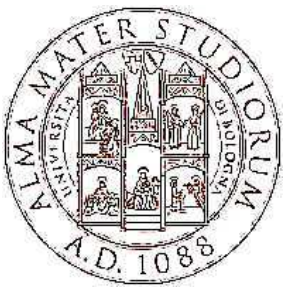


Alma Mater Studiorum
Università degli Studi di Bologna

Dipartimento di Scienze Economiche

Dottorato di ricerca in Economia

XX ciclo

Settore scientifico disciplinare: SECS-P/01 Politica Economica

The Gender Wage Gap in Developing and Transition Countries

Simona Valmori

Relatore: Prof. Hartmut Lehmann

Coordinatore: Prof. Andrea Ichino

Esame finale anno 2008

Acknowledgments

Al termine di questi tre anni di dottorato, ringrazio in modo particolare il Prof. Hartmut Lehmann che mi ha dato la possibilità di cominciare questo percorso con la sua supervisione e mi ha aiutato a terminarlo nel modo migliore. Ringrazio in modo speciale i miei colleghi Norberto Pignatti, Randolph Bruno e Anzelika Zaiceva per il loro invalutabile supporto a 360 gradi e per la loro disponibilità illimitata e per tutto il tempo che mi hanno dedicato nell'ultimo anno.

Ringrazio il Prof. Lachaud e il governo Comoriano, soprattutto l'unità statistica, per avermi dato accesso al prezioso dataset che ho utilizzato per scrivere il capitolo sulle Comore.

Ringrazio i miei colleghi dell'aula dottorandi, in modo particolare Davide Dragone, che mi ha seguito, consigliato e supportato nel periodo di creazione e scrittura della tesi.

Ringrazio le mie amiche "africane" Chiara, Shireen e Nele che condividono con me una grande passione per l'Africa.

Ringrazio tutte le persone che ho conosciuto e mi hanno seguito nell'ultimo anno, in particolare Adriana, per avermi consentito di arrivare e vivere questo momento in modo sereno e consapevole.

Ringrazio Lorenzo per il supporto tecnico e l'assistenza che mi ha sempre dato nei momenti di difficoltà.

Ringrazio i miei genitori ed Elisa che mi sono stati sempre vicini, mi hanno sempre sostenuto ed incoraggiato. Ora posso condividere con loro la gioia di questo momento.

Contents

Overview

The existence of a differential payment for labor market services between men and women is taken as a universal phenomenon in almost all countries regardless of the nature and structure of the economic system. The first prominent empirical studies on this issue were published for the U.S. on wage discrimination between blacks and whites and males and females by Oaxaca (1973) and Blinder (1973). Since then, a large number of studies on the estimation of wage discrimination or unjustified wage differentials have been produced for many countries, mostly Western industrialised countries. The consensus is that the most important factors in explaining the fall of the gap from 40% to 20% were improvements in women's human capital, a change in women working sectors and a sharp decline of the "unexplained" part of gender wage gap.

While a consolidated knowledge exists on this issue with reference to western countries, the interest on the gender wage gap in transition countries has developed since the beginning of reforms from planned to market economies and in developing countries is still very marginal. According to Weichselbaumer and Winter-Ebmer (2005) only 3% of all existing studies on gender wage gap since the 90s are on Africa and the ones already existing mainly suggest that there are significant gender wage gaps in some African countries and none or very low in some others. In transition countries the main findings refer to the change in earning differentials over the transition from centrally based economies to free market economies. After the spectacular reduction in the gap in the early nineties, there is not a consensus on the current trend of the gap - especially on the situation of women in these countries - and on the mechanisms underlying the evolution. Moreover, in some countries the reduction of the gap was partly explained by a withdrawal of the less skilled women from the labour market.

The aim of my dissertation is to study the gender wage gap with a specific focus on developing and transition countries. These are the contexts where the contribution to the knowledge base about this phenomenon would be more relevant. Studying this issue in these countries could give very important results in terms of policy advice and recommendations.

In developing countries this kind of differential treatment has being recognized as having a detrimental impact on economic growth, poverty and other macro-economic outcomes. Most empirical studies support the hypothesis that narrowing gender inequality stimulates per capita growth and overall living standards by increasing average human capital and lowering population growth. At the micro-level, labour markets in Africa are central to the livelihoods of people both in

and outside the labour force, because there are not formal social insurance mechanisms. Women tend to be more disadvantaged, because they are more limited on access to productive assets such as land and consequently they are limited in their options for engaging in sustainable livelihood activities.

In transition countries, the peculiar changes in labor market institutions and competitive environment have been reshaping the working lives of women. During the transition process, the labour market has changed from an exclusively formal and public institution to one that is extremely polarized, with pre-transition formal employment in the public sector coexisting with an often highly informal emerging unregulated private market. At the same time, the burden of parental duties, especially child care, has shifted increasingly away from the state and into the household. New demands upon women have risen involving a re-assessment of their lifetime decisions, such as how much to invest in education and when to start a family.

The fast changing nature of the labour market in transition economies and the idiosyncratic and heterogeneous situation of labour markets in Africa make the gender wage gap a very interesting research topic. If we take into account its transversal linkages to issues such as poverty reduction, labour market segmentation, labour market institutions, the interest on the issue becomes even more relevant.

In the first chapter of my dissertation I present the main existing theories proposed to analyse the gender wage gap. The human capital theory is the reference model for all decomposition techniques, while other theories give difference explanations to the observed gender discrimination in the labour market. I provide a review of the literature and the open issues faced by the literature in estimating the gender wage gap in developing and transition countries. As regards the methodology, labour economists commonly define wage discrimination by comparing wages for equally productive workers. The raw wage gap is then decomposed into a portion explained by differences in human capital endowments and a residual or unexplained part which is the difference in human capital prices. Most of the discussion in the literature over the techniques of wage decomposition starts from the specification issues on which the decomposition is based. This is the most challenging point from an applied econometrician's perspective and depends on the application of appropriate methods of estimation. Moreover, further important questions arise in measuring the gender wage gap in developing and transition countries because of the specificities of their labour markets and of their socio-economic situation.

The second chapter is an empirical analysis of the gender wage gap in a developing countries, the Union of Comoros, using data from the multidimensional household budget survey

"Enquete integrale auprès des ménages" (EIM) run in 2004. The interest of my work is to provide a benchmark analysis for further studies on the situation of women in the Comorian labour market and to contribute to the literature on gender wage gap in Africa by making available more information on the dynamics and mechanism of the gender wage gap, given the limited interest on the topic in this area of the world. Enhancing the gender gap literature on developing countries, especially on the poorest ones, is crucial for several reasons. First, as already mentioned, there are manifest shortcomings of studies on African countries, particularly due to the shortage of available information. Second, gender inequality effects are stronger where markets do not function efficiently and the States lack the resources for introducing corrective policies. Third, understanding the roots of inequalities between the sexes and reducing gender gaps have a central place in term of policies in these countries. This lack of information is even more serious if we consider that in many African countries finding a job and having a decent wage is the major escaping route from poverty for women (Sender 2002), especially women who are head of household.

The third chapter is an applied analysis of the gender wage gap in a transition country, Poland, using data from the Labour Force Survey (LSF) collected for the years 1994 and 2004. I provide a detailed examination of how gender earning differentials have changed over the period starting from 1994 to a more advanced transition phase in 2004, when market elements have become much more important in the functioning of the Polish economy than in the earlier phase. The interest is to observe the effects of reforms over a large period and to investigate if previous findings on the evolution of the gender wage gap and on labour market dynamics are confirmed.

The main contribution of my dissertation is the application of the econometrical methodology that I describe in the beginning of the second chapter. First, I run a preliminary OLS and quantile regression analysis to estimate and describe the raw and conditional wage gaps along the distribution. Second, I estimate quantile regressions separately for males and females, in order to allow for different rewards to characteristics. Third, I proceed to decompose the raw wage gap estimated at the mean through the Oaxaca-Blinder (1973) procedure. In the second chapter I run a two-steps Heckman procedure by estimating a model of participation in the labour market which shows a significant selection bias for females. Forth, I apply the Machado-Mata (2005) techniques to extend the decomposition analysis at all points of the distribution. The general idea is to generate the female wage distribution that would emerge if women were given men's labour market characteristics but remained to receive returns to those characteristics like women. The gap between identical men and women in terms of their characteristics could then be attributed to unequal gender treatment. In the case of Poland, I can realise a deeper and wider analysis because I dispose of data

for two years. Therefore, I can study the evolution of the wage gap over the period in exam and I can decompose the change in the wage gap in each year and also across the two years.

The preliminary analysis of the wage gap gives very different results for the two countries. The estimates for Comoros show a very high raw wage gap decreasing along the wage distribution. When controlling for human capital and other characteristics, it becomes higher in the top deciles. In Poland the raw gap is quite low in both years with respect to international standards and when I control for characteristics the gap is larger and increasing all along the distribution. In 2004 the difference between the raw and the adjusted gap becomes more important, which means that women's great improvement in characteristics over the decade have not translated into higher wages. Therefore, women have better characteristics than men at the top of the distribution in Comoros and at any point of the distribution in Poland. In Comoros, difference in characteristics between genders can partly explain the wage gap only in the first half of the distribution. Moreover, in Comoros the wage gap is higher at the bottom of the wage distribution (glass ceiling effect) while in Poland at the top of the distribution (ceiling floor). This means that while in Comoros low skilled women are the more discriminated, in Poland discrimination affects more women at the top of the distribution.

Quantile estimations run separately for men and women show important differences in both countries. Among the human capital controls, education is the most important variable determining wages. In Comoros returns to education are very similar for both genders but significant for women only in the second half of the distribution, while in Poland they are much higher for women. This reflects the different level of development of the two countries: while in developing countries such as Comoros gender discrimination begins from the access to basic assets such as education and training, in transition countries women possess important advantages in productive characteristics, mainly education. However, from the descriptive statistics of the sample and from official Polish statistics emerges that a significant share of women with low education dropped from the labour market since the beginning of the transition. This evidence is consistent with the observed rising female wages, therefore it should be taken seriously into account when studying policies aimed at the reduction of the wage gap. Tenure and age show low returns in both countries and their level of significance is quite variable along the distribution in Comoros. In Poland, it is evident that the experience acquired after the beginning of the transition tends to be more remunerated in 2004, as expected.

The Oaxaca-Blinder decomposition of the wage gap at the mean give similar results in both countries. In Comoros most of the gap is unexplained, while in Poland none of the gap is explained

by differences in endowment between male and female workers and if women were remunerated according to their characteristics, their wage would be higher than men's wage. That means, the gap may depend from other unobservable factors or it can be due to pure discrimination. It is important to notice that having a cross section dataset I can't account for unobserved heterogeneity in personal characteristics that may stem from omitted common variables or global shocks that affect each individual unit differently. An important concern for Comoros is the selection bias since the proportion of women in the active labour force is very low also with respect to other African countries. The two-steps Heckman procedure reveals a positive selection effect because women who actually work in Comoros are those with the highest potential wages. This means that if all women worked the wage gap would be much higher.

The counterfactual analysis sheds more light on the preliminary analysis at the quantile and reinforces the role of gender discrimination as a main explanation of the wage gap in both countries. In Comoros, if I apply male returns to female characteristics the counterfactual gap is much lower at any point of the distribution and women would be paid more than men at the top. The important reduction of the gap at the bottom confirms the glass floor effect affecting women at the bottom of the distribution. If I apply female returns to male characteristics, the counterfactual gap is very similar to the observed gap at any point of the distribution, meaning that the difference in the distribution of characteristics between genders is not so relevant in explaining the wage gap.. Overall, difference in the pay structure is far more important then the differences in their characteristics along the all distribution

In Poland, in each of the two years, if women had been paid at men's returns the gap would have been much lower than the observed gap or even positive, meaning that the difference between male and female earnings would have been lower or women would have been paid more than men. Overall, the differences between men's and women's pay structure are far more important than the differences in their characteristics in explaining each of the gaps. Moreover, in both years the gap would have fallen more at the top than at the bottom if women had men's betas. This confirms that discrimination was higher at the top than at the bottom.

In Poland I can also implement the Juhn, Murphy and Pierce (1991) decomposition over the period 1994-2004, to account for effects to the pay gap due to changes in overall wage dispersion beyond Oaxaca's standard decomposition. This is particularly important in the context of transition countries since I expect the rise in overall wage inequalities to be a consequence of the transition from the centrally planned to the market economy. The analysis shows that the observed component is the most important factor involving the narrowing of the wage gap over the period. Higher levels of education attained by women together with an increase in the returns to education over the period

in exam would cause an important reduction in the wage gap. This improvement was partly offset by the growth of the unexplained part which has penalized women relative to men.These results show that in the beginning of the new century factors working against a narrowing of the gap are offsetting the improvements made in the previous phase, which explains why in 2004 the gap shows a very important increase once I control for differences in characteristics.

Chapter 1

A survey of the literature and the open issues on the gender wage gap in developing and transition countries

Introduction

According to mainstream economists, labour market discrimination exists whenever workers with identical productive characteristics receive different rewards for their attributes because of the population group to which they belong. This is called "post-entry" discrimination as it occurs after a worker enters the labour market and neglects discrimination that occurred previously (i.e. different probabilities of accessing enhancing assets such as education and training). This kind of discrimination may take three different forms: wage, employment and occupation or job discrimination.

Gender wage discrimination occurs when women receive systematically a lower wage than the men regardless the fact that they have equal productivity related characteristics. Thus, gender pay differentials are systematically greater than differences in productive endowments. This is a particularly important issue as the unequal treatment of men and women in the labour market is an important potential source of economic inefficiency as workers with equal productivity end up receiving different rewards. The potential resulting costs to the economy increase when the workers affected are those with the highest productivity potential (Tzannatos 1994). In addition, any existing gap in the price of time for equally productive men and women has enormous consequences for their respective incomes, the domestic division of labour, the incentives for girls to be educated and their parents' incentives to invest in their health and fertility choices. It also deepens women's vulnerability to poverty and their financial dependence on the family. Therefore, this chain represents the main pathway through which gender inequality affects economic growth and impedes poverty reduction.

Especially in developing countries, this kind of discrimination has being recognized as having a detrimental impact on economic growth, poverty and other macro-economic outcomes. Most empirical studies support the hypothesis that narrowing gender inequality stimulates per capita growth and overall living standards by increasing average human capital and lowering population growth (Forsythe et alii 2000; Morrison and Jutting 2005). At the micro-level, labour markets in Africa are central to the livelihoods of people, both in and outside the labour force, because there are not formal social insurance mechanisms. Most poor people are in fact dependent on employed people through informal social networks for their livelihood. Women tend to be more disadvantaged, because they are more limited on access to productive assets such as land and consequently they are limited in their options for engaging in sustainable livelihood activities. In addition to the burden of the household, they are subject to different forms of discrimination.

In transition countries, the peculiar changes in labor market institutions and competitive environment have been reshaping the working lives of women. In the pre-transition era, equal labor market participation did not lead to gender equality, contrary to the perception of socialist ideology. Women were performing long hours of unpaid work at home in addition to their paid contribution in the labour market and this dual role had importantly affected women's labour market opportunities. In addition many studies have shown that there was occupational segregation under the socialism. During the transition process, the labour market has changed from an exclusively formal and public institution to one that is extremely polarized, with pre-transition formal employment in the public sector coexisting with an often highly informal emerging unregulated private market. At the same time, the burden of parental duties, especially child care, has shifted increasingly away from the state and into the household. The relative value of time spent in productive and nurturing activities have been substantially altered. New demands upon women have risen involving a re-assessment of their lifetime decisions, such as how much invest in education and when to start a family. All these changes influenced the new structure of the transitional labour market.

In the first section of the survey I present the main existing theories proposed to analyse the gender wage gap. The human capital theory is the reference model for all empirical decomposition techniques, while other theories give different explanations to the observed gender discrimination in the labour market. The second section is a review of the empirical literature on the gender wage gap in developing and transition countries. The third section proposes an overview of the open issues in the estimation of the gender wage gap. I discuss empirical questions related to the estimation of the wage equations and its implications on decomposition techniques. In the forth section I focus on the issues specific to developing and transition countries and how the literature have taken them into account when measuring the gender wage gap.

1. The theories on the gender wage gap

The conventional explanations proposed to analyse the gender wage gap refer to different theories. While the human capital theory is the model used to identify the "explained" part of the wage gap providing the link between investment in human capital and earnings, the subsequent theories focus on different dimensions of discrimination in order to justify the existence of the "unexplained" part of the gap.

The human capital theory (Becker 1967) links expected lifetime labor force participation to one's incentive to acquire marketable training. The acquired training in school and on the job will in turn determine earning potential. Therefore, incentives to accumulate human capital vary along the life cycle and they are directly proportional to the time one expects to work over the lifetime. Since on average women work fewer hours throughout their lives, one expects women to purchase less human capital investment than men and this translates into lower per hour relative women's wages. This idea that women may face different incentives to accumulate human capital than men can be referred to the influential work of Mincer and Polachek (1974), where the authors provide evidence that married women tend to interrupt their labour market attachment with periods of non-participation. Using a regression framework, they find that the expected career interruptions do have an impact on the human capital investment of young women. Polachek (2004) uses this model to explain the reduction in the gender wage gap since the '80s. According to his work, in the last twenty years lifetime work expectations have become more similar between men and women. Men's lifetime labor force participation has been diminishing while women's is rising, therefore males invest less in human capital while females invest more.

Concerning gender based discrimination, a first stream of the literature argues this phenomenon is a demand-driven process. According to traditional employers' discrimination theory, men may receive preferential treatment because i) employers prefer men to women, ii) male workers dislike working alongside women, especially in a hierarchically inferior position, or iii) customers have a disutility by being served by women. This is the "taste-based" theory of discrimination proposed by Becker (1957). If we assume the validity of this theory, we have to suspend the employers maximizing profits hypothesis in favour of a more general form of utility maximization. The utility of the employer is increasing in profit and decreasing in the proportion of women among the labour force. Women are employed only if the attendant disutility can be compensated by increasing profit. That is, if women are paid less than men. Both scenarios are inefficient in the sense they are based upon a misallocation of resources and they cannot survive in the long run unless the employer has some monopoly power or the taste for discrimination is common to all employers. In the case of employee discrimination, the employment of women would cause the marginal value product of other employees to decline, while with consumer discrimination it is the consumer's preference for men which reduces the marginal value product.

The theory of statistical discrimination provides a different explanation of why rational employers with no taste for discrimination might finally discriminate in favour of men (Arrow

1972, Phelps 1972, Aigner and Cain 1977). This theory suggests that in a world of imperfect information on an individual worker's productivity it may be too costly for employers to obtain sufficient information, in order to efficiently estimate the productivity for each perspective worker. A profit maximization employer would thus have a preference for hiring men if he expects they are on average more productive or he will require higher credentials for female job applicants (Coate and Loury, 1993). Therefore, the extent of statistical discrimination depends on the difference between male and female net average productivity, turnover and other costs and finally on the difference in the signals provided by the two groups (i.e. interviews and aptitude scores).

A further stream of the literature explores the dimension of job segregation (Edgeworth 1922, Bergman 1971, Zellner 1972). If women tend to be segregated into a relatively small number of occupations and/or firms, the excess of supply of labour in these occupations leads to wages falling while it gives the employer some degree of monopsony power. The "crowding hypothesis" argues that if workers are equally productive, the lower wage of women is a result of their excess of supply in some occupations. In this framework, gender is used as a signal to divide workers into two groups and allocate them to the crowded and uncrowded sector. Men and women are "not competing" groups of workers and women are prevented to access the uncrowned male sector by discrimination or other existing barriers. Theories of occupational segregation of the labor market in the human capital model argue that men and women move into different fields, a process that can be explained by self selection model (Polachek 1981). Women who plan to withdraw from the labour market when they have children, may choose occupations with low skill depreciation and low penalties from interruptions (Polachek 1976). Therefore, men benefit from less competition and higher wages in the primary sector while women concentrate in relatively few 'female' occupations, mostly in the secondary segment of the labour market.

These theories are also consistent with the idea of dual labour market theories (Chiswick 1973, Stiglitz 1973), which assume the labour market as segmented into a primary or formal sector characterized by 'good' jobs in terms of pay, security, working conditions, and a secondary or informal sector offering poor working conditions, low pay and no protections or guarantees (Doeringer and Piore 1971). The most recent literature distinguishes between two types of occupational segregation and it also applies to the different developments of formal and informal labour markets. Horizontal segregation refers to the distribution of men and women across occupations and sectors, for example in the communist countries women were concentrated in the service sectors while men were equally divided between industry and services (World Bank 2001).

Vertical segregation is explained by the hierarchy or status of women within each occupation or sector, for example if men are more likely to be in supervision/managerial positions (Hakim 1992, Blackburn et al. 2001, Charles and Grusky 1995). Within the formal sector, women tend to be underrepresented among private sector workers and are more likely to be employed in the public sector. However, structural adjustment programs have importantly reduced the size of the public sector reducing employment opportunities for women and causing their progressive marginalization (Standing 1999, Chen et alii 2005). Within the informal sector, women tend to be disproportionally represented in low-pay, low productive positions in low skilled occupations, contrary to men (Das 2003, Amuedo-Dorantes 2004, Chen and al. 2005).

2. Empirical studies on the gender wage gap in developing and transition countries

The first prominent empirical studies on this issue were published for the U.S. on wage discrimination between blacks and whites and males and females by Oaxaca (1973) and Blinder (1973). Since then, a large number of empirical studies on the estimation of wage discrimination or unjustified wage differentials have been produced for many countries, mostly Western industrialised countries. Blau and Khan (1992, 1996) show that in Western countries a substantial gender wage gap exists, which has been decreasing since the late 1970s. Over the period, the gap has dropped from 40% to 20%. According to their studies, the most important factors in explaining this change were improvements in women's human capital, a decline in gender differences in experience and a change in women's working sectors. Moreover, a major reason for the convergence in men's and women's wages was the sharp decline of the "unexplained" gender wage gap. That is, women improved their unmeasured skills relative to men or discrimination against them decreased.

With respect to developing countries, a substantial literature now exists on countries in Latin America and East Asia. Among the several empirical works, Psacharopoulos and Tzannatos (1992) find that on average discrimination accounts for about 88% of the male advantage in pay in 15 different Latin American countries, after controlling for selectivity bias. Claudio Montenegro (2001) analyses the gender wage gap in Chile with a quantile regression approach, and he finds that the unexplained gap steadily increases from 10% to 40% as one moves from the bottom to the top of the distribution. Moreover, returns to education and experience are significantly different at different quantile of the distribution. Horton's (1996) book on a seven-country study of women in

East Asian labour market reported that in India women get 55.8% of men as urban employees and 50.6% as rural employees.

By contrast, the literature on gender wage gap in Africa is small. Weichselbaumer and Winter-Ebmer (2005) estimate in their meta-analysis study that only 3% of all existing studies on gender wage gap since the 90s are on Africa, and the ones already existing mainly suggest that there are significant gender wage gaps in some African countries and none or very low ones in some others. Glick and Sahn (1997) analyze gender differences in earnings in Guinea Conakry. They find that differences in characteristics account for 45% of the male-female gap in earnings from self-employment and 25% of the differences in earnings from public sector employment, while in the private sector women actually earn more than men do. Agesa (1999) used urban workers data for Kenya and found that the relative pay for women as a percentage of men's wage is 63% and out of this total wage gap, about 60% is unexplained.

Appleton et alii (1999) investigate the gender wage gap in Ethiopia, Uganda and Cote d'Ivoire by pointing to the importance of sectoral disaggregation when applying decomposition techniques. While for industrialized countries a distinction is often made between unionized and non-unionized sectors, for Africa a more important distinction is between the public and private sectors. While government pay policies may be insulated from competitive pressure, private-sector earnings may be less affected by the political economy influence determining public sector pay. They find a substantial gap in Ethiopia and Uganda (33% and 24% respectively), mainly attributable to differences in returns to wage generating characteristics. However, the gender wage gap is narrower than it would be otherwise observed because of the overrepresentation of women in the better paid public sector job.

Temegesen and Zeufack (2002) analyzed a manufacturing survey data pooled from four Sub Saharan countries and they find that women in these countries get on average about 87% of men's earnings, only about 29% of this gap is explained by human capital. Armitage and Sabot (1991) found gender inequality in the public sector of Tanzania, but no gender discrimination in Kenya's labour market. Similarly, Glewee (1990) finds no wage discrimination against women in Ghana. On the contrary, female seem better off than males in the public sector. Siphambe and Thokweng-Bakwena (2001) show that in Botswana most of the wage gap in the public sector is due to differences in characteristics (especially education and training), while mainly to discrimination (i.e. different rewards for the same characteristics) in the private sector.

A group of studies on developing countries had a different approach by focusing on the manufacturing sector with matched worker-firm datasets providing extended information on firms

and introducing controls for characteristics of workplaces. Temesgen (2006) finds a significant gap in Ethiopia, where women employees in the manufacturing sector earn on average 73% of men's earning, despite having higher average educational qualifications. The unexplained component is around 60%, but controlling for establishment characteristics result into an even higher "discrimination" component. Fafchamps et alii (2006) investigate the phenomenon of job sorting in African labor markets through a matched employer-employer data set from eleven African countries, and they find that there is a large gender wage gap in all studied countries, a large proportion of which is explained by selection into low wage occupations and firms. Nordman and Roubaud (2006) show that the gender gap is in favour of males in the formal private and informal sector while women seem to be better of than men in the public sector. Moreover, the weak representation of women in the higher paying public sector maintains the gap greater than it otherwise would be.

However, very few studies have looked at this topic with a distributional approach, which is very much important given the high level of inequality that characterizes these countries. Nordman and Wolff (2007) find that controlling for firm-specific effects increases the value and significance of the gender dummy coefficient along the earning distribution in Morocco, which is significantly different at different points of the distribution. Nielsen and Rosholm (2001) use quantile regression to study the evolution of the gender wage gap during a period of economic transition in Zambia during the '90s. They find that the gender pay gap in the private sector was lower at the bottom of the distribution, but in 1996 it was similar across quantiles. However, this study does not conduct a decomposition analysis of the wage gap, therefore the factors driving the gender wage gap remain unknown. Overall, the application of quantile regression in the context of the gender wage gap in developing countries is still in its infancy.

As regard female participation in the labour market, the literature shows that in developing countries the predominance of the informal activity together with the limited role of the public sector in providing stable jobs may give rise to significant selection effects when entering the labour market. The pattern of participation and sectoral distribution importantly differs across sectors, i.e. women participate to a much lesser extent than men in wage employment. In most of the countries examined by Lachaud (1994), around one third of the total male population of households works in wage employment, while the proportion of women is between 10 and 15%. Regarding self employment, the situation of men and women is very similar but when a substantial amount of capital is required, men are usually advantaged. Therefore, labour market segmentation and professional segregation are the rule rather than the exception (Cogneau 1999). This may reflect

discriminatory practices (sexist recruitment methods, stereotypes and prejudice against women, etc.) which prevent women from having access to certain well-paid segments or profession.

In transition countries the issue of gender wage gap has attracted researchers' attention since the beginning of reforms from planned to market economies. Now there is a large empirical literature exploring labour market outcomes in the transitional economies and most of this literature focuses on comparisons between the two pre- and post- reform periods and across countries, by investigating the wage position of women in the transitional economies. The consensus is that the gender pay gap has either remained stable or has actually decreased in Central Europe while it has increases in the countries of the former Soviet Union.

Newell and Reilly (1996) highlighted the dramatic rise in wage inequality experienced in Russia over the post-reform era (1992-1996). They find that no other transitional economy has experienced such a widening of wage differentials. Nonetheless, Reilly (1999) finds that the unadjusted gender pay gap between 1992 and 1996 exhibited some degree of stability and in contrast to Breinard, his analysis suggests that wage dispersion plays a modest role in generating pay differentials by gender. He also finds that while women benefited from a closure of the gender gap in observable skills and prices differentials, this effect was nearly offset by an increase in the unexplained part of the differential.

Breinard (2000) analysed several East European countries around the year when market reforms where introduced (for Poland, January 1990). In the case of Poland, she used the household budget surveys for 1986 and 1992, with an analysed sample of men aged 18-64 and women 18-59, including agricultural and self-employed workers. She found that women's relative wages have increased on all of the Eastern Europe countries surveyed because of improvement in gender specific factors, and Poland has experienced the highest improvement in the region. Therefore, despite the difficulties of moving to a market economy, women have benefited relative to men in the labor market in that region because the returns to human capital characteristics, especially education, have increased with the onset of the market economy. Differently, Russia and Ukraine experienced a substantial decline in female relative wages because of a tremendous widening of the wage distribution.

Ganguli and Terrel (2005a) test the hypothesis that gender gap was lower in Soviet times and at the beginning of transition than in today's Ukrainian transition economy, because of the egalitarian principle of communist state. However, the authors find that the raw gap actually declined as a result of a decline in the gaps in the lower part of the distribution but differences in rewards are the most important factor in explaining the gender gap, at the mean and along the

distribution. They also apply the Machado-Mata decomposition method to data from the Longitudinal Monitoring Survey (ULMS) and they find that there important differences of the gap along the distribution and in different years. At the bottom, improvements in the wage gap in 2003 are driven by the better composition of women's characteristics relative to men's. Another study of these authors (2005b) on the gender wage gap in Ukraine finds that the change in the structure of wages and in observable labour characteristics resulted in an increase of male as well as female inequality.

Orazem and Vodopivec (2000) analysed changes in women's relative wages in Slovenia (1987-1992) and Estonia (1989-1994). They find that the improvement in women's relative wages was mainly due to i) an increase in returns to human capital, favouring more women; ii) a shift in the labour demand toward predominantly female sectors (health, education, financial services, retail trade) and away from traditionally male sectors (agriculture, manufacturing, mining, transportation); iii) the exit from the labor market of low skilled women, especially in Estonia, while women who remained employed had higher average education level.

Newell and Reilly (2001) investigate the path through the 1990s of the gender pay gap in a number of former communist countries of Eastern Europe and the Soviet Union. Regarding Poland, they perform an econometric analysis for 1992 and 1996 of earnings perceived by the main job. The main findings are that: i) in general the gap in pay between men and women has remained low by international standards; ii) most of the average gender difference in pay is attributed to treatment effects over differences in endowment, iii) there is evidence of larger gaps in higher paid jobs relative to lower paid jobs. These findings are also common to the other countries under study.

Hunt (2002) has highlighted the importance of selection effects when interpreting movements in the gender pay gap. Using GSOEP data for 1990-1994, she finds that the wages of East German women rose by ten percentage points relative to men. However, half the relative wage gain was due to the selective withdrawal from the labor market of less qualified females, which reflected a general fall in demand for low skilled workers. Her conclusion points to the question whether the apparent stability of the average gender pay gap observed for other transitional economies hides important movements in the labour market status of women.

Grajek (2001, 2003) focuses on the analysis of the earnings gap in Poland over the period 1987-1996, that is in the last year of central planning and during the period of transition to market economy. He applies the Oaxaca's standard decomposition followed by the Juhn, Murphy and

Pierce (1991) decomposition to a large dataset of Household Budget Survey. He finds that 1989[1] was a breakthrough in the relative improvement for women (the wage gap decreased by 10.2 log points) and the explained and unexplained factors in the decomposition analysis were roughly equally responsible for the falling pay gap. The returns to observed human capital were as important as other factors including returns to un observed human capital, discrimination and sectoral shifts, when the state owned enterprises were reducing their costs. After 1992, the situation has stabilized and even reversed in the following years due to discrimination practices and rising overall wage inequality offsetting the advantages of females due to observed skills. Also, women suffered job and pay cuts, especially in the public service sector.

Using data from the 1993-1997 Polish labour force survey, Adamchik and Bedi (2003) find that the male-female wage differential was quite stable and most of the explained portion of the gap is due to industrial and occupational affiliation of women, which could also be viewed as a consequence of pre- and post-labour market discrimination based on gender stereotyping. Throughout the period, a large portion of the gap remains anyway unexplained.

Newell and Socha (2005) analyse the changes in the size distribution of wages in Poland over the period 1992-2002. They find that until 1998 there was a relatively constant level of inequality, mainly driven by privatisation which was generating major increases in the relative wages of professional and managerial workers. At the same time, the decline in labour market participation of less educated workers was limiting the widening of inequality. The end of the fall in participation around 2000 contributed therefore to an increase in wage inequality. Newell and Socha (2007) find that the proximate causes of the rise in wage inequality after 1998 where shifts in returns to education, increases in private sector employment and rises in the share of young people in some low paid sectors of the labour market.

The review of the evidence in transitional economies suggests that the pay gap between men and women is low by international standards and women have benefited throughout the systemic change from command to market economy. This fall in gender earnings differential was mainly due to the high productivity endowments held by women in these countries with respect to men. Indeed, a large part of the total differential can only be attributed to gender differences in the returns to their characteristics.

[1] It was the year of the first democratic parliamentary elections which resulted in forming the first non-communist government. Poland became the first member of the former Sovietic bloc to re-establish political democracy and market economy.

3. The estimation of the gender wage gap: open issues

Labour economists most commonly define wage discrimination by comparing wages for equally productive workers. This is implemented through the estimation of wage differentials conditional on human capital characteristics that reflect productivity potential. The raw wage gap is then decomposed into a portion explained by differences in human capital endowments and a residual or unexplained part which is the difference in human capital prices. Once taking into account identification issues, decomposition techniques have mostly been mechanically applied to decompose the raw gender wage gap and provide very detailed descriptive results on factors that contribute to wage differences between men and women in the labour market (Wright and Ermish 1991).

Therefore, all the methodologies designed to calculate gender pay gaps rely on the specification of a well-defined human capital model augmented by factors which should capture differentials and monopoly rents associates with an individual's job or sector of attachment. The fundamental methodological issues may be summarized in two points (Kunze 2000): defining precise measures of individual's human capital characteristics, which depends on the availability of data; deriving consistent estimates of the coefficients of the human capital variables in the wage regression model in order to make male and female outcomes comparable, which is the most challenging point from an applied econometrician's perspective and depends on the application of appropriate methods of estimation.

Therefore, most of the discussion associated with these methodologies start from the specification of the earning equations on which the decomposition is based. Some authors (Montenegro 2001) point to the fact that the earning equation is based on some restrictive assumptions, such as assuming that individuals are of equal abilities and face equal opportunities (i.e. perfect capital and labour markets). The consensus is that controls for productivity– related factors is fundamental, but there is no universally accepted set of conditioning variables that should be included in the log earning equation. Nonetheless, there has been much discussion on the choice of the variables to include in the wage regression (Cain 1986; Appleton et alii 1999; Nordman and Robaud 2006). Two issues are particularly crucial.

First, whether the included characteristics are themselves affected by discrimination. An example is the variable controlling for occupation. Given that in most Western industrialised economies strong occupational segregation is observed between genders, taking the distribution of men and women across occupations as exogenous may result in assuming away an important dimension of discrimination. Results from case studies in Africa (Glewwe 1990, Armitage and

Sabot 1991, Appleton and al. 1999) suggest the importance both of sectoral choice and of workers' job status, for analysing differences of wage determination between male and females workers. In many African countries, the predominance of informal activity as well as the decreasing role of the public sector in providing stable jobs for qualified workers may involve significant selection effects at the labour market entry. Indeed this professional segregation may reflect discriminatory practices preventing women from having access to certain well-paid segments or professions.

Second, whether included variables adequately account for productivity (Weichselbaumer and Winter-Ebmer 2005). If omitted variables are correlated with the so called "discrimination effect", then it may capture not only discrimination but also unobserved differences in productivity. Conventional earning functions do not directly measure cognitive skills or reasoning ability, which is very likely to be correlated with schooling. As a result, estimated returns to schooling will be biased upward. In the meanwhile, it is very difficult to conceptualize and measure ability and there is no consensus as to whether it is significant enough to differentiate earnings. Moreover, evidence suggests that the variables of the model are not perfectly correlated with education (Boissière et alii 1985).

As regard experience, Kunze (2007) finds that most of the empirical studies reviewed in her survey do not contain information on the actual work histories of individuals, such as actual work experience, hence proxies are used instead. The most desirable measure to use is actual labour force experience, accounting for labour force interruption, but this is generally unavailable given data constraints and available information. However, most of the studies on transition countries use either individual's age or potential work experience (measured as age minus years of formal schooling minus six, the age at which children usually start going to school) as a proxy for labour experience. The major problem with the potential measure is that it correlates poorly with the actual female labor force experience. This may be particularly relevant in the case of estimation of wage regressions for samples of females as well as young workers. These two groups may have in fact working life cycles characterised by more frequent interruptions (Antecol and Bedard 2004). Given that women are likely to acquire less labor force experience than men through the effects of family formation and nurturing responsibilities, a failure to account for these factors may have important consequences. Proxy measures tend to overstate women's actual work experience by not accounting for interruptions related to parenting, or for any restrictions on the number of hours worked per week (Paci and Reilly 2005). Therefore, it is not surprising to find that women appear underpaid with respect to men for comparable experience. Alternative approaches use imputed experience instead of potential work experience as a proxy for actual work experience for females (Miller

1987, Zabalza and Arrufat 1985) or exclude from the sample married females (Greenhalgh 1980) rather then pooling single and married females. Moreover, it has been argued that less job training, less experience, greater time in housework and lower occupational choices may be voluntary choices made by women which are not captured in the data and may e responsible for some proportion of the unexplained residual (O'Neill 1985, Light and Ureta 1995).

A further potential problem relates to the sample wage observations which may not be randomly drawn from the population, because they represent the outcome of a selection process in terms of unmeasured worker characteristics (Blau 1998, Kunze 2007). This is defined as the sample selection problem and it may violate consistency of OLS applied to wage level equation. The set of observed labour force participants could be either positively selected (i.e. an over-representation of workers with high wage offers having controlled for measured characteristics) or negatively selected (i.e. an over-representation of workers with low wage offers having controlled for measured characteristics). While in Western industrialised countries traditionally almost all men work continuously, independently of their environment and individual circumstances, labour force participation rates within the group of women may vary considerably and hence modelling the decision of women to work is a quite complex issue. Their decision may depend on various observed factors such as the number of children, provision of child facilities, cost of child care, income of the husband or partner, institutional framework and unobserved factors and identification of the earning equation parameters crucially depends on the exclusion restrictions made. Variables selected as instruments to identify the selection effect shouldn't affect the level of the wage.

The discussion suggests that exogeneity of all popular instrument used in the literature seem to depend on assumptions that require further testing to justify their use. Therefore, the standard correction procedures adopted in the face of this type of problem have been subject of much criticism given their sensitivity to identifying restrictions and their general lack of robustness (Manski 1989, Bushway et alii 2007).

4. Measuring the gender wage gap in developing and transition countries

When applying gender wage gap analysis to developing countries, the most crucial issues regard the preliminary phase of data collection and the identification of variables and models which can sute at best the shape and mechanisms of African labour markets.

In most Sub-Saharan African countries any analysis of the labour market is made difficult by the lack of available data. Information on the labour market is fragmentary and irregular, especially with regard to the informal sector. Concerning the urban labour market, studies carried out by statistical services of the ministries of planning or labour are very scarce in number and allow only a partial exploration of the mechanisms ruling the labour market. Relatively very little information is provided on remuneration from self-employment and the benefits of wage earners in the modern production sector.

An additional constraint arises at the sub-national level when household surveys do not allow for appropriate disaggregation or differentiation at the household level. In Africa, households tend to be large and complex and they are involved in a wide range of income-generating activities (Deaton 1997), especially women who co-produce with the breadwinner. The working time follow seasonality, therefore income is variable and part earned in kind (this is one of the reasons why consumption is a more adapt measure of well being).

The definition of the labour force itself is problematic and according to different criteria there may be very different results. The labour force framework, the major measurement device of the labour market which is used in household surveys, fits mainly situations where the dominant type of employment is "regular full time employment" (ILO 1990). Structural labour market conditions in Southern Africa are such that only a minority of the labour force is in regular full time paid employment (Sparreboom 2001). In the urban areas of Africa, unemployment as defined by the ILO does not fully encompass the scope of inactivity. Some persons exercise a secondary and/or main activity while they are looking for a different job. Most studies try to adapt employment categories taken from standard international classifications. Although this approach could be justified by the need of comparable studies across countries, it is not appropriate to analyse the stratification of the labour market in developing countries. The labour categories used fail to identify precarious form of work which overlaps the formal-informal dichotomy and to differentiate the different degrees of vulnerability of certain groups, i.e. females.

Moreover, gender discrimination in the labour market takes place also through mechanisms different from the ones occurring in developed countries, which are also very often neglected by the literature on labour market. The role of traditional institutions in accessing the

labour market increases as employment becomes more precarious. Individual mechanisms of the labour market in Africa have a social effect which goes further than individual well-being (Lachaud 1994). Social norms tend to require women to declare themselves only in unpaid domestic work, which place them outside the coverage of the labour force. Women are more likely to be discouraged workers in the context of limited opportunities because men tend to be served first and they are mostly in informal employment, where they are more likely to be among the working poor (UNECA 2005). As a result, the gender gap is even more difficult to monitor.

As regards transition countries, the most peculiar features of the gender wage gap analysis relate to the nature of the turmoil following the reforms implemented in the first phase of transition and to the structural change of the labour market.

Reilly (1999) faces some of these important methodological issues in his empirical paper on gender pay gap in Russia. He applies the Oaxaca-Blinder decomposition to Russian data over the period 1992-1996 and he finds that the analysis is importantly affected by the high degree of instability introduced in the Russian wage structure by disruptions associated with the transition. Similar conclusions are reached by Glinskaya and Mroz (1996) and Newell and Reilly (1997), as a result of the poor performance of conventional wage equations when fitted to Russian data. Indeed, the volatile and disrupted nature of the Russian labour market over the period considered is in marked contrast to conditions usually found in the more established labour markets in more developed economies. Newell and Reilly (2001) reach the same conclusions after the application of decomposition methodologies to a set of countries undergoing a transition al change.

Therefore, it may be regarded as inappropriate to use such wage equations to inform the evolution of the gender pay gap in countries where the labour markets does not conform to traditional conditions. However, Newell and Reilly (2001) estimate that once this criticism is well taken into account it is still possible to maintain and apply this empirical approach, because it provides a framework within which the effect of wage structure on gender pay gap can be assessed, being specific to the Russian experience. It also marks a benchmark for future work on the country.

A different kind of methodological issues concerns the measurement of men and women's wages. Usually, the pay measure used relates to earnings received in the main job by employees and excludes from the analysis the treatment of secondary earnings. This is quite relevant considering that the incidence of secondary job holding has increased as a result of the changes brought by the transition. Some studies use hourly pay, while others use a monthly measure but control for hours worked as an explanatory variable. However, many studies using the second approach do not take into account the potential problems of endogeneity of hours worked. In addition, the choice between

a pay measure gross or net of taxes may have relevant consequences if the income tax structure in a specific country is highly progressive.

The issue of payment arrears may also have important gender implications in transition countries. Many authors have undertaken gender pay gap analysis but they have ignored arrears while only a few have attempted to explore the impact of arrears on the gender pay gap (Gerry et alii 2002). Enterprise managers might have used wage arrears as a discriminatory tool favouring male workers. Glinskaya and Mroz (1996) conclude that the effect of payment arrears on the gender wage gap is ambiguous, but the inclusion of occupational controls may mitigate such effects. Lehmann and Wadsworth (2007) find that since women are less likely to be observed with wage arrears, if everyone were paid in full the wage gap would increase from 20% to 30%.

Conclusion

The aim of this survey is to present the literature and the open issues faced by the literature in estimating the gender wage gap in developing and transition countries. This is a very important issue, given the socio-economic situations and the mechanisms at work in the labour market in these countries. In Africa, the labour market is central to the livelihood of people in and outside the labour force, and women are very likely to be among the most vulnerable groups. In transition countries, the reforms associated to changes in labour market institutions had a tremendous impact on women's opportunities in the labour market and lifetime decisions.

In the first part of this chapter I present the main existing theories proposed to analyse the gender wage gap. The human capital theory is the reference model for all empirical decomposition techniques, while other theories give difference explanation to the observed gender discrimination in the labour market. The review of the literature reveals that the phenomenon of the gender wage gap has been deeply investigated in industrialised countries and consolidated evidence exists on the evolution of the wage gap since the late '70s. Differently, in transition countries the interest has developed quite recent and in developing countries is still very marginal.

In transition countries, the main findings apply to the change in earning differentials over the transition from centrally based economies to free market economies. The consensus is that nowadays the pay gap between men and women is low by international standards and women have benefited throughout the systemic change from command to market economy. A large part of the

total differential still existing can only be attributed to gender differences in the returns to their characteristics. In developing countries, studies on the gender wage gap mainly suggest that there are significant gender wage gaps in some African countries and none or very low in some others. In some countries, they also find important differences across the public/private sectors and across occupations and firms. Further research on the topic is therefore crucial in order to understand better the forms of discrimination acting against women in the labour market and to adopt the best policies to deal with this phenomenon.

Focusing on the empirical techniques adopted to analyse the gender wage gap, two key-points are of crucial importance: defining precise measures of individual's human capital characteristics and deriving consistent estimates of the coefficients of the human capital variables in the wage regression model, in order to make male and female outcomes comparable. Most of the discussion in the literature over the techniques of wage decomposition stars from the specification issues on which the decomposition is based. This is the most challenging point from an applied econometrician's perspective and depends on the application of appropriate methods of estimation. Moreover, further important issues arise in measuring the gender wage gap in developing and transition countries, because of the specific nature of their labour market and socio-economic situation.

The contribution of my dissertation is twofold: contributing to the literature on the gender wage gap in developing and transition countries and providing more information on the nature of the labour market and on the issue of the wage gap in Comoros and in Poland. As already seen, there is a serious lack in the literature on the gender wage gap in developing countries, which also poses serious concerns given the important linkages between the conditions of the labour market and the extent and incidence of poverty.

In transition countries most existent studies do not go beyond the period of the late '90s when analysing gender wage gap, while the changing nature of the labour market would require a close monitoring and forecasting of the phenomenon.

Chapter 2

The gender wage gap in the Union of Comoros

Introduction

Gender discrimination arises when women and men have different opportunities in the access to resources, such as education, health services, employment, career and decision making process. Especially in developing countries, this kind of discrimination has being recognized to have a detrimental impact on economic growth, poverty and macro-economic outcomes. Most empirical studies support the hypothesis that narrowing gender inequality stimulates per capita growth and overall living standards by increasing average human capital and lowering population growth (Forsythe et alii 2000; Morrison and Jutting 2005). They focus on the importance of gender equality/inequality issues both as outcome variables of overall economic activities as well as determinants to overall well being of society.

In fact, the economic role of women is heavily influenced by laws, norms, codes of conduct and tradition in developing countries (Morrison and Jutting 2004). Economic growth is regarded as a factor that can increase their vulnerability, especially if they are excluded from the transformation process of the economy. According to the UNECA report 2005, female workers in Sub-Saharan Africa are mostly in the informal employment sector where they are more likely to be among the working poor. 84 per cent of female non-agricultural workers are in the informal sector compared with 63 per cent of male non-agricultural workers. In the meantime, access to decent employment and decent wage is reckoned to be a major route out of poverty and situations of intra-household deprivation and violence.

How much is the magnitude of the gender wage gap in developing countries? What is the proportion of gender wage differences attributed to differences in characteristics between men and women, and what is the proportion attributed to discrimination? The purpose of this contribution is to answer these important questions in the context of the Comorian labour market, using data from the last household budget survey (EIM 2004). The main result I obtain by analysing the difference in earnings between males and females at the mean and along the entire distribution is that there exists an important gender wage gap which decreases at the top of the distribution and which is mostly not explained by differences in characteristics between males and females. More importantly, were all women employed in the labour marked, the gap would be much higher.

The situation of Comoros is quite peculiar with respect to the other African countries studied in the literature, which focus on the manufacturing sector. A small and isolated island archipelago, Comoros has not yet achieved take off of the manufacturing or service sector with high enough

productivity to absorb excess labour in agriculture, while in the last decade there has been an explosion of the informal sector.

In 2005 the government engaged under a twelve months IMF Staff Monitored Program (SMP) and committed to set the country on a path of sustainable growth and poverty reduction through the implementation of the National Poverty reduction Strategy Paper (DSRP), with a view to making progress towards the Millennium Development Goals. With the support of donors, the government is actively engaged in order of promote Goal 3, which is about promoting gender equality: "Female education, training and empowerment are identified as the most important factors to remove gender discrimination" (Comores-PNUD 2007, page 10).

Therefore, the challenge of Comoros is how to restructure its economy in order to make it viable and at the same time alleviate poverty and improving the condition of life of the population, especially the more disadvantaged. A deeper knowledge of the functioning of labour market and pay structures faced by various population groups in the economy is thus a key element to implement reform in the labour market and design a mechanism to improve income distribution and allocative efficiency of labour.

In my approach, I investigate the difference in hourly earnings between female and male using two different estimation techniques, OLS and quantile regression methods, with a twofold objective: contributing to the literature on gender wage gap in Africa and shading some light on the situation of discrimination of women in the Comorian labour market. The preliminary estimate of the raw gap shows that the difference in earning differentials between men and women is quite high with respect to international standards and this gap is decreasing along the wage distribution. Applying quantile regression method allows to get deeper and more specific in the analysis of the wage gap: OLS seem to under-estimate the wage gap in the bottom of the distribution and overestimate it in the top. When controlling for covariates, the departure from OLS estimate is even more marked in the middle of the distribution. The main finding is that in the first half of the distribution of wages the gap stems from observed differences, while in the second half the estimated values of the gap increases controlling for personal characteristics and other variables. That means, men have better characteristics[2] than women till the 60th percentile, but afterwards women are better endowed than men. Thus, introducing the covariates makes the wage gap even greater than it would be in the top deciles.

Preliminary to the application of the Oaxaca-Blinder decomposition (Oaxaca 1973, Blinder 1973), I run two separate earning regressions for males and females. In order to take into account

[2] I indicate as skills: education, tenure, age as proxy of potential experience.

sample selection issues, I introduce in my regressions a participation equation that provides important information on the characteristics of the Comorian labour market. Indeed the most relevant point is the difference in the gender participation rate into the labour market. This is the first dimension of discrimination: women who actually work in Comoros are those with the highest potential wages. If all women worked, the wage gap would be much higher. Therefore, a model of intra-household labour allocation decision seems to fit well the mechanism of female participation into the labour market and to explain the results emerging from the participation equation. However, having a cross section dataset I can't account for unobserved heterogeneity in personal characteristics that may stem from omitted common variables or global shocks that affect each individual unit differently.

The most important result I obtain by decomposing the wage gap is that very little[3] of the difference in female and male predicted wages can be explained by the variables used by literature following to the human capital model. Most of the gap is unexplained and if I account for sample selection this part is even greater. More importantly in terms of policy implications, if all women worked, the size of the gender wage gap would be much larger. One of the main conjection of my analysis is that the bulk of this part is due to discrimination. If we believe in this, a fall of the unexplained component would have a positive effect of increasing female participation in the labour market through the rise of wage, relying on the underlying hypothesis that women supply is elastic to wage changes.

A first implication of this result is the necessity to improve research in developing countries in order to: i) deeper investigate on the nature of this unexplained component, ii) isolate the portion of pure discrimination from potential omitted variables related to unobserved characteristics, iii) have better and more complete data, to account for unobserved heterogeneity.

Finally, running the counterfactual decomposition I see that differences in rewards to characteristics between males and females are overall much more important than the difference in the distribution of characteristics between genders. Estimates of the counterfactual gap at different quantiles show in fact that difference in the distribution of characteristics between genders is not so relevant.

The chapter is organized as follows. Section one provides some background information on Comoros and the dataset. Section two presents the econometric methodology I apply throughout my dissertation (OLS and quantile regression analysis, wage decomposition and counterfactual analysis). Section three shows the empirical results of my analysis. Finally, in section four I draw together the main findings and conclude.

[3] Between 14.4% and 20.2%

1. Background on Comoros

1.1 Comoros Context

Comoros is an archipelago lying in the Indian Ocean between the northern tip of Madagascar and the African mainland. It covers a total area of .236 km² and it consists of four main islands: Grande Comore (Ngazidja), Mohéli (Mwali), Anjouan (Ndzuani) and Mayotte (Maore). Formerly a French colony, the islands are independent since 1975, except for Mayotte, which is an overseas territory of France. Today, Comoros belong to Least Developed Countries and to Small Island Developing States (SIDS), being faced with two main hindrances to development - isolation and small markets. After four years of institutional and political crisis, a new constitution was voted in a referendum in 2001 for the Union of Comoros, entailing q greater autonomy for each of the three islands in the archipelago. However, a marked political instability seems to be a constant element in the life of the country.

The situation of the Union of Comoros is quite peculiar with respect to other African countries, the specificity of the country is made of the following elements characterising its economic and social situation. Its economic development is hampered by a very unstable macroeconomic framework with an unsustainable external debt, which the intervention of international institutions in the '90s through stabilization policies and structural adjustment programmes has failed to fix. The rate of growth, which is again positive since 2000 after a decade of decline, is sustained only by internal consumption and external transfers. The ratio of investment on the GDP has steadily decreased with the reduction of external aid in the '90, while the low development of basic infrastructure together with political uncertainty and the small size of the domestic market have limited financial development and affected negatively competitiveness and business climate. Notwithstanding the reversal of the current account deficit in the last few years, the economy of the Comoros in general—and the balance of payments in particular—remains highly fragile, given its dependence on transfers and the high concentration of exports in a few commodities with international prices subject to considerable volatility (vanilla, ylang-ylang, cloves). In Appendix A I report some descriptive statistics on the country.

The structure of the economy (Commisariat Général au Plan 2005) is characterised by a dualism between a predominant agrarian sector, mainly subsistent and not enough productive which contributes to 40% of GDP and a developed tertiary sector mainly based on import trade which

provides on average 52% of GDP. The industrial basis is marginal. Trade policy remains highly restrictive as a result of high tariffs and duties on imports, which are the main source of public revenue.

The agrarian sector, fishing, animal husbandry and the category of familiar aid offers 57.4% of jobs, most of them based in rural area. 46.9% of agrarian job are performed by women.

The informal non agrarian sector is made by micro-entrepreneurs and independents of the informal sector. It contributes to total employment by 16.2 % and activities are mainly located in agrarian zones. The importance of the sector is the greatest in Grande Comore, where it represents 17.5% of total employment, while its relative importance is limited in Mohéli.

The modern sector is represented by wage workers of the private and public sector, although 75% are non protected workers. 51.4% of jobs are located in the agrarian sector. It is very important in the main island, Grande Comore, especially in the capital given its administrative function. As regards women, this sector employs only 13.7% of all women participating in the labour market, most of them are unprotected workers.

Production costs, and in particular labour costs, are relatively high. This may explain why exports outside the three key commodities are very limited. Salaries in the Comoros are higher relative to GDP per capita—Purchasing Power Parity (PPP) adjusted to reflect differences in cost of living—than in its neighbouring competitor countries (IMF 2004). Utilities like water and electricity are also produced at comparatively high cost, and frequent power outages result in an important disruption of production and trade. Structural factors that negatively affect the competitiveness of the Comoros also arise from its geographic location and characteristics befitting a small island country. In particular, its remoteness and insularity result in high transportation costs. A narrow resource base and small domestic market also hinder economic diversification and limit capacity in the private sector.

The population is severely affected by this adverse economic situation. According to the last estimates, 44.9% of the population lives under the poverty line while vulnerability affects a much wider proportion of the population. The social dimension of development is very fragile, though most of the indicators are above the average for Sub Saharan. The estimated level of Human Development Index in 2005 (0.55) corresponds to a life expectancy at birth of 63 years, a combined gross enrolment ratio for primary, secondary and tertiary schools of 46% and an adult literacy rate of 57%.

Under these circumstances, the country launched in 2003 a participatory process aimed at defining a national poverty reduction strategy (I_PRSP transmitted to the IMF in October 2005), which provides a long-term vision of sustainable economic growth and poverty reduction. The

success of the implementation of this strategy together with a satisfactory performance of the economic reforms started in 2005 under the supervision of the IMF, is a prerequisite for negotiating a program that could be supported by the Fund under the Poverty Reduction and Facility (PRGF) program and for obtaining debt relief under the enhanced Heavily Indebted Poor Countries HIPC initiative (IMF 2006).

1.2 Data and descriptive statistics

The data used for this analysis are drawn from a multidimensional household budget survey, "Enquete integrale auprès des ménages" (EIM) run in 2004. The aim of the EIM was updating data on poverty, following indicator on households well being, addressing all the dimensions of poverty and strengthening the capacity and coordination of different institutional actors involved in poverty alleviation. The scope of the survey is thus comprehensive and the data collected can be used to analyse a variety of issues related to population well being and economic activities.

The sample is drawn from a stratified two-stage survey plan with unequal probabilities[4] (which are used during the analysis), in order to be representative of the different reality of the three islands. It covered a total of 2988, households out of 83614 households registered in the last census in September 2003. Of the sampled households, 1391 are in Grande Comore, 1212 in Anjouan and 384 in Mohéli. Of the total population in working age (9728 individuals), 36% are active wage earner (46.6% of men and 27% of women) but only 30% reported information on the income received/earned, 6% are unemployed in search for a job and 58% are inactive. In appendix I report descriptive statistics for the total sample of people in working age (Appendix B), for the sub-sample of employed people (Appendix C) and for the final sub-sample of employed[5] people (Appendix D). The rate of labour supply is very low if compared to Sub Saharan Africa, which is mainly based on agrarian activities, and also to Middle east and North Africa in the '90s (Commisariat Général au plan 2005).

Section five of the questionnaire provides data on occupation, wages and working conditions. The following information were provided: wages, frequency of payments (daily, weekly, every 15 days, monthly every three months, annually), number of hours worked for payment, tenure. Within my analysis, I select individuals aged between 15 and 65 years old (population in working wage) who are currently employed. Earnings declared have been converted

[4] In the first part of my analysis, I accounted for the sampling procedures by calculating OLS estimates through clustering and weights. In the second part, I introduced quantile regressions but without accounting for sampling.
[5] I refer to the sub-sample of my wage regressions

into hourly earnings using the available information on the usual number of hours worked per week in the questionnaire. In figure 1 I draw the wage density by gender.

An important step in the development of this paper has been the preparation of the dataset. Starting from the raw data, it has been necessary to:

- test for the general consistency of the information provided between the different sections of the questionnaire and solving eventual incongruence through matching and crossing information.
- build the variables to be used in the analysis using the cleaned raw data. As soon as some potential problem was identified, the available data where cross-checked with answers in the original questionnaires and the problem solved.

At this stage, several variables classifying the sample of the currently employed people according to their labor market status were built and tested. Different criterions of aggregation were also tested. The final classification is the most adequate to investigate the issue of the gender wage gap taking into account the characteristics of the Comorian labour market and economy.

2: Methodology

2.1 Gender wage gap: analysis at the mean

I investigate the wage situation of the sample of currently employed people by comparing male and female earnings at their mean. The absolute raw differential is estimated by $\widehat{\beta}_1$ in the regression:

Eq 1: $\ln(W) = \beta_0 + \beta_1 F + e$

where *Wt* is the log of hourly wage, *F* is the dummy variable for gender (female=1, male=0) and *e* is the disturbances term..

This measure provides a first estimate of the size of the difference in earnings between male and female workers. However, it does not allow disentangling the part of the gap which is explained by differences in productivity and personal characteristics from the part which is explained by different wage structures existing in the labour market. In order to capture more effectively wage effects due exclusively to gender discrimination, it is useful to control for differences in productivity or other differences that may exist between the individual and job characteristics of the

two groups. This requirement demands the use of multivariate econometric analysis that allows to control or hold constant other factors, while exploring the effect of the relevant characteristics (gender) on the variable of interest (wage). Educational level and labor force experience are the factors that are usually held constant. As a consequence, wage differences mediated through the gender control reflect wage differences for broadly comparable workers.

Following the Human capital model, the log of earnings of an individual *i* is assumed to be a function of a person's productive characteristics which are approximations of marginal productivity and the return to theses characteristics. The earning function formally defined by Mincer (1974) is based on a life-cycle earnings model and has the following form:

Eq 2: $\ln(W_i) = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^{\,2} + \beta_4 Z_i + ui$

where $\ln(Y_i)$ is the natural log of hourly wage, S_i the years of schooling, E_i potential experience (post school investment in human capital), E_i^2 the square of the potential experience, included to account for the declining age-earning profile effect, for a given level of experience, and Z_i represents vectors for other variables. Schooling is particular important for developing countries where returns to education are expected to be higher (Sahn and Alderman 1988, Schultz 2004). With this specification, the introduction of a gender dummy may be interpreted as the effect of gender on log earnings at the various percentiles once one controls for any differences in observed labour market characteristics. By controlling for other variables I would expect the coefficient on the gender dummy to be lower, if the gender gap is explained in some measure by these variables.

In my model the specification of the earning equation is made using the following variables. First, I include the set of human capital characteristics: years of completed education, age as a proxy for potential experience, age squared, tenure in the current job and tenure squared.

Second, I control for the working status of my sub-sample. I divide the sample into four categories: wage workers (working omitted categories), self-employed in the agrarian sector and self-employed in the non agrarian sector.

Third, I control for occupational categories by aggregating the nine occupational classes into six (omitted category: plant and machine operators, assemblers, elementary occupations) because of difficulties of small sample size (Sutherland and Alexander 2002). Military personnel were omitted from the sample. In Appendix A I give details of the construction of the six occupational groups.

Forth, I control for the presence of a union in the working place. The hypothesis is that

union members may earn significantly more than non-union members (Butcher and Rouse 2001).

Finally, I control for the urban/rural living environment (omitted category: rural) and the island (omitted category: Grande Comore, the main highland). In A8-A10, I report some descriptive statistics on wage and on the characteristics of work participants. The most relevant element is the disparity between men and women with respect to education. According to traditional and cultural factors, in the intra-household division of work it is the men (husband, maternal uncle or other male members) who are responsible for the means of living of the family. Until recently, female jobs were mainly domestic aid at home and aid in the fields, and young girls' education was not considered important because their future husband would care about them.

An important concern is the frequent phenomenon of sample selection bias which arises in the estimation wage equation (Heckman 1979; Green 1997). If women choose to work for wages based on reservation wage, those with "no market" income are excluded from wage regression and the sample includes only individuals whose wage is greater than the reservation wage. Thus, the sample of observed wages is biased upward. A solution can be found if there are some variables that strongly affect the probability of being into the labour force but not the offered wage.

Heckman modelled a solution by identifying two equations: a selection equation, which models the probability of engaging in paid work, and an outcome equation, which applies only to those who are observed in paid work. Heckman's procedure, which has become known as "Heckit", consist of first estimating a participation equation for all the observations in the sample using standard probit/maximum likelihood technique model, in order to single out the employed from the non participants. Second, using the parameter estimates from the participation equation to compute a variable λ (the inverse Mills ratio[6], called lambda) which is then included in a second step, as an additional regressor in the outcome equation:

Selection eq: $z_i^* = \gamma y_i + u_{1i}$, where $z_i = 0,1$

Earning eq: $W_i = \beta X_i + u_{2i}$, observed only if $z_i = 1$

Where $u_{1i} \sim N(0, 1)$ $u_{12} \sim N(0, \sigma)$ corr(u_{1i} , u_{12})= ρ

The second step provides also a test for sample selectivity as well as an estimation technique. Standard regression techniques would in fact yield biased estimates when ρ is

[6] The inverse Mills ratio equals the ratio of the probability density function to the cdf evaluated at time t for observation i (Maddala 1985, p. 224).

significantly different from 0[7].

I will introduce a sample selection equation to estimate OLS regression models in order to apply the gender wage decomposition at the mean. Studies on women in developing countries confirm that it is a common practice of the literature on the economics of female labour supply to account for individual and household characteristics that limit an individual labour supply. I'm aware of the debate in the literature on the validity of the exclusion restriction of the instruments that are usually used to account for participation in the labour market (Kunze 2007). However, the relative low participation of women in the Comorian labour market makes Comoros a classical case study where not accounting at all for participation would result in a too biased estimate. In this framework controlling for participation in the labour market gives very interesting results, because it makes possible to capture the effect of a very important dimension of gender discrimination in developing countries.

Following the literature referring to Heckman (1979), I model the participation equation as a function of age, age squared, education, presence of children aged below six years in the household, marital status and island. Given the social, cultural and economic context of the country, I also include variables which control for intra-household labour substitutability or complementarity. Therefore, the number of children aged between six and fifteen years, the number of employed members and the number of adult women in the household are also included in my specification. Only a few studies have addressed the issue of intra-household labour supply decision with this approach. In developing countries, they find that child labor and adult labor are substitute (Ranjan 2000a and 2000b; Shahina 2006; Amin et alii 2006).

2. 2 The gender wage gap along the distribution: an application of quantile regression model (QMR)

Quantile regression technique provides a powerful tool to estimate wage equations because it allows to increase the number of point in the earnings distribution at which the wage gap is evaluated and it has important advantages with respect to OLS estimates.

In fact, usual wage gap measures and decomposition techniques fail to take into account the dimension of distribution and inequality. The classical methodology i) limits the analysis to the

[7] The coefficient on the selectivity regressor is sigma $\sigma * \rho$. Since $\sigma \neq 0$, the ordinary t statistics for testing the hypothesis that ρ =0 can be used, and it will be asymptotically distributed as N(0,1) under the null hypothesis.

calculation of the mean wage gap at an aggregate level, ii) it is not robust to the presence of outliers, while quantile regression results are characteristically robust to outliers and heavy tailed distribution and the estimates are invariant to outliers of the dependent variable. Finally, a quantile regression approach avoids the restrictive assumption that the error terms are identically distributed at all points of the conditional distribution.

The quantile regression model first introduced by Koenker and Bassett (1978) denoted by $Q_\theta(w \mid X)$ the θ^{th} quantile of the distribution of the log wage given the vector of covariates X, where $\theta \in (0,1)$. They model these conditional quantiles as

Eq 3: $Q_\theta(w \mid X) = X'\beta(\theta)$,

where X is a $kx1$ vector of covariates and $\beta(\theta)$ is a conformable vector of quantile regression (QR) coefficients. For given $\theta \in (0,1)$, $\beta(\theta)$ can be estimated by minimizing in β,

$$n^{-1}\sum_{i=1}^{n}\rho_\theta(w_i - X'\beta)$$

with

$\rho_\theta(\mu) = \theta\mu$ for $\mu \geq 0$

$\rho_\theta(\mu) = (\theta - 1)\mu$ for $\mu \langle 0$

The latter expression is referred to as the check "function" because the weight applied to μ will be shaped like a "check" with the inflection point at $X_i'\beta = 0$. Median (50%) quantile regression results correspond to the Minimum Absolute Deviation (MAD) estimator. Using this methodology, the log wage equation is estimated conditional on a given specification and then calculated at various percentiles of the residuals by minimizing the sum of the absolute deviations of the residuals from the conditional specification.

The estimates obtained for the coefficient on the gender dummy variable allow establishing the magnitude of the ceteris paribus gender pay gap at different points of the wage distribution. The log wage quantile regression at, for instance, the 30th percentile predicts the value of the log wage at the 30th percentile rather than at the mean and the interpretation of the gender dummy is the percent (or log point) difference, ceteris paribus, between the male and female 30th percentile wage. When I estimate the raw gap, there is one covariate X_i, the female dummy variable, in equation 3.

Therefore, the raw gender gap is calculated as the difference in log hourly wages between female and male workers at various points of the distribution.

Finally, quantile regression coefficient can be interpreted as the partial derivative of the conditional quantile of y_i with respect to particular regressors $\Delta Q_\theta(y_i \mid x_i)\Delta x$. That is, the derivative is interpreted as the marginal change in *y* at the θ^{th} conditional quantile due to marginal change in a particular regressor (Yasar et alii, 2006).

2.3 Wage decomposition: the Oaxaca-Blinder procedure

This method of decomposition (Oaxaca 1973, Blinder 1973) allows measuring and decomposing the observed gender wage differential into two components: human capital or productivity and returns to human capital. This latter part is defined as treatment or discrimination effect and represents the focus of policy makers, providing information on the prevalence and magnitude of unequal treatment or compensation between men and women in the labour market.

The decomposition technique involves two case scenarios, based on which men-women ratio would prevail in the absence of discrimination: i) men would face the same wage structure currently faced by women; ii) women would face the same wage structure currently faced by men. That is, in a situation of no discrimination women would receive the same wage as they presently do, but if discrimination occurs, men would receive on average more than what they would be awarded in a non-discriminating labour market; vice versa, in a situation of no discrimination men would receive the same wage, but if discrimination arises, women would receive on average less than what they would be awarded in a non-discriminating labour market. Therefore, it is possible to use either men's or women wage prevailing structure as a base in order to calculate gender wage gaps.

Once the earning function is run for men and women using equation n.1, it is possible to estimate the average log of wages $\ln(\overline{W})$ for each gender group of the sample, by calculating the fitted value of wages or earnings at the means of the independent variables, indicated by $\overline{X}'\hat{\beta}$. The difference in the fitted values can be decomposed into "explained" and "unexplained" components in the two possible specifications:

Eq 4: $\ln(\overline{W}_m) - \ln(\overline{W}_f) = \hat{\beta}_m(\overline{X'}_m - \overline{X'}_f) + \overline{X'}_f(\hat{\beta}_m - \hat{\beta}_f)$

Eq 5: $\ln(\overline{W}_m) - \ln(\overline{W}_f) = \hat{\beta}_f(\overline{X'}_m - \overline{X'}_f) + \overline{X'}_m(\hat{\beta}_m - \hat{\beta}_f)$

In equation 4, the male wage structure is taken as the non-discriminatory benchmark and the male coefficient as the non-discriminatory wage structure. The first term on the right stands for differences in human capital characteristics, the endowment effect, evaluated at male returns, and the second term represents differences in returns to men and women evaluated at the mean of women's characteristics. Under discrimination, males are paid competitive wages but female are underpaid

In equation 5, the female wage structure is taken as the non-discriminatory benchmark and the female coefficient as the non-discriminatory wage structure. The first term on the right stands for differences in human capital characteristics, evaluated at female returns, and the second term represents differences in returns to men and women evaluated at the mean of men's characteristics. Under this scenario, women are paid competitive wages but discrimination arises under the form of male being over-paid (i.e. nepotism).

The explained part corresponds to the average difference in wage level that is expected on the basis of the differences between women and men's human capital characteristics. The unexplained part shows the differences in wages related to differences in returns to these human capital characteristics. Depending on which wage structure would prevail in the absence of discrimination, there will be a different estimate of the discrimination component. This choice has been treated as an index number problem, where

Eq 6: $\beta^* = \Omega\beta_m + (I - \Omega)\beta_f$

The literature has proposed different weighting schemes to deal with the problem. Oaxaca obtained the estimates for both specifications and used the results to establish a range in which the "true" values of the component lie.

Neumark (1988) reduced the decomposition to the two Oaxaca's specifications in case of no discrimination in the wage structure, that is $\beta^* = \beta_m$ or $\beta^* = \beta_f$ He proposed using the coefficients from a pooled model for both groups:

Eq 7: $\ln \overline{W}_m - \ln \overline{W}_f = \beta^* (\overline{X}_m - \overline{X}_f) + \left[(\beta_m - \beta^*) \overline{X}_m + (\beta^* - \beta_f) \overline{X}_f \right]$

Neumark shows that β^* can be estimated using the weighted average of the wage structure of male and female. The first term is the gender wage gap attributable to differences in characteristics while the second and the third terms represent the difference between actual and pooled return for men and women. In the context of OLS regression, the method proposed by Neumark is equivalent to using the weighting matrix

Eq 8: $W = (X_m^{'} X_m + X_f^{'} X_f)^{-1} (X_m^{'} X_m)$

where X_m and X_f are the matrices of observed values for the two samples (Oaxaca and Ransom 1994).

Usually, in the absence of discrimination population's characteristics give raise to a distribution of wages intermediate between the majority and minority distribution. The amount of the difference in wages attributable to majority overpayment is determined by the belief as to how much weight should be placed on the majority wage structure and the minority wage structure, respectively, in determining the distribution that would prevail in the absence of discrimination. In this case, the pooled model is closer to the model with women as reference group.

It is to note that the derivation of consistent estimates of the total explained part of a wage differential and an estimate of discrimination depend on consistency of the measures of human capital characteristics included in the wage model and consistency of the estimates of the parameters of interests. Consideration of standard errors of the parameter estimates can be used to evaluate efficiency of the estimated component of the decompositions. Moreover, while estimation and interpretation of the total explained part of the gap is straightforward, estimation of the contribution of single factors is only possible for variables included in the explained part. This interpretation is not possible for the unexplained part. While the wage gap due to the sum of the differences in all coefficients is well defined, the wage gap due to differences of a subset of coefficients is not (Kunze 2000)[8].

[8] This problem is even more relevant if dummy variables are included in the vector of regressors, since the decomposition may critically depend on the chosen reference point (Oaxaca and Ransom 1999).

2.4 Conterfactual decomposition of changes in wage distribution : the Machado-Mata approach

The Oaxaca-Blinder decomposition is based on the ordinary least squares property that the mean wage conditional on the average characteristics of the sample equal to the unconditional mean wage. Therefore, it is possible to obtain an exact decomposition of the average wage gap between both groups of workers. However, in the context of quantile regression the unconditional θ^{th} quantile wage is equal to its θ^{th} quantile wage conditional on the vector of average characteristics of individuals at that percentile plus the mean of those individuals' error terms. But this error term is not zero, thus it is not possible to perform an exact decomposition of the wage differential at different quantiles.

Machado and Mata (2005) propose a method that extends the traditional Oaxaca decomposition of effects on mean wages to the entire wage distribution by allowing to overcome the above described problem. This method is based on the estimation of the marginal density function of wages in a given year implied by counterfactual distributions of some or all the observed variables of interest. Their main methodological contribution is to marginalize the conditional wage distribution estimated through quantile regressions run separately for men and women using different scenario for the distribution of workers attributes, which allows obtaining the distribution that would prevail if all workers had the same observable characteristics.

The basic idea underlying their model is drawn on the inverse probability integral transformation theorem, which states that if x is a random variable with a cumulative distribution function F(x), then $F^{-1}(x)$ ~ U(0,1), therefore $X_i\beta_\theta$ has the same distribution as y for a given X_i and a random θ ~ (0,1). That means that it is possible to create random sample while maintaining the conditional relationship between the log wages and the covariates.

On these premises, I create counterfactual densities where women are given men's characteristics in one scenario, and women are given men's "rewards" using both quantile regression and bootstrapping techniques following the approach below:

1. I draw m random numbers from a uniform distribution on (0, 1): $\theta_{1,}, \theta_2 \ldots\ldots \theta_m$; here I set m=500, as the quantile I will estimate.

.

2. For each θ_i, where i = 1,2,…,m, I estimate the quantile regression coefficient $\beta(\theta_i)$ for men and

women from the model $Q_\theta(w \mid X) = X'\beta(\theta_i)$

3. I generate random samples of male and female characteristics X , by making m draws separately for men and women with replacement.

4. I generate the predicted counterfactual wages $W_f^* = X_m\hat{\beta}_f(\theta_i)$, $W_f^* = X_f\hat{\beta}_m(\theta_i)$ and construct counterfactual gaps $X_m\beta_m - X_f\beta_m, X_m\beta_m - X_m\beta_f$

Therefore, I can compare the observed gap to counterfactual gap to learn whether the gap would be lower or higher.

3. Empirical results

3.1 The raw gender wage gap

I first run OLS and quantile regressions on my pooled male-female dataset with no controls being X_i the female dummy variable only, in order to estimate the raw gender gap at different point of the distribution. I report the result of this first estimation in table 1 and I plot the gender log wage gap at each percentile in figure 2.

[Insert table 1] *[Insert figure 2]*

Albeit preliminary, these results lead to several conclusions. First, the gap is quite high compared to international standards, its values ranging from 0.73 to 0.29 log points. Second, the gap it is not constant along the distribution and it follows a decreasing path. It starts from its highest value in the bottom of the distribution, then it slowly reduces towards the median value (0.53) and it fluctuates around it between the 40th and the 70th percentiles. In the remaining part of the distribution, it further reduces by reaching its lowest value of 0.30 log points in the top percentile.

By comparing quantile with OLS estimation, I can see that the gap at the 50th percentile is 0.5log points higher than the mean wage gap (0.48). Quantile confidence intervals are almost perfectly overlapping OLS confidence intervals between the 20th and the 70th percentile. However, outside this interval, quantile estimates are different from OLS estimates. That is, OLS tend to underestimate the gap in the bottom of the distribution and overestimate it in the top. However, if we look at the confidence intervals they are not statistically different. Hence, these preliminary findings suggest OLS estimates - which are not robust to extreme observations or non-gaussian distributions of residuals - may be biased and therefore it matters to assess the magnitude of the gender earnings gap not only at the mean of the sample but along the entire distribution.

It is important to note that this decreasing profile along the distribution points to the existence of a glass floor effect which in Comoros works against women receiving low wages. Precisely, there is 'sticky floor'[9], in the sense that the raw gap is higher in the lower half of the distribution compared to the higher half. Again, the implications of this finding for Comoros are very important in terms of policies and there are strictly related with labour supply decisions. The previous interpretation focusing on female labour supply decision may explain why women do not get into the labour market: if they did it, their wage would be too low.

This result is at odds with recent finding of the literature on gender wage gap Europe end Latin America. The conclusion reached by these studies (Arulampalam et alii 2004; Gardeazabal and Ugidos 2005) is that there is an acceleration of the gap in the upper tail of wage distribution, which is interpreted as a "sticky" or "glass" ceiling effect. In Ukraine, Ganguli and Terrel (2005a) found the same evidence of a persisting ceiling effect, with a gender gap exhibiting a decrease in the bottom of the distribution in 2003. In Chile, Montenegro (2001) finds that the unexplained wage gap is not constant along the conditional wage distribution, but it increases from 10% to 40% as the conditional wage distribution moves from the lower to the upper part.

[9] The term 'sticky floor' is discussed in Arulampalam et alii (2004) and it refers to the gap at the bottom of the distribution and how persistent it is. They define a glass ceiling as occurring when the 90th percentile wage gap is higher than the gap in other parts by at least 2 points.

3.2 The conditional gender wage gap

Wage regressions account for various factors that may explain differences in individual wages, including gender. The estimated coefficients for a gender dummy from wage regression captures the extent to which the wage gap between women and men remains unexplained after checking for other individual differences and control variables. In this section, I look at the gender gap introducing the set of covariates of my earning equation (table 3). In table 5 I present results from quantile regressions on the pooled male and female dataset, with the underlying hypothesis is that returns to labour market characteristics are the same at various quantiles for men and women. A preliminary observation, is that Comorian data present an high volatility, which also makes OLS and quantile estimation confidence intervals very much overlapping. Therefore, all comments apply to the comparison between punctual estimation points, while if we look at the variability of confidence intervals the two estimates are not statistically different.

[Insert table 2]

From table 1 and 2 I see that the two estimates of the OLS wage gap are very similar (-0.48 log points). Looking at the distribution at the quantile, I find that the gender wage gap is higher in the first two quantile, then it declines till the 70th percentile and it reaches again the same values as the 30-40th percentiles. Compare the two measures of raw and adjusted gap, the most important result is that the raw gap is higher than the adjusted one till the 60th percentile of the distribution, while in the last three deciles the reported conditional wage gap is lower with respect to raw wage gap. These estimates show that part of the earning gap stems from observed differences in only in the first part of the distribution, while in the upper part introducing the covariates makes the wage gap even greater than it would be without controlling for them in the top deciles. That means, men have better skills than women till the 60th percentile, but afterwards women's characteristics are better. However, the gap is still much higher at the bottom than at the top of the distribution, confirming the previous finding of glass floor effect working against less skilled women at the

bottom of the distribution.

[Insert figure 3]

In figure 3, I see that the variability of the conditional wage gap in the tails is more compressed with respect to OLS confidence intervals than in figure 1. That is, OLS under-estimation and overestimation effects respectively in the bottom and in the top of the distribution are reduced with respect to quantile estimates. The gap estimated at the 50^{th} percentile is now equal to 0.38 while the mean estimated gap is nearly similar to the gap estimated at the 25^{th} percentile. After this threshold is overtaken, the line fluctuates above the OLS estimate. Interestingly, if we control for observable characteristics, OLS underestimate the wage gap in the first two quantiles but after the 25^{th} percentiles it systematically overestimate it, whereas in the previous case there was a fluctuation around the OLS mean and not such a marked departure from it.

3.3 Quantile regressions

The assumption of equal returns to observable characteristics for women and man may not hold in reality. Coefficients from wage regressions that have been estimated separately for women and men point to the unequal rewards to labour market characteristics, if substantially different. To confirm this hypothesis, I perform an earning equation where I interact all the variables with gender, in order to run a Chow test. According to the result, I can reject the null hypothesis that the regression coefficients are all equal to zero and the returns to characteristics for men and women are significantly different. Therefore, I can proceed by going a step further in depicting the gender wage gap through isolating a part of the wage gap that could be explained by the difference in observable characteristics between women and men from the part which is due to differences in returns to these characteristics.

Tables 6 and 7 present the quantile regressions results for women and men. Figures 4-6 allows a visual appreciation of the returns to various characteristics for male and female for the most significant and interesting coefficients.

[Insert table 6 and 7] *[Insert figures 4-6]*

As expected, the proxy for education has a positive and significant effect for male and it rises across the quantiles. For women, it is significant only from the 50th percentile onwards, and the magnitude of the coefficient is nearly the same as in the male regression. For men, this result is consistent with finding from Morocco (Nordmann and Wolf 2007) and Portugal (Machado and Mata 2001), where all aspects of human capital are more valued specifically for high paying jobs. In figure 3, we see that the pattern of return to education for men is quite flat and fluctuating around mean estimates. Quantile estimates differ from OLS only in the upper quantile. Conversely, female return to education show an increasing trend and it differ significantly across the quantiles, crossing OLS estimate at the 40th percentile.

Age is significant for men till the 40th percentile with a decreasing value of the coefficient. It looks like potential experience is rewarded only in the first half of the distribution and much more than education, while in the second half the effect of education is the stronger and experience is not significant. For women, age is significant only in the first two percentile. With respect to women, we can notice an analogous switch off effect between experience and education but less clear, because between the 30th and 40th percentile they are both not significant. In figure 4 we see that age it is not significantly different for men from OLS estimates between the 20th to the 40th percentile, while in the bottom it‘s returns to it are higher. For women, it is significantly different from OLS estimates in the bottom of the distribution, it means OLS would underestimate it.

As regard tenure, it is significant between the 20th and 60th percentile and in the 80th for men with decreasing coefficient, while for women it has a negative and significant effect till the 30th percentile. This result is at odd with the existing literature, because it means that the higher is women’s tenure, the lower is their wage. This may be do to a problem of quality of data[10]. Returns to tenure are significant only for men, positive and with a limited variability decreasing as in the case of age (table 7).

Union is positive and significant as expected for both samples. For men, it is significant in all the distribution but the top and bottom percentile, with an increasing effect. For women, it is very significant till the 70th percentile with a decreasing effect. The value of the coefficient is between three and two times higher than for men, and it’s highest value is in the bottom percentile.

Working in an urban environment has a positive and very significant effect for male which

[10] There was a problem of reporting in the questionnaire regarding the definition of this variable.

is higher in the bottom and the top of the distribution, while it appears not significant for women except than in the 10th percentile.

Being self-employed in the agrarian sector is positively related to wage for men between the 30th and the 80th quantile, while being self-employed in the non agrarian sector has a negative significant effect on the wage level in the first half of the distribution. For women, being self-employed in the agrarian sector has a positive and significant effect in the 60th and 70th percentile, while being self-employed in the non-agrarian sector has a negative and significant effect in the first two percentiles.

As regard occupational group for men workers, being in group one[11] and five[12] is positively correlated with the level of wage especially in the 30th and 40th percentile. Being in group two[13] has a positive and significant effect in the 20th and between the 40th and 60th percentile. For women, being in group one, two and three[14] has a positive effect except that in the last top percentile. The effect is higher in the first half of the distribution and for group three, also in the 80th percentile. However, being in group four[15] has a negative and significant effect in the 60th and 70th percentile.

3.4 Gender wage gap decomposition

3.4.1 Participation and labour supply decisions

In order to apply the Oaxaca-Blinder decomposition, I specify my sample selection equation. The earning regression run separately for male and female is reported in table 4, with and without selectivity correction. Several comments apply to these results.

[Insert table 4]

First, the specified model of participation in the labour market[16] is very significant for women but not for men, and not in the model with pooled data. A possible explanation for this

[11] Legislators, administrators, managers, professionals.
[12].Trade workers
[13] Technicians, associated professionals, clerks.
[14] Service workers, sale workers.
[15] Agricultural workers.
[16] I reported the variables used in the model in table 12

disparity can be found by looking at the composition of the sample. In fact, 47% of men in working age are into the labour market versus 27% of women, that is the participation of women in the labour market is much more limited and probably affected by external variables that should be taken into account in order to limit the bias.

As we see from the estimated value of lambda, for women there is a positive correlation between the probability of being into the active labour force and having a high wage. This is a really a key point: women who actually work in Comoros are those with the highest potential wages, while if all women worked the wage gap would be much higher.

Second, the results of the selection equation are very interesting in order to understand some of the mechanisms of participation into the labour market. As expected, the probability of men having a job is globally higher than the probability of women. However, this difference is more marked for certain categories. Married men are the more likely to be employed with respect to divorced or never married men. The same is true for women, but the probabilities are much lower. In general, the most disadvantaged in accessing to the labour market, are young people and especially young women. They are more likely to enter the labour market by the agrarian and informal sector, which will lead them towards marginal activities. Differently, young men face a more diversified offer, including especially wage earning which can be important in their professional career (Commisariat Général au Plan 2005).

The number of children and small children is not significant for women. This is a puzzle with respect to the hypothesis of the traditional theory concerning the negative influence of the number of children on the probability of women being employed. However, this result is consistent with the evidence found by Lachaud (1994) in his study on Sub Saharan Africa, where he finds evidence of the traditional hypothesis only in one out of six countries. As regard men, the number of children is significantly negative while the number of small children is significantly positive. This could be explained by a model of inter-household labour allocation decision. According to this, the family has a considerable influence on the behaviour of its members and it is possible to assume that family structure may influence the labour supply of an individual. In this context, men are more likely to work when they have small children but when children reach a certain age[17] and can start working, they might reduce their labour supply.

The number of persons working in the family has a positive effect on the probability of being employed, for both men and women. This effect which apparently is opposite to expectations, may have two possible explanations. It may be due to the high number of families of agrarian self-employed workers in the sample. Or it may be due to a sort of network effect, that is existence of a

[17] The phenomenon of infant work is important in Comoros (Commisariat Géneral au plan, 2004).

positive network of personal relations around a person (usually the husband) is more conducive to find jobs offers to other members of the family. Again, the family seems to be a key unit of decision at the labour supply allocation level, especially if we consider that individual mechanisms of the labour market in Africa have an indirect social effect far beyond individual well being. Lachaud (1994) finds that even if the existence of technical skills is taken into account, the form of employment of the head of the household is a major determinant of the nature of the social relations which structure the work environment and in particular the process governing access to employment. Thus, the vulnerability on the labour market of the head of the household may well create external negative factors in terms of employment and income for other members of the household.

The number of adult women is very significant both for male and female but with a negative sign. It confirms the existence of an interaction effect between the individual labour supply within households. As the number of adult females in working age increase in the household, both men and women tend to reduce the labour supply.

3.4.2 Empirical results of the decomposition

The results of the wage decomposition for the two models are presented in table 8, with and without sample selection correction. In the model without selection the predicted log hourly wage for men and women is respectively 5.80 and 5.28 Comorian Franc (FC) FC. The wage gap amounts to 0.52 log points[18]. When men are considered as the reference group, that is they are rewarded for some specific characteristics, the explained part of the wage is 17% while 83% remains unexplained. When women are the reference group, that means that they are paid less than men because they are discriminated for some specific characteristics. Also in this case the explained part of the wage gap is very low, while the unexplained part accounts for 8.6%. By using a weighted wage structure, the explained part amount to 20.2% and the unexplained, corresponding to differences between actual and pooled returns for men and women is 79.8%.

In the model with selection the log hourly wage for men is 5.81 Comorian Franc (FC) and for women is 4.20 FC. As expected, the wage gap is much larger than in the model without

[18] The formula I utilise: `Wage gap= (Wm-Wf)/Wf*100=(Wm/Wf-1)*100, lnWm-lnWf=ln(Wm/Wf)=+0.48 Wage gap= " 100*(exp(0.48)-1)`

selection. That means, had all women worked in the Comoros, the gender gap would have been considerably larger. By correcting for the selection bias we take into account some factors that may influence the wage gap through female participation into the labour market. When men are the reference group, the explained part represents 4.5% while 95.5% is the unexplained part. When women are the reference group, the explained part is 9.8% of the wage gap, while the unexplained part accounts for 90.2%. If we use a weighted wage structure, the explained part amounts to 11.4% and the unexplained is 89%. A further important consideration is that if we make that the bulk of the unexplained part emerged from is due to discrimination, then a fall of this component would have a positive effect of increasing female participation in the labour market through the rising wage, if women supply is elastic to wage changes.

The most important result, is that very little of the measured wage gap can be explained by the variables used by literature following the human capital model. Most of the gap is unexplained, but if I account for sample selection the importance of this component is even greater. The range of variation of the explained part of gender wage gap calculated at the mean is between 4.4% and 17%, while the remaining 83-95% is unexplained. Moreover, if all women worked, the size of the gender wage gap would be much larger.

The explanatory power of the variables of the model is very low with respect to other developing countries and especially to developed countries. As already mentioned, studies on gender wage gap in industrialised countries has monitored a decrease of the unexplained part from 1980 to 1990 from 40 to 20%, suggesting that women improved their unmeasured skills relative to men or that discrimination against them decreased (Blau and Khan 2007).

3.5 Counterfactual gender wage gap

The quantile regression analysis undertaken in section 3 provided an insight into the ceteris paribus gender pay gap as we move across the distribution. This general result represent an important issue and suggests that it could be very useful to estimate separate quantile regression models by gender and provide decomposition at different quantiles of the distribution.

[Insert table 9]

Looking at table 9, several points should be noted. First, the predicted male wage for the 90th percentile is approximately 1.75 times the predicted male wage for the 10th percentile. In the case of women, the difference is 2.03 times. If I apply male returns to female characteristics, the predicted female wage for the 90th percentile is 1.87 times the predicted female wage for the 10th percentile. If I apply female returns to male characteristics (row n.5), the predicted wage for the 90th percentile is 2.16 times the predicted female wage for the 10th percentile. However, the wages calculated in the 10th and 90th percentile would be respectively 80% and 98% of observed men's wage. This suggests that female wage distribution is less compressed than male wage distribution, but this is due to an higher dispersion of female wages in the bottom of the distribution.

Second, if I compare the distributions of predicted wages, I see that the value of the female predicted wage at the 50th percentile is situated in the 30th percentile of the distribution of the predicted wage for men. If I apply male returns to female characteristics, the predicted counterfactual wage would correspond to the wage sited at the 50th percentile of the male distribution. That means, when applying male returns to female characteristics the ranking of female wage in the male distribution improves importantly.

Third, if I apply male returns to women characteristics (row 4), the counterfactual gap (row n.6) is negative till the 70th percentile while it becomes positive in the two top deciles. The most relevant point is that this counterfactual gap is much lower than the observed gap, especially at the bottom, which confirms the previous finding of floor ceiling effect. Moreover, there would have been an "inverse" wage gap in the top two deciles. That is, women would have been paid higher wages than men at the top of the distribution, while in the rest of the distribution the earning differential between males and females would have been importantly smaller. Also, this counterfactual gap follows a quite regular decreasing path (in absolute value) along the distribution, meaning that the difference in the Betas is important all along the distribution but in the upper part of the distribution women's characteristics tend to improve with respect to the lower part. In figure 7 I draw the distribution of this counterfactual, as explained part of the decomposition of the wage gap.

Forth, if I apply female returns to male characteristics (row n.5) the counterfactual gap (row n.8) is negative all along the distribution and it follows a decreasing path very similar to the observed gap. Indeed, the difference between the counterfactual and the observed gap amounts to 0-1%. Only in the 30th, the 40th percentiles the counterfactual gap is slightly higher than the observed gap, while at other points of the distribution the counterfactual tend to be similar or slightly higher than the observed gap.

This means that at the men and women have very similar characteristics along the distribution. In figure 8 I draw the distribution of this counterfactual, as unexplained part of the decomposition of the wage gap when women are the reference group. From the graph, I see that the unexplained part almost overlap with the total wage gap, meaning that most of the gap is unexplained.

Therefore, the bulk wage gap can be attributed to women's lower returns to labour market characteristics as compared to men's return. At the top of the distribution, only a very small portion of the gap can be attributed to men having better characteristics than women .

The size of the counterfactual gaps in row n. 2 and 4 relative to the observed gap (row n.1) can also be interpreted as a term in separate decompositions. The ratios between counterfactual 1 and observed gap $(X_f - X_m)\beta_m / WG_o$ indicates the relative importance of the differences in men's and women's characteristics in explaining the observed gap. The ratio between gap counterfactual 2 and observed gap $(\beta_f - \beta_m)X_m / WG_o$ (row 9) indicates the importance of different returns to characteristics in explaining the observed gap. As expected, difference in the pay structure is far more important then the differences in their characteristics along the all distribution. This result confirms the finding reached through the counterfactual decomposition of wages in Morocco (Nordman and Wolff, 2007), where women are less rewarded for their observed endowments than males and this is all the more true when they reach top position.

5. Conclusion

Starting from the preliminary evidence of a very high raw wage gap, I fist analyzed the data with OLS and quantile regression approach. A first important finding is that the gap shows a decreasing path from the bottom to the top of the distribution. This fact points to the existence of a glass floor against women receiving low wages. In that sense, this result evidenced for a developing country like Comoros are different from those observed in European and African countries like Morocco, where the gender gap is often observe to be higher in the top of the distribution. By controlling for the set of covariates, the conditional gap is lower than the raw gap till the 60th percentile, but in the top of the distribution its becomes wider. Therefore, the main finding is that in the first half of the distribution part of the earning gap is explained by observed differences in characteristics, while in the upper part taking controlling for personal characteristics and other

variables increases the estimated values of the gap. That means, men have better characteristics than women till the 60th percentile, but afterwards women are better endowed than men. Thus, introducing the covariates makes the wage gap even greater than it would be in the top deciles.

Quantile estimations for separate regressions for men and women are shown in order to illustrate gender differences in returns to labour market characteristics. Education, whose returns are expected to be higher in developing countries, is significant for female only after the median, with an increasing trend. Tenure is negative but not significant for female. Age, a proxy for experience, is significant only in the first two percentile. It looks like the model used in the literature can explain very little of the determination of female wages. However, when controlling for occupational group, being in the highest group (professional, managers and technicians) has a positive and significant returns, which may offset the importance of education being this kind of jobs very much related to education.

In order to implement the Oaxaca-Blinder decomposition, I run a two-steps Heckman procedure with a model of participation in the labour market which shows a significant selection bias for women. The most relevant point is in fact the difference in the gender participation rate into the labour market. I found that the first dimension of discrimination against women in the labour market is related to those social, cultural and economic factors that prevent women to participate into the active labour force. In this context, a model of intra household labour allocation decision allows to explain some of the dynamic of female participation into the labour market emerging from the participation equation. The most disadvantaged, are young not married women, while it seems to be an important network effect which is specific of a segmented labour market.

The results of the wage decomposition at the mean confirm the evidence of a wage gap of 61%, which in case of correction for selectivity bias becomes much higher. That is, if all the women excluded not having any economic activity would join the labour market, the average wage would be much lower. Interestingly, most of the gap estimated is unexplained, that is it may depend from other unobservable factors or it can be due to pure discrimination. It is important to notice that having a cross section dataset I can't account for unobserved heterogeneity in personal characteristics that may stem from omitted common variables or global shocks that affect each individual unit differently. All that means it is necessary to improve research in Comoros as well as in other developing countries in order to: i) deeper investigate on the nature of this unexplained component, ii)isolate the portion of pure discrimination from potential omitted variables related to unobserved characteristics, iii)have better and more complete data, to account for unobserved heterogeneity.

Progress in this direction is also necessary to help policy makers to design effective labour market policies and to fight against discrimination. This is particularly important, because a general policy aimed to simply increase participation in the labour market would not necessary increase female well-being or decrease the wage gap. A policy aimed at promoting female education would not do it either.

In order to have a more precise understanding of the mechanisms governing the labour market, it would be necessary to focus on both the supply and demand side. According to which of them is more relevant, there are different policy implications on labour market and gender wage gap.

Please, note that this paper is an analysis of the supply side. As far as the demand is concerned, the only information available refers to the macroeconomic background of the country. The lack of manufacturing sector, the low development of basic infrastructure and the very limited provision of services by the State make the labor demand very weak.

A possible interpretation of the previous results focusing on female labour supply is that the opportunity cost of working in the market is too high with respect to the cost of accomplishing their domestic daily tasks. That is, the value of the wage they would receive is too low with respect to the value of the domestic and caring tasks which are reserved to them. Therefore, in order to make more women getting into the labour market there are two possibilities: having an increase in the wage or reducing the opportunity cost of working, for example trough the improvement of basic infrastructure, access to drinking water and the availability of maternal schools.

The counterfactual decomposition allows extending the Oaxaca-Blinder wage decomposition method to quantile regression. If I apply male returns to female characteristics, the counterfactual gap is much lower at any point of the distribution and women would be paid more than men at the top. The important reduction of the gap at the bottom confirms the glass floor effect affecting women at the bottom of the distribution.

If I apply female returns to male characteristics, the counterfactual gap is very similar to the observed gap at any point of the distribution, meaning that the difference in the distribution of characteristics between genders is not so relevant in explaining the wage gap. At the top, women have slightly lower skills than men. Overall, difference in the pay structure is far more important then the differences in their characteristics along the all distribution

Tables

Table 1: Raw gender wage gap

Perc.	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	OLS.
RAW	-0.732	-0.619	-0.511	-0.511	-0.532	-0.511	-0.352	-0.336	-0.300	-0.481
	(0.109)	(0.063)	(0.054)	(0.045)	(0.034)	(0.050)	(0.065)	(0.085)	(0.125)	0.053

Standard errors are reported in parenthesis

Table 2:Conditional gender wage gap, by OLS

Perc.	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	OLS
ADJUST	-0.548	-0.525	-0.443	-0.426	-0.388	-0.381	-0.430	-0.454	-0.435	-0.484
	(0.077)	(0.067)	(0.062)	(0.049)	(0.058)	(0.047)	(0.064)	(0.059)	(0.131)	(0.053)

Standard errors are reported in parenthesis

Table 3: Pooled OLS Wage regressions with and without sample selection equation (dependent variable: log hourly earning 2004)

	Log hourly earning 2004		
	Heckman with Probit eq		Regression with no selection
	wage	occ	wage
female DV	-0.526***	-1.026***	-0.484***
	(0.125)	(0.046)	(0.053)
age	0.094**	0.179***	0.045***
	(0.041)	(0.010)	(0.015)
age^2	-0.001**	-0.002***	-0.001***
	(0.000)	(0.000)	(0.000)
years of education	0.025***	0.006	0.026***
	(0.008)	(0.005)	(0.007)
tenure	0.001		0.009
	(0.010)		(0.008)
tenure^2	-0.000		-0.000
	(0.000)		(0.000)
union	0.345***		0.304***
	(0.102)		(0.086)
urban	0.247***		0.230***
	(0.074)		(0.059)
	(0.108)		(0.092)
Self-emp A	0.117		0.165**
	(0.081)		(0.074)
Self-emp NA	-0.209		-0.332***
	(0.131)		(0.089)
Anjouan	-0.068	0.403***	-0.106*
	(0.105)	(0.064)	(0.055)
Moheli	-0.445***	0.394***	-0.504***
	(0.147)	(0.071)	(0.076)
married		0.956***	
		(0.074)	
divorced		0.874***	
		(0.091)	
n. children		-0.026*	
		(0.015)	
n. small		0.020	

children			
		(0.019)	
n. employed		0.181***	
		(0.040)	
n. females		−0.078***	
		(0.022)	
Constant	3.668***	−4.077***	4.693***
	(0.970)	(0.176)	(0.274)
Observations	3669	3669	2799
rho	0.12		
sigma	1.29		
Mills'lambda	0.15		

This regression includes controls for: occupation.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4: OLS Wage regressions for male and female subsamples, with and without sample selection equation (dependent variable: log hourly earning 2004)

	Male sub-sample			Female sub-sample		
	Heckman with Probit eq.		Regression with no selection	Heckman with Probit eq.		Regression with no selection
	wage	occ[19]	wage	l_sal_h	occ	wage
age	0.071*	0.210***	0.073**	0.182***	0.151***	0.181***
	(0.038)	(0.012)	(0.035)	(0.057)	(0.014)	(0.062)
age^2	-0.001*	-0.002***	-0.001**	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
years of education	0.025***	-0.019***	0.026***	0.037***	0.029***	0.044***
	(0.009)	(0.005)	(0.008)	(0.014)	(0.007)	(0.015)
tenure	0.019		0.020*	-0.024		-0.017
	(0.012)		(0.012)	(0.017)		(0.016)
tenure^2	-0.000		-0.000*	0.000		0.000
	(0.000)		(0.000)	(0.000)		(0.000)
union	0.308***		0.300***	0.516***		0.540***
	(0.115)		(0.115)	(0.146)		(0.143)
urban	0.286***		0.289***	0.150		0.133
	(0.076)		(0.076)	(0.117)		(0.113)
Self-emp A	0.220**		0.207**	-0.043		-0.085
	(0.098)		(0.099)	(0.149)		(0.143)
Self-emp NA	-0.132		-0.128	-0.348*		-0.381*
	(0.166)		(0.161)	(0.201)		(0.194)
Anjouan	-0.179*	0.301***	-0.181*	0.432**	0.472***	0.471**
	(0.093)	(0.065)	(0.094)	(0.182)	(0.077)	(0.194)
Moheli	-0.437***	0.417***	-0.428**	-0.254	0.361***	-0.233
	(0.148)	(0.074)	(0.145)	(0.195)	(0.091)	(0.205)
married		1.469***			0.387***	
		(0.074)			(0.093)	
divorced		0.631***			0.626***	
		(0.122)			(0.104)	
n. children		-0.049***			0.014	
		(0.016)			(0.020)	
n. small children		0.046**			-0.029	
		(0.022)			(0.022)	
n. employed		0.236***			0.206***	
		(0.041)			(0.043)	
n. females		-0.118***			-0.110***	
		(0.024)			(0.022)	
Constant				0.603	-4.230***	
				(1.409)	(0.254)	
Observations	2169	2169	1801	1500	1500	981
rho	-0.03			0.64		
sigma	1.26			1.50		
Mills lambda[20]	-0.03			0.97		
OLS-	-0.03		-0.018	0.97		1.110***

[19] Participation equation; the dependent variable **occ** is 1 if the person has currently a job, 0 otherwise
[20] Selection term

estimated Mills lambda[21]						
			(0.135)			(0.409)

This regression includes controls for: occupation.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: Pooled quantile regression (dependent variable: log of hourly wages)

Percent.	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**
Female DV	-0.55***	-0.52***	-0.44***	-0.43***	-0.39***	-0.38***	-0.43***	-0.45***	-0.44***
	(0.08)	(0.07)	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)	(0.13)
age	0.09***	0.07***	0.05***	0.04***	0.02	0.01	0.01	-0.01	-0.00
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)
age^2	-0.00***	-0.00***	-0.00***	-0.00***	-0.00*	-0.00	-0.00	-0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
years of education	0.01	0.02**	0.02***	0.02***	0.02***	0.03***	0.03***	0.04***	0.04**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
tenure	0.01	0.02	0.01	0.01*	0.02**	0.02**	0.01	0.02*	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
tenure^2	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
union	0.43***	0.27**	0.22**	0.25***	0.28***	0.36***	0.42***	0.37***	0.25
	(0.13)	(0.11)	(0.10)	(0.08)	(0.09)	(0.08)	(0.10)	(0.10)	(0.21)
urbain	0.43***	0.32***	0.28***	0.20***	0.16**	0.16***	0.26***	0.26***	0.13
	(0.08)	(0.07)	(0.07)	(0.05)	(0.06)	(0.05)	(0.07)	(0.06)	(0.14)
g_occ1	0.46**	0.55***	0.61***	0.62***	0.55***	0.35***	0.30**	0.17	0.13
	(0.18)	(0.15)	(0.14)	(0.11)	(0.13)	(0.11)	(0.15)	(0.13)	(0.30)
g_occ2	0.28	0.36**	0.42***	0.48***	0.47***	0.33***	0.26	0.14	0.34
	(0.19)	(0.17)	(0.16)	(0.12)	(0.15)	(0.12)	(0.16)	(0.15)	(0.32)
g_occ3	0.23	0.34**	0.37***	0.45***	0.45***	0.37***	0.50***	0.58***	0.58**
	(0.17)	(0.14)	(0.13)	(0.10)	(0.12)	(0.10)	(0.13)	(0.12)	(0.27)
g_occ4	-0.26*	0.02	0.07	0.11	0.05	-0.06	0.11	0.05	0.08
	(0.15)	(0.13)	(0.12)	(0.10)	(0.11)	(0.09)	(0.13)	(0.12)	(0.27)
g_occ5	0.13	0.18	0.27**	0.29***	0.31***	0.11	0.12	0.01	-0.01
	(0.13)	(0.11)	(0.11)	(0.08)	(0.10)	(0.08)	(0.11)	(0.10)	(0.23)
Self-emp A	-0.05	-0.00	0.11	0.19***	0.22***	0.26***	0.26***	0.20**	0.27
	(0.11)	(0.09)	(0.09)	(0.07)	(0.08)	(0.07)	(0.09)	(0.08)	(0.18)
Self-emp NA	-0.56***	-0.58***	-0.40***	-0.29***	-0.14	-0.09	-0.08	-0.08	-0.06
	(0.12)	(0.11)	(0.10)	(0.08)	(0.10)	(0.08)	(0.11)	(0.10)	(0.23)
Constant	2.50***	3.32***	3.88***	4.26***	4.79***	5.53***	5.75***	6.64***	7.21***
	(0.40)	(0.35)	(0.33)	(0.25)	(0.30)	(0.24)	(0.33)	(0.30)	(0.65)
Obs.	2799	2799	2799	2799	2799	2799	2799	2799	2799

This regression includes controls for: island occupation. * significant at 10%; ** significant at 5%; *** significant at 1%.
g_occ1: Legislators, administrators, managers, professional; g_occ2: Technicians, associate professionals, clerks; g_occ3: Service workers, sales workers; g_occ4: agricultural and fisheries workers; g_occ5: Trade workers.

[21] In order to make a further check , I put the estimated Mill's Lambda into the wage regression.

Table 6: Quantile regression for women (dependent variable: log of hourly wages)

Percent.	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
age	0.11**	0.06*	0.04	0.01	−0.01	−0.00	−0.00	−0.01	0.03
	(0.05)	(0.03)	(0.04)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.06)
age^2	−0.13**	−0.07*	−0.04	−0.02	0. 10	−0.00	−0.01	0.00	−0.04
	(0.06)	(0.04)	(0.05)	(0.03)	(0.02)	(0.02)	(0.03)	(0.04)	(0.07)
years of education	−0.02	−0.00	0.01	0.02	0.03***	0.02**	0.04***	0.05***	0.05*
	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
tenure	−0.06**	−0.04**	−0.04*	−0.02	0.00	0.01	0.01	0.02	0.01
	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
$tenure^2$	0.13**	0.08**	0.09*	0.05	0.01	0.01	0.01	−0.01	0.02
	(0.05)	(0.04)	(0.05)	(0.04)	(0.02)	(0.02)	(0.03)	(0.04)	(0.07)
union	0.89***	0.64***	0.50*	0.59***	0.50***	0.51***	0.55***	0.36	−0.28
	(0.26)	(0.21)	(0.26)	(0.20)	(0.12)	(0.13)	(0.15)	(0.24)	(0.48)
urban	0.33*	0.17	0.18	0.08	0.02	0.00	−0.04	−0.16	−0.12
	(0.18)	(0.13)	(0.15)	(0.12)	(0.07)	(0.08)	(0.09)	(0.14)	(0.23)
g_occ1	0.81***	0.88***	0.99***	0.79***	0.66***	0.49***	0.12	0.13	0.72
	(0.41)	(0.30)	(0.36)	(0.28)	(0.17)	(0.19)	(0.22)	(0.36)	(0.61)
g_occ2	0.70*	0.72***	0.94***	0.87***	0.86***	0.55***	0.39*	0.50	0.67
	(0.42)	(0.32)	(0.36)	(0.28)	(0.17)	(0.19)	(0.22)	(0.34)	(0.57)
g_occ3	0.54*	0.58***	0.71***	0.89***	0.83***	0.61***	0.48***	0.77***	0.63
	(0.33)	(0.24)	(0.27)	(0.21)	(0.13)	(0.13)	(0.16)	(0.25)	(0.42)
g_occ4	0.15	−0.10	−0.00	−0.02	−0.16	−0.42***	−0.42***	−0.41	−0.57
	(0.37)	(0.26)	(0.29)	(0.23)	(0.13)	(0.14)	(0.16)	(0.27)	(0.46)
g_occ5	−0.09	−0.25	−0.08	0.07	−0.08	−0.32**	−0.27*	−0.14	−0.25
	(0.29)	(0.22)	(0.26)	(0.20)	(0.12)	(0.13)	(0.15)	(0.23)	(0.42)
selfempa	0.20	−0.00	0.07	−0.05	0.02	0.26***	0.35***	0.06	−0.04
	(0.24)	(0.17)	(0.20)	(0.15)	(0.09)	(0.10)	(0.11)	(0.18)	(0.31)
selfempna	−0.83***	−0.56***	−0.34	−0.28	−0.08	0.00	−0.08	−0.31	−0.43
	(0.28)	(0.20)	(0.23)	(0.18)	(0.10)	(0.11)	(0.13)	(0.21)	(0.36)
Constant	1.99***	3.52***	3.97***	4.56***	5.18***	5.49***	5.85***	6.43***	6.56***
	(0.88)	(0.59)	(0.69)	(0.52)	(0.31)	(0.33)	(0.39)	(0.57)	(1.03)
Obs.	990	990	990	990	990	990	990	990	990

This regression includes controls for: island occupation. * significant at 10%; ** significant at 5%; *** significant at 1%.

g_occ1: Legislators, administrators, managers, professional; g_occ2: Technicians, associate professionals, clerks; g_occ3: Service workers, sales workers; g_occ4: agricultural and fisheries workers; g_occ5: trade workers.

Table 7: Quantile regression for men (dependent variable: log of hourly wages)

Percent.	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
age	0.08**	0.06***	0.05***	0.04**	0.02	0.02	0.02	−0.02	−0.02
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
age^2	−0.09**	−0.07***	−0.06***	−0.04**	−0.02	−0.03	−0.03	0.01	0.03
	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.05)
years of education	0.02*	0.02***	0.02***	0.02***	0.03***	0.03***	0.03***	0.04***	0.06***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
tenure	0.03	0.04***	0.04***	0.03***	0.03***	0.03**	0.02	0.02*	−0.00
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
$tenure^2$	−0.08*	−0.09***	−0.08***	−0.06***	−0.05**	−0.04*	−0.02	−0.02	0.02
	(0.05)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.05)
union	0.22	0.24**	0.25***	0.23**	0.26***	0.32***	0.32***	0.34***	0.17
	(0.18)	(0.11)	(0.09)	(0.10)	(0.09)	(0.11)	(0.11)	(0.13)	(0.21)
urban	0.38***	0.25***	0.30***	0.20***	0.22***	0.25***	0.34***	0.38***	0.42***
	(0.12)	(0.08)	(0.07)	(0.07)	(0.06)	(0.08)	(0.08)	(0.09)	(0.16)
g_occ1	0.38	0.44***	0.51***	0.51***	0.46***	0.34**	0.36**	0.24	0.17
	(0.25)	(0.15)	(0.13)	(0.14)	(0.12)	(0.15)	(0.15)	(0.17)	(0.32)
g_occ2	0.20	0.30*	0.19	0.37**	0.30**	0.29*	0.19	0.12	0.26
	(0.29)	(0.17)	(0.15)	(0.16)	(0.14)	(0.17)	(0.18)	(0.20)	(0.35)
g_occ3	0.13	0.06	0.02	0.19	0.19	0.11	0.29*	0.22	0.26
	(0.27)	(0.17)	(0.14)	(0.15)	(0.13)	(0.16)	(0.16)	(0.19)	(0.34)
g_occ4	−0.32	0.11	0.16	0.13	0.13	0.12	0.24*	0.16	0.54*
	(0.23)	(0.14)	(0.12)	(0.13)	(0.11)	(0.14)	(0.14)	(0.16)	(0.29)
g_occ5	0.38*	0.36***	0.42***	0.41***	0.36***	0.26**	0.23*	0.06	0.25
	(0.20)	(0.12)	(0.10)	(0.11)	(0.10)	(0.12)	(0.12)	(0.14)	(0.26)
Self-emp A	−0.09	0.08	0.24***	0.23**	0.31***	0.33***	0.28***	0.24**	0.27
	(0.17)	(0.10)	(0.09)	(0.09)	(0.08)	(0.10)	(0.10)	(0.12)	(0.21)
Self-emp NA	−0.46**	−0.65***	−0.44***	−0.30***	−0.17*	−0.09	−0.07	−0.02	−0.04
	(0.20)	(0.12)	(0.11)	(0.11)	(0.10)	(0.12)	(0.12)	(0.15)	(0.25)
Constant	2.33***	3.14***	3.67***	4.12***	4.92***	5.11***	5.50***	6.81***	7.33***
	(0.61)	(0.39)	(0.33)	(0.35)	(0.30)	(0.38)	(0.38)	(0.44)	(0.78)
Obs.	1809	1809	1809	1809	1809	1809	1809	1809	1809

This regression includes controls for: island occupation. * significant at 10%; ** significant at 5%; *** significant at 1%.

g_occ1: Legislators, administrators, managers, professional; g_occ2: Technicians, associate professionals, clerks; g_occ3: Service workers, sales workers; g_occ4: agricultural and fisheries workers; g_occ5: trade workers.

Table 8: Oaxaca decomposition at the mean

	Predicted wage For male	**Predicted wage for Female**	**Total gap**	Explained part $(\overline{X_m} - \overline{X_f})b$	Unexplained $(b_m - b_f)\overline{X}$
Model without selection	5.80	5.28	-0.52		
* If men are the reference gr.				17%	83%
*If women are the reference gr.				14.4%	85.6%
*Weighted wage structure (Cotton Newmark)				20.2%	79.8%
Model with Selection	5.81	4.20	-1.61		
*If men are the reference gr.				4.5%	95.5%
*If women are the reference gr.				9.8%	90.2%
*Weighted wage structure				11.4%	89%

Table 9: Machado Mata method for the counterfactual distribution

2004	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
(1)Yf[22]	3.467	4.160	4.646	5.011	5.234	5.521	5.927	6.348	7.070
(2)Ym[23]	4.200	4.780	5.157	5.521	5.766	6.032	6.279	6.685	7.370
(3)Observed gap in 2004, Yf-Ym=WG$_0$	-0.732	-0.619	-0.511	-0.511	-0.532	-0.511	-0.352	-0.336	-0.300
(3)XfBm	3.980	4.796	5.143	5.467	5.782	6.113	6.423	6.834	7.470
(5)XmBf	3.384	4.258	4.636	5.001	5.300	5.702	6.151	6.563	7.338
(6) Gap with counterfactual 1 (XfBm -XmBm)	-0.143	-0.070	-0.052	-0.046	-0.035	-0.025	-0.019	0.000	0.019
(7)Counterfactual 1/obs. '04	0.195	0.113	0.102	0.090	0.065	0.049	0.053	-0.001	-0.064
(8) Gap with counterfactual 2 (XmBf- XmBm)	-0.667	-0.586	-0.539	-0.530	-0.506	-0.457	-0.385	-0.317	-0.271
(9) Counterfactual 2/obs. '04	0.910	0.947	1.055	1.037	0.952	0.894	1.093	0.942	0.904

[22] Predicted wage estimated through a OLS regression including only the female dummy variable
[23] Predicted wage estimated through a OLS regression including only the male dummy variable

Figures

Figure 1: Wage density in 2004, by gender

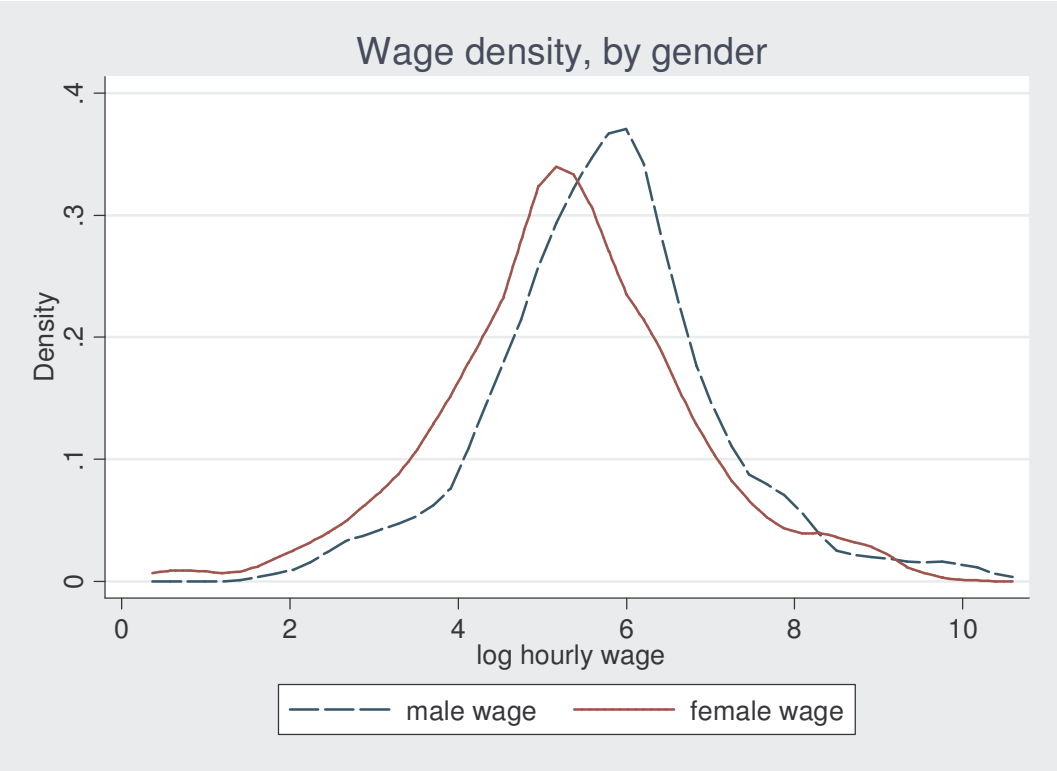
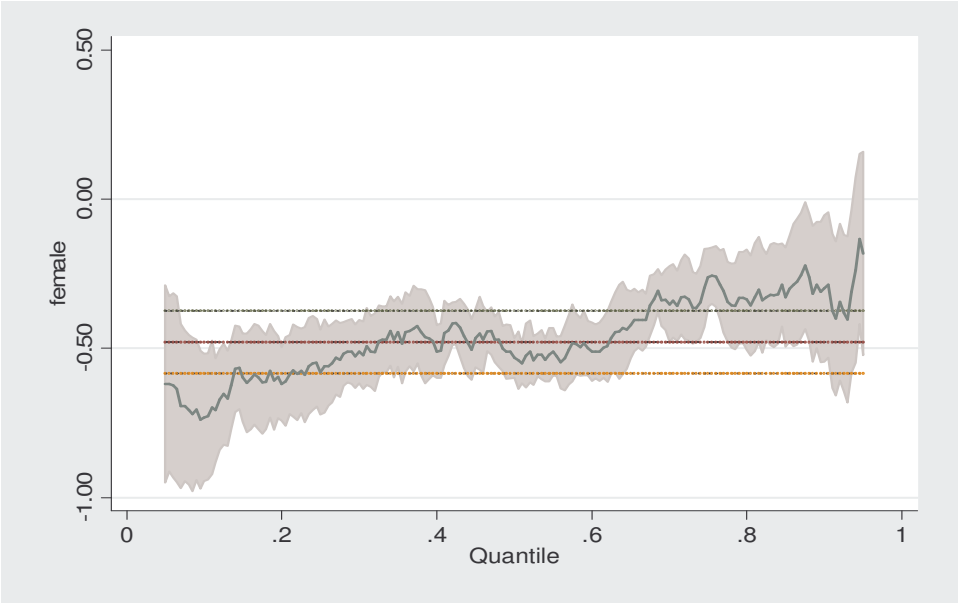
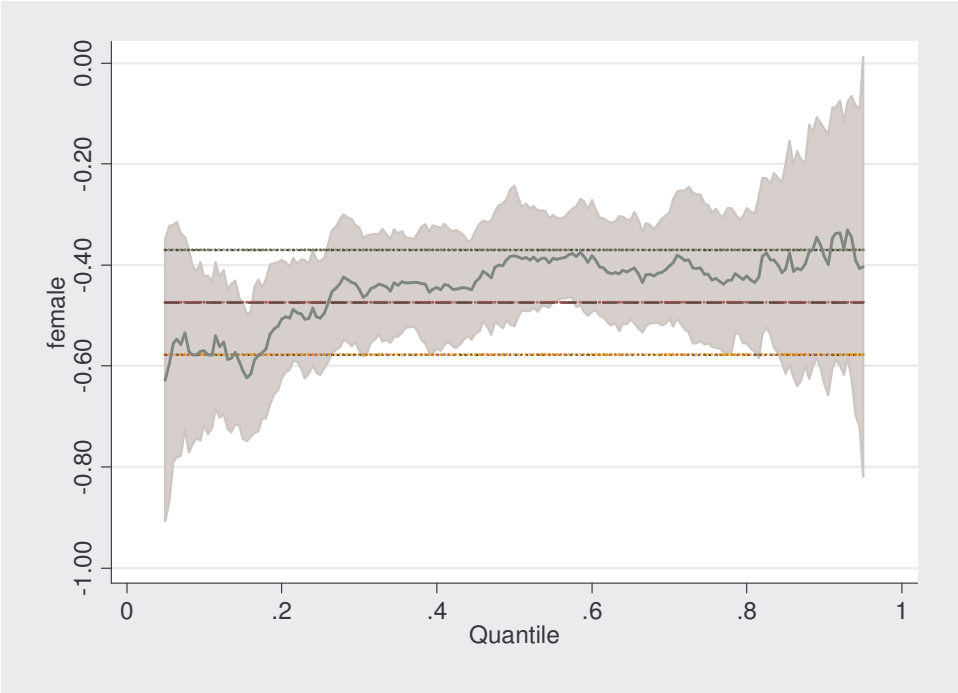


Figure 2: Raw gender wage gap



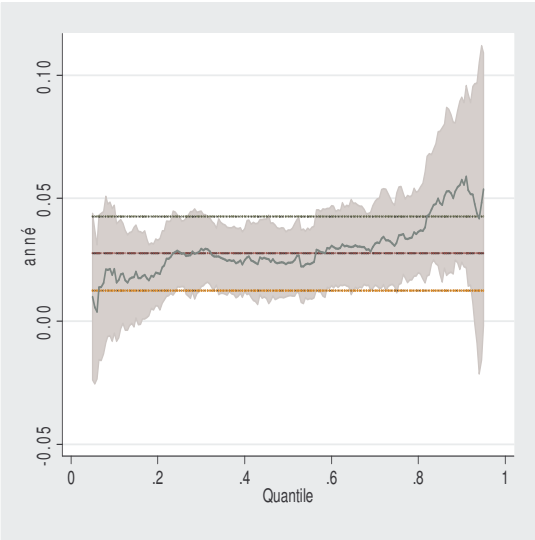
Confidence intervals extend to 95% in either direction. Horizontal lines represent OLS estimates with 95% confidence intervals.

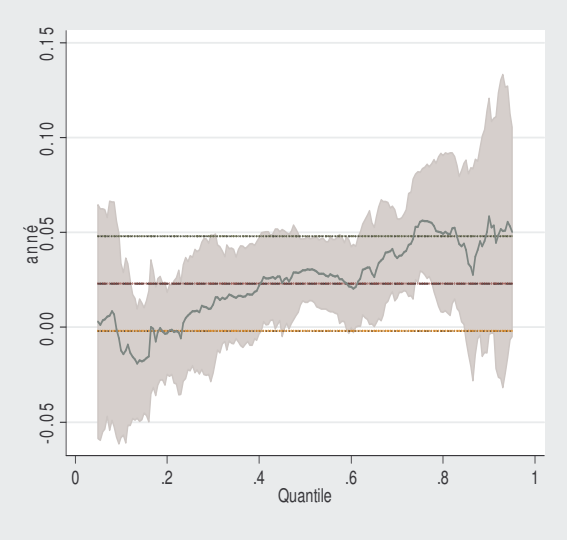
Figure 3: Conditional gender wage gap



Confidence intervals extend to 95% in either direction. Horizontal lines represent OLS estimates with 95% confidence intervals.

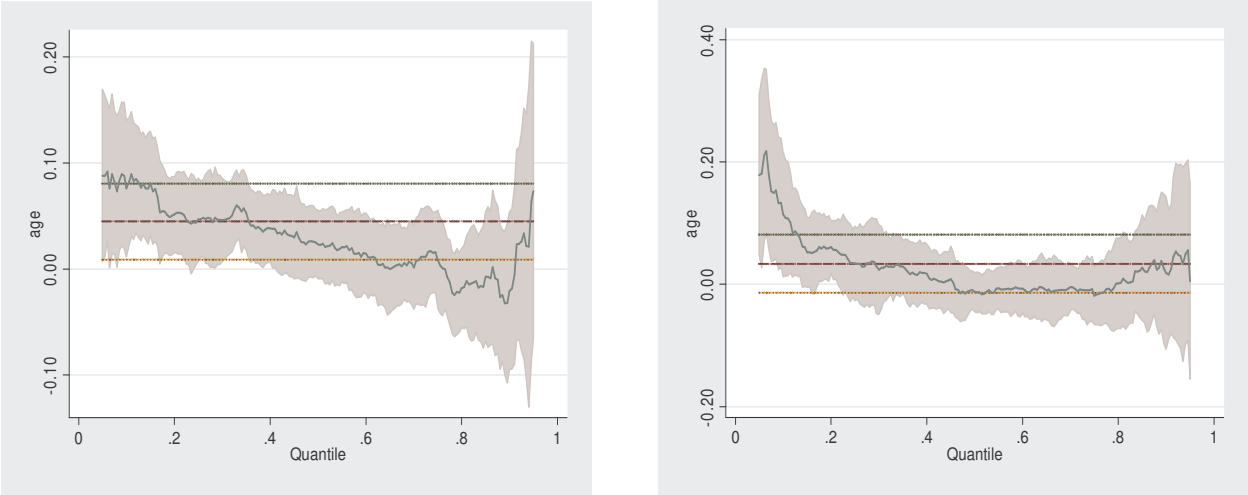
Figure 4: Returns to education for men and women





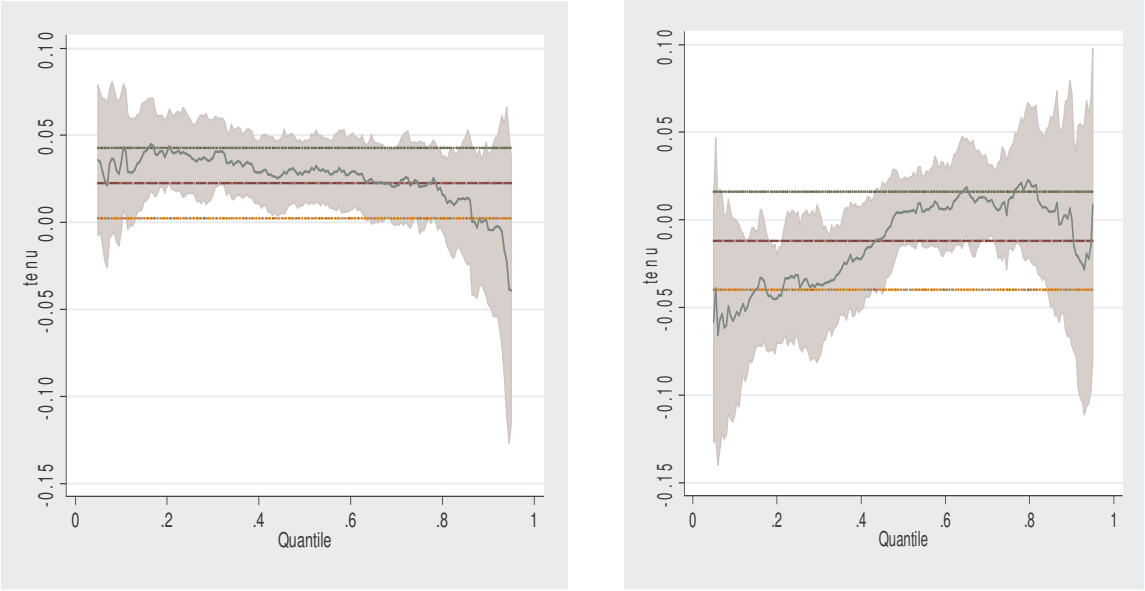
Confidence intervals extend to 95% in either direction. Horizontal lines represent OLS estimates with 95% confidence intervals.

Figure 5: Return to age for men and women



Confidence intervals extend to 95% in either direction. Horizontal lines represent OLS estimates with 95% confidence intervals.

Figure 6: Return to tenure for men and women



Confidence intervals extend to 95% in either direction. Horizontal lines represent OLS estimates with 95% confidence intervals.

Figure 7: Counterfactuals gender wage gap, explained part (men are the reference group)

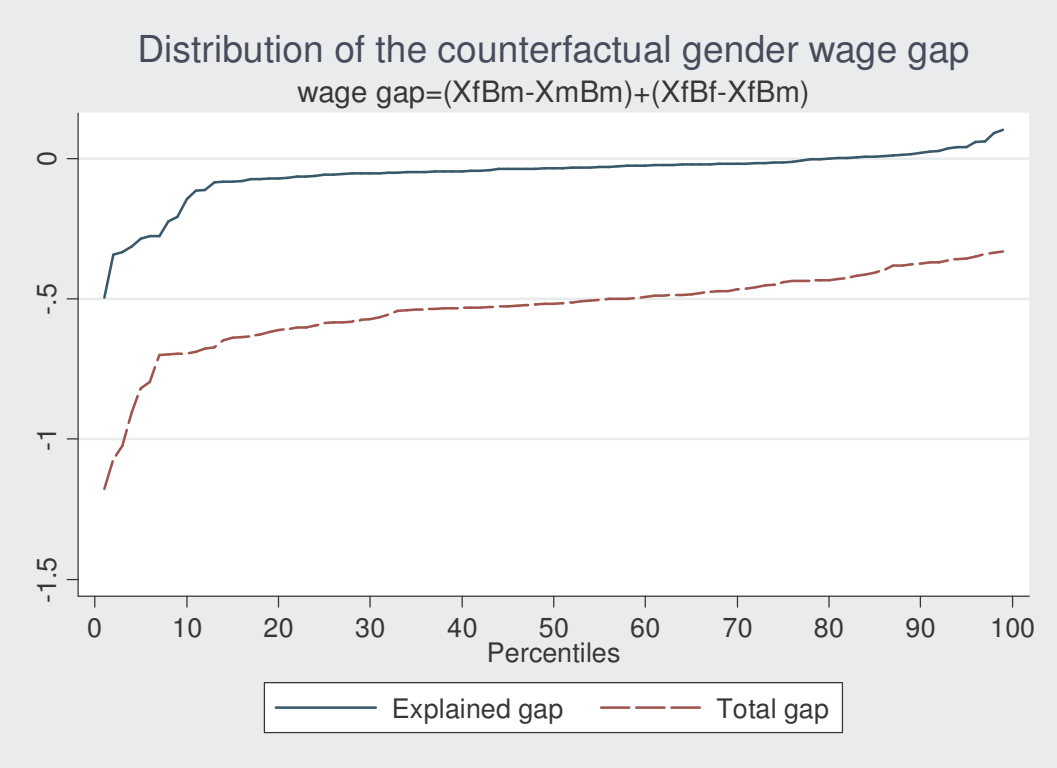
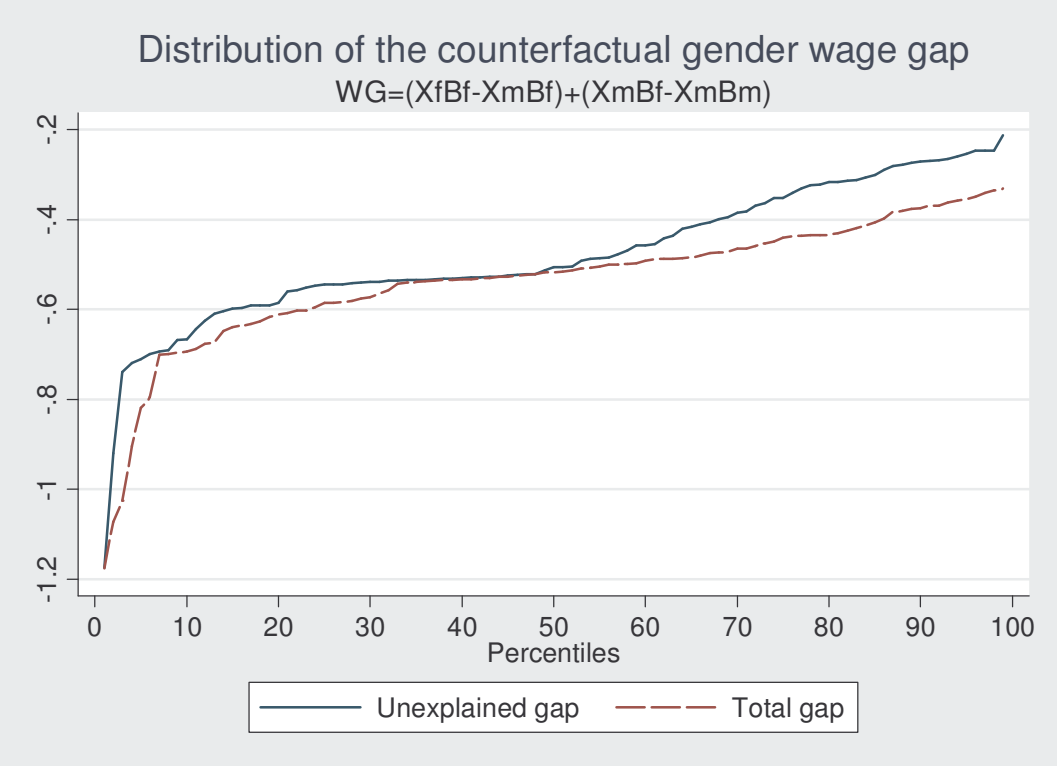
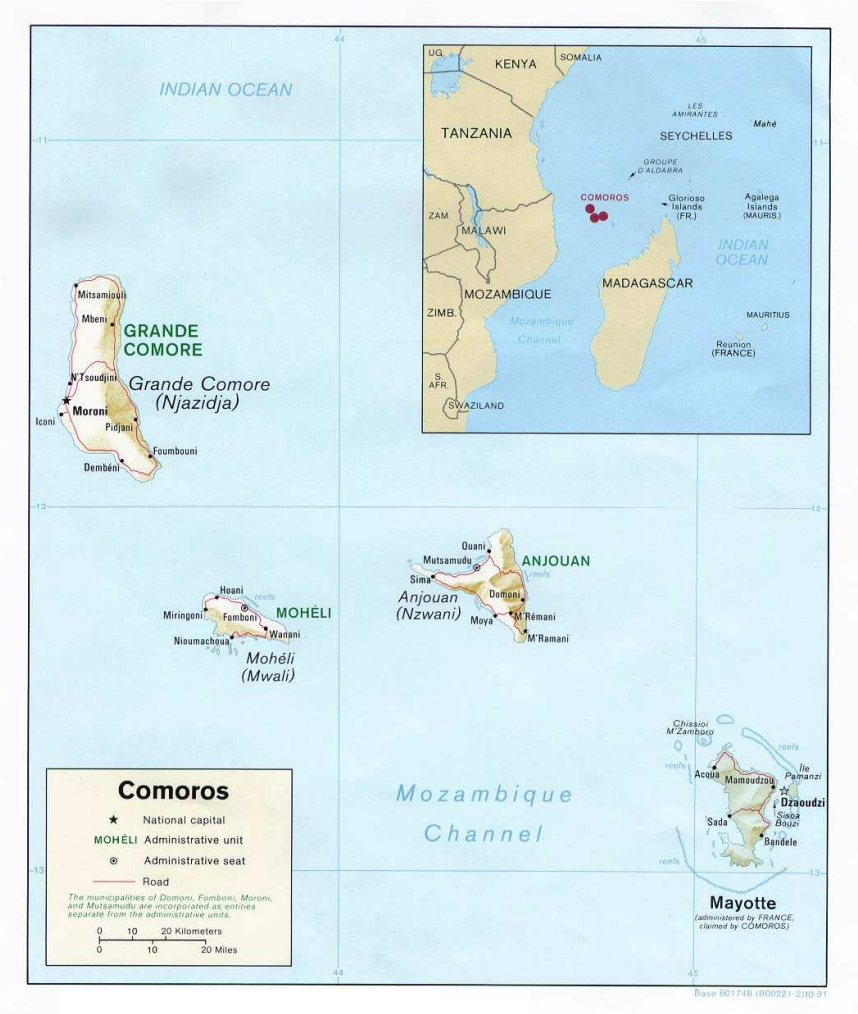


Figure 8: Counterfactuals gender wage gap, unexplained part (women are the reference group)



APPENDIX A

A.1 Comoros map



A .2 Socio-Economic Indicators

Indicators	valeur	année
area	1.812 KM² (without Mayotte)	-
population	575 .639 habitants	2003
Population annual rate of growth	2,1%	1992-2003
Life expectancy at birth	60,6 years	2003
GDP pro capita	450 dollars	
GDP	140.7 billions Comorien franc	2004
Export	11.695 billions Comorien franc	2003
Import	30.311 billions Comorien franc	2003
Population living under the poverty line	44,8%	2004
Net rate of primary schooling	73%	2003
Female net rate of primary schooling enrollement	66,4%	2001
Girls/boys parity in primary school enrollement	0,9	2001
Mortality rate for children less than 5 years old	74 for 1000 babies borne alive	2000
Maternal Mortality rate (delivery)	517 for 100.000 babies borne alive	2000

A .3 Evolution the Human Development Index[24] in Comoros

	1980	1985	1990	1995	2000	2003
HDI	0,480	0,498	0,504	0,517	0,533	0,547
World ranking	136	137	138	138	134	132

Source : Human Development Report 2005

[24] A composite index measuring average achievement in three basic dimensions of human development—a long and healthy life, knowledge and a decent standard of living.

A.4 Definition of the variables used in the regression

Variables	Definition
married	Married
divorced	divorced, widow, separated
single	Never married[25]
n. children	number of children aged between 6 and 14
n. small children	number of children, small than 6 years old
n. employed	Number of people working in the family
n. females	Number of women in working age
education	Years of schooling
tenure	Tenure
union	Being affiliated to a union
urban	Working in urban environment
rural	Working in rural environment[26]
Self-emp A	Self-employed in the agrarian sector
Self-emp NA	Self-employed in the non-agrarian sector
Grande Comore	Dummy variable for the first island[27]
Anjouan	Dummy variable for the second island
Moheli	Dummy variable for the third island

A.5 Occupational categories

	Occupational group	
Original group	**Description**	**Merged group**
1	Legislators Administrator Managers	Occ_1
2	Professional	
3	Technicians Associate professionals	Occ_2
4	Clerks	
5	Service workers Sales workers	Occ_3
6	Agricultural workers Fisheries workers	Occ-4
7	Trade workers	Occ_5
8	Plant & Machine Operators Assemblers	Occ_6
9	Elementary occupations	

[25] Reference category for civil status
[26] Reference category for working environment
[27] Reference category for location

APPENDIX B - Statistics for the total sample of people in working age (15-65 years old)

B.1 Labour force participation

Distribution by gender within working status

	male	female
employed	61.40	38.60
unemployed	33.57	66.43
inactive	41.24	58.76
Total	48.17	51.83

Distribution by working status within gender

	male	female	Total
employed	46.61	27.23	36.56
unemployed	4.01	7.38	5.76
inactive	49.38	65.39	57.68

Composition of the sample of employed people answering to the wage question

	employed	
Wage	**male**	**female**
NO	18.59	29.64
YES	81.41	70.36

B.2 Labour force participation by family status

	male				female			
status	**employed**	**unemployed**	**inactive**	**Total**	**employed**	**unemployed**	**inactive**	**Total**
single	16.90	72.87	77.57	49.10	11.87	26.34	46.28	35.44
married	79.67	21.28	19.23	47.48	69.77	60.48	43.77	52.08
divorced.	3.43	5.85	3.20	3.41	18.35	13.17	9.95	12.48
Total	2,184	188	2,314	4,686	1,373	372	3,297	5,042

Age class, by occupational status and gender

	male				female			
age	**employed**	**unemployed**	**inactive**	**Total**	**employed**	**unemployed**	**inactive**	**Total**
15-19	3.34	26.06	41.01	22.86	4.59	18.28	28.42	21.18
20-29	17.90	43.09	34.23	26.97	23.09	37.10	33.00	30.60
30-39	31.68	13.30	10.20	20.34	30.59	19.89	17.02	20.92
40-49	24.31	5.85	4.97	14.02	21.27	9.95	8.61	12.16
50-59	14.79	6.38	4.80	9.52	15.22	11.02	7.89	10.12
60-65	7.97	5.32	4.80	6.30	5.24	3.76	5.07	5.02
Total	2,184	188	2,314	4,686	1,373	372	3,297	5,042

APPENDIX C - Statistics for the sub-sample of employed people

C.1 Age class, by gender and year

Gender distribution within age class

	male	female
15-19	50.29	49.71
20-29	44.9	55.1
30-39	46.64	53.36
40-49	51.72	48.28
50-59	47.13	52.87
60-65	54.2	45.8
Total	48.09	51.91

Distribution by age class within gender

	male	female	Total
15-19	23.33	21.37	22.32
20-29	26.75	30.41	28.65
30-39	19.91	21.1	20.53
40-49	14.38	12.44	13.37
50-59	9.51	9.884	9.704
60-65	6.116	4.789	5.427

C.2 Educational attainments, by gender and year

Distribution by gender within educational attainments

Last class	male	female
Not finished primary	43.7	56.3
Primary	55.86	44.14
Secondary 1st	51.6	48.4
Secondary 2nd	57.91	42.09
University	69.29	30.71
Post University	80.19	19.81
Total	47.4	52.6

Distribution by educational attainments within gender

Last class	male	female	Total
Not finished primary	65.03	75.49	70.53
Primary	14.39	10.25	12.21
Secondary 1st	10.09	8.529	9.269
Secondary 2nd	6.581	4.31	5.386
University	3.116	1.244	2.131
Post University	.7926	.1764	.4685

Education by Working Status

Last class	Wage workers		Selfemp NA		Self-emp A		Not working	
	male	female	male	female	male	female	male	female
Not finished primary	47.77	53.37	62.04	72.46	78.63	92.44	67.28	75.14
Primary	16.18	9.09	19.44	14.75	13.67	5.85	13.56	11.22
Secondary 1st	13.50	13.49	9.57	8.52	5.97	1.14	10.08	8.95
Secondary 2nd	11.72	13.49	6.48	2.95	1.30	0.43	7.02	4.06

University	8.59	9.09	1.23	0.66	0.33		1.69	0.60
Post University	2.23	1.47	1.23	0.66	0.11	0.14	0.38	0.03
Total	896	341	324	305	922	701	2,124	3,351

C.3 Working status, by gender

Distribution by gender within working status

job	male	female
Wage workers	72.46	27.54
Self-emp. NA	52.41	47.59
Self-emp A	55.36	44.64
Total	60.94	39.06

Distribution by working status within gender

job	male	female	Total
Wage workers	42.72	25.33	35.93
Self-emp. NA	16.42	23.26	19.09
Self-emp A	40.86	51.41	44.98

C.4 Occupational groups, by gender and year

Distribution by gender within occupations

	male	female
g_occ1	72.94	27.06
g_occ2	67.9	32.1
g_occ3	45.04	54.96
g_occ4	53.64	46.36
g_occ5	66.04	33.96
g_occ6	70.03	29.97
Total	58.73	41.27

Distribution by occupations within gender

	male	female	Total
g_occ1	10.12	5.346	8.151
g_occ2	5.401	3.634	4.672
g_occ3	6.649	11.55	8.671
g_occ4	48.1	59.16	52.67
g_occ5	17.97	13.15	15.98
g_occ6	11.76	7.162	9.862

C.5 Average years of tenure, by age classes

	male	female	Total
15-19	5.20	4.67	4.94
20-29	6.49	6.42	6.45
30-39	10.57	10.49	10.54
40-49	16.27	17.92	16.87
50-59	23.35	22.49	22.98
60-65	32.27	29.53	31.42
Total	14.68	13.58	14.22

APPENDIX D - Statistics for the final sub-sample of employed[28] people

D.1 Age class, by gender

Gender distribution within age class

	gender	
	male	female
15-19	57.75	42.25
20-29	57.71	42.29
30-39	64.28	35.72
40-49	68.94	31.06
50-59	61.08	38.92
60-65	75.41	24.59
Total	64.15	35.85

Distribution by age class within gender

	gender		
	male	female	Total
15-19	3.038	3.978	3.375
20-29	17.11	22.43	19.02
30-39	32.36	32.18	32.3
40-49	26.22	21.13	24.39
50-59	14.1	16.08	14.81
60-65	7.177	4.189	6.106

D.2 Educational attainments, by gender

Distribution by gender within educational attainments

Last class	male	female
Not finished primary	59.28	40.72
Primary	77.52	22.48
Secondary 1st	73.46	26.54
Secondary 2nd	71.3	28.7
University	68.8	31.2
Post University	75.92	24.08
Total	64.15	35.85

Distribution by educational attainments within gender

Last class	male	female	Total
Not finished primary	61.96	76.18	67.06
Primary	15.14	7.858	12.53
Secondary 1st	10.29	6.653	8.986
Secondary 2nd	6.73	4.849	6.056
University	4.577	3.715	4.268
Post University	1.311	.7442	1.108

Educational attainments within working status and gender

Last class	Wage workers		Selfemp NA		Self-emp A	
	female	male	female	male	female	male

[28] I refer to the sub-sample of my wage regressions

Not finished primary	47.90	53.54	62.67	74.52	77.81	91.95
Primary	15.68	9.43	18.84	13.13	14.50	6.10
Secondary 1st	13.95	12.46	10.27	8.49	5.77	1.22
Secondary 2nd	11.85	12.79	5.48	2.70	1.48	0.73
University	8.52	10.44	1.37	0.39	0.30	
Post University	2.10	1.35	1.37	0.77	0.15	

D.3 Working status, by gender

Distribution by gender within working status

	male	female
Wage workers	73.47	26.53
Self-emp. NA	54.36	45.64
Self-emp A	59.76	40.24
Total	64.15	35.85

Distribution within working status by gender

	gender		
job	**male**	**female**	**Total**
Wage workers	46.26	29.9	40.4
Self-emp. NA	17.92	26.93	21.15
Self-emp A	35.82	43.17	38.46

D.4 Occupational groups, by gender and year

Distribution by gender within occupations

	gender	
g_occ	**male**	**female**
g_occ1	74.1	25.9
g_occ2	69.14	30.86
g_occ3	48.1	51.9
g_occ4	60.78	39.22
g_occ5	67.84	32.16
g_occ6	73.69	26.31
Total	64.15	35.85

Distribution by occupations within gender

	gender		
g_occ	**male**	**female**	**Total**
g_occ1	11.43	7.15	9.895
g_occ2	6.11	4.881	5.669
g_occ3	7.362	14.22	9.82
g_occ4	41.42	47.83	43.72
g_occ5	21.01	17.83	19.87
g_occ6	12.67	8.096	11.03

D. 5 Average years of tenure, by age classes

	male	female	Total
15-19	4.53	4.38	4.47
20-29	6.17	5.97	6.09

30–39	10.41	9.48	10.08
40–49	16.23	16.6	16.35
50–59	22.44	22.09	22.31
60–65	31.90	25.08	30.22
Total	14.27	12.67	13.70

Chapter 2

The gender wage gap in Poland (1994-2004)

Introduction

The reforms associated to the transition from a centrally planned economy to a free market economy in the '80s and '90s had a dramatic impact on the lives of male and female workers in those countries. The labour market is in fact the conduit through which reform policies impact on a country's standard of living and it is the market through which many of the rewards of transition are transmitted (Horton, Kambur and Mazumdar, 1991). Labour market earnings are an important component of household income in all transitional countries and therefore they represent a crucial link between household welfare and economic activities.

Overall, the welfare of both men and women was crucially affected by the pace of recovering and developing of transition economies. While it is well known that after 1998 Poland finally achieved macroeconomic stabilization and overtook its pre-transition level of output, it is not clear whether the economic situation of Polish women improved along with the positive economic performance of the country. All the changes brought by transition had important consequences for the gender division of labor within the household and on gender equality in economic opportunities. Loss of stability in the labour market, unemployment, reduction of social services enabling them to reconcile their family and employment responsibilities were among the main costs faced by women during the transition.

The "Women in the labour market" report (Karat Coalition 2003) shows that despite some countries[29] of the CEE and CIS area experienced quite a significant economic growth, no drastic improvements have been noted and in some countries the situation has even deteriorated. This is also a very difficult situation to monitor because some of the aspects of the situation of women in the labour market do not appear in statistic and macroeconomic analysis. A common phenomenon is hidden discrimination, as a result of the economic processes taking places in these countries. Overall, less skilled women are experiencing an increased marginalization, increasing unemployment and falling into poverty.

Poland is among the EU member states with the smallest annual average earnings including benefits and allowances, although it is an economy with most extreme disparities as well (EC 2005). Poland has also adopted the equal pay directive which prohibits pay discrimination on the basis of gender, in the context of the accession to the EU in 2004 (Zielinska 2005). Nonetheless, the issue of gender pay gap is not visible on the political, social or economic agenda. As reported by CSO (2003), the Polish labour market is characterised by 20,6% female unemployment rate and the

[29] The report focuses on Bulgaria, Georgia, Macedonia, Poland, Serbia & Montenegro and Slovenia.

main problem faced by women in this country is a visible decrease of their ability to compete for employment and particularly employment which could guarantee them economic independence.

Despite the lack of specific policies aiming at the reduction of the gender pay gap, there has been a clear trend in the narrowing of the distance between incomes of women and men over the transition period. The previous literature (Breinard 2000, Grajek 2003, Adamchik and Bedi 2003, Newell and Socha 2005) attributed this change over the pre-post reform period[30] to improvements in gender specific skills, increase in returns to observed human capital, change in the industrial and occupation affiliation of women and changes in relative wages and employment across industries. However, once controlling for the determinants of the gap in the 90's, the literature finds that most of the gap was due to unequal treatment effects[31] of men and women over differences in endowments. As a result, the situation of women was deteriorating because of discrimination and increasing wage inequality.

The aim of this chapter is to refine the previous findings concerning gender pay gap in Poland and to enquire how the market-based reforms have affected the wages of women relative to men over the period 1994-2004. Differently from most previous papers, I focus on the change of the gap over the second phase of transition, when market elements have become much more important in the functioning of the Polish economy than in the earlier phase.

By using the Labour Force Survey (LSF) data collected by the Polish Central Statistical Office, I investigate which factors influenced the wage gap and the extent to which this gap may be attributed to differences in observable characteristics and to differences in returns to the same characteristics. The emphasis is on relative wage rewards within formal labour markets while informal activity, though potentially important in some transitional economies, is ignored.

The paper is organised as follows. In section one, I describe the subsample of data I selected to perform my analysis and the most important changes in the structure and characteristics of male and female workers over the period in exam.

In section two, I define the specification of my earning equation and the most relevant variables and controls.

In section three, I start performing the econometric analysis by applying OLS and quantile regression methods. I find evidence that the raw gap has decreased over the period in exam. Overall, women have benefited from transition. However, the most relevant point is that once I

[30] The most important decline was in 1989, in the eve of the major markets reforms.
[31] Returns to male and female characteristics differs from the non-discriminatory returns.

control for human capital and other characteristics the gap ends to be much larger in both years and increasing along the distribution.

In section four, I enquire on the determinants of this situation. Quantile regressions run separately for male and female workers in both years reveal that there are important differences in the returns to characteristics by gender, which are remarkably stable in the two years. As expected, many wage determinants have larger proportional impact on wages in the upper part of the distribution.

In section five and six, I apply the Oaxaca-Blinder (1973)and the Juhn, Murphy and Pierce (1991) decompositions techniques in order to examine more in depth the components of the wage gap. The first methodology confirms the preliminary findings concerning adjusted and unadjusted measures of the wage gap. That is, the raw gap is entirely "unexplained" and if we controls for characteristics, it is even wider. The second methodology, which decomposes the change in the wage gap over the two years 1994 and 2004, reveals that increasing returns in education and a better female endowment in education were the two main forces driving down the gender wage gap.

In section seven, I perform the Machado Mata counterfactual analysis which sheds more light on the female-male earning differential at the quantile. Finally, I report my main results and conclusions.

1. Data and statistics

My analysis relies on the data from the Labour Force Survey (LSF) collected quarterly by the Central Statistical Office (Glowny Urzad Statyczny, GUS) for the years 1994 and 2004[32].
The sample coverage is representative and quite large (Appendix A and B). It amounts to 53,785 individuals in 1994 and 47,232. For the purpose of the paper, I use the sub-sample of the wage employed aged between 15 and 65 who were reported as working full time for pay in the reference week. This is due to the fact that only full time hired employees answer a question about their net monthly earnings at the main workplace, while self employed individuals, assisting members in the family business and part-time hired employees do not answer this question. In 2004, there is a small subsample of part-time workers declaring a wage. However, for reason of comparability I do not include them in my subsample. Wages are calculated as hourly wage rate, by dividing the net monthly wage[33] by number of hours worked in a month[34]. The measure of wage I use is adjusted

[32] I used data from the forth round, conducted in September 1994 and September 2004.
[33] This measure of earnings in net monthly earnings in an individual's main job. Tax rules for males and females are the same, thus using net earnings should not affect a comparison across gender.
[34] In the questionnaire, workers answered to the question: "how many hours do you normally work in the week?"

for i) the currency change which happened in 1995[35], because of the hyperinflation in the beginning of the '90s and for ii) the rate of inflation, by using a consumer price index for Poland (1994=100).

Two issues are to take into account. First, there was an important fall in the proportion of employees in the PLFS who give an answer to the wage question. In 1994, 89.56% of the employees report a wage and the percentage drops to 61.46% in 2004. Second, it is to mention the reliability of the reported earnings. Under the centrally planned regime earnings data reported by respondent were verified with that reported by employers. This was no longer the case in the 1990th, thus it is possible that despite guarantees of confidentiality there may be a tendency to under-report wages and top earners may have taken advantage of this situation more than the middle ones. Since there is a greater proportion of males at the upper end of the wage distribution, it is possible that the gender wage gap will be underestimated. However, as long as the temporal tendency to under-report remains unchanged, trends in the wage gap are not likely to e affected (Adamchik and Bedi, 2003).

The size of the final sub-sample and some descriptive statistics are reported in Appendix A. It is very interesting to note how the composition of my sample of workers has changed over the years 1994 2004. The share of male and females workers is quite stable over the decade, the male percentage has increased of 1% while the female one has decreased of the same amount. However, the numerosity of the sample has importantly declined. If I take the total sub-sample of workers, in 2004 I have 12% less of men and 13% less of women with respect to 1994. If I take the sub-sample of workers answering to the wage question, in 2004 I have 30% less men and 31% less women. The percentage of women aged 15 -19 years has decreased of about 10%, while the percentage of males of the same age has increased. The same trend has been reported by CSO (2003) which reports that the most visible change during the years 1992 2002 was a marked decrease of economic activity of persons from age groups 18-19 and 20-24, especially women. One of the reasons could be the observed growth of educational aspirations among younger generations. The CSO reports also that also the drop of active women aged between 35 and 44 years, who are the most economically active. Restructuring of the economy to compete with the EU economies had a significant impact on the drop of economic activity among women with a low level of education (Karat Coalition 2003).

In my 2004 sample I register an important fall in the share of workers with a low education level. I do not have anymore workers with elementary education and the share of workers with basic education dropped from 486 to about 10% and from 23% to 7% respectively for male and females workers. Women education is also closely related to their rate of unemployment. Generally speaking, the higher education level attained by a woman, the lower is the risk of

[35] The 1/01/ 1995: 10,000 "old"zloty (PLZ) = 1 "new" zloty (PLN)

unemployment (in 2002, the rate of unemployment of women with tertiary education was 8.3% in comparison with women with primary education, 24.8% (CSO 2003). In 2004, I have 29% more of men and 47% more of women having attained university and post education level, while 18% more of men and 11% less of women with a secondary vocational diploma.

The dataset contains information on individual's age, education, tenure in the main job, occupation (ISCO '88 classification) , sector of occupation (Polish Classification of Activities, national version of NACE[36] rev.1), type of contract (permanent or temporary), ownership of the firm (public or private) and region. As regard the variable for education, individual educational attainments were harmonised following the changed in the educational system happened in 1999 (see Appendix C). Concerning region, the basic geographical unit in Poland is the province (voivodship). In 1999, the existing 49 voivodships have been aggregated by the Polish statistical office into sixteen voivodships which reflect historical, cultural and geographical distinctions within Polish society. For my analysis, I aggregated the existing voivodships in the eight regions (makroregion) in both years following the 1999 administrative change.

2. Specification of the earning equation

In order to allow comparisons over time, my specification of the earnings equation is limited by the variables available for the two years of my analysis.

First, I include conventional human capital characteristics: education, age as a proxy of potential experience and tenure, which reflects year of work experience with the current employer. Some studies include controls on marital status (Adamchik and Bedi, 2003) in order to capture personal characteristics or any premium or penalty that may be associated with family size.

Second, I control for job related characteristics. I introduce a set of seven occupational indicators (reference category: elementary occupations; Appendix C) and a set of nine controls for industries (reference category: mining and manufacturing) I follow the literature treating job characteristics as potential factors explaining the wage gap between men and women rather than a manifestation of employment discrimination.

Third, I control for the type of contract, permanent or temporary (reference category: temporary).

[36] Classification of Economic Activities in the European Community

Forth, several studies report that that wages and firm size are positively correlated (Oi, 1990; Newell and Socha 2005). If men tend to work in larger firms, than a part of the wage gap may be attributed to firm size. In order to control for this effect, I introduce a four firm size variables (reference category: a firm with less than 20 employees).

Fifth, wages in Poland may vary across the public and the private sectors (Newell and Socha, 2005; Adamchick and Bedi 2003). Therefore, I include an indicator for the public sector work. This dummy may also serve as a control for the lower degree of flexibility in the pay structure in this sector (Joshi and Paci 1998).

Sixth, I introduce a set of eight controls for regional labour market conditions (reference category: region n.8; Appendix C).

Finally, the dependent variable is the logarithm of net hourly earnings, as reported in the questionnaire. In order to account for sampling variability[37], I run my earning equation introducing clustering which controls for type of occupation, sector and region.

In the context of the potential problems associated with sample selection bias, standard correction procedures can be adopted (Heckman 1979). The first major problem is the basic requirement for an adequate set of instruments to identify the selection effects and other wage equation parameters. However, the absence of adequate instruments in my dataset (I have very few information at the household level) restrict my ability to address the issue in a satisfactory manner. Thus, the econometric analysis reported in my paper does not deal with the issue of selectivity bias. This choice is followed in the most of the studies on gender wage gap in transition countries (Paci and Reilly 2005).

3. The observed wage gap: unconditional and conditional measures

3.1 Unconditional gap

To begin my analysis, I first run OLS and quantile regressions on my pooled male and female data separately for 1994 and 2004 with no controls, in order to estimate the raw gender gap at different points of the distribution (table 1). By estimating quantile regression, I can estimate the θ^{th} quantile of y_i, the log wage, conditional on covariates, where the θ^{th} quantile of the distribution of y_i given X_i is:

[37] Units within clusters are similar to each other in terms of the variable of interest.

Eq 9: $Q_\theta(y_{i|} \mid X_i) = X_i\beta(\theta)$,

where there is only one covariate X_i , the female dummy variable.

[Insert table 1]

There are two notable findings on the raw mean gender wage gaps. First, the gap is relatively low for the two years, (0.125 and 0.033 at the mean respectively in 1994 and 2004) compared to Blau and Khan's (2003) estimates of mean raw gaps in 21 countries. Therefore, women are paid in average 12.5% less than men in 1994 and 3.3% less in 2004, without imposing any control in the wage equation. As regard 1994, the estimated value is slightly higher than raw measure of the wage gap estimated by Newell and Reilly (2001) for Poland (0.095)

Second, the observed mean gap in 1994 is 9.2% higher than in 2004 while at the median it is only 2% higher. Overall, the gap is lower in 2004 at any point along the distribution except than at the 40th percentile where the gap is the same in same in both year. If we look at the value of the log wage gaps at different quintiles, we see that the fall in the mean gap in 2004 is the result of a decline in the gap in the tails and especially in the higher half of the distribution.

[Insert figure 3 and 4]

In figure 3 and 4 I report the evolution of the log gender earning gap along the earnings distribution and I add confidence band measure at the 95 percent level. In both years the gap follows a quite irregular path and the confidence intervals are overlapping with OLS estimates just in a few points. Moreover, the OLS value is importantly lower than the median value in 2004 while in 1994 the two values are identical. These preliminary findings suggest it is important to assess the gender earnings gap not only at the mean of the samples but also all along the distribution, in order to have a more correct picture.

In 1994 the gap tends to be higher than the mean gap since the 20th percentile, it reaches its highest value at the 60th and 70th percentile and it drops to its lowest value in the last percentile. In 2004, it is higher than its mean value from the 20th percentile but it starts decreasing from its median value, for reaching the same value as OLS estimates at about the 70th percentile.

In both years, the gap falls importantly below the mean value in the lower and the upper tails. With respect to OLS estimates, the difference in the lower tail is more important in 1994. In the upper tail, the drop in the wage gap is more important in 2004. In this year, women at the top of the distribution are paid 9.5% more than men.

3.2 Unconditional gap

In table 4 I report the gender pay gaps based on estimate from a pooled earning equation (table 2 and 3), specified as indicated in the previous section.

The most relevant result is the rise in the adjusted pay gap relative to the unadjusted measure, especially in 2004. Indeed, the estimates at the mean are 3% and 12, 5% higher than the raw gap respectively in 1994 and 2004 while at the median the difference is about 4-5%.

This difference between adjusted and unadjusted gap suggests that women's personal and productive characteristics are higher than men's, so controlling for this reveals a larger gender pay gap. Newell and Reilly (2001) find the same results in the case of Poland as well as Latvia, Russia and Ukraine in a model where they include only human capital controls (education and the proxies for labour force experience). Nestic (2007) finds a 15% in the raw and adjusted pay gap in Croatia in 2003.

[Insert figure5 and 6]

In figure 5 and 6 I plot the adjusted measure of the gender log wage gaps. Now the slopes follow a smooth decreasing path in both years and the gap reaches its highest value in the upper tails. The evident increase in the gap demonstrates a "glass ceiling" effect, in the sense that the gap is higher in the upper half of the distribution compared to the lower half.

If we compare the two adjusted gaps, we see that now the estimated values at the mean are almost the same. Along the distribution, the gap is about 1%-2% higher in 2004 till the median while there is an inversion afterwards. Indeed, the gap is higher in 1994 from the 60^{th} till the 80^{th} percentile but at the top the value is almost the same.

In 1994 the slope of the gap is steeper and decreases at a constant pace almost all along the distribution. Quantile estimates are overlapping with OLS estimates just in a small portion. In 2004, the value of the gap varies very little from the 50^{th} till the 80^{th} percentile, and then it reaches its highest value at the top. OLS estimates correspond to the values estimated in the 40^{th} and the $30^{t/h}$

/40th percentiles respectively in 1994 and 2004. This highlights the danger of relying on mean regression procedures when the data may describe a more complex and varied situation.

The same evidence of glass ceiling effect is found by Ganguli and Terrel (2005a) in the case of Ukraine when analysing the raw gap. Also Newell and Reilly (2001) find a steady rise in the estimated gender pay gap as they move from low paid to high paid jobs in most of the transitional countries studied. The persistence of glass ceiling could be due to women being persistently less productive (compared to men) at the top of the distribution or it could be generated by continuous discrimination (low returns to their characteristics). In the following part of the paper I look more in details at returns to characteristics for both gender along the distribution and I apply decomposition analysis to wages in order to identify what is driving this path of wage gap along the distribution.

4. Quantile regression analysis

In table 5 and 6 I report the output of the quantile regressions estimated from a pooled earning equation in 1994 and 2004. The implementation of the 'index number' decomposition assumes the separation of the sample by gender is statistically justified. Therefore, I test this assumption by performing an earning equation where I interact all the variables with the gender dummy variable in order to run a Chow test. According to the result, I can reject the null hypothesis that the regression coefficients are all equal to zero suggesting that gender differentials in return to productivity characteristics are important and the returns to characteristics for men and women are significantly different. This confirms the need for a further analysis of the gender wage gap.

In table 5-8 I report the output of the quantile regressions estimated at the quantile in 1994 and 2004, separately for male and female[38].

In 1994 education yields positive returns and very significant both for male and female workers. As regards males, there is a constant return of 2% till the median while in the second half of the distribution it goes up to 3%. Regarding females, there is a 3-4% return for each additional year of education till the 50th percentile, and then it grows to 5% till the top.

Age is significant for females only in the 40th percentile, with a 1% return. For males, there is a 1% return in the 40th and 50th percentile, then it increases to 2% and it reaches a 3% value in the

[38] All the evaluations of returns to the dummy variables are made with respect to the reference category, as described in section 2.

top of the distribution. A possible explanation is that the experience gathered under socialism was less relevant in market economy and the erosion of the "socialist" experience value might have offset the expected increase in experience premiums. Cichomsky (2005) finds that during the transition the general human capital attained through modernised school education gains in importance over human capital gained through work experience. This trend is reinforced by the entrance in the labour market of new generations with better education. Therefore, these changes in returns to observed human capital benefited women, who are on average better educated than men (Newell and Reilly 2001).

Returns to tenure are very low (1%), they are significant for women only in the bottom percentile and for men in the central/ upper part of the distribution.

It is interesting to note that working in a public firm has negative returns for both genders. For males, estimated coefficients are significant only in few quantile while for females, they are always significant and their absolute value increases along the distribution, from 5% to 9%. This points to a significant difference between working in public or private sector. In Ukraine, Ganguli and Terrell (2005a) find that the mean gender gap in the public sector is higher and in the private sector is similar to a set of other European countries.

In 2004, age is significant for males with 2%-3% returns along the distribution. As regard women, there is a 2% returns significant till the 70^{th} percentile. With respect to 1994, this means that workers gained working experience over the transition period and this experience tended to be valued at higher rewards.

Returns to education are again higher for women than for men. This is a constant finding of the literature on Poland. Brainerd (2000) found that over the transition women continue to earn higher returns to education in comparison with men, using data from 1986 and 1992. With respect to 1994, both men and women get much higher return especially in the top of the distribution (for 8% and 4% respectively for women and men). This phenomenon of increasing returns to education can be seen in the context of the very rapid structural and technological change of the Polish economy and the concurrent recession (Newell and Socha 2007). Also, this reflects the shift in relative demand for skilled labour that has happened at a great pace in most of the transition countries. Therefore, the marked increase in returns to female education can also suggest that women tend to adapt much faster to the changing conditions in the labour market. Newell and Socha (2005) find that the rise in wage inequality in the period 1998-2002 was associated to rapid rising returns to education for highly qualified workers in highly skilled occupations and falling relative wages for those with only primary education.

Tenure has still very low returns as in 1994, which are stable across the distribution and more significant for men than for women. Looking at sample statistics, it is also interesting to note that the mean value of years of tenure has importantly dropped over the decade.

Working in the public sector is not significant anymore for females, while for male workers it has a positive value decreasing from 10% in the bottom till 4% in the 70th percentile. Regarding males, this confirms the finding that private sector firms pay less at the low end of the wage spectrum and more at the top end. An explanation might be that the public sector is more unionised and unions usually prefer egalitarian wage structures (Newell and Socha 2005 and 2007). Instead, for females I do not find any significant difference.

My analysis shows that there are important differences in the returns to characteristics by gender, which are remarkably stable in the 2 years. As expected, many wage determinants have larger proportional impact on wages in the upper part of the distribution. Newell and Socha (2005) reach very similar results applying quantile regression estimates to Polish data in 2002.

5. Decomposition at the mean: the Oaxaca-Blinder methodology

[Insert table11]

In table 11 I report the results of the Oaxaca decomposition. As I mentioned in section 4, the total gender wage gap in Poland was quite low in 1994 and it has decreased to 3.3% in 2004. In 1994, the decomposition predicts a mean gender wage difference of 9.1%, significant at 5% level. Regardless of the reference group I take, the explained part of the wage gap is negative and not significant. This analysis confirms the preliminary findings regarding conditional and unconditional gaps. That is, once characteristics are taken into account, the unexplained portion of the wage gap becomes more important.

In 2004, the mean prediction of the gender wage difference is 3.3%. In this year, regardless of the reference group we take (men, women, and weighted group) the explained part is negative and significant. Again, if women were paid according to their characteristics with either men's prices their wage would be higher than men's wage. If we further decompose the explained part (table 12), the most relevant characteristics contributing to this part are education, especially for women, and tenure in a lower measure. These estimates confirm the relatively high unexplained

wage differential when differences in endowment are taken into account. Women should earn more than men because they are better endowed, more importantly they have more education than men, on average. As regard the unexplained part, it is estimated to be greater than the overall difference, as expected from the explained part being negative.

These results are consistent with the work of Newell and Reilly (2001), which find that most of the gender difference in pay is attributed to treatment effects over difference in endowment. As they point out, this finding could partially be due to an inappropriate measurement of female labour force experience. It has been shown that this could involve assigning a greater portion of the actual wage difference to the unexplained component. Grajek (2001) performs the same decomposition at the mean for the years 1987-1996 and he finds that the unexplained part is greater than the overall difference in predicted wages. The conclusion found by Plomien (2006) in her review of national studies of gender based wage differential is that the most frequently factor benefiting female wages is the attainment of higher education. In Poland women are, on average, better educated than men and they benefit from higher rewards to their education. However, the overall higher human capital does not fully translate into a stronger position in the labour market in terms of wages.

6. Juhn, Murphy and Pierce decomposition

A very convenient tool for studying the gender pay gap is the technique of Blau and Khan (1997), inspired by Juhn, Murphy and Pierce (1991). This decomposition accounts for effects to the pay gap due to changes in overall wage dispersion beyond Oaxaca's standard decomposition. This is particularly important in the context of transition countries since I expect the rise in overall wage inequalities as a consequence of the transition from the centrally planned to the market economy.

The procedure is based on the coefficients obtained from male wage regressions for the years of interest, implicitly assuming that female returns would be the same in the absence of discrimination. Thus β_{94} and β_{04} are obtained from the regressions for males:

$$\ln w_{it} = X_{it}\beta_{it} + \sigma_t\theta_{it}$$

where θ_{it} is a standardized residuals and σ_t is the standard deviation of wage residuals. The male-female wage gap for year t is:

$$D_t = \overline{\ln w_{mt}} - \overline{\ln w_{ft}} = \Delta X_t\ \beta_t\ + \sigma_t \Delta\theta_t$$

where Δ represents the difference in male-female averages in X_t and θ_t . The change in the gap between 1994 and 1990 can therefore be decomposed:

$$D_{04} - D_{94} = (\Delta X_{04} - \Delta X_{94})\beta_{04} + \Delta X_{94}(\beta_{04} - \beta_{94}) + (\Delta\theta_{04} - \Delta\theta_{94})\sigma_{04} + \Delta\theta_{94}(\sigma_{04} - \sigma_{94})$$

The first term reflects the contribution of changes in characteristics of female relative to males (labor market skills) at given returns and it is known as the "Observed *Xs*" effect. Usually, this effect evolves slowly over time influenced by new cohorts coming to the labour market. Large "Observed *Xs*" effect would suggest asymmetric perturbations to labor market participation of males and females with equal skills.

The second term, known as the "Observed prices" effects, reflects changes in returns to characteristics (represented by male skills returns) given that the distribution of characteristics differs between men and women. In transition to market economy increasing returns to skills are likely to be expected due to the abandoning of the artificially suppressed central wage setting system. Since women tend to be better educated than men, the increase is likely to be beneficial for women.

The third term represents relative movements in position in a distribution of residuals with a given standard deviation (known as the "Gap effect"). It accounts for the change in unexplained gender gap differential corrected for change in unexplained wage dispersion (represented by standard deviation of the male residual distribution). The idea is to disentangle the changing relative performance of women due to changes in relative unobserved skills (managerial abilities, willingness to work overtime, to accept stress, etc…) and discrimination from changes driven by dispersion of the wage distribution. Given that changes in relative wages accounts for the change in gender pay gap due to changing dispersion, they enter the "Gap" effect as well.

The forth term, the "Unobserved price effect", represents how an increase in the standard deviation of residuals affects wages given the different relative position of men and women in the residual distribution. This term is meant to illustrate the effect of rising returns to unobserved skills and rising inequality across industries during transition.. If women acquire less unobserved skills on average or are discriminated (segregated) against, the rise in wage dispersion will work against them. Furthermore, female domination in low paying industries will reinforce the effect (Grajek 2003).

The second and the forth terms represent observed and unobserved components of the effects of changes in wage structure, while the first and third terms represent the effect of changes in observed and unobserved relative characteristics of men and women ("Gender specific effects").

As explained in detail by Blau and Khan (1997), the $\Delta\theta_t\sigma_t$ terms are calculated based on the mean of the residuals for women computed from the male wage equation. The mean of the male residuals from the male wage equation is zero, hence $\Delta\theta_t\sigma_t$ is equal to the negative of the mean female residuals. The term $\Delta\theta_{94}\sigma_{04}$ is trickier, in the sense it involves calculating what the mean 1994 female residuals would be if the standard deviation of residuals were that of 2004 (for men the mean is again zero). Each women is thus assigned first a percentile in the 1990 male residual distribution, based on her 2004 residual, then she is assigned the residual that corresponds to that percentile in the 2004 distribution. The negative of the mean of these is $\Delta\theta_{94}\sigma_{04}$.

[Insert table 13]

The results of this decomposition are shown in table 13. The reduction in the wage gap between 1994 and 2004 amounts to 0.91 log points. The explained part is negative, -0.096 log points, meaning that improvement in observed characteristics and prices over the period caused the fall in the raw gap. Changes in the observed Xs account for - 0.036 log points. This effect can be attributed to rising shares of female workers attaining higher education level. The Observed prices effect account for 0.06 log points fall. From table 6-10 we saw that returns to education have importantly increase over the period 1994-2004 and especially for women in the higher part of the distribution. Before the transition the wage distribution was artificially compressed, while once reforms have taken place rewards for observed skills have started to gradually augment as wage determination decentralized. Those with more skills have benefited disproportionally (Breinard 2000).

However, this gap reducing effect was partly offset by the increase in the unexplained part which is anyway less important than the explained part. Rising returns to unobserved human capital characteristics (0.01%) means that female labour market skills have worsened or that labor market discrimination against women has risen. As suggested by Blau and Khan (1997), it is possible to say that women have been "swimming upstream". This result confirms the finding of Grajek (2003), who finds that after an important upward movement of mean female position in male wage distribution in 1990, the situation stabilized or even reversed because of the discrimination practices and rising of overall wage inequality. Differently, Breinard (2000) found that in the previous 1986-1992 period there was an important improvement in the mean female percentile in the male distribution in Poland and in other transition countries, which strongly contributed to the reduction of the wage gap.

Third, the "Unobserved Prices" effect accounts for 0.004 log points decrease in the gender pay gap. Even if the effect is small, this result is interesting because it marks an inversion in the trend remarked by the previous literature. Grajek (2003) found that the trend of increasing wage dispersion was actively working against the reduction of the pay gap till 1996. Differently, my results show that a narrowing of the distribution of male wage residuals, holding constant the gap in male female unmeasured skills, was happening over the decade.

A further interesting point is that changes in the gender pay gap could also be due to supply and demand shifts that have adversely affected women relatively to men. This is an expected result in transition to market economy. Breinard (2000) found that over the transition in Poland the net supply of women fell while that of men rose. Grajek (2003) and Newell and Socha (2005) finds also evidence of an important drop in employment over the years 1987-1996. As described in section 1, there was an important drop of the numerosity of the sample between 1994 and 2004. However, the fall in the share of workers with elementary and vocational education from 1994 to 2004 was much higher for men (especially for basic education). Therefore, it seems that males workers were more hit from unemployment, but there is also a significant share of women with low education dropping out over the period examined. And this could have also contributed to the fall of the wage gap. Thus, my data confirm Schultz's (1975) hypothesis that the least educated are likely to be the least successful in transition. This important drop of low wage female workers is also consistent with the evidence found by literature in other countries. Hunt (2002), who finds that the withdrawal from employment of low earners, mainly women, explained 40% of the drop in the wage gap in Eastern part of Germany. Ganguli and Terrel (2005a) attribute the fall in wages in the lower part of the distribution for women in Ukraine to a large exodus of less skilled women over the period 1986-2003.

7. Counterfactual analysis: the Machado-Mata decomposition

Following Ganguli and Terrel (2005a), I apply the Machado-Mata decomposition method to create counterfactual densities, using quantile regression and bootstrapping techniques Following their procedure, I create counterfactual densities for each of the two years where women are given men's characteristics (*Xi*) in one scenario, and women are given men's returns to characteristics (β_i) in another to learn the extent to which it is differences in productive characteristics or

differences in returns that explain the gap within each year. In this way, I obtain i) the female log wage density that would arise if women were given men's characteristics but were paid as women, ii) the density that would arise if women retained their own characteristics but were paid as men. I also create counterfactual where women in 2004 are given their characteristics in 1994 to investigate to what extent the change in female's characteristics over time explain the change in the gaps. Then, I repeat the same for men in 2004.

I create the counterfactual distributions with the following steps:

1. I randomly draw 2500 numbers from a standard uniform distribution, U(0,1) as the quantile we will estimate.
2. Using the male and female data for each year and gender, I estimate 2500 quantile regression coefficients $\beta(\theta_i)$ for i=1,…, 2500, for men and women ($\beta^M(\theta)$ and $\beta^F(\theta)$).
3. I generate random samples of the male and female 1994 and 2004 covariates (*Xi)* by making 2500 draws of men and women with replacement from each year.
4. With my *Xi* and β_i generated for men and women in each year, I can compute the predicted counterfactual wages and construct counterfactual gaps.

7.1 Counterfactual gaps in 1994 and 2004

In table 14 I present the observed gender gaps and the two counterfactual gaps for each of the two years 1994 and 2004 and for ten points of the wage distribution. The counterfactual in row n.2 assume that women have men's βs in that year, and the counterfactual in row n. 4 assume that women had men's *X*s in that year.

[Insert table 14]

Counterfactual 1994: If women had been paid at men's returns (counterfactual one, row n.2), the gap would have been decreasing along the distribution and negative till its median value. Therefore, with respect to the observed gap in 1994, it would have been much lower in the first half of the distribution, while in the second half there would have been an "inverse" gap because women's wages would gain higher wages than men.

If women had men's *Xs* (counterfactual two, row n.4) the gap would have been negative and decreasing all along the distribution. With respect to the observed gap, it would have been

importantly larger till the median but smaller in the top two deciles. That means that in 1994 men were less skilled than women in the lower half of the distribution, while in the upper half there was not a significant difference between male and female characteristics. At the top, men had slightly better characteristics than women.

Counterfactual 2004: If women had been paid at men's returns (row n.2), I find that there would have been an "inverse" wage gap increasing all along the distribution. That is, women would have been paid higher wages than men at any point of the distribution. Women would have been paid 2% and 20% more than men respectively at the bottom and at the top.

If women had men's *Xs*, the gap would have been larger than the observed gap all along the distribution, especially at the bottom. That means, women had better skills than men especially in the bottom deciles. Moreover, this counterfactual gap follows a decreasing path and in the upper part of the distribution the difference with respect to the observed gap is less important. Indeed, the counterfactual gap that is wider than the raw one is not a common feature of advanced market economies and this reflects situations where women possess advantages in productive characteristics, mainly education, as it is often the case in the CEE economies (Nestic 2007).

If I compare rows two in the two years, I see that the contribution of the Betas to the counterfactual gap is much more important in 2004. The "inverse" wage gap has in fact become larger over the period. Differently, If I compare rows four there is not a significant difference between the counterfactual in the two years. That means, the distribution of male characteristics is quite stable with respect to female's one over the period.

The size of the counterfactual gaps in row n. 2 and 4 relative to the observed gap (row n.1) can also be interpreted as a term in separate decompositions. The ratios in rows n. 3 represent the importance of differences in men's and women's *X*s in explaining the observed gap while the ratio in rows n. 5 represent the importance of differences in men's and women's β_i with respect to the observed gap. I find that the difference in men's and women's wage structure is overall much more important than the difference in their characteristics, especially at the bottom of the distribution.

Moreover, in both years the gap would have fallen more at the top than at the bottom if women had men's betas. This means discrimination was higher at the top than at the bottom. Looking at characteristics, women have better skills than men especially at the bottom while in the second half of the distribution the difference become less significant.

7.2 Counterfactual gaps between 1994 and 2004

In table n.15 I present the counterfactuals with the following scenarios: i) women in 2004 are given their characteristics in 1994, ii) women in 2004 are given their returns in 1994, iii) men in 2004 are given their characteristics in 1994, iv) men in 2004 are given their returns in 1994. I would like to investigate why the observed gap is higher in the two bottom deciles of the distribution while in the top deciles it decreases importantly, except than in the 80th decile where its value is equal to the value in 1994.

[Insert table15]

If women's X_s had not changed from 1994 ceteris paribus (row n.4), the gap would have been higher than the observed gap in 2004 and decreasing all along the distribution. This implies that women's X_s have been importantly improving over the period and especially in the lower part of the distribution. Overall, in 2004 women have much better characteristics then in 1994. This is consistent with the hypothesis that changes in women's X_s contributed to the fall of the gap especially at the bottom and in a lower measure at the top. If we look at row number 6, we see that the ratio between the counterfactual and the observed gap with respect to the gap in 1994 is much higher at the bottom percentiles.

If women in 2004 had the same *Bs* as in 1994 (row n.7), the gap would have been higher than the observed gap. The difference between the bottom and the top percentile of this counterfactual gap is also relatively small. At the bottom percentile women would have been paid 18% less than men while in the top percentile 14%less than men. With respect to the observed gap in 2004, the impact of applying 1994 betas is different along the distribution. Women at top are the ones who would gain the most. This finding is consistent with a great deal of evidence that the transition process rewards people at the top of the skill distribution but penalizes the less-skilled. With respect to counterfactual one (row n.4), the gap is lower till the 40th percentile. Therefore, the change in characteristics was more important than the changes in the betas till the 40th percentile, will in the upper part of the distribution the difference in the *Bs* is much more important.

If in 2004 men had 1994 *X*s (row n.10) the gap would have been larger than the observed gap in 2004 till the 30th percentile (between 5% and 14%) and in the 70th and the 80th (between 2 and 4%). In the central part of the distribution it would have been very similar. This means that men's characteristics have importantly improved at the bottom and in a lower proportion at the top and in the central part of the distribution. At the top, women would have still been paid more than

men, but much less than in 2004. This result is quite puzzling and requires further research. A possible explanation could be that the best skilled male workers have emigrated abroad.

Finally, if men in 2004 had 1994 *Bs* (row n.13) the gap would have been positive and increasing all along the distribution. Women would have been paid 5% more than men at the bottom of the distribution and 23% more at the top. That means that 1994 male betas were much worse than 2004 betas and over the period in exam there was an important increase in rewards to male characteristics, especially at the top. Therefore, the important increase in male returns to characteristics contributed to maintain the existent wage gap.

In figure 7 and 8 I draw the distribution of the counterfactual gender wage gap in 1994 and 2004. It is interesting to note that while the unexplained component is quite constant across the period, the total gap is lower at any point of the distribution and it becomes positive at the top. This is explained by the important improvement in female characteristics.

Overall, in each of the two years I find that if women had been paid at men's returns the gap would have been much lower than the observed gap or even positive, meaning that the difference in male and female earning differentials would have been lower or women would have been paid more than men.

If women had men's *Xs*, the gap would have been negative in both year and larger than the observed gap. At the bottom, it would have been much larger in both years while the difference becomes less significant at the top. I also find that overall the difference in men's and women's pay structure is far more important than the difference in their characteristics in explaining each of the gaps. Moreover, in both years the gap would have fallen more at the top than at the bottom if women had men's betas. This means discrimination was higher at the top than at the bottom. Looking at characteristics, women have better skills than men especially at the bottom while in the second half of the distribution the difference become less significant.

What explains the fall of the gap all along the distribution from 1994 and 2004 and the persistence of the ceiling is the important improvement in women's productive characteristics, especially at the bottom. Differently, the improvement of women's *Bs* was more important at the top. This confirms the evidence that people at the top of the skill distribution are the most rewarded by the transition process. As regards the contribution of the male characteristics and returns structure, the distribution of male characteristics has not tremendously changed while returns to their characteristics have importantly improved. This may be a possible explanation for the persistence of the gap.

A further possible explanation for the reduction of the wage gap may also be the large drop of low skilled female from the labour force over the period reported by official statistics and in section 1 in this chapter.

Conclusion

In this paper I provide a detailed examination of how gender earning differentials have changed over the period starting from 1994 to a more advanced transition phase in 2004, when market elements have become much more important in the functioning of the Polish economy than in the earlier phase.

My analysis confirms earlier findings that the gross gender pay gap in Poland is relatively small and it has maintained its previous decreasing trend over the period 1994-2004. Overall, the gap is lower in 2004 at any point along the distribution. At the mean, in 2004 it is 9.2% lower than in 1994 while at the median there is a 2% difference. However, once I control for human capital and other characteristics the gap is larger in both years and increasing all along the distribution. The difference between raw and adjusted gap is wider in 2004.

Looking at returns to characteristics in both years along the distribution, my analysis shows that there are important differences in the returns to characteristics by gender which are remarkably stable in the two years. As expected, many wage determinants have larger proportional impact on wages in the upper part of the distribution. Education is the variables showing the highest returns, especially for women, while experience is more important in 2004.

The Oaxaca decomposition technique confirms my most relevant finding: once I control for characteristics, the gap increases importantly. The explained part is negative, meaning that women are better endowed in characteristics than men and the entire gap ends up being unexplained. That means that none of the gap is explained by differences in endowment between male and female workers. If women were remunerated according to their characteristics, their wage would be higher than men's wage. Moreover, the explained part is significant only in 2004 regardless of the reference group I choose.

The Juhn, Murphy and Pierce technique sheds more light on the evolution of the gap over the period 1994-2004. The observed component is the most important factor involving the narrowing of the wage gap over the period. Higher level of education attained by women together with an increase in the returns to education over the period in exam would cause an important reduction in the wage gap. However, this improvement was partly offset by the growth of the

unexplained part which has penalized women relative to men. An interesting finding is the fall in the unobserved price effect even if it is very small, meaning that a narrowing of the distribution of male wage residuals holding constant the gap in male female unmeasured skills was happening over the decade. This also marks an inversion with respect to the previous trend.

These results refine the previous findings of the literature. After a great reduction of the wage gap in the early '90, previous work found that the increase of discrimination against women and overall wage inequality had caused the stagnation of the gap in the second half of the '90, rather than gradual driving out of the gap. My results show that in the beginning of the new centuries factors working against a narrowing of the gap are offsetting the improvements made in the previous phase, which explains why in 2004 the gap shows a very important increase once I control for differences in characteristics. A further interesting consideration concerns the change in the composition of the sample over the period. The withdrawal from the labour market of less skilled and low wages women reported in my sample analysis and confirmed by official statistics is consistent with the observed rising female wages.

The counterfactual analysis reinforces the role of gender discrimination as a main explanation of the wage gap. In each of the two years, if women had been paid at men's returns the gap would have been much lower than the observed gap or even positive, meaning that the difference in male and female earning differentials would have been lower or women would have been paid more than men. Overall, the difference in men's and women's pay structure is far more important than the differences in their characteristics in explaining each of the gaps. Moreover, in both years the gap would have fallen more at the top than at the bottom if women had men's betas. This means discrimination was higher at the top than at the bottom.

What explains the fall of the gap all along the distribution from 1994 and 2004 is the important improvement in women's productive characteristics, especially at the bottom of the distribution A possible explanation could also be the large drop of low skilled female from the labour force over the period reported by official statistics. Differently, the improvement of women's betas was more important at the top. This finding is also consistent with a great deal of evidence that the transition process rewards people at the top of the skill distribution but penalizes the less-skilled. As regards the contribution of the male characteristics and returns structure, the distribution of male characteristics has not tremendously changed while returns to their characteristics have importantly improved. This may be a possible explanation for the persistence of the gap.

TABLES

Table 1: Raw wage gap, at the mean and at the quantile[39]

	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	OLS
1994	-0.049	-0.128	-0.121	-0.118	-0.125	-0.174	-0.174	-0.159	-0.087	-0.125
	(0.002)	(0.002)	(0.009)	(0.007)	(0.006)	(0.009)	(0.015)	(0.004)	(0.010)	(0.027)
2004	-0.008	-0.061	-0.078	-0.118	-0.105	-0.095	-0.041	0.000	0.095	-0.033
	(0.003)	(0.011)	(0.002)	(0.003)	(0.003)	(0.004)	(0.005)	(0.007)	(0.022)	(0.025)

Standard errors are reported in parenthesis

Table 2: Wage regression 1994

Log hourly earning 1994			
	Pooled data	Male subsample	Female Subsample
female DV	-0.155***		
	(0.019)		
age	0.010***	0.016***	0.004
	(0.004)	(0.005)	(0.004)
age^2	-0.000*	-0.000**	0.000
	(0.000)	(0.000)	(0.000)
years of education	0.035***	0.024***	0.044***
	(0.002)	(0.002)	(0.003)
tenure	0.007***	0.006*	0.005**
	(0.002)	(0.003)	(0.002)
$tenure^2$	-0.000*	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Permanent	0.165***	0.175***	0.134***
	(0.016)	(0.021)	(0.025)
public firm	-0.023	-0.025	-0.057***
	(0.019)	(0.019)	(0.012)
	(0.025)	(0.030)	(0.025)
firm size:20-50	0.048***	0.079***	0.024**
	(0.009)	(0.013)	(0.011)
firm size:50-100	0.038***	0.065***	0.022*
	(0.010)	(0.013)	(0.012)
firm size>100	0.095***	0.179***	0.010
	(0.019)	(0.018)	(0.021)
Constant		2.764***	2.797***
		(0.089)	(0.091)
Occupational controls	YES	YES	YES
Sector controls	YES	YES	YES
Regional controls	YES	YES	YES
Observations	15254	8181	7073
R-squared	0.36	0.32	0.42

Standard errors are reported in parenthesis

[39] Coefficient on the gender dummy variable, estimated from a pooled equation with no controls

Table 3: Wage regression 2004

Log hourly earning 2004			
	Pooled data	**Male subsample**	**Female Subsample**
female DV	-0.158***		
	(0.010)		
age	0.022***	0.023***	0.018***
	(0.003)	(0.004)	(0.005)
age^2	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
years of education	0.053***	0.035***	0.069***
	(0.005)	(0.003)	(0.007)
ten	0.007***	0.009***	0.007***
	(0.002)	(0.002)	(0.002)
tenure^2	-0.000**	-0.000**	-0.000
	(0.000)	(0.000)	(0.000)
permanent	0.121***	0.126***	0.102***
	(0.012)	(0.015)	(0.018)
	(0.024)	(0.035)	(0.033)
firm size:20-50	0.043***	0.037**	0.052***
	(0.012)	(0.015)	(0.017)
firm size:50-100	0.046***	0.047***	0.056***
	(0.011)	(0.016)	(0.015)
firm size>100	0.079***	0.154***	0.000
	(0.017)	(0.015)	(0.022)
Constant		2.625***	2.154***
		(0.088)	(0.135)
Occupational controls	YES	YES	YES
Sector controls	YES	YES	YES
Regional controls	YES	YES	YES
Observations	8132	4405	3727
R-squared	0.46	0.43	0.52

Standard errors are reported in parenthesis

Table 4: Adjusted wage gap, at the mean and at the quantile[40]

	0.10	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**	**OLS**
1994	-0,107	-0,127	-0,141	-0,158	-0,169	-0,178	-0,189	-0,186	-0,197	-0.155
	(0,009)	(0,008)	(0,007)	(0,007)	(0,008)	(0,007)	(0,009)	(0,010)	(0,011)	(0.019)
2004	-0,116	-0,140	-0,150	-0,168	-0,172	-0,173	-0,178	-0,171	-0,194	-0.158
	(0,013)	(0,010)	(0,010)	(0,009)	(0,009)	(0,012)	(0,010)	(0,013)	(0,015)	(0.010)

Standard errors are reported in parenthesis

[40] Coefficient on the gender dummy variable, estimated from a pooled earning equation.

Table5: Quantile regression 1994, dependent variable is log wage 1994

Percentile	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
female DV	-0.11***	-0.13***	-0.14***	-0.16***	-0.17***	-0.18***	-0.19***	-0.19***	-0.20***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
age	0.00	0.00	0.01**	0.01***	0.01***	0.01***	0.01**	0.01**	0.02***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
age^2	0.00	-0.00	-0.00	-0.00*	-0.00**	-0.00***	-0.00*	-0.00*	-0.00**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
years of education	0.03***	0.03***	0.03***	0.03***	0.03***	0.04***	0.04***	0.04***	0.04***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
tenure	0.01***	0.01***	0.01***	0.01***	0.01***	0.01**	0.01***	0.01**	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$tenure^2$	-0.00***	-0.00**	-0.00**	-0.00**	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
permanent	0.14***	0.16***	0.17***	0.16***	0.15***	0.16***	0.17***	0.18***	0.22***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
public firm	-0.02	-0.03***	-0.03***	-0.02***	-0.02**	-0.03***	-0.03***	-0.04***	-0.03**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
firm size:20-50	0.04***	0.05***	0.04***	0.04***	0.04***	0.04***	0.04***	0.05***	0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
firm size:50-100	0.03**	0.04***	0.04***	0.04***	0.03***	0.03***	0.03**	0.05***	0.07***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
firm size>100 workers	0.08***	0.10***	0.09***	0.09***	0.09***	0.09***	0.09***	0.11***	0.13***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant	2.73***	2.74***	2.74***	2.79***	2.81***	2.85***	2.94***	2.98***	3.01***
	(0.08)	(0.07)	(0.06)	(0.06)	(0.07)	(0.07)	(0.08)	(0.09)	(0.11)
Observations	15254	15254	15254	15254	15254	15254	15254	15254	15254

This regression includes controls for: occupation, sectors, regions. Standard errors are reported in parenthesis
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6: Quantile regression 2004, dependent variable is log wage 2004

Percentile	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
female DV	-0.12***	-0.14***	-0.15***	-0.17***	-0.17***	-0.17***	-0.18***	-0.17***	-0.19***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
age	0.03***	0.03***	0.02***	0.02***	0.02***	0.02***	0.02***	0.02***	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
age^2	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
years of education	0.03***	0.04***	0.04***	0.05***	0.05***	0.05***	0.06***	0.06***	0.06***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
tenure	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$tenure^2$	-0.00***	-0.00**	-0.00**	-0.00	-0.00	-0.00	-0.00**	-0.00*	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
permanent	0.11***	0.10***	0.11***	0.12***	0.13***	0.13***	0.12***	0.12***	0.11***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
public firm	0.09***	0.08***	0.07***	0.06***	0.05***	0.05***	0.04***	0.04**	0.02
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
firm size:20-50	0.06***	0.04***	0.04***	0.05***	0.04***	0.02	0.03**	0.03*	0.05**
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
firm size:50-100	0.06***	0.05***	0.04***	0.04***	0.04***	0.04***	0.03**	0.03*	0.06***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
firm size>100	0.09***	0.08***	0.07***	0.08***	0.08***	0.08***	0.08***	0.11***	0.15***
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Constant	2.36***	2.36***	2.47***	2.49***	2.50***	2.49***	2.55***	2.54***	2.75***
	(0.09)	(0.07)	(0.07)	(0.07)	(0.07)	(0.09)	(0.08)	(0.10)	(0.12)
Observations	8132	8132	8132	8132	8132	8132	8132	8132	8132

This regression includes controls for: occupation, sectors, and regions. Standard errors are reported in parenthesis
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7: Quantile regression for Polish women, dependent variable is log wage 1994

Percentiles	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
age	0.00	−0.00	0.00	0.01*	0.01	0.01	0.01	0.01	−0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
age^2	0.00	0.00	0.00	−0.00	−0.00	−0.00	−0.00	−0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
years of education	0.03***	0.04***	0.04***	0.04***	0.04***	0.05***	0.05***	0.05***	0.05***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
tenure	0.01***	0.01***	0.01**	0.00*	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$tenure^2$	−0.00**	−0.00**	−0.00	−0.00	−0.00	0.00	0.00	0.00	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
permanent	0.11***	0.12***	0.12***	0.12***	0.13***	0.14***	0.16***	0.16***	0.18***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)
public firm	−0.05***	−0.04***	−0.05***	−0.04***	−0.06***	−0.07***	−0.07***	−0.08***	−0.09***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)
firm size:20–50	0.03**	0.02	0.02*	0.02	0.02	0.02	0.03*	0.02	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
firm size:50–100	0.03*	0.02	0.02	0.02*	0.02	0.01	0.03	0.04**	0.03
	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
firm size>100	0.03**	0.01	0.01	0.00	0.00	0.00	0.02	0.02*	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Constant	2.64***	2.73***	2.75***	2.66***	2.70***	2.74***	2.84***	2.91***	3.17***
	(0.10)	(0.10)	(0.10)	(0.08)	(0.10)	(0.10)	(0.12)	(0.11)	(0.13)
Observations	7073	7073	7073	7073	7073	7073	7073	7073	7073

This regression includes controls for: occupation, sectors, regions. Standard errors are reported in parenthesis
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8: Quintile regression for Polish men, dependent variable is log wage 1994

Percentiles	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
age	0.01	0.01*	0.01	0.01**	0.01**	0.02***	0.02***	0.02***	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
age^2	−0.00	−0.00	−0.00	−0.00	−0.00**	−0.00***	−0.00***	−0.00***	−0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Years of education	0.02***	0.02***	0.02***	0.02***	0.02***	0.03***	0.03***	0.03***	0.03***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
tenure	0.01	0.00	0.01***	0.01**	0.01**	0.00	0.01	0.01*	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$tenure^2$	−0.00	−0.00	−0.00***	−0.00	−0.00	−0.00	−0.00	−0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
permanent	0.17***	0.18***	0.19***	0.17***	0.18***	0.17***	0.16***	0.18***	0.19***
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
public firm	−0.01	−0.02	−0.03***	−0.03***	−0.02	−0.03**	−0.02	−0.03**	−0.04**
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
firm size:20-50	0.05**	0.06***	0.08***	0.09***	0.08***	0.07***	0.08***	0.07***	0.09***
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
firm size:50-100	0.05**	0.06***	0.07***	0.07***	0.06***	0.06***	0.07***	0.07***	0.09***
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
firm size>100	0.15***	0.16***	0.17***	0.19***	0.18***	0.18***	0.18***	0.18***	0.21***
	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)
Constant	2.62***	2.63***	2.74***	2.75***	2.80***	2.74***	2.77***	2.87***	2.86***
	(0.13)	(0.11)	(0.09)	(0.10)	(0.10)	(0.11)	(0.10)	(0.12)	(0.14)
Observations	8181	8181	8181	8181	8181	8181	8181	8181	8181

This regression includes controls for: occupation, sectors, regions. Standard errors are reported in parenthesis
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 9: Quintile regression for Polish women, dependent variable is log wage 2004

Percentiles	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
age	0.02***	0.02***	0.02***	0.02***	0.02***	0.02***	0.02**	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
age^2	−0.00**	−0.00**	−0.00***	−0.00**	−0.00**	−0.00*	−0.00*	−0.00	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
years of education	0.04***	0.05***	0.05***	0.06***	0.06***	0.07***	0.07***	0.08***	0.08***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
tenure	0.01**	0.01***	0.01***	0.00	0.01**	0.00	0.01***	0.01***	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$tenure^2$	−0.00	−0.00	−0.00	0.00	−0.00	−0.00	−0.00*	−0.00**	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
permanent	0.10***	0.08***	0.07***	0.09***	0.09***	0.11***	0.09***	0.07***	0.10***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
public firm	0.02	0.02	0.01	0.02	−0.00	−0.02	−0.01	0.01	0.01
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
firm size:20−50	0.03	0.02	0.04**	0.05***	0.04**	0.04**	0.04**	0.06**	0.08**
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
firm size:50−100	0.04*	0.05**	0.04**	0.03*	0.04*	0.04*	0.04*	0.04*	0.08**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
firm size>100	−0.00	−0.00	−0.00	−0.01	−0.02	−0.02	−0.02	0.02	0.09***
Constant	2.14***	2.26***	2.22***	2.22***	2.26***	2.27***	2.29***	2.52***	2.60***
	(0.14)	(0.14)	(0.11)	(0.11)	(0.12)	(0.12)	(0.13)	(0.14)	(0.19)
Observations	3727	3727	3727	3727	3727	3727	3727	3727	3727

This regression includes controls for: occupation, sectors, regions. Standard errors are reported in parenthesis
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 10: Quintile regression for Polish men, dependent variable is log wage 2004

Percentiles	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
age	0.03***	0.03***	0.02***	0.02***	0.02***	0.02***	0.02***	0.03***	0.02***
	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
age^2	−0.00***	−0.00***	−0.00***	−0.00***	−0.00***	−0.00***	−0.00***	−0.00***	−0.00**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
years of education	0.02***	0.02***	0.03***	0.03***	0.03***	0.04***	0.04***	0.04***	0.04***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
tenure	0.01***	0.01***	0.01***	0.01***	0.01**	0.01***	0.01***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$tenure^2$	−0.00	−0.00**	−0.00	−0.00*	−0.00	−0.00*	−0.00*	−0.00	−0.00*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
permanent	0.12***	0.11***	0.13***	0.13***	0.15***	0.15***	0.13***	0.12***	0.09***
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
public firm	0.10***	0.10***	0.08***	0.08***	0.06***	0.07***	0.04**	0.02	−0.01
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)
firm size:20-50	0.07***	0.04**	0.04*	0.04***	0.03*	0.03	0.02	0.02	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
firm size:50-100	0.05**	0.05***	0.05**	0.04***	0.04**	0.04***	0.03*	0.03	0.05*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
firm size>100	0.18***	0.15***	0.14***	0.13***	0.15***	0.16***	0.17***	0.18***	0.19***
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Constant	2.32***	2.51***	2.64***	2.73***	2.62***	2.65***	2.65***	2.66***	2.94***
	(0.11)	(0.09)	(0.11)	(0.08)	(0.10)	(0.09)	(0.11)	(0.12)	(0.15)
Observations	4405	4405	4405	4405	4405	4405	4405	4405	4405

This regression includes controls for: occupation, sectors, regions. Standard errors are reported in parenthesis
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 11: Oaxaca-Blinder Decomposition

year	**1994**	**204**
Overall Change in the wage gap	0.125	0.033
	(0.027)	(0.032)
Females are the reference group:		
Observed characteristics	−0.039[41]	−0.134
	(0.024)	(0.034)
Observed prices	0.163	0.168
	(0.013)	(0.015)
Males are the reference group:		
Observed characteristics	−0.013[42]	−0.109
	(0.022)	(0.029)
Observed prices	0.137	0.143
	(0.015)	(0.014)
Weighted group as reference group:		
Observed characteristics	0.015[43]	−0.079
	(0.021)	(0.028)
Observed prices	0.109	0.113
	(0.011)	(0.009)

Table 12: Details of the Oaxaca-Blinder decomposition

	Reference group		
Decomposition of the explained part 2004	Female	Male	Weighted
Age	−0,014	−0.017	−0.013
education	−0.088	−0.044	−0.070
tenure	−0.002	−0.012	−0.009
public firm	NS	−0.010	−0.010

[41] Not significant
[42] Not significant
[43] Not significant

Table 13: JMP Decomposition

Decomposition	Men as a benchmark group
	Wages
Total change in the wage gap	-0.91
Explained part	-0.096
Are obtained from due to:	
observed characteristics (1)	-0.036
observed prices (2)	-0.06
Unexplained part	0.006
due to:	
gap effect (3)	0.01
unobserved prices (4)	-0.004
Sum gender-specific: (1)+(3)	0.046
Sum wage structure: (2)+(4)	-0.064

Table 14: Counterfactual gaps 1 (Machado-Mata decomposition)

1994	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
(1) Observed gap in 1994	-0.049	-0.128	-0.121	-0.118	-0.125	-0.174	-0.174	-0.159	-0.087
(2) Gap with counterfactual 1 ($X_fB_m - X_mB_m$)	-0.022	-0.014	-0.010	-0.009	-0.005	0.005	0.009	0.014	0.022
(3)Counterfactual 1/obs. '04	0.460	0.112	0.086	0.076	0.039	-0.026	-0.049	-0.088	-0.250
(4) Gap with counterfactual 2 ($X_mB_f - B_mX_m$)	-0.207	-0.195	-0.185	-0.181	-0.176	-0.173	-0.148	-0.140	-0.119
(5) Counterfactual 2/obs. '04	4.244	1.524	1.527	1.536	1.409	0.989	0.850	0.881	1.365
2004	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
(1)Observed gap in 2004	-0.008	-0.061	-0.078	-0.118	-0.105	-0.095	-0.041	0.000	0.095
(2) Gap with counterfactual 1 ($X_fB_m - X_mB_m$)	0.024	0.044	0.059	0.080	0.098	0.124	0.140	0.161	0.196
(3)Counterfactual 1/obs. '04	-2.893	-0.721	-0.755	-0.676	-0.934	-1.306	-3.435		2.061
(4) Gap with counterfactual 2 ($X_mB_f - B_mX_m$)	-0.222	-0.204	-0.194	-0.183	-0.168	-0.153	-0.144	-0.127	-0.105
(5) Counterfactual 2/obs. '04	26.705	3.370	2.483	1.553	1.590	1.604	3.539		-1.102

Table 15: Counterfactual gaps 2 (Machado-Mata decomposition)

	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
(1) Observed gap in **1994, Yf-Ym**	-0.049	-0.128	-0.121	-0.118	-0.125	-0.174	-0.174	-0.158	-0.087
(2) Observed gap in **2004, Yf-Ym**	-0.008	-0.061	-0.078	-0.118	-0.105	-0.095	-0.041	0.000	0.100
(3) Observed '04/'94	0.170	0.474	0.642	1.000	0.842	0.547	0.234	0.000	-1.150
Counterfactuals for women									
(4) Gap with counterfact. 1 ($X_f^{94}B_f^{04} - X_m^{04}B_m^{04}$)	-0.203	-0.185	-0.176	-0.167	-0.144	-0.121	-0.092	-0.057	-0.039
(5) Counterfact. 1/obs. '94	4.159	1.451	1.453	1.415	1.152	0.691	0.529	0.362	0.446
(6) (5)/(3)	24.453	3.060	2.262	1.415	1.369	1.265	2.258		-0.388
(7) Gap with counterfact. 2 ($X_f^{04}B_f^{94} - X_m^{04}B_m^{04}$)	-0.184	-0.167	-0.162	-0.160	-0.156	-0.152	-0.149	-0.146	-0.144
(8) Counterfact. 2/obs. '94	3.765	1.303	1.337	1.360	1.247	0.872	0.852	0.924	1.656
(9) (8)/(3)	22.134	2.748	2.081	1.360	1.481	1.594	3.639		-1.440
Counterfactuals for men									
(10) Gap with counterfact. 3 ($X_f^{04}B_f^{04} - X_m^{94}B_m^{04}$)	-0.147	-0.136	-0.121	-0.107	-0.100	-0.077	-0.060	-0.044	0.014
(11) Counterfact. 3/obs. '94	3.010	1.064	0.995	0.904	0.800	0.441	0.346	0.276	-0.156
(12) (11)/(3)	17.695	2.243	1.549	0.904	0.950	0.807	1.478		0.135
(13) Gap with counterfact. 4 ($X_f^{04}B_f^{04} - X_m^{04}B_m^{94}$)	0.046	0.054	0.062	0.066	0.072	0.098	0.125	0.170	0.235
(14) Counterfact. 4/obs. '94	-0.936	-0.426	-0.508	-0.561	-0.572	-0.563	-0.715	-1.077	-2.696
(15) (14)/(3)	-5.503	-0.899	-0.790	-0.561	-0.680	-1.030	-3.055		2.344

Figures

Figure 1: Wage density in 1994, by gender

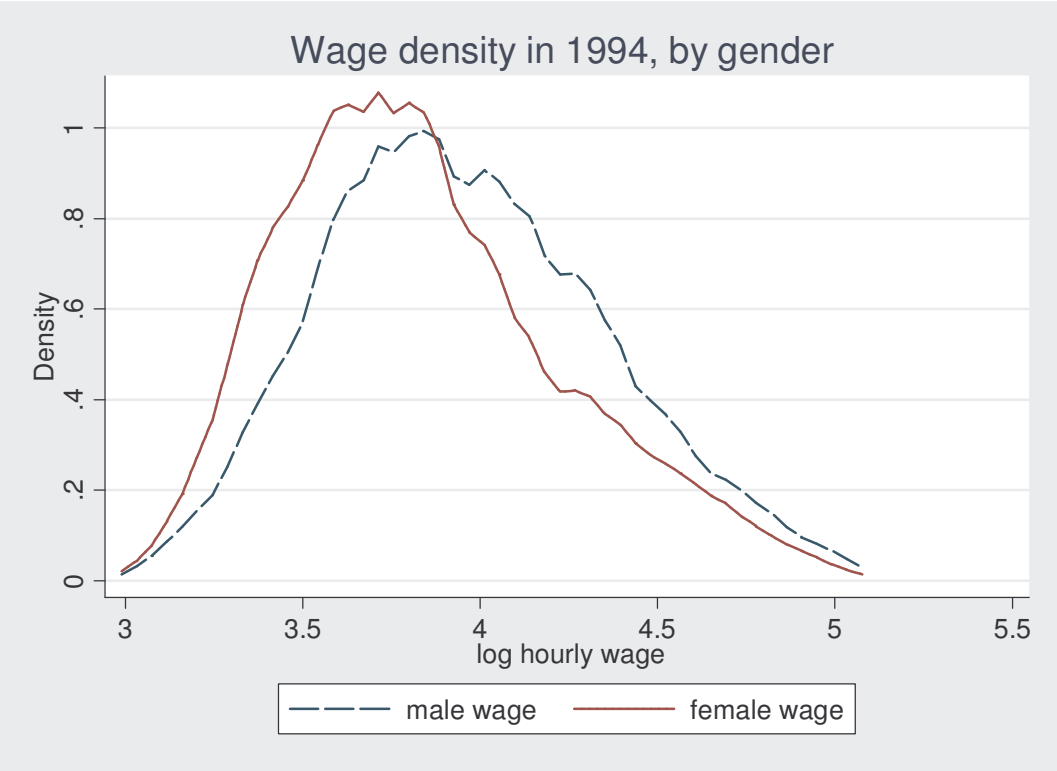


Figure 2: Wage density in 2004, by gender

Figure 3: Raw gap in 1994

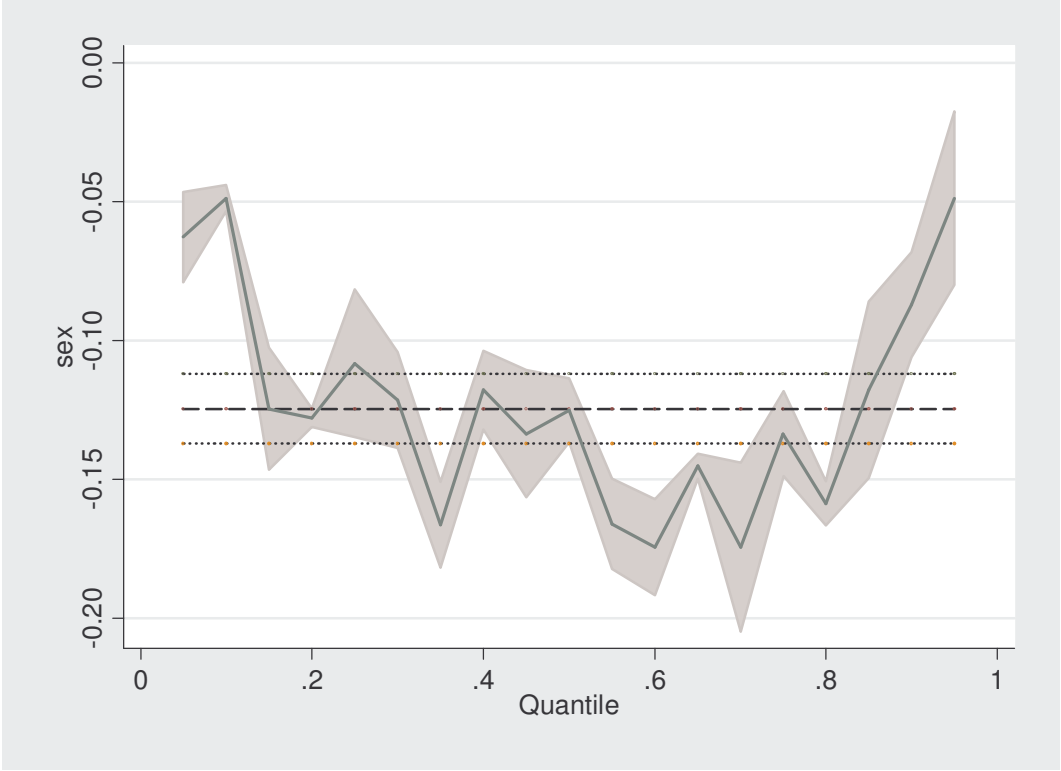


Figure 4: Raw gap in 2004

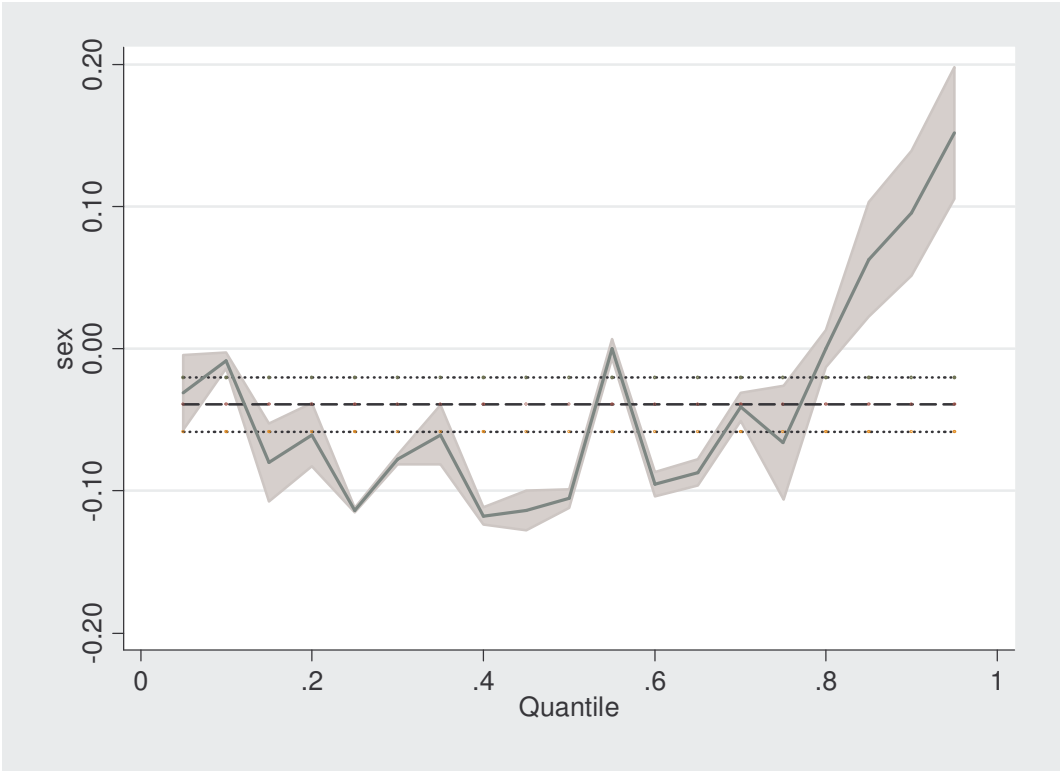


Figure 5: Adjusted gap in 1994

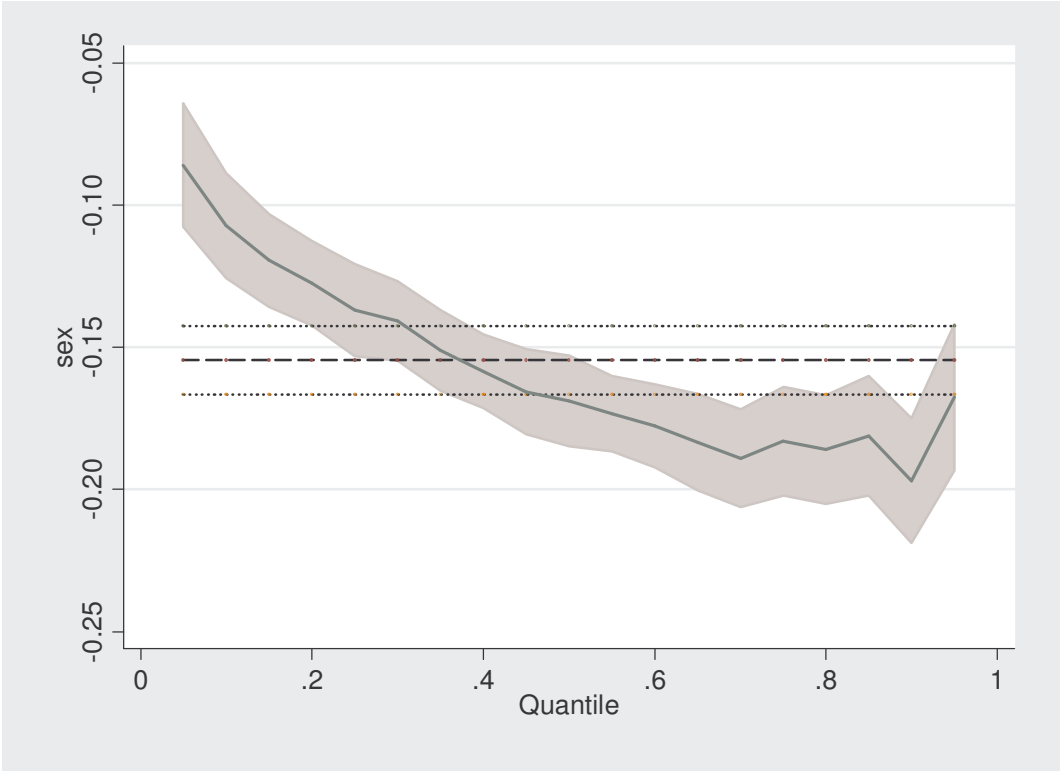


Figure 6: Adjusted gap in 2004

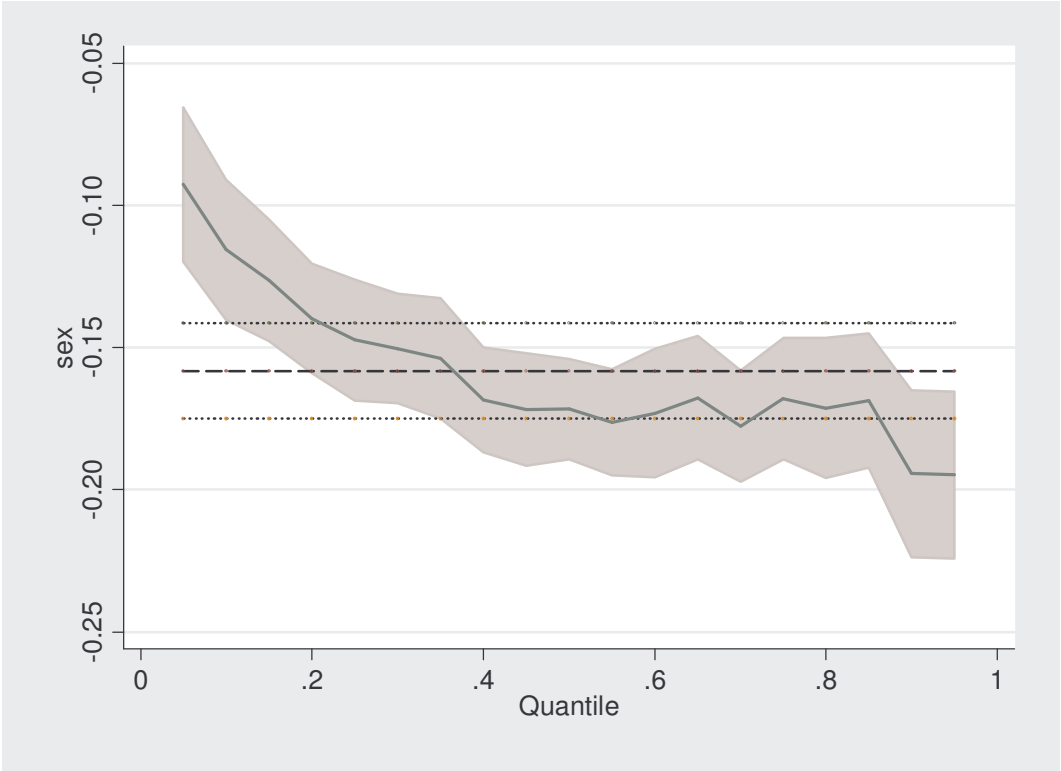
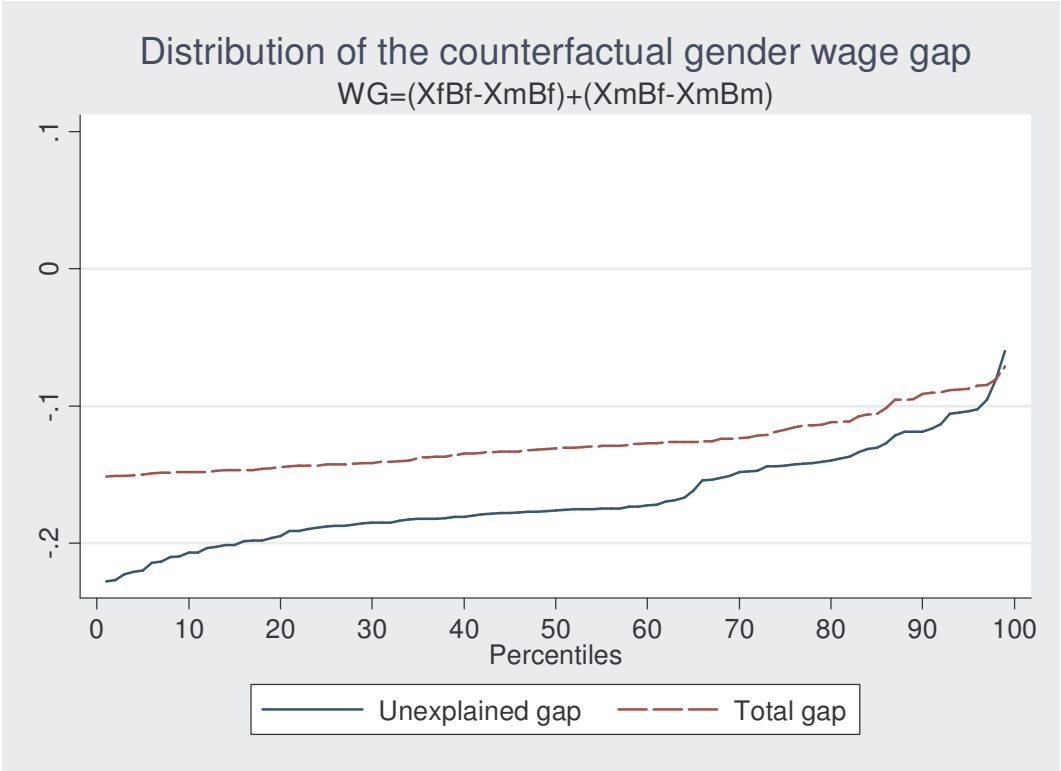
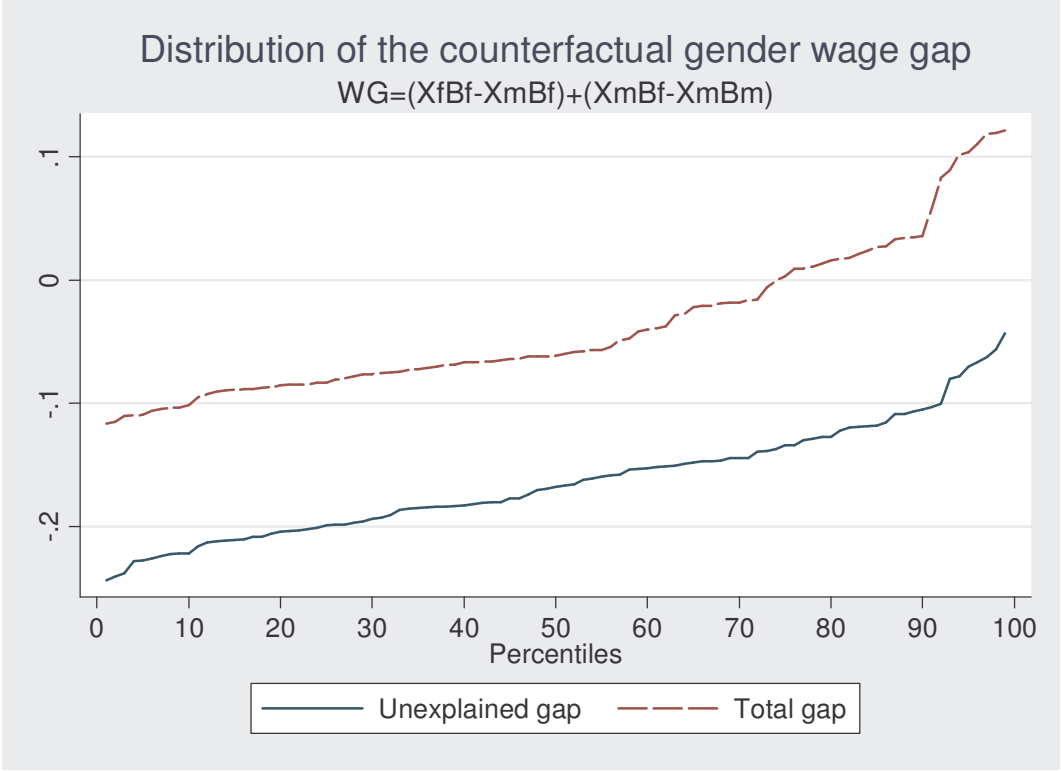


Figure 7: Counterfactuals gender wage gap, 1994



Note: women are the reference group

Figure 8: Counterfactuals gender wage gap, 2004



Note: women are the reference group

APPENDIX A- Statistics for the final sub-sample of dependent employees

A1. Response rate to the wage question (full time employees)

Response rate distribution within year

	Wage reported	
year	NO	YES
1994	10.44	89.56
2004	38.36	61.64
Total	22.63	77.37

Distribution by year within response rate

	Wage reported		
year	NO	YES	Total
1994	26.00	65.21	56.34
2004	74.00	34.79	43.66

A2. Composition of the total sample of employees

Distribution by gender within year

	gender	
year	male	female
1994	53.55	46.45
2004	54.42	45.58
Total	53.93	46.07

Distribution by year within gender

	gender		
year	male	female	Total
1994	55.94	56.80	56.34
2004	44.06	43.20	43.66

A3. Composition of the total sample of workers, answering the wage question

Distribution by gender within year

	gender	
year	male	female
1994	53.63	46.37
2004	54.17	45.83
Total	53.82	46.18

Distribution by year within gender

	gender		
year	male	female	Total
1994	65.00	65.49	65.23
2004	35.00	34.51	34.77

A.4 Age class, by gender and year

Gender distribution within age class

	1994		2004	
age	male	female	male	female
15–19	58.73	41.27	68.97	31.03
20–29	58.95	41.05	58.59	41.41
30–39	52.58	47.42	54.62	45.38
40–49	49.08	50.92	48.24	51.76
50–59	58.17	41.83	56.87	43.13
60–65	75.65	24.35	81.82	18.18
Total	53.63	46.37	54.17	45.83

Distribution by age class within gender

	1994			2004		
age	male	female	Total	male	female	Total
15–19	1.81	1.47	1.65	0.91	0.48	0.71
20–29	23.43	18.87	21.32	25.86	21.60	23.91
30–39	32.91	34.33	33.56	27.92	27.42	27.69
40–49	29.86	35.84	32.63	28.31	35.90	31.79
50–59	10.93	9.09	10.08	15.98	14.33	15.22
60–65	1.06	0.40	0.75	1.02	0.27	0.68

A.5 Educational attainments, by gender and year

Distribution by gender within educational attainments

	1994		2004	
	male	female	male	female
University and post	45.90	54.10	43.44	56.56
Post-secondary	18.51	81.49	33.45	66.55
Secondary general	48.05	51.95		
Secondary vocational	23.68	76.32	71.32	28.68
Basic vocational	70.31	29.69	63.72	36.28
Elementary	56.23	43.77		
Not completed primary	50.00	50.00		
Total	53.63	46.37	54.17	45.83

Distribution by educational attainments within gender

	1994			2004		
	male	female	Total	male	female	Total
University and post	10.48	14.28	12.24	39.75	61.18	49.57
Post-secondary	1.64	8.34	4.75	4.45	10.46	7.21
Secondary general	24.15	30.20	26.96			
Secondary vocational	3.08	11.48	6.98	46.47	22.08	35.29
Basic vocational	45.96	22.45	35.06	9.33	6.28	7.93
Elementary	14.62	13.16	13.94			
Not completed primary	0.07	0.08	0.08			

A.6 Average years of tenure, by gender and year

	1994			2004		
age class	male	female	Total	male	female	Total
15-19	1.27	1.0	1.16	1.09	1.31	1.15
20-29	5.87	4.81	5.43	3.06	3.17	3.10
30-39	15.23	14.12	14.70	7.78	8.18	7.96
40-49	24.26	22.57	23.4	12.14	14.29	13.25
50-59	33.28	28.81	31.42	15.86	17.18	16.43
60-65	38.14	31.86	36.61	19.28	20.33	19.47
Total	17.70	16.60	17.20	9.14	10.58	9.80

A.7 Occupational groups, by gender and year

Distribution by gender within occupations

	1994		2004	
Occupation	male	female	male	female
Legislators, Administrators, Managers, Professionals, Technicians	39.15	60.85	34.84	65.16
Clerks	23.24	76.76	36.38	63.62
Service workers Sales workers	28.64	71.36	36.29	63.71
Agricultural workers Fisheries workers	72.26	27.74	78.57	21.43
Trade workers	79.40	20.60	81.22	18.78
Plant & Machine Operators Assemblers	86.74	13.26	86.78	13.22
Elementary occupations	46.30	53.70	43.96	56.04
Total	53.63	46.37	54.17	45.83

Distribution by occupations within gender
g

	1994			2004		
Occupation	male	female	Total	male	female	Total
Legislators, Administrators, Managers, Professionals, Technicians	24.21	43.53	33.17	20.43	45.16	31.76
Clerks	4.12	15.74	9.51	5.61	11.59	8.35
Service workers Sales workers	4.28	12.33	8.01	8.17	16.96	12.20
Agricultural workers Fisheries workers	1.37	0.61	1.02	1.00	0.32	0.69
Trade workers	39.12	11.73	26.42	33.58	9.18	22.39
Plant & Machine Operators Assemblers	17.20	3.04	10.63	22.79	4.11	14.23
Elementary occupations	9.71	13.02	11.24	8.42	12.69	10.38

A.8 Categories of contract (Permanent/ temporary)

Distribution by gender within categories of contract

	1994		2004	
	Male	female	Male	female
permanent	53.29	46.71	52.42	47.58
temporary	65.36	34.64	60.08	39.92
Total	53.63	46.37	54.17	45.83

Distribution by categories of contract within gender

	1994			2004		
	Male	female	Total	Male	female	Total
permanent	96.54	97.88	97.16	74.62	80.06	77.12
temporary	3.46	2.12	2.84	25.38	19.94	22.88

A.9 Sectors, by gender and year

Distribution by gender within sectors

	1994		2004	
sector	male	female	male	female
agriculture & fishing	76.03	23.97	76.79	23.21
mining & manufacturing	63.62	36.38	67.26	32.74
recycling and construction	85.11	14.89	91.78	8.22
retails & trade, hotels	37.92	62.08	41.52	58.48
transport	67.33	32.67	75.81	24.19
financial intermediation	40.70	59.30	48.79	51.21
public administration	58.85	41.15	48.15	51.85
education and health	20.93	79.07	18.55	81.45
other service & activities	55.54	44.46	52.33	47.67
Total	53.63	46.37	54.17	45.83

Distribution by sector within gender

	1994			2004		
sector	male	female	Total	male	female	Total
agriculture & fishing	4.96	1.81	3.50	4.13	1.48	2.91
mining & manufacturing	40.46	26.76	34.11	37.50	21.57	30.20
recycling and construction	15.30	3.10	9.64	15.21	1.61	8.98
retails , trade & hotels	7.54	14.28	10.67	11.62	19.35	15.16
transport	8.67	4.86	6.90	9.60	3.62	6.86
financial intermediation	3.72	6.26	4.90	5.93	7.35	6.58
public administration	7.60	6.15	6.93	6.49	8.26	7.30
education and health	7.52	32.86	19.27	6.45	33.46	18.83
other service & activities	4.23	3.92	4.08	3.06	3.30	3.17

A.10 Ownership of the firm, by gender and year

Distribution by gender within firm's ownership

	1994		2004	
Firm's ownership	**male**	**female**	**male**	**female**
public	51.27	48.73	41.81	58.19
private	60.74	39.26	62.76	37.24
Total	53.63	46.37	54.17	45.83

Distribution by firm's ownership within gender

	1994			2004		
Firm's ownership	**male**	**female**	**Total**	**male**	**female**	**Total**
public	71.75	78.88	75.06	31.65	52.05	41.00
private	28.25	21.12	24.94	68.35	47.95	59.00

APPENDIX B- Statistics for the total sample of full time dependent employees

B1. Age class, by gender and year

Gender distribution within age class

	1994		2004	
age	male	female	male	female
15-19	58.90	41.10	66.67	33.33
20-29	57.75	42.25	58.38	41.62
30-39	52.52	47.48	55.08	44.92
40-49	49.28	50.72	49.02	50.98
50-59	58.85	41.15	55.92	44.08
60-65	76.51	23.49	75.83	24.17
Total	53.55	46.45	54.42	45.58

Distribution by age class within gender

	1994			2004		
age	male	female	Total	male	female	Total
15-19	1.89	1.52	1.71	0.97	0.58	0.80
20-29	23.11	19.49	21.43	25.90	22.05	24.15
30-39	32.49	33.88	33.14	27.77	27.04	27.44
40-49	29.96	35.56	32.56	27.84	34.56	30.90
50-59	11.30	9.11	10.29	16.24	15.29	15.81
60-65	1.25	0.44	0.87	1.27	0.48	0.91

B2. Educational attainments, by gender and year

Distribution by gender within educational attainments

	1994		2004	
	male	female	male	female
University and post	47.79	52.21	46.06	53.94
Post-secondary	18.47	81.53	33.66	66.34
Secondary vocational	47.58	52.42	72.03	27.97
Secondary general	23.52	76.48		
Basic vocational	69.80	30.20	62.32	37.69
Elementary	56.52	43.48		
Not completed primary	53.85	46.15		
Total	53.55	46.45	54.39	45.61

Distribution by educational attainments within gender
g

	1994			2004		
	male	female	Total	male	female	Total
University and post	11.63	14.65	13.03	45.61	63.71	53.86
Post-secondary	1.61	8.20	4.67	4.81	11.31	7.77
Secondary vocational	23.56	29.93	26.52	41.72	19.32	31.50
Secondary general	3.05	11.43	6.94			
Basic vocational	45.30	22.60	34.76	7.86	5.67	6.86
Elementary	14.78	13.11	14.00			
Not completed primary	0.08	0.08	0.08			

B3. Occupational groups, by gender and year

Distribution by gender within occupations

	1994		2004	
job	male	female	male	female
1	39.89	60.11	39.07	60.93
2	23.18	76.82	34.36	65.64
3	28.59	71.41	36.60	63.40
4	70.48	29.52	78.95	21.05
5	78.93	21.07	82.78	17.22
6	86.67	13.33	85.90	14.10
7	46.76	53.24	46.39	53.61
Total	53.55	46.45	54.42	45.58

Distribution by occupations within gender

	1994			2004		
job	male	female	Total	male	female	Total
1	25.06	43.55	33.65	25.84	48.09	35.98
2	4.05	15.46	9.35	5.61	12.79	8.89
3	4.36	12.56	8.17	7.80	16.12	11.59
4	1.28	0.62	0.97	1.04	0.33	0.72
5	38.40	11.82	26.06	31.32	7.78	20.59
6	16.96	3.01	10.48	20.52	4.02	13.00
7	9.89	12.98	11.33	7.88	10.87	9.24

B4. Average years of tenure, by gender and year

	1994		2004	
age	male	female	male	female
15-19	1.126	1.21	.97	.69
20-29	2.97	3.03	5.79	4.82
30-39	7.73	8.21	15.15	14.06
40-49	12.29	14.17	24.19	22.52
50-59	16.02	17.21	33.32	28.80
60-69	20.78	23.33	38.09	31.68

B5. Categories of contract (Permanent/ temporary)

Distribution by gender within length of contract

	1994		2004	
	male	female	male	female
permanent	53.16	46.84	52.75	47.25
temporary	65.38	34.62	60.20	39.80
Total	53.55	46.45	54.42	45.58

Distribution by length of contract within gender

	1994			2004		
	male	female	Total	male	female	Total
permanent	96.09	97.61	96.79	75.22	80.44	77.60
temporary	3.91	2.39	3.21	24.78	19.56	22.40

B6.Sectors, by gender and year

Distribution by gender within sectors

	1994		2004	
sector	male	female	male	female
agriculture & fishing	75.73	24.27	76.18	23.82
mining & manufacturing	63.02	36.98	67.77	32.23
recycling and construction	85.23	14.77	91.67	8.33
retails , trade & hotels	38.00	62.00	43.50	56.50
transport	67.37	32.63	74.08	25.92
financial intermediation	40.68	59.32	49.05	50.95
public administration	59.01	40.99	45.55	54.45
education and health	21.27	78.73	20.04	79.96
other service & activities	56.59	43.41	48.67	51.33
sector	53.55	46.45	54.42	45.58

Distribution by sector within gender

1994				2004		
sector	male	female	Total	male	female	Total
agriculture & fishing	4.86	1.79	3.43	3.61	1.35	2.58
mining & manufacturing	39.54	26.76	33.61	35.51	20.16	28.51
recycling and construction	15.50	3.10	9.74	15.79	1.71	9.37
retails , trade & hotels	7.86	14.79	11.08	12.39	19.21	15.50
transport	8.76	4.89	6.96	9.55	3.99	7.01
financial intermediation	3.78	6.36	4.98	6.86	8.51	7.61
public administration	7.68	6.16	6.97	6.56	9.36	7.83
education and health	7.54	32.20	18.99	6.70	31.89	18.18
other service & activities	4.47	3.96	4.23	3.05	3.84	3.41

B7. Ownership of the firm, by gender and year

Distribution by gender within firm's ownership

	1994		2004	
Firm's ownership	male	female	male	female
public	51.25	48.75	42.34	57.66
private	60.26	39.74	62.82	37.18
Total	53.55	46.45	54.42	45.58

Distribution by firm's ownership within gender

	1994			2004		
Firm's ownership	**male**	**female**	**Total**	**male**	**female**	**Total**
public	71.22	78.12	74.43	31.92	51.89	41.02
private	28.78	21.88	25.57	68.08	48.11	58.98

APPENDIX C

Educational categories

1994				2004			
wyksz		Years of education	NEW	wyksz [44]		Years of education	NEW
1	university	18	20	1	doctoral	21	20
2	Technical College	16	16	2	Post diploma	18	20
3	High school	14	14	3	master	17	16
4	Technical High school	13	13	4	university	15	16
5	Technical Training	11	11	5	School after lyceum	14	14
6	Primary	8	7	6	Middle voc. S.	13	13
.				8	Basic voc. S.	12	11
				9	gymnasium	9	11
				10	elementary	6	7

*Wyksz is the original variable for education in 1994 and 2004.
*Years of education is the variable indicating years of schooling accomplished (calculated on the basis of Wyksz)
*****NEW is the harmonised variable for years if schooling.

Geographical Subdivision of Voivodships into regions

Region 1	Region 2	Region 3	Regio 4
01 Warszawskie 13 Ciechanowskie 47 Lodzkie 59 Piotrkowskie 61 Plockie 67 Radomskie 73 Sieradzkie 75 Skierniewickie	09 Bydgoskie 25 Kaliskie 31 Koninskie 57 Pilskie 63 Poznanskie 87 Torunskie 91 Wloclawskie	07 Bielskie 15 Czestochowskie 27 Katowickie 53 Opolskie	21 Gorzowskie 23 Jeleniogorskie 39 Legnickie 41 Leszczynskie 89 Walbrzyskie 93 Wroclawskie 97 Zielonogorskie

Region 5	Region 6	Region 7	Region 8
17 Elblaskie 19 Gdanskie 33 Koszalinskie 77 Slupskie 81 Szczecinskie	05 Bialostockie 45 Lomzynskie 51 Olsztynskie 55 Ostroleckie 79 Suwalskie	03 Bialskopodlaskie 11 Chelmskie 43 Lubelskie 71 Siedleckie 95 Zamojskie	29 Kieleckie 35 Krakowskie 37 Krosnienskie 49 Nowosadeckie 65 Przemyskie 69 Rzeszowskie 83 Tarnobrzeskie 85 Tarnowskie

Occupational categories

Original group	Occupational group Description	Merged grous
1	Legislators Administrator	Occ1
2	Managers Professional Technicians	
3	Associate professionals	
4	Clerks	Occ2
5	Service workers Sales workers	Occ3
6	Agricultural workers Fisheries workers	Occ4
7	Trade workers	Occ5
8	Plant & Machine Operators Assemblers	Occ6
9	Elementary occupations	Occ7

Bibliography

Adamchik V. and Bedi A. (2003), Gender pay differentials during the transition in Poland , *Economics of Transition*, 11(4), pp 697-726.

Aigner D. J. and G.C. Cain (1977), Statistical theories of discrimination, *Industrial and Labor Relations Review,* 30 (2), pp. 175-187.

Agesa R. U. (1999), The urban gender wage gap in an African country: findings from Kenya, *Canadian Journal of Development Studies,* 20(1), pp. 59-76

Albrecht J., Björklund A. and Vroman S. (2003), Is there a glass ceiling in Sweden?, *Journal of Labor Economics,* 21, pp. 145-177.

Amin S., Quayes S. and Rives J. M. (2006), Are children and parents substitutes or complements in the family labor supply decision in Bangladesh?, *The Journal of Development Areas*, 40 (1).

Amuedo-Dorantes C. (2004), Determinants of poverty and implications of informal sector work in Chile, *Economic Development and Cultural Change*, 52 (2), pp. 347-368.

Antecol H. and Bedard K. (2004),The Racial Wage Gap: The Importance of Labor Force Attachment Differences across Black, Mexican, and White Men**,** *The Journal of Human Resources*, 39 (2), pp. 564-583.

Appleton S., Hoddinott J. and Krishnan P. (1999), The gender wage gap in three African countries, *Economic Development and cultural Change,* 47(2): 289-312.

Armitage J. and Sabot R. (1991), "Discrimination in East African's urban Labour market", in Birdsall and Sabot eds.), *Unfair advantage: labor market discrimination in developing countries*, Washington D.C., The World Bank.

Arrow K. J. (1972), "Models of job discrimination", in *Racial Discrimination in Economic Life*, Lexingon, Mass.: Anthony H. Pascal (ed.), Heath.

Arumlampalam W., Booth A. L. and Bryan M. L. (2004), Is there a glass ceiling over Europe? Exploring the gender pay gap across the wages distribution*, IZA Discussion Paper* n.1373.

Becker G. (1957), *The Economics of Discrimination*, second edition, Chicago: University of Chicago Press

Becker, G**.** (1967), *Human Capital and the Personal Distribution of Income*, Ann Arbor: University Of Michigan Press.

Bergman B. (1971), The effect on white incomes of discrimination in employment, *Journal of Political Economy*, 79, pp.103-110.

Blackburn R. M., B. Brooks and J. Jarman (2001), Occupational stratification: the vertical dimension of occupational segregation, *Work, Employment and society*, 15 (3), pp. 511-538.

Blau F. D. and Khan L. M. (1992), The gender earnings gap: learning from international comparisons, *American Economic Review*, 82, pp. 533-538.

Blau F. D. and Khan L. M. (1996), Wage structure and gender earnings differentials, *Economica*, (*Supplement on Economic Policy and Income distribution*), 63, pp. S29-S62.

Blau F. D. and Khan L. M. (1997), Swimming upstream: trends in the gender wage differential in the 1980's, *Journal of labour Economics*, 15(1), pp 1-42.

Blau F. D. (1998), Trends in the well-being of American Women, 1970-1995, *Journal of Economic literature,* 36 (March), pp. 112-165.

Blau, F. D. and L. M. Kahn (2003), "Understanding International Differences in the Gender Pay Gap", *Journal of Labor Economics*, 21(1), pp. 106-144.

Blauh F. D and Khan M (2007), The gender pay gap, *Economist Voice,* 4 (4).

Blinder A. S. (1973), Wage Discrimination: Reduced Form and Structural Variables, *Journal of Human Resources,* 8, pp. 436-455.

Boissière M., Knight J.B. and Sabot R. H. (1985), Earnings, Schooling, Ability, and Cognitive Skills, *The American Economic Review*, 75(5).

Brainerd E. (2000), Women in transition: changes in gender waged differentials in Eastern Europe and the former Soviet Union, *Industrial and Labor Relations Review,* 54(1), pp 138-162.

Bushway S., Johnson B. D. and Slocum L. A. (2007), Is the magic still there? The use of the Heckman two-step correction for selection bias in criminology, *Journal of Quantitative Criminology,* 23, pp. 151-178.

Butcher K. F. and C. Rouse (2001), Wage effects of unions and industrial councils in South Africa, *World Bank Policy Research Working paper*: 2520.

Cain G. (1986), "The economic analysis of labour market discrimination: a survey", In *Handbook of Labour Economics vol. 1*, editors: O. Ashenfelter and R. Layards, Amsterdam: North-Holland, pp. 693-785.

Charles M. and D. B. Grusky (1995), Models for describing the underlying structure of sex segregation, *American Journal of Sociology*, 100, pp.931-971.

Chen M., J Vanek, F. Lund, J. Heintz, R. Jhabvala and C. Bonner (2005), *The progress of the world's women: women, work and poverty*, New York, NY: United Nations Development Fund for Women.

Chiswick B. R. (1973), "Schooling, screening and income", in L. C. Solomon and P. J. Taubman (eds.), *Does college matter?,* New York: Academic Press.

Cicomsky B. (2005), *Raport 5: Wynagrodzenia Kobiet i Mezczyzn w Latach 1996-2002. Projekt: Wzmocnienie Polityki Rownego Traktowania Kobiet i Mezczyzn w ramach* Projektu Phare PL 2002/000-580-01.04.02, in Plomien A. (2006), *The gender pay gap in Poland Report*, European

Commission's Expert Group on Gender, Social Inclusion and Employment report for the Equal Opportunities Unit, DG Employment.

Coate S. and Loury G. (1993), Will affirmative action policies eliminate negative stereotypes?, *American Economic Review*, 83 (5), pp. 1220-1240.

Cogneau D. (1999), La formation du revenue des ménages à Antananarivo: une microsimulation en équilibre général pour la fin du siècle , *Economie de Madagascar*, 4, pp. 131-155.

Cotton J (1988), On the decomposition of wage differentials, *Review of Economic and Statistics* 70, pp. 236-243.

Comores-PNUD(2003), *Rapport nationale sur les OMD*.

Comores-PNUD (2007), *Rapport nationale sur le developpement humain aux Comores-le genre et le développement humain* (RNDH).

Commisariat Général au plan -Union des Comores (2005), *Pauvreté, inegalité et marché du travail dans l'Union des Comores – Element s d'analyse fondés sur l'enquête integrale auprès des ménages de 2004.*

Commisariat Général au plan-Union des Comores (2005), *Document de stratégie de croissance et de réduction de la pauvreté – Document intérimaire actualisé.*

CSO (2003), *Labour force survey, IV quarter 2002*, Central Statistical Office, Warszawa.

Das M. B. (2003), The other side of self-employment: household enterprises in India, *Social Protection Discussion Paper Series*, Report n. 27873, Washington D.C.: The World Bank.

Deaton A. (1997), *The analysis of household surveys- A microeconometric approach to development policy*, Baltimore and London: Johns Hopkins University Press for the World Bank.

Doeringer P. B. and M. J. Piore (1971), *International labour markets and manpower analysis*, Lexington, MA, D.C.: Health and Co.

Edgeworth F.Y. (1922), Equal pay to men and women for equal work, *Economic Journal* 32, pp. 431-457.

European Commission (2005), *Employment and social affairs. Employment in Europe 2005. Recent Trends and Prospects.* Manuscript completed in September 2005.

Fafchamps M., Söderborn M. and Benhassine N. (2006), Job sorting in African labor markets, Mimeographed, University of Oxford.

Forsythe N., Korzeniewicz R. P. and V. Durrant (2000), Gender Inequalities and Economic Growth: A Longitudinal Evaluation, *Economic Development and Cultural Change*, 48 (3), pp. 573-617.

Ganguli I. and K. Terrel (2005a), Institutions, market and men's and women's wage inequality: evidence from Ukraine, *IZA Discussion Paper 1724*.

Ganguli I. and K. Terrell (2005b), Wage ceilings and floors: the gender gap in Ukraine's transition,

IZA Discussion Paper n.1776.

Gardeazabal J. and A. Ugidos (2005), Gender wage discrimination at quantiles, *Journal of Population Economics*, 18(1), pp. 165-179.

Gerry C.J., B.Y Kim and C.A. Li (2002), The Gender Wage Gap and Arrears in Russia: Evidence from the RLMS, Mimeo, Department of Economics, University of Essex, United Kingdom.

Glewwe P. (1990), Schooling skills and the return to education: an econometric exploration using data from Ghana, *Living Standards Measurement Working Paper* n.76, Washington D.C.. The World Bank

Glick P. and Sahn D. (1997), Gender and education impact on employment and earnings from Conakry, Guinea, *Economic Development end Cultural Change,* 45, pp. 793-824

Glinskaya E. and T. Mroz (1996), The gender wage gap in Russia from 1992 to 1995, Mimeo, Carolina Population Center, The University of North Carolina.

Grajek M. (2001), Gender pay gap in Poland, *WZB Discussion paper* FS IV 01-13, Berlin: Wissenschaftszentrum Berlin für Sozialforschung.

Grajek M. (2003), Gender pay gap in Poland, *Economics of Planning*, 36(1), pp. 23-44.

Green W. H. (1997), *Econometric analysis, third edition*, Upper Saddle River, NJ: Prentice Hall.

Greenhalgh C. (1980), Male-female wage differentials in Great Britain – Is the marriage an equal opportunity?, *Economic Journal* 90 (363), pp. 751-775.

Hakim C. (1992), Explaining trends in occupational segregation: the measurement, causes and consequences of the sexual division of labour, *European Sociological Review*, 8 (2), pp. 127-152.

Heckmann J. (1979), Sample selection bias as a specification error, *Econometrica, 4*7(1), pp. 153-161.

Horton S, Kambur R. and Mazumdar D.(1991), Labor markets in an era of adjustment: evidence from 12 developing countries, *International Labor Review*, 130, pp. 532-558.

Horton S. (1996), *Women and industrialization in Asia*, London and New York : Routledge.

Hunt J. (2002), The transition in East Germany: when is a ten point fall in the gender wage gap bad news?, *Journal of Labour Economic*s, 20(1), pp 148-169.

Joshi H. and P. Paci (1998), *Unequal Pay for Men and Women. Evidence from British Cohort Studies*. Cambridge MA, MIT Press.

Juhn C., Murphy K. and B. Pierce (1991), "Accounting for the slowdown in Black-White Wage Convergence", in: Kosters M. (ed.) *Workers and Their wages*, Washington D.C.: AEI Press.

Karat Coalition (2003), *Gender assessment of the impact of European Union accession on the status of women in the labour market in central and eastern Europe, National study: Poland*, Ed: Kinga Lohmann, Anita Seibert, Karat Coalition 2003.

Karat Coalition (2005), *Women in the labour market in Bulgaria, Georgia, Macedonia, Poland, Serbia & Montenegro and Slovenia*, Infosheet for 49th session of CSW, New York.

Koenker R. and G. Basset (1978), Regression quantiles, *Econometrica*, 46, pp. 33-50.

Kunze A. (2000), The determination of wages and the gender wage gap: a survey, *IZA discussion paper* n.193.

Kunze A .(2007), Gender wage gap studies: consistency and decomposition, forthcoming in *Empirical Economics.*

ILO (1990), Surveys *of economically active population, employment, unemployment and underemployment. An ILO manual on concepts and methods.* R. Hussmanns, F. Mehran and V. Verma. ILO, Geneva.

IMF (2004), Union *of the Comoros: Selected Issues and Statistical Appendix*, Country Report N. 04/233.

IMF (2005), *Union of the Comoros: 2005 Article IV Consultation and First Review Under the Staff*, Country Report N. 05/299

IMF (2006), *Union of the Comoros: Interim Poverty Reduction Strategy Paper—Joint Staff Advisory Note.*

Lachaud J. P.(1994), *The African Labour Market*, Genève, ILO.

Lehmann H. and J. Wadsworth (2007), Wage arrears and the inequality in the distribution of pay: lessons from Russia", *Research in Labor Economics*, 26, pp. 125-155.

Light A. and M. Ureta (1995), Early-career work experience and gender wage differentials, *Journal of Labour Economics,* 13 (1), pp. 121-154.

Machado J. A. and Mata J. (2001), Earning functions in Portugal 1982-1984: Evidence from quantile regressions, *Empirical Economics,* 26, pp. 115-134.

Machado J. A. and Mata J. (2004), Counterfactual decomposition of changes in the wage distribution using quantile regression, *Journal of Applied Econometrics,* 20, pp 445-465.

Maddala G. S. (1985), *Limited-Dependent and Qualitative Variables in Econometrics*, Econometric Society Monographs. Cambridge University Press: New York, NY.

Manski C. (1989), Anatomy of the selection problem, *Journal of Human Resources,* 24, pp. 343-360.

Miller P. (1987), The wage effect of the occupational segregation of women in Britain, *Economic Journal*, 97, pp. 885-896.

Mincer J. (1974), *Schooling, experience and earnings*, New York: the National Bureau of Economic Research.

Mincer J. and S. W. Polachek (1974), Family investments in human capital: earnings of women, *Journal of Political Economy,* 82 (2), pp. 76-108.

Montenegro C. (2001), Wage distribution in Chile: does gender matter? A quantile regression approach, *Policy research report on Gender and Development – The World Bank* , WP n. 20.

Morrisson C. and J. Jütting (2004), The Impact of Social Institutions on the Economic Role of Women in Developing Countries, *Working Paper 234*, OECD Development Centre, Issy-les-Moulineaux, France.

Morrison C. and J. Jütting (2005), Women's discrimination in developing countries: A new data set for better policies, *World Development*, 33 (7), pp. 1065-1081.

Nestic D. (2007), Differing characteristics or differing rewards: what is behind the gender wage gap in Croatia?, *Working paper 704*, The Institute of Economics Zagreb.

Neumark D. (1988), Employer's Discriminatory Behavior and the Estimation of Wage Discrimination, *Journal of Human Resources,* (23), pp. 693-709.

Newell A. and Reilly B. (1996), The gender wage gap in Russia: some empirical evidence, *Labour Economics*, 3(3), pp 337-356.

Newell A. and B. Reilly (1997), Rates of returns to educational qualifications in the transitional economies, *Education Economics*, 7 (1).

Newell A. and Reilly B. (2001), The gender wage gap in the transition from communism: some empirical evidence, *Economic System*, 25, pp. 287-304.

Newell A. and M. W. Socha (2005), The distribution of wages in Poland, 1992-2002, *IZA Discussion Paper* 1485.

Newell A. and M. W. Socha (2007), The polish wage inequality explosion, *IZA Discussion Paper* 2644.

Nielsen H. K. and M. Rosholm (2001), The public-private sector wage gap in Zambia in the 1990's: a quantile regression approach, *Empirical Economics,* 26, pp. 169-182.

Nordman C. and F. Roubaud (2006), Reassessing the Gender Wage Gap: Does Labour Force Attachment Really Matter? Evidence from Matched Labour Force and Biographical Surveys in Madagascar, *DIAL Working Paper*, IRD-Paris.

Nordman C. J. and Wolff F. C. (2007), Is there a glass ceiling in Morocco? Evidence from matched worker-firm data, paper presented at the IZA/World Bank conference on employment in developing countries in Berlin, February 2007.

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

Oaxaca R. and M.R. Ransom (1994), On discrimination and Decomposition of wage Differentials, *Journal of Econometrics*, 61 (1), pp. 5-21.

Oaxaca R. and M.R. Ransom (1999,, Identification in detailed wage decompositions, Review of Economics and Statistics 81, pp. 154-157.

Oi W. (1990), Employment relations in dual labor markets (It's nice work if you can get it). *Journal of Labor Economics* 8, pp. S124–S149.

O'Neill J. (1985), The trend in the male-female wage gap in the United States, *Journal of Labour Economics*, 3 (1), pp. 91-116.

Orazem P. and M. Vodopivec (2000), Male female differences in labor market outcomes during the early transition to market, *Journal of Population Economics*, 13(2).

Paci P. and Reilly B. (2005), Does economic liberalization reduce gender inequality in the labour market: the experience of the transition economies of Europe and central Asia, Report prepared for the Poverty Reduction & Economic Management Group. Eastern Europe && Central Asia Region, World Bank, Washington D.C.: World Bank.

Phelps E. S. (1972), The statistical theory of racism and sexism, *American Economic Review*, 62, pp. 659-661.

Polachek S. W. (1976), Occupational segregation: an alternative hypothesis, *Journal of Contemporary Businness* 5, pp. 1-12.

Polachek S. W. (1981), Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure, *Review of Economics and Statistics,* 63(1), pp. 60-69.

Polachek S. W. (2004), How the human capital model explains why the gender wage gap narrowed, *IZA discussion paper 1102*.

Plomien A. (2006), *The gender pay gap in Poland Report*, European Commission's Expert Group on Gender, Social Inclusion and Employment report for the Equal Opportunities Unit, DG Employment.

Psacharopoulos G. and Z. Tzannatos eds. (1992), *Women's Employment and Pay in Latin America; Overview and Methodology*, Washington, D.C.: World Bank.

Ranjan P. (2000a), Analysis of child labour in Peru and Pakistan: a comparative study, *Journal of Population Economics,* 13, pp. 3-19.

Ranjan P. (2000b), Child labour, child schooling and their interaction with adult labour: empirical evidence from Peru and Pakistan, *The World Bank Economic Review,* 14, pp. 347-367.

Reilly B. (1999), The gender pay gap in Russia during the transition, 1992-1996, *Economics of transition,* 7(1), pp. 245-264.

Sahn D. E. and Alderman H. (1988), The effects of human capital on wages and the determinants of labour supply in a developing country, *Journal of Development Economics,* 29, pp. 157-183.

Schultz T.W. (1975), The value of the ability to deal with disequilibria, *Journal of Economic Literature,* 13, pp. 827–846.

Schultz T. P. (2004), Evidence of returns to schooling in Africa from household surveys: monitoring and restructuring the market for education, *Journal of African Economies,* 10(2), pp. 127-142.

Sender J. (2002), Women's Struggle to Escape Rural Poverty in South Africa*, Journal of Agrarian Change*, 2 (1), pp. 1-49.

Siphambe H. K., Thokweng-Bakwena M. (2001), The wage gap between men and women in Botswana's formal labour market, *Journal of African Economies,* 19 (3), pp. 127-142

Sparreboom T. (2001), An assessment of labour market information systems in Southern Africa, Africa Development XXVI (3,4), pp. 149-181.

Standing G. (1999), Global feminization through flexible labor: a theme revisited, *World Development*, 27, pp. 583-602.

Stiglitz J. E. (1973), Approaches to the Economics of Discrimination, *American Economic Review Papers and Proceedings*, 63, pp. 287-295.

Sutherland H., Alexander W. R. (2002), The occupational distribution of Maori: 1997-2000, *University of Otago Economics Discussion Papers* n. 0204

Temesgen T. and Zeufack A. G. (2002), *Gender pay inequality in African manufacturing: evidence from linked employer-employee establishment-level data*, Washington, DC: The World Bank Development Research Group.

Temesgen T. (2006), Decomposing gender wage differentials in Urban Ethiopia: evidence from linked employer-employee (LEE) Manufacturing survey data, *Global Economic Review*, 35(1), pp. 43-66.

Tzannatos Z. (1994), The costs of ethnic discrimination to society and minorities: a note on the methodology and results from Guatemala, Mimeo, The World Bank.

UNDP (2005), Human Development Report 2005.

UN Economic Commission for Africa , UNECA (2005), *Economic Report for Africa 2005.*

Yasar M., Nelson C.H., and R.M. Rejesus (2006) Productivity and Exporting Status of Manufacturing Firms: Evidence from Quantile Regressions , *Review of World Economics* , 142(4).

Weichselbaumer D. and Winter-Ebmer R. (2005), A meta-analysis of the international gender wage gap, *Journal of Economic Surveys,* 19(3): 479-511

World Bank (2001), *Engendering Development- Through Gender Equality in Rights,Resources and Voice*. World Bank Policy Research Report. New York: Oxford University Press.

Wright R. E. and J. F. Ernish (1991), Gender discrimination in the British labour market: a reassessment, The Economic Journal, 101(406), pp. 508-22.

Zabalza A. and J. L. Arrufat (1985), *The extent of sex discrimination in Great Britain, Women and equal pay-The effects of legislation on female employment and wages in Britain*, editors: Zabalza and Tzannatos, Cambridge University Press.

Zellner H. (1972), Discrimination against women, occupational segregation and the relative wage, *American Economic Review*, 62, pp. 157-160.

Zielinska E. (2005), "National report on Poland in Report on equal pay(draft)", EC's Legal Experts Network, December 2005.