf{P}_{Sj}\left(\mathbf{z}_i'\hat{\gamma}\right) = \left(\mathbf{z}_i'\hat{\gamma}\right)^{j-1}$ and $\mathbf{P}_{Sj}\left(\mathbf{z}_i'\hat{\gamma}\right) = \left(\frac{\hat{f}\left(\mathbf{z}_i'\hat{\gamma}\right)}{\hat{F}\left(\mathbf{z}_i'\hat{\gamma}\right)}\right)^{j-1}$.

included in the regression, as in the case of Buchinsky (1998, 2001); Albrecht et al. (2009), is two[11].

An additional problem of the equation 7 lies on the fact that it is not possible, without additional assumptions, to consistently estimate the constant term in $\beta(\theta)$, $\beta_0(\theta)$, separately from the constant term in the polynomial $\mathbf{P}_S\left(\mathbf{z}_i^{'}\hat{\gamma}\right)\hat{\delta}(\theta)$, $\hat{\delta}_0(\theta)$, since we can define $P_{S1}(\hat{g}) \equiv 1$. We can follow the solution outlined in Albrecht et al. (2009), Buchinsky (1998) and Andrews and Schafgans (1998), where $\beta_0(\theta)$ can be estimated through an identification at infinity approach, i.e. from a subsample of observations with values of the observables such that the probability of working given those values is close to one. However, provided that we have used a method (probit) which is a substitute for the most correct method to correct for sample selection, results presented here do not include this identification[12].

3.2 Conditional wage distribution decomposition

Most of the literature that studies the gender wage gap consists on OB type decompositions (Oaxaca 1973). Put succinctly, this allows to decompose differences in wages between 2 groups. Let the wages (for our case, males and females) be expressed as

$$w_s = \mathbf{x}_s^{'}\beta_s + \varepsilon_s \text{ for } s = M, F$$

where w are the observed wages, $\mathbf{x}$ is a vector of covariates and β is a vector of parameters. Under reasonable regularity conditions[13], the gap between the *mean* wages across the two groups can be expressed, after simple grouping, as the sum of the difference between the actual wage of each group and a *counterfactual term*:

$$\overline{w}_M - \overline{w}_F = \overline{\mathbf{x}}_M^{'}\hat{\beta}_M - \left[\overline{\mathbf{x}}_M^{'}\hat{\beta}_F - \overline{\mathbf{x}}_M^{'}\hat{\beta}_F\right] - \overline{\mathbf{x}}_F^{'}\hat{\beta}_F = \underbrace{\overline{\mathbf{x}}_M^{'}\left(\beta_M - \beta_F\right)}_{\text{Returns effect}} + \underbrace{\left(\overline{\mathbf{x}}_M^{'} - \overline{\mathbf{x}}_F^{'}\right)\hat{\beta}_F}_{\text{Covariates effect}} \tag{8}$$

[11] Addition of more terms to the series expansion usually change little the results and controls for departures to the normality assumption of the errors, which is inherent in the Probit regression. However, as in the typical OLS framework, this generates severe multicollinearity problems.

[12] This is due to implausibility of the results found which can be attributed to the fact that the constant estimated from the subsample found after running a probit and retaining those individuals with high probability of working under a Probit differs from that sample found under a SLS model. Hence, results correcting for selection should be interpreted with caution.

[13] According to Firpo et al. (2007), these are: mutually exclusive groups; outcomes defined according to a definite structural form (worker i belonging to either group is paid according to the wage structure which are functions of the workers observable, $\mathbf{x}_s^{'}$, and unobservable, ε_s, characteristics); feasibility of a simple counterfactual treatment (counterfactuals can be constructed based on the alternative wage structure, i.e. using the observed wage structure of A as a counterfactual for B); existence of an overlapping support (the effect of manipulations of the distribution of observables $\mathbf{x}_s$ will not be confounded by changes in the distribution of ε_s); invariance of conditional distributions (construction of the counterfactual for B workers that would have prevailed if they were paid like A workers assumes that the conditional wage distribution apply or can be extrapolated).

The first term on the right hand side, the returns effect, measures the contribution of the difference in returns to the male-female gap. The second term, the covariates effect, measures the contribution of the differences in values of the covariates to the gap. In cases where group membership is linked to some immutable characteristics of the workers, such as gender, the return effect has also been called the unexplained part of the wage differentials or the part due to discrimination. Two key points must be highlighted from this latter expression. On the hand, OB implies the creation of an (arbitrarily chosen) counterfactual mean wage, given in this case for $\overline{\mathbf{x}}_M^{'}\hat{\beta}_F$; this can be interpreted as the average wage predicted for the males if they were paid according to the labour market schedule for females but retain their own characteristics. On the other hand, since the marginal returns to characteristics are estimated by a regression to the mean, OB only decomposes *mean* differences. Nevertheless, the size of the gap varies at different points of the distribution: in some countries women face glass ceilings while in others women stand on sticky floors. Additionally, note that decomposition methods inherently follow a partial equilibrium approach.

Studies for Peru have, almost exclusively, relied on this approach to decompose the gender wage gap and some of them are worth to consider. Firstly, Montes (2007) analyses the period 1997-2000 for the urban areas and finds that there was a small but significant wage discrimination in favour of women in 1997, which disappeared by 2000. Castillo (2011) studies the 2003-2009 period for the whole country (using a different dataset than the prior one) and finds that, taking alternative specifications, the returns effect explains most of the gap in Peru: it represents as high as half of the 22% gap between men and women for that period. This is in line with what MTPE (2014) reports: out of the 33% gender gap in 2012, 4% is due to covariate effect and the remaining 29% is due to discrimination effect. Indeed, this result also holds for year 2007 (inter-gender income differential was 28% and that part due to discrimination was 4%). Additionally, Yamada et al. (2013), for 2010, confirms the fact that men have a higher return to cognitive abilities which generates an increase in the gender wage gap. Finally, the survey for Latin American 18 countries, undertaken by Atal et al. (2009), reveals that out of the 18% gap, discrimination against women accounts for nearly 20%; the covariate effect, favouring females, counter this tendency. The magnitude of discrimination factor is the fifth highest among the countries analysed.

Provided the limited understanding of the gender gap OB approach provides, few methods have been developed to extend this decomposition to distributional parameters other than the mean. One of the earliest attempts to generalize the analysis of gaps in the *entire* density of wages corresponds to DiNardo et al. (1996) (DFL) decomposition. By viewing each wage observation as a vector composed of the wage itself, a set of individual (t_x) and unobservable attributes (m), they propose a semi-parametric procedure that is innovative in two aspects. The first, they rely on weighted-kernel densities to estimate the counterfactuals, rendering a visual representation of the impact of explanatory variables. Specifically, if $f(w;t_w = A;t_x = A;m_A)$ refers to the observed distribution of wages for group A with the distribution of attributes as in A and returns and unobservable attributes as in A, the counterfactual density $f(w;t_w = A;t_x = B;m_A)$ (wages for A with the distribution of attributes as in B but returns as A) can be written as $\int f(w|z,t_w = A;m_A)\Psi(z)dF(Z|t_z = A)$, where the reweighting function $\Psi(z)$ maps the group B distribution of covariates into A's. Put simpler, they start with group A and then replace the distribution of covariates of this group with the distribution of covariates of group B using the reweighting function, which can be estimated by a kernel function. The second innovation, in the context of their original proposal, is that they not only analyse the impact of individual characteristics but also impact of variables which are *not* priced by the market but instead spillover across the distribution (such as unionization and minimum wages). This approach, despite of its flexibility and elegance, suffers from a capital problem: since it relies on a non-parametric method, it prevents the estimation of population parameters and, hence, it is unable to separate that part attributable to returns effects of that from covariates effect[14].

A more recent approach uses the *Recentered Influence Function* (RIF) regressions proposed by Firpo et al. (2007). These are defined as $RIF(w;v) = v(F_w) + IF(w;v)$, where $IF(.)$ is an influence function corresponding to observed wage w for the distributional statistic of interest $v(F_w)$; in its simplest form, the conditional expectation of $RIF(w;v)$ can be modelled as a linear function of the explanatory variables estimable by OLS. The idea is to use the RIF for the distribution statistic of interest (since $\int RIF(w;v)dF(w) = v(F_w)$) instead of the usual outcome variable w as the left hand side variable in a regression. A primary advantage is that, unlike distributional decomposition techniques (including MM), the estimated coefficients of the RIF regression can be used to perform the detailed decomposition in the same way as in the standard OB decomposition. A second advantage is that decompositions of quantiles can then be obtained by inverting back proportions into quantiles by using a simple first order approximation. The downside of this approach is that RIF regression coefficients only provide a local approximation for the effect of changes in the distribution of a covariate on the statistic of interest, which may could produce approximation errors. In the case of wage distribution, characterized by humps at lower parts of the distribution, this approximation may be imprecise (Fortin et al. 2010). A second limitation is that this method is based on the estimation of unconditional quantile regressions in the presence of exogenous covariates and does not consider the possible presence of endogeneity (Salardi 2012).

The Machado and Mata (2005) decomposition (MM) constitutes an alternative method. It consist on estimating models for the quantiles of the conditional wage distribution to estimate counterfactual densities consistent with the conditional model and covariate

[14] Fortin et al. 2010 also state a further problem: in the program evaluation literature, reweighting can have some undesirable properties in small samples when there is a problem of common support.

densities. Thus, estimation of a grid of parameters for different quantiles allows to address some of the inherent problems of the approaches mentioned. On the one hand, as shown by Autor et al. (2005), it nests the kernel reweighting proposed by DFL and corrects the shortcomings of an alternative full distribution accounting method introduced by Juhn et al. (1993) (JMP) [15]. Unlike JMP, MM does not rely solely the conditional mean of the wage distribution to characterize the whole distribution but instead models each quantile based on a conditional distribution (Autor et al. 2005). On the other hand, it allows dealing with the sample selection problem, which tends to be a pervasive problem in the study of labour market outcomes and about which DFL and RIF methods remain silent. Since we use quantile regressions to estimate the counterfactual distribution, we can resort to the non-parametric sample selection correction presented in the previous subsection with the modification of the original MM method proposed by Albrecht et al. (2009). This will provide a more precise and theoretically correct estimation of the contribution of endowments and returns along the distribution of wages.

However, it is important to acknowledge that this approach is not free of drawbacks. One particular problem is that linear specification can be restrictive and finding the correct functional form for the conditional quantile regressions can be tedious (Fortin et al. 2010). Another point to consider is that under MM we can compute sub-components of the returns effect but not those of the covariates effect[16]. Nevertheless, our interest here lies on assessing the total bulk of discrimination against women independently of its causes, considering that the decomposition is not devised to recover behavioural relationships or structural parameters (Fortin et al. 2010). Therefore, given that the MM decomposition constitutes a more complete approach, we now show how to implement it.

The first step is to build the counterfactual distribution, which involves estimating the marginal density function of wages. Despite that we can estimate a marginal wage density directly from the data on wages, this would not necessarily correspond to the conditional distribution modelled in equation 7, because the expected value of the conditional expectation does not equals the unconditional expectation (i.e. the iterated expectations property does not hold). Notwithstanding, we can simulate a sample from the estimated conditional distribution by resorting on the *Probability Integral Transformation* theorem: if U is a uniform random variable defined on $[0,1]$, then $F^{-1}(U)$

[15] This method decomposes changes in the wage distribution between two years into three components: changes in returns ($\Delta\beta$), changes in quantities ($\Delta g(\mathbf{x})$) and changes in the residual distribution. Hence, they model wage inequality as $w_{it} = x_{it}^{'}\hat{\beta}_t + F_t(\theta_{it})$, i.e. a function of the distribution of covariates ($g(\mathbf{x})$), the vector of between-group prices (β) and the cumulative distribution of the residual, $F(\theta)$ where $\theta_{it} = F_t^{-1}(\varepsilon_{it}) \in (0,1)$.

[16] According to Fortin et al. 2010, MM suggest an unconditional reweighting approach to do so but it does not provide a consistent effect since the effect of the reweighted covariate of interest gets confounded by other covariates correlated with that same covariate.

has density F. Then, if θ_1, ..., θ_m are drawn from $U(0,1)$ distribution, the corresponding m estimates of the quantiles of wages conditional on $\mathbf{x}$, $w \equiv \left\{\mathbf{x}'\hat{\beta}(\theta_j)\right\}_{j=1}^{m}$, constitutes a random sample from the estimated distribution of wages conditional on $\mathbf{x}$.

Based on this, we can generate two counterfactual densities which constitute random samples from the wage density: one that would prevail if females were rewarded according to their characteristics but taking the male distribution of covariates, and one that would prevail if women retained their own labour market characteristics but were paid like men. Applied to the gender gap analysis, the original MM procedure for generating the first counterfactual distribution can be expressed as:

1. Sample θ from $U[0,1]$
2. For the data set of females estimate $Q_\theta(w|\mathbf{x}_F)$ and save the vector of estimates $\hat{\beta}_F(\theta)$
3. Generate a random sample with replacement from the empirical distribution of $\mathbf{x}_M$
4. Compute the linear combination $w_{MF} = \mathbf{x}_M\hat{\beta}_F(\theta)$
5. Repeat steps 1 to 4 M times

Even though the resulting sample is based on estimates rather than on the true parameters, the quantiles computed converge to population quantiles of w_{MF} as the number of observations for the sample of males and females becomes large[17]. In order to estimate the second counterfactual density, w_{FM}, we must reverse the roles of male and females in steps 2 (estimate from the data of males) and 3 (generate the sample from $\mathbf{x}_F$).

Nevertheless, unlike the original MM procedure, we are correcting for sample selection in the quantile regression context (as outlined in the previous subsection). Hence, we can rely on the extension devised by Albrecht et al. (2009)

1. Estimate γ using a single index method such as SLS
2. Sample θ from $U[0,1]$
3. For the data set of females estimate $Q_\theta(w|\mathbf{x}_F,\mathbf{z})$ for each $j = 1,...,m$ by applying equation 7 and save the vector of estimates $\hat{\beta}_F(\theta)$
4. Generate a random sample with replacement from the empirical distribution of $\mathbf{x}_M$ taking the observations of only those who work

[17] Albrecht et al. (2009) proves that, under reasonable assumptions, the qth quantile estimated under the MM procedure, $\hat{\rho}(q)$, is a consistent estimator of the qth quantile of the unconditional distribution of w_{MF}, $\rho(q)$. Furthermore, $\sqrt{M}(\hat{\rho}(q)-\rho(q))$ is asymptotically normal.

5. Compute the linear combination $w_{MF} = \mathbf{x}_M \hat{\beta}_F(\theta)$

6. Repeat steps 2 to 5 M times

Similar as Autor et al. (2005); Pham and Reilly (2007); Albrecht et al. (2003); Nguyen et al. (2007); Aktas and Uysal (2012); Rica et al. (2008), we adopt in this study a variant of these procedures on the grounds of computational feasibility. Instead of sampling θ from a standard uniform distribution, we estimate $\hat{\beta}_F(\theta_j)$ for a grid of θs (in intervals of one centile beginning with 0.01), then repeat the steps 4 and 5 (taking 1,000 elements of x_M) for each value of θ and stack them into a vector of 99,000 ($= 99 \times 1000$) elements. This eliminates the sampling error that is inherent in the step 2 and, in practice, yields the same estimates as the original MM procedure (Albrecht et al. 2009).

Denote by $f(w_s)$ an estimator of the marginal density of w for s based on the observed sample and by $f^*(w_{MF})$ an estimator of the counterfactual density. Given the linearity of quantile regression, the differences between the distributions for men and females for a given percentile $\rho(.)$ can be decomposed without a residual[18] as

$$\rho(f(w_M)) - \rho(f(w_F)) = \underbrace{[\rho(f(w_M)) - \rho(f^*(w_{MF}))]}_{\text{Returns effect}} + \underbrace{[\rho(f^*(w_{MF})) - \rho(f(w_F))]}_{\text{Covariates effect}} \quad (9)$$

In other words, we can separate the overall difference for a given percentile in that part attributable to the discrimination and that part attributable to the different endowments for any point we are interested. Finally, note that taking the second counterfactual, w_{FM}, instead of w_{MF} provides different contributions of the two terms of the decomposition. Consequently, as a robustness check, we will reverse the order of the decomposition to make sure that the results are resilient to counterfactual decomposition chosen.

3.3 Applications of the MM decomposition

Most of the studies which apply the MM method to unveil the sources behind labour market gaps have been applied to regions other than Latin America. In the case of Europe, Arulampalam et al. (2007) analyses 11 countries using data covering the 1995-2001 period. They find that the unconditional gender gap across the wage distribution varies considerably across selected European Union countries: moving up the distribution, it decreases in some countries (Ireland, Italy and Spain) and increases in others

[18] Originally, they also provide the procedure for generating a marginal density implied by the model (a random sample that arises for sex s if the model were true and the covariates were distributed as those for sex s). Since this is density is an approximation to the observed one, it equals $f(w_s)$ plus an error. For the sake of simplicity, we chose to include the observed densities on the right hand side of expression 9 instead of those implied by the model so that the exact decomposition holds.

(Finland and Denmark). When considering those employed in the public and in the private sector separately, measures of the raw gap exhibit a similar pattern, such that in some countries there is a hint of a glass ceilings while in others there are indications of sticky floors. Decomposition results suggest that differences in returns are noticeable and even higher than the observed gap itself. Nevertheless, the authors acknowledge that limitations of the dataset prevent them from addressing the issue of self selection.

Some other studies have focused on particular European countries. Albrecht et al. (2003) shows that in Sweden the raw gender gap in 1998 increases throughout the distribution, notably in the upper tail; this suggests a strong glass ceiling effect women. Indeed, they find that this holds only when comparing males and females (and not, for example, when comparing natives and migrants). MM decomposition results indicate that returns effect accounts for the most part of the gap; adding covariants to the basic specification results on this effect being the dominant at the bottom of the distribution. Like Arulampalam et al. (2007), they do not control for non-random selection. In contrast, two other studies turned their attention to this problem. First, Rica et al. (2008), using 1999 data for Spain, find that the gap increases as we move up the distribution only for those with college and tertiary education while for those with lower education the gap is lower at the top. Conditioning on a set of covariates, tenure and secondary education yields a higher return for females at the lower quantiles than males. Correcting for selection makes more acute the reduction in the gap for those less educated. Decomposition results suggest that differences in observed characteristic explain about one half of the gap at the top of the distribution and discrimination seems to be an important factor driving the gap at the bottom. Second, Albrecht et al. 2009 study the gap across the distribution for men and women who work full time in the Netherlands, being the selection in this case related to the decision of working full time versus partial time. After correcting for selection and for gender differences in the distribution of observed characteristics, they find a significant positive gap across the entire distribution, being higher at the highest quantiles. Decomposition results suggest that most of the gender gap across the distribution is accounted for by the effects of returns, however one third of the counterfactual difference is due to differences in covariates.

Few studies have undertaken this decomposition for Asia. Aktas and Uysal (2012) analyse the Turkish labour market in 2006. While OLS shows a 3% unconditional gap, quantile regression results show that there is no raw gap at the lower end of the wages distribution. However, a different outcome emerges at other points of the distribution: while the median men earn around 6.5% more than females, a female at the top earns 5% higher wages than men. Including additional control variables to the basic Mincer framework results on a reversal of the (conditional) gap at the top of the distribution: females now earn 3% lower wages than men. Additionally, quantile regression indicate that returns to labour market characteristics differ for males and females. Application of the decomposition indicates that most of the gap stems from differences in returns to labour market characteristics. As the authors recognize, they do not provide any correction on selection. In turn, Pham and Reilly (2007), using data from a Vietnam household survey for 2002, turn their attention to the ethnic wage gap (Kihns -the majority of the country- and other ethnic minority groups). Results confirm existence of wage inequality between the majority and minority ethnic groups. Particularly, the latter group secures lower returns in the labour market for their endowments than the former. Decomposition results reveal that the ethnic wage gap can be attributed mainly to the effect of returns at most of the quantiles of the conditional distribution. Differences in selection effects between the two groups are negligible.

To our best knowledge, only one study has applied the MM approach to analyse the gender wage gaps to a Latin America country. Salardi (2012) analyses Brazil for 1986 and 2006 and finds that, conditioning on several variables, males in 1987 have a greater advantage on wages at the bottom and at the top of the distribution; by 2006 the U shaped pattern has disappeared although their advantage remains. Decomposition results suggest that in both years wage gaps were attributable to the effects of returns mainly at the extremes of the wage distribution (which reflects gender based discrimination in the labour market), although these component has declined in time. Two other recent researches, which applied different distributional decomposition methods, stand out. On the one hand, Pacheco (2013) uses the RIF regression method for urban Nicaragua in 2005 and 2009. Application of decomposition for mean wages shows that discrimination accounted for a large share of the differences; when applying the decomposition across the wage distribution, returns effects have a different effect on the wage gap across the distribution while in 2009 this effect has been reduced at the lower and upper part. On the other hand, Arceo-Gomez and Campos-Vazquez (2014) apply DFL method and, unlike the previous two studies, they do correct or selection. They focus on Mexico during the 1990-2010 period and find that the mean gap in 1990 was around 0.4% but increased in 2010 to 6%. However, this average estimate hides the sticky floor and glass ceiling patterns over the period. DFL decomposition reveals that most of the wage gap is due to the returns effect and, when correcting for selection, they find that there is positive selection of females into labour market.

Taking these studies into consideration, the current research represents an opportunity to apply the MM decomposition for a Latin American country but considering two aspects that, in most cases, have been absent: correction for sample selection and comparison of the components of the decomposition in time. Before turning to the results, we discuss in the next section the characteristics of dataset and some descriptives.

4 Data

The dataset used in this study corresponds to the National Household Survey (ENAHO, according to its initials in Spanish) collected by the National Institute of Statistics and

Informatics (INEI, according to its initials in Spanish). This survey takes as study population the set of all private dwellings and its residents in the urban and rural area. Three reasons underlie the choice of this dataset. Firstly, this constitutes Peru's main primary source of information for elaboration of official indicators on living conditions, poverty and employment, since it covers a wide range of dimensions (demographics, education, health, labour, household production, etc.) for the whole national population. Secondly, it allows to obtain comparable estimates thorough the years since the survey design has been unaltered since year 2004, allowing for a set of independent yearly samples (from 2004 to 2014) used to some extent although emphasis is placed on 2004 and 2014 samples. Thirdly, it features detailed information related to labour wages, which allow for a reasonably precise approximation of this key variable for our study. To allow comparability between years, deflated monthly labour wages are considered (see Appendix B for details).

The participation equation in the first stage of the sample selection correction model takes as dependent variable an indicator of the working status of the individual and, as independent variables, a set of individual characteristics as well as a set of variables which impacts on the probability that the person works but not on their expected wage. This corresponds to

$$employed = \Phi\left(\theta_0 + \gamma_1 sex + \theta_1 \mathbf{age} + \theta_2 \mathbf{educ} + \gamma_2 urban + \theta_3 kids_{0-6} + \theta_4 kids_{7-18} + \gamma_3 monoparental\right) \tag{10}$$

where $\Phi(.)$ is the standard normal distribution operator and $employed$ is a dummy variable which takes the value 1 if the individual works and 0 otherwise. Note that this model implicitly assumes that a person who wants to work can find a suitable job. The validity of this assumption can be, at first glance questionable, given that unemployment depends on the demand of employees exerted by firms. However, it does not seems to be the case in Peru: studies (Aliaga 2010; Freije 2002; Saavedra 1999) suggest that informal economy (i.e. those under unprotected jobs and unregulated enterprises) are an important source of employment for most of the population in periods of high unemployment or in order to diversify their sources of income, given the non-existent barriers to entry and the lack of regulation and law enforcement of these activities. The first set of independent variables correspond to individual characteristics: sex (a dummy which equals 1 if the individual is male and 0 otherwise), $\mathbf{age}$ (years of age as a vector including the first and second degree terms), $\mathbf{educ}$ (years of education as a vector including the first and second degree terms) and $urban$ (a dummy which equals 1 if the individual is located in urban area and 0 otherwise)[19]. The second set of variables includes the number and age composition of the children in the household, extending the idea presented in Gronau (1973). Following the models analysed by Mroz (1987), these are $kids_{0-6}$ and $kids_{7-18}$ (number of children in the household aged 0 to 6 years

[19] We included a set of dummies representing Peru's 25 regions in order to control for specific effects on each of these. However, these change little the results and, given our preference for a more parsimonious model, we retain the urban-rural dummy.

old and 7 to 18 years old) as well as *monoparental* (a dummy which equals 1 if the individual lives in a monoparental household). This latter variable is based on the fact that heads of monoparental households face difficulties to combine domestic work and paid activities, resulting on a lower probability on entering the labour market (OIT and PNUD 2009).

The wage equation in the second stage of the sample selection correction model corresponds to a variant on the basic Mincer's model.

$$\ln w = \beta_0 + \delta_1 sex + \beta_1 \mathbf{age} + \beta_2 \mathbf{educ} + \delta_2 urban + \mathbf{\Phi labour\,characteristics} + \upsilon \tag{11}$$

where the dependent variable is the natural log of the (monthly) labour wage deflated of the individual and sex, **age**, **educ** and $urban$ are defined and before. We do not rely on a unique basic model but, instead, following Arulampalam et al. (2007), Rica et al. (2008) and Albrecht et al. (2003), we estimate different models where we progressively add covariants in order to obtain a better estimate of the gender gap across the wage distribution. These are included in the **labour characteristics** vector and correspond to: a dummy of informal status (because of its pervasiveness in the economic structure), a vector of 8 industry dummies as well as a vector of 8 occupation dummies (which approximate characteristics of the labour demand) and a vector of 3 dummies capturing firm size (since these reflects, on average, differences in productivity). Construction of these variables are explained in detail in Appendix B. In the end, we have a set of 5 different models where first one only includes basic individual characteristics and the last one includes all the previously mentioned variables. Note that this implies a set of 5 different decompositions, whose comparison will allow us to assess how robust results are to alternative specifications.

Two important considerations are necessary to state. First, some of the variables added to the basic model arguably can be considered as endogenous characteristics. E.g., Dolton and Kidd (1994) suggest that if the difference between male and female distribution is a result of discriminatory practices, it is not legitimate to take the distribution as given; accordingly, they model occupation as endogenously determined and instrument out the wage. Furthermore, Dolton and Makepeace (1987) analyse the case where selection is not only related to the decision of working but also to decision to enroll in a union. Nevertheless, as an accounting exercise, it is useful to know the extent to which the gender gap at different quantiles can be explained by these variables. Second, the set of variables in the outcome equation constitute a subset of those in the selection equation. These exclusion restrictions allows to identify the slope coefficients in the outcome equation. It is worth keeping in mind that these reduced forms estimates should be interpreted as the sample's best linear predictors; causal interpretation for these coefficients is valid only if the underlying models for both equations are truly linear (Buchinsky 2001).

Before turning to the econometric estimation, it is useful to acknowledge that an important factor driving the results may be changes in of labour force composition between 2004 and 2014 (Table 1). Related to age profiles of the labours, it has remained steady for males and females, for the largest age group in both genders has been those between 26-35 years and 36-45 years. These two age categories account, each, for 30% of the labour force analysed here. As stated in section 2, informality characterizes most of employment in Peru, not only for female labours but also for males. Indeed, females have shown a (statistically significant at 5%) higher incidence of informality than males not only in 2004 but also in 2014. However, informality declined during this period, a result which is consistent with what ILO (2014) reports. A different pattern for both females and males emerges when we focus in the economic activity they perform. Considering the sector where they females employed, most have been allocated in Wholesale and Retail, Hotels and Restaurants sector (near 40% of the labours) followed by Community, Social and Personal Services (around 30%); considering the occupation they carry out, Service and sales (25% in 2004 and 29% in 2014) and Elementary occupations (33% in 2004 and 27% in 2014) account for more than half of the jobs. According to OIT and PNUD (2009), these are characterized for high rates of informality and hence it may explain its higher prevalence in females. For males, there is a more widespread distribution of employment around the categories shown. The share of workers in Agriculture sector has decreased (nearly 30% in 2004 to 23% 2014) while the share of Construction sector has increased (from around 10% to approximately 15%) and by 2014 this latter is the second largest sector. The share of Mining and Quarrying sector, characterized for higher productivity and incomes, increased in 0.7 percentage points (higher than in the case of women) but still holds a low share: 2.4%. Several occupations depict a high concentration of males workers, being the highest Elementary occupations (26% in 2004 and 23% in 2014). Finally, both male and female labour have been allocated, mainly, in micro firms, with women having a larger share in this category in 2014 (76.29% and 69.36% for males). Nevertheless, between 2004 and 2014 labour employed in micro firms experienced a relative reduction which contrasts with the increase of allocation in large firms (this increased from 5.1% to 9.7% for females and 9.1% to 13.7% for males). As World Bank (2015) and Távara et al. (2014) show, large firms, on average, show a higher productivity in Peru; then, the shift to these activities is an important contributor to the increase of wages during this period.

Table 2 shows summary statistics for the variables which will used in the estimations. Since we are not using (only) a mean regression method but, instead, a quantile regression approach, we focus on the average values at each quintile of the wages distribution instead of their unconditional means. A first noticeable fact is that the average wage for women are systematically lower than those of males both at any point of the distribution. This holds for 2004 and 2014. Nevertheless, an important phenomenon arises: the relative gap is reduced as we move up the distribution. In 2004, at the first quintile,

Table 1 – Distribution of working females and males by different characteristics, 2004 and 2014

	Females				Males			
	2004		2014		2004		2014	
Age								
18-25	18.00	(0.54)	17.46	(0.39)	18.01	(0.39)	17.52	(0.31)
26-35	29.62	(0.64)	26.07	(0.43)	29.48	(0.49)	25.54	(0.36)
36-45	28.84	(0.65)	27.89	(0.45)	26.60	(0.51)	27.84	(0.38)
46-55	15.45	(0.46)	17.97	(0.32)	15.99	(0.35)	17.33	(0.27)
56-65	8.10	(0.35)	10.62	(0.27)	9.91	(0.33)	11.77	(0.24)
Informal	78.82	(1.03)	65.47	(0.56)	72.67	(0.88)	55.85	(0.49)
Sector								
Agriculture, Forestry, and Fishing	10.96	(0.46)	10.23	(0.29)	30.34	(0.63)	23.08	(0.39)
Mining and Quarrying	0.18	(0.07)	0.25	(0.05)	1.69	(0.23)	2.40	(0.18)
Manufacturing and Public Utilities	10.99	(0.51)	10.14	(0.34)	12.54	(0.42)	11.17	(0.29)
Construction	0.59	(0.10)	1.36	(0.11)	9.78	(0.33)	15.17	(0.34)
Wholesale and Retail, Hotels and Restaurants	42.96	(0.78)	40.71	(0.52)	12.81	(0.38)	13.11	(0.32)
Transport, Storage, and Communications	1.45	(0.24)	2.27	(0.17)	10.92	(0.35)	13.44	(0.32)
Finance, Insurance, and Real Estate	3.55	(0.31)	6.12	(0.27)	6.33	(0.36)	6.92	(0.25)
Community, Social and Personal Services	29.31	(0.69)	28.92	(0.49)	15.58	(0.45)	14.72	(0.32)
Occupation								
Managers, professionals and armed forces	11.71	(0.55)	12.74	(0.39)	9.29	(0.41)	8.95	(0.28)
Technicians and associates	6.47	(0.38)	7.15	(0.30)	7.32	(0.35)	8.65	(0.26)
Clerks	6.82	(0.43)	10.24	(0.34)	4.04	(0.22)	6.21	(0.24)
Service and sales workers	25.93	(0.64)	28.96	(0.47)	6.70	(0.28)	8.52	(0.25)
Skilled agricultural and fishery workers	7.33	(0.35)	6.58	(0.22)	22.36	(0.52)	16.73	(0.32)
Craft and related trades workers	7.26	(0.45)	6.66	(0.28)	14.96	(0.43)	14.68	(0.36)
Plant and machine operators and assemblers	1.42	(0.17)	1.04	(0.11)	9.80	(0.34)	13.46	(0.31)
Elementary occupations	33.06	(0.73)	26.63	(0.49)	25.52	(0.56)	22.80	(0.38)
Firm size								
Micro	83.44	(0.67)	76.29	(0.51)	75.92	(0.62)	69.36	(0.47)
Small	5.22	(0.39)	7.06	(0.31)	7.68	(0.36)	7.81	(0.25)
Medium	6.17	(0.42)	6.91	(0.29)	7.28	(0.35)	9.08	(0.29)
Large	5.17	(0.38)	9.74	(0.37)	9.12	(0.43)	13.75	(0.36)

Note: Sample include individuals between 18 and 65 years. Observations weighted by expansion factor and VCE corrected according to survey's complex sample design. Standard errors in parenthesis.
Elaborated by the author based on INEI - National Household Survey (2004-2014)

the relative (unconditional) difference in averages[20] at the bottom quintile is 2.1 which shrinks to 1.4 at the top quintile; in 2014, the relative difference at the bottom is 2.4 and this is reduced to 1.4 at the top. In other words, the stability of the relative wages between 2004 and 2014 depicted on Figure 1 hides an important heterogeneity across the wages distribution: namely, the decrease of the gap as we move up the distribution; however, the male's advantage holds in general.

Taking into consideration individual characteristics variables, the average age of labours is similar across quantiles but it increases as we move from the middle quintile

[20]If we define $\ln(M) \equiv m$ and $\ln(F) \equiv f$, it follows that the relative gap between males and females in the usual metric, $\frac{M}{F}$, equals $\frac{M}{F} = \exp\left(\ln\left(\frac{M}{F}\right)\right) = \exp(\ln M - \ln F) \equiv \exp(m - f)$, i.e. exponentiation of the difference of the values in the table.

to the upper parts of the distributions: the richest group's average is around 40 years. Also, consistent with what human capital theories suggest, there is an association of years of schooling and income. Nevertheless, two points are important to notice. First, males show, on average, higher education levels in all the quintiles but the last. This may be a factor underlying the reduction of the relative gap at the top shown before. Second, there is a general increase in educational levels between 2004 and 2014 for both genders, irrespective of their position in the wages distribution. For instance, females at the fifth quintile went from 8.8 years of education in 2004 to 9.9 years in 2014 while males in the same group went from 9.4 years to 10.2 years. Considering the location of households, a consistent finding arises: urbanization is more prevalent as we move up the distribution and the differences between the top and bottom quintiles are striking. E.g. for 2014, differences in urbanization rates between those at the top and at the bottom is 40 and 26 percentage points in the case of males and females, respectively. This disparity, although not analysed thoroughly in this study, can be attributed to the fact that rural men are prone to migrate to urban areas in order to provide remittances for their households, given the higher wages they can earn in the latter area. Hence, more females are able to remain employed in rural areas, earning lower wages and locating at lower parts of the distribution.

Considering labour characteristics of the workers, informality is negatively associated with income. E.g. the top quintile in 2014 shows on average 25% of informality whereas the poorest one shows on average 88%. There are also important differences between males and females related to the sector where they work. On the one hand, in 2004 and 2014 most of the females are in the tertiary sector although at the lowest quintiles they are located mainly in the primary sector (agriculture). This is consistent with the view that most females are located in the rural areas, where wages are lower. On the other hand, only males in the lowest quintiles are predominantly located in the agricultural sector; beyond this point, there is more diversified allocation. Focusing in the occupation, females of the first four quantiles have been engaged mainly in elementary occupations as well as clerk and sales workers, being these usually characterized for a low productivity. Only those of the top quintile have an important participation in the manager, professional or armed forces group. A less uniform result in both years is found when analysing men. Those at the lowest quintile have been working mostly as agriculture and fishery workers as well as elementary occupation workers (these account, approximately, for 50% and 25% of the total of employed, respectively). Those at the third and fourth quintile have worked mainly as craft workers and plant operators. In the top quintile, managers and professionals are the prevalent. Considering that, as stated before, largest firms are the most productive in Peru, those in larger firm sizes show higher incomes, a pattern which holds consistently for 2004 and 2014.

Finally, regarding household level variables, as we move from the lower to the upper part of the wages distributions the number of underaged members decreases. This is more noticeable when we analyse children between 7 and 18 years old. An interesting result emerges when we analyse monoparentality: its prevalence increases as we go from the poorer to the richer females, a result which holds for both years. All in all, the information presented in these two tables will be useful to discuss the results of the estimations and decompositions, shown in the next section.

Table 2 – Within quintiles means of variables for female and male samples, 2004 and 2014

	Females					Males				
	Q1	Q2	Q3	Q4	Q5	Q1	Q2	Q3	Q4	Q5
2004										
Dependent variable										
(Ln.) Labour income of the individual	4.40	5.61	6.26	6.78	7.56	5.13	6.17	6.65	7.06	7.89
Individual level variables										
Age of the individual	38.21	37.78	35.35	36.03	38.50	37.28	36.45	35.92	37.23	40.65
Years of schooling of the individual	6.77	7.48	8.80	10.31	13.14	7.14	7.98	9.44	10.50	12.67
If household is located on a urban zone (d)	0.63	0.77	0.86	0.92	0.93	0.43	0.58	0.78	0.86	0.90
Labour characteristics variables										
If the labour is informal (d)	0.96	0.96	0.93	0.79	0.36	0.93	0.91	0.83	0.64	0.34
If the labour is in primary sector (d)	0.22	0.17	0.10	0.06	0.02	0.65	0.49	0.25	0.16	0.13
If the labour is in secondary sector (d)	0.20	0.10	0.08	0.12	0.09	0.12	0.19	0.29	0.26	0.23
If the labour is in tertiary sector (d)	0.59	0.73	0.82	0.82	0.89	0.22	0.32	0.46	0.58	0.64
If the labour is a manager, professional or armed forces (d)	0.01	0.03	0.04	0.09	0.38	0.01	0.01	0.03	0.08	0.31
If the labour is a technician (d)	0.01	0.02	0.04	0.08	0.16	0.03	0.03	0.05	0.09	0.15
If the labour is a clerks or sales worker (d)	0.23	0.34	0.36	0.38	0.32	0.04	0.06	0.11	0.14	0.17
If the labour is an agric. and fishery worker (d)	0.17	0.12	0.05	0.03	0.01	0.54	0.35	0.15	0.10	0.06
If the labour is a craft worker or a plant operator (d)	0.19	0.09	0.06	0.08	0.03	0.13	0.21	0.32	0.35	0.21
If the labour is an elementary occupation worker (d)	0.38	0.41	0.45	0.34	0.10	0.25	0.34	0.34	0.25	0.10
Labours in the firm where the individual works	5.04	7.08	32.65	82.86	277.06	7.98	28.66	52.74	120.01	376.98
Household level variables										
Number of children of HoH between 0 and 6 years	0.43	0.36	0.29	0.24	0.24	0.50	0.50	0.40	0.40	0.39
Number of children of HoH between 7 and 18 years	1.19	1.20	1.08	0.97	0.78	1.24	1.22	1.06	1.06	1.03
If household has only a single parent (d)	0.27	0.31	0.34	0.34	0.30	0.21	0.20	0.22	0.21	0.18
2014										
Dependent variable										
(Ln.) Labour income of the individual	4.69	6.02	6.67	7.11	7.85	5.58	6.63	7.07	7.45	8.17
Individual level variables										
Age of the individual	39.98	39.21	36.60	36.69	39.94	38.31	37.04	37.48	38.94	40.98
Years of schooling of the individual	7.75	8.68	9.90	11.37	13.58	8.20	9.23	10.21	11.16	12.85
If household is located on a urban zone (d)	0.69	0.80	0.89	0.93	0.95	0.51	0.71	0.83	0.87	0.90
Labour characteristics variables										
If the labour is informal (d)	0.90	0.88	0.77	0.51	0.27	0.86	0.77	0.59	0.42	0.23
If the labour is in primary sector (d)	0.23	0.16	0.08	0.05	0.03	0.57	0.33	0.18	0.12	0.14
If the labour is in secondary sector (d)	0.14	0.09	0.12	0.14	0.09	0.15	0.22	0.31	0.32	0.29
If the labour is in tertiary sector (d)	0.63	0.75	0.80	0.81	0.88	0.29	0.45	0.51	0.55	0.57
If the labour is a manager, professional or armed forces (d)	0.01	0.05	0.07	0.13	0.35	0.02	0.01	0.04	0.10	0.25
If the labour is a technician (d)	0.02	0.03	0.06	0.09	0.15	0.04	0.04	0.07	0.11	0.16
If the labour is a clerks or sales worker (d)	0.28	0.38	0.44	0.45	0.40	0.06	0.13	0.15	0.17	0.20
If the labour is an agric. and fishery worker (d)	0.18	0.11	0.04	0.02	0.01	0.48	0.22	0.09	0.06	0.05
If the labour is a craft worker or a plant operator (d)	0.15	0.08	0.08	0.07	0.02	0.16	0.30	0.34	0.34	0.25
If the labour is an elementary occupation worker (d)	0.36	0.35	0.32	0.24	0.07	0.24	0.29	0.30	0.22	0.09
Labours in the firm where the individual works	6.06	22.46	76.15	212.12	447.36	13.13	60.23	153.39	219.45	496.40
Household level variables										
Number of children of HoH between 0 and 6 years	0.34	0.27	0.22	0.19	0.17	0.34	0.32	0.32	0.29	0.32
Number of children of HoH between 7 and 18 years	0.94	0.89	0.79	0.72	0.60	1.03	0.91	0.84	0.80	0.78
If household has only a single parent (d)	0.28	0.32	0.34	0.36	0.35	0.22	0.23	0.23	0.21	0.18

Note: (Ln.) Wages measured in constant 2014 Soles. Sample include individuals between 18 and 65 years. Observations weighted by expansion factor. Primary sector includes Agriculture, forestry, and fishing, Mining and quarrying sectors. Secondary sector includes Manufacturing, Public Utilities, Construction, Wholesale and Retail, Hotels and Restaurants sectors. Terciary sector includes Transport, Storage, and Communication, Finance, Insurance, and Real Estate, Community, Social and Personal Services sectors. (d)=Dummy variable.
Elaborated by the author based on INEI - National Household Survey (2004-2014)

5 Results

This section, on the first place, presents the results of quantile regression in order to analyse the gender wage gaps at different points of the distribution considering an unconditional model as well as a conditional model, assuming both equal and differing returns to their characteristics. On the second place, it presents results of MM decompositions, which will allow to unveil what part of the male-females gender wage gap can be attributed to a discrimination factor against women considering different points of the wage distribution. Considering different models will provide us with a robustness check and a clearer picture of gender inequality.

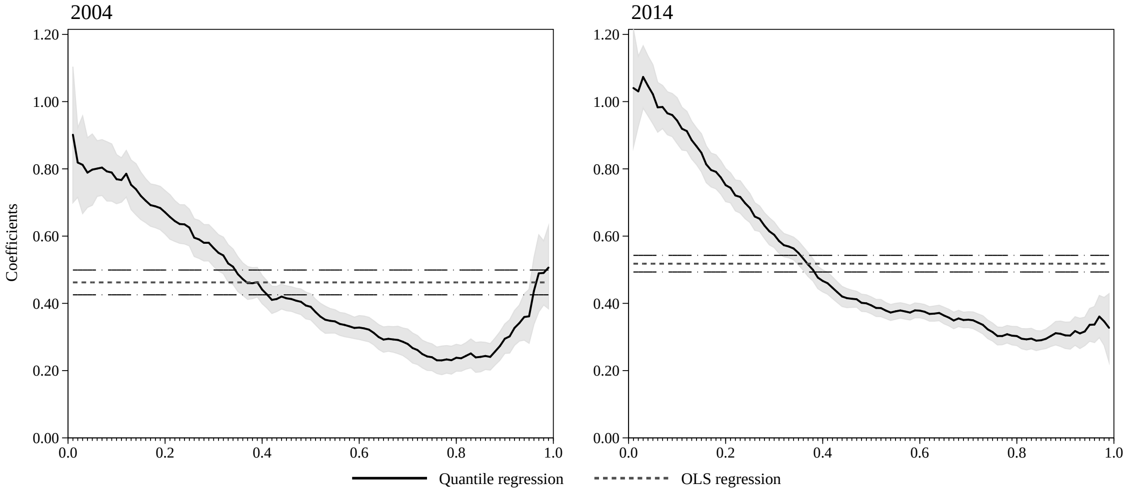
5.1 Estimation of quantile regressions

A first approach to the study of gender gaps throughout the distribution is to analyse the observed (unconditional) differences in wages (Figure 3). Focusing first on the mean difference, results show that the OLS coefficient of the gender dummy (horizontal dotted lines) amounts to 0.46 in 2004 and to 0.52 in 2014, a change which is marginally significant at 5%. Put differently, the gender-prime of males over females, on average, has increased from 59% to 68% between those 2 years[21]. Nevertheless, as stated before, this does not takes into account the bewildering variety of relative differences at different point of the distribution. The estimates of the gender dummies from the quantile regression at each percentile (and their associated confidence intervals) account for this (solid lines). An inspection of these show three remarkable characteristics. In the first place, each one of the estimates are over the value of zero, which implies that the gender wage gap favours men not only when we look at the mean but also when analysing different parts of the distributions. In the second place, the (unconditional) wage gaps show a U pattern for 2004 and a more convex-from-the-origin shape for 2014. I.e. the (raw) gender advantage of males over females decrease as we move from the lowest-payed employees to those more well-of (e.g., for 2014, the gap is 98% at the 25th percentile, 48% at the median and 37% at the 75th percentile) until we reach a point high enough in the distribution (around the 85th percentile) after which the gap increases again but at a very moderate rate. This means that at the higher parts of the distribution the advantage of men increases again although slightly. However, this non-linear behaviour is more blatant in 2004 than in 2014. In the third place, gaps are higher for year 2014 and 2004, except at the last decile (due to the differences in this area outlined). This difference is more noticeable at the first three deciles and at the 6th, 7th and 8th decile (on average, 9 percentage points higher).

Although revealing, these estimates are not purged from differences in credentials

[21] Throughout this subsection we are using the exact change formula, gap $= [(\exp(\beta) - 1) \times 100]\,\%$, instead of the usual -although inexact in the case of dummies- approximation, gap $= [100\beta]\,\%$.

Figure 3 – Raw gender (log) wage gaps across percentiles, 2004 and 2014



Note: (Ln.) Wages measured in constant 2014 Soles. Gender dummy equals 1 if individual is male and 0 otherwise. Sample include individuals between 18 and 65 years. Observations weighted by expansion factor and VCE corrected according to survey's complex sample design. Confidence intervals for the quantile regression coefficients (shaded area) and the OLS regression (dotted line) coefficient correspond to the 95% level.
Elaborated by the author based on INEI − National Household Survey (2004−2014)

of the individuals; i.e. we are comparing averages of people with different preparation, age and labour attributes. In order to examine the effects of differences in characteristics on the gender gap at different points of the distribution, we carry out a series of quantile regressions correcting for sample selection and focus on the coefficients of the gender dummy (Table 3)[22]. In the first place, we net out basic characteristics (age, education and area of residence). The average wage gap results in 68% in 2004 and 78% in 2014 (column labelled Heckman). This central tendency indicator contrasts with the estimates under quantile regression at the representative percentiles presented in the Table. For example, under this simple model, the gender wage gap in 2004 goes from 92% in the 10th percentile to 61% in the median and 55% in the 90th percentile. For 2014, the gap is 98% at the 1st decile, 65% at the median and 77% at the 9th decile. This distributional change in the estimates is consistent the pattern shown in the last Figure.

However, individual characteristics are not the only factors intervening the payment that the labour receives. In order to obtain the gender wage gaps taking workers which are comparable in terms of their labour characteristics we include in the remaining rows, one by one, relevant variables which allow to refine these estimates. Under the model including only basic characteristics and informality, estimates are lower than

[22] Results without correcting for selection are shown in Table A2 in Appendix A. Comparison of results shows that uncorrected estimates of mean regression are of similar magnitude than those corrected by sample selection. However, uncorrected estimates are higher when we focus in the coefficients estimated by quantile regression at the different percentiles showed. Nevertheless, the same pattern and relative differences between 2004 and 2014 hold.

when including only personal characteristics. The average wage gap is 55% in 2004 and 67% in 2014, and at other points of the distribution the gap decreases as we move from the the poorest to the richest individuals. For 2004, it goes from 76% at the 10th percentile, 16% at the median and 16% at the third quartile (significant only at 10%). Note that the coefficient at the 90th percentile is not significantly different from zero at conventional levels of significance. Gender differences for 2014 are higher than in 2004: 79% at the 10th percentile, 54% at the median and 55% at the top decile. Inclusion of sectoral dummies to the latter model does not modify results in 2014 while in 2004 this reduces differences at the 10th (68%) and at the 25th percentile (60%). Similar as in the previous model, the wage gap at the top of the distribution for this year is statistically indistinguishable from zero. When we add a set of occupation dummies, mean gaps increase relative to the latter model (it is now 64% for 2004 and 68% for 2014) and also there is a reduction in the gap as we move up the distribution. Only in 2004, gaps change noticeably at different deciles: it is now 101% at the bottom decile and 9% at the 90th percentile (although statistically not different from zero). Comparing the two years, the differences have increased remarkably at the top (it went from 29% to 53% for the 75th percentile and from 9% to 54% at the 90th percentile between 2004 and 2014). Finally, inclusion of firm size variables does not change the mean gender wage gap estimates but changes the gaps at the bottom.

The essential message is clear. The mean gender wage gaps and those at the selected quantiles experienced a generalized increase between 2004 and 2014 which favours males. Across the different models considered, the mean wage gap ranges from 55% to 68% for 2004 and from 64% to 78% for 2014. Estimates for the 10th percentile go from 68% to 101% in 2004 and from 67% to 98% in 2014; for the median, from 16% to 41% in 2004 and for 52% to 65% in 2014 and for the top decile, from 4% to 55% in 2004 and 55% to 77% in 2014. However, gaps at the top of the distribution are not significant in 2004. This suggest a strong sticky floor effect in 2004 and a sticky floor as well as a glass ceiling effect in 2014, although the former effect is stronger in 2014. Comparison of the the observed gap (top row) with the gap netting out individual and labour characteristics reveals that considering a more elaborate model leads to a slight increase in the mean gap for both years as well as in lower gaps for the bottom of the distribution. From the 50 percentile onwards, adding the full set of covariants results in higher gaps in year 2014 and lower differences in 2004, to the point of losing statistical significance.

The results just outlined are constrained in a way: they assume that returns of the characteristics are the same for males and females. However, since our ultimate interest lies on being able to assert what part of the gap is due to discrimination against women, i.e. lower returns to their characteristics, this homogeneity assumption should be ruled out. Hence, instead of carrying out regressions considering a pooled dataset including both gender and males, we now estimate the models taking males separately

Table 3 – Gender wage gaps under alternative models at selected quantiles, 2004 and 2014

	Heckman	Quantile regression					Obs.
		10th	25th	50th	75th	90th	
Year 2004							
Observed	0.462	0.769	0.625	0.390	0.240	0.295	28,121
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 1 (Basic controls)	0.520	0.651	0.466	0.475	0.465	0.436	28,119
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 1 + Informality	0.437	0.565	0.262	0.323	0.151	0.040	14,244
	[0.000]	[0.009]	[0.093]	[0.001]	[0.085]	[0.787]	
Model 2 + Sector	0.464	0.516	0.472	0.369	0.156	0.184	14,244
	[0.000]	[0.008]	[0.002]	[0.000]	[0.083]	[0.171]	
Model 3 + Occupation	0.492	0.698	0.429	0.361	0.257	0.087	14,244
	[0.000]	[0.000]	[0.001]	[0.000]	[0.002]	[0.467]	
Model 4 + Firm size	0.491	0.673	0.327	0.364	0.192	0.084	12,577
	[0.000]	[0.000]	[0.036]	[0.001]	[0.043]	[0.508]	
Year 2014							
Observed	0.518	0.944	0.683	0.394	0.315	0.305	45,752
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 1 (Basic controls)	0.577	0.682	0.474	0.502	0.570	0.569	45,745
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 1 + Informality	0.511	0.582	0.486	0.434	0.467	0.436	45,745
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 2 + Sector	0.494	0.528	0.476	0.421	0.439	0.434	45,745
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 3 + Occupation	0.519	0.510	0.493	0.434	0.422	0.431	45,744
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 4 + Firm size	0.530	0.543	0.547	0.487	0.422	0.466	39,466
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	

Note: (Ln.) Wages measured in constant 2014 Soles. Gender dummy equals 1 if individual is male and 0 otherwise. Discrete effects for the coefficient of gender dummy (evaluated at the mean) are reported. Sample include individuals between 18 and 65 years. Observations weighted by expansion factor and VCE corrected according to survey's complex sample design. P-values of the gender coefficients shown in brackets. Observed model includes only the gender dummy; basic controls include gender, years of schooling (2nd degree polynomial), age (2nd degree polynomial) and a dummy of area of residence.
Elaborated by the author based on INEI - National Household Survey (2004-2014)

from females. For the sake of brevity, we only consider results of the first model (individual characteristics only) and the full model (individual and the set of labour characteristics) correcting for selection, but the results for the intermediate models are available upon request. Considering the simpler model (Table 4)[23] in 2004, mean estimates describe a (significant) non linear relationship of age and wage for both genders, with the inflection point at 32 years old for females and 45 years old for males. The marginal effect (evaluated at the mean values of covariants) is higher for males (1.8% versus 1.2% for females). Returns for education show, for 2004, a (statistically significant) convex pattern for both genders; however, the inflection point occurs at negative years of education and, hence, for almost all the individuals sample, we can consider that education

[23] Estimates without correcting for sample selection are shown in table A3 in Appendix A. These are qualitatively similar to the uncorrected estimates, although the significance of the IMR term in mean and quantile regression implies that it is methodologically correct to consider the corrected estimates.

has a monotonic (positive) effect on wages[24]. Mean regression estimates suggest that marginal returns of education (evaluated at the means) are very similar for females and for males (9.6% and 9.1%, respectively). Quantile regression estimates show that returns are higher for females only at the 25th and 50th percentile (e.g. 10% for females and 8% for males in both percentiles). For 2014, this (statistically significant) convexity of returns to education holds for estimates under Heckman method and show, again, higher marginal returns for females than males (9.6% and 9.0%, respectively). Quantile regressions for this year show a similar pattern of that depicted for 2004. Urbanization variable exerts a positive influence on the mean wage in 2004, being the discrete effect similar for females and males: a urban woman or a male earns 75% more compared with an individual with the same characteristics but living in the rural area. However, the effect is higher for males at the different quantiles but decreases as we move from the bottom to the top of the distribution. For the 10th percentile, it is 110% for females and 131% for males; for the 90th percentile, it is 41% and 68% for females and males respectively. This disparity between the urban returns for each gender holds also in 2014: on average and ceteris paribus, living in the urban area implies a wage 70% higher for females but only 56% higher for males. The same decrease of returns throughout the distribution is found in both genders but, contrary to what happens in 2004, females earn a higher prime.

Results of the full model for 2004 (Table 5)[25] suggest that the effect of age is very similar to what was found with the basic model: both females and males experience a (statistically significant) non-linear relationship under mean regression technique. Also, returns to age (evaluated at the mean) are higher for males (1.4% vs 0.7%). Coefficients of education depict, for the most part, a (significant) concave pattern except at the upper part of the wage distribution for males. Returns to education under Heckman model (evaluated at the mean) are higher for males than for females (3.5% versus 2.2%) and the same holds when we consider the quantile regression estimates (e.g. at the 1st quartile is 3.5% for males and 1% for females; at the 90th percentile is 4.3% for males and 2.8% for females). Unlike the previous model, urbanization shows a remarkably higher return for females not only considering the mean but also the quantile regression estimates. Again, the effect decreases in the upper parts of the distribution.

The first labour characteristic, informality, has again a negative effect and it represents a burden of 31% in females and of 16% in males. Indeed, the penalization in females represents twice the magnitude (in percentage points) compared to that accruing to men at different quantiles. This negative effect is not reduced at higher parts of the distribution. The set of coefficients related to labour sector estimated by Heck-

[24]This can be explained because the estimation sample includes people with only kindergarten, i.e. 0 years of schooling, who represents 6% of the sample in 2004 and 4% in 2014.

[25]See Tables A4 and A5 in Appendix A for results without sample selection correction. However, the fact that sample selection correction terms (IMR) are significant lead us to focus only on those corrected estimates.

man method (jointly significant for both genders) taking as reference the agricultural sector, shows that only females located in Mining and Quarring, Manufacture and Construction sectors obtain higher returns than those in agriculture (83%, 23% and 79% respectively), while in case of males all sectors exhibit higher returns except in Services and Finance. Through the percentiles, there is only a clear pattern for males, since the sectors with higher returns compared to agriculture are the same as in mean regression. Considering the vector of coefficients related to occupation (statistically significant for both genders) taking as reference group those who carry out elementary occupations, mean and quantile regression estimates show that females who work at the different occupations (with exception of Skilled agricultural and fishery workers) earns higher wages than the reference category (being the highest differential found in the managers group and technicians group), ceteris paribus. A similar pattern emerges when we consider coefficient for males and an interesting regularity is found when comparing mean estimates with those at the 90th percentile: the coefficients for all occupations in the former are notably higher than for the latter which can be signalling higher education for men engaged in these occupations at the top of the distribution. For both sexes a rule emerges when analysing the firm size coefficients: the higher the size of the firm, the biggest the wage differential relative to micro-sized firms. The estimated prime associated to the different categories is higher for females under mean regression but this effect shrinks as we move from the bottom to the top of the distribution and remain higher for females only until the 75th percentile. However, at the 90th percentile there is no difference between the wage of a woman working on a small sized firm and an otherwise comparable women in a micro firm, but males in small sized forms earn 22% more than those in a micro firm.

Finally, results of the full model for 2014 (Table 6) suggest that the effect of age is very similar to what was found in the year 2004 in terms of the significance of the parameters and the marginal effects. Coefficients of education depict a different non-linear pattern in both genders: it is mainly concave for females (except at the upper part of the distribution) and convex for males at the upper part (but the inflection point occurs at 4 years of education and so for most of this sub-sample returns are positive). Returns to education under Heckman model (evaluated at the mean) are similar but these are higher for males when we consider quantile regression estimates (e.g. at the 1st quartile is 1.8% for males and 0.1% for females; at the 90th percentile is 4.5% for males and32.8% for females). Similar to results in 2004, urbanization estimates show a higher return for females under both regression techniques (although they are similar at the 90th percentile). E.g. Heckman correction estimate provides a 47% effect on wages for females and 15% for males and quantile regression at the median provides a 45% effect on wages for females and 11% for males.

Informality keeps exerting a negative effect in the wages of workers independently of their gender, but this time the effect is similar. At the 10th quartile, penalization is

26.51% for females and 25.32% for males; at the median, 18% for females and 17% for males. The set of coefficients related to sector where the individual works estimated by Heckman method (jointly significant in both genders), taking as reference the agricultural sector, show a similar pattern to 2004. Females located in Mining and Quarring, Manufacture, Construction and Finance sectors obtain higher returns than agriculture (127%, 11%, 23% and 16% respectively), while males exhibit higher returns in all sector compared to agriculture (being the highest effect found in Mining and Quarring, 87%). Considering the vector of coefficients related to occupation (statistically significant in both genders at any level of significance), taking as reference group elementary occupations, mean and quantile regression estimates show that on average females who work at the different occupations earn different wages than those at the base category; this time, those working as the Skilled agricultural, Craft and related and Plant and machine operators earn lower wages than those in elementary occupations, ceteris paribus. Consistently across estimates, men working as Skilled agricultural labours earn lower wages than the base category. For both sexes the same rule emerges: the higher the size of the firm, the biggest the wage differential relative to micro-sized firms. The average prime associated to the different categories is higher for females than for males. For example, men in large firms earn, on average and ceteris paribus, 56% more than those in small firms while females earn 73% more. Note that most of the coefficients are lower than those found in 2004 and that the differential also shrinks as we move from the bottom to the top of the distribution. At the 90th percentile there is no difference between the wage of a woman working on a small sized firm and an otherwise comparable women on micro firm.

These results indicate that returns to observed characteristics of the workers exhibit a similar pattern when we compare those results for 2004 and 2014. Also, returns to personal characteristics are different for males and females in terms of returns to education (higher for males), returns to urbanization (higher for females only in the full model). When focusing on returns to labour characteristics, informality exerts a similar burden in 2014 but not in 2004 (it was higher for females). Occupation coefficients follow different patterns for both males and females and returns to larger firm sizes are higher for females. Indeed, quantile regression estimates show that returns to the different characteristics, for the most part, vary throughout the wages distribution and hence this regression technique adds to the understanding of difference in returns between genders. At this point, it would be useful to decompose the gender gap into two parts: that arising from differences in characteristics and those arising from differences in returns.

Table 4 – Regressions by gender under basic model at selected quantiles, 2004 and 2014

	Females						**Males**					
	Heckman	**Quantile regression**					**heckman**	**Quantile regression**				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
2004												
Age of the individual	−0.030***	0.008	0.006	0.037***	0.048***	0.098***	0.050***	0.011	0.031*	0.031***	0.009	−0.011
	(0.008)	(0.028)	(0.020)	(0.012)	(0.011)	(0.013)	(0.006)	(0.020)	(0.018)	(0.012)	(0.013)	(0.016)
Age of the individual^2	0.000***	0.000	0.000	0.000**	0.000***	−0.001***	−0.001***	0.000	0.000	0.000**	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling of the individual	0.047***	−0.016	0.029	0.031***	0.021*	0.001	−0.003	0.010	0.005	0.001	−0.012	−0.039***
	(0.014)	(0.024)	(0.019)	(0.012)	(0.011)	(0.015)	(0.010)	(0.012)	(0.011)	(0.008)	(0.009)	(0.011)
Years of schooling of the individual^2	0.002***	0.006***	0.004***	0.004***	0.004***	0.005***	0.005***	0.005***	0.004***	0.004***	0.005***	0.007***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
If household is located on a urban zone (d)	0.565***	0.746***	0.723***	0.530***	0.303***	0.343***	0.547***	0.838***	0.718***	0.539***	0.467***	0.520***
	(0.043)	(0.068)	(0.056)	(0.039)	(0.034)	(0.037)	(0.025)	(0.044)	(0.038)	(0.028)	(0.031)	(0.037)
Constant	6.226***	4.959***	5.375***	5.000***	5.126***	4.510***	4.776***	4.595***	4.749***	5.271***	6.226***	6.996***
	(0.177)	(0.650)	(0.487)	(0.283)	(0.263)	(0.348)	(0.120)	(0.416)	(0.370)	(0.252)	(0.273)	(0.331)
Observations	17,946	10,417	10,417	10,417	10,417	10,417	20,625	17,702	17,702	17,702	17,702	17,702
F test IMR	−3.087	11.006	18.287	12.070	2.271	0.829	−4.265	23.376	24.151	19.742	23.482	25.122
	[0.002]	[0.000]	[0.000]	[0.000]	[0.103]	[0.437]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test education	204.759	62.366	185.266	412.664	387.632	156.354	344.060	265.909	325.060	600.411	455.445	420.798
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test model	181.963	73.265	215.886	350.249	271.963	154.120	470.959	404.256	467.293	713.570	485.619	431.298
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
2014												
Age of the individual	−0.045***	−0.082***	−0.034**	0.016**	0.041***	0.057***	0.024***	0.007	−0.019	−0.026**	−0.016	−0.021
	(0.006)	(0.024)	(0.015)	(0.007)	(0.009)	(0.011)	(0.004)	(0.022)	(0.016)	(0.010)	(0.011)	(0.015)
Age of the individual^2	0.001***	0.001***	0.000**	0.000*	0.000***	−0.001***	0.000***	0.000	0.000	0.000***	0.000*	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling of the individual	0.011	−0.004	0.010	0.010	−0.013	−0.026**	−0.014**	0.006	0.005	−0.017**	−0.046***	−0.050***
	(0.009)	(0.021)	(0.013)	(0.008)	(0.009)	(0.011)	(0.007)	(0.013)	(0.010)	(0.007)	(0.007)	(0.009)
Years of schooling of the individual^2	0.003***	0.004***	0.004***	0.004***	0.005***	0.006***	0.005***	0.004***	0.004***	0.005***	0.007***	0.007***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
If household is located on a urban zone (d)	0.532***	0.797***	0.728***	0.521***	0.314***	0.262***	0.444***	0.800***	0.665***	0.474***	0.324***	0.237***
	(0.031)	(0.055)	(0.039)	(0.027)	(0.025)	(0.031)	(0.018)	(0.036)	(0.027)	(0.019)	(0.018)	(0.026)
Constant	7.029***	7.227***	6.343***	5.659***	5.729***	5.720***	5.847***	5.078***	6.199***	7.061***	7.416***	7.927***
	(0.118)	(0.611)	(0.367)	(0.194)	(0.224)	(0.291)	(0.083)	(0.493)	(0.365)	(0.240)	(0.242)	(0.352)
Observations	29,263	18,835	18,835	18,835	18,835	18,835	30,865	26,910	26,910	26,910	26,910	26,910
F test IMR	−4.189	29.461	32.971	15.489	5.228	0.274	−7.011	22.228	42.385	73.749	40.314	21.024
	[0.000]	[0.000]	[0.000]	[0.000]	[0.005]	[0.760]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test education	362.109	42.417	181.165	497.032	326.743	235.598	788.881	271.722	535.010	847.914	840.850	609.862
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test model	314.516	140.851	357.916	572.994	300.654	202.153	688.121	445.985	643.310	796.252	598.028	398.137
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

Note: (Ln.) Wages measured in constant 2014 Soles. Sample include individuals between 18 and 65 years. Observations weighted by expansion factor and VCE corrected according to survey's complex sample design. Heckman model estimated by Maximum Likelihood. Standard errors in parenthesis and p-values of the F-test in brackets. (d)=Dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level.

Table 5 – Regressions by gender under full model at selected quantiles, 2004

	Females						**Males**					
	Heckman	**Quantile regression**					**Heckman**	**Quantile regression**				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
Individual level variables												
Age of the individual	−0.043***	0.034	0.010	0.020	0.024*	0.022	0.034***	0.045	0.018	0.001	0.026**	0.025
	(0.013)	(0.024)	(0.026)	(0.017)	(0.014)	(0.018)	(0.008)	(0.034)	(0.022)	(0.016)	(0.013)	(0.020)
Age of the individual^2	0.001***	0.000	0.000	0.000	0.000	0.000	0.000***	−0.001	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling of the individual	0.097***	0.083***	0.096***	0.084***	0.068***	0.071***	0.022	0.058***	0.036**	0.022*	0.028***	0.000
	(0.021)	(0.026)	(0.023)	(0.018)	(0.018)	(0.018)	(0.013)	(0.017)	(0.015)	(0.011)	(0.011)	(0.014)
Years of schooling of the individual^2	−0.004***	−0.004***	−0.005***	−0.004***	−0.003***	−0.002**	0.001	−0.002*	0.000	0.000	0.000	0.002***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
If household is located on a urban zone (d)	0.557***	0.713***	0.641***	0.549***	0.395***	0.293***	0.239***	0.243***	0.231***	0.223***	0.146***	0.173***
	(0.075)	(0.076)	(0.083)	(0.062)	(0.053)	(0.068)	(0.047)	(0.085)	(0.054)	(0.040)	(0.035)	(0.049)
Labour characteristics variables												
If the labour is informal (d)	−0.368***	−0.359***	−0.363***	−0.322***	−0.418***	−0.365***	−0.169***	−0.177***	−0.165***	−0.169***	−0.221***	−0.183***
	(0.074)	(0.083)	(0.076)	(0.064)	(0.075)	(0.059)	(0.034)	(0.053)	(0.039)	(0.032)	(0.024)	(0.028)
Sector												
Mining and Quarrying (d)	0.609***	1.092*	0.713	0.457	−0.003	1.036	0.664***	0.760***	0.500***	0.590***	0.843***	0.770***
	(0.186)	(0.637)	(0.544)	(0.510)	(0.171)	(1.621)	(0.134)	(0.160)	(0.082)	(0.138)	(0.266)	(0.062)
Manufacturing and Public Utilities (d)	0.205*	0.444**	0.143	0.303***	0.123	0.429***	0.267***	0.377***	0.238***	0.219***	0.282***	0.215***
	(0.106)	(0.174)	(0.136)	(0.099)	(0.139)	(0.093)	(0.059)	(0.146)	(0.070)	(0.053)	(0.050)	(0.043)
Construction (d)	0.580***	0.282*	0.341*	0.542*	0.746***	0.729***	0.286***	0.379**	0.340***	0.244***	0.280***	0.207***
	(0.180)	(0.163)	(0.184)	(0.304)	(0.094)	(0.182)	(0.052)	(0.149)	(0.061)	(0.047)	(0.054)	(0.046)
Wholesale and Retail, Hotels and Restaurants (d)	0.089	−0.044	−0.109	0.167*	0.084	0.251***	0.241***	0.262**	0.118*	0.169***	0.261***	0.289***
	(0.086)	(0.140)	(0.130)	(0.093)	(0.072)	(0.083)	(0.051)	(0.123)	(0.071)	(0.045)	(0.049)	(0.063)
Transport, Storage, and Communications (d)	0.253	0.160	−0.094	0.327*	0.075	0.688	0.254***	0.298**	0.206***	0.213***	0.264***	0.155***
	(0.180)	(0.162)	(0.217)	(0.185)	(0.129)	(0.696)	(0.058)	(0.138)	(0.077)	(0.052)	(0.036)	(0.054)
Finance, Insurance, and Real Estate (d)	0.115	−0.065	0.040	0.346***	0.165	0.326	0.032	0.014	0.096	0.087	0.083*	0.113
	(0.114)	(0.220)	(0.219)	(0.117)	(0.143)	(0.236)	(0.074)	(0.141)	(0.102)	(0.058)	(0.048)	(0.110)
Community, Social and Personal Services (d)	−0.061	−0.148	−0.243*	0.101	−0.070	0.042	0.002	−0.140	−0.080	0.028	0.094	0.090
	(0.086)	(0.134)	(0.132)	(0.090)	(0.080)	(0.080)	(0.064)	(0.131)	(0.111)	(0.067)	(0.078)	(0.082)
Occupation												
Managers, Professionals and Armed forces (d)	0.476***	0.544***	0.700***	0.457***	0.304***	0.293***	0.768***	0.837***	0.493***	0.666***	0.991***	1.070***
	(0.108)	(0.150)	(0.130)	(0.106)	(0.096)	(0.109)	(0.105)	(0.095)	(0.114)	(0.102)	(0.096)	(0.100)
Technicians and associates (d)	0.586***	0.526***	0.617***	0.505***	0.604***	0.532***	0.405***	0.444***	0.268***	0.321***	0.425***	0.671***
	(0.078)	(0.112)	(0.111)	(0.090)	(0.067)	(0.075)	(0.065)	(0.129)	(0.070)	(0.052)	(0.119)	(0.096)
Clerks (d)	0.279***	0.355***	0.455**	0.264***	0.259***	0.219***	0.325***	0.289***	0.231**	0.303**	0.414***	0.514***
	(0.075)	(0.104)	(0.179)	(0.076)	(0.058)	(0.083)	(0.079)	(0.078)	(0.090)	(0.126)	(0.035)	(0.076)
Service and sales worker (d)	0.200***	0.178*	0.272***	0.149***	0.151***	0.166**	0.247***	0.320***	0.255***	0.203***	0.161***	0.339***
	(0.053)	(0.101)	(0.077)	(0.057)	(0.042)	(0.069)	(0.046)	(0.073)	(0.058)	(0.044)	(0.022)	(0.064)
Skilled agricultural and fishery workers (d)	−0.104	−0.133	−0.367**	−0.156	−0.161	−0.042	−0.121**	−0.328***	−0.463***	−0.266***	−0.029	0.166***
	(0.101)	(0.121)	(0.146)	(0.112)	(0.103)	(0.141)	(0.055)	(0.112)	(0.064)	(0.043)	(0.042)	(0.053)

Continued on next page

Table 5 – Regressions by gender under full model, 2004 (*continued from previous page*)

	Females						Males					
	Heckman	Quantile regression					Heckman	Quantile regression				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
Craft and related trades worker (d)	−0.590***	−1.162***	−0.891***	−0.646***	−0.349**	−0.462***	0.082*	0.155	0.065	0.066*	0.025	0.152***
	(0.125)	(0.161)	(0.146)	(0.158)	(0.157)	(0.100)	(0.043)	(0.101)	(0.044)	(0.040)	(0.038)	(0.029)
Plant and machine operators and assemblers (d)	−0.688***	−1.592***	−1.248***	−0.661	−0.401**	−0.687***	0.186***	0.339***	0.198***	0.174***	0.151***	0.222***
	(0.199)	(0.308)	(0.135)	(0.458)	(0.169)	(0.142)	(0.047)	(0.086)	(0.055)	(0.039)	(0.032)	(0.036)
Size												
Small (d)	0.181**	0.565***	0.266	0.228**	0.003	−0.040	0.281***	0.498***	0.321***	0.209***	0.159***	0.194***
	(0.071)	(0.085)	(0.189)	(0.093)	(0.050)	(0.151)	(0.042)	(0.049)	(0.054)	(0.033)	(0.037)	(0.034)
Medium (d)	0.427***	0.985***	0.611***	0.362***	0.306***	0.129**	0.414***	0.631***	0.424***	0.304***	0.297***	0.300***
	(0.075)	(0.109)	(0.088)	(0.087)	(0.067)	(0.062)	(0.057)	(0.063)	(0.040)	(0.034)	(0.041)	(0.024)
Large (d)	0.749***	1.163***	0.902***	0.825***	0.577***	0.401***	0.553***	0.792***	0.517***	0.506***	0.435***	0.370***
	(0.092)	(0.102)	(0.145)	(0.069)	(0.045)	(0.075)	(0.043)	(0.076)	(0.065)	(0.046)	(0.030)	(0.090)
Constant	7.092***	4.579***	5.569***	5.850***	6.452***	6.877***	5.351***	4.180***	5.456***	6.245***	5.978***	6.297***
	(0.313)	(0.613)	(0.591)	(0.425)	(0.373)	(0.457)	(0.177)	(0.716)	(0.468)	(0.334)	(0.284)	(0.412)
Observations	11,983	4,454	4,454	4,454	4,454	4,454	11,046	8,123	8,123	8,123	8,123	8,123
F test IMR	−1.996	8.152	9.252	9.945	8.568	8.831	−4.750	5.698	11.684	17.954	9.580	4.040
	[0.046]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.003]	[0.000]	[0.000]	[0.000]	[0.018]
F test education	15.926	5.486	10.210	12.769	11.849	19.901	38.527	16.206	22.870	37.793	49.304	50.569
	[0.000]	[0.004]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test sector	5.626	5.260	3.538	2.512	26.549	7.964	10.476	9.602	10.458	6.226	14.190	24.085
	[0.000]	[0.000]	[0.001]	[0.014]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test occupation	16.555	15.833	31.260	9.486	15.089	18.728	15.187	12.969	14.398	16.707	34.396	30.697
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test firm size	23.035	46.959	20.501	50.348	61.650	12.368	63.953	47.618	42.326	58.982	81.077	53.924
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test model	41.787	262.733	103.551	96.371	75.520	85.464	86.527	121.034	257.133	122.018	326.904	196.139
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

Note: (Ln.) Wages measured in constant 2014 Soles. Sample include individuals between 18 and 65 years. Observations weighted by expansion factor and VCE corrected according to survey's complex sample design. Standard errors in parenthesis and p-values of the F-test in brackets. (d)=Dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level.

Table 6 – Regressions by gender under full model at selected quantiles, 2014

	Females						Males					
	Heckman	**Quantile regression**					**Heckman**	**Quantile regression**				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
Individual level variables												
Age of the individual	−0.037***	−0.026	−0.025**	0.020**	0.042***	0.060***	0.034***	−0.003	−0.008	−0.013	−0.011	−0.016
	(0.006)	(0.016)	(0.011)	(0.009)	(0.008)	(0.009)	(0.004)	(0.020)	(0.012)	(0.010)	(0.010)	(0.012)
Age of the individual^2	0.000***	0.000	0.000**	0.000**	0.000***	−0.001***	0.000***	0.000	0.000	0.000	0.000	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling of the individual	0.043***	0.053***	0.061***	0.050***	0.025***	0.007	−0.004	0.000	−0.002	−0.020***	−0.023***	−0.029***
	(0.010)	(0.015)	(0.011)	(0.009)	(0.009)	(0.010)	(0.008)	(0.014)	(0.009)	(0.007)	(0.008)	(0.007)
Years of schooling of the individual^2	−0.001*	−0.003***	−0.003***	−0.002***	0.000	0.001**	0.002***	0.001	0.001**	0.002***	0.003***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
If household is located on a urban zone (d)	0.385***	0.564***	0.578***	0.371***	0.206***	0.127***	0.138***	0.231***	0.164***	0.102***	0.119***	0.128***
	(0.035)	(0.053)	(0.037)	(0.030)	(0.027)	(0.033)	(0.022)	(0.034)	(0.023)	(0.018)	(0.019)	(0.021)
Labour characteristics variables												
If the labour is informal (d)	−0.206***	−0.308***	−0.259***	−0.205***	−0.194***	−0.190***	−0.199***	−0.292***	−0.231***	−0.188***	−0.147***	−0.142***
	(0.028)	(0.046)	(0.031)	(0.026)	(0.025)	(0.028)	(0.016)	(0.030)	(0.019)	(0.014)	(0.017)	(0.018)
Sector												
Mining and Quarrying (d)	0.821***	0.882***	0.760**	0.748***	0.616***	0.783	0.626***	0.406***	0.547***	0.662***	0.632***	0.618***
	(0.126)	(0.113)	(0.350)	(0.147)	(0.071)	(0.791)	(0.051)	(0.094)	(0.060)	(0.036)	(0.034)	(0.053)
Manufacturing and Public Utilities (d)	0.109*	0.226*	0.000	−0.030	−0.077	0.113*	0.137***	0.070	0.075**	0.124***	0.154***	0.105***
	(0.064)	(0.124)	(0.069)	(0.052)	(0.066)	(0.065)	(0.032)	(0.050)	(0.036)	(0.028)	(0.032)	(0.035)
Construction (d)	0.212***	0.309**	0.172*	0.137*	0.156	0.338***	0.294***	0.286***	0.313***	0.350***	0.309***	0.257***
	(0.074)	(0.153)	(0.095)	(0.073)	(0.113)	(0.064)	(0.027)	(0.050)	(0.036)	(0.022)	(0.021)	(0.022)
Wholesale and Retail, Hotels and Restaurants (d)	−0.045	−0.058	−0.199***	−0.094*	−0.037	0.002	0.113***	0.067	0.084**	0.125***	0.145***	0.158***
	(0.048)	(0.106)	(0.067)	(0.053)	(0.044)	(0.049)	(0.031)	(0.059)	(0.034)	(0.028)	(0.033)	(0.027)
Transport, Storage, and Communications (d)	0.155**	0.071	0.029	0.086	0.064	0.230***	0.073**	0.126**	0.073**	0.069**	0.030	0.002
	(0.067)	(0.169)	(0.092)	(0.083)	(0.049)	(0.067)	(0.032)	(0.055)	(0.037)	(0.028)	(0.030)	(0.033)
Finance, Insurance, and Real Estate (d)	0.076	0.128	0.031	0.017	0.015	0.049	0.080**	−0.053	0.060	0.111***	0.131***	0.165***
	(0.055)	(0.111)	(0.071)	(0.055)	(0.054)	(0.059)	(0.036)	(0.061)	(0.042)	(0.031)	(0.035)	(0.051)
Community, Social and Personal Services (d)	−0.067	−0.029	−0.156**	−0.107*	−0.096**	−0.110**	−0.206***	−0.424***	−0.177***	−0.083***	−0.086***	−0.107***
	(0.048)	(0.114)	(0.068)	(0.055)	(0.044)	(0.052)	(0.040)	(0.101)	(0.042)	(0.032)	(0.030)	(0.036)
Occupation												
Managers, Professionals and Armed forces (d)	0.440***	0.578***	0.425***	0.325***	0.337***	0.475***	0.648***	0.613***	0.580***	0.599***	0.724***	0.802***
	(0.054)	(0.068)	(0.051)	(0.051)	(0.063)	(0.048)	(0.051)	(0.078)	(0.045)	(0.040)	(0.052)	(0.036)
Technicians and associates (d)	0.505***	0.682***	0.567***	0.438***	0.392***	0.530***	0.337***	0.322***	0.298***	0.285***	0.344***	0.457***
	(0.044)	(0.060)	(0.061)	(0.029)	(0.051)	(0.058)	(0.031)	(0.057)	(0.029)	(0.026)	(0.032)	(0.050)
Clerks (d)	0.368***	0.634***	0.449***	0.351***	0.303***	0.303***	0.293***	0.271***	0.262***	0.220***	0.292***	0.299***
	(0.038)	(0.060)	(0.046)	(0.041)	(0.035)	(0.036)	(0.032)	(0.058)	(0.030)	(0.028)	(0.040)	(0.039)
Service and sales worker (d)	0.335***	0.430***	0.461***	0.327***	0.239***	0.290***	0.235***	0.220***	0.197***	0.168***	0.212***	0.272***

Continued on next page

Table 6 – Regressions by gender under full model, 2014 (*continued from previous page*)

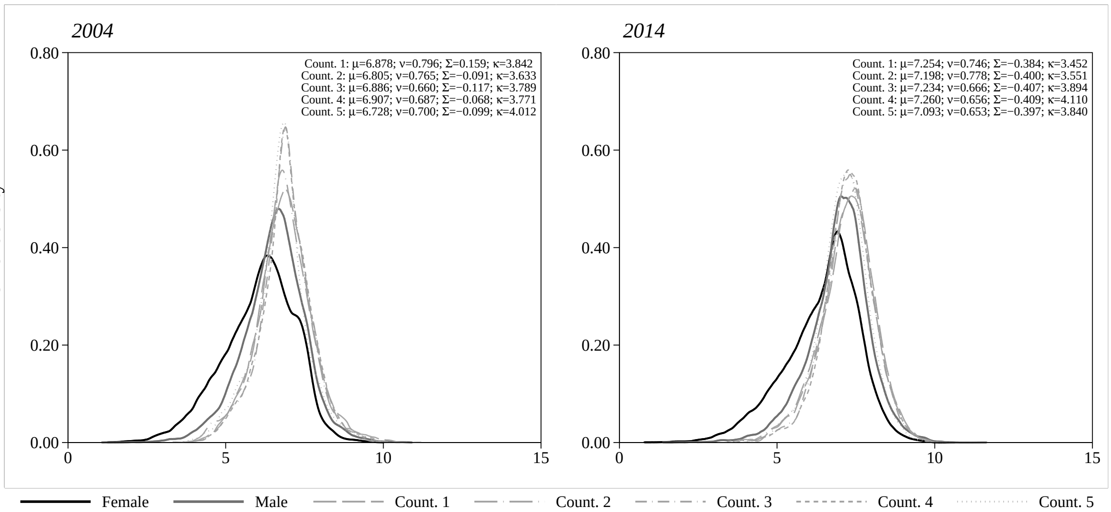
	Females						Males					
	Heckman	**Quantile regression**					**Heckman**	**Quantile regression**				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
	(0.028)	(0.052)	(0.038)	(0.028)	(0.024)	(0.028)	(0.028)	(0.057)	(0.028)	(0.030)	(0.031)	(0.039)
Skilled agricultural and fishery workers (d)	−0.274***	−0.132	−0.294***	−0.393***	−0.397***	−0.287***	−0.215***	−0.562***	−0.469***	−0.312***	−0.091***	0.106***
	(0.053)	(0.111)	(0.070)	(0.053)	(0.049)	(0.060)	(0.027)	(0.049)	(0.035)	(0.024)	(0.025)	(0.026)
Craft and related trades worker (d)	−0.314***	−0.710***	−0.404***	−0.172**	−0.080	−0.190***	0.142***	0.131***	0.151***	0.133***	0.133***	0.171***
	(0.065)	(0.107)	(0.080)	(0.077)	(0.060)	(0.058)	(0.023)	(0.048)	(0.029)	(0.019)	(0.023)	(0.027)
Plant and machine operators and assemblers (d)	−0.404***	−1.218***	−0.687**	−0.173	0.038	−0.214*	0.236***	0.297***	0.262***	0.215***	0.239***	0.242***
	(0.102)	(0.188)	(0.277)	(0.186)	(0.111)	(0.122)	(0.026)	(0.039)	(0.023)	(0.023)	(0.026)	(0.032)
Size												
Small (d)	0.240***	0.918***	0.528***	0.263***	0.083**	−0.029	0.204***	0.446***	0.276***	0.180***	0.127***	0.076**
	(0.035)	(0.068)	(0.040)	(0.032)	(0.034)	(0.037)	(0.026)	(0.035)	(0.025)	(0.019)	(0.020)	(0.032)
Medium (d)	0.427***	1.069***	0.673***	0.453***	0.226***	0.194***	0.279***	0.568***	0.343***	0.242***	0.166***	0.112***
	(0.039)	(0.051)	(0.040)	(0.035)	(0.034)	(0.041)	(0.022)	(0.039)	(0.023)	(0.018)	(0.023)	(0.022)
Large (d)	0.548***	1.161***	0.829***	0.605***	0.432***	0.315***	0.443***	0.734***	0.500***	0.364***	0.328***	0.323***
	(0.036)	(0.057)	(0.049)	(0.033)	(0.038)	(0.039)	(0.023)	(0.035)	(0.021)	(0.020)	(0.022)	(0.026)
Constant	7.001***	5.612***	6.396***	6.097***	6.149***	6.055***	5.891***	6.065***	6.597***	7.154***	7.346***	7.705***
	(0.132)	(0.420)	(0.297)	(0.221)	(0.216)	(0.243)	(0.091)	(0.459)	(0.284)	(0.223)	(0.231)	(0.273)
Observations	26,335	15,907	15,907	15,907	15,907	15,907	27,514	23,559	23,559	23,559	23,559	23,559
F test IMR	−4.360	24.540	38.035	39.722	61.490	17.489	−6.653	17.954	31.453	38.737	34.614	19.858
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test education	29.312	6.046	15.596	20.255	39.704	40.077	75.854	8.842	20.563	73.971	90.010	115.536
	[0.000]	[0.002]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test sector	12.085	72.455	8.008	8.828	19.234	13.434	58.824	15.923	41.273	100.131	89.254	44.793
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test occupation	43.951	45.180	42.904	45.058	32.809	39.675	41.062	34.011	63.145	67.057	45.148	77.482
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test firm size	84.298	165.400	121.935	120.780	45.656	29.073	133.501	148.906	196.616	129.640	74.057	50.380
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test model	134.281	419.330	214.408	387.273	120.022	87.471	272.184	362.503	328.523	306.536	205.266	227.326
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

Note: (Ln.) Wages measured in constant 2014 Soles. Sample include individuals between 18 and 65 years. Observations weighted by expansion factor and VCE corrected according to survey's complex sample design. Standard errors in parenthesis and p-values of the F-test in brackets. (d)=Dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level.

5.2 Decompositions considering alternative models

The set of counterfactual densities generated under the MM decomposition applying the extension for sample selection correction is shown in Figure 4[26]. Within each year, the 5 counterfactual densities shown, built after considering alternative models, have similar moments. For the year 2004, the counterfactual densities show a more leptokurtic shape than the observed densities (between 3.6 and 4), for it has a higher concentration of values around the mean of the distribution (which is around 6.9 for the densities under the first four models and 6.7 for the density built after the model 5 which includes all the covariates). Their standard deviations are fairly similar and the counterfactuals have a less negative skewness (ranging from -0.11 to -0.06) than the observed distributions (-0.47 for males and -0.31 for females). The skewness and kurtosis tests suggest to reject the null hypothesis that these distributions have similar higher order moments than those of a normal distribution. Considering the counterfactual densities for year 2014, they also show a higher kurtosis than those empirical distributions (between 3.5 and 4.1) and also a higher mean (ranging from 7.09 to 7.25). Standard deviations are fairly similar (around 0.81 except for the density estimated under model 2) and also have a less negative skewness (around -0.39) than the observed distributions (-0.66 for females and -0.53 for males). Normality is rejected in each of these 5 counterfactual distributions according to the skewness and kurtosis test.

Figure 4 – Observed and counterfactual densities under alternative models, 2004 and 2014



Note: Counterfactual distributios calculated taking covariants of females and coefficients of males under alternate models. Model 1 corresponds to the model with only basic characteristics; the remaining models are defined as in subsection 5.1. μ represents the mean; ν, the variance; Σ, the skewness and κ, the kurtosis.
Elaborated by the author based on INEI − National Household Survey (2004−2014)

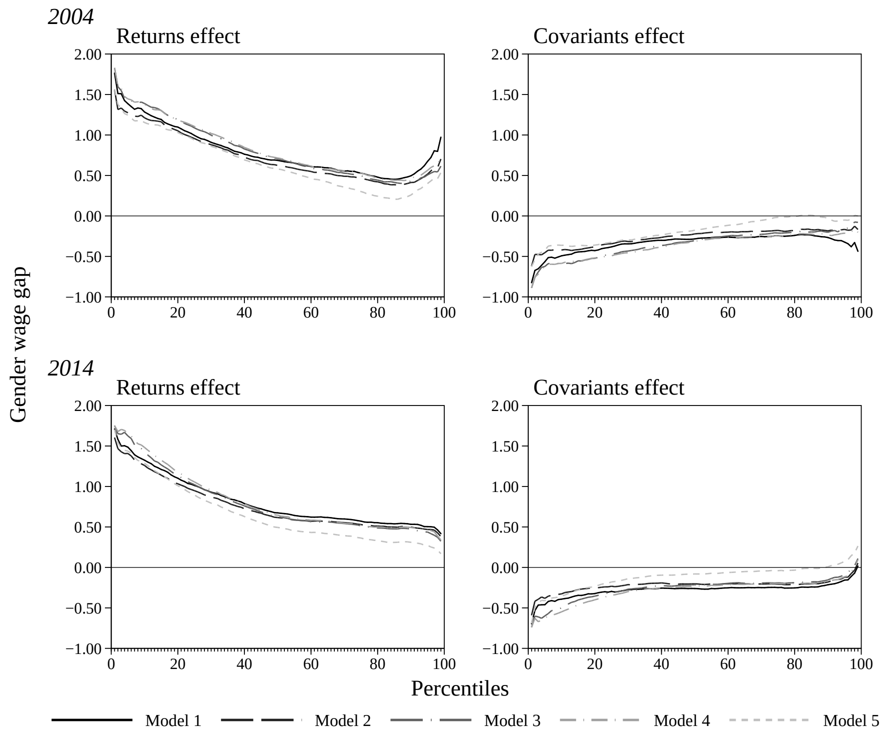
[26]Counterfactual densities estimated without accounting for sample selection correction is shown in A6 in Appendix A. These densities show very similar characteristics than those shown here although these latter have a higher kurtosis than those uncorrected.

These densities were used to decompose the gender wage gap a different percentiles of the distribution within each year following equation 9. Before turning to this, it is useful to consider a graphic representation of how log wage densities differ in absolute terms between males and females and how this can be accounted by the covariates and returns effect (this is shown in Figures A7 and A8 in Appendix A). For both years, the figures show that the biggest differences occur around the centre of the density; this holds for the different models considered. The male-female difference is negative between the 4.5 and 5.5 interval and turns positive at the upper parts of the distribution. Decomposition of this difference in terms of its two components reveals that returns effect mimic the total difference but it shows a notably higher advantage of males at the upper part of the distribution than females. In contrast, the covariants effects shows an inverse pattern: the difference between males and females on this effect is positive at the lower part of the distribution and negative at higher parts. In other words, both terms have the opposite effect and tend to offset the effect exerted by the other.

Focusing, in the decomposition results (Figure 5)[27], there are 4 remarkable regularities. First, the returns effect, associated with discrimination against women, is positive at every single point of the distribution and shows a decreasing pattern. Only in 2004, approximately at the 80th percentile, it increases again at a high speed. Second, the covariants effect has an offsetting negative effect, suggesting that differences in credentials favour women at different parts of the wage distribution. This increases as we move from the bottom to the top of the distribution and only in 2014 it increases at a high pace from the 90 percentile onwards. Third, results are similar in 2004 and 2014 in terms of their basic descriptions (decreasing returns effects, increasing covariants effects). Fourth, these conclusions are robust under the different models considered, suggesting that, independently of the way that we choose how to model log wages, returns effect accounts for the most part of the wage gap and covariant effect offset this influence.

[27] Figure A9 and Table A6 in Appendix A show the decomposition results without considering the sample selection correction. Note that the basic story is basically the same, although the effect of the two components is, in general, lower. Yet, the total values of the gaps (after adding the returns are covariants effect) are similar.

Figure 5 – Decomposition of the gender wage gap on covariate and returns effects at each percentile under alternative models across percentiles, 2004 and 2014



Note: Model 1 corresponds to the model with only basic characteristics; the remaining models are defined as in subsection 5.1.
Elaborated by the author based on INEI − National Household Survey (2004−2014)

Table 7 presents the values of the two components in detail for different percentiles. For 2004, the returns effect ranges from 1.25 to 1.45 at the 5th percentile (in other words, this effect itself results on a higher wage for males at the 5th percentile which can be as low 250% and as high as 326%), from 0.57 to 0.7 at the median and from 0.26 to 0.47 at the 90th percentile. The coefficient effect ranges from -0.625 to -0.425 at the 5th percentile (in other words, this effect itself results on a lower wage for males which can be as low as 36% and high as 46%), from -0.313 to -0.177 at the median and from -0.27 to -0.03 at the 90th percentile. Note that these two effects are lower when we consider the more complete model. For 2014, the returns effect ranges from 1.41 and 1.65 at the 5th percentile (put differently, this effect itself results on a higher wage for males at the 5th percentile which can be as low as 309% and as high as 420%), from 0.494 to 0.674 at the median and from 0.311 to 0.532 at the 90th percentile. The coefficient effect ranges from -0.55 to -0.33 at the 10th percentile (put differently, this effect itself results on a lower wage for males at the 10th percentile which can be as low as 28% and as high as 42%), from -0.26 to -0.08 at the median and from -0.21 to 0.01 at the 9th decile. For the values shown in the Table, the values of the two effects are higher in 2004 than in 2014 and, in general, the highest part of the gender wage gap is explained mainly by the discrimination factor which favour men and goes against females.

Table 7 – Decomposition of the gender wage gap on covariate and returns effects under alternative models at selected percentiles, 2004 and 2014

	Model 1		Model 2		Model 3		Model 4		Model 5	
	C	R	C	R	C	R	C	R	C	R
2014										
5	−0.566	1.387	−0.453	1.275	−0.625	1.446	−0.624	1.445	−0.425	1.246
10	−0.493	1.284	−0.421	1.212	−0.598	1.389	−0.587	1.378	−0.363	1.154
25	−0.382	0.994	−0.344	0.956	−0.475	1.087	−0.493	1.105	−0.325	0.937
50	−0.284	0.686	−0.222	0.624	−0.304	0.706	−0.313	0.715	−0.177	0.579
75	−0.245	0.529	−0.179	0.463	−0.212	0.495	−0.246	0.530	−0.018	0.301
90	−0.267	0.496	−0.184	0.413	−0.202	0.431	−0.232	0.461	−0.029	0.258
95	−0.323	0.674	−0.166	0.517	−0.156	0.507	−0.213	0.563	−0.050	0.401
2014										
5	−0.460	1.486	−0.382	1.408	−0.599	1.625	−0.624	1.650	−0.411	1.437
10	−0.392	1.325	−0.327	1.260	−0.497	1.431	−0.548	1.482	−0.354	1.288
25	−0.296	1.012	−0.232	0.948	−0.306	1.022	−0.340	1.056	−0.179	0.896
50	−0.260	0.674	−0.203	0.616	−0.214	0.628	−0.232	0.645	−0.080	0.494
75	−0.248	0.572	−0.206	0.530	−0.191	0.514	−0.198	0.522	−0.040	0.364
90	−0.214	0.532	−0.180	0.499	−0.154	0.473	−0.175	0.493	0.008	0.311
95	−0.156	0.504	−0.120	0.468	−0.089	0.437	−0.114	0.463	0.075	0.273

Note: C=Covariates effect, R=Returns effect. Model 1 corresponds to the model with only basic characteristics; the remaining models are defined as in subsection 5.1.
Elaborated by the author based on INEI - National Household Survey (2004-2014)

6 Conclusions and policy recommendations

This study applied the Machado and Mata (2005) decomposition method in order to decompose the (log) wage gender gaps between males and females in that part attributable to differences in characteristics and that part attributable to differences in returns to these characteristics, being the latter usually considered as the portion of the gap due to discrimination against women. Unlike the previous studies for Peru and, in general, for Latin America, we take into account sample selection with the extension of Albrecht et al. (2003) and apply this decomposition to years 2004 and 2014 in order to assess how inequality and its components changed during this period.

Results here suggest that the raw (unconditional) gender wage gap shows a decreasing tendency as we move to the upper parts of the distribution (excepting year 2004 were they increase beyond the 85th percentile suggesting that at the higher parts of the distribution the advantage of men increases again although slightly), and it is statistically different from zero at any point of the distribution. Also, unconditional gaps are higher in 2004 than in 2014. Conditioning the wages on a set of individual and labour characteristics, we find that the mean gender wage gaps and those at the different quantiles experienced a generalized increase between those years which favours males. However, gaps at the very top of the distribution are not significant in 2004. A strong sticky floor effect is found in 2004 and a sticky floor as well as a glass ceiling effect is found in 2014. Adding covariates to the simplest model leads to a slight increase in the mean gap for both years as well as in lower gaps for the bottom of the distribution; from the 50 percentile onwards this results in non-significant differences in 2004. If we allow differences in the returns of the characteristics of males and females, we find that these are different in terms of education (higher returns for males), returns to urbanization (higher returns for females). Informality exerts a similar burden for the wage of males and females in 2014 but not in 2004 (when it was higher for females) and returns for those located in bigger firm sizes are higher for females. Quantile regression estimates show that returns to the different characteristics vary throughout the wages distribution.

Results of the decomposition of the gender wage gap shows that the effect associated with discrimination against women is positive at every point of the distribution and decreases as we move to the top (excepting 2004 where it increases again at a high speed once we go beyond the 80th percentile). Also, the covariants effect has an negative effect which offsets the influence of the returns effect. This increases as we move from the bottom to the top of the distribution and only in 2014 it grows at a high pace from the 90 percentile onwards. Results are similar in 2004 and 2014 in terms of their basic descriptions (although for key percentiles the values of the effects are higher in 2004) and, importantly, these are remarkably similar under the five specifications chosen to model the wages.

One particular problem that this study finds is the identification of the constant and the tremendous computational burden that implies applying the SLS regression method (which justified the use of a probit model instead). Future studies which analyse labour market outcomes should try to deal with the sample selection using this most (theoretically) correct approach. Another limitation is that we are considering that Peruvian males and females face an endogenous choice of participating in the labour market and that some attributes of their jobs, such as informality, is exogenous. This is rather a non-realistic assumption because, given the high informality rates and its pervasiveness, it is expected that decision to participate in the labour market as formal or informal is, instead, endogenous. Whether it is done as a simultaneous decision problem or as a decision process done in different stages only adds complexity to the estimation methods but does not change the fact that it is necessary to consider this in future researches. Furthermore, we make no account for the difference of people participating into the labour market as full time workers or part time workers. Admittedly this can make a difference in the results (presumably at the bottom end) for we are considering only monthly wages which tend to hide the differences in hourly wages for those who work as part time workers and for those who as work full time workers. Finally, this study analyses thoroughly the gender wage gasp for two years and compares the results found within each of these. Notwithstanding, this does not allow to unsnarl the causes underlying this change: is it because of a change in the inequality within groups? due to a change in labour force composition? due to variations in the minimum wage? Adopting the between-year decomposition proposed by Autor et al. (2005) would allow us to decompose this inter-year change and to provide a better assessment of the behaviour of inequality in time.

Despite of these limitations, we believe that the results presented here are robust enough to provide solid evidence that the gender wage gap is a problem which, despite of all the efforts undertaken by the Peruvian government, still remains at a high level during the period 2004-2014. In fact, discrimination against women is the most important factor driving these gaps (regardless of taking corrected or uncorrected estimates), which casts doubt on the aggregate efficiency of policies put in place in order to alleviate this problem. Furthermore, glass ceilings and sticky floors are still present during this ten year lapse, being the latter a more important problem for policy makers since it involves people whose wages are low enough to keep them within a poverty situation and to experience vulnerability to macroeconomic and idiosyncratic shocks. A solution to this issue would require a more coordinated effort than what has been taken so far and to stop considering as the only indicator of the gender wage gap the difference in mean wages because, as has been repeatedly stressed here, this hides an large part of the complex portrait of inequality.

References

Adams, R. (2003). *Economic Growth, Inequality, and Poverty. Findings from a New Data Set*. Policy Research Working Paper 2972. The World Bank. Poverty Reduction and Economic Management Network. Poverty Reduction Group.

Aktas, A. and Uysal, G. (2012). *Explaining the Gender Wage Gap in Turkey Using the Wage Structure Survey*. BETAM working paper series 005.

Albrecht, J., Björklund, A., and Vroman, S. (2003). Is there a glass ceiling in Sweden? *Journal of Labor Economics*, 21(1).

Albrecht, J., VanVuuren, A., and Vroman, S. (2009). Counterfactual distributions with sample selection adjustments: Econometric theory and an application to The Netherlands. *Labour Economics*, 16(4).

Aliaga, L. (2010). *Informal Economy Budget Analysis in Peru and Metropolitan Lima*. Wiego Working Paper No 13. Cambridge, MA: Harvard Kennedy School.

Andrews, D. and Schafgans, M. (1998). Semiparametric estimation of the intercept model. *Review of Economic Studies*, 65:497–517.

Arceo-Gomez, E. and Campos-Vazquez, R. (2014). Evolution of the gender wage gap in México. *El Trimestre económico*, 81(323):619–653.

Arulampalam, W., Booth, A. L., and Bryan, M. L. (2007). Is there a glass ceiling over Europe? exploring the gender pay gap across the wage distribution. *Industrial and Labor Relations Review*, 60(2):163–186.

Atal, J., Ñopo, H., and Winder, N. (2009). *New Century, Old Disparities Gender and Ethnic Wage Gaps in Latin America*. American Development Bank, Research Department.

Autor, D., Katz, L., and Kearney, M. (2005). *Rising wage inequality: the role of composition and prices*. NBER Working Paper No. 11628.

Bassett, G. and Koenker, R. (1982). An empirical quantile function for linear models with i.i.d. errors. *Journal of the American Statistical Association*, 77:407–415.

Bassett, G. and Koenker, R. (1986). Strong consistency of regression quantiles and related empirical process. *Econometric Theory*, 2:191–201.

Beteta, E. and Del Pozo, J. M. (2016). Aplicación de una metodología para el análisis de las desigualdades socioeconómicas en acceso a servicios de salud y educación en Perú en 2005-2012. In Francke, P. and Rodriguez, J., editors, *Políticas públicas en un contexto de crecimiento*. Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú.

Buchinsky, M. (1998). The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. *Journal of Applied Econometrics*, 13:1–30.

Buchinsky, M. (2001). Quantile regression with sample selection: Estimating women's return to education in the u.s. *Empirical Economics*, 26:87–113.

Castillo, W. (2011). *Discriminación salarial por género en el Perú : 2003-2009*. Thesis for the degree of magister in Applied Economics. Universidad De Chile - Facultad De Ciencias Físicas y Matemáticas.

CEPAL, FAO, ONU, PNUD, and OIT (2013). *Trabajo decente e igualdad de género. Políticas para mejorar el acceso y la calidad del empleo de las mujeres en América Latina y el Caribe*. Santiago de Chile: Comisión Económica para América Latina y el Caribe (CEPAL).

CIES (2011). *La investigación económica y social en el Perú. Balance 2007-2011 y agenda 2012-2016*. Lima: Consorcio de Investigación Económica y Social (CIES).

Coppola, A. and Calvo-Gonzalez, O. (2011). *Higher Wages, Lower Pay. Public vs. Private Sector Compensation in Peru*. Policy Research Working Paper 5858. The World Bank Latin America and the Caribbean Region.

Deaton, A. (1997). *The Analysis of Household Surveys: The Microeconometric Approach to Development Policy*. Baltimore MD: Johns Hopkins University Press.

Deininger, K. and Squire, L. (1998). New ways of looking at old issues: inequality and growth. *Journal of Development Economics*, 57(259-287).

DiNardo, J., Fortin, N., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5).

Dolton, P. and Kidd, M. (1994). Occupational access and wage discrimination. *Oxford Bulletin of Economics and Statistics*, 56(4).

Dolton, P. and Makepeace, G. (1987). Interpreting sample selection effects. *Economics Letters*, 24:373–379.

Firpo, S., Fortin, N., and Lemieux, T. (2007). *Decomposing Wage Distributions using Recentered Influence Function Regressions*. Mimeo, University of British Columbia.

Fortin, N., Lemieux, T., and Firpo, S. (2010). *Decomposition methods in economics*. National Bureau Of Economic Research .Working Paper 16045.

Freije, S. (2002). *Informal Employment in Latin America and the Caribbean: Causes, Consequences and Policy Recommendations*. Inter-American Development Bank. Sustainable Development Department, Social Development Division, Labor Markets Policy Briefs Series.

Garavito, C. (2011). Desigualdad en los ingresos: género y lengua materna. In Iguíñiz, J. and León, J., editors, *Desigualdad distributiva en el Perú: dimensiones*. Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú.

Gronau, R. (1973). *Wage comparisons - A selectivity Bias*. National Bureau Of Economic Research. Working Paper 13.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Ichimura, H. (1993). Semiparametric Least Squares (sls) and weighted SLS estimation of single-index models. *Journal of Econometrics*, 58:71–120.

ILO (2012). *Global Employment Trends for Women*. Geneva: International Labour Organization.

ILO (2014). *Trends in informal employment in Peru: 2004 - 2012*. ILO. Notes on formalization. Regional Office for Latin America and the Caribbean.

INEI (2014a). *Perú: Brechas de género, 2001-2013. Avances hacia la igualdad entre hombres y mujeres*. Lima: Instituto Nacional de Estadística e Informática (INEI).

INEI (2014b). *Perú: Evolución de los Indicadores de los Objetivos de Desarrollo del Milenio al 2012*. Lima: Instituto Nacional de Estadística e Informática (INEI).

Jaramillo, M., Ñopo, H., and Díaz, J. J. (2007). La investigación sobre el mercado laboral peruano: Instituciones, capacitación y grupos desfavorecidos. In GRADE, editor, *Investigación, políticas y desarrollo en el Perú*. Lima: Grupo de Análisis para el Desarrollo.

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

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Koenker, R. and Hallock, K. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4):143 – 156.

Machado, J. and Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20(4):445–465.

Montes, J. (2007). *Evolution of the gender wage gap in Peru, 1997-2000*. Thesis for the degree of Master of Science submitted to the Office of Graduate Studies of Texas A&M University.

Mroz, T. (1987). The sensitvity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, 55(4):765–799.

MTPE (2006). *Informe Anual 2005. La Mujer en el Mercado Laboral Peruano*. Lima: Ministerio De Trabajo y Promoción Del Empleo (MTPE).

MTPE (2010). *Informe Anual 2009. La Mujer en el Mercado Laboral Peruano*. Lima: Ministerio De Trabajo y Promoción Del Empleo (MTPE).

MTPE (2012). *Políticas Nacionales de Empleo*. Lima: Ministerio De Trabajo y Promoción Del Empleo (MTPE).

MTPE (2014). *Informe Anual 2012. La Mujer en el Mercado Laboral Peruano*. Lima: Ministerio De Trabajo y Promoción Del Empleo (MTPE).

MTPE (2015). *Informe Anual 2014. La Mujer en el Mercado Laboral Peruano*. Lima: Ministerio De Trabajo y Promoción Del Empleo (MTPE).

Newey, W. (1988). *Two step series estimation of sample selection model*. Presented at the 1988 European meeting of the Econometric Society.

Nguyen, B., Albrecht, J., Vroman, S., and Westbrook, M. (2007). *A quantile regression decomposition of urban - rural inequality in Vietnam*, volume 83.

Oaxaca, R. (1973). Male-female differentials in urban labor markets. *International Economic Review*, 14(693-709).

OIT and PNUD (2009). *Trabajo y Familia: Hacia nuevas formas de conciliación con corresponsabilidad social*. Santiago de Chile: Oficina Internacional del Trabajo.

Pacheco, E. (2013). *Gender Wage Gap in Urban Nicaragua: Evidence from Decomposition Analysis*. Master in Economic Development and Growth. Lund University.

PCM (2013). *Perú: tercer informe nacional de cumplimiento de los Objetivos de Desarrollo del Milenio.* Lima: Presidencia del Consejo de Ministros (PCM), Sistema de Naciones Unidas en el Perú.

Persson, T. and Tabellini, G. (1994). Is inequality harmful for growth? *The American Economic Review*, 84(3):600–621.

Pham, T. and Reilly, B. (2007). *Ethnic Wage Inequality in Vietnam: Empirical Evidence from 2002*. MPRA Paper No. 6477.

Rica, S., Dolado, J., and Llorens, V. (2008). Ceilings or floors? gender wager gaps by education in Spain. *Journal of Population Economics*, 21(3):751–776.

Ross, K. and Peschiera, J. (2015). Explaining the peruvian growth miracle. In Santos, A. and Werner, A., editors, *Peru: Staying the Course of Economic Success*. Washington D.C.: International Monetary Fund.

Saavedra, J. (1999). *La dinámica del mercado de trabajo en el Perú antes y después de las reformas estructurales*. Serie Reformas Económicas, No 27. Comisión Económica para América Latina y el Caribe (CEPAL).

Saavedra, M., Moreno, H., and Hernández, Y. (2008). Caracterización de las mipymes en latinoamérica: un estudio comparativo. *Revista Internacional La Nueva Gestión Organizacional*, 4(7):57–74.

Salardi, P. (2012). *Wage Disparities and Occupational Intensity by Gender and Race in Brazil: An Empirical Analysis Using Quantile Decomposition techniques*. University of Sussex, Job market paper.

Távara, J., González de Olarte, E., and Del Pozo, J. M. (2014). Heterogeneidad estructural y articulación productiva en el Perú: evolución y estrategias. In Infante, R. and Chacaltana, J., editors, *Hacia un desarrollo inclusivo. El caso del Perú*. Santiago de Chile : Comisión Económica para América Latina y el Caribe (CEPAL).

Webb, R. (2013). *Conexión y despegue rural*. Lima: Fondo Editorial de la Universidad San Martín de Porres.

World Bank (2010). *Peru - Recent economic development in infrastructure investing in infrastructure as an engine for growth : spending more, faster and spending better. Vols. I and II*. Washington, D.C.: The World Bank.

World Bank (2015). *Peru: Building on Success. Boasting productivity for faster growth*. Washington, D.C.: The World Bank.

Yamada, G., Lavado, P., and Velarde, L. (2013). *Habilidades No Cognitivas y Brecha de Género Salarial en el Perú*. Banco Central de Reserva del Perú. DT. N° 2013-014. Working Paper series.

A Statistical appendix

Table A1 – Descriptive statistics of wages, 2004 and 2014

	2004				**2014**			
	Statistic	SE	UB	LB	Statistic	SE	UB	LB
Total								
Mean	1,094.51	27.79	1,040.01	1,149.00	1,460.00	15.76	1,429.10	1,490.90
Percentile								
p10	160.95	3.43	154.22	167.68	239.83	3.86	232.26	247.41
p25	365.42	6.83	352.03	378.82	563.49	7.14	549.49	577.48
p50	720.65	8.93	703.14	738.16	1,067.57	7.10	1,053.65	1,081.50
p75	1,264.71	16.42	1,232.52	1,296.91	1,772.69	15.01	1,743.27	1,802.11
p90	2,074.11	41.67	1,992.45	2,155.78	2,857.06	41.95	2,774.84	2,939.28
Gini	0.52	0.01	0.50	0.54	0.47	0.00	0.46	0.48
Female								
Mean	831.30	18.73	794.58	868.02	1,139.26	15.88	1,108.12	1,170.40
Percentile								
p10	104.12	3.24	97.78	110.46	143.38	4.34	134.87	151.89
p25	243.69	7.15	229.67	257.71	376.53	8.05	360.76	392.30
p50	553.85	12.45	529.45	578.24	836.75	11.58	814.05	859.45
p75	1,065.50	22.30	1,021.78	1,109.21	1,429.32	21.81	1,386.59	1,472.06
p90	1,722.82	30.88	1,662.30	1,783.34	2,340.56	42.57	2,257.13	2,423.99
Gini	0.51	0.01	0.50	0.53	0.49	0.00	0.48	0.50
Male								
Mean	1,249.65	39.78	1,171.66	1,327.64	1,679.66	20.07	1,640.32	1,719.00
Percentile								
p10	224.72	4.65	215.61	233.82	368.40	6.69	355.29	381.51
p25	455.45	5.98	443.72	467.18	745.83	7.16	731.80	759.86
p50	817.77	9.11	799.92	835.63	1,240.93	11.00	1,219.37	1,262.50
p75	1,354.12	19.32	1,316.25	1,391.99	1,957.93	15.93	1,926.71	1,989.16
p90	2,313.63	50.49	2,214.68	2,412.58	3,174.75	50.41	3,075.95	3,273.55
Gini	0.51	0.01	0.48	0.53	0.44	0.00	0.43	0.45

Note: (Ln.) Wages measured in constant 2014 Soles. Standard errors of coeficients corrected according to survey's complex sample design. UB and LB refers to upper bound and lower bound of the confidence intervals at 95% of significance.
Elaborated by the author based on INEI - National Household Survey (2004-2014)

Table A2 – Gender wage gaps under alternative models (uncorrected quantile regressions), 2004 and 2014

	OLS reg.	**Quantile regression**					**Obs.**
		10th	25th	50th	75th	90th	
Year 2004							
Observed	0.462	0.769	0.625	0.390	0.240	0.295	28,121
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 1 (Basic controls)	0.518	0.882	0.627	0.394	0.346	0.342	28,119
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 1 + Informality	0.435	0.733	0.487	0.362	0.281	0.243	14,244
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 2 + Sector	0.461	0.771	0.516	0.373	0.284	0.259	14,244
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 3 + Occupation	0.490	0.696	0.520	0.390	0.312	0.277	14,244
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 4 + Firm size	0.487	0.674	0.579	0.414	0.289	0.302	12,577
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Year 2014							
Observed	0.518	0.944	0.683	0.394	0.315	0.305	45,752
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 1 (Basic controls)	0.576	1.005	0.640	0.459	0.412	0.391	45,745
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 1 + Informality	0.510	0.810	0.518	0.408	0.379	0.341	45,745
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 2 + Sector	0.492	0.777	0.500	0.393	0.345	0.313	45,745
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 3 + Occupation	0.518	0.736	0.530	0.404	0.368	0.327	45,744
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	
Model 4 + Firm size	0.527	0.756	0.548	0.422	0.367	0.336	39,466
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	

Note: (Ln.) Wages measured in constant 2014 Soles. Gender dummy equals 1 if individual is male and 0 otherwise. Sample include individuals between 18 and 65 years. Observations weighted by expansion factor and VCE corrected according to survey's complex sample design. P-values of the gender coefficients shown in brackets. Observed model includes only the gender dummy; basic controls include gender, years of schooling (2nd degree polynomial), age (2nd degree polynomial) and a dummy of area of residence.

Elaborated by the author based on INEI - National Household Survey (2004-2014)

Table A3 – Quantile regressions by gender under basic model (uncorrected quantile regressions), 2004 and 2014

	Females						Males					
	OLS	**Quantile regression**					**OLS**	**Quantile regression**				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
2004												
Age of the individual	0.068***	0.063***	0.069***	0.068***	0.063***	0.089***	0.104***	0.145***	0.110***	0.084***	0.088***	0.088***
	(0.007)	(0.016)	(0.011)	(0.007)	(0.006)	(0.007)	(0.004)	(0.011)	(0.007)	(0.004)	(0.005)	(0.006)
Age of the individual^2	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.002***	−0.001***	−0.001***	−0.001***	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling of the individual	−0.003	−0.054**	−0.012	0.010	0.014	0.005	−0.001	0.012	0.011	0.001	−0.007	−0.042***
	(0.012)	(0.023)	(0.018)	(0.012)	(0.009)	(0.014)	(0.009)	(0.014)	(0.011)	(0.008)	(0.009)	(0.014)
Years of schooling of the individual^2	0.006***	0.009***	0.007***	0.005***	0.004***	0.005***	0.005***	0.004***	0.004***	0.004***	0.005***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
If household is located on a urban zone (d)	0.473***	0.654***	0.647***	0.491***	0.279***	0.347***	0.464***	0.610***	0.605***	0.453***	0.343***	0.355***
	(0.038)	(0.062)	(0.052)	(0.037)	(0.030)	(0.038)	(0.024)	(0.034)	(0.026)	(0.019)	(0.021)	(0.025)
Constant	3.799***	2.619***	3.040***	3.920***	4.643***	4.560***	3.650***	1.773***	3.105***	4.188***	4.544***	4.936***
	(0.142)	(0.293)	(0.224)	(0.141)	(0.120)	(0.143)	(0.090)	(0.208)	(0.141)	(0.084)	(0.099)	(0.123)
Observations	10,417	10,417	10,417	10,417	10,417	10,417	17,702	17,702	17,702	17,702	17,702	17,702
Pseudo R^2	0.245	0.097	0.137	0.161	0.171	0.154	0.303	0.151	0.173	0.175	0.185	0.207
F test education	458.461	77.472	199.405	530.639	466.076	168.187	327.714	215.568	301.323	639.157	423.324	282.288
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test model	320.651	85.339	182.082	396.183	296.997	153.207	519.021	355.777	517.151	772.503	476.098	327.104
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
2014												
Age of the individual	0.048***	0.045***	0.050***	0.045***	0.049***	0.057***	0.082***	0.128***	0.080***	0.065***	0.062***	0.068***
	(0.005)	(0.012)	(0.008)	(0.005)	(0.005)	(0.006)	(0.004)	(0.007)	(0.005)	(0.003)	(0.003)	(0.004)
Age of the individual^2	−0.001***	−0.001***	−0.001***	0.000***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling of the individual	−0.021***	−0.076***	−0.026**	0.000	−0.018**	−0.026***	−0.002	0.024*	0.018**	−0.001	−0.033***	−0.032***
	(0.008)	(0.019)	(0.011)	(0.008)	(0.008)	(0.010)	(0.006)	(0.013)	(0.009)	(0.007)	(0.007)	(0.006)
Years of schooling of the individual^2	0.006***	0.011***	0.008***	0.005***	0.005***	0.006***	0.004***	0.003***	0.003***	0.004***	0.006***	0.006***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
If household is located on a urban zone (d)	0.481***	0.627***	0.631***	0.500***	0.303***	0.266***	0.392***	0.666***	0.584***	0.384***	0.242***	0.156***
	(0.027)	(0.049)	(0.037)	(0.025)	(0.023)	(0.030)	(0.018)	(0.030)	(0.021)	(0.016)	(0.015)	(0.018)
Constant	4.532***	3.176***	3.808***	4.719***	5.369***	5.637***	4.543***	2.378***	4.019***	5.020***	5.662***	5.898***
	(0.098)	(0.251)	(0.167)	(0.095)	(0.094)	(0.123)	(0.075)	(0.152)	(0.101)	(0.072)	(0.070)	(0.089)
Observations	18,835	18,835	18,835	18,835	18,835	18,835	26,910	26,910	26,910	26,910	26,910	26,910
Pseudo R^2	0.226	0.104	0.134	0.141	0.143	0.144	0.237	0.140	0.139	0.129	0.140	0.166
F test education	994.362	206.191	841.226	1,220.328	709.207	449.583	818.505	284.718	581.753	752.840	892.890	895.282
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test model	659.423	194.836	668.987	892.444	489.436	295.691	802.239	443.593	710.878	742.401	652.581	577.879
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

Note: (Ln.) Wages measured in constant 2014 Soles. Sample include individuals between 18 and 65 years. Observations weighted by expansion factor and VCE corrected according to survey's complex sample design. Standard errors in parenthesis and p-values of the F-test in brackets. (d)=Dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level.

Table A4 – Regressions by gender under full model (uncorrected quantile regressions) at selected quantiles, 2004

	Females						Males					
	OLS	Quantile regression					OLS	Quantile regression				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
Individual level variables												
Age of the individual	0.061***	0.083***	0.072***	0.062***	0.052***	0.055***	0.106***	0.148***	0.112***	0.082***	0.080***	0.080***
	(0.010)	(0.016)	(0.014)	(0.010)	(0.009)	(0.009)	(0.006)	(0.012)	(0.008)	(0.006)	(0.005)	(0.005)
Age of the individual^2	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.002***	−0.001***	−0.001***	−0.001***	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling of the individual	0.066***	0.051**	0.054***	0.069***	0.055***	0.026	0.032***	0.052***	0.044***	0.025**	0.027***	0.006
	(0.016)	(0.025)	(0.018)	(0.017)	(0.016)	(0.016)	(0.012)	(0.016)	(0.014)	(0.010)	(0.010)	(0.013)
Years of schooling of the individual^2	−0.003**	−0.002	−0.002**	−0.003***	−0.002**	0.001	0.000	−0.002*	−0.001	0.000	0.000	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
If household is located on a urban zone (d)	0.489***	0.602***	0.558***	0.454***	0.348***	0.239***	0.088**	0.081	0.058	0.091***	0.062**	0.071**
	(0.060)	(0.071)	(0.069)	(0.055)	(0.059)	(0.069)	(0.043)	(0.060)	(0.046)	(0.030)	(0.025)	(0.028)
Labour characteristics variables												
If the labour is informal (d)	−0.331***	−0.336***	−0.406***	−0.301***	−0.435***	−0.391***	−0.165***	−0.185***	−0.165***	−0.185***	−0.221***	−0.159***
	(0.071)	(0.086)	(0.063)	(0.068)	(0.069)	(0.077)	(0.035)	(0.045)	(0.037)	(0.031)	(0.028)	(0.027)
Sector												
Mining and Quarrying (d)	0.715***	1.143***	0.881	0.628	0.092	0.993	0.662***	0.736***	0.533***	0.554***	0.861***	0.819***
	(0.154)	(0.321)	(0.719)	(1.101)	(0.430)	(1.816)	(0.130)	(0.159)	(0.120)	(0.122)	(0.247)	(0.063)
Manufacturing and Public Utilities (d)	0.186*	0.363*	0.110	0.316***	0.186	0.366***	0.257***	0.336***	0.257***	0.200***	0.263***	0.221***
	(0.107)	(0.208)	(0.083)	(0.114)	(0.170)	(0.124)	(0.061)	(0.130)	(0.083)	(0.052)	(0.051)	(0.042)
Construction (d)	0.534***	0.374	0.480*	0.567*	0.729***	0.586**	0.283***	0.411***	0.325***	0.221***	0.254***	0.230***
	(0.193)	(0.300)	(0.255)	(0.329)	(0.073)	(0.286)	(0.054)	(0.129)	(0.081)	(0.050)	(0.058)	(0.039)
Wholesale and Retail, Hotels and Restaurants (d)	0.066	0.010	−0.131	0.249**	0.083	0.258***	0.227***	0.293***	0.131	0.155***	0.232***	0.339***
	(0.087)	(0.133)	(0.097)	(0.100)	(0.066)	(0.094)	(0.052)	(0.098)	(0.080)	(0.050)	(0.052)	(0.073)
Transport, Storage, and Communications (d)	0.187	0.112	−0.079	0.383***	0.074	0.544	0.247***	0.365***	0.226***	0.184***	0.213***	0.131**
	(0.174)	(0.309)	(0.162)	(0.117)	(0.091)	(0.437)	(0.060)	(0.122)	(0.081)	(0.055)	(0.039)	(0.053)
Finance, Insurance, and Real Estate (d)	0.090	0.056	−0.015	0.455***	0.146*	0.316**	0.015	0.067	0.121	0.067	0.062	0.081
	(0.122)	(0.143)	(0.142)	(0.153)	(0.087)	(0.123)	(0.075)	(0.122)	(0.089)	(0.058)	(0.050)	(0.082)
Community, Social and Personal Services (d)	−0.092	−0.112	−0.232**	0.135	−0.038	0.034	−0.020	−0.151	−0.079	−0.004	0.056	0.071
	(0.090)	(0.129)	(0.102)	(0.096)	(0.068)	(0.093)	(0.065)	(0.109)	(0.111)	(0.069)	(0.077)	(0.090)
Occupation												
Managers, Professionals and Armed forces (d)	0.487***	0.477**	0.649***	0.384***	0.270**	0.314**	0.772***	0.820***	0.503***	0.675***	0.981***	1.098***
	(0.118)	(0.199)	(0.113)	(0.107)	(0.107)	(0.148)	(0.104)	(0.128)	(0.131)	(0.104)	(0.112)	(0.103)
Technicians and associates (d)	0.595***	0.520***	0.612***	0.488***	0.556***	0.542***	0.401***	0.404***	0.248***	0.321***	0.417***	0.640***
	(0.084)	(0.123)	(0.103)	(0.096)	(0.056)	(0.093)	(0.064)	(0.102)	(0.058)	(0.052)	(0.122)	(0.049)
Clerks (d)	0.318***	0.270	0.494***	0.284***	0.208***	0.179*	0.320***	0.192**	0.220**	0.283*	0.377***	0.530***
	(0.081)	(0.225)	(0.129)	(0.076)	(0.063)	(0.093)	(0.081)	(0.077)	(0.090)	(0.152)	(0.077)	(0.099)
Service and sales worker (d)	0.153***	0.132	0.236***	0.099*	0.145***	0.160**	0.259***	0.305***	0.218***	0.202***	0.173***	0.295***
	(0.057)	(0.106)	(0.083)	(0.053)	(0.047)	(0.069)	(0.047)	(0.059)	(0.048)	(0.044)	(0.030)	(0.108)
Skilled agricultural and fishery workers (d)	−0.100	0.011	−0.367***	−0.087	−0.141	0.011	−0.171***	−0.333***	−0.483***	−0.281***	−0.060	0.141***
	(0.097)	(0.130)	(0.119)	(0.113)	(0.099)	(0.136)	(0.055)	(0.092)	(0.073)	(0.043)	(0.043)	(0.048)

Continued on next page

Table A4 – Regressions by gender under full model (uncorrected quantile regressions), 2004 (*continued from previous page*)

	Females						**Males**					
	OLS	**Quantile regression**					**OLS**	**Quantile regression**				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
Craft and related trades worker (d)	−0.695***	−1.087***	−0.928***	−0.590***	−0.423**	−0.421***	0.091**	0.158*	0.054	0.073*	0.022	0.139***
	(0.120)	(0.198)	(0.139)	(0.180)	(0.194)	(0.127)	(0.044)	(0.096)	(0.051)	(0.040)	(0.038)	(0.041)
Plant and machine operators and assemblers (d)	−0.856***	−1.501***	−1.306***	−0.676*	−0.393*	−0.537	0.198***	0.320***	0.170***	0.185***	0.159***	0.249***
	(0.203)	(0.212)	(0.144)	(0.395)	(0.229)	(0.423)	(0.048)	(0.067)	(0.051)	(0.040)	(0.033)	(0.039)
Size												
Small (d)	0.259***	0.619***	0.246***	0.216**	0.022	−0.035	0.297***	0.509***	0.302***	0.194***	0.195***	0.203***
	(0.072)	(0.101)	(0.094)	(0.099)	(0.048)	(0.139)	(0.042)	(0.050)	(0.041)	(0.033)	(0.051)	(0.039)
Medium (d)	0.551***	1.151***	0.593***	0.404***	0.288***	0.131	0.423***	0.646***	0.411***	0.290***	0.273***	0.299***
	(0.077)	(0.082)	(0.076)	(0.088)	(0.061)	(0.086)	(0.054)	(0.073)	(0.046)	(0.044)	(0.052)	(0.034)
Large (d)	0.882***	1.349***	0.950***	0.847***	0.563***	0.375***	0.580***	0.809***	0.503***	0.478***	0.451***	0.370***
	(0.086)	(0.135)	(0.083)	(0.077)	(0.071)	(0.108)	(0.043)	(0.054)	(0.070)	(0.037)	(0.032)	(0.060)
Constant	4.229***	2.647***	3.729***	4.198***	5.235***	5.568***	3.816***	2.039***	3.478***	4.534***	4.881***	5.087***
	(0.216)	(0.332)	(0.268)	(0.222)	(0.195)	(0.203)	(0.127)	(0.263)	(0.181)	(0.127)	(0.097)	(0.114)
Observations	4,454	4,454	4,454	4,454	4,454	4,454	8,123	8,123	8,123	8,123	8,123	8,123
Pseudo R^2	0.313	0.188	0.195	0.184	0.188	0.211	0.372	0.222	0.225	0.215	0.228	0.274
F test education	17.419	3.168	5.365	13.841	11.918	19.726	30.575	13.120	21.471	34.393	45.931	53.093
	[0.000]	[0.042]	[0.005]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test sector	6.599	2.847	3.351	3.281	37.502	3.884	10.382	9.865	6.088	5.485	8.821	29.784
	[0.000]	[0.006]	[0.001]	[0.002]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test occupation	16.855	11.435	35.175	6.882	16.400	8.527	15.686	12.067	11.740	17.524	20.131	38.109
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test firm size	37.456	73.656	60.035	41.823	27.020	4.450	70.261	85.093	38.655	63.084	68.342	33.594
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.004]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test model	58.479	92.305	143.497	115.739	82.270	235.043	88.254	124.048	102.103	115.527	152.070	240.368
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

Note: (Ln.) Wages measured in constant 2014 Soles. Sample include individuals between 18 and 65 years. Observations weighted by expansion factor and VCE corrected according to survey's complex sample design. Standard errors in parenthesis and p-values of the F-test in brackets. (d)=Dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level.

Table A5 – Regressions by gender under full model (uncorrected quantile regressions) at selected quantiles, 2014

	Females						**Males**					
	OLS	**Quantile regression**					**OLS**	**Quantile regression**				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
Individual level variables												
Age of the individual	0.052***	0.046***	0.052***	0.051***	0.051***	0.057***	0.082***	0.117***	0.079***	0.064***	0.057***	0.056***
	(0.005)	(0.007)	(0.006)	(0.005)	(0.005)	(0.005)	(0.003)	(0.006)	(0.004)	(0.003)	(0.003)	(0.004)
Age of the individual^2	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of schooling of the individual	0.035***	0.014	0.034***	0.030***	0.016**	−0.001	0.009	0.023**	0.017*	−0.004	−0.011	−0.013
	(0.009)	(0.014)	(0.012)	(0.008)	(0.008)	(0.009)	(0.007)	(0.010)	(0.009)	(0.006)	(0.007)	(0.009)
Years of schooling of the individual^2	0.000	0.001	0.000	0.000	0.001**	0.002***	0.001**	0.000	0.000	0.001***	0.002***	0.003***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
If household is located on a urban zone (d)	0.356***	0.483***	0.490***	0.327***	0.195***	0.111***	0.073***	0.123***	0.066***	0.023	0.046***	0.043**
	(0.033)	(0.047)	(0.038)	(0.029)	(0.027)	(0.030)	(0.021)	(0.029)	(0.020)	(0.015)	(0.016)	(0.020)
Labour characteristics variables												
If the labour is informal (d)	−0.213***	−0.276***	−0.242***	−0.203***	−0.183***	−0.185***	−0.204***	−0.301***	−0.235***	−0.189***	−0.155***	−0.140***
	(0.028)	(0.032)	(0.030)	(0.026)	(0.025)	(0.029)	(0.016)	(0.032)	(0.018)	(0.014)	(0.016)	(0.019)
Sector												
Mining and Quarrying (d)	0.880***	0.934***	0.769***	0.787***	0.608***	0.768	0.616***	0.380***	0.580***	0.660***	0.632***	0.637***
	(0.115)	(0.129)	(0.265)	(0.072)	(0.084)	(0.696)	(0.051)	(0.075)	(0.061)	(0.035)	(0.038)	(0.055)
Manufacturing and Public Utilities (d)	0.121*	0.162	0.010	−0.024	−0.060	0.114*	0.131***	0.046	0.078**	0.131***	0.148***	0.109***
	(0.065)	(0.108)	(0.074)	(0.052)	(0.067)	(0.067)	(0.032)	(0.040)	(0.037)	(0.026)	(0.033)	(0.040)
Construction (d)	0.212**	0.360***	0.153*	0.168**	0.155	0.362***	0.298***	0.253***	0.322***	0.348***	0.297***	0.263***
	(0.082)	(0.137)	(0.085)	(0.074)	(0.120)	(0.053)	(0.027)	(0.042)	(0.035)	(0.021)	(0.021)	(0.028)
Wholesale and Retail, Hotels and Restaurants (d)	−0.078	−0.046	−0.211***	−0.088*	−0.045	−0.020	0.104***	0.033	0.091***	0.110***	0.137***	0.151***
	(0.052)	(0.085)	(0.075)	(0.047)	(0.047)	(0.048)	(0.032)	(0.049)	(0.035)	(0.026)	(0.031)	(0.028)
Transport, Storage, and Communications (d)	0.164**	0.143	0.012	0.070	0.074	0.207***	0.075**	0.114***	0.077**	0.081***	0.032	0.008
	(0.076)	(0.128)	(0.103)	(0.095)	(0.068)	(0.066)	(0.033)	(0.043)	(0.035)	(0.028)	(0.030)	(0.034)
Finance, Insurance, and Real Estate (d)	0.087	0.148	0.026	0.028	−0.001	0.037	0.069*	−0.059	0.063	0.099***	0.138***	0.149***
	(0.059)	(0.098)	(0.085)	(0.051)	(0.054)	(0.074)	(0.037)	(0.059)	(0.044)	(0.031)	(0.038)	(0.041)
Community, Social and Personal Services (d)	−0.066	−0.058	−0.159**	−0.090**	−0.098**	−0.104**	−0.215***	−0.446***	−0.211***	−0.091***	−0.081***	−0.135***
	(0.053)	(0.094)	(0.075)	(0.045)	(0.047)	(0.049)	(0.041)	(0.069)	(0.037)	(0.032)	(0.029)	(0.049)
Occupation												
Managers, Professionals and Armed forces (d)	0.454***	0.523***	0.408***	0.309***	0.359***	0.491***	0.650***	0.637***	0.600***	0.594***	0.713***	0.809***
	(0.050)	(0.079)	(0.055)	(0.051)	(0.056)	(0.051)	(0.051)	(0.093)	(0.050)	(0.041)	(0.052)	(0.052)
Technicians and associates (d)	0.541***	0.645***	0.559***	0.461***	0.381***	0.517***	0.341***	0.335***	0.302***	0.278***	0.337***	0.477***
	(0.042)	(0.066)	(0.052)	(0.039)	(0.061)	(0.070)	(0.032)	(0.047)	(0.028)	(0.029)	(0.031)	(0.056)
Clerks (d)	0.436***	0.538***	0.478***	0.362***	0.310***	0.287***	0.304***	0.280***	0.251***	0.211***	0.279***	0.296***
	(0.037)	(0.061)	(0.047)	(0.040)	(0.035)	(0.041)	(0.032)	(0.042)	(0.033)	(0.035)	(0.044)	(0.030)
Service and sales worker (d)	0.365***	0.410***	0.447***	0.322***	0.245***	0.286***	0.239***	0.225***	0.197***	0.181***	0.213***	0.255***

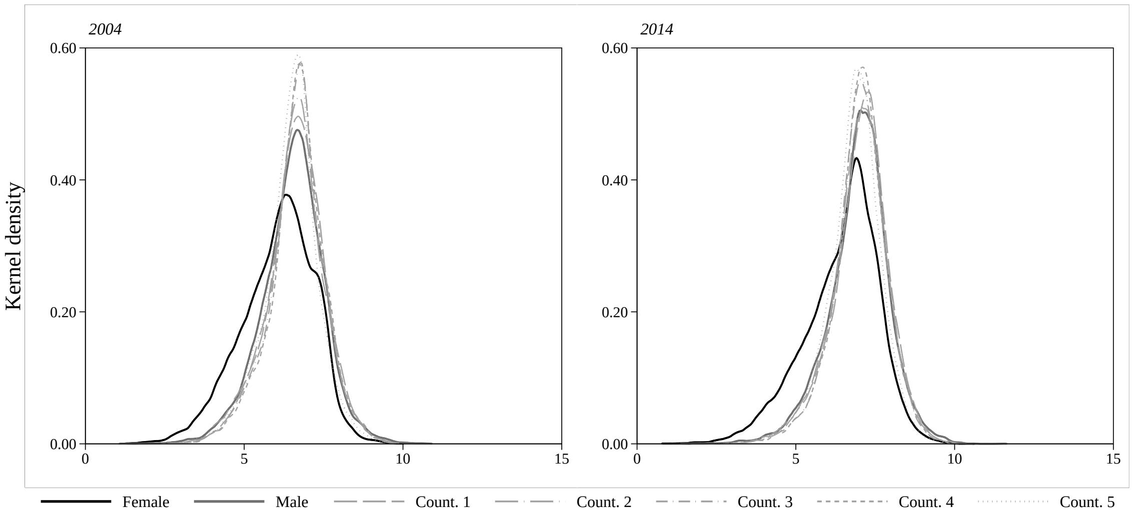
Continued on next page

Table A5 – Regressions by gender under full model (uncorrected quantile regressions), 2014 (*continued from previous page*)

	Females						**Males**					
	OLS	**Quantile regression**					**OLS**	**Quantile regression**				
		10th	25th	50th	75th	90th		10th	25th	50th	75th	90th
	(0.030)	(0.044)	(0.044)	(0.031)	(0.025)	(0.033)	(0.028)	(0.049)	(0.032)	(0.027)	(0.028)	(0.038)
Skilled agricultural and fishery workers (d)	−0.222***	−0.111	−0.312***	−0.349***	−0.382***	−0.305***	−0.249***	−0.562***	−0.476***	−0.310***	−0.089***	0.091***
	(0.056)	(0.088)	(0.077)	(0.050)	(0.054)	(0.056)	(0.027)	(0.042)	(0.033)	(0.024)	(0.024)	(0.029)
Craft and related trades worker (d)	−0.409***	−0.651***	−0.433***	−0.158**	−0.086	−0.192***	0.144***	0.121***	0.146***	0.134***	0.145***	0.165***
	(0.068)	(0.113)	(0.074)	(0.071)	(0.061)	(0.053)	(0.024)	(0.044)	(0.030)	(0.018)	(0.023)	(0.030)
Plant and machine operators and assemblers (d)	−0.537***	−1.156***	−0.765**	−0.224	0.029	−0.242	0.244***	0.298***	0.266***	0.209***	0.234***	0.223***
	(0.129)	(0.141)	(0.356)	(0.148)	(0.086)	(0.164)	(0.026)	(0.042)	(0.025)	(0.024)	(0.027)	(0.032)
Size												
Small (d)	0.371***	0.937***	0.580***	0.255***	0.101***	−0.039	0.214***	0.442***	0.282***	0.192***	0.120***	0.078***
	(0.033)	(0.040)	(0.044)	(0.030)	(0.032)	(0.043)	(0.027)	(0.051)	(0.021)	(0.020)	(0.023)	(0.026)
Medium (d)	0.560***	1.202***	0.731***	0.445***	0.247***	0.181***	0.297***	0.581***	0.343***	0.250***	0.170***	0.118***
	(0.035)	(0.047)	(0.040)	(0.034)	(0.037)	(0.049)	(0.022)	(0.044)	(0.024)	(0.019)	(0.026)	(0.025)
Large (d)	0.700***	1.293***	0.876***	0.584***	0.428***	0.305***	0.460***	0.704***	0.494***	0.372***	0.323***	0.337***
	(0.035)	(0.044)	(0.044)	(0.036)	(0.035)	(0.042)	(0.022)	(0.032)	(0.022)	(0.018)	(0.022)	(0.028)
Constant	4.592***	3.334***	4.043***	4.849***	5.512***	5.826***	4.826***	3.357***	4.623***	5.389***	5.819***	6.071***
	(0.111)	(0.163)	(0.154)	(0.115)	(0.108)	(0.117)	(0.077)	(0.141)	(0.102)	(0.066)	(0.069)	(0.083)
Observations	15,907	15,907	15,907	15,907	15,907	15,907	23,559	23,559	23,559	23,559	23,559	23,559
Pseudo R^2	0.292	0.203	0.199	0.168	0.155	0.174	0.341	0.238	0.224	0.201	0.199	0.226
F test education	44.888	15.983	21.855	45.433	78.124	62.016	61.964	10.116	17.234	64.001	89.965	92.931
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test sector	15.979	21.888	11.519	31.829	13.594	27.503	58.724	21.396	52.896	101.692	76.607	35.658
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test occupation	55.050	53.785	44.368	35.922	30.168	34.949	45.568	41.867	64.815	65.721	44.708	44.699
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test firm size	152.649	294.226	138.506	96.746	51.454	22.311	149.538	162.213	168.511	158.932	75.630	47.816
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
F test model	220.285	383.444	325.669	245.844	139.115	171.574	293.467	381.075	328.405	334.034	212.177	167.248
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

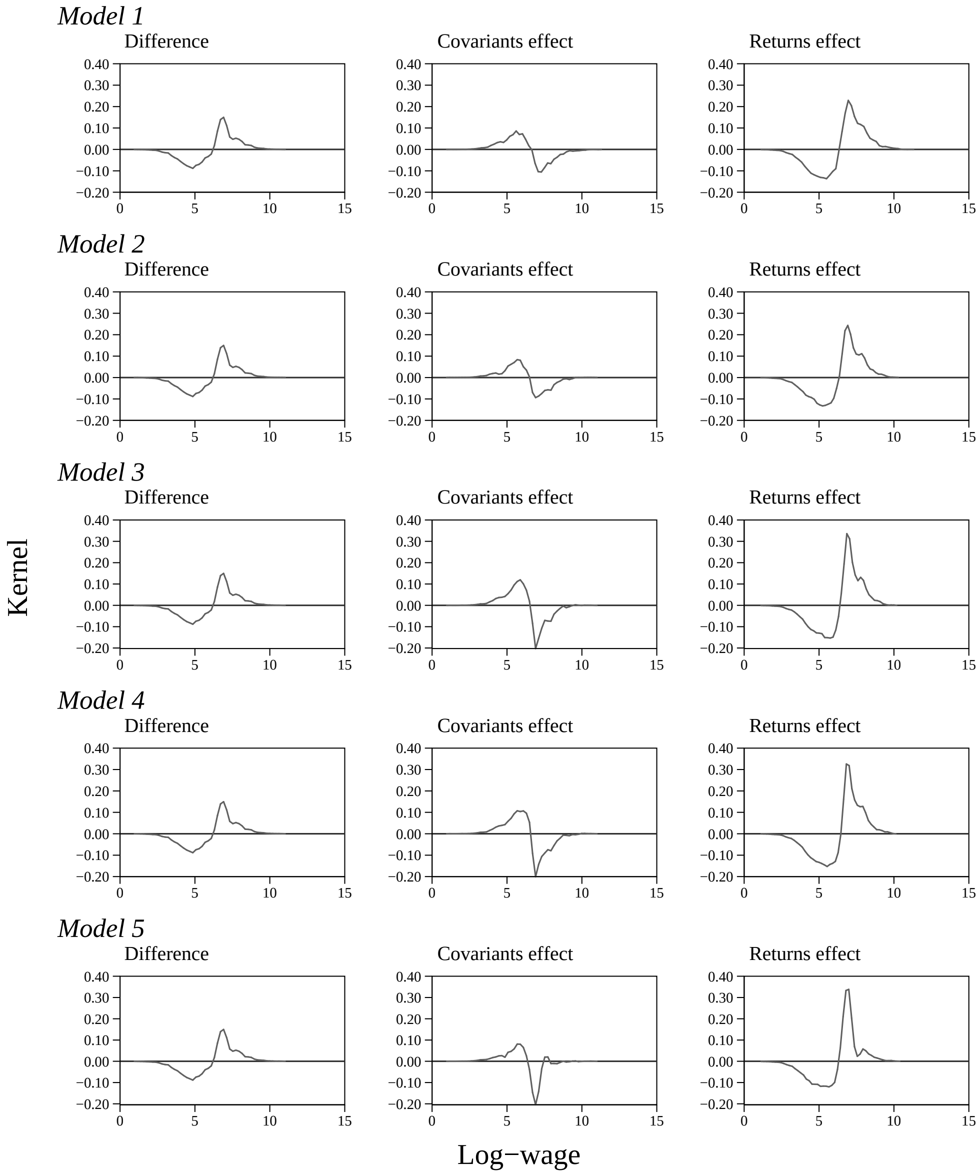
Note: (Ln.) Wages measured in constant 2014 Soles. Sample include individuals between 18 and 65 years. Observations weighted by expansion factor and VCE corrected according to survey's complex sample design. Standard errors in parenthesis and p-values of the F-test in brackets. (d)=Dummy variable. * Significant at 10% level; ** Significant at 5% level; *** Significant at 1% level.

Figure A6 – Observed and counterfactual densities (uncorrected quantile regressions) under alternative models, 2004 and 2014



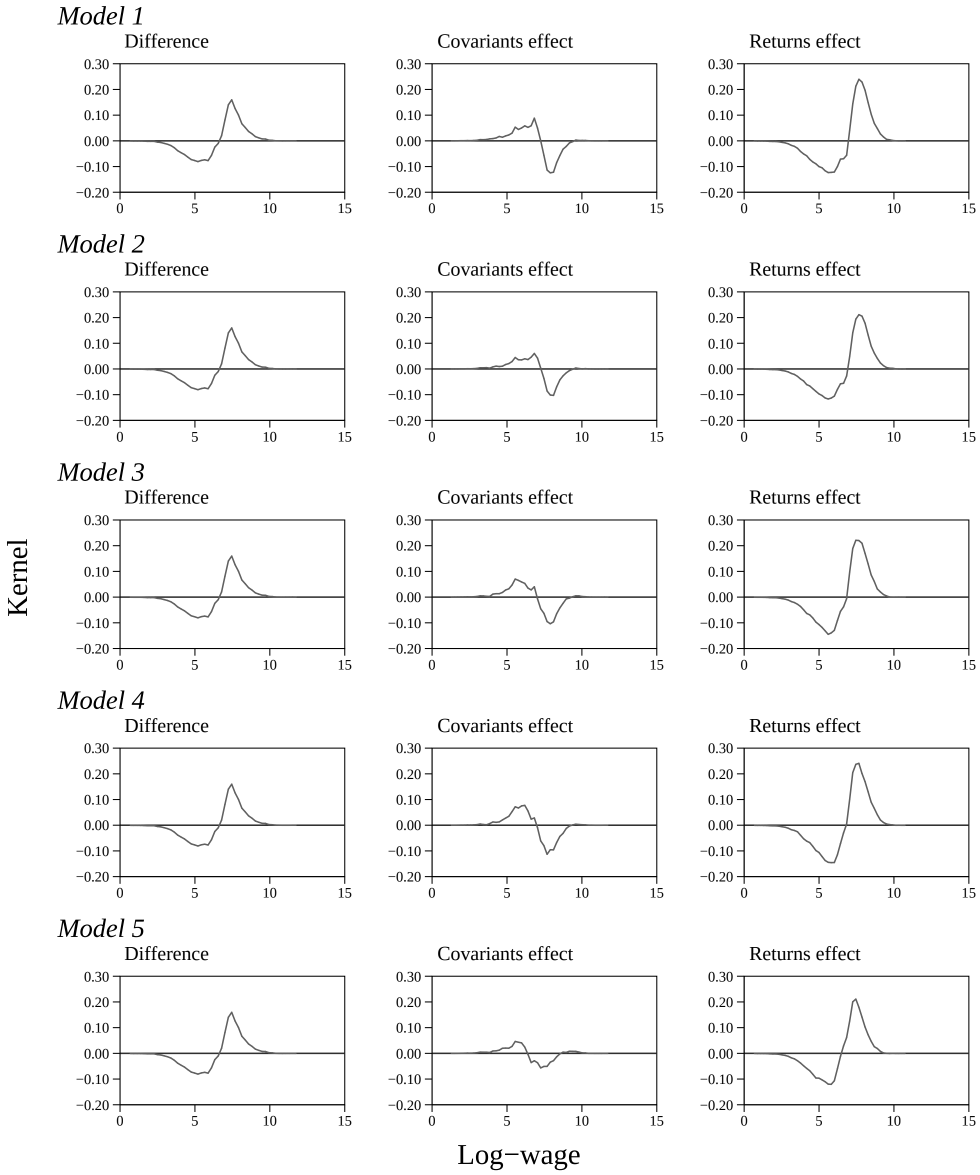
Note: Counterfactual distributios calculated taking covariants of females and coefficients of males under alternate models. Model 1 corresponds to the model with only basic characteristics; the remaining models are defined as in subsection 5.1.
Elaborated by the author based on INEI − National Household Survey (2004−2014)

Figure A7 – Differences in densities (corrected quantile regressions) under alternative models, 2004



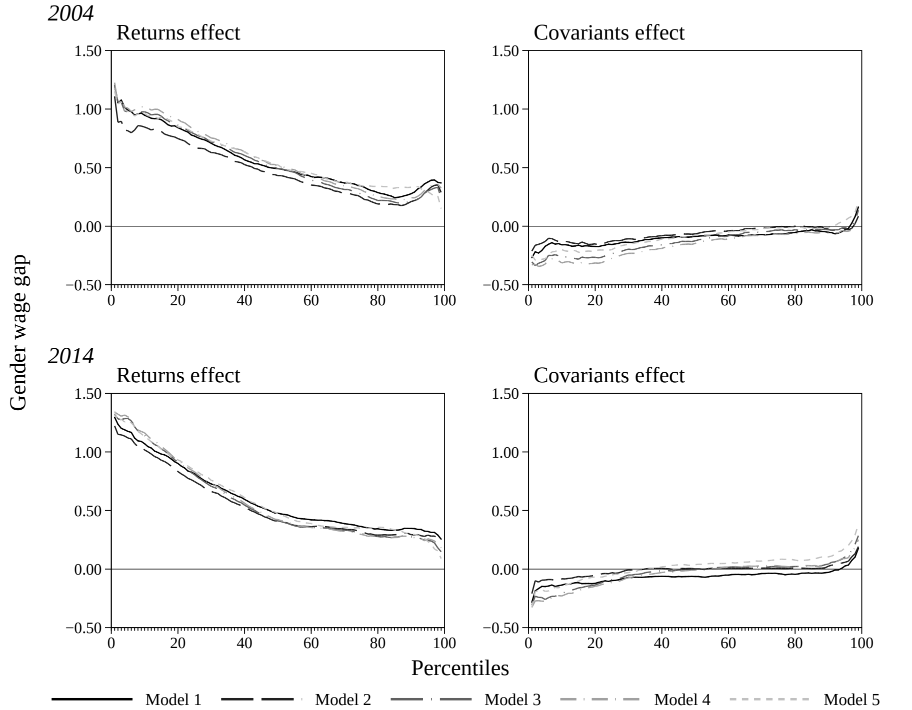
Nota: In order to estimate differences each density was evaluated at the same points in a range which contains some of the points. The kernel function corresponds to the Epanechnikov funcion.
Elaborated by the author based on INEI − National Household Survey (2004−2014)

Figure A8 – Differences in densities (corrected quantile regressions) under alternative models, 2014



Nota: In order to estimate differences each density was evaluated at the same points in a range which contains some of the points. The kernel function corresponds to the Epanechnikov funcion.
Elaborated by the author based on INEI − National Household Survey (2004−2014)

Figure A9 – Decomposition of the gender wage gap on covariate and returns effects at each percentile (uncorrected quantile regressions) under alternative models, 2004 and 2014



Note: Shaded area represent confidence intervals corresponding to percentiles 2.5 and 97.5 of bootstrap estimates. Model 1 correspon
to the model with only basic characteristics; the remaining models are defined as in subsection 5.1.
Elaborated by the author based on INEI − National Household Survey (2004−2014)

Table A6 – Decomposition of the gender wage gap on covariate and returns effects at selected percentiles (uncorrected quantile regressions) under alternative models, 2004 and 2014

	Model 1		**Model 2**		**Model 3**		**Model 4**		**Model 5**	
	C	R	C	R	C	R	C	R	C	R
2014										
5	−0.171	0.990	−0.130	0.814	−0.293	0.977	−0.323	1.008	−0.278	0.981
10	−0.158	0.947	−0.141	0.847	−0.271	0.977	−0.314	1.020	−0.200	0.961
25	−0.156	0.769	−0.127	0.681	−0.233	0.788	−0.274	0.829	−0.199	0.803
50	−0.086	0.491	−0.066	0.432	−0.125	0.491	−0.147	0.513	−0.076	0.523
75	−0.066	0.344	−0.005	0.247	−0.034	0.275	−0.069	0.310	−0.015	0.346
90	−0.052	0.278	−0.026	0.211	−0.031	0.216	−0.058	0.243	−0.011	0.333
95	−0.031	0.377	−0.021	0.309	−0.013	0.300	−0.043	0.330	0.048	0.295
2014										
5	−0.150	1.176	−0.093	1.119	−0.260	1.286	−0.274	1.300	−0.190	1.268
10	−0.137	1.070	−0.083	1.017	−0.207	1.141	−0.229	1.163	−0.146	1.130
25	−0.098	0.814	−0.031	0.747	−0.090	0.806	−0.114	0.830	−0.051	0.845
50	−0.064	0.477	0.002	0.412	0.001	0.412	−0.009	0.422	0.038	0.474
75	−0.040	0.363	0.006	0.318	0.019	0.305	0.028	0.296	0.083	0.349
90	−0.028	0.347	0.025	0.294	0.043	0.275	0.042	0.276	0.104	0.298
95	0.027	0.321	0.058	0.290	0.105	0.243	0.090	0.258	0.183	0.236

Note: C=Covariates effect, R=Returns effect. Model 1 corresponds to the model with only basic characteristics; the remaining models are defined as in subsection 5.1.
Elaborated by the author based on INEI - National Household Survey (2004-2014)

B Variables construction appendix

ENAHO dataset is conducted on a continuous basis by INEI by means of interviews to the households members. This survey is characterized by a probabilistic, stratified and multi-staged sample (similar to World Banks' LSMSs) in order to reduce sampling error and provide representative estimates for each strata.

It is important to point out some special considerations ENAHO. In the first place, the standard used by INEI to define those under the Labour Force includes individuals over 14 years old[28]. However, wage for under-aged individuals and for those over retirement age (65 in the Peruvian case), presumably, follow a different data generation process than what is normally assumed under the Mincer regression framework adopted here. In the same vein as Aktas and Uysal (2012); Albrecht et al. (2003, 2009); Buchinsky (1998) and others, the analysis is restricted for a subset of those in Labour Force: 18 to 65 years. In the second place, the complex design of the survey influences the estimation of parameters and standard errors in two different manners. On the one hand, observations are expanded by using a weight variable which reflects its probabil-

[28] Article 51º of Law Nº 27337 (modified in year 2001) states that the minimum age to authorize teenage work is 14 years old, with some exceptions for those from 12 years old under parental authorization and as long as their duties do not harm their health or development or interfere with their educational process.

ity weight (the inverse of the probability that the observation is included in the sample, such that it is weighted more heavily if it has a very small probability of selection). On the other hand, standard errors are properly adjusted given that the sample has been divided in 8 strata and different sample units for urban and rural areas. In the third place, despite that for each year ENAHO data is reported on a quarterly basis and on a yearly basis, only the latter sample is considered. This is because only the yearly samples include information of all the individuals for the given year, which results in an increase of the number of observations each year and, thus, a more precise characterization of the population. Finally, to allow comparability monthly labour wages are deflated temporally and spatially. Temporal deflators translate wages for any given year in terms of 2014 Soles to net out the effect of inflation. Spatial deflators translate wages of urban and rural areas in each of the 25 regions in the sample in terms of Metropolitan Lima (capital city) Soles to net out the spatial differences in costs of living. Accordingly, after applying these two deflators, wages are expressed in Metropolitan Lima 2014 Soles.

Regarding the variables in the participation equation (equation 10), the dependent variable, $employment$, is constructed by INEI applying two criteria. Among those who had a job, if the hours worked were more than 15 per week; among those who did not had a job, if they we looking for a job or if they were not looking for a job but were engaged in productive activities anyway. Only the former are considered since only they declare positive income. Note that this variable does not takes into consideration information about whether the individual works part time or full time. Admittedly, this could have an impact on wages, for monthly wages would be different under those two regimes. We acknowledge this limitation and proceed without taking this difference into account since we are interested on the grounds of simplicity of the model.

Among the independent variables, years of education was constructed based on the information declared by the household members. More specifically, individuals declare their level of education (e.g. second grade of primary, fourth grade of secondary, fourth year of university, etc.) and, based on this, we imputed the minimum necessary years of education in order to attain that level. Hence, it could censor the number of years of education in two ways. For those individuals reporting primary or secondary school, we do not consider the extra number of years the individual studied because of failing one or more years. For those individuals reporting higher education, we do not consider the extra number of years the individual studied beyond the standard length of a professional career in Peru. For university education, it is 5 years and for technical occupations, 2 years. We believe that this censoring is necessary in order to make educational attainment comparable among individuals and that it doesn't affect estimates in a relevant way.

Regarding variables in the outcome equation (equation 11), the labour characteristic variables follow international classifications in order to facilitate comparability with other studies. Informal status of the individual is approximated by the lack of affiliation to a pension system of the worker (declared by the worker itself). Admittedly this definition can be considered debatable since there is not a unique definition of this working status. Nevertheless, according to Freije (2002), p. 2: "Informal workers lack almost every form of social protection [...] No access to the pensions system protection make informal workers unable to retire and force them to work longer perhaps under decreasing productivity of their human capital". Since this description characterizes an important part of workers settled in the informal sector in Peru rather than alternative definitions (e.g. working on a firm without accounting books, not receiving an invoice for their professional services, working less than 40 hours per week, etc.), we choose to take it as our indicator of informality.

The vector of industry dummies is a reduced version of the International Standard Industrial Classification of All Economic Activities (ISIC), Revision 3, defined by United Nations. Description of the categories, divisions, groups and classes available http://unstats.un.org/unsd/c Although the original classification considers 17 major groups, we add them into 8: Agriculture, forestry, and fishing; Mining and quarrying; Manufacturing and Public Utilities; Construction; Wholesale and Retail Trade, Hotels and Restaurants; Transport, Storage, and Communication; Finance, Insurance, and Real Estate; Community, Social and Personal Services. The vector of occupation dummies is a reduced version of the International Standard Classification of Occupations (ISCO-08) from the International Labour Organization. Description of the 10 major groups, sub-major groups, minor groups and unit groups are available at http://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm. Provided that the number of females as managers and armed forces who are non-missing in terms of the covariates considered was null for both years, we merge these major groups with that of Professionals. In the end, we end up with with 8 major groups: Managers, Professionals and Armed forces, Technicians and associates, Clerks, Service and sales workers, Skilled agricultural and fishery workers, Craft and related trades workers, Plant and machine operators and assemblers and Elementary occupations. As Dolton and Kidd (1994), we recognize that the range of occupational classification might affect empirical results, however we chose to this classification because it allows comparability with other studies. Finally, the vector of firm size includes those categories considered by INEI and shown in Saavedra et al. 2008: micro (from 1 to 9 workers), small (from 10 to 20 workers), medium (from 21 to 100 workers) and large (more than 101 workers) firms.

ÚLTIMAS PUBLICACIONES DE LOS PROFESORES
DEL DEPARTAMENTO DE ECONOMÍA

- *Libros*

José Rodríguez y Pedro Francke (editores)
2017 Exclusión e inclusión social en el Perú. Logros y desafíos para el desarrollo. Lima, Fondo Editorial, Pontificia Universidad Católica del Perú.

Iván Rivera
2017 Principios de Macroeconomía. Un enfoque de sentido común. Lima, Fondo Editorial, Pontificia Universidad Católica del Perú.

Ismael Muñoz, Marcial Blondet y Gonzalo Gamio (Editores).
2017 Ética, agencia y desarrollo humano. V Conferencia de la Asociación Latinoamericana y del Caribe para el Desarrollo Humano y el Enfoque de Capacidades. Lima, Fondo Editorial, Pontificia Universidad Católica del Perú.

Waldo Mendoza y Janneth Leyva
2017 *La economía del VRAEM. Diagnósticos y opciones de política.* Lima, USAID-CIES.

Félix Jiménez
2017 *Macroeconomía. Enfoques y modelos.* Lima, Editorial Macro.

Máximo Vega-Centeno
2017 *Ética y deontología.* Lima, Fondo Editorial, Pontificia Universidad Católica del Perú.

Mario D. Tello
2017 *Análisis de equilibrio general. Modelos y aplicaciones para países en desarrollo*. Lima, Fondo Editorial, Pontificia Universidad Católica del Perú.

Roxana Barrantes, Elena Borasino, Manuel Glave, Miguel Ángel La Rosa y Karla Vergara
2016 *De la Amazonía su palma. Aportes a la gestión territorial en la región Loreto*. Lima, Instituto de Estudios Peruanos, IEP, Grupo de Análisis para el Desarrollo, Grade y Derecho, Ambiente y Recursos Naturales, DAR.

- *Documentos de Trabajo*

No. 441 "Crecimiento y desindustrialización prematura en Perú. Un análisis Kaldoriano". Félix Jiménez. Junio, 2017.

No. 440 "La economía de PPK. Promesas y resultados: la distancia que los separa". Waldo Mendoza Bellido y Erika Collantes Goicochea. Mayo, 2017.

No. 439 "¿Es la educación un mecanismo de salida hacia mejores ocupaciones para las trabajadoras del hogar?" Cecilia Garavito. Mayo, 2017.

No. 438 “¿Caminos distintos y destinos iguales?: Análisis de la convergencia en patrones de uso de internet entre diferentes grupo etarios”. Roxana Barrantes y Eduardo Vargas. Abril, 2017.

No. 437 “Factores discrecionales y no discrecionales de la eficiencia educativa: evidencias para el caso peruano”. Guillermo Jopen Sanchez. Abril, 2017.

No. 436 “Empirical Modeling of Latin American Stock and Forex Markets Returns and Volatility using Markov-Switching GARCH Models”. Miguel Ataurima Arellano, Erika Collantes y Gabriel Rodríguez, Marzo, 2017.

No. 435 “Estimation of the Sovereign Yield Curve of Peru: The Role of Macroeconomic and Latent Factors”. Alejandra Olivares Ríos, Gabriel Rodríguez y Miguel Ataurima Arellano, Marzo, 2017.

No. 434 “El Estado Rentista. El ejemplo del Perú en el siglo XIX”. Héctor Omar Noejovich. Enero 2017.

No. 433 “Una alternativa al IS-LM-AD-AS: El modelo IS-MR-AD-AS”. Waldo Mendoza. Diciembre, 2016.

- *Materiales de Enseñanza*

No. 2 “Macroeconomía: Enfoques y modelos. Ejercicios resueltos”. Felix Jiménez. Marzo, 2016.

No. 1 “Introducción a la teoría del Equilibrio General”. Alejandro Lugon. Octubre, 2015.

Departamento de Economía - Pontificia Universidad Católica del Perú
Av. Universitaria 1801, Lima 32 – Perú.
Telf. 626-2000 anexos 4950 - 4951
http://departamento.pucp.edu.pe/economia/