Article

Not peer-reviewed version

Pseudo-Panel Decomposition of the Blinder–Oaxaca Gender Wage Gap

Jhon James Mora * and Diana Yaneth Herrera Duque

Posted Date: 9 June 2025

doi: 10.20944/preprints202506.0536.v1

Keywords: pseudo-panel; Blinder-Oaxaca decomposition; selection bias; human capital

doi:10.20944/preprints202506.0536.v1

Article

Pseudo-Panel Decomposition of the Blinder–Oaxaca Gender Wage Gap

Jhon James Mora [1,*] and Diana Yaneth Herrera [2]

[1] Department of Economics, Universidad Icesi, Cali, Colombia
[2] Department of Economics, ICESI University, Cali, Colombia
* Correspondence: jjmora@icesi.edu.co

Abstract: This article introduces a novel approach to decomposing the Blinder-Oaxaca gender wage gap using pseudo-panel data. In many developing countries, panel data is not available; however, understanding the evolution of the gender wage gap over time requires tracking individuals longitudinally. When individuals change across time periods, estimators tend to be inconsistent and inefficient. To address this issue, and building upon the traditional Blinder-Oaxaca methodology, we propose an alternative procedure that follows cohorts over time rather than individuals. This approach enables the estimation of both the explained and unexplained components – "endowment effect" and "remuneration effect"– of the wage gap, along with their respective standard errors, even in the absence of true panel data. We apply this methodology to the case of Colombia, finding a gender wage gap of 14.6% in favor of male cohorts. This gap comprises a −5.6% explained component and a 20% unexplained component.

Keywords: pseudo-panel; Blinder-Oaxaca decomposition; selection bias; human capital

JEL Classification: J15; J16; J31; J70; C23

1. Introduction

Measuring wage gaps between groups over time ideally requires panel data. However, such data are often unavailable in most developing countries due to the high costs associated with tracking individuals over time. Instead, these countries typically rely on repeated cross-sectional household surveys that are representative at each point in time but do not follow the same individuals across periods. For example, Colombia conducts periodic household surveys with representative samples, but the individuals surveyed differ across survey waves.

Some researchers address this limitation by pooling cross-sectional data and including time dummies to estimate consistent parameters. However, this approach is inefficient in the presence of measurement error arising from unobserved heterogeneity that varies across time. Under these circumstances, estimators may become inconsistent, and a preferable alternative is the use of pseudo-panel data, as proposed by Deaton (1985) and further developed by Mora and Muro (2014).

Consider now the estimation of wage gaps between two groups—typically men and women—using pseudo-panel data. Measurement errors persist because gender wage gaps often reflect unobserved individual heterogeneity that varies across groups. Furthermore, these estimates may be inconsistent in the presence of selection bias.

In cross-sectional data, the Blinder-Oaxaca decomposition is commonly applied to estimate the wage gap, and adjustments for selection bias are often included. To enhance efficiency, Jann (2005) provides a method to estimate the variance-covariance matrix of the decomposition. In the context of panel data, Kröger and Hartmann (2021) present an approach that extends the Kitagawa-Oaxaca-Blinder decomposition method to analyze wage differentials over time. However, to date, no consistent and efficient methodology exists for estimating the gender wage gap and its twofold

 doi:10.20944/preprints202506.0536.v1

decomposition using pseudo-panel data. This is due to the fact that individuals in period t differ from those in period t–1, and unobserved heterogeneity is not constant over time.

Wage gaps between men and women are a matter of global importance. Equal pay is one of the guiding principles of the International Labour Organization (ILO) and a key target of the United Nations Sustainable Development Goals. Moreover, Buchely (2013) argues that gender inequality in the labor market imposes inefficiencies on society, as the costs associated with women's disadvantage are externalized through the social security system, particularly in health and pensions.

The literature on the gender wage gap is extensive. Among international studies, Paz (1998) estimates the wage gap in Greater Buenos Aires and the Northwest of Argentina using data from the Permanent Household Survey. The income disparity between women and men is 0.70 for the overall population and 0.60 among individuals with spouses. The Blinder-Oaxaca decomposition indicates that approximately 90% of the wage gap remains unexplained by differences in human capital. Di Paola and Berges (2000) analyze gender income differences in Mar del Plata, Argentina, employing the Blinder-Oaxaca decomposition and correcting for selection bias using the Heckman method. Their findings suggest that 78% of the wage gap is explained by human capital endowments, while the remaining 22% is attributable to discrimination.

Johansson et al. (2005) study the gender wage gap in Sweden during the 1980s and 1990s using cross-sectional data from the Swedish Household Income Survey. Their results show a gap of approximately 13% in the 1980s and 15% in the 1990s, with the unexplained portion ranging between 5% and 9%. Watson (2010) analyzes the gender wage gap among full-time managers in Australia from 2001 to 2007, using data from the Household, Income and Labour Dynamics in Australia Survey. He finds that female managers earn about 27% less than their male counterparts, and that the unexplained portion of the gap –remuneration effect– ranges from 65% to 90%, depending on the decomposition method used.

Biltagy (2014) examines wage disparities in Egypt using data from the 2006 Egyptian Labour Market Panel Survey. The Blinder-Oaxaca decomposition reveals a gender wage gap of 25%, attributed entirely to discrimination against women. Blau and Kahn (2017) study changes in the gender wage gap in the United States between 1980 and 2010 using microdata from the Panel Study of Income Dynamics. They find that the unexplained component of the wage gap –remuneration effect– declined from 49% in 1980 to 38% in 2010.

Several studies also focus on the Colombian labor market. Baquero (2001) applies the Oaxaca (1973) decomposition using data from the National Household Survey and finds a wage gap of approximately 34% in favor of men in 1999. Abadía (2005) examines statistical discrimination by gender using data from the Continuous Household Survey for the second quarter of 2003, distinguishing between public and private sector workers. While no discrimination is found in the public sector, evidence of discrimination exists in the private sector, particularly against married or cohabiting women. Bernat (2005) analyzes hourly wage differences in Colombia's seven major cities from 2000 to 2003 and concludes, based on a Blinder-Oaxaca decomposition, that gender discrimination persists. Fernández (2006), using Quality of Life Survey data and quantile regressions for 1997–2003, shows that wage differences favoring men are concentrated in the upper percentiles of the wage distribution, while in lower percentiles, the differences tend to favor women.

This article contributes to two strands of the literature on the gender wage gap. First, it extends the Blinder (1973) and Oaxaca (1973) decomposition to pseudo-panel data, enabling analysis in contexts where traditional panel data are unavailable. Second, it adapts the correction proposed by Jann (2005) for estimating the variance-covariance matrix of the decomposition to the pseudo-panel framework. Finally, the proposed methodology is applied to the Colombian case, serving as an illustrative example for a developing country context.

The remainder of the article is organized as follows. Section 2 reviews the literature on the Blinder-Oaxaca decomposition. Section 3 presents the proposed pseudo-panel adaptation of the Blinder-Oaxaca decomposition and the corresponding variance-covariance matrix following Jann

(2005). Section 4 applies this methodology to estimate the gender wage gap in Colombia. Section 5 concludes.

2. Blinder–Oaxaca Decomposition

The most widely employed technique for assessing the gender wage gap is the Blinder–Oaxaca decomposition (1973). This method disaggregates the observed wage differential into two main components. The first component reflects differences in the returns to observable productivity-related characteristics (e.g., education, experience), while the second component captures disparities due to unobservable factors, including discrimination.

Consider two groups: men and women. The first step in the decomposition involves estimating wage equations for each group $g \in \{M, W\}$, where individual wages are modeled as follows:

$$\begin{gathered} LnW_g = f(S_{g,}, Exp_g, Exp_g^2) \\ g = men, women \end{gathered} \quad (1)$$

In Equation (1), ln$W$ denotes the natural logarithm of hourly wages, S represents years of schooling, Exp corresponds to potential labor market experience—calculated as age minus years of schooling minus six—and Exp^2 is the square of potential experience. This specification follows the human capital framework proposed by Mincer (1974), who argued that the returns to education can be quantified through an income equation based on an individual's educational attainment and work experience. The Mincer equation predicts a positive relationship between years of schooling and earnings. However, in the case of Colombia, this theoretical expectation does not fully materialize for women. Despite having, on average, higher levels of schooling than men, women continue to experience lower wages. This indicates that the returns to human capital differ significantly between men and women, suggesting the presence of structural inequalities in the labor market that may not be explained solely by differences in observable characteristics.

With respect to Equation (1), the difference in average wages between the two groups can be expressed as follows:

$$(\overline{w}_m - \overline{w}_w) = (\bar{X}_m - \bar{X}_w)\hat{\beta}_m + (\hat{\beta}_m + \hat{\beta}_w)\bar{X}_w \quad (2)$$

where $\overline{w}_g$ and $\bar{X}_g$ denote the mean of the logarithm of wages and the control characteristics for group g, respectively, and $\widehat{\beta_g}$ is the estimated parameter from Equation (1). The wage gap can thus be decomposed into two components: the explained component or "endowment effect" which reflects differences in observable productive characteristics between the groups, and the unexplained component or "remuneration effect", which captures the portion of the wage differential that cannot be attributed to such characteristics—often interpreted as a result of discrimination or other unobserved factors.

In recent decades, applications of the Blinder–Oaxaca decomposition have often omitted statistical inference information, such as standard errors and confidence intervals. However, interpreting decomposition results without reference to the precision of the estimates significantly limits their reliability and analytical value.

Oaxaca and Ransom (1998) and Greene (2003) proposed methods to approximate these standard errors, assuming fixed regressors. This assumption neglects a critical source of statistical uncertainty, which may lead to biased inference in most empirical applications (Jann, 2005). In particular, treating the regressors as non-stochastic tends to substantially underestimate the standard errors associated with the explained component –endowment effect– of the wage gap.

In response to these limitations, Jann (2005) developed unbiased variance estimators for the components of the Blinder–Oaxaca decomposition. Suppose that

$$\tilde{Y} = \tilde{X}'\hat{\beta} \quad (3)$$

where $\tilde{X}$ is a vector of sample means and $\hat{\beta}$ is a vector of regression coefficients. The sample variance $V(\tilde{X}'\hat{\beta})$ can be estimated as follows:

a) If the covariates are fixed, then $\tilde{X}$ has no sampling variance. If the regressors are fixed, then $\tilde{X}$ is constant. Therefore, $V(\tilde{X}'\hat{\beta}) = \tilde{X}'V(\hat{\beta})\tilde{X}$.

b) However, in most applications, the regressors and $\tilde{X}$ are stochastic. Since $\tilde{X}$ and $\hat{\beta}$ are not correlated (as long as this is true, then $Cov(\epsilon, X) = 0$), the sampling variance is as follows (Jann, 2008),

$$\hat{V}(\tilde{X}'\hat{\beta}) = \tilde{X}'\hat{V}(\hat{\beta})\tilde{X} + \tilde{\beta}'\hat{V}(\tilde{X})\hat{\beta} + trace\{\hat{V}(\tilde{X})\hat{V}(\hat{\beta})\} \quad (4)$$

Where $trace\{\hat{V}(\tilde{X})\hat{V}(\hat{\beta})\}$ disappears asymptotically and $\hat{V}(\hat{\beta})$ is the variance-covariance matrix obtained from the regression process.

3. Pseudo-Panel Approach to the Blinder–Oaxaca Decomposition

In the absence of longitudinal panel data that track the same individuals over time, it is necessary to employ pseudo-panel data methods. Pseudo-panels consist of observations drawn from different individuals across various time periods—that is, the individuals observed at time t differ from those observed at time t−1. Utilizing the pseudo-panel approach enables the consistent and efficient application of the Blinder–Oaxaca decomposition when only cohort-level tracking is feasible.

Deaton (1985) introduced the concept of pseudo-panels as a method for exploiting repeated cross-sectional surveys. This approach entails grouping individuals into synthetic cohorts based on time-invariant and exogenous characteristics, such as age and gender.

As previously noted, the foundational idea of pseudo-panels is to construct cohorts composed of individuals who exhibit similar behavioral patterns (Guillerm, 2017). For instance, in the case of Colombia, where the age of legal adulthood is 18, it is appropriate to form nine five-year cohorts spanning the working-age population, specifically individuals aged 18 to 63. These cohorts approximate different stages of the employment life cycle.

Estimating returns to human capital using pooled cross-sectional regressions introduces an errors-in-variables problem, primarily due to time-varying unobserved individual heterogeneity. Additionally, such estimations are subject to inconsistency in the presence of selection bias.

To address these concerns, consider the following pseudo-panel specification for estimating returns to human capital (Mincer, 1974; Mora and Muro 2014):

$$Y_{i(t)} = \beta_1 X_{i(t)} + \rho\, \lambda_{i(t)} + f_i + \epsilon_{i(t)} \quad (5)$$

where $Y_{i(t)}$ denotes income, $X_{i(t)}$ represents the set of explanatory variables consistent with human capital theory—namely, education and potential experience— $\lambda_{i(t)}$ accounts for potential selection bias, and f_i captures individual-specific unobserved heterogeneity. The subscripts $i(t)$ indicate that the data originate from independent, representative cross-sectional surveys in which individuals are observed only once, in a single time period.

In this context, Deaton (1985) demonstrates that when individuals differ across time periods, estimations based on Equation (5) yield inconsistent results. To overcome this limitation, Deaton proposes a pseudo-panel estimation strategy that involves constructing cohorts based on invariant characteristics.

Building on this approach, Mora and Muro (2014) develop a methodology for addressing pseudo-panel data in the presence of selection bias. Specifically, they propose using the Generalized Method of Moments (GMM) to account for the measurement error problem inherent in pseudo-panel data. This methodology, hereafter referred to as GMMC (GMM with correction for measurement error), leads to the following equation:

$$E[(Y_{it} - X'_{it}\beta_1 - Z'_{0i}\delta - \rho\lambda_{ct})h(Z_{0i}, Z_{1it})] = B\beta + b \quad (6)$$

where Z_0 denotes a matrix of fictitious cohort indicators; Z_{1it} are instrumental variables that vary over time (they do not contain Z_0); $h(\cdot)$ is a known function —typically comprising time effects and cohort-by-time interaction terms—although other time-varying variables may also be incorporated.; $\beta = (\beta'_1\delta'\rho)'$; and B, b depends on the covariance matrix of measurement errors.

Regarding the selection mechanism, a panel probit model is employed to characterize the selection process, specified as follows:

$$E[(s_{it} - Z'_{1it}\gamma_t)A_t] = 0 \quad (7)$$

 doi:10.20944/preprints202506.0536.v1

where s_{it} is the selection process, and A is a cohort mean operator, $(Z_0'Z_0)^{-1}Z_0'$.

Definition 1. The cohort-level expression for the moment conditions specified in Equations (6) and (7) can be formulated as follows:

$$E\left[s_{ct} - Z_{1ct}^{'}\gamma_t\right] = 0;\ t = 1, \dots, T,\ c = 1, \dots C, \tag{8}$$

$$E[(\Delta Y_{ct} - \Delta X_{ct}'\beta_1 - \rho\Delta\lambda_{ct})\Delta W_{ct}] = B\beta + b \tag{9}$$

Here, $\Delta W_{ct} = (\Delta X_{ct}', \Delta\lambda_{ct})'$. Equation (8) is a system of T cross-sectional linear regressions. First, differences from the synthetic panel (Deaton, 1985) are used in Equation (9). By substituting $\widehat{\gamma_{ct}}$ into Equation (9), we obtain the following:

$$E\left[\left(\Delta Y_{ct} - \Delta X_{ct}'\beta_1 - \rho\Delta\,\widehat{\lambda_{ct}}\right)\Delta X_{ct}\right] = B\beta + b. \tag{10}$$

Finally, the GMMC estimator is as follows:

$$\hat{\beta} = \left[\sum_{c=1}^{C}(\Delta W_c'\Delta W_c + B')D_c\sum_{c=1}^{C}(\Delta W_c'\Delta W_c + B)\right]^{-1}\left[\sum_{c=1}^{C}(\Delta W_c'\Delta W_c + B')D_c\sum_{c=1}^{C}(\Delta W_c'\Delta Y_c - b)\right] \tag{11}$$

where $\Delta W_c = (\Delta W_{c2}, \Delta W_{c3}, \dots, \Delta W_{CT})'$ and $\Delta Y_c = (\Delta Y_{c2}, \Delta Y_{c3}, \dots, \Delta Y_{CT})'$. The optimal choice of D_c is any consistent estimator of the inverse of the covariance matrix of $\Delta W_c'\Delta W_c$ (Hansen, 1982).

The asymptotic distribution of the GMMC estimator, for B, b, ΔW_c known, can be derived using standard assumptions and GMM theory (Mora and Muro, 2014). Following Deaton (1985), Newey and McFadden (1994), and Mora and Muro (2014), the following is a convenient expression for an upper limit of the covariance matrix V_β^{MM}:

$$V_\beta^{MM} = [M_{WW} - \Sigma]^{-1}\left[\Sigma_{WW}\left(\sigma_\mu^2 + \sigma_{00} + \theta'\Sigma\theta - 2\sigma'\theta\right) + (\sigma - \Sigma\theta)(\sigma - \Sigma\theta)'\right][M_{WW} - \Sigma]^{-1} + \Pi'\hat{V}\Pi \tag{12}$$

In Equation (12), the first additive term corresponds to the covariance matrix associated with the pseudo-panel data model (Deaton, 1985). The second term is the correction matrix designed to adjust for selectivity bias, which is essential for obtaining consistent estimators in the pseudo-panel framework. This correction term reflects an estimated regressor—rather than the true regressor—in the second stage of the two-step Generalized Method of Moments with Measurement Error Correction (GMMC) estimation procedure (Mora and Muro, 2014). Moreover, the covariance matrix of the parameter estimates is further adjusted for bias using the approach proposed by Newey and McFadden (1994). A comprehensive demonstration of this methodology is provided in Mora and Muro (2014).

For instance, to estimate the returns to education within each group, extending the Mincer (1974) earnings equation to the pseudo-panel context can be expressed as follows:

$$LnW_{g,ct} = \alpha_{ct} + \beta_1 S_{g,ct} + \beta_2 Exp_{g,ct} + \beta_3 Exp_{g,ct}^2 + \lambda Sel_{g,ct} + \mu_{g,ct} \qquad c = 1, \dots, C\ t = 1, \dots, T\ g = men, women \tag{13}$$

$$Sel_{g,ct} = f(Married_{g,ct}, Head\_Household_{g,ct}, Ch6_{g,ct}, N\_ind_{g,ct}) \tag{14}$$

In this context, lnW_{ct} denotes the natural logarithm of hourly wages for cohort c in year t. The variable S represents years of schooling, while Exp denotes potential labor market experience, calculated as age minus years of schooling minus six. Exp^2 is the square of potential experience, capturing the nonlinear (diminishing) returns to experience. The term α accounts for unobserved heterogeneity across cohorts, and μ is the error term. The inverse Mills ratio λ is included in the wage equation to correct for selection bias, as wages are only observed for employed individuals. Excluding

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 June 2025 doi:10.20944/preprints202506.0536.v1

individuals who are not currently working (e.g., unemployed) but have invested in human capital introduces selection bias in the estimation of returns to education.

The parameter β_1 captures the return to an additional year of education, while β_2 and β_3 represent the returns to an additional year of experience and its diminishing effect, respectively. Equations (13) and (14) are estimated using the GMMC approach that corrects for selection bias, as specified in Equation (11).

Regarding the selection equation, $Sel_{i,ct}$ is a binary indicator for labor force participation, equal to one if the individual is either employed or unemployed (i.e., actively participating in the labor market), and zero otherwise. The covariates used in the selection equation include: *Married*, a binary variable equal to one if the individual is married; *Head_household*, a dichotomous variable equal to one if the individual is the head of their household; *Ch6*, a continuous variable measuring the number of children under the age of six in the household; and *N_ind*, which denotes the total number of individuals residing in the household.

According to the ILO (2020a), marital status has a differential impact by gender on labor market outcomes, particularly in labor force participation, job types, and underemployment. Being the head of household entails greater financial responsibilities and thus influences the decision to participate in the labor market (Budlender, 2003). Similarly, the presence of young children[1] and the overall household size are relevant determinants of labor market participation, as highlighted by Tobón and Rodríguez (2015), Cools et al. (2017), ILO (2020b), and Baranowska-Rataj and Matysiak (2022).

Definition 2. The counterpart of the Blinder–Oaxaca (1973) decomposition in the pseudo-panel data framework for two groups is defined as follows:

$$LnW_{m,ct} - LnW_{w,ct} = \underbrace{(X_{m,ct} - X_{w,ct})\hat{\beta}_{m,ct}}_{A} + \underbrace{X_{w,ct}(\hat{\beta}_{m,ct} - \hat{\beta}_{w,ct})}_{B} + \underbrace{(\epsilon_{m,ct} - \epsilon_{w,ct})}_{C} \tag{15}$$

$$m{:}\, men;\ w{:}\, women$$

In Equation (15), the first term (A) represents the explained component–endowment effect–, which captures the portion of the wage differential attributable to observable differences in productive characteristics between the two groups. The second term (B) corresponds to the unexplained component–remuneration effect–, which reflects differences in the returns to these characteristics and is often associated with discrimination or unobserved heterogeneity. The final term (C) tends to converge to zero, as the evaluation of Equation (15) at the mean of the logarithm of the hourly wage distribution implies that the linear combination of the error terms has an expected value of zero.

For instance, the explained component—based on education, experience, and squared experience as proxies for human capital accumulation—can be expressed as:

$$\tilde{\beta}^{S}_{m,ct}(\bar{S}_{m,ct} - \bar{S}_{w,ct}) + \tilde{\beta}^{Exp}_{m,ct}(\overline{Exp}_{m,ct} - \overline{Exp}_{w,ct}) + \tilde{\beta}^{Exp^2}_{m,ct}\left(\overline{Exp^2}_{m,ct} - \overline{Exp^2}_{w,ct}\right) \tag{16}$$

Similarly, the unexplained component is expressed as:

$$\alpha^{m}_{ct} - \alpha^{w}_{ct} + (\tilde{\beta}^{S}_{m,ct} - \tilde{\beta}^{S}_{w,ct})\bar{S}_{w,ct} + (\tilde{\beta}^{Exp}_{m,ct} - \tilde{\beta}^{Exp}_{w,ct})\overline{Exp}_{w,ct} + (\tilde{\beta}^{Exp^2}_{m,ct} - \tilde{\beta}^{Exp^2}_{w,ct})\overline{Exp^2}_{w,ct} \tag{17}$$

Definition 3. The variance-covariance matrix counterpart for the pseudo-panel data model, following Jann (2008), for two groups is specified as follows:

For the explained component–endowment effect–, the variance-covariance matrix is given by:

$$\hat{V}^{MM}{}_{explain}\left\{\left(\tilde{X}_{m,ct}-\tilde{X}_{w,ct}\right)'\hat{\beta}_{m,ct}\right\}\approx$$
$$\left(\tilde{X}_{m,ct}-\tilde{X}_{w,ct}\right)'\hat{V}\left(\hat{\beta}_{m,ct}\right)\left(\tilde{X}_{m,ct}-\tilde{X}_{w,ct}\right)+\hat{\beta}'_{m,ct}\left\{\hat{V}\left(\tilde{X}_{m,ct}\right)+\hat{V}\left(\tilde{X}_{w,ct}\right)\right\}\hat{\beta}_{m,ct} \tag{18}$$

For the unexplained component–remuneration effect–, the variance-covariance matrix is given by:

$$\hat{V}^{MM}{}_{unexplain}\left\{\tilde{X}'_{w,ct}\left(\hat{\beta}_{m,ct}-\hat{\beta}_{w,ct}\right)\right\}\approx$$
$$\tilde{X}'_{w,ct}\left\{\hat{V}\left(\hat{\beta}_{m,ct}\right)+\hat{V}\left(\hat{\beta}_{w,ct}\right)\right\}\tilde{X}_{w,ct}+\left(\hat{\beta}_{m,ct}-\hat{\beta}_{w,ct}\right)'\hat{V}\left(\tilde{X}_{w,ct}\right)\left(\hat{\beta}_{m,ct}-\hat{\beta}_{w,ct}\right) \tag{19}$$

The $trace\{\,\hat{V}(\tilde{X}_{m,ct})\,\hat{V}(\hat{\beta}_{m,ct})\,\}$ disappears when we use cohort's as instruments and NT/C→ ∞.

4. Blinder-Oaxaca Wage Gap Decomposition: The Case of Colombia

The Colombian labor market continues to exhibit significant gender disparities. For instance, according to the World Economic Forum's Global Gender Gap Report (2021), substantial wage differences persist between women and men in Colombia. Among 156 countries, Colombia ranks 120th on the equal pay index for comparable work, with a score of 0.56 (where 1 indicates full parity). Despite increased female labor force participation and longer average years of schooling among women, their earnings remain significantly lower than those of men. Data from the National Administrative Department of Statistics (DANE—its Spanish acronym) reveal that women's labor force participation rate rose from approximately 46% in 1991 to 54% in 2019, while men's participation rate has remained steady at around 75%. Additionally, Piñeros (2009) notes that the educational attainment gap between men and women began to narrow in the 1970s, with women surpassing men in average years of schooling during the 1980s.

Peña et al. (2013) emphasize that the predominant emerging family structure in Colombia is the female-headed single-parent household, where gender inequalities negatively affect family income and human capital accumulation, thereby limiting social mobility for members of such families.

The gender wage gap in Colombia has been examined at various points in time using cross-sectional data (e.g., Bernat, 2005; Fernández, 2006; Badel & Peña, 2010). For example, Fernández (2006) reports that the average wage differential was 19% in 1997 and decreased to 13% in 2003. However, a comprehensive understanding of the underlying causes for the persistence of this gap over time remains insufficient.

Badel and Peña (2010) examine the gender wage gap in Colombia's seven largest cities using quantile regression techniques. Their findings indicate that men earn more than women, and the wage gap exhibits a U-shaped pattern, with women's wages falling further below men's at the extremes of the wage distribution compared to the middle. Similarly, Galvis (2011) investigates regional and gender wage differentials in Colombia employing quantile regressions. The results reveal consistent positive wage differentials favoring men. Furthermore, a Blinder–Oaxaca decomposition suggests that these wage gaps are not fully explained by observable individual characteristics; rather, they primarily arise from differences in the returns to these characteristics (e.g., education) and unobserved factors.

Mora and Arcila (2014) analyze the wage gap between Afro-descendant and White individuals in Cali, utilizing data from the 2013 Employment and Quality of Life Survey. When incorporating variables such as migration status and perceived discrimination into the selection equation for Afro-descendants, they estimate a wage gap of 42%, of which 9% is attributable to differences in human capital characteristics, while 33% is linked to labor market discrimination.

To the best of our knowledge, although prior studies have documented the existence of the gender wage gap in Colombia (e.g., Baquero, 2001; Fernández, 2006; Peña & Badel, 2010), the present study is the first to analyze the evolution of this gap over time. Specifically, it employs a pseudo-panel dataset combined with decomposition methods and selectivity correction techniques.

To estimate the gender wage gap over time, we constructed a pseudo-panel comprising a time series of independent and representative cross-sectional samples spanning from 2016 to 2021. This pseudo-panel is based on data from the Large Integrated Household Survey (GEIH– its Spanish acronym), a multipurpose survey conducted by Colombia's official statistics agency, DANE. The GEIH regularly monitors the labor market and provides monthly labor statistics at the national, departmental, and major city levels.

Since the observations consist of independent cross-sectional data for each period, nine 5-year cohorts of individuals aged 18 to 63 have been defined. The sample comprises a total of 840,499 individuals. Table 1 displays the distribution of the sample by cohort and year. Each cohort includes more than 5,500 individuals. The cohort representing the youngest age group has an average of 16,155 individuals per year, whereas the oldest cohort has an average of 8,876 individuals per year.

Table 1. Number of Individuals by Cohort.

Cohort, $C_{i(t)}$	2016	2017	2018	2019	2020	2021	Total
18–22 years old	18,986	18,402	17,679	17,974	9,138	14,750	96,929
23–27 years old	22,024	21,729	21,909	22,677	11,746	18,789	118,874
28–32 years old	20,119	20,198	20,255	21,709	11,638	18,627	112,546
33–37 years old	19,545	19,358	19,324	20,011	10,667	17,193	106,098
38–42 years old	16,405	16,449	16,954	18,636	9,953	16,588	94,985
43–47 years old	16,028	15,396	14,931	16,039	8,342	13,827	84,563
48–52 years old	15,814	15,448	15,432	15,601	8,206	13,465	83,966
53–58 years old	15,888	16,018	16,312	17,093	9,198	14,771	89,280
59–63 years old	8,904	9,353	9,662	10,283	5,600	9,456	53,258
Total	153,713	152,351	152,458	160,023	84,488	137,466	840,499

Source: author's own calculations.

Regarding the number of individuals per cohort, Verbeek and Nijman (1992) assert that including at least 100 individuals per cohort is sufficient to mitigate sampling errors. Descriptive statistics of the variables are provided in Appendix A.

Gender wage gaps in Colombia have been notable for their persistence over time. Despite increases in women's average years of schooling and labor market participation in recent decades, empirical evidence consistently shows that men continue to receive higher remuneration than women.

Table 2 presents the results of the gender wage gap decomposition without accounting for selection bias. The models considered include a pooled cross-section, a pooled cohort, and a pseudo-panel with Deaton's correction. It is important to note that without applying Deaton's correction, the model essentially becomes an error-in-variables model, wherein all explanatory variables (except dummy variables) are subject to measurement error (Deaton, 1985).

Table 2. Decomposition Results without Selection Bias Correction.

Without Selection Bias			
	Pool	**Pseudo Panel – Pooled Cohort²**	**Pseudo Panel (Deaton)**
Differential	10.12987***	0.14467***	0.20106***
	(0.00181)	(0.02554)	(0.0000779809)
Explained – Endowment Effect	−0.08622***	−0.05064**	−0.02020***
	(0.00100)	(0.02434)	(0.0000000016)

 doi:10.20944/preprints202506.0536.v1

Unexplained –Remuneration Effect	0.21609***	0.19531***	0.22126***
	(0.00152)	(0.00887)	(0.0000779794)
NT, Cohorts	771,194	108	108

Note: *** Statistically significant at the 0.01 level; ** statistically significant at the 0.05 level; * statistically significant at the 0.1 level. Standard errors appear in parentheses ().

When calculating the wage gap in the Colombian urban labor market, it is found that women earn, on average, 13% less than men in the pooled configuration. The endowment component is approximately −8.6%, indicating that the difference in observable characteristics favors women. In this regard, women possess superior attributes that enhance productivity (e.g., human capital and work experience) compared to men. This finding corroborates previous studies that have documented women's higher average years of schooling relative to men (Abadía, 2005; Galvis, 2011).

Regarding the unexplained component, the estimated effect is 21.6%. This suggests that if men and women had equivalent endowments, a substantial wage gap would still persist, indicating that gender differences in wages cannot be fully accounted for by productivity-related attributes or other supply-side factors.

In the pseudo-panel configurations, the wage differential is 14.5% without correction for measurement error, and increases to 20.1% when such correction is applied. This implies that male cohorts earn, on average, 20% more than female cohorts in the Colombian urban labor market. The endowment effect in this setting is −2%, while the remuneration effect accounts for 22%, indicating that the unexplained component exceeds the total observed wage differential between men and women cohorts.

It is important to note that these regression results are subject to bias, as they do not adjust for selection bias, given that not all individuals participating in the labor market receive wages (Heckman, 1979).

Table 3 presents the results of the wage gap estimation with correction for selection bias, following Mora and Herrera (MH):

Table 3. Decomposition Results with Selection Bias Correction.

	With Selection Bias	
	Pool	**Pseudo Panel (MH)**
Differential	0.21186***	0.14617***
	(0.00203)	(0.00005)
Explained– Endowment Effect	−0.08622***	−0.05636***
	(0.00100)	(0.000000025)
Unexplained –Remuneration Effect	0.29809***	0.20254***
	(0.00179)	(0.00005)
NT, Cohorts	840,499	108

Note: *** Statistically significant at the 0.01 level; ** statistically significant at the 0.05 level; * statistically significant at the 0.1 level. Standard errors appear in parentheses.

When adjusting for selection bias, the results from the pooled configuration become more pronounced, with the total wage differential increasing to 21%. Notably, while the endowment effect remains negative, indicating that women possess, on average, more productive characteristics, the remuneration effect rises to approximately 30%. In the pseudo-panel framework, the estimated wage gap stands at 14.6% in favor of male cohorts, with the explained component at −5.6% and the unexplained component at 20%.

Within this context, the explained component captures gender differentials associated with variations in returns to individuals' observable characteristics. The residual unexplained component is often interpreted as a proxy for labor market discrimination. However, it is important to emphasize

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 June 2025 doi:10.20944/preprints202506.0536.v1

that these estimates are indicative rather than definitive, since the unexplained portion may also encompass differences in unobservable attributes not captured by the model.

6. Conclusions

The Blinder–Oaxaca methodology is a widely recognized approach for estimating gender wage differentials between two groups. Jann (2005, 2008) enhances this methodology by providing a correction for the variance–covariance matrix, which ensures efficiency in cross-sectional estimations. While Kroger and Hartmann (2021) discuss the decomposition effects in panel data, such extensions are not directly applicable to pseudo-panel data. This limitation arises because the individuals observed in period t are not the same as those observed in period t–1, and unobserved individual heterogeneity may vary across time.

Many developing countries, such as Colombia, lack true panel data structures but possess independent repeated cross-sectional data. In this context, the pseudo-panel approach provides a practical alternative for analyzing labor market outcomes over time, particularly in the absence of longitudinal tracking of individuals.

As in many other countries, gender wage disparities persist in Colombia. Our empirical findings, based on a pseudo-panel configuration with corrections for both measurement error and selection bias, consistently show wage differentials favoring men in the Colombian urban labor market. The results indicate that female cohorts earn, on average, 15% less than their male counterparts. Importantly, this wage gap is not primarily attributable to differences in observable attributes such as education or experience. Instead, the bulk of the gap is explained by differential returns to these attributes and potentially by unobservable factors, highlighting the presence of labor market discrimination.

The Colombian labor economics literature has repeatedly documented that women have increased their labor force participation and now exhibit, on average, higher educational attainment than men. Nevertheless, the gender gap in the returns to human capital remains significant, suggesting that women are not equally rewarded for their skills and qualifications in the labor market.

The review of previous studies and empirical evidence supports the conclusion that wage differentials between men and women persist in Colombia and that human capital endowments explain only a limited portion of this gap.

To address the gender wage gap, a range of policy interventions and institutional efforts can be implemented. These include promoting equal access to education and vocational training to ensure women acquire skills aligned with labor market demands. Scholarship programs and mentorship initiatives can also encourage women's participation in male-dominated fields of study. Furthermore, workplace-level reforms are critical. These include the enforcement of anti-discrimination policies in hiring and compensation, as well as measures to support work–life balance, such as flexible work arrangements and remote work options. Such policies can facilitate women's sustained engagement in the labor market and contribute to reducing the persistent gender pay gap.

Data availability statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest: The authors declare that they have no conflict of interest.

Appendix A

Table A1. Descriptive statistics by year.

	Variable	$C_{i(t)}$	mean	Std . dev .	min	Max		Variable	$C_{i(t)}$	mean	Std . dev .	min	Max
2016	LnW_men	9	8.338058	0.1245341	8.035196	8.432405	2017	LnW_men	9	8.408438	0.1203705	8.118181	8.498527
	S_men	9	10.93983	1.164774	9.069445	12.35118		S_men	9	11.00728	1.18347	9.130664	12.48615
	Exp_men	9	29.20804	14.86316	9.020644	51.70675		Exp_men	9	29.15555	14.88868	8.99544	51.67625
	Exp2_men	9	1076,076	915.6101	92.67495	2712.067		Exp2_men	9	1073.002	916.5095	91.70228	2708.582
	LnW_women	9	8.150793	0.1214262	7.943633	8.285857		LnW_women	9	8.237926	0.1204845	8.048492	8.371805
	S_women	9	11.62916	1.744159	8.663899	13.75622		S_women	9	11.71577	1.723189	8.829645	13.83775
	Exp_women	9	28.5162	15.45154	7.635523	52.12373		Exp_women	9	28.4318	15.42173	7.621135	51.89346
	Exp2_women	9	1051,439	936.6272	67.30845	2754,966		Exp2_women	9	1045.286	931,609	66.8628	2730.813
	Married_men	9	0.5842465	0.2132395	0.1192248	0.7230747		Married_men	9	0.5760853	0.212277	0.1159624	0.7180166
	Head_men	9	0.5503303	0.2342715	0.0946546	0.7713839		Head_men	9	0.5390329	0.2303175	0.0942274	0.7532307
	Ch6_men	9	0.3380849	0.1352057	0.1891162	0.5493901		Ch6_men	9	0.3341325	0.1264513	0.1976904	0.5283061
	N_ind_men	9	0.2505849	0.051748	0.2015876	0.3492341		N_ind_men	9	0.2541971	0.0552963	0.2006257	0.3659145
	Married_women	9	0.5212701	0.1220463	0.2378164	0.622967		Married_women	9	0.5153326	0.1216957	0.2327136	0.6176041
	Head_women	9	0.289074	0.1262674	0.0703748	0.4455163		Head_women	9	0.2943324	0.1281973	0.0743065	0.4519494
	Ch6_women	9	0.1721031	0.086282	0.0880289	0.2974964		Ch6_women	9	0.1657271	0.0820278	0.0860906	0.2940533
	N_ind_women	9	0.0330633	0.0106235	0.0195133	0.0541875		N_ind_women	9	0.033518	0.0094911	0.0246975	0.0541904
	Sel_men	9	0.8958672	0.1040352	0.6712098	0.9704566		Sel_men	9	0.8946891	0.1060642	0.6594643	0.9698599
	Sel_women	9	0.7036599	0.1388076	0.4290726	0.8216574		Sel_women	9	0.70193	0.1403489	0.4323498	0.8267639
2018	LnW_men	9	8.445422	0.1180687	8.159978	8.545579	2019	LnW_men	9	8.475406	0.1285858	8.16323	8.568756
	S_men	9	11.13864	1.156181	9.289095	12.51258		S_men	9	11.21744	1.135674	9.473826	12.58708
	Exp_men	9	29.01479	14.84102	8.98698	51.51229		Exp_men	9	28.94681	14.80999	9.066341	51.34639
	Exp2_men	9	1062,879	911.2461	91.50246	2690.313		Exp2_men	9	1057,985	907.84	92.34877	2672,988
	LnW_women	9	8.296329	0.1236226	8.101408	8.434343		LnW_women	9	8.338293	0.1190225	8.136816	8.463536
	S_women	9	11.89469	1.723275	8.905832	13.98808		S_women	9	12.0792	1.585988	9.441266	14.00982
	Exp_women	9	28.26562	15.41085	7.662983	51.89826		Exp_women	9	28.08228	15.26059	7.72789	51.40152
	Exp2_women	9	1035.102	929.6148	67.33604	2730.174		Exp2_women	9	1020,683	915.0557	68.05177	2680,442
	Married_men	9	0.5681512	0.2084602	0.1166601	0.7080629		Married_men	9	0.5667591	0.209994	0.1172997	0.7156888
	Head_men	9	0.5289457	0.2265079	0.0928693	0.7390612		Head_men	9	0.5252956	0.2250406	0.0925015	0.7454165

	Ch6_men	9	0.3327424	0.1287793	0.1768762	0.5322238		Ch6_men	9	0.331331	0.1293698	0.1777108	0.5234528
	N_ind_men	9	0.2569053	0.0536286	0.2153275	0.3665618		N_ind_men	9	0.2685936	0.0520334	0.2227821	0.3709244
	Married_women	9	0.5185747	0.1230933	0.2298088	0.6170303		Married_women	9	0.5143612	0.1229235	0.2285872	0.6170322
	Head_women	9	0.293476	0.1261038	0.0730442	0.4463721		Head_women	9	0.2994137	0.1250416	0.0776866	0.4414686
	Ch6_women	9	0.1665284	0.0856539	0.077763	0.2974278		Ch6_women	9	0.1577043	0.0800504	0.0756345	0.2802293
	N_ind_women	9	0.0337784	0.0097015	0.0231715	0.0525967		N_ind_women	9	0.0354072	0.0095942	0.0249269	0.0527509
	Sel_men	9	0.8877906	0.1115676	0.6329794	0.9648866		Sel_men	9	0.8831719	0.1121991	0.6234337	0.964309
	Sel_women	9	0.6930691	0.1415745	0.4252475	0.8230862		Sel_women	9	0.6840926	0.1453998	0.417093	0.8179824
	LnW_men	9	8.451991	0.1255018	8.145081	8.555027		LnW_men	9	8.47214	0.1247396	8.172038	8.570444
	S_men	9	11.35415	1.061942	9.707747	12.54652		S_men	9	11.44236	1.049986	9.814035	12.68768
	Exp_men	9	28.80064	14.67451	9.100624	51.03596		Exp_men	9	28.75464	14.65834	9.186735	51.00526
	Exp2_men	9	1045,842	897.0726	92.93246	2641,078		Exp2_men	9	1041,916	895.3744	93.51768	2635.418
	LnW_women	9	8.3469	0.1261775	8.129876	8.501195		LnW_women	9	8.353196	0.1280188	8.140433	8.490309
	S_women	9	12.33815	1.553494	9.696412	14.17378		S_women	9	12.37938	1.522375	9.717836	14.1562
	Exp_women	9	27.84481	15.19051	7.795722	51.12937		Exp_women	9	27.79751	15.10841	7.875242	51.04327
	Exp2_women	9	1004.76	907.1082	68.33904	2652.217		Exp2_women	9	999.5638	901.2829	69.90714	2643.216
2020	Married_men	9	0.5629667	0.2038836	0.1207701	0.707759	2021	Married_men	9	0.5545357	0.2055342	0.1129931	0.6995674
	Head_men	9	0.5084341	0.2216666	0.0748535	0.7233106		Head_men	9	0.4992236	0.2173066	0.0814947	0.7124131
	Ch6_men	9	0.318702	0.1272451	0.1639886	0.5214967		Ch6_men	9	0.2951451	0.1233416	0.1492987	0.4826719
	N_ind_men	9	0.3616837	0.0668421	0.3057756	0.4843956		N_ind_men	9	0.331973	0.0668036	0.279148	0.4584355
	Married_women	9	0.5096974	0.1181309	0.2340668	0.6071287		Married_women	9	0.5021007	0.1186671	0.2241888	0.5943267
	Head_women	9	0.3081704	0.1309637	0.065785	0.4649241		Head_women	9	0.3263767	0.1318954	0.0757831	0.4637941
	Ch6_women	9	0.1507033	0.0777763	0.0745363	0.2614055		Ch6_women	9	0.1371094	0.0768648	0.0589467	0.253281
	N_ind_women	9	0.0587062	0.0153298	0.0412137	0.0909816		N_ind_women	9	0.0539215	0.0152554	0.0368529	0.0838601
	Sel_men	9	0.8729501	0.1160677	0.6118618	0.9588074		Sel_men	9	0.8725552	0.118324	0.6102073	0.9583185
	Sel_women	9	0.663856	0.145117	0.3924115	0.7944374		Sel_women	9	0.6606521	0.1521143	0.371933	0.796875

Source: author's own calculations.

doi:10.20944/preprints202506.0536.v1

Notes

1. The number of children under six years of age in a household is related to the participation decision but not to wage, as used by Heckman (1974).
2. Since two groups of nine cohorts each are observed over six time periods, the total number of cohort-period observations used in the decomposition amounts to 108.

References

Abadía, L. K. (2005). Discriminación salarial por sexo en Colombia: un análisis desde la discriminación estadística. Documentos de Economía, Universidad Javeriana-Bogotá.

Badel, A., & Peña, X. (2010). Decomposing the gender wage gap with sample selection adjustment: evidence from Colombia. Revista de análisis económico, 25(2), 169-191.

Baquero, J. (2001) Estimación de la discriminación salarial por género para los trabajadores asalariados urbanos de Colombia (1984-1999). Informe técnico, Universidad del Rosario, Facultad de Economía.

Baranowska-Rataj, A., & Matysiak, A. (2022). Family Size and Men's Labor Market Outcomes: Do Social Beliefs About Men's Roles in the Family Matter? Feminist Economics, 28(2), 93-118, doi: 10.1080/13545701.2021.2015076

Bernat, L. (2005). Análisis de género de las diferencias salariales en las siete principales Áreas Metropolitanas colombianas: ¿Evidencia de discriminación? Documento PNUD.

Biltagy, M. (2014). Estimation of Gender Wage Differentials using Oaxaca Decomposition Technique. Topics in Middle Eastern and North African Economies, 16(1), 17–42.

Blau, F. D., & Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. Journal of Economic Literature, 55(3), 789–865. doi: 10.1257/jel.20160995

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. The Journal of Human Resources, 8(4), 436-455. doi: 10.2307/144855

Buchely, L. (2013). Overcoming Gender Disadvantages. Social Policy: Analysis of urban middle-class women in Colombia. Revista de Economía del Rosario, 16(2), 313-340.

Budlender, D. (2003). The Debate about Household Headship. Social Dynamics, 29(2), 48–72. doi: 10.1080/02533950308628675

Cools, S., Markussen, S., & Strøm, M. (2017). Children and Careers: How Family Size Affects Parents' Labor Market Outcomes in the Long Run. Demography, 54, 1773–1793. doi: 10.1007/s13524-017-0612-0

Deaton, A. (1985). Panel data from a time series of cross-sections. Journal of Econometrics, pp. 30, 109–125.

Di Paola, R., & Berges, M. (2000). Sesgo de selección y estimación de la brecha por género para Mar del Plata. Nülan. Deposited Documents, 891, Universidad Nacional de Mar del Plata, Facultad de Ciencias Económicas y Sociales, Centro de Documentación.

Fernández, M. (2006). Determinantes del diferencial salarial por género en Colombia, 1997-2003. Revista Desarrollo y Sociedad, 58, 165-208.

Galvis, L. A. (2011). Diferenciales salariales por género y región en Colombia: una aproximación con regresión por cuantiles. Revista de Economía del Rosario, 13(2), 235-277.

Greene, W. H. (2003). Econometric Analysis. 5th Edition, Prentice Hall, Upper Saddle River.

Guillermo, M. (2017). Pseudo-panel methods and an example of application to Household Wealth data. Economie et Statistique / Economics and Statistics, 491-492, 109-130.

Hansen, L. P. (1982). Large Sample Properties of Generalized Methods of Moments Estimators. Econometrica, 50(4), 1029-1054. doi: 10.2307/1912775

Heckman, J. (1974). Shadow Prices, Market Wages, and Labor Supply. Econometrica, 42(4), 679. doi:10.2307/1913937

Heckman, J. (1979). Sample selection bias as a specification error. Econometrica, 47(1).

International Labour Organization. [ILO]. (2020a, May 15th). International Day of Families: How marital status shapes labour market outcomes. ILO. https://ilostat.ilo.org/international-day-of-families-how-marital-status-shapes-labour-market-outcomes/

International Labour Organization. [ILO]. (2020b, March 3rd). Having kids sets back women's labour force participation more so than getting married. ILO. https://ilostat.ilo.org/having-kids-sets-back-womens-labour-force-participation-more-so-than-getting-married/

Jann, B. (2005, April 8th). Standard Errors for the Blinder-Oaxaca Decomposition. German Stata Users Group Meetings, Stata Users Group, Berlín, Alemania.

Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. The Stata Journal, 8(4), 453–479. doi: 10.1177/1536867X0800800401

Johansson, M., Katz, K., & Nyman, H. (2005). Wage Differentials and Gender Discrimination: Changes in Sweden 1981–98. Acta Sociologica, 48(4), 341–364. doi: 10.1177/0001699305059946

Kroger, H., & Hartmann, J. (2021). Extending the Kitagawa Oaxaca Blinder decomposition approach to panel data. The Stata Journal: Promoting Communications on Statistics and Stata, 21(2), 360410. https://doi.org/10.1177/1536867X211025800

Mincer, J. (1974). Schooling, Experience, and Earnings. New York: National Bureau of Economic Research.

Mora, J.J., & Arcila, A.M. (2014). Brechas salariales por etnia y ubicación geográfica en Santiago de Cali. Revista de Métodos Cuantitativos para la Economía y la Empresa, Universidad Pablo de Olavide, 18(1), 34-53.

Mora, J.J., & Muro, J. (2014). Consistent estimation in pseudo panels in the presence of selection bias. Economics, 8, 1-25. doi: 10.5018/economics-ejournal.ja.2014-43

Newey, W.K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In R. Engle & D. McFadden (Eds.), Handbook of Econometrics (2111–2245). North-Holland. doi: 10.1016/S1573-4412(05)80005-4

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. International Economic Review, 14(3), 693–709.

Oaxaca, R., & Ransom, M. R. (1998). Calculation of approximate variances for wage decomposition differentials. Journal of Economic and Social Measurement, 24, 55–61. doi: 10.3233/JEM-1998-0143

Paz, J. (1998). Brecha de ingresos entre géneros (Comparación entre el Gran Buenos Aires y el Noroeste Argentino). Anales de la AAEP.

Peña, X., Cárdenas, J.C., Ñopo, H., Castañeda, J.L., Muñoz, J.S., & Uribe, C. (2013). Mujer y movilidad social. Documentos CEDE. Universidad de los Andes.

Piñeros, L. A. (2009). Las Uniones Maritales, los Diferenciales Salariales y la Brecha Educativa en Colombia, Revista Desarrollo y Sociedad, 64, 55-84.

Tobón, C., & Rodríguez, F. L. (2015). Factores que determinan la probabilidad de participación laboral en el área metropolitana de Medellín [tesis de maestría, Universidad EAFIT]. Universidad EAFIT. https://repository.eafit.edu.co/server/api/core/bitstreams/e15372c3-81b2-4e62-b53b-94dbd9e2ef4f/content

Verbeek, M., & Nijman, T. (1992). Testing for Selectivity Bias in Panel Data Models. International Economic Review, 33(3), 681.

Watson, I. (2010). Decomposing the Gender Pay Gap in the Australian Managerial Labour Market. Australian Journal of Labour Economics, 13(1), 49–79.

World Economic Forum (2021). Global Gender Gap Report 2021. Insight Report. https://www3.weforum.org/docs/WEF_GGGR_2021.pdf