Gender wage gap when women are highly inactive: Evidence from repeated imputations with Macedonian data

Petreski, Marjan and Mojsoska-Blazevski, Nikica and Petreski, Blagica

University American College Skopje, Association for economic research, advocacy and policymaking “Finance Think” - Skopje

2014

Gender wage gap when women are highly inactive: Evidence from repeated imputations with Macedonian data

Marjan Petreski
School of Business Economics and Management
University American College Skopje

CERGE-EI Career Integration Fellow, Prague, Czech Republic
marjan.petreski@uacs.edu.mk

Nikica Mojsoska Blazevski
School of Business Economics and Management
University American College Skopje
nikica@uacs.edu.mk

Blagica Petreski
Association for economic research, advocacy and policymaking
"Finance Think" - Skopje
blagica.petreski@financethink.mk

Abstract

The objective of this research is to understand if large gender employment and participation gaps in Macedonia can shed some light on the gender wage gap. A large contingent of inactive women in Macedonia including long-term unemployed due to the transition process, female remittance receivers from the male migrant, unpaid family workers in agriculture and so on, is outside employment, but is not necessarily having the worst labour-market characteristics. In addition, both gender wage gap and participation gap enlarge as education decreases, revealing the importance of non-random selection of women into employment. Though, the standard Heckman-type correction of the selectivity bias suggests that non-random selection exists, but the resulting wage gap remains at the same level even when selection has been considered. Instead, we perform repeated wage imputations for those not in work, by simply making assumptions on the position of the imputed wage observation with respect to the median. Then, we assess the impact of selection into employment by comparing estimated wage gaps on the base sample versus on an imputed sample. The main result is that selection explains most of the gender wage gap in the primary-education group (75%), followed by the secondary-education group (55%). In the tertiary group, the small initial gap vanishes once selection considered. This suggests that indeed non-working women are not those with the worst labour-market characteristics. Results suggest that gender wage discrimination in Macedonia is actually between 5.4% and 9.8% and does not exist for the highly-educated women. The inability of the Heckman-type correction to document a role for selection in explaining the gender wage gap may be due to the criticisms to the exclusion restrictions and the large amount of missing wages.

Keywords: gender wage gap, gender participation gap, selection bias, repeated imputations

JEL classification: J16, J31, E24

Acknowledgement: This research has been generously supported by the Global Development Network and the Government of Japan within the Japanese Award for Outstanding Research on Development 2013. The authors thank for the guidance and useful comments of Vladimir Gligorov and the hosts of the Institute for East and Southeast European Studies, Regensburg, Germany, during Marjan Petreski's stay, 15.1-1.2.2014. All remaining errors are solely the authors'.

1. Introduction

Macedonia suffers a large gender wage gap of about 12.5% of males' wage. This gap amplifies to 28.4% for the low-skilled workers, i.e. those with primary or without school. The so-called adjusted wage gap which considers the different characteristics of the working men and women increases to 17.3%, given that employed females on average have better (observable) human-capital characteristics than men. So, workers' characteristics cannot explain the gender wage gap. However, it may still hide an issue which is prevalent in Macedonia and even in all transition economies: the role of woman as a housewife and unpaid family worker in the otherwise large agricultural sector, the inactivity of females due to having permanent income from the male migrant working abroad; females exposed to long-term unemployment with depreciated skills, non-linear careers due to the care for children, and other traditional decisions and stereotypes pertinent to women (European Commission, 2011; Mojsoska Blazevski et al. 2013). These groups of females are persistently inactive and hence outside the wage distribution. As a matter of fact, at the end of 2010, the gender employment gap amounted to more than 18 p.p. against a gender unemployment gap of less than 1 p.p., suggesting that the difference is due to the large inactivity of women. Hypothetically, if they were part of the wage distribution, they were more likely to be on the left side of the wage distribution, but not necessarily at the left corner, i.e. not with the worst labour-market characteristics.

In this paper, we argue that the large gender wage gap in Macedonia may be driven by the absence of the above groups of females from the labour market, i.e. by the large gender participation gap; and that studies failing to address this issue deceptively conclude that the gap cannot be rationally explained, pointing to discrimination against female workers. If selection into employment is non-random, it could be that the selection is important in affecting the observed gender wage gap. If women who are not featuring into the wage distribution due to inactivity tend to have prevalently, but not exclusively, low-wage characteristics, then lower female employment rates become consistent with the large adjusted-for-characteristics gender wage gap in Macedonia. Namely, populating the left-hand side of the wage distribution by those inactive women may drive the gender wage gap below the unadjusted level and therefore explain the inflated adjusted-for-characteristics gender wage gap.

The literature on the gender wage gap is vast and dates back to the 1970s (e.g. Gronau, 1974) and mainly treats two topics: i) workers' and other characteristics which can affect female wages; and ii) gender pay discrimination whereby a man and a woman with the same characteristics are paid differently (Altonji and Blank, 1999). While the effects of the 'explained' part – i.e. of factors like education, experience, age and so on – of the gender wage gap have been well estimated in the empirical literature (for a summary see, e.g. Echrenberg and Smith, 2003), the empirical work on the 'unexplained' part remains with mixed results. It could be due to unobservable factors (like risk aversion, attitude to work, access to social networks) or to discrimination against women (more on the discrimination in Petersen and Saporta, 2004).

The literature on gender gaps in labour force participation rates is also vast. Altonji and Blank (1999) give an overall survey, while Blau and Kahn (2003) offer international comparison. Other pertinent literature includes Beblo et al. (2003); Albrecht et al. (2004); Neal (2004); Fotin (2005); Azmat et al. (2006); Machado (2012) and others. However, although there exists substantial literature on gender wage gaps on one hand, and gender employment gap on the other hand, to our knowledge the variation in both quantities and prices in the labour market, and in particular relating it with the level of skills, has seldom been simultaneously exploited. Few studies in this respect include: Petrongolo and Olivetti (2006); Neal (2004); Blau and Kahn (2004).

The research of the gender wage and employment/participation gaps in transition economies has been scarce (for a summary refer to Arandarenko et al. 2013), and even scarcer for the Western Balkan countries. To our knowledge, only the paper of Blunch (2010) considers some Western Balkan countries in such a context. Studies analysing Macedonia in particular are constrained to Angel-Urdinola (2008), Angel-Urdinola and Macias-Essedin (2008) and Arandarenko et al. (2013). The former two document that gender wage gap is not necessarily explained by the labour-market segmentation (whereby women enter lower-paying sectors) nor by differences in returns to education, but more likely by labour-market discrimination. Similar conclusion is reached by the third study, nevertheless it concludes that there is no selectivity bias in employment among women.

The objective of this research is to understand if gender employment and participation gaps in Macedonia can shed some light on the gender wage gap. In particular, we explore this view by estimating selection-corrected wage gaps for different education levels from repeated imputations. By doing so, the research will contribute to the current sparse of knowledge and policy debate on the inequality of men and women in Macedonia, especially with regard to their employment and earnings opportunities. Namely, a multitude of activists and civil organizations argue that gender discrimination with regard to economic outcomes in Macedonia is on the stage and urge that the government takes active measures to close the gender wage gap, as well to improve opportunities for more women to be employed. Recently, the government undertook wide measures to establish Equal Opportunities Committees at the central and local level and it endorsed the ILO Convention for equal wage for work of equal value. However, observable progress may be lacking. While the gaps themselves may be indeed urging for active policy measures in this domain, we argue that the picture may be considerably different when gender wage gap has been corrected for the gender employment and participation gaps.

The study is organized in the following way. Section 2 presents the underlying survey used in the analysis as well the economic model. Section 3 reviews the gaps in a descriptive manner and then establishes a causal links with the workers' and job characteristics. Section 4 presents the methodology to be employed in correcting the selectivity bias. Section 5 presents the results and offers discussion. The last section concludes.

2. Data and model

The analysis is based on the Survey of Income and Labour Conditions (SILC). This is longitudinal survey of a representative sample of Macedonian individuals and their households, performed in accordance to the Eurostat SILC. It has been performed in Macedonia for the first time in 2010. The survey has a representative sample of about 13,800 individuals, which gives sufficiently rich set for analysis.

As suggested in the literature (e.g. Petrongolo and Olivetti, 2006), we restrict our analysis to individuals aged 16-64, having excluded non-employed persons under 18 and over 64 years of age, students, pensioners, persons with disability due to inflexible labour supply; employed with zero wages as these are likely not the result of their human capital, but a specific situation on the labour market; and self-employed due to the different factors affecting their wages. Finally, our sample consists of 5,642 individuals, which is the sample we work with. Their labour market status is given in Table 1. As suggested in the introduction, while males have both higher rates of employment and unemployment, bulk of females is outside the labour market in Macedonia, i.e. inactive.

Table 1 – Labour market status of persons capable to work (%)

Labour market status	Males	Females
Employed	55.61	36.87
Unemployed	43.23	28.62
Inactive	1.16	34.51
Source: Authors' calculations based on SILC 2010		

The wage concept we use throughout the analysis is the net hourly wage, as the gross wage which is more used in the studies is not directly available in the survey. To estimate the gender wage gap, we rely on Mincer's (1974) human capital earnings function which relates the log of individual wage to gender, level of education, age, years of work experience and other labour-market characteristics, within a specification which is linear in education and quadratic in age[1]:

$$\ln(y_i) = \alpha + \beta gender_i + \gamma_1 edu_{i1} + \gamma_2 edu_{i2} + \gamma_3 age_i + \gamma_4 age_i^2 + \gamma_5 exper_i + X_i' \delta_k + \varepsilon_i$$

Whereby $\ln(y_i)$ is the log of the net hourly wage; $gender_i$ is a dummy variable taking a value of 1 for females and zero for males; edu_{i1} and edu_{i2} are dummy variables for completed education, categorized into three levels (primary or less, secondary and tertiary); primary education is the omitted category; age_i refers to years of age, while $exper_i$ to years of work experience; and X_i' is s vector of other labour-market characteristics, such as type of contract, occupation and industry. The type of employment

[1] Frequently, studies include either age or work experience and their quadratic term. Here, we use only the quadratic term of age as it is likely to be sufficient in capturing the turning point of the wage over the life cycle.

contract comes in two forms: permanent or temporary contract; and with written or without written contract (i.e. informal employment). We consider ten occupational categories based on the two-digit codes of the International Standard Classification of Occupations (ISCO) and eight economic sectors grouped according to the National Classification of Activities NKD Rev.2 (2009). Ethnicity, though very important for the Macedonian case, is not included as the variable is not available in SILC.

Our main interest is the coefficient β, which measures to what extent a female with the same labour-market characteristics as a male is paid less (or more). It is expected that β is negative.

3. Stylised facts and workers' characteristics

In 2010, the unadjusted gender wage gap in Macedonia – calculated as a difference between what average working man and woman are paid as a percentage of man's wage – has been 12.5%. This is still below the level in EU-27 (16% in 2011), Norway (16%) and Switzerland (17.5%) and could be attributed to females exiting labour markets during the (early) transition process due to factors like lack of appropriate skills and smaller opportunity costs of home production (Grajek, 2003). The calculated gap of 12.5% in 2010 is largely in line with the one calculated in Arandarenko et al. (2013) of 13.4% by using the Labour Force Survey 2008-2012; and in Blunch (2010) of 17.5% by using the UNDP Social Exclusion Survey 2010.

When we disaggregate the wage gap according to the skills level as approximated by the level of education, we obtain a result whereby the gap is wider for low- and medium-skilled workers and narrows for high-skilled ones.

Table 2 - Unadjusted gender wage gap (%)

Education level	%
Primary education or less	-28.4
Secondary education	-21.6
Tertiary education	-8.2
Source: Authors' calculations based on SILC 2010	

On the other hand, the employment and participation rates of male workers outpace the ones of female workers by nearly 18.7 and 33.4 percentage points, respectively, pointing out to the very large employment and participation gaps. This is driven by a multitude of factors, including Grajek's (2003) argument of women exiting the market in the early transition due to skill deficiency, but also factors inherent to the Eastern European economies: the traditional role of the woman in the society as a housewife and child-raising person; rural women who engage as an unpaid family worker; the large emigration whereby labour market participation by women is "traded-off" with the remittances sent by the male migrant; long-term unemployment causing further skill erosion and so on (Mojsoska Blazevski et al. 2013). Many of those groups of inactive women, are likely to have high reservation wage and hence stay outside the labour market. Table 3 suggests that working women have better labour-market characteristics than their non-working counterparts, i.e. are more educated, experienced and slightly younger, on average. However, note that the difference in the average education is not that large: the average working woman has slightly more than completed secondary education, while the average non-working counterpart has a third way to completion of a secondary education. This may suggest that non-working women are likely not with the worst labour-market characteristics although they would still be located on the left side of the wage distribution.

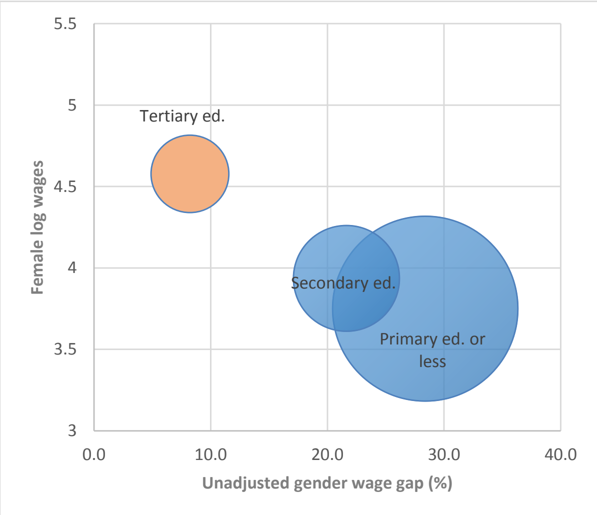
Table 3 – Labour-market characteristics of working and non-working males and females

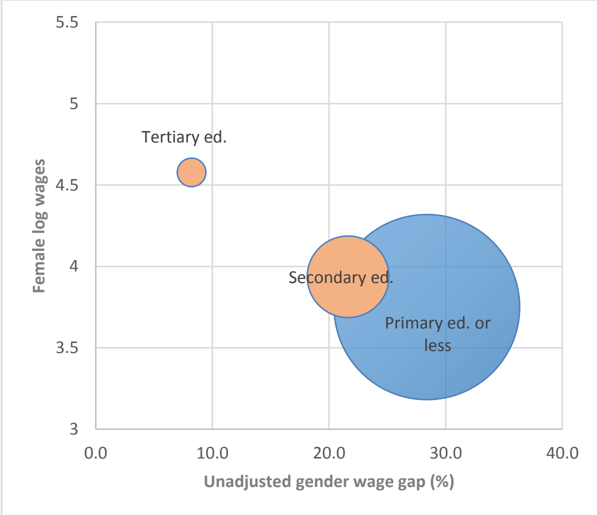
		FEMALE		MALE	
	Variable	**Mean**	**Std.Dev**	**Mean**	**Std.Dev**
WORKING	**Education***	3.19	0.70	3.06	0.66
	Experience	14.62	10.41	15.97	11.19
	Age	41.11	10.50	41.36	11.45
NON-WORKING	**Education***	2.29	0.83	2.67	0.68
	Experience	3.11	7.39	9.06	12.15
	Age	41.50	12.65	39.16	14.24

Source: Authors' calculations based on SILC 2010.
** Education average is a simple average of: 0 – no education; 1 – incomplete primary; 2 – complete primary; 3 – complete secondary; 4 – complete tertiary; 5 – complete post-tertiary.*

Table 3 reveals additional interesting fact: the average working woman in Macedonia has better education characteristics than the average working man. This is a consequence of the low female employment rate, in the presence of selective participation into employment and has been documented for the Southern European countries also (Petrongolo and Olivetti, 2006). A look at Figure 1 suggests that this may be true: the employment gap is very wide for primary or less educated individuals, but more strikingly, the participation gap soars at 68 p.p. for this group of individuals. Namely, the largest circle on the right figure is comparatively much larger than the other two, as compared to the left figure.

Figure 1 – Gender wage gap against i) gender employment gap (left) and ii) gender participation gap (right), at different levels of education





Source: Authors' calculations based on SILC 2010

Note: The size of the circles represents the size of the respective gaps. Orange circles represent those characteristics which can be found more frequently among employed/active women than employed/active men, (e.g. employed women more frequently have tertiary education than employed men; active women more frequently have secondary or tertiary education than active men) and vice versa for the blue circles.

Figure 1 reveals one additional interesting aspect, that is the positive correlation between the gender wage gap on the one hand and the gender employment and participation gaps on the other (correlation coefficient of 0.86 and 0.85 respectively). Namely, circles get larger as we move to the right on the x-axis. Such positive correlation between the gaps may reveal significant sample selection effects in observed wage distributions. This may suggest that women tend on average to be more negatively selected into employment than men, which is likely explained by the high reservation wage.

If the claim that employed women tend to have better education characteristics than employed men (Table 3), then controlling for workers' and other characteristics might not reduce or even inflate the wage gap. Some evidence of inflating the adjusted gender wage gap is provided in Table 4. Column (1), whereby the gender is the only variable in a Mincer earnings function, reproduces the unadjusted gender wage gap, while the subsequent columns are variations of the Mincer equation: column (2), in addition to gender, considers human-capital characteristics, while the subsequent columns add the type of contract, occupation and industry, respectively. All coefficients pertinent to Table 4 are reported in Table A1 in Appendix 1, along a brief discussion of their signs and magnitudes. It is evident from Table 4 that all these characteristics inflate the gender wage gap, probably the most of the inflation is made by the human-capital characteristics (where education belongs). It suggests that once workers' and other characteristics relevant for wage have been controlled for, the potential gender pay discrimination worsens, i.e. that the explained part of the gender wage gap is negative (similarly as in Arandarenko et al. 2013 and Blunch, 2010). This finding of a limited role of workers' and job characteristics to explain the gender wage gap in Macedonia is in line with the findings in Angel-Urdinola (2008) and Angel-Urdinola and Macias-Essedin (2008), who attribute the unexplained part to gender pay discrimination. However, this would have been a valid conclusion, should the selection of women into inactivity (Table 1) be fully random.

Table 4 – Unadjusted and adjusted gender wage gap – Mincer earnings equation

	Unadjusted GWG	Adjusted for workers' characteristics, and industry GWG			
Dependent variable: Log of the net hourly wage					
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Female	-0.125***	-0.173***	-0.175***	-0.160***	-0.175***
Workers' characteristics					
Education, age and experience		Yes	Yes	Yes	Yes
Job characteristics					
Contract			Yes	Yes	Yes
Occupation				Yes	Yes
Industry					Yes
Sample	2,613	2,613	2,547	2,420	2,420
Adjusted R-square	0.011	0.259	0.277	0.31	0.356

*Source: Authors' calculations. *, ** and *** denote statistical significance at the 10, 5 and 1% level, respectively. Estimates are robust to heteroskedasticity.*

Table 5 presents the estimated gender wage gap in the same fashion as Table 4, by education cohorts. For convenience, the other information is removed from the table and is available on request. The unadjusted gap reported in column (1) is a reproduction of Table 2, while the subsequent columns adjust the gap for different worker's and job characteristics. Contrary to the overall case, it is evident that observables do exert some explanatory power over the gender wage gap. However, since we observe education cohorts here, the reduction of the adjusted wage gap within the cohorts actually confirms that employed women have only better education characteristics than employed men in Macedonia, but this is not the case with the other characteristics, as we documented in Table 3. In the cases of primary and secondary education, all observables reduce the gender wage gap, i.e. explain about 28% and 17% of it, respectively. In the case of tertiary education, it may be that after the observable characteristics have been taken into account, gender wage difference does not exist anymore – as the significance of the estimated coefficient is reduced – but the finding is not robust. Still, even if we consider the cases when observables explain a largest share of the gender wage gap (column 5 for primary education and column 4 for secondary education), the resulting gender wage gaps are still very large. Does this mean that gender pay discrimination has been committed within these two groups? Only if the selection into employment has been fully random.

Table 5 – Unadjusted and adjusted gender wage by level of education

	Unadjusted GWG	Adjusted for workers' characteristics, and industry GWG			
Dependent variable: Log of the net hourly wage					
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
		(1) + Age and experience	(2) + Contract	(3) + Occupation	(4) + Industry
Female, primary	-0.284***	-0.256***	-0.283***	-0.245***	-0.205***
Female, secondary	-0.216***	-0.201***	-0.203***	-0.179***	-0.210***
Female, tertiary	-0.082**	-0.070*	-0.061	-0.077*	-0.098**
*Source: Authors' calculations. *, ** and *** denote statistical significance at the 10, 5 and 1% level, respectively. Estimates are robust to heteroskedasticity.*					

The high reservation wage of the majority of the non-working women results in a non-random selection into employment. If this is true, it makes sense to worry about the way in which selection may affect the resulting gender wage gap (Gronay, 1974; Manski, 1989; Vella, 1998). In particular, if Macedonian non-working women tend to have low-to-medium wage characteristics, low female employment rates for low-skilled women may become consistent with high gender wage gaps (Table 1) simply because low-to-medium wage female workers would not feature in the observed wage distribution.

4. Methodology

Given non-working women in Macedonia have not chosen randomly not to work or participate on the labour market, their non-existence in the wage distribution makes inferences about the estimated gender wage gap false. Now fairly standardly, wage equations estimate the Heckman (1976, 1979) sample-selection model – whereby selection problem is treated as an omitted-variable problem – to correct for the non-random selection into employment. An alternative empirical approach we employ in this paper recuperates the counterfactual wage distribution that would prevail had the selection into work been fully random. We use repeated imputation technique based on median regressions (Rubin, 1987) which does not require assumptions on the actual level of missing wages, as usually required in the matching approach, nor it requires arbitrary exclusion restrictions and lack of robustness (Manski, 1989) raised in two-stage Heckman (1979) sample-selection models. Afterwards, the gender wage gap, which constitutes our base sample, is compared with the sample based on the imputed wages, i.e. with the sample where all individuals are assumed to be employed, i.e. all individuals have an observed wage.

The impact of selection into employment on estimated gender wage gaps will be evaluated by comparing estimates obtained from repeated imputations vis-à-vis the baseline specification. One plausible characteristic of the median regressions is that, if missing wage observations fall completely on one or the other side of the median regression line, the results are only affected by the position of wage observations with respect to the median, and not by the precise values of imputed wages. Hence, we can make an assumption referring to the economic theory on whether an individual who is not in work should have a wage observation below or above the median wages for their gender. This assumption relates to individual's observable characteristics, including: education, experience, age, marital status and the number of children (but not gender). By doing so, we allow for, so-called, selection on observables. In doing so, we will estimate the probability of each individual belonging above or below their gender-specific median. Then, the missing wage values are replaced by simulated versions, and independent simulated datasets are obtained. Despite early suggestions (e.g. Schafer and Olsen, 1998; Schafer, 1999) that 3 to 5 imputations are sufficient to obtain good results, some more recent contributions (Graham et al. 2007) document that increasing the number of imputations increases the efficiency of the estimations. Hence, we use variants of 5, 50 and 100 imputations.

Selection on observed characteristics is exploited in the matching approach also, which consists in imputing wages for the non-employed by assigning them the observed wages of the employed with matching characteristics (see Blau and Beller, 1992 and Juhn, 1992; 2003). Under this approach, each unemployed/inactive person obtains only one wage given its characteristics, i.e. the assignment of wages to those individuals is conducted in a fully deterministic way. While our approach partially relies on the selection of observables, as explained above, we extend the framework of Johnston et al. (2000) and Neal (2004) by using probability models (probit) to assign individuals on either side of the median of the wage distribution.

However, since here we try to recover the wages of individuals for whom we do not have any guidance on their wages (unlike in panel datasets where they could have earned in some years but not in others), we need some guidance on the magnitude of the problem in median gender wage gaps it may inflict. If women in Macedonia are on average less attached to the labour market than men (Table 1), and if individuals who are less attached have on average higher wage characteristics than the fully attached low-skilled individuals (due to issues like child-raising, high reservation wage, reliance on remittances)[2], then the difference between the gender wage gap on the imputed and the actual wage distribution tends to be lower, the higher the proportion of imputed wage observations. In the final step, we use the estimated gender wage gaps from each of the simulated datasets to obtain the part of the variance reflecting missing-data uncertainty. This method has the advantage of using all available information on the characteristics of the non-employed and of taking into account uncertainty about the reason for missing wage information (Rubin, 1987; Petrongolo and Olivetti, 2006).

In what follows, we first present the results with the Heckman-type correction; then the results with the repeated-imputations correction and then offer a discussion.

[2] For example, in our sample, the average education of a non-working woman is 2.3, while that of working low-skilled woman is 1.8, and former are younger by about 5 years than the latter.

5. Results and discussion

5.1 Heckman-type correction

Table 6 reports the results with the correction for selectivity a-la Heckman (1979). For convenience, the first row of results reproduces the uncorrected-for-selectivity gender wage gap, while different columns include different workers' and job characteristics. Results of the Heckman estimation procedure show that self-selection is a neither statistically nor economically significant factor. The inverse mills ratio is insignificant in all cases, while the estimated coefficients are of almost the same magnitudes, i.e. the difference is in the third or fourth decimal. Hence, results imply that selection into the labour force has not role to place for the difference between men and women in Macedonia. This is hardly convincing, but in line with the findings of Arandarenko et al. (2013) for women.

Table 6 – Results with Heckman correction for selection

Dependent variable: Log of the net hourly wage					
Variable		**(1)**	**(2)**	**(3)**	**(4)**
OLS	**Female**	-0.173***	-0.175***	-0.160***	-0.175***
Heckman	**Female**	**-0.173*** **	**-0.175*** **	**-0.160*** **	**-0.175*** **
Workers' characteristics					
Education, age and experience		Yes	Yes	Yes	Yes
Job characteristics					
Contract			Yes	Yes	Yes
Occupation				Yes	Yes
Industry					Yes
Inverse mills ratio (lambda) Selection bias (standard error in parenthesis)		0.036 (0.058)	0.039 (0.057)	0.024 (0.095)	0.032 (0.068)
*Source: Authors' calculations. *, ** and *** denote statistical significance at the 10, 5 and 1% level, respectively. Estimates are robust to heteroskedasticity.*					

When we look at the results by education group, some deviations from the aggregate picture emerge. Table 7 reproduces the unadjusted and adjusted-for-characteristics gaps (columns 1 and 2) and then adds the Heckman-corrected gaps in column (3). Under each coefficient in column (3), we give the inverse mills ratio and its significance in squared brackets. For convenience, the table reports only the gender wage gaps, while other coefficients are available on request. Selectivity bias becomes statistically significant in all three education cohorts. The negative coefficients indicate that those with relatively low earnings prospects within the cohorts tend to self-select into employment. It is plausible explanation, given that we suspected that those who decide not to participate (predominantly women) are not with the worst labour-market characteristics.

Table 7 – Results after Heckman correction, by education

Dependent variable: Log of the net hourly wage			
Variable	**(1)**	**(2)**	**(3)**
	Unadjusted GWG	Adjusted for characteristics GWG	**Adjusted for characteristics and selectivity GWG (Heckman)**
Female, primary	-0.284***	-0.255***	**-0.256*** [-0.451***]**
Female, secondary	-0.216***	-0.201***	**-0.190*** [-0.537***]**
Female, tertiary	-0.082**	-0.070*	**-0.057 [-0.517***]**
*Source: Authors' calculations. *, ** and *** denote statistical significance at the 10, 5 and 1% level, respectively. Estimates are robust to heteroskedasticity. Lambda is in square parentheses.*			

However, the obtained coefficients suggest that the gap in the primary-education cohort remains even after correction for selectivity, while in the secondary-education cohort, selectivity explains only 1.1 percentage points of the gap. The coefficient in the tertiary-education cohort loses significance. Again, it is unlikely that gender participation gap, especially among women in the primary education group, cannot explain the gender wage gap.

5.2 Repeated-imputation correction

Given the suspects of the Heckman-type correction that selection does not explain the gender wage gap in Macedonia, we present the results of a correction based on repeated imputations. Results are reported in Table 8. Again, for convenience, the first column reproduces the OLS estimates. Column (2) reports the estimates with 5 imputations, column (3) with 50, while column (4) with a hundred. Note that imputations include worker's characteristics and exclude job-related characteristics, as there is no sufficient information for the latter to be imputed.

Before looking at the coefficients of interest, note that below each estimated coefficient there are two pieces of information given in italic: the number in squared brackets represents the relative efficiency of the multiple-imputation inference, while the percentage relates to the share of between-imputation variance – i.e. the one due to missing observations – in the total variance. The relative efficiency of the multiple imputation inference is determined by the amount of missing information and the number of imputations. Our results land some evidence in line with Graham et al. (2007) that increasing the number of imputations increases the relative efficiency, since the numbers in parentheses approximate unity as we move from 5 to 100 imputations. The between-imputation variation is fairly large in the majority of variables, which is expected given more than a half of our sample were individuals who were unemployed or inactive, hence without wage. This prevents the uncertainty due to the missing information in our sample to be small.

Turning to the results of our interest – the gender wage gap – the repeated imputations give a gap which is almost three times smaller than the adjusted-for-workers'-characteristics gap and twice smaller than the unadjusted gap. This is the first evidence that majority of the inflated adjusted-for-characteristics gender wage gap in Macedonia is rather due to non-random selection of females into employment, and not due to gender discrimination. More precisely, the lower wage gaps on imputed rather than actual wage distributions suggest, as expected (recall the discussion related to Figure 1), that women in Macedonia tend on average to be more negatively selected into work than men. In other words, women which are outside the labour market are not the ones with the worst characteristics, as we documented in Table 3,[3] but also with the Heckman-corrected results (Table 7). Our finding of Table 8, column (5) suggests that once workers' characteristics and selectivity bias into employment have been taken into consideration, the unexplained gender wage gap reduces to about 5.7%, which could be labelled as gender wage discrimination in Macedonia against women. This finding is robust to including different sets of predictors of the missing wage in the probit regression (Table A2 in Appendix 1)[4].

[3] Should have this been the case, we will have observed that the selectivity-corrected gender wage gap increases rather than decreases.

[4] As an additional, yet indirect, robustness check, we performed the repeated imputation with both the probit regression predicting missing wage position with respect to the median and the Mincer's function having only gender as explanatory variable. If only gender determines the missing wage and the way of imputing wages depends only on gender, then the unadjusted gender wage gap should be reproduced. Indeed, we obtain an estimate of 12.35%, which is very close to our estimate of 12.47% in Table 4 (the difference is only due to randomness inflicted by the imputations).

Table 8 – Results after imputation

Variables	No imputations	5 imputations	50 imputations	100 imputations
Dependent variable: Log of the net hourly wage				
	(1)	**(2)**	**(3)**	**(4)**
Female	-0.173***	-0.060***	-0.059***	-0.057***
		[0.97]	*[0.99]*	*[1.00]*
		11.5%	*34.9%*	*34.3%*
Secondary education	0.168***	0.0373*	0.0351	0.0366*
		[0.93]	*[0.99]*	*[1.00]*
		30.2%	*35.5%*	*35.2%*
Tertiary education	0.731***	0.551***	0.551***	0.551***
		[0.89]	*[0.99]*	*[1.00]*
		49.3%	*34.6%*	*34.4%*
Age	0.0140*	0.00778	5.84E-03	5.55E-03
		[0.92]	*[0.99]*	*[1.00]*
		32.4%	*31.5%*	*37.9%*
Age^2	-0.0002***	-8.40E-05	-5.80E-05	-5.38E-05
		[0.95]	*[0.99]*	*[1.00]*
		18.8%	*29.7%*	*38.1%*
Experience	0.0156***	0.00501***	0.00520***	0.00517***
		[0.87]	*[0.99]*	*[1.00]*
		58.3%	*36.2%*	*41.4%*
Constant		3.830***	3.858***	3.863***
		[0.93]	*[0.99]*	*[1.00]*
		27.5%	*33.9%*	*37.9%*
Observations	2,613	5,642	5,642	5,642
R-square	0.259			
Imputations		3029	3029	3029

*Source: Authors' calculations. *, ** and *** denote statistical significance at the 10, 5 and 1% level, respectively. Estimates are robust to heteroskedasticity.*

Table 9 details the analysis at the education level. It brings out only the variable of interest due to space. Results are very revealing and aligned with our discussion that women absent of the labour market are likely at the right side of the primary education cohort and spread over the secondary-education cohort (on the left side of the wage distribution, but not necessarily at the left corner). Namely, after repeated imputations, the gender wage gap in the primary-education group reduces to only 5.4% out of 28.4% (unadjusted) and 25.6% (adjusted for workers' characteristics). This suggests that 2.8 p.p. of the gender gap in this group could be explained by workers' characteristics, but the largest amount, 21.2 p.p. (or 75%), is due to selectivity bias into employment. As suggested earlier, this is the contingent of women who tend to have low-to-medium-wage characteristics and who purposely stay outside the labour market from a variety of reasons: child-raising, care of elderly, women who receive remittances from the male migrant, unpaid family worker and so on. The large wage gap in this group, namely, discourages labour market participation of these women relatively more than in the other education groups. And, this is consistent with the sizeable impact of selection in the lowest education group. The residual of the gender gap of 5.4% could be labelled as gender pay discrimination in this educational group.

Table 9 – Results after imputation, by education

Dependent variable: Log of the net hourly wage			
Variable	**(1)**	**(2)**	**(3)**
	Unadjusted GWG	Adjusted for characteristics GWG	**Adjusted for characteristics and selectivity GWG (100 imputations)**
Female, primary	-0.284***	-0.255***	**-0.054***
Female, secondary	-0.216***	-0.201***	**-0.098*****
Female, tertiary	-0.082**	-0.070*	**-0.045**
*Source: Authors' calculations. *, ** and *** denote statistical significance at the 10, 5 and 1% level, respectively. Estimates are robust to heteroskedasticity.*			

More appealing, in authors' view, are the results for the secondary-education group. Here, only 1.5 p.p. of the gender wage gap could be associated with the workers' characteristics, *inter alia*, because the power of the education to inflate the gap may be working more than in the previous group. While, 11.8 p.p., or 55%, of the gender wage gap could be associated with the selectivity bias into participation. Given the gender participation gap in this group has been much smaller than in the primary-education group (Figure 1, right), its explanatory power over the gender wage gap is expectedly smaller. However, selectivity having to explain 38% of the gender wage gap in this group again suggests that the categories of women identified before are with low-to-medium-wage characteristics. It is a plausible explanation that some of the female remittance receivers or those having high reservation wages are actually not that low-skilled, but decide not to work, simply for the benefits afforded by remittances or by the other circumstances. The identified gender pay discrimination in this group, of 9.8%, may also suggest that these women are more deterred from the labour market, since they may be exposed to larger wage discrimination than their primary- or tertiary-education peers (as the unobserved characteristics among education group cannot be assumed different).

Finally, the gender wage gap in the tertiary-education group vanishes, as not only the magnitude of the coefficient declines, but it also entirely loses significance. This suggests that once selectivity bias into employment is considered (i.e. highly-educated women who decide not to work are taken into account), the labour market does discriminate the wage of the highly-educated females.

5.3 Why Heckman fails to explain the gap?

Overall, both Heckman-corrected and imputations-corrected estimates suggest a presence of selectivity bias, i.e. both document a role of the gender employment and participation gaps for the gender wage gap in Macedonia. However, the Heckman procedure fails to find that selection actually explains a portion of the gap, which is found under the repeated imputations technique. Hence, the question of why this may be the case. Heckman (1979) himself considered this estimator to be useful for giving good starting values for maximum likelihood estimation and that "given its simplicity and flexibility, the procedure outlined … is recommended for exploratory empirical work" (p.160). But, the usage of the method, especially in wage estimation, has been widespread. There are, though, some critical points which may reduce the efficiency of the Heckman procedure to explain the gender wage gap even though selectivity has been documented.

A first line of criticism is that estimated coefficients are sensitive to the distributional assumptions placed on the error term in the outcome equation and especially in the selection equation (Little and Rubin, 1987). However, the Monte-Carlo simulations summarized in Puhani (2000) and conducted to examine this do not find superiority of an estimator. However, the correlation between the error terms of the outcome and selection equations has been found to reduce Heckman's model efficiency.

A second – and probably most important – line of criticism is related to the exclusion restrictions, i.e. the variables explaining the selection equation, but not the outcome one. In the practical work, the selection equation needs variables which are not included in the outcome equation, i.e. affect the decision to participate in the labour market, but do not affect the wage. We included the marital status and the number of children as they may be affecting the reservation wage and hence the decision to participate in the labour market. But, there is no guarantee that they do not affect the wage also, nor that they are a good predictors of the labour-market status. Moreover, we are lacking additional variables – like the amount of remittances received or the scope of agricultural work of the household, or availability of childcare facilities – which may explain the decision not to participate, but not the wage directly. Leung and Yu (1996) investigated this issue in details and find that the collinearity between the outcome-equation regressors and the inverse Mills ratio may be the main source of high inefficiency of the Heckman estimator. It could be caused either by the exclusion restrictions, which we have already thrown doubt on that may not be independent of the wage; or by the large share of missing data, which is also the case in our sample (about 60% of working-age individuals do not have a wage in the sample). To further confirm that this may be the case, we regressed the inverse mills ratio on the regressors of the outcome equation and obtained a very high R-square ranging up to 0.99, which is a clear sign that this may be the most severe source of inefficiency of the Heckman-corrected estimates in our case.

Finally, even early debates sparked by Duan et al. (1983) claimed that the predictive power of the OLS model is at least as good as that of the Heckman model. Therefore, the considerations about the distributional properties of the errors and the role of the exclusion restrictions in our case suggest that there is a potentially important and large source of inefficiency in the Heckman-type estimator, which explains why it yields to a result whereby the gender wage gap remains unexplained even after selectivity has been taken into account in an economy with large participation gap for women.

6. Conclusions and recommendations

Low-skilled females in Macedonia are paid much less than low-skilled males with the same labour-market characteristics. Although at first glance this fact seems to suggest evidence of a highly discriminatory pay across genders in this group, appearances can be deceptive, given the large gender participation gaps. In this paper we noted that gender wage gaps by education cohorts in Macedonia are positively correlated with the gender employment and participation gaps, hence elucidating the importance of non-random selection of women into employment in understanding the variation in gender wage gaps by educational cohorts. Though, the standard Heckman-type correction of the selectivity bias suggests that non-random selection exists, but the resulting wage gap remains at the same level even when selection has been considered. Instead, we performed repeated wage imputations for those not in work, by simply making assumptions on the position of the imputed wage observation with respect to the median. Then, we estimated gender wage gaps on imputed wage distributions, and assessed the impact of selection into employment by comparing estimated wage gaps on the base sample with those obtained on a sample enlarged with wage imputation. Imputation has been performed on observable human-capital characteristics of missing wage observations. We argued that repeated imputations technique is likely more efficient than Heckman technique due to issues related to the distributional properties of the errors in the outcome and selection equations, the exclusion restrictions imposed in the latter technique as well its inefficiency to handle large amount of missing data. The analysis utilized the Survey on Income and Living Conditions in Macedonia 2010.

We find lower gender wage gaps on imputed rather than actual wage distributions, suggesting that women in Macedonia on average tend to be more negatively selected into work than men. We document that the gender wage gap, after accounting for the selection bias, dwindles to about 5.7% from 17.3%. More indicatively, within education cohorts, selection explains most of the gender wage gap in the primary-education group (75%), followed by the secondary-education group (55%). In the tertiary group, once selection considered, the earlier non-robust finding that gender wage gap does not exist is confirmed. Overall, **results suggest that selection into the labour market in Macedonia has large role to play in explaining the gender wage gap**. Findings, however, suggest that women who self-select not to engage on the labour market dominantly belong to the primary-education group (low skills), but also to the secondary-education group, i.e. on average they are not the women with the worst labour-market characteristics. This could be explained by several characteristics of non-working women in an ex-communist society as Macedonia is: long-term unemployed due to losing jobs in the early transition; female household heads receiving remittances from the male migrants; unpaid family workers in agriculture; women caring for dependents, all of those circumstances likely resulting in high reservation wage. The consideration of non-workers into the estimates of the gender wage gap suggests that the gender wage discrimination in Macedonia is actually between 5.4% and 9.8% and does not exist for the highly-educated women.

The result that the gender participation gap has a large role to play in explaining gender wage gap in Macedonia and the application of repeated imputations technique outpacing some drawbacks of the standard Heckman-type correction of the selectivity bias into labour-market participation are the two contributions the paper makes to the current sparse of literature.

7. References

Albrecht, J., A. van Vuuren and S. Vroman (2004) Decomposing the Gender Wage Gap in the Netherlands with Sample Selection Adjustment. IZA DP No. 1400.

Altonji, J. and R. Blank (1999) Race and Gender in the Labor Market. In: O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*, North-Holland, volume 3C: 3141-3259.

Angel-Urdinola, D. F. (2008) Can the Introduction of a Minimum Wage in FYR Macedonia Decrease the Gender Wage Gap? Social Policy Discussion Paper No. 0837, Washington DC: World Bank.

Angel-Urdinola, D. F. and Macias, V. (2008) FYR Macedonia Labour Market Profile 2004-2007: A Policy Note. Washington, DC: World Bank

Arandarenko, M., Vujic, S., Vladisavljevic, M. (2013) *Gender pay gap in Western Balkans: Evidence from Serbia, Macedonia and Montenegro*. Belgrade: Foundation for Advancement of Economics.

Arulampalan, W., A. Booth and M. Bryan (2007) Is There a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution. *Industrial and Labor Relations Review,* 60(2), p.163-186.

Azmat, G., M. Güell and A. Manning (2006) Gender Gaps in Unemployment Rates in OECD Countries. *Journal of Labor Economics,* 24(1), p.1-37.

Beblo, M., D. Beninger, A. Heinze and F. Laisney (2003) Measuring Selectivity-corrected Gender Wage Gaps in the EU. ZEW DP No. 03-74, Mannheim.

Blau, F. and A. H. Beller (1992) Black-White Earnings over the 1970s and 1980s: Gender Differences in Trends. *Review of Economics and Statistics*, 72(2): p.276-286.

Blau, F. and L. Kahn (1996) Wage Structure and Gender Earnings Differentials: An International Comparison, *Economica,* 63, p.S29-S62.

Blau, F. and L. Kahn (2003) Understanding International Differences in the Gender Pay Gap. *Journal of Labor Economics*, 21, p.106—144.

Blau, F. and L. Kahn (2004), The US gender Pay Gap in the 1990s: Slowing Convergence? NBER WP No 10853.

Blunch, N.H. (2010) *The Gender Earnings Gap Revisited: A Comparative Study for Serbia and Five Countries in Eastern Europe and Central Asia*. Washington, DC: World Bank.

Buchinsky, M. (1998) The Dynamics of Changes in the Female Wage Distribution in the USA: A Quantile Regression Approach. *Journal of Applied Econometrics*, 13, p.1-30.

Duan, N., Manning, W.G., Morris, C.N., and Newhouse, J.P. (1983) A Comparison of Alternative Models for the Demand for Medical Care. *Journal of Business and Economic Statistics,* 2(3), p.283-289.

Echrenberg, R. G. and Smith, R. S. (2003) *Modern Labor Economics: Theory and Public Policy* (8th edition). Pearson Education, Inc.

European Commission (2011) Exchange of good practices on gender equality: Reducing the gender pay gap. European Commission, Summary report. Available at: http://ec.europa.eu/justice/gender-equality/files/exchange_of_good_practice_de/summary_report_de_2011_en.pdf

Fortin N. (2005) Gender Role Attitudes and the Labour Market Outcomes of Women across OECD Countries. *Oxford Review of Economic Policy*, 21(3), p.416-438.

Graham, J.W., Olchowski, A.E. and Gilreath, T.D. (2007) How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. *Prev Sci,* 8, p.206-213.

Grajek, M. (2003) Gender Pay Gap in Poland. *Economics of Planning*, 36, p.23-44.

Gronau, R. (1974) Wage Comparison - A Selectivity Bias. *Journal of Political Economy*, 82(6), p.1119-1143.

Heckman, J. (1976) The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic Social Measurement,* 5(4), p.475-492.

Heckman, J. (1979) Sample Selection Bias as a Specification Error. *Econometrica*, 47, p.153- 163.

Johnson, W., Y. Kitamura and D. Neal (2000) Evaluating a Simple Method for Estimating Black-White Gaps in Median Wages. *American Economic Review*, 90, p.339-343.

Juhn, C. (1992) Decline of Labor Market Participation: The Role of Declining Market Opportunities. *Quarterly Journal of Economics,* 107, p.79-122.

Junh, C. (2003) Labor Market Dropouts and Trends in the Wages of Black and White Men. *Industrial and Labor Relations Review*, 56(4), p.643-662.

Little, R.J.A. and Rubin, D.B. (1987) *Statistical Analysis with Missing Data.* New York: John Wiley and Sons.

Machado, C. (2012) Selection, Heterogeneity and the Gender Wage Gap. IZA Discussion Paper No. 7005

Manski, C. F. (1989) Anatomy of the Selection Problem. *Journal of Human Resources*, 24, p.343-360.

Mincer, J. (1974) *Schooling, Experience and Earnings*. Columbia University Press, New York.

Neal, D. (2004) The Measured Black-White Wage Gap among Women is Too Small. *Journal of Political Economy*, 112, p.S1-S28.

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

Petersen, T. and Saporta, I. (2004) The Opportunity Structure for Discrimination. *American Journal of Sociology*, 109(4), p.852-901.

Petrongolo, B. and Olivetti, C. (2006) Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps. CEPR Discussion Paper No. 5506.

Puhani, P.A. (2000) The Heckman Correction for Sample Selection and its Critique. *Journal of Economic Surveys,* 14(1) p.53-68.

Rubin, Donald B. (1987) *Multiple Imputation for Nonresponse in Surveys, Wiley Series in Probability and Mathematical Statistics.* Wiley & Sons, New York.

Schafer, J.L. (1999) Multiple imputation: A primer. *Statistical Methods in Medical Research,* 8, p.3-15.

Schafer, J.L. and Olsen, M.K. (1998) Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective. *Multivariate Behavioral Research,* 33(4), p.545-571.

Vella, F. (1998) Estimating Models with Sample Selection Bias: A Survey. *Journal of Human Resources*, 33, p.127-169.

Appendix 1 – Mincer earnings functions

Table A1 – OLS estimates (Table 3 with details)

	Unadjusted GWG	Adjusted for workers' characteristics and workplace characteristics GWG			
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
		(1) + Education, age and experience	(2) + Contract	(3) + Occupation	(4) + Industry
Female	-0.125***	-0.173***	-0.175***	-0.160***	-0.175***
Secondary education		0.168***	0.173***	0.176***	0.147***
Tertiary education		0.731***	0.728***	0.684***	0.578***
Age		0.0140*	0.00751	0.0134*	0.0152**
Age^2		-0.0002***	-0.0001*	-0.0002**	-0.0002**
Experience		0.0156***	0.0130***	0.0109***	0.0098***
In the employee has a full-time contract			0.233***	0.194***	0.197***
Occupations (comparative group: elementary occupations)					
Armed forces				0.102***	0.0987***
Managers				0.142**	0.143**
Professionals				0.207***	0.193***
Technician professionals				0.0173	0.0041
Clerks				0.0666	0.061
Service and sales workers				0.0165	0.0344
Agriculture workers				0.00206	0.000508
Craft workers				-0.0305	-0.0143
Machinists				-0.0168	0.00722
Economic sectors (comparative group: industry)					
Agriculture					-0.0661
Mining					0.425***
Public sector					0.276***
Construction					0.116***
Services					0.0607**
Financial sector					0.424***
Other services					0.102***
Constant	4.249***	3.581***	3.537***	3.418***	3.324***
Observations	2,613	2,613	2,547	2,420	2,420
R-squared	0.011	0.259	0.277	0.31	0.356
*Source: Authors' calculations. *, ** and *** denote statistical significance at the 10, 5 and 1% level, respectively. Estimates are robust to heteroskedasticity.*					

Table A1 presents the results of Table 3 in detail. Aside gender, results suggest that return to education increases with the level of education. Additional year of age brings, on average, higher salary by about 1% with a turning point which is practically insignificant (turning point at about 350 years of age).

Similarly, additional year of experience is associated with about 1% higher wage. Full-time contract brings a wage premium of about 20% against a part-time contract.

Only wages in three groups of occupations: armed forces, managers and professionals differ than compared to the baseline group. These three have, on average, higher wages by 10%, 14% and 19%, respectively, than compared to the elementary occupations. On the other hand, all sectors except agriculture pay different wages: financial sector and mining pay about 42% higher wage each than industry, on average. Public sector pays about 28% higher wage than industry and so on.

***

Table A2 – The adjusted for characteristics and for selectivity gender wage gap, under different explanatory sets of the missing wage

Dependent variable: Log of the net hourly wage					
	The probit regression predicting missing wage includes:				
	All human capital char.	Only education	Education and age	Education, age and marriage/kids	Age, experi-ence and marriage/kids
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Female	-0.057***	-0.057***	-0.057***	-0.057***	-0.059***
Secondary education	0.037	0.0137	0.0756***	0.0137	-0.00415
Tertiary education	0.551***	0.562***	0.559***	0.562***	0.395***
Age	0.0055	0.00513	-0.00135	0.00513	0.00293
Age^2	-0.0001*	-4.57E-05	6.55E-05	-4.57E-05	-6.55E-06
Experience	0.0052***	0.00266**	0.00299***	0.00266**	0.00512***
Constant	3.863***	3.920***	3.955***	3.920***	3.968***
*Source: Authors' calculations. *, ** and *** denote statistical significance at the 10, 5 and 1% level, respectively. Estimates are robust to heteroskedasticity.*					