The Gender Wage Gap in Developing Countries

Chengyu Si[1], Denis Nadolnyak[2], and Valentina Hartarska[2]

[1] Mavenform Inc. (dba Pathrise), San Francisco, California, USA

[2] Auburn University, Auburn, Alabama, USA

Correspondence: Valentina Hartarska, Department of Agricultural Economics and Rural Sociology, Auburn University, Auburn, Alabama, USA.

Abstract

The labor literature documents a gender wage gap between the earnings of men and women in the developed countries but research on the gap in developing countries is less common and usually focuses on individual countries. Research typically identifies a part of the wage gap that is due to observable characteristics, or endowments, and an unexplained part attributable to differences in gender wage equation coefficients sometimes interpreted as gender bias. The focus of recent gender wage gap research has been on understanding how the gender gap varies along the earnings distribution. In this paper, we investigate the determinants of the gender wage gap across 12 developing countries using data from the World Bank's STEP (Skills Toward Employment and Productivity) Measurement survey that provides information on household demographic characteristics, education and training, employment, and job skills. We apply the classical Blinder-Oaxaca decomposition and combined with quantile regression and also use decomposition methods based on the recentered influence function to obtain distributional decomposition of the wage gap by covariates and by quantiles. Our results show that there are significant gender pay differences that generally decrease with earnings and are mostly due to differences in the intercepts and a slightly offsetting difference in returns to education that are higher for women. The explained part of the gender wage gap, although much smaller, is attributed almost entirely to lower educational level of women and the motherhood penalty. The main conclusion is that improvements in women's education and training, as well as child friendly labor policies, remain the most important factors that can help to narrow the gender gap in developing countries, which is consistent with previous work.

Keywords: gender wage gap, Blinder-Oaxaca decomposition, quantile regression, recentered influence function

1. Introduction

A gap between the earnings of men and women — the gender wage gap — has been widely documented across counties and time. Much of the research attributes the gender wage gap to the differences in educational level, experience, occupational choices, and gender discrimination. Within the household, the distribution of family duties and responsibilities also plays a significant role, particularly in families with children when outside the family childcare options are limited. In societies with significant gender wage gaps, and at lower income levels, women may specialize in performing household duties and raising children, which gives men the opportunity to work more and incentivizes them to improve their earning capacity. The family wage structure is likely connected to the spouses educational and occupational choices. Depending on the income level, age, and a number of other factors, the dynamics of gender specialization may have complex positive or negative feedbacks, reinforcing or curbing the gender wage gap that itself can lead to significant macroeconomic losses due to distortions in the occupational choices by men and women and gaps in labor force participation.

While measuring the gender wage gap is important, understanding its distributional effects in developing countries is of particular interest because it helps to identify the source of social inequality and to inform policy makers on efficient ways to raise household earnings, such as improving access to education and training programs for certain groups of the population or encouraging parent friendly job creation. In this regard, analyzing the gender wage gap at different wage distributional levels is timely and important and is the focus of this research. More concretely, we investigate the factors that determine the gender wage gap at different income levels in developing countries.

Gender wage discrimination has been the subject of research for several decades. In a summary of early research, Altonji, Joseph & Blank (1999) discuss the causes of the gender wage gap and note the role of pre-market human capital in education and family background, different return to experience and seniority, job characteristics, and the "unexplained gap", which is associated with gender wage discrimination. More recently, Blau and Kahn (2017) analyzing US data up to 2010 find that conventional human capital variables did not explain much of the gender wage gap in the US. Their analysis of the distribution of the wage gap showed that the gap declines much more slowly at the top of the wage distribution than at the middle or bottom and that, by 2010, it was noticeably higher at the top. Summarizing the vast literature, the authors suggest that some of the gender wage differences were attributable to gender roles and the gender division of labor. Furthermore, their review of experimental evidence reveals significant role for discrimination, with psychological attributes or non-cognitive skills used to explain the gender differences in outcomes. Uncovering the importance of labor force participation interruption and shorter hours in high skills occupations suggests that, among other factors, the number of children within a family has an important negative effect on women's wages.

The research is clear that women in the labor market are subject to motherhood penalty (Anderson, Deborah, Binder & Krause, 2002; Kleven, Henrik, Camille & Søgaard, 2019; Agüero, Jorge, Marks & Raykar, 2020). Motherhood is associated with lower hourly pay for reasons such as sacrificing job experience, lower productivity at work, and trade-offs for mother-friendly jobs or other discriminations by employers. The childbearing itself has negative, large, and long-lasting effects on earning, the effects being driven by earnings and not by labor supply, and stronger at the intensive margin (Lundborg, Petter, Plug & Rasmussen, 2017). In the US, the motherhood penalty is associated with a wage penalty of 7% per child (Budig and England, 2001). More recent evidence shows contradictory results about the trend of the motherhood penalty by skill level. For example, according to Glauber (2018), the overall trend decreased between 1980-2014 in the US and, while it continued to exist at low-earning levels, it was all but eliminated for high-earning female workers. Yet, England, Bearak, Budig & Hodges (2016) find that, for the period 1979 - 2010, motherhood penalty continued to exist among women with high skill and high incomes, while differences attributable to employer discrimination were negligible.

Wage inequality might be driven by different factors in developing and developed countries. First, work focused on the returns on investment in education has found differences in return to education by gender, which is important as education is a main factor affecting wages and the wage gap (Psacharopoulosa and Patrinos, 2018). Gonzalez and Miles (2001), for example, argue that differences in increased returns to education explain the wage gap in Uruguay, a developing country, while in developed countries the wage inequality between men and women is attributable to differences in the real minimum wage. In Latin America, evidence shows a widening of the wage differential between skilled (more educated) and unskilled workers since 1980s as a result of its increased trade openness affecting disproportionately men and women (Wood, 1997). A recent analysis of the gender wage gap of in highly skilled physicians in Peru points out that the unexplained component, which is associated with gender discrimination, seem to be the main factor driving the gap (Amaya & Mougenot, 2019). Overall, however, the research on the wage gap in developing countries is limited and dispersed. The present research is filling in a gap by evaluating the gender wage differences along the earnings distribution for 12 developing and middle-income countries and, by identifying the main factors that drive this gap, provides fresher and more nuanced policy relevant insights.

One of the most popular methods to evaluate wage discrimination is the Blinder-Oaxaca decomposition (Blinder, 1973; Oaxaca, 1973). It decomposes the difference in the mean wages of men and women into two components, one part explained by workers' characteristics (endowments) and another unexplained component, which is due to differences in the coefficient estimates (returns to endowments) that is usually regarded as gender discrimination. The Oaxaca-Blinder decomposition has several limitations. Ideally, a perfect evaluation method would address the issue of self-selection into labor force, account for characteristics of unmatched men and women, and allow to account for different wage distribution. There are indeed few studies, if any, that meet all the three requirements simultaneously. For example, Nopo's (2008) method is based on matching but lacks distributional analysis. Others performed various decomposition with or without accounting for self-selection but did not account for the characteristics of the unmatched men and women (e.g., Albrecht, Van Vuuren & Vroman, 2009).

The present work is closer to research that focuses on the wage inequality at different distributional levels addressing the challenge to explore the distributional differences in the wage gap among different quantiles. The simplest way to evaluate distributional differences is estimating them across different wage levels (Melly, 2005; Albrecht, Björklund, & Vroman, 2003). The alternative, on which we base our work, is to analyze the wage gap along the distribution using the approach by Machado and Mata (2005). The basic idea is to decompose the changes in the wage distribution in several factors contributing to those changes and to perform counterfactual exercises to identify the corresponding explained and unexplained sources of wage inequality by quantile. However, the basic Machado-Mata decomposition does not tell

us what the individual contributions of the repressors are. To find these individual influences, we estimate the Re-centered Influence Functions (RIFs) proposed by Firpo, Fortin, and Lemieux (2009).

There are numerous studies applying this and related methods to a single developing or developed country. Nguyen, Albrecht, Vroman, & Westbrook, (2007) investigate the welfare inequality between urban and rural areas in Vietnam and find that the gap is due to the difference in education, ethnicity, and age for the lower quantiles and the difference between urban and rural sectors for the higher quantiles. Montes-Rojas, Siga, & Mainali (2017) investigate the caste wage differentials in Nepal and find that the effects of occupation and firm size are uniform across quantiles while, for the low quantiles, education has a large effect on the wage gap. Gardeazabal and Ugidos (2005) measure the gender wage gap in Spain and find that gender wage discrimination reaches the highest point at ninetieth percentile.

In this paper, we explore the causes of wage inequality at different income levels in 12 developing and middle-income countries using Oaxa-Blinder and Machado-Mata decompositions, as well as the RIF regressions proposed by Firpo et al. (2009). The objective is to provide detailed information about the explained and unexplained parts of the wage gap and characterize the distributional characteristics of this gap for the overall sample and by subgroups of countries. We use the World Bank's STEP (Skills Toward Employment and Productivity) Measurement household survey data from 12 developing countries in several world regions to estimate the gender-specific wage equation based on Mincer's (1974) theory. The distributional effect is captured by our use of quantile regressions. To break the gender wage gap into explained (i.e., due to differences in the explanatory variables) and unexplained (due to differences in coefficients) parts, we apply the Blinder-Oaxaca wage decomposition method and then estimate quantile decomposition models to find distributional differences in the gender bias.

Our results show that the significant gender differences within earnings quantiles are mostly due to unexplained factors usually attributed to gender bias and their share in the gap increases with the earnings quantiles. We also find that traditional education, as well as additional certifications and training, have the highest explanatory power suggesting that improving these attributes can help close the gender gap in developing countries, which is consistent with previous work. Also consistent with previous work, our findings show that child friendly labor policies would also help narrow the gender gap. The rest of the paper is organized as follows. Section 2 and 3 describe the data and methodology used in the analysis. Section 4 presents discussion of the results, and Section 5 concludes.

2. Data

The data come from the World Bank's STEP (Skills Toward Employment and Productivity) Measurement household survey data collected in either 2012 or 2013 in 12 developing countries. The survey provides information on the distribution of job skills and the demand for these skills in the labor markets of developing countries. The scope of the survey includes household demographic characteristics, education and training, employment, job skill requirements, and family characteristics. The countries include Colombia, Ghana, Kenya, Sri Lanka, Ukraine, Armenia, Lao PDR, Macedonia, Vietnam, China, Bolivia, and Georgia that cover the regions of Central America, Sub-Saharan Africa, Eastern & Southern Europe, and South & East Asia. These regions have different cultures, institutional environments, and level of development, which makes the dataset a good sample for searching for causes and remedies for the gender wage gap in the developing world.

Table 1 shows variable definitions used in the regression and includes the log of hourly earnings, education and job related training and certificates, age and job experience (tenure), number of children, and rural/urban residence.

Table 1. Variable Definitions

Variable Name	Definition
ln_earnings_usd	USD log of Hourly earnings, adjusted by PPP
edu	Number of years of education
age	Age in years
tenure	Number of years in current job
training	Participated in a training course such as work-related training or private skills training that lasted at least 5 days/ 30 hours (not part of the formal educational system) in the last 12 months, 1=yes, 0=no.
certificate	An industry-recognized or government certificate in a particular field (not from a formal ed. Institution). 1=yes, 0=no.
children	Number of children under 6 years old
urban	1=yes, 0=no

Figure 1 shows the kernel density plot of hourly earnings for males and females. While the wage distributions are similar, the density of females' wages is larger at low earnings levels and lower at higher levels, indicating that, for the entire sample, female's hourly earnings are lower than males.
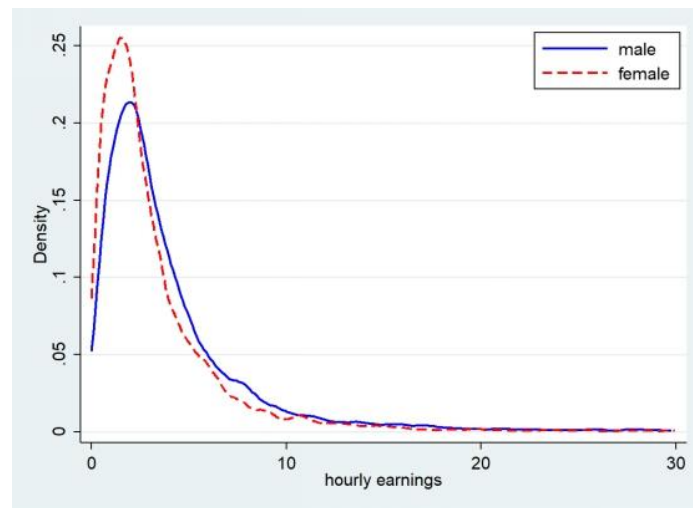

Figure 1. Kernel Density Plot of Hourly Earnings

Table 2. Summary Statistics

Male					
Variable	Obs	Mean	Std. Dev.	Min	Max
ln_earning~d	8,788	1.0328	1.0175	-7.708	6.5461
edu	14,180	11.061	4.054	0	28
age	14,322	36.503	13.751	15	64
tenure	9,890	8.0192	9.0851	0	70
training	14,314	0.0975	0.2967	0	1
certificate	14,322	0.1184	0.3231	0	1
children	14,322	0.3313	0.6341	0	7
urban	14,322	0.926	0.2618	0	1
Female					
Variable	Obs	Mean	Std. Dev.	Min	Max
ln_earning~d	9,580	0.7596	1.073	-7.1484	6.2947
edu	20,840	10.709	4.3482	0	28
age	21,247	37.457	13.632	15	64
tenure	11,011	8.0255	9.1858	0	53
training	21,241	0.0752	0.2637	0	1
certificate	21,247	0.0977	0.2969	0	1
children	21,247	0.4251	0.6862	0	6
urban	21,247	0.9272	0.2598	0	1

Table 2 shows summary statistics. The mean value of log hourly earnings is 1.0328 for males and 0.7596 for females (the dollar values are adjusted for PPP). Men, on average, have slightly higher education relative to women: 11.06 versus 10.7 years. The average age of male and female respondents is 36.5 and 37.5, respectively. Tenure, defined as the number of months in current job which can be considered as a proxy for experience is close to 96.2 for both males and females. Training and certificate are dummy variables, where 1 denotes "has training/certificate in the past 12 months" and 0 denotes no. The average values show that 9.7% and 7.5% of male and female respondents participated in a training course in 12 months prior to the survey, whereas 11% of men and 9.8% of women had a certificate. The average number of children reported by men is 0.33, compared to 0.43 reported by women.

4. Methodology

4.1 Wage Equation

First, we estimate a simple OLS wage equation based on Mincer's (1974) theory and estimated separately for men and women. The equation is where denotes gender. The estimated coefficients provide the percent change in the hourly wage corresponding to a unit change in the independent variables. The regression is run with country fixed effects to capture country-specific heterogeneity.

4.2 Quantile Regression

Next, we explore the gender gap differences within the earnings distribution using quantile regressions. The idea of quantile regression was first developed by Koenker and Bassett (1978). Unlike the OLS that is based on conditional means, quantile regression allows for distributional (quantile) variation of regressors making the estimation more robust to outliers. Quantile regressions estimate conditional quantiles of , which is written as where is the quantile of y such that and the marginal effect is calculated as

4.3 Wage Decomposition

The purpose of this analysis is to show what difference in wages can be explained by the differences in the in observable characteristics of men and women (endowments) and by the differences in coefficients that is usually attributed to unexplained factors that can be related to gender discrimination. For the basic analysis, we apply the Blinder-Oaxaca decomposition, which is based on the wage equation that is estimated on conditional mean: where and are the estimated coefficients from the male/female wage equations (2) and (3). Equations (4) and (5) are two possible decompositions assuming that the wages of males or females are non-discriminatory. The first terms on the right-hand side of equations (4) and (5) represent the gender wage difference in explanatory variables (i.e., the explained difference in endowments) whereas the second terms show the unexplained difference (due to the difference in coefficients) usually attributed to gender discrimination.

Since we are interested not only in the average effect but also in the distributional wage decomposition, we apply the Machado and Mata (MM) approach (Machado and Mata, 2005; Nguyen et al. 2007). Similarly to the Blinder-Oaxaca method, the MM approach decomposes the gender wage gap into two factors: the contribution of the difference in the male and female endowment attributes (explanatory variables in X) and the gender difference in the coefficients including the intercepts (the unexplained difference in returns to endowments attributed to gender bias). The idea of the MM approach is a counterfactual analysis that is based on quantile regression. The approach generates the counterfactual distribution of females' wages as if they have males' return to endowment rates (the coefficients in male's wage equation). The procedure of the MM decomposition is as follows: 1) generate a random sample from uniform distribution with sample size ; 2) using data on males, run regression at to get m estimates of quantile regression coefficients for each ; 3) generate an *m* size random sample with replacement from the rows of from the female data; 4) construct the counterfactual distribution with the coefficients from step 2 and the sample from step 3, which can be written as . Then the aggregate decomposition at each quantile is written as where the first term on the right-hand side represents the covariates effect that indicates the difference in characteristics/endowments and the second term represents the difference in the coefficients effect (gender bias).

4.4 Recentered Influence Function Method

The Machado and Mata's (2005) approach is similar to that of DiNardo, Fortin & Lemieux (1995) in capturing the contribution of individual covariates. DiNardo et al. (1995) use a reweighting procedure to construct a counterfactual distribution of the dependent variable (wage, Z-score), which is best suited for capturing the individual contributions of the coefficients when the covariates are dummy variables. Firpo et al. (2009) propose a further improvement, which we use here, that is based on the RIF (re-centered influence function). The first step in the RIF is to construct the counterfactual distribution of the nutrition measure (*y*). Then, for the th quantile of the distribution , the RIF is in which is the influence function, and is the density of *y* evaluated at . As the expected value of the influence function is zero, the expectation of the RIF is Combining the RIF model with the counterfactual measure of earnings in the decomposition equation can be written as where is the distributional statistic of interest, and represents the approximate error. The explained part of the traditional Blinder-Oaxaca decomposition is called the composition effect, while the unexplained part is called the structural effect. Parallel to the classical decomposition, the first term in the equation above represents the part of the gap that is due to differences in characteristics (endowments) and the second term represents the part attributed to differences in the coefficients.

5. Results and Discussion

Table 3 shows estimation results for the wage equation (1), both OLS and quantile by gender, with country fixed effects. The first two columns are the results from the OLS regression for males and females, the rest of the table showing the quantile regression results.

Table 3. Results from OLS and Quantile Regression with Country Fixed Effect

Quantiles	OLS		10		20		30		40	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
edu	0.0690***	0.0745***	0.0770***	0.0808***	0.0795***	0.0795***	0.0704***	0.0794***	0.0665***	0.0773***
Edu^2	0.0120***	0.0164***	0.00204	0.00217	0.00665**	0.00963***	0.00968***	0.0105***	0.0119***	0.0158***
age	0.0403***	0.0323***	0.0573***	0.0184***	0.0496***	0.0298***	0.0428***	0.0332***	0.0411***	0.0300***
Age^2	-0.0005***	-0.0004***	-0.0008***	-0.0003***	-0.0007***	-0.0004***	-0.0005***	-0.0004***	-0.0005***	-0.0004***
tenure	0.0058***	0.0065***	0.0077***	0.0114***	0.0105***	0.0105***	0.0105***	0.0114***	0.0086***	0.0107***
training	0.155***	0.190***	0.207***	0.213***	0.160***	0.237***	0.150***	0.207***	0.148***	0.188***
certificate	0.0918***	0.121***	0.0919**	0.0802*	0.0693**	0.0605*	0.0864***	0.0956***	0.0702***	0.0971***
children	0.0207	0.0162	0.0259	-0.0586**	0.0298*	-0.0163	0.0262	-0.0116	0.0188	0.0029
Urban	0.0836	0.110*	0.0674	0.328***	0.148**	0.306***	0.0704***	0.0794***	0.0546	0.122*
_cons	-0.751***	-1.156***	-1.749***	-1.704***	-1.541***	-1.794***	-1.152***	-1.589***	-0.862***	-1.364***
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R2/ Pseudo R2	0.1778	0.2196	0.1882	0.2543	0.1662	0.2207	0.1473	0.1928	0.1319	0.1693

Quantiles	50		60		70		80		90	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
edu	0.0635***	0.0749***	0.0660***	0.0757***	0.0671***	0.0776***	0.0673***	0.0745***	0.0676***	0.0610***
Edu^2	0.0137***	0.0169***	0.0142***	0.0199***	0.0164***	0.0229***	0.0144***	0.0226***	0.0177***	0.0296***
age	0.0421***	0.0318***	0.0392***	0.0339***	0.0352***	0.0311***	0.0204***	0.0252***	0.0212***	0.0273**
Age^2	-0.0005***	-0.0004***	-0.0005***	-0.0004***	-0.0004***	-0.0003***	-0.0002***	-0.0003**	-0.0002**	-0.0003**
tenure	0.0069***	0.0105***	0.0074***	0.0079***	0.0067***	0.0076***	0.0057***	0.0050***	0.0064**	0.0006
training	0.146***	0.187***	0.168***	0.164***	0.149***	0.134***	0.132***	0.132***	0.113**	0.100**
certificate	0.0607**	0.121***	0.0616*	0.120***	0.0328	0.119***	0.101***	0.126***	0.120**	0.131***
children	0.0252*	0.0307	0.0336***	0.0287*	0.0368**	0.0401**	0.0262*	0.0512*	0.0173	0.0892***
Urban	0.0006	0.0758	-0.0093	0.0361	-0.0024	-0.0661	0.0132	-0.0525	0.0501	0.0701
_cons	-0.618***	-1.229***	-0.515***	-1.090***	-0.361***	-0.866***	0.0636	-0.560**	0.309	-0.218
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj R2/ Pseudo R2	0.1202	0.1509	0.1091	0.1360	0.0959	0.1186	0.0825	0.0942	0.0592	0.0629
N									8656	9369

Standard errors in parentheses

$^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

The OLS results show that, generally, education has a positive but slightly different effect on hourly earnings. One more year of education at the mean increases males' earnings by 3% and females' earnings by 4% suggesting higher return to education for females, consistent with previous work (summarized in Psacharopoulosa and Patrinos, 2018). Age, training, and having a certificate all have positive effects on the earnings of both genders. Training increases male and female earnings by 15.5% and 19.0% respectively and certificate increases the earnings by 9.18% and 12.1%. The main negative impact on females' earnings is attributed to the difference in the intercept estimates: -1.156 vs. -0.751 suggesting a pure gender bias. At the mean, the effect of the number of children on earnings is not significant but it is important in the quantile regression to which we turn next.
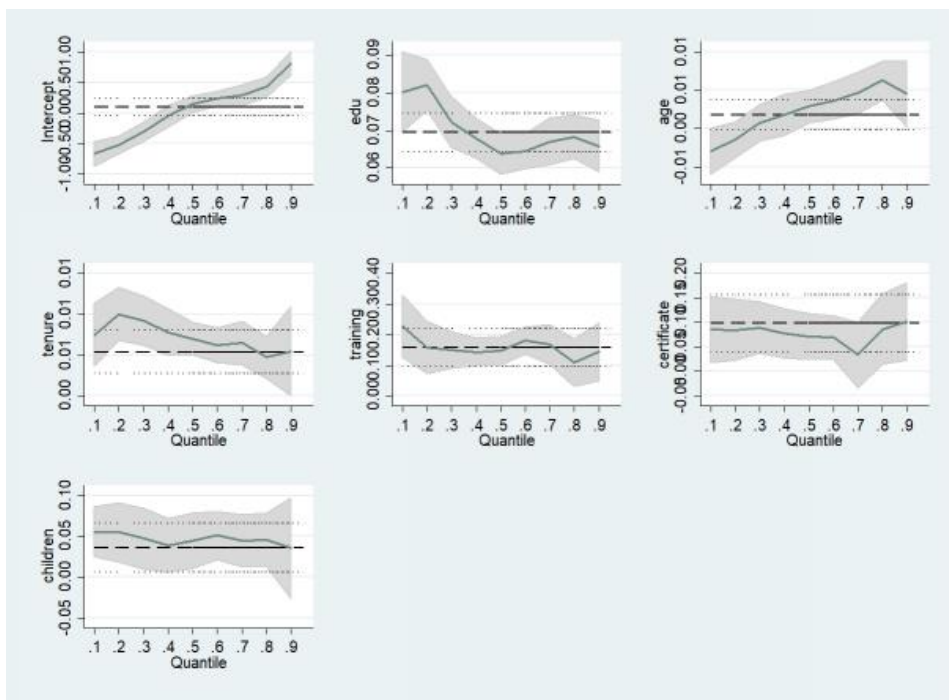
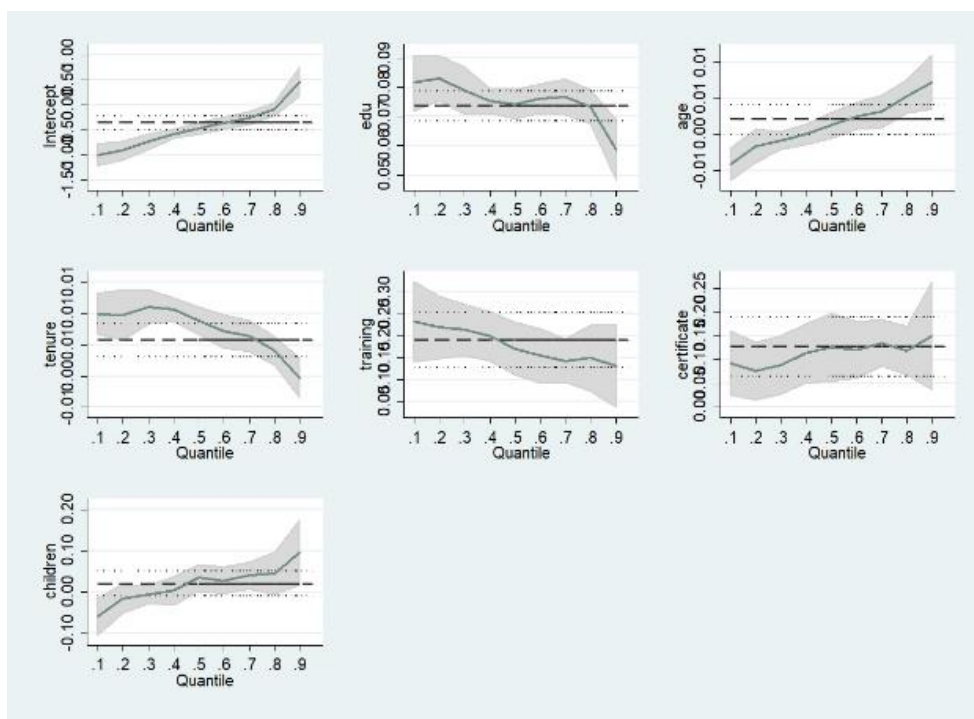Figure 2. Quantile Regression Plot of Men's Log Hourly Earnings

Figure 3. Quantile Regression Plot of Women's Log Hourly Earnings

The results of the quantile regression presented in Figures 2 and 3 and in in Table 3 show that the effect on hourly earnings varies noticeably across the quantiles. Returns to education are higher at lower quantiles for both males and females. The slightly positive impacts of both tenure and training are declining with quantiles for both genders. Having a certificate has positive effect on both male and female earnings from 10 to 40 quantile but, from 50 to 90 quantiles, certificate only has positive effects on females' earnings. The positive impact of age is increasing with quantiles for both genders. The effect of the number of children also varies by quantiles. The number of children positively impacts

males' earnings in quantiles 10 through 80, whereas it has a negative impact in low and positive in high quantiles for females, possibly exemplifying the difference between poverty traps and the incentives to earn more at the low and high ends of the income spectrum. At the 10th quantile, an additional child decreases females' earnings by 5.86%. The positive effect of children in quantiles 50 to 90 is difficult to interpret as it is hard to argue about causality (i.e., whether higher earnings lead to more children or vice versa).

It is interesting to look at the regional differences by running separate regressions. Below we report regional regression results while not showing the corresponding tables due to space limitation. In Latin America, which is represented by Colombia and Bolivia in the sample, the effect of education on earnings is positive and significant. One more year of education increases hourly earnings by 6.73% for males and by 7.23% for females, which is larger than in the overall sample. Training also has a positive effect which is slightly larger for males. In the Sub-Saharan Africa represented in the sample by Ghana and Kenya, the overall effect of education is slightly larger than in Latin America but the difference between coefficients for males and females is only 0.01%. Besides education, working experience and training also have significant effect on hourly earnings Sub-Saharan in Africa. One-month increase in experience increases hourly male earnings by 1.49% and female earnings by 1.22%. The positive effect of training for females, while having a certificate has a positive effect on males' earnings only. The effect of education is larger at middle quantiles. As in the full sample, children are associated with lower wages for women in lower quantiles and higher wages in higher quantiles. In the Eastern & Southern Europe represented by Ukraine, Macedonia, Georgia, and Armenia, the effect of education increases with quantiles and is smaller than in Sub-Saharan Africa but the positive difference between males and females is larger (1.69%). The effects of age, tenure, training, and certificate are also smaller than those in Sub-Saharan Africa. The effect of children on males' wage is positive and significant at most quantiles, but there is no effect on females' earnings. In the South and East Asia represented by China, Vietnam, and Lao PDR the difference of educational effect between males and females is the largest of all, with an additional year in education associated with 6.4% and 8.43% higher hourly earnings for males and females, although both decrease in higher quantiles. Tenure and training are insignificant but certificate and urban employment are significant. The effect of the urban residence dummy is 36.4% for males' and 19.9% for females' earnings. The number of children has a negative effect on females' earnings at quantile 20 but positive at quantile 70 and 90, with no effect on males' earnings.

Table 4. Results from Blinder-Oaxaca Decomposition with Country Fixed Effects

Overall		Differences			
Ln(earnings) usd		explained		unexplained	
Men	1.032***	edu	0.0192***	edu	-0.0450
	(0.0109)		(0.00476)		(0.0450)
Women	0.760***	age	-0.000508	age	-0.00619
	(0.0111)		(0.000371)		(0.0590)
Raw difference	0.272***	tenure	-0.000364	tenure	-0.00527
	(0.0156)		(0.000790)		(0.0153)
explained	0.0316***	training	0.00123	training	-0.00393
	(0.00730)		(0.000858)		(0.00505)
unexplained	0.241***	certificate	0.00139**	certificate	-0.00371
	(0.0141)		(0.000620)		(0.00538)
		children	-0.000796*	children	0.00654
			(0.000427)		(0.00911)
		urban	-0.00209*	urban	-0.0266
			(0.00108)		(0.0848)
				_cons	0.497***
					(0.130)
Country FE	Yes				
N	18104				

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4 shows results from the traditional Blinder-Oaxaca decomposition. The raw difference of male and female log of hourly earnings is 0.272 and significant. Only 0.032 of this difference is attributed to endowment (explanatory variables) differences, whereas 0.24 is due to the differences in coefficients and represents the "bias" that cannot be explained by the available variables. Gender differences in education and having a certificate (both higher for men), the number of children (higher for women), and urban residency account for most of the explained portion of the gender wage gap. The unexplained difference in earnings is entirely attributed to the difference in the intercept estimates making the gender bias truly unexplainable.

Table 5. Results from Machado & Mata Approach with Country Fixed Effects

Quantile	Raw	Characteristics	Coefficients
10	0 .408***	0. 0873***	0. 321***
	(0. 0151)	(0. 0216)	(0. 0209)
20	0. 320***	0. 0486***	0. 271***
	(0. 00978)	(0. 0150)	(0. 0147)
30	0.277***	0. 0295**	0. 248***
	(0. 00917)	(0. 0127)	(0. 0122)
40	0. 255***	0. 0196*	0. 235***
	(0. 00927)	(0. 0116)	(0. 0107)
50	0. 240***	0. 0159*	0. 224***
	(0. 0101)	(0. 0105)	(0. 00957)
60	0. 229***	0. 0178*	0. 212***
	(0. 0105)	(0. 0108)	(0. 00997)
70	0. 217***	0. 0195*	0. 197***
	(0. 0112)	(0. 0111)	(0. 0103)
80	0. 211***	0. 0203*	0. 190***
	(0. 0131)	(0. 0123)	(0. 0116)
90	0. 224***	0. 0338**	0. 190***
	(0. 0190)	(0. 0169)	(0. 0150)

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5 shows the results from the Machado and Mata (2005) approach that provides an aggregate decomposition of explained and unexplained gender differences by quantiles of the log of hourly earnings. The raw differences are all significant and decline steadily with quantiles, ranging from 0.4 to 0.22. The explained differences attributed to differences in endowments (characteristics) are also mostly declining with the quantiles and account for only 10%-20% of the raw difference. The unexplained part of the gender wage gap is all significant at the 1% level and is also declining with the quantiles, representing roughly 80% to 90% of the raw difference. These results are consistent with the Blinder-Oaxaca decomposition but provide more insight by showing that the gender gap in hourly earnings is decreasing with earnings quantiles yet the unexplained part of the difference is slightly increasing. This, combined with the results of the Blinder-Oaxaca decomposition, suggests that, although the gender differences in educational endowments decrease with income, the differences in the intercept increase making the gender bias at higher quantiles hard to explain even in terms of coefficients, the only important factor being the "endowment" of gender.

Table 6. Decomposition from RIF (Recentered Influence Function) Regression

	(1) rif_y10	(2) rif_y20	(3) rif_y30	(4) rif_y40	(5) rif_y50	(6) rif_y60	(7) rif_y70	(8) rif_y80	(9) rif_y90
overall									
Men	0.0278	0.414***	0.665***	0.845***	1.003***	1.192***	1.388***	1.614***	1.954***
Women	-0.0835***	0.344***	0.567***	0.721***	0.891***	1.046***	1.234***	1.492***	1.782***
difference	0.111***	0.0703***	0.0984***	0.124***	0.112***	0.147***	0.154***	0.122***	0.172***
explained	-0.142***	-0.130***	-0.110***	-0.101***	-0.0955***	-0.0838***	-0.0692***	-0.0396***	-0.0122
unexplained	0.253***	0.200***	0.208***	0.225***	0.208***	0.231***	0.224***	0.161***	0.184***
explained									
years_educ_act	-0.0482***	-0.0452***	-0.0424***	-0.0430***	-0.0469***	-0.0515***	-0.0514***	-0.0556***	-0.0514***
age	0.0021	0.0028*	0.0017	0.0008	0.0001	-0.0005	-0.0026**	-0.0053***	-0.0063***
tenure	-0.0054**	-0.0071***	-0.0073***	-0.0063***	-0.0068***	-0.0071***	-0.0059***	-0.0049***	-0.0032*
training	-0.0018	-0.0018*	-0.0022*	-0.0015*	-0.0021*	-0.0022*	-0.0029*	-0.0026*	-0.0029*
pub_emp	-0.0057	-0.0119***	-0.0080***	-0.0093***	-0.0130***	-0.0121***	-0.0119***	-0.0062*	0.0042
certificate	-0.0004	-0.0005	-0.0003	-0.0004	-0.0001	-0.0001	0.0000	-0.0001	-0.0003
children	0.0014	0.0017*	0.0013*	0.0017*	0.0016*	0.0014*	0.0016*	0.0017*	0.0017
Bolivia	0.0001	0.0000	0.0000	0.0000	-0.0001	-0.0001	-0.0001	-0.0003	-0.0003
Colombia	0.0004	0.0007	0.0006	0.0004	0.0003	0.0003	0.0004	0.0007	0.0009
Georgia	0.0067	0.0003	-0.0027	-0.0004	-0.0001	-0.0017	-0.0033	-0.0088***	-0.0117***
Ghana	-0.0494***	-0.0300***	-0.0198***	-0.0140***	-0.0081***	-0.0026	0.0010	0.0097***	0.0116***
Kenya	-0.0484***	-0.0374***	-0.0252***	-0.0197***	-0.0094***	-0.0024	0.0071*	0.0216***	0.0236***
Laos	-0.0012	-0.0003	-0.0017	-0.0042**	-0.0045**	-0.0035**	-0.0020*	0.0024*	0.0042**
Macedonia	0.0005	0.0012	0.0015	0.0013	0.0017	0.0021	0.0026	0.0028	0.0013
Sri Lanka	0.00254	0.00304	0.00152	-0.00147	-0.00117	0.00145	0.00466	0.0193***	0.0243***
Ukraine	0.0033	-0.0040	-0.0041	-0.0028	-0.0048	-0.0022	-0.0031	-0.0069	-0.0008
Vietnam	0.0019	-0.0011	-0.0028*	-0.0016	-0.0018	-0.0028*	-0.0029	-0.0071***	-0.0072**

Yunnan (China)	-0.0001	-0.0002	-0.0002	-0.0002	-0.0002	-0.0000	-0.0000	0.0001	0.0001
unexplained									
years_educ_act	0.190*	-0.0196	-0.198***	-0.218***	-0.212***	-0.247***	-0.279***	-0.243***	-0.222**
age	-0.231**	-0.146*	-0.108	-0.0726	-0.0521	-0.0793	-0.0758	-0.0548	-0.154
tenure	0.0716**	0.0523**	0.0279	-0.0023	-0.0073	-0.0049	-0.0065	0.0021	0.0187
training	0.0033	-0.0034	0.0009	-0.0079	-0.0069	-0.0069	0.0008	-0.0030	-0.0083
pub_emp	-0.0141	-0.0104	-0.0072	0.0003	0.0182	-0.0050	-0.0172	-0.0520**	-0.0401
certificate	0.0170	0.0191*	0.0133	0.0137*	-0.0067	-0.0150*	-0.0241***	-0.0242**	-0.0189
children	0.0342**	0.0314***	0.0202**	0.0216**	0.0185*	0.0156	0.0142	0.0098	-0.0073
Bolivia	-0.0151	-0.0070	-0.0175**	-0.0208***	-0.0187***	-0.0269***	-0.0305***	-0.0170*	-0.0042
Colombia	0.0130	0.0135	-0.0102	-0.0294***	-0.0306***	-0.0312***	-0.0353***	-0.0209**	-0.00825
Georgia	-0.0020	0.0120	0.0035	-0.0092	-0.0103	-0.0137*	-0.0119	0.0078	0.0258**
Ghana	0.0155**	0.0074	-0.0020	-0.0121***	-0.0139***	-0.0206***	-0.0246***	-0.0195***	-0.0125*
Kenya	-0.0051	-0.0037	-0.0230***	-0.0354***	-0.0349***	-0.0437***	-0.0477***	-0.0344***	-0.0275**
Laos	-0.0077	-0.0106*	-0.0230***	-0.0202***	-0.0245***	-0.0219***	-0.0217***	-0.00814	-0.0052
Macedonia	-0.0023	-0.0055	-0.0262**	-0.0439***	-0.0429***	-0.0445***	-0.0447***	-0.0452***	-0.0497***
Sri Lanka	0.0122	0.0200***	0.0017	-0.0137***	-0.0214***	-0.0289***	-0.0298***	-0.0088	0.0046
Ukraine	-0.0195	-0.0141	-0.0370***	-0.0252**	-0.0089	-0.0167	0.0010	0.0165	0.0250
Vietnam	-0.0182	-0.0164	-0.0284***	-0.0366***	-0.0450***	-0.0484***	-0.0570***	-0.0260*	-0.0205
Yunnan (China)	-0.0164	-0.0402***	-0.0529***	-0.0437***	-0.0387***	-0.0453***	-0.0398***	-0.0177*	0.0015
_cons	0.227	0.321**	0.674***	0.780***	0.746***	0.915***	0.953***	0.700***	0.687***
N	12405	12405	12405	12405	12405	12405	12405	12405	12405

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6 shows decomposition results from the RIF (Recentered Influence Function) Regression offering even deeper insights by providing estimates of explained and unexplained differences by covariates and by quantiles. The general picture differs slightly from the findings reported above since this is a different approach. The raw differences are about a half of the previous estimations, likely due to a decrease in the negative education coefficient difference and a slight decrease in the negative endowment differences at the higher quantiles, and are larger for the upper quantiles.

Women are found to have higher observable educational attainments, tenure and training than men as well as observable work in the public sector and the differences are relatively stable across quantiles. However, the number of children under 6 is associated with higher earnings by men mostly in the middle of the earnings distribution. In terms of the coefficients, educational differences between men and women are negative and sizeable in all but the lowest quantile. The impacts of tenure, training, and working in the public sector endowments are marginally significant and negative for all. The number of children is positive for only up to the 50th percentile and does not seem to affect women at higher level of wages. While the unexplained part of the gender pay gap is still mostly due to the intercept difference, a part of it appears to be offset by higher returns to education for women, particularly in the middle and high earnings levels, and the number of children that is still punishing.

6. Conclusion

In this paper, we investigate the determinants of the gender wage gap across 12 developing countries using data from the STEP (Skills Toward Employment and Productivity) Measurement survey that provides information on household demographic characteristics, education and training, employment, job skill requirements, and family characteristics. We apply the classical Blinder-Oaxaca decomposition and combine it with quantile regression to find the distributional effects. We also use decomposition methods of Machado and Mata (2005) and Firpo et al. (2009) based on the recentered influence function to obtain distributional decomposition of the wage gap by covariates and by quantiles.

The results indicate the existence of significant gender wage gaps at all wage levels. Both Blinder-Oaxaca and Machado and Mata Decompositions show that roughly 80% to 90% of the wage difference across all wage quantiles is unexplained by the data (i.e., is due to differences in the coefficients) with the largest difference being in the intercepts. The explained part of the wage gap is mostly due to difference in education and training and motherhood penalty. Quantile analysis shows that education and training have the largest positive impacts on the wages of both genders, with larger coefficients for females indicating greater returns to skill acquisition. Interestingly, the number of children has a negative effect on female earnings in the lower wage quantiles suggesting a motherhood penalty in the low income categories and a positive effect in the higher quantiles of both male and female male earnings.

In terms of policy relevance, our results suggest that improving education is perhaps the most efficient way to reduce the observed gender wage gap as women seem to have higher rates of return to education relative to men. Considering the motherhood penalty in the low wage quantiles, efforts to offer motherhood friendly jobs and to increase government support to low income families with more children seem to be relevant policy measures. Judging by the signs and

magnitudes of our coefficients, encouraging mothers at low wage levels to take part in the training programs and/or get training certificates appears to be an efficient way to increase their earnings and reduce the gender wage gap. None of these measures, however, would eliminate the underlying unexplained causes for labor market inequality/ gender wage disparity that are revealed by the differences in the intercept estimates. This calls for more research experimenting with other, non-traditional, explanatory variables.

References

Agüero, J. M., Mindy, M., & Neha, R. (2020). *Economic Development and the Motherhood Wage Penalty.* Working papers 2020-06, University of Connecticut, Department of Economics.

Albrecht, J., Björklund, A., & Vroman, S. (2003). Is there a glass ceiling in Sweden? *Journal of Labor Economics*, *21*(1), 145-177. https://doi.org/10.1086/344126

Albrecht, J., Van Vuuren, A., & Vroman, S. (2009). Counterfactual distributions with sample selection adjustments: Econometric theory and an application to the Netherlands. *Labour Economics*, *16*(4), 383-396. https://doi.org/10.1016/j.labeco.2009.01.002

Altonji, J. G., & Rebecca, M. B. (1999). "Race and Gender in The Labor Market" In Handbook Of Labor Economics, Volume 3, Edited By O. Ashenfeher And D. Card. https://doi.org/10.1016/S1573-4463(99)30039-0

Amaya, E., & Mougenot, B. (2019). The gender differences in highly paid wage: a case study of Peruvian physicians. *Cadernos de saude publica*, *35*, e00043018. https://doi.org/10.1590/0102-311x00043018

Anderson, D. J., Binder, M., & Krause, K. (2002). The motherhood wage penalty: Which mothers pay it and why?. *American economic review*, *92*(2), 354-358. https://doi.org/10.1257/000282802320191606

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature, 55*(3), 789-865. https://doi.org/10.1257/jel.20160995

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, 436-455. https://doi.org/10.2307/144855

Budig, M. J., & England, P. (2001). The wage penalty for motherhood. *American sociological review*, 204-225. https://doi.org/10.2307/2657415

DiNardo, J., Fortin, N. M., & Lemieux, T. (1995). *Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach* (No. w5093). National bureau of economic research. https://doi.org/10.3386/w5093

England, P., Bearak, J., Budig, M. J., & Hodges, M. J. (2016). Do Highly Paid, Highly Skilled Women Experience the Largest Motherhood Penalty? *American Sociological Review*, *81*(6), 1161-1189. https://doi.org/10.1177/0003122416673598

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, *77*(3), 953-973. https://doi.org/10.3982/ECTA6822

Gardeazabal, J., & Ugidos, A. (2005). Gender wage discrimination at quantiles. *Journal of population economics*, *18*(1), 165-179. https://doi.org/10.1007/s00148-003-0172-z

Glauber, R. (2018). Trends in the motherhood wage penalty and fatherhood wage premium for low, middle, and high earners. *Demography*, *55*(5), 1663-1680. https://doi.org/10.1007/s13524-018-0712-5

González, X., & Miles, D. (2001). Wage inequality in a developing country: decrease in minimum wage or increase in education returns. *Empirical Economics*, *26*(1), 135-148. https://doi.org/10.1007/s001810000056

Kleven, H., Camille, L., & Jakob, E. S. (2019). Children and Gender Inequality: Evidence from Denmark. *American Economic Journal: Applied Economics, 11*(4), 181-209. https://doi.org/10.1257/app.20180010

Koenker, R., & Bassett, J. G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33-50. https://doi.org/10.2307/1913643

Lundborg, P., Erik, P., & Astrid, W. R. (2017). Can Women Have Children and a Career? IV Evidence from IVF Treatments. *American Economic Review, 107*(6), 1611-37. https://doi.org/10.1257/aer.20141467

Machado, J. A., & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, *20*(4), 445-465. https://doi.org/10.1002/jae.788

Melly, B. (2005). Decomposition of differences in distribution using quantile regression. *Labour economics*, *12*(4), 577-590. https://doi.org/10.1016/j.labeco.2005.05.006

Mincer, J. (1974). *Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2.*

Montes-Rojas, G., Siga, L., & Mainali, R. (2017). Mean and quantile regression Oaxaca-Blinder decompositions with an application to caste discrimination. *The Journal of Economic Inequality*, *15*(3), 245-255. https://doi.org/10.1007/s10888-017-9355-9

Nguyen, B. T., Albrecht, J. W., Vroman, S. B., & Westbrook, M. D. (2007). A quantile regression decomposition of urban–rural inequality in Vietnam. *Journal of Development Economics*, *83*(2), 466-490. https://doi.org/10.1016/j.jdeveco.2006.04.006

Ñopo, H. (2008). Matching as a tool to decompose wage gaps. *The review of economics and statistics*, *90*(2), 290-299. https://doi.org/10.1162/rest.90.2.290

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, 693-709. https://doi.org/10.2307/2525981

Psacharopoulosa, H., & Anthony, P. (2018) Returns to investment in education: a decennial review of the global literature. *EDUCATION ECONOMICS, 26*(5), 445-458. https://doi.org/10.1080/09645292.2018.1484426

Wood, A. (1997). Openness and wage inequality in developing countries: the Latin American challenge to East Asian conventional wisdom. *The World Bank Economic Review*, *11*(1), 33-57. https://doi.org/10.1093/wber/11.1.33