UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

School of Economics and Finance

ECON 7008A
Wage Differentials across Female-dominated, Male-dominated and Mixed-jobs - A Case for South Africa

Author:

Esther Bhosha

Student Number: 842902

Email address: 842902@students.wits.ac.za

Supervisor:

Professor Miracle Benhura

Date: 31 August 2020

Submitted in Support of Master of Economic Science (CCA11, Weighting 50%)

Contents

Abstract

One of the most commonly cited reasons to explain the gender wage gap is women's selection into relatively low paying jobs while men do the same in higher paying jobs. This, in turn, results in female-dominated and male-dominated jobs. Hence, this study investigates the extent to which female-dominated occupations pay less than male-dominated and mixed jobs in South Africa, as this has not yet received significant attention in the literature. Propensity Score Matching techniques, and PALMS and LMDSA 2015 datasets are utilised for the analysis. When confining the analysis to individuals aged 18-60 years, results for the overall sample show that male-dominated and mixed jobs indeed pay more than female-dominated jobs. On the contrary, being in female-dominated jobs pays more than being in a male-dominated job when the sample is generally restricted to men and particularly to those in the non-public sector. For men in the public-sector, being in a female-dominated job pays less than being in a mixed- job, while the reverse is the case for those in the non-public sector. Regarding women, female-dominated occupations tend to pay them less than male-dominated or mixed occupations. This is especially true in the non-public sector. This evidence suggests that the observation that female-dominated jobs pay less than other jobs holds when labour market segmentation is taken into account. Thus it is crucial for informing policy that aims at bringing wage equality and improving the experiences of women (and some men) in the labour force.

Keywords: wage disparities, male-, female-dominated, mixed occupations, propensity score matching

1.0 Introduction

The importance of studying wage gaps across female-, male-dominated and mixed occupations partly hinges on the gender pay gap and the associated quest for wage equality. The gender pay gap has received considerable scholarly attention for a number of decades. This has been spurred by a need to construct policies that encourage women's participation in the labour market. At a broad scale, there is a general consensus on the reduction of this earnings gap worldwide (Blau & Kahn, 1992, 2006; Selmi, 1999; Harkness, 1996; Zhan et al., 2008). Though numerous authors agree on the steady reduction of the gender earnings gap over the past decades, Mandel & Semynov (2005) conclude that this reduction is mostly experienced in developed welfare states and is attributed to their more egalitarian wage structures rather than to their family policies. Blau & Khan (2006) together with Gunderson (1989) further explain that despite a substantial increase in women's earnings relative to men's, the earnings gap has been shrinking at a decreasing rate and eventually reached a constant level. Various researchers have also suggested several reasons that explain the existence of the gender earnings gap. Most commonly cited ones include women's low education levels, shorter work hours, women's career interruptions, employers' expectations and perceptions of women's productivity, and gender differences in occupational choices (Polachek, 1981; Blau & Ferber*.,* 1991; Selmi, 1999; Bonjor & Gerfin, 2001; Drolet, 2001; Bertrand *et al*., 2010; Blau & Kahn, 2017). In turn, that gender differences in occupational choices explain the gender wage gap is partially based on the notion that the occupational choices culminate into differently paid female-dominated, male-dominated and mixed occupations.

The current study is an addition to literature on this strand of labour market inequalities as it investigates wage differentials across female-, male-dominated and mixed occupations in South Africa. This follows as the gender wage differential has been persistent in South Africa, and it is partly attributed to women's concentration in low paid jobs compared to those for men (c.f. Casale and Posel, 2011, Mosomi, 2019). Such female-dominated jobs are, hence, less paying in comparison to male-dominated and perhaps mixed jobs. To this author's knowledge, currently there is no existing South African study that has exclusively focussed on whether male-, female-dominated and mixed jobs undeniably have different pay structures such that women's crowding in female-dominated jobs indeed explains the gender pay gap. Hence, this study aims to quantitatively investigate the extent to which the assertion that female-dominated jobs in South Africa pay less than male-dominated jobs is true. Where possible, the results will inform the direction of measures to reduce such wage inequalities in African labour markets where such studies are currently scarce.

South Africa presents an interesting context for this topic due to its rich historical legacy of racial and gender-based segregation. This dates back to the Apartheid era and contributed to the contemporary spatial distribution of people and wealth (Mariotti & Fourie, 2014). The spatial segregation was also implemented in the form of education when, in 1953, the Bantu Education Act was passed by the government. It aimed to change the education system and curriculum of black Africans to suit low-wage labour that would minister the needs of the white economy (Nkomo, 1981). In subsequent years, other such laws were passed for other non-white racial groups. These had a significant impact on spatial distribution of factors such as employment, wealth and residential opportunities in South Africa. Regrettably, the legacy of these spatial segregation policies largely persists in post-apartheid South Africa (Seekings, 2010; Mariotti & Fourie, 2014).

Narrowing this historical legacy to South African women, studies show that women, black Africans in particular, faced an additional barrier of entry into the labour market in the form of a structural patriarchal system of power that men, both African and white, tirelessly fought to maintain through customary and statutory law (United Nations Centre Against Apartheid, 1978; Bond, 2010). These obstacles were less felt by women of other racial groups. In general, black women were labelled as both poor and uneducated (Hopfer, 1997) and for the subset who somewhat overcame the existing barriers imposed by the patriarchal system and the racial segregation features, their success in entering the job market was limited to low-skilled jobs that leaned more towards their innate abilities, for example, domestic work (Fester, 2000; Gaitskell *et al*. 1983, Cock, 1985; Landis 1975). Very few black women were able to enter the medical and teaching professions because of the financial burden of university fees and the racial-based prohibitive practices (United Nations Centre Against Apartheid, 1978; Martineau, R., 1997; Martin & Jarret-Kerr - South African History Online, 2019). The remaining fraction of black women remained in former homelands and rural areas practising subsistence agricultural production while men left their homes to find work in the cities and mines (South African History Online: Land, Labour and Apartheid). Black men dominated in different jobs such as those in the mining and agricultural sectors (South African History Online-Land, Labour and Apartheid). For the small fraction of well-educated black men, Apartheid legislation excluded them from most senior-level jobs while the same legislation authorised the "reservation" of many skilled jobs and managerial positions for white men primarily (Barnard, 2009). White women were most often employed in service industries and clerical positions while very few held managerial positions in both the private and public sector (Meer, 1985). This occupational sorting has culminated into female-, male-dominated and mixed jobs in South Africa. Nonetheless, the post-Apartheid South African government has embarked on some initiatives to improve the labour-market position of women. Some of these measures, discussed below, are indirectly linked to bringing wage equality between female- and male-dominated jobs and mixed occupations.

As a commitment to furthering women's economic position South Africa participates in treaties such as the Beijing Declaration and Platform for Action since 1995, the African Union Solemn Declaration on Gender Equality in Africa since 2000, and the United Nations Development Programme (UNDP). The mandate, amongst others, is to promote gender equality and the empowerment of women. The South African government has also passed some legislation such as the Employment Equity Act (1988) with the sore aim of ensuring that the labour-force expands in line with demographic categories. In addition, the government and other private organisations, for example Higher Education Resource Services South Africa (HERS SA), have committed to or have been formed with the sore aim of sponsoring female educational programmes. These sought to improve women's human capital and empower them towards being self-sustaining, relevant and competitive in the labour market. In regard to efforts which necessarily deal with wage equalisation between male- and female-dominated occupations, Mosomi (2019) finds that the minimum wage implementation has had a positive impact at the top of the income distribution. Legislative efforts such as the Employment Equity Regulation (2014) and ILO Equal Remuneration Convention ratified in 1995 attempt to ensure that work of equal value yields equal pay. This indirectly encompasses male-dominated, mixed and female-dominated occupations. Nonetheless, despite these efforts to promote the working conditions of women and bridge the wage gaps between female- and male-dominated or mixed occupations, the challenges still persist (International Women's Conference, 2017). In the third quarter of 2018, women only comprised 32 percent of managers in South Africa (Statistics South Africa, 2018). They also dominated domestic work, clerical and technician occupations whereas men dominated the rest, which tend to be relatively better paying.

Above statistics unsurprisingly reinforce the relevance of earlier South African literature on gender disparities in the labour market. For instance, Casale (2004) and Casale & Posel (2002, 2005) find that despite a notable increase in women's participation in the labour market since the mid-1990s, women continue to be dominant in low-paying and less secure employment. Mathur-Helm (2005) finds that reaching top level positions is still uncommon for South Africa's women, as its corporate environment is not yet ready to accept women as professional equals. Bhorat & Goga (2013) re-examine the gender wage gap using a more nuanced methodology (the recentred influence function) and find that, among other interesting results, the gender wage gap is higher at the bottom of the wage distribution than at the top. This somewhat suggests a wage penalty for women as they dominate in lowly paid jobs as compared to men. Consequently, this study aims to concretise this assertion by examining wage differentials across three occupational groups that are broadly defined by gender composition in South Africa.

The running hypotheses of the study are that:

i. In South Africa wage differentials persist in female- versus male-dominated occupations and mixed occupations after correcting for the relevant non-discriminatory controls.
ii. Male-dominated jobs pay more than female-dominated jobs
iii. Mixed jobs pay more than female-dominated jobs

The remainder of the study is structured as follows. Section 2 discusses theoretical and empirical literature that underpins this study. Section 3 discusses the methodology while data and descriptive statistics are presented in section 4. Section 5 discusses the results and section 6 gives a final analysis and conclusion of the study.

2.0 Female-Dominated, Male-Dominated and Mixed Occupations

As aforementioned, a commonly cited reason why women earn less than men is their concentration in low paying and low skill-requirement but relatively flexible jobs such as domestic work. In contrast, there is a relatively higher concentration of men in higher paying and more skilled jobs (Cross & Bagilhole, 2002; Bagilhole, 1994; Eccles, 1986). In this section, we draw a theoretical link between low (high) wages and female-(male) dominated jobs.

A. The theory of compensating wage differentials

The theory of compensating wage differentials or equalising difference can be adopted as a potential link between job composition and wages. It posits that all individuals are risk averse with women being relatively more risk averse than men (Jianakoplos & Bernasek, 1998; Bernasek & Shwiff, 2001). Thus women are more likely to choose relatively less risky jobs than men. Riskier jobs presumably pay higher wages than safer jobs to compensate workers for loss of the safety job amenity. As a result, that male-dominated jobs pay higher than female-dominated jobs is partly underlain by compensation for risk-taking.

B. Madden's Household Responsibility Theory

Madden's Household Responsibility Theory (1981) also referred to as the Spatial Entrapment Hypothesis (SEH) is another alternative theoretical link. This SEH describes women's traditional roles and responsibilities in the household, such as caregiving and day to day chores, which discourage them from finding occupations that require them to work far from their homes. Thus in general women may then have to settle or be concentrated in jobs with more employment stability and flexible working conditions such as teaching and domestic work (Lankau & Scandura, 1997; Kwan 1999; Rapino & Cooke, 2011; Madden, 1981) even if the jobs are low-paying. In some cases women may be overqualified or untrained for these jobs, and the incentives may be misaligned with their human capital endowments. This has been found to contribute towards worsening the gender wage gap (Jacobs, 1999; Himmelweit & Sigala, 2003). Thus, the SEH may partly contribute to women's concentration in low paying occupations leading to women's inferior position in the labour market relative to men's.

C. Ability and Physical Stamina

Ability and fortitude required in male-dominated versus female-dominated jobs are other potential factors for the earning differentials. This is partly linked to women's lower chances of enrolling in Science, Technology, Engineering and Mathematics (STEM) programmes than men. In turn women are less likely to work in STEM associated jobs which tend to be demanding and highly paid. Hill *et al*. (2010) discovered that among first-year college students, females voiced fewer intentions to major in STEM courses. Moreover at graduation, males outnumbered females in nearly every STEM field, and there was a further decline of women's participation in STEM related jobs. This observation was credited to factors related to ability which are propagated by social and environmental barriers. These include beliefs about intelligence, gender stereotyping, individual self-assessment, spatial skills, workplace bias and the college student experience. This finding can be extended to occupational fields such as plant and machine operation or extraction and building which require physical stamina, a characteristic more common in men compared to women. However, more recently we have observed a cultural shift in the family structure, division of household and non-household related work (Gunderson, 1989) in addition to more women becoming educated (Kabeer, 2010; Harkness 1996). This is inclusive of women's increased entrance in jobs formerly believed to be strictly for men e.g. bus driving. As a result some women are slowly being rescued from low-paying jobs.

D. Occupational Crowding Theory

Finally, we can also consider the gender discrimination theory of occupational crowding. This theory stems from the neo-classical concepts of demand and supply and explains that, as a result of discrimination, women are restricted to a smaller subset of occupations which consequentially become female-dominated. As a result, in these occupations, there is excessive supply of labour which lowers overall earnings (Bergmann 1974; Cotter *et al.,* 2003). Conversely, in the male occupations from which women are restricted from entering, there is a lower supply of labour hence an increase in the overall earnings.

Other non-empirical literature on social and psychological factors of why men and women self-select into female- and male-dominated occupations includes Frome *et al*., 2006; Evans & Steptoe, 2002; Cross & Bagilhole, 2002; Bradley, 1993; Jome and Toker, 1998 and Simpson, 2004. Bradley (1993) suggests that men in non-traditional jobs may be undermining historical gender relations by responding to transformations in the job market. In light of these transformations, Frome *et al*. (2006) show that despite social movements and increased efforts to open occupational doors to traditional male-jobs for women, in comparison to men, women still bear a higher burden of balancing career and family thus steering them away from more demanding male-dominated occupations in which their ambitions may lie. Cross *et al.* (2002) and Torre (2018) posit that men entering women dominated occupations may suffer lower wages and social status compared to male-dominated occupations, while, in contrast, Simpson (2004) finds that men in female-dominated occupations, though in minority, might have an advantage over women in the same broad class through employers' assumption that men are more career-orientated and possess enhanced leadership qualities. This ties to the phenomena known as "glass escalator" which refers to the observation that heterosexual white men are put to higher up positions when entering female dominated sex-segregated professions. Evans *et al*. (2002) focus on the dimension of health impairments on males in female-dominated jobs and females in male-dominated occupations and find that females in male-dominated occupations are likely to experience more anxiety among other observations.

2.1 Gender, Earnings and Job-Types in Africa

This empirical literature review serves to highlight that labour markets in African countries are replete with gender wage disparities. To some extent this gap has been associated with differences in men and women's job characteristics such as women's dominance in low-paid jobs than those of men. However, currently no enquiry has been made to ascertain the extent of wage differences across these broadly defined occupation groups. Yet the findings may be instrumental for strategies to reduce wage inequalities in general and gender wage pay gaps in particular. As such, the subject matter of gender earnings gap has received considerable attention in several studies that use African data. Earlier literature includes Knight & Sabot (1982) who investigate the gender earnings gap using 1971 data for Tanzanian manufacturing firms and find that this gap is attributed to differences in productivity linked characteristics between men and women. Using private and public sector data of Ethiopia, Uganda and Cote d'Ivore, Appleton *et al.* (1999) also find that this gap is a result of discrimination against females. Neitzert (1994) arrives to the same conclusions using Kenyan data while Agesa (1999) further refines this result by concluding that the gender earnings gap is more pronounced in urban than rural areas of Kenya. More recently, Nordman *et al.* (2011) find that gender earnings gap exist in all the seven West African capitals under their investigation[1]. Moreover, differences in sector allocation contribute, on average, one third of gender earnings gaps while differences in characteristics of men and women explain approximately 40% of the raw gender gap averaging across the cities.

Using the 2007 and 2010 Swaziland Labour Force Surveys, Brixiová & Kangoye (2016) find that women continue to have lower employment and labour force participation rates, and are dominant in low-productivity activities. Again using Kenyan data, Agesa *et al.* (2013) investigate the contribution of gender differences in characteristics or the return of these characteristics towards the gender pay gap. They find that gender differences in characteristics and the returns thereof widen the gender pay gap at lower end of wage distribution while the top distribution is only influenced by differences in characteristics. In addition, the drivers of gender differences in characteristics and their accompanying returns are driven by industry, occupation, higher education and region covariates. Comblon *et al.* (2017) use labour force data collected in Cameroon and Mali to explore gender differentials in labour market outcomes. They conclude that for both countries, gender wage differentials are predominantly driven by differences in education levels. Moreover, while men are more likely to be salaried workers, women are more often occupied in unpaid family work.

Arbache *et al.* (2010) use data for several African countries[2] spread across the continent and document that the existing gender wage gap is not only attributed to discrimination but to factors such as access to education and credit, cultural values and household duties and labour market conditions, for example firm and sector-level characteristics. Ntuli & Kwenda (2020) analyse

[1] Abidjan, Bamako, Cotonou, Dakar, Lome, Niamey and Ouagadougou

[2] Burkina Faso, Burundi, Côte d'Ivoire, Cameroon, Ethiopia, The Gambia, Ghana, Guinea, Kenya, Madagascar, Malawi, Mauritania, Mozambique, Nigeria, São Tomé and Príncipe, Sierra Leone, Uganda, Zambia

gender gaps in employment and wages from a global viewpoint of countries in sub-Saharan Africa. Their analysis attests to similar patterns and correlates of these gaps across sub-Saharan Africa. In another study, Magadla *et al.* (2019) investigates if working mothers earn less than non-mothers for the black African community of South Africa using National Income Dynamics Study (NIDS) data. They conclude that there is a motherhood wage penalty at lower levels of the income distribution while at high income levels, mothers earn higher than their child-free counterparts, especially if they are married. These studies explore various dimensions of an existing wage disparity. However, none of them explore the dimension of how the wage disparity intersects with male-dominated or mixed occupations versus female-dominated occupations.

Studies such as Arabsheibani & Manfor (2002), hint on the aforesaid link by posing the question of how Libyan women would fare in a male-dominated society. Yet even this does not necessarily answer the question of whether male-dominated jobs necessarily pay more than female-dominated jobs merely because of their classification. To our knowledge, there are but a few African studies which come closest to our subject of interest. Interestingly they are all South African cases which could be a result of the easier availability of labour-market data. Shepherd (2008) investigates the evolution of gender discrimination over the period 1995 to 2005. Her analysis takes the form of a logarithmic equation which expresses the difference between male and female wages being dependent on differences in male and female productive characteristics and discrimination proxied by gender difference in wage ratios. Amongst others, Shepherd finds that male-dominated industries paid better than female-dominated industries; and African men hold top-paying positions in female-dominated industries.

Another relevant study is Mosomi (2019) who investigates the evolution of the gender earnings gap using 1993-2015 Post-Apartheid Labour Market Series data. Using descriptive data trend analysis, Mosomi concludes that the South African gender wage inequality has been high because there were more women in low-paying occupations and that the wage gap at the top of the distribution does not only indicate discrimination but the type of work women do, which is often more administrative and less technical than occupations dominated by men. She also concludes that the implementation of the minimum wage has caused a reduction of the existing gap at the bottom of the income distribution. Finally, we can consider Ncube & Tregenna (2013) who partly investigate if increased employment of women in male- dominated industries necessarily lowers the earnings. They find that in almost all male-dominated industries, earnings reduce as more women enter these industries. Though these findings are undoubtedly in line with our current analysis, they also expose the dearth of rigorous empirical work that specifically investigates the extent to which occupational choices as captured by female-, male-dominated, mixed jobs lead to wage inequality in an African labour market. Importantly, unlike the studies discussed above, the current study narrows the scope of analysis to thoroughly focus on the earnings of male-, female-dominated and mixed occupations. In addition, we apply another arguably appropriate methodology, that is, Propensity Score Matching without restricting our analysis to black individuals as done by Shepherd.

3.0 Methodology

The study's ultimate goal is to compare hourly wages for individuals employed in female-dominated occupations versus 'similar' individuals employed in either male-dominated or mixed occupations. However we cannot escape the problem of selection bias that is presented with this study. The selection bias presents itself in two ways:

i. via selection of labour force into employment
ii. via selection into female-dominated, male-dominated and mixed jobs

In this particular context, we may regard the first source of selection bias negligible since we are concerned with comparing wage outcomes of people that are necessarily employed. This would not be the case if, for example, we were investigating the SEH phenomenon which requires that we consider both the employed and unemployed. The second source however is more worrisome and raises the need to choose a methodology which corrects for such biases. The Propensity Score Method (PSM) is arguably the most appropriate method given that it bases the comparison on individuals (in treatment versus control groups) with similar relevant characteristics or covariate factors. PSM is an attempt to mimic the randomized assignment to treatment and comparison groups by choosing for the comparison group those units that have similar propensities to the units in the treatment group; thus it belongs to the category of quasi-experimental methods satisfactory randomisation. Any differences between those individuals, for example in preferences for job amenities, are merely by chance and not systematic. Given that matching is performed based on similar observable characteristics, the selection problem is greatly minimised though perhaps not completely resolved due to the possible existence of unobservables which may affect the outcome (wages) or the shift from the outcome to occupational choice.

PSM rests on two main underpinnings:

i. Selection on Observables/Conditional Independence Assumption
ii. Overlap or Common Support

The first assumption allows for the creation of a counterfactual for the treated individuals but is however limited in that it rules out the possibility of selection on unobservable factors. The second assumption ensures that there is a positive probability that all covariate patterns are assigned to each treatment state. Rosenbaum & Rubin (1983) call these assumptions together 'strong ignorability' because of their highly restrictive nature. The method's ability to mimic randomisation combined with a reasonable number of observable characteristics (see descriptive statistics in Appendix) contained in the dataset enables us to reasonably argue for the satisfaction of the above-mentioned two fundamental conditions.

The PSM method encompasses the Logit (or Probit) Model which is used to calculate propensity scores before choosing a matching algorithm. The Logit Model is specified as:

$$fdomjob_i = \varphi_i + \beta x_i' + u_i \quad (1)$$

where (i) $fdomjob_i$ is the binary indicator taking the value "1" if one is employed in a female-dominated job otherwise "0" and (ii) x is a vector of control variables such as age in its linear and non-linear form, level of education, marital status, place of worker's residence, household-related factors such as household size, presence of children under the age of 14 in a household and a career-interruption factor proxied by maternal/paternal leave. The outcome variable, real earnings per hour variable, is calculated after merging the Post-Apartheid Labour Market Series (PALMS) and the Labour Market Dynamics South Africa (LMDSA) 2015 dataset. This merge allows us to combine monthly real earnings found in the PALMS dataset and daily hours worked found in the LMDSA which is then used to approximate hours per month and consequently hourly real wage[3]. The age variable is the individual's age; the inclusion of the age-squared stems from the evidence that occupational choice may vary over one's lifetime. Clark *et al.* (1996) empirically support the U-shaped relationship between age and job satisfaction. Extending this by assuming job satisfaction as a suitable proxy for occupational choice, age-squared can thus be used to account for occupational choice. Education is measured as the number of years of education attained; since marital statuses are categorised as "Married", "Living together like Husband and wife", "Widow/widower", "Divorced or Separated" or "Never Married", for simplicity sake, we combined the categories of "Married" and "Living together like Husband and wife" to both fall under the married category while the rest fall under the unmarried category. The dummy variable "maternity" indicates whether one's job makes provision for maternity or paternity leaves; the urban-rural categorical variable, accounts for an individual's area of residence as being in an urban, traditional, farming or mining area. Finally the number of children variable NC is a set of household structure variables with i = 1, 2. Specifically, NC indicates household with no children; NC1 indicates presence of children aged below 6 years; and NC2 indicates presence of children aged 7-14 years in a household.

The inclusion of these variables is supported by standard theory discussed in earlier paragraphs and the work of other researchers. For example, the SEH motivates the inclusion of household-related duties such as care-giving to the elderly (pensioners) or children young enough to require care-giving in households in which they are present. It is more likely that need for more care-giving or generally increased household tasks necessitates performers of these tasks, usually women, to acquire jobs with flexible working hours. Given this need, the probability of getting employment in a female-dominated occupation is likely to be higher. Education is also important given that one of the cited reasons to explain why women are concentrated in lower paying jobs is their lower educational attainment. Besides the theoretic reasoning, other related work apply the given control variables. Closely related to the current study is the work of Brown *et al.* (1980) who use multinomial logit model to predict occupational choices for individuals based on an

[3] To approximate hourly real wage, we divide the number of weeks in the year 2015 (50 weeks) by 12 (months in a year) and multiply the outcome (4.167 weeks per month) by the average working hours per week. The outcome is then divided into the monthly real earnings to obtain the hourly real wage.

individual's characteristics and qualifications. Other applications include Beller (1982), Irfan *et al.* (2013), Mon (2017), Shepherd (2008), Lehrer & Stokes (1985) and Agesa *et al.* (2013). Finally, the inclusion of paternal or maternal leave dummy variable is motivated by the afore-mentioned literature which identifies women's career interruptions such as giving birth, as a potential cause of entry into female-dominated jobs which are more likely to present such opportunities.

Other studies of this nature include a variable that accounts for earnings as part of the explanatory variables in the logit model; however this inclusion may be viewed as problematic given reverse causal relationship inferred. On one hand, earnings may impact the outcome variable that is, being in a female-dominated job while on the other hand being in a female-dominated job may impact your earnings. Addressing this challenge is not a trivial task but for simplicity this study will not include the earnings regressor. Moreover, given that earnings form our final outcome variable, our approach is justified.

For the matching exercise, the treatment group will comprise of individuals working in female-dominated jobs. There are two separate control groups, i.e., individual workers in male-dominated and mixed jobs. Matching algorithms to be considered include the Nearest Neighbour (NN), Caliper and Radius (C&R), Local Linear Regression, Kernel, and Stratification (Gertler *et al.,* 2016*;* Khandker *et al.,* 2009*;* Becker & Ichino, 2002; Baser, 2006; Caliendo & Kopeinig, 2008; Baum, 2013). As the name suggests, the NN algorithm matches an individual in the treatment group with a control group participant whose propensity score is closest. If for each individual in the treatment group there is a single match in the control group, then it is said to be a 1:1 matching otherwise a 1:n matching. The NN method, though simple to carry out, has the disadvantage that not every NN is necessarily the best match. One way to resolve this is to employ a modification which is the C&R matching method. The modification entails defining a tolerance on the maximum radius distance between an individual in the treatment group and one in the control group. Austin (2011) recommends using calipers of width equal to 0.2 of the standard deviation (log likelihood) of the logit of the propensity score.

As an alternative to Radius matching, which rules out matches beyond the threshold of the caliper, the Kernel and Local-Linear methods are nonparametric methods which use all units in the control group to construct a match for each programme participant. Higher weight is placed on observations or persons who are in close proximity in terms of the propensity scores. Conversely lower weight is placed on observations that are not in close proximity to each other. The drawback however lies in that, since the Kernel algorithm does not discard any information or matches, there is a possibility of using "bad" observations as matches (Caliendo & Kopeinig, 2008) in addition to the trade-off between lower variance and increased bias. A similar trade-off is encountered in defining the width of the C&R in the C&R matching algorithm. Finally, the Stratification algorithm involves partitioning the common support into a set of intervals (or strata) and matching the treatment and control subjects within each strata. The drawback lies in evaluating a "correct" number of intervals. In general, each matching method has its advantages and disadvantages and as Caliendo & Kopeinig (2008) demonstrate, no single algorithm will dominate in all data situations; consequently, the matching algorithms' performance may vary from case to case. Therefore this study considers all these methods to assess robustness of the findings.

The outcome of interest is the average treatment effect on the treated (ATT) which is defined as the mean difference in outcomes between treated individuals and their matched counterparts in the control group. For this study, this definition translates to the mean difference in log hourly wages between individuals employed in female-dominated occupations and their 'twins' employed in either male-dominated or mixed occupations. The ATT can also be arithmetically expressed as follows: given an outcome Y, where Y_1 is the outcome of the treated and Y_0 the outcome of the untreated, and a dummy variable D which takes the values "1" for being treated or "0" otherwise, over X covariate factors with propensity score P(X),

$$\tau_{ATT} = E_{P(X)|D=1}\{E[(Y_1 \mid D=1,\ P(X)] - E[(Y_0 \mid D=0,\ P(X)]\} \qquad (2)$$

Our primary focus on the ATT lies on the coefficient sign, the magnitude which is indicative of the penalty's size, and statistical significance.

Our model also raises the need to define thresholds that determine the category under which an occupation falls, and consequently an individual in the labour market. Therefore, we will define this threshold to be at least two-thirds. If for example, in the occupation for teaching, we find 30 teachers in our dataset, where 20 of them are female, this occupation will be labelled as female-dominated and have the value "1" in the Logit model. If however, of the 30 teachers, only 10 are female, then the occupation will be labelled as male-dominated. If an individual falls in neither of these categories, this automatically leads them to fall under a mixed occupational group. These categories and the associated hourly wages are used to analyse whether male- and female-dominated or mixed jobs pay differently. In general, various authors use different cut-off points in their definition of male- or female-dominated jobs. Tophoven *et al*. (2015) use at least 80%; Hayes (1986) uses the at least 70% threshold; and Corcoran *et al.* (1984) use at least 60% threshold.

This study considers the nine major occupational categories listed below together with their respective sub-major occupations:

i. **Legislators, Senior Officials and Managers**
   - legislators and senior officials
   - Corporate Managers
   - General Managers.

ii. **Professionals**
   - Physical, mathematical and engineering science professionals
   - Life science and health professionals
   - Teaching professionals
   - Other professionals.

iii. **Technical and associate professionals**

- Natural and engineering science associate professionals
- Life science and health science and health associate professionals
- Teaching associate professionals
- Other associate professionals

iv. **Clerks**
- Office clerks
- Customer services clerks

v. **Service workers and shop and market sales workers**
- Personal and protective services workers
- Models, salespersons and demonstrators

vi. **Skilled agricultural and fishery worker**
- Market-orientated skilled agricultural and fishery workers
- Subsistence agricultural and fishery workers

vii. **Craft and related trades workers**
- Extraction and building trades workers
- Metal, machinery and related trades workers
- Precision, handicraft, printing and related trades workers
- Other craft and related trades workers

viii. **Plant and machine operators and assemblers**
- Stationary-plant and related operators
- Machine operators and assemblers
- Drivers and mobile plant operators

ix. **Elementary Occupations**
- Elementary sales and services occupations
- Agricultural, fishery and related labourers
- Labourers in mining, construction, manufacturing and transport

In the data, the sub-major categorisation can be done using the first-two digits of the occupations variable, to decode the sub-groupings. Moreover, in instances where, given any major group, the number of individuals in a sub-major group is less than 20, we combine the group in question with another closely related sub-major group in the same major group.

Finally we reiterate and discuss other potential weakness of the methodology. Aforementioned, PSM does not adequately account for unobservable factors such as one's risk tolerance and other inherited or innate characteristics which may be unidentifiable or immeasurable. The unobservable factors may bias estimation of the treatment effect however the method works to minimise the bias by comparing outcomes of treated and control subjects who are as similar as possible. Thus PSM reduces but does not completely eliminate but minimises the bias generated by unobservable confounding factors. Furthermore, the general rule regarding observable covariate factors is to include all available variables which are thought to be related to outcomes from empirical or theoretical research and which differ across groups. However the more covariates are included in the model, the more degrees of freedom the model loses. Reichardt *et al*. (1980) propose limiting analyses to variables that have theoretical significance to treatment selection and those previously proven to predict outcomes. Rosenbaum's (2002) proposes including all pre-treatment covariates whose group differences met a low threshold for significance, for example, $|t|>1.5$ in the logistic regression. Moreover, variables should be stable over time, or deterministic, or, where possible, measured before participation, to avoid becoming confounded with outcomes. However, the dataset at hand does not sufficiently allow us to control for predetermined characteristics.

4.0 Data and Descriptive Analysis

This cross-sectional analysis uses the Post-Apartheid Labour Market Series (PALMS) data which covers the years 1993-2017 complemented with the Labour Market Dynamics in South Africa (LMDSA) data produced by Statistics South Africa. Thanks to the Data First Resource Unit at the University of Cape Town for providing PALMS data. Specifically we use the 2015 data because of the greater availability of variables of interest. Our dataset is appropriate since it contains comprehensive information and details that are crucial for this study. This includes individuals' information on gender, age, marital status, education, employment status, monthly salary, main occupation, number of hours worked per week in an individual's main occupation, among others.

The raw 2015 PALMS data comprises of 40,809 households with 286,782 individuals of which 52.43 percent are female. Of the 286,782 individuals, 115,281 make up the labour force with only 76,489 employed while the rest are either unemployed or discouraged job seekers. 39,770 of the employed are male while 36,719 are female with the average age of the employed approximating to 39 years for men and 40 years for women. To improve on the quality of our analysis, we confine our data to the age range of 18 to 60 and exclude unemployed individuals and individuals with missing values for the monthly real earnings. We also create a variable for total hours worked in a month after eliminating outliers and thus use this variable to calculate real earnings per hour. This reduces the dataset to approximately 17,000 households and 31,000 individuals.

For the sake of tracing the trends of male- or female-dominated jobs and if this might have changed overtime in support of factors such as the progressive increase in number of educated women or the changes in beliefs surrounding gender-based division of labour, we will present a comparison using data for 2008. Importantly, the data allows us to identify different occupations (or sub-major occupational groups) within Major occupational groups using the first two digits of the major-group occupations list[4]. For an example, under the major group "Legislators, senior officials and Managers", individuals with the first two digits "11" belong to the occupation or sub-major group of Legislators and senior officers, while the digits "12" account for Corporate Managers.

[4] Classifications adopted from International Standard Classification of Occupations (ISCO-88). The same classifications are used for both the public and non-public sectors.

The sub-major categorisation which is done using the first-two digit method to decode the sub-groupings reveals the following:

TABLE 1: Categorised Subgroup Occupations for 2015

Male-dominated sub-group occupations:	**Female-dominated sub-group occupations:**	**Mixed sub-group occupations:**
• General Managers • Physical, mathematical and engineering science professionals • Natural and engineering science associate professionals • Market-orientated skilled agricultural and fishery workers • Extraction and building trades workers • Metal, machinery and related trades workers • Precision, handicraft, printing and related trades workers • Stationary-plant and related operators • Drivers and mobile plant operators • Agricultural, fishery and related labourers • Labourers in mining, construction, manufacturing and transport	• Life science and health science and health associate professionals • Teaching associate professionals • Office clerks • Customer services clerks • Elementary sales and services occupations	• Legislators and senior officials • Teaching professionals • Corporate Managers • Life science and health professionals • Other professionals. • Other associate professionals • Subsistence agricultural and fishery workers • Personal and protective services workers • Models, salespersons and demonstrators • Other craft and related trades workers • Machine operators and assemblers

As noted in earlier paragraphs, it is of interest to observe if this categorisation may have changed over time. We conduct this analysis by comparing to the 2008 dataset and results are presented in Table 2 below.

TABLE 2: Categorised Subgroup Occupations for 2008

Male-dominated sub-group occupations:	**Female-dominated sub-group occupations:**	**Mixed sub-group occupations:**
• General Managers • Corporate Managers	• Teaching professionals • Life science and health professionals	• Other professionals. • Other associate professionals • Office clerks

• Legislators and senior officials • Physical, mathematical and engineering science professionals • Natural and engineering science associate professionals • Market-orientated skilled agricultural and fishery workers • Extraction and building trades workers • Metal, machinery and related trades workers • Precision, handicraft, printing and related trades workers • Stationary-plant and related operators • Drivers and mobile plant operators • Agricultural, fishery and related labourers • Labourers in mining, construction, manufacturing and transport	• Life science and health science and health associate professionals • Teaching associate professionals • Customer services clerks • Elementary sales and services occupations	• Personal and protective services workers • Models, salespersons and demonstrators • Subsistence agricultural and fishery workers • Other craft and related trades workers • Machine operators and assemblers

Evidently, there were no major changes in sub-occupations that fall under male-, female-dominated or mixed occupations over time. On the one hand, some sub-major groups have maintained male dominance, for example Labourers in mining, construction, manufacturing and transport; Extraction and building trades workers; Metal, machinery and related trades workers; Stationary-plant and related operators; and Drivers and mobile plant operators. On the other, Life science and health science and health associate professionals; Customer services clerks; and Elementary sales and services occupations have maintained female dominance. In addition, an increase in the number of women that work as Legislators and senior officials is observed.

Table 3 below displays summary statistics for hours worked and wages across broadly defined male-, female-dominated and mixed occupations (before the matching analysis). Full tables of Summary statistics by occupation group, and for the overall sample are presented in Section B of the Appendix, the variable definitions are presented in Section A of the Appendix.

TABLE 3: Summary Statistics Across Male-dominated, Female-dominated and Mixed Occupations

Variable	Number of Observations	Mean/proportion	Std. Dev.	Min	Max
Female-Dominated Occupations					
Average hours worked per week	9,318	40.00011	12.58092	1	84
Wages per hour	9,318	43.65071	64.53526	5.021605	966.6589
Log of Wages per hour	9,318	3.23617	.9654075	1.61375	6.873846
Male-dominated Occupations					
Household size	11,745	6.714432	6.097515	1	64
Average hours worked per week	11,745	43.72771	11.8785	1	84
Wages per hour	11,745	39.58796	62.03844	5.000984	977.3953
Log of Wages per hour	11,745	3.152735	.917017	1.609635	6.884891
Mixed Occupations					
Average hours worked per week	9,917	45.71645	12.61239	1	84
Wages per hour	9,917	66.03255	95.90459	5.012305	992.2997
Log of Wages per hour	9,917	3.542748	1.096821	1.611896	6.900025

Interestingly, female-dominated occupations have on average more working hours per week and higher wages per hour than male-dominated occupations; the opposite is true when comparing female-dominated jobs with mixed jobs. Table 4 below reports averages of real monthly earnings across gender and hourly wages which are not disaggregated by sex, for all occupations. It is apparent that within the various occupational groups, men generally have higher real monthly earnings than women. Moreover, the average real monthly earnings of men in female-dominated occupations are higher than those of men in male-dominated occupations. In contrast, average earnings of women in female-dominated occupations are lower than those for women in male-dominated jobs. A comparison of average wages between female-dominated and mixed-occupations shows that the former generally have lower real monthly earnings. As in Table 3, the average hourly wage is relatively higher in female-dominated jobs when compared to the male-dominated. However female-dominated jobs fare worse than mixed jobs.

TABLE 4: Mean Real-Earnings and Wages-per-hour for all occupations

Occupation	**Real monthly earnings**	**Real monthly earnings**	**Log of Hourly Wages**
Female-Dominated Occupations			
	MALE	**FEMALE**	
Life science and health science and health associate professionals	11659.61	11512.36	3.7412
Teaching associate professionals	13682.83	11795.83	4.027472
Office clerks	11086.99	9936.091	3.683159
Customer services clerks	10355.98	6289.169	3.309619
Elementary sales and services occupations	5245.361	2973.383	2.879846
Average	8148.099	5930.35	3.5282592
Male-Dominated Occupations			
General Managers	13722.52	11894.11	3.690924
Physical, mathematical and engineering science professionals	25732.78	23076.4	4.544812
Natural and engineering science associate professionals	14501.01	11676.17	3.813112
Market-orientated skilled agricultural and fishery workers	8877.422	5977.173	3.212955
Extraction and building trades workers	6324.557	6383.03	3.19074
Metal, machinery and related trades workers	9186.016	5924.954	3.425279
Precision, handicraft, printing and related trades workers	5748.426	7143.16	3.160711
Stationary-plant and related operators	7930.117	5452.982	3.377017
Drivers and mobile plant operators	6595.422	5817.746	3.125528
Agricultural, fishery and related labourers	2928.265	2479.753	2.644804
Labourers in mining, construction, manufacturing and transport	5594.153	4544.429	3.023491
Average	6990.332	6258.102	3.382670273
Mixed Occupations			
Legislators and senior officers	27360.74	19012.81	4.369973
Corporate Managers	21330.24	18752.27	4.240543
Life science and health professionals	29133.68	18538.21	4.402335
Teaching professionals	20976.08	18401.49	4.4952
Other professionals	21458.08	18398.32	4.349859
Other associate professionals	13820.32	12640.74	3.797202
Personal and protective services workers	7013.356	6359.352	3.134713
Models, salespersons and demonstrators	7224.503	6761.896	3.100443
Other craft and related trades workers	5455.003	3751.884	2.998629
Machine operators and assemblers	7227.677	5347.592	3.174033

Average	11666.85	10850.11	3.806293

The summary statistics above may be regarded as preliminary evidence for the differences in pay structures in male-, female-dominated and mixed occupations. Hence, we proceed to further unpack these differentials using Propensity Score Matching analysis.

5.0 Results

Subsequent to performing several simulations of different covariate combinations or discriminatory factors, our final model was composed of the following covariates as this combination better satisfied the balancing properties and displayed sufficient common support relative to other covariate combinations: household size, age, marital status, level of education, maternity/paternity dummy variable, a categorical variable of the presence of children below the age of 7 and presence of children between the ages of 7 and 14, both relative to households with no children under the age of 14 and finally the rural/urban worker's residence variable. In most instances, we found the 5 different matching algorithms used herein yielding analogous results.

5.1 Balancing Property and Common Support

For the sake of brevity, some results for the logit models and balancing tests followed by the complementing common support graphs are presented in Sections C and D of the Appendix. The analysis is conducted by running the 5 algorithms on 9 different samples of data namely: pooled data (that is, including both men and women), two samples of men only and women only; two samples of the public and non-public sectors; two samples of men in the public- and non-public sectors, and finally two samples of women in the public- and non-public sectors. This sample breakdown was done for both the comparison of female- versus male-dominated occupations and female-dominated versus mixed occupations. Importantly, the breakdown between public and non-public sector allow for a more nuanced and thorough investigation given the different pay structures that the two sectors normally present. The non-public sector which is usually driven by market forces rather than a need for equity, is expected to present higher disparities, while in the public sector, due to easier enforcement of employment equity regulations and being strong advocates of an all-inclusive work-place environment, these disparities are expected to be almost non-existent.

In determining whether or not the balancing property is satisfied we consider the t-statistic of each covariate and the p>chi2 of the joint model. The t-statistic must show insignificance i.e. less than the rule of thumb value of |2| as this implies no significant difference between the average of the treatment and control group for each covariate. If this is true for most if not all covariates, this would be translated into a joint insignificant p>chi2 value, i.e. a value greater than 10% at predetermined levels of significance. These conditions were met by all models with the exception of the sample of women in the public sector under the comparison of female- versus male-dominated occupations. In section D of the Appendix, the graphs for common support before and after matching indicate that the condition for common support is satisfied for all cases. In some cases the graphs show an almost complete overlap of the treatment and support curves after matching which is desirable. Given that our model satisfies the balancing properties and has sufficient common overlap, we can confidently proceed to analyse and interpret our findings.

5.2 Comparison of Male-Dominated Versus Female-Dominated Occupations

Tables 5.1, 5.2 and 5.3 below report PSM results for male-dominated versus female-dominated occupations.

TABLE 5.1: Comparison for Pooled Data and Split by Gender

Panel A: Pooled						
Matching Algorithm	Sample Size of Treated	Treated	Control	ATT	SE	t-statistic
Nearest Neighbour	9,000	3.219	3.257	-0.039*	0.018	-2.110
Radius Caliper	8,973	3.217	3.267	-0.051*	0.015	-3.420
Kernel	9,000	3.219	3.263	-0.044*	0.014	-3.100
Local linear regression	9,000	3.219	3.277	-0.059*	0.022	-2.700
Stratification	9,032	-	-	-0.047*	0.015	-3.138
Panel B: Male Only						
Nearest Neighbour	3,575	3.388	3.317	0.071*	0.025	2.830
Radius Caliper	3,559	3.385	3.319	0.065*	0.021	3.160
Kernel	3,575	3.388	3.323	0.065*	0.020	3.220
Local linear regression	3,575	3.388	3.337	0.051***	0.030	1.730
Stratification	3577	-	-	0.058*	0.020	2.872
Panel C: Female Only						
Nearest Neighbour	5,381	3.109	3.096	0.013	0.032	0.410
Radius Caliper	5,344	3.102	3.132	-0.029	0.027	-1.090
Kernel	5,381	3.109	3.133	-0.023	0.025	-0.920
Local linear regression	5,381	3.109	3.158	-0.048	0.037	-1.290
Stratification	5390	-	-	-0.037	0.028	-1.344

Note: the stratification algorithm code spills an already calculated ATT.

TABLE 5.2: Comparison across Public & Non Public Sector

Panel A: Public Sector						
Matching Algorithm	Sample Size of Treated	Treated	Control	ATT	SE	t-statistic
Nearest Neighbour	1,960	3.651	3.549	0.102	0.057	1.800
Radius Caliper	1,960	3.654	3.590	0.064	0.051	1.260
Kernel	1,960	3.651	3.584	0.067	0.047	1.420
Local linear regression	1,960	3.651	3.588	0.063	0.065	0.970
Stratification	1,960	-	-	0.068	0.050	1.368
Panel B: Non-Public Sector						
Nearest Neighbour	7,066	3.095	3.185	-0.091*	0.018	-4.800
Radius Caliper	7,066	3.094	3.193	-0.099*	0.015	-6.600
Kernel	7,071	3.098	3.187	-0.089*	0.014	-6.120
Local linear regression	7,071	3.097	3.185	-0.087*	0.022	-3.900
Stratification	7,071	-	-	-0.088*	0.015	-5.936
Note: *-1 percent level of significance **-5 percent level of significance ***-10 percent level of significance						

Note: the stratification algorithm code spills an already calculated ATT.

TABLE 5.3: Comparison across Public, Non Public Sector and Gender

Panel A: Male and Public Sector						
Matching Algorithm	Sample Size of Treated	Treated	Control	ATT	SE	t-statistic
Nearest Neighbour	704	3.682	3.591	0.091	0.075	1.210
Radius Caliper	691	3.678	3.612	0.065	0.066	1.000
Kernel	704	3.682	3.602	0.080	0.062	1.290
Local linear regression	704	3.682	3.596	0.086	0.088	0.970
Stratification	719	-	-	0.101	0.071	1.420
Panel B: Male and Non Public Sector						
Nearest Neighbour	2,891	3.318	3.251	0.067*	0.027	2.510
Radius Caliper	2,884	3.317	3.253	0.064*	0.022	2.930

Kernel	2,891	3.318	3.254	0.064*	0.021	2.970
Local linear regression	2,891	3.318	3.257	0.060*	0.031	1.950
Stratification	2892	-	-	0.062*	0.021	2.943
Panel C: Female and Public Sector						
Nearest Neighbour	1,181	3.618	3.498	0.120	0.101	1.200
Radius Caliper	1,176	3.624	3.523	0.101	0.093	1.080
Kernel	1,181	3.618	3.519	0.099	0.086	1.150
Local linear regression	1,181	3.618	3.552	0.066	0.110	0.600
Stratification	1212	-	-	0.088	0.081	1.086
Panel D: Female and Non-Public						
Nearest Neighbour	4,227	2.951	3.033	-0.082*	0.034	-2.410
Radius Caliper	4,210	2.951	3.040	-0.090*	0.029	-3.130
Kernel	4,227	2.951	3.050	-0.099*	0.026	-3.750
Local linear regression	4,227	2.951	3.063	-0.112*	0.039	-2.860
Stratification	4,240	-	-	-0.104*	0.029	-3.611
Note: *-1 percent level of significance **-5 percent level of significance ***-10 percent level of significance						

Note: the stratification algorithm code spills an already calculated ATT.

In Table 5.1, pooling both men and women, the ATT is negative and statistically significant at the 1 percent level across the matching algorithms. The ATT ranges between -0.059 and -0.039 hourly log wage points. This result is consistent with the hypothesis that wage differentials persist in female- versus male-dominated occupations and male-dominated occupations pay more than female dominated occupations. Restricting the analysis to men only shows that men in female-dominated occupations are paid more than men in male-dominated occupations with an ATT ranging from 0.051 to 0.071 hourly log wage points. This finding is consistent with Simpson (2004). The ATT for women only is not statistically significant at any of the predetermined levels of significance. In Table 5.2, the ATT for individuals employed in the public sector is statistically insignificant however in the non-public sector, it is significant at the 1 percent level ranging between -0.099 and -0.087 hourly log wage points. This result supports the hypothesis that male-dominated occupations pay more than female dominated occupations. In Table 5.3, the ATT for men in the public sector has no statistical significance while for men in the non-public sector, the ATT is both positive and statistically significance at the 1 percent level with a narrower range between 0.060 and 0.067 hourly log wage points. For women in the public sector, the ATT remains insignificant however in the non-public sector, women in female-dominated occupations earn less than those in male-dominated occupations. The ATT ranges between -0.112 and -0.082 and is significant at the 1 percent level of significance.

5.3 Comparison of Mixed Versus Female Dominated

Tables 5.4, 5.5 and 5.6 below report the PSM results for mixed versus female-dominated occupations.

TABLE 5.4: Comparison for Pooled Data and Split by Gender

Panel A: Pooled						
Matching Algorithm	Sample Size of Treated	Treated	Control	ATT	SE	t-statistic
Nearest Neighbour	9,028	3.226	3.316	-0.090*	0.020	-4.460
Radius Caliper	9,022	3.226	3.329	-0.103*	0.017	-6.120
Kernel	9,028	3.226	3.318	-0.092*	0.016	-5.600
Local linear regression	9,028	3.226	3.308	-0.082*	0.024	-3.410
Stratification	9,028	-	-	-0.096*	0.016	-6.028
Panel B: Male Only						
Nearest Neighbour	3,555	3.380	3.401	-0.021	0.029	-0.730
Radius Caliper	3,550	3.381	3.406	-0.026	0.024	-1.060
Kernel	3,555	3.380	3.406	-0.025	0.024	-1.060
Local linear regression	3,555	3.380	3.386	-0.006	0.033	-0.180
Stratification	3,555	-	-	-0.031	0.023	-1.369
Panel C: Female Only						
Nearest Neighbour	5,509	3.104	3.208	-0.103*	0.028	-3.640
Radius Caliper	5,494	3.106	3.216	-0.110*	0.025	-4.470
Kernel	5,509	3.104	3.200	-0.095*	0.024	-4.030
Local linear regression	5,509	3.104	3.212	-0.107*	0.033	-3.270
Stratification	5,511	-	-	-0.105*	0.022	-4.696

TABLE 5.5: Comparison across Public & Non Public Sector

Panel C: Public Sector						
Nearest Neighbour	1,980	3.649	3.789	-0.139*	0.049	-2.860
Radius Caliper	1,964	3.651	3.751	-0.100*	0.042	-2.360
Kernel	1,964	3.649	3.755	-0.105*	0.040	-2.610
Local linear regression	1,985	3.649	3.744	-0.094***	0.056	-1.671
Stratification	1,964	-	-	-0.106*	0.044	-2.428
Panel D: Non-Public Sector						
Nearest Neighbour	6, 965	3.093	3.203	-0.111*	0.022	-4.991
Radius Caliper	6, 927	3.094	3.205	-0.111*	0.019	-5.923
Kernel	6,950	3.093	3.204	-0.112*	0.018	-6.221
Local linear regression	6,950	3.0927	3.220	-0.127*	0.026	-4.860
Stratification	6,965	-	-	-.0117*	0.017	-6.806
Note: *-1 percent level of significance **-5 percent level of significance ***-10 percent level of significance						

TABLE 5.6: Comparison across Public, Non Public Sector and Gender

Panel A: Male and Public Sector						
Matching Algorithm	Sample Size of Treated	Treated	Control	ATT	SE	t-statistic
Nearest Neighbour	744	3.684	3.771	-0.086	0.071	-1.230
Radius Caliper	730	3.693	3.795	-0.102***	0.061	-1.660
Kernel	744	3.684	3.822	-0.138*	0.059	-2.320
Local linear regression	744	3.684	3.804	-0.120	0.081	-1.480
Stratification	750	-	-	-0.168*	0.063	-2.688
Panel B: Male and Non Public Sector						
Nearest Neighbour	2,784	3.330	3.266	0.064*	0.032	2.030
Radius Caliper	2,771	3.331	3.273	0.058*	0.027	2.160
Kernel	2,784	3.330	3.281	0.049*	0.026	1.890

Local linear regression	2,784	3.330	3.277	0.054*	0.036	1.480
Stratification	2,788	-	-	0.045*	0.039	2.100
Panel C: Female and Public Sector						
Nearest Neighbour	1,213	3.615	3.649	-0.034	0.070	-0.480
Radius Caliper	1,212	3.615	3.662	-0.048	0.062	-0.770
Kernel	1,213	3.615	3.673	-0.058	0.057	-1.020
Local linear regression	1,213	3.615	3.691	-0.076	0.078	-0.980
Stratification	1,216	-	-	-0.050	0.060	-0.834
Panel D: Female and Non-Public						
Nearest Neighbour	4,266	2.962	3.154	-0.192*	0.031	-6.140
Radius Caliper	4,227	2.964	3.131	-0.167*	0.027	-6.170
Kernel	4266	2.962	3.130	-0.168*	0.026	-6.570
Local linear regression	4,266	2.962	3.150	-0.188*	0.037	-5.080
Stratification	4,266	-	-	-0.168 *	0.024	-6.984
Note: *-1 percent level of significance **-5 percent level of significance ***-10 percent level of significance						

In Table 5.4, using pooled data, the ATT is negative and statistically significant at 1 percent level and ranges from -0.103 to -0.082 hourly log wage points. This supports the hypothesis that female-dominated occupations pay less in comparison to mixed-occupations. The ATT for men bears no statistical significance however for women, the ATT is negative and significant at the 1 percent level ranging from -0.110 to -0.095 hourly log wage points. In Table 5.5, the ATT for the public sector is negative and statistically significant at the 1 and 10 percent levels with a range between -0.139 and -0.094 hourly log wage points. Similarly, the non-public sector has a negative and statistically significant ATT which ranges from -0.127 to -0.111 hourly log wage points. For men in the public sector (Table 5.6), the significant ATT is negative and ranges between -0.168 and -0.102. This finding supports the claims of Cross *et al.* (2002) and Torre (2018). In contrast however, the ATT for men in the non-public sector is both positive and significant ranging from 0.045 to 0.064; this is supportive of Simpson (2004). The ATT for women in the public sector is not statistically significant at the predetermined levels of significance. In the non-public sector, it is both negative and statistically significant at the 1 percent level ranging from -0.192 to -0.167 hourly log wage points.

6.0 Final Analysis and Conclusion

Calculating at the lower bounds of the ATT range, the results above numerically suggest that male-dominated occupations pay approximately R1.0397 per hour[5] more than female-dominated occupations. This amounts to a weekly penalty of approximately R44.7101[6] which then yields a monthly penalty of approximately R187[7]. If this numerical reasoning is applied to all ATT (in monthly wages), men in female-dominated jobs are paid at least R189 more than men in male-dominated jobs. In the non-public sector, the penalty amounts to approximately R198, and for men only, the penalty amounts to approximately R190. As Simpson (2004) posits, men in female-dominated occupations may benefit from the employers' belief of their greater career-orientation and superior leadership qualities, resulting in higher earnings. As for women in the non-public sector, we observe that women in female-dominated occupations receive approximately R195 less than women in male-dominated occupations. On the public-sector front, the disparity of earnings in male- versus female-dominated occupations is less transparent for both men and women as the ATT is insignificant. This may be indicating more efficient administering of policies which curb wage-disparities in the public-sector where administration is generally easier to conduct and monitor as opposed to the non-public sector where monitoring may be more taxing hence a more noticeable disparity.

The comparison of female-dominated versus mixed occupations reveals a penalty of approximately R195 for working in a female-dominated job at a broad level. In both the public and non-public sector, the penalty for working in a female-dominated occupation is approximately R206 and R203 respectively. For the two samples of women only and women in the non-public sector, the penalty of working in a female-dominated occupation is approximately R197 and R211 respectively. Notably, the highest penalty is recorded for women in the non-public sector. The sample of men aggregated in one basket does not yield significant results. However a sample division of men in the public sector and men in the non-public sector reveals that men in the public sector and in female-dominated occupations receive R199 less than men in mixed occupations. Conversely, men in the non-public sector that work in female-dominated jobs are paid approximately R187 more than men in mixed occupations. The public sector result may be suggesting that men who are not employed in mixed occupations are unwillingly[8] entering female-dominated occupations. They may be pushed by limited job availability, rather than occupational alignment to their innate ability, human capital endowment or preference. This may result in unsatisfied workers who are likely to under-perform, and thus earn lower earnings. The latter refers us back to the inferences of Simpson (2004).

[5] exponential of |-0.039|

[6] Multiplying R1.0397 by the average of 43hours/work per week calculated in Appendix Section B

[7] Multiplying R44.7101 by 4.167 weeks which is the calculated average weeks per month for year 2015

[8] This is linked to the negative stigma and lower social status attached to the idea of a man performing a "woman's" job (Cross et al. 2002; Torre 2018).

In the non-public sector, the common observation is that women in male-dominated and mixed jobs earn more than their counterparts in female-dominated jobs. This may not necessarily be a result of the employers' relatively higher "esteem" of women in male-dominated or mixed occupations, but a dissimilar perception of their occupation. Once again, this may be indicative of a less-efficient administration and monitoring system of wage-disparity curbing policies in the non-public-sector.

In conclusion, the current study employed state of the art econometric methods to the question of whether male-dominated, mixed and female-dominated occupations irrefutably have different pay structures with implications on the SA gender pay gap. Using the 2015 PALMS and LMDSA datasets, we applied Propensity Score Matching to compare hourly earnings of workers in female-dominated jobs versus comparable workers in male-dominated and mixed jobs. Our choice of covariate factors was informed by empirical and theoretical research which we deemed relevant for the particular outcomes of this study and the availability of relevant data. More importantly, selection of the final model was guided by the model's ability to satisfy the balancing and common support properties while keeping in mind the need to maintain reasonable degrees of freedom. Though our methodology does not completely rid the risk of confounding factors which result from unobservables, the ability to satisfy balancing properties and achieve satisfactory common support allow us to confidently assert our findings.

Our findings suggest that women (and at times, men) bear a penalty of being employed in female-dominated occupations thus contributing to women's inferior positions in the labour market and society in general. To some extent, this explains why female-headed households are poorer than male-headed households in South Africa. This realisation invokes the need for government to formulate more regulation or improved techniques of monitoring already existing legislation which ensures equality within and across occupations after accounting for relevant discriminatory factors. This is especially required within the non-public or private sector where the disparity under investigation is most transparent i.e. women in female-dominated occupations are paid less than women in male-dominated and mixed occupations. Complementing this recommendation are concerted strategies between government and its stakeholders towards strengthening programmes or initiatives towards educating more South African women. This should not simply end at basic education but education which adequately prepares women to be highly skilled and competitive in the labour market. An example is offering bursaries or any such opportunities which focus specifically on STEM educational programmes at tertiary institutions. This opens an avenue for more women to penetrate some currently male-dominated occupations such as Physical, mathematical and engineering science professionals or Natural and engineering science associate professionals. Admittedly, innate capabilities may limit entry into occupations such as Extraction and building trades workers or Stationary-plant and related operators, however where possible, women must be empowered to move into higher paying occupations

The South African government has already embarked on a journey of encouraging the study of STEM-related subjects by expanding the STEM curriculum in primary and secondary education, and establishing National Schools of Specialisation that offer new and other skills-based subjects such as Aviation and Maritime studies. The benefits of such developments can only be realised to maximum potential if inclusiveness on the fronts of both gender and race is ensured. This will act to encourage females from elementary stages of their livelihoods, to enter STEM or highly specialised occupations which earn higher incomes. Generally, there is also a need for changing social perceptions and norms that have historically locked women in female-dominated jobs that are congruent to their socially designed roles and responsibilities.

References

Agesa, R.U., 1999. The urban gender wage gap in an African country: Findings from Kenya. *Canadian Journal of Development Studies/Revue canadienne d'études du développement*, *20*(1), pp.59-76.

Agesa, R.U., Agesa, J. and Dabalen, A., 2013. Sources of the persistent gender wage gap along the unconditional earnings distribution: Findings from Kenya. *Oxford Development Studies*, *41*(1), pp.76-103.

Austin, P.C., 2011. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical statistics*, *10*(2), pp.150-161.

Anker, R., 1998. *Gender and jobs: Sex segregation of occupations in the world*. International Labour Organization.

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

Arabsheibani, G.R. and Manfor, L., 2002. From farashia to military uniform: Male-female wage differentials in Libya. *Economic Development and Cultural Change*, *50*(4), pp.1007-1019.

Arbache, J.S., Kolev, A. and Filipiak, E. eds., 2010. *Gender disparities in Africa's labor market*. The World Bank.

Bagilhole, B., 1994. Women, Work and Equal Opportunity. *Alder shot: Avebury.*

Barnard, G., 2009. Realising South Africa's Employment Potential.

Baser, O., 2006. Too much ado about propensity score models? Comparing methods of propensity score matching. *Value in Health*, *9*(6), pp.377-385.

Baum, C.F., 2013. Propensity Score Matching Regression Discontinuity Limited Dependent Variables. *http://fmwww. bc. edu/ECC S*, *2013*.

Becker, S.O. and Ichino, A., 2002. Estimation of average treatment effects based on propensity scores. *The stata journal*, *2*(4), pp.358-377.

Beller, A.H., 1982. Occupational segregation by sex: Determinants and changes. *Journal of Human Resources*, pp.371-392.

Bergmann, B.R., 1974. Occupational segregation, wages and profits when employers discriminate by race or sex. *Eastern Economic Journal*, *1*(2), pp.103-110.

Bernasek, A. and Shwiff, S., 2001. Gender, risk, and retirement. *Journal of economic issues*, *35*(2), pp.345-356.

Bertrand, M., Goldin, C. and Katz, L.F., 2010. Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, *2*(3), pp.228-55.

Bhorat, H. and Goga, S., 2013. The gender wage gap in post-apartheid South Africa: A re-examination. *Journal of African Economies*, *22*(5), pp.827-848.

Blau, F.D. and Ferber, M.A., 1991. Career plans and expectations of young women and men: The earnings gap and labor force participation. *Journal of Human Resources*, pp.581-607.

Blau, F.D. and Kahn, L.M., 1992. The gender earnings gap: learning from international comparisons. *The American Economic Review*, *82*(2), pp.533-538.

Blau, F.D. and Kahn, L.M., 2006. The US gender pay gap in the 1990s: Slowing convergence. *ILR Review*, *60*(1), pp.45-66.

Blau, F.D. and Kahn, L.M., 2017. The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, *55*(3), pp.789-865.

Bond, C.L., 2010. The Revolution Will Be Recognized: Black South African Women's Fight Against Patriarchy and Apartheid.

Bonjour, D. and Gerfin, M., 2001. The unequal distribution of unequal pay–An empirical analysis of the gender wage gap in Switzerland. *Empirical economics*, *26*(2), pp.407-427.

Bradley, H., 1993. Across the great divide: The entry of men into “women’s jobs.”. *Doing “women’s work”: Men in nontraditional occupations*, pp.10-27.

Brixiová, Z. and Kangoye, T., 2016. Gender disparities in employment and earnings in Sub-Saharan Africa: Evidence from Swaziland.

Brown, R.S., Moon, M. and Zoloth, B.S., 1980. Occupational attainment and segregation by sex. *ILR Review*, *33*(4), pp.506-517.

Caliendo, M. and Kopeinig, S., 2008. Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, *22*(1), pp.31-72.

Casale, D. and Posel, D., 2002. The continued feminisation of the labour force in South Africa: An analysis of recent data and trends. *South African Journal of Economics*, *70*(1), pp.156-184.

Casale, D., 2004. What has the feminisation of the labour market ‘bought’ women in South Africa? Trends in labour force participation, employment and earnings, 1995–2001. *Journal of Interdisciplinary Economics*, *15*(3-4), pp.251-275.

Casale, D. and Posel, D., 2005. Women and the economy: how far have we come?. *Agenda*, *19*(64), pp.21-29.

Casale, D. and Posel, D., 2011. Unions and the gender wage gap in South Africa. *Journal of African Economies,* 20 (1), pp.27-59.

Clark, A., Oswald, A. and Warr, P., 1996. Is job satisfaction U-shaped in age?. *Journal of occupational and organizational psychology*, *69*(1), pp.57-81.

Cock, J., 1985. Black Women Under Apartheid.

Comblon, V., Robilliard, A.S. and Roubaud, F.R.A.N.Ç.O.I.S., 2017. *Gender Analysis of Labour Market Outcomes in Sub-Saharan Africa: Recent Evidence from Cameroon and Mali*. UN Women.

Corcoran, M., Duncan, G.J. and Ponza, M., 1984. 1 0 Work Experience, Job Segregation, and Wages. *SPONS AGENCY*, p.178.

Cotter, D.A., Hermsen, J.M. and Vanneman, R., 2003. The effects of occupational gender segregation across race. *The Sociological Quarterly*, *44*(1), pp.17-36.

Cross, S. and Bagilhole, B., 2002. Girls' jobs for the boys? Men, masculinity and non-traditional occupations. *Gender, Work & Organization*, *9*(2), pp.204-226.

Drolet, M., 2001. The persistent gap: New evidence on the Canadian gender wage gap. *Statistics Canada Analytical Studies Branch Working Paper*, (157).

Eccles, J.S., 1986. Gender-roles and women's achievement. *Educational Researcher*, *15*(6), pp.15-19.

Elu, J.U. and Loubert, L., 2013. Earnings inequality and the intersectionality of gender and ethnicity in Sub-Saharan Africa: The case of Tanzanian manufacturing. *American Economic Review*, *103*(3), pp.289-92.

Evans, O. and Steptoe, A., 2002. The contribution of gender-role orientation, work factors and home stressors to psychological well-being and sickness absence in male-and female-dominated occupational groups. *Social science & medicine*, *54*(4), pp.481-492.

Fester, G., 2000. Women writing for their rights. *Agenda*, *16*(46), pp.42-47.

Frome, P.M., Alfeld, C.J., Eccles, J.S. and Barber, B.L., 2006. Why don't they want a male-dominated job? An investigation of young women who changed their occupational aspirations. *Educational Research and Evaluation*, *12*(4), pp.359-372.

Gaitskell, D., Kimble, J., Maconachie, M. and Unterhalter, E., 1983. Class, race and gender: Domestic workers in South Africa. *Review of African Political Economy*, *10*(27-28), pp.86-108.

Gertler, P.J., Martinez, S., Premand, P., Rawlings, L.B. and Vermeersch, C.M., 2016. *Impact evaluation in practice*. The World Bank.

Grün, C., 2004. Direct and indirect gender discrimination in the South African labour market. *International Journal of Manpower*, *25*(3/4), pp.321-342.

Gunderson, M., 1989. Male-female wage differentials and policy responses. *Journal of Economic Literature*, *27*(1), pp.46-72.

Harkness, S., 1996. The gender earnings gap: evidence from the UK. *Fiscal Studies*, *17*(2), pp.1-36.

Hayes, R., 1986. Gender nontraditional or sex atypical or gender dominant or… research: Are we measuring the same thing?. *Journal of Vocational Behavior*, *29*(1), pp.79-88.

Hill, C., Corbett, C. and St Rose, A., 2010. *Why so few? Women in science, technology, engineering, and mathematics*. American Association of University Women. 1111 Sixteenth Street NW, Washington, DC 20036.

Hopfer, C., 1997. Empowering adult education in Namibia and South Africa during and after apartheid. *International review of education*, *43*(1), pp.43-59.

Irfan, M., Anwar, S., Akram, W., Ramzan, S. and Waqar, I., 2013. Determinants of Occupational Choice among Four Provinces of the Pakistan. *International Journal*, *2*(6).

Himmelweit, S. and Sigala, M., 2003. *Internal and External Constraints on Mother's Employment: Some Implications for Policy*. Economic and Social Research Council.

Hinks, T., 2002. Gender wage differentials and discrimination in the New South Africa. *Applied Economics*, *34*(16), pp.2043-2052.

Jacobs, J.A., 1999. The sex segregation of occupations: Prospects for the 21st century. *Handbook of gender and work*, pp.125-144.

Jianakoplos, N.A. and Bernasek, A., 1998. Are women more risk averse?. *Economic inquiry*, *36*(4), pp.620-630.

Jome, L.M. and Tokar, D.M., 1998. Dimensions of masculinity and major choice traditionality. *Journal of Vocational Behavior*, *52*(1), pp.120-134.

Kabeer, N., 2010. Women's empowerment, development interventions and the management of information flows. *ids Bulletin*, *41*(6), pp.105-113.

Khandker, S., B. Koolwal, G. and Samad, H., 2009. *Handbook on impact evaluation: quantitative methods and practices*. The World Bank.

Knight, J.B. and Sabot, R.H., 1982. Labor market discrimination in a poor urban economy. *The Journal of Development Studies*, *19*(1), pp.67-87.

Kwan, M.P., 1999. Gender, the home-work link, and space-time patterns of nonemployment activities. *Economic geography*, *75*(4), pp.370-394.

Land, Labour and Apartheid 2019, viewed 26 March 2020, <https://www.sahistory.org.za/article/land-labour-and-apartheid>.

Landis, E.S., 1975. *African Women Under Apartheid*. Africa Fund.

Lehrer, E.L. and Stokes, H., 1985. Determinants of the female occupational distribution: A log-linear probability analysis. *The Review of Economics and Statistics*, pp.395-404.

Madden, F J., 1981. Why women work closer to home. *Urban studies*, *18*(2), pp.181-194.

Magadla, S.S., Leibbrandt, M. and Mlatsheni, C., 2019. Does a motherhood penalty exist in the post-apartheid South African labour market?.

Mandel, H. and Semyonov, M., 2005. Family policies, wage structures, and gender gaps: Sources of earnings inequality in 20 countries. *American sociological review*, *70*(6), pp.949-967.

Mariotti, M. and Fourie, J., 2014. The economics of apartheid: An introduction.

Martin, F R. and Jarret-Kerr, C R. 2019, Apartheid in Nursing- A Challenge, viewed 26 March 2020, <https://www.sahistory.org.za/sites/default/files/DC/asoct58.8/asoct58.8.pdf>

Mathur-Helm, B., 2005. Equal opportunity and affirmative action for South African women: a benefit or barrier?. *Women in Management Review*, *20*(1), pp.56-71.

Martineau, R., 1997. Women and education in South Africa: Factors influencing women's educational progress and their entry into traditionally male-dominated fields. *Journal of Negro Education*, pp.383-395.

Meer, F., 1985. Women in the apartheid society.

Mon, M., 2017. Determinants of female labour force participation in Burma: An empirical Analysis of socio-economic survey data. *ABAC Journal*, *21*(1).

Mosomi, J., 2019. An empirical analysis of trends in female labour force participation and the gender wage gap in South Africa. *Agenda*, pp.1-15.

Ncube, P. and Tregenna, F., 2013. An empirical investigation of the gender wage gap in South Africa.

Neitzert, M., 1994. A Woman's Place: Household Labour Allocation in Rural Kenya. *Canadian Journal of Development Studies/Revue canadienne d'études du développement*, *15*(3), pp.401-427.

Nkomo, M., 1981. The contradictions of Bantu education. *Harvard educational review*, *51*(1), pp.126-138.

Nordman, C.J., Robilliard, A.S. and Roubaud, F., 2011. Gender and ethnic earnings gaps in seven West African cities. *Labour Economics*, *18*, pp.S132-S145.

Ntuli, M. and Kwenda, P., 2020. Gender Gaps in Employment and Wages in Sub-Saharan Africa: A Review. In *Women and Sustainable Human Development* (pp. 183-203). Palgrave Macmillan, Cham.

Nwosu, C. and Ndinda, C., 2018. Female household headship and poverty in South Africa: an employment-based analysis. Economic Research Southern Africa: Working Papers.

Polachek, S.W., 1981. Occupational self-selection: A human capital approach to sex differences in occupational structure. *The review of Economics and Statistics*, pp.60-69.

Rapino, M.A. and Cooke, T.J., 2011. Commuting, gender roles, and entrapment: A national study utilizing spatial fixed effects and control groups. *The Professional Geographer*, *63*(2), pp.277-294.

Reichardt, C.S., Minton, B.A. and Schellenger, J.D., 1980. The analysis of covariance (ANCOVA) and the assessment of treatment effects. *Prevention Evaluation Research Monograph II-Outcome (Section D4, chap. VIII). Denver: University of Denver, Department of Psychology*.

Rosenbaum, P.R., 2002. Overt bias in observational studies. In *Observational studies* (pp. 71-104). Springer, New York, NY.

Rosenbaum, P.R. and Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), pp.41-55.

Williams, C.L. ed., 1993. *Doing" women's work": Men in nontraditional occupations* (Vol. 3). Sage Publications.

Scandura, T.A. and Lankau, M.J., 1997. Relationships of gender, family responsibility and flexible work hours to organizational commitment and job satisfaction. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *18*(4), pp.377-391.

Seekings, J., 2010. Race, class and inequality in the South African city.

Selmi, M., 1999. Family leave and the gender wage gap. *NCL Rev.*, *78*, p.707.

Shepherd, D., 2008. Post-apartheid trends in gender discrimination in South Africa: Analysis through decomposition techniques. *Stellenbosch, University of Stellenbosch. Retrieved October*, *5*, p.2011.

Simpson, R., 2004. Masculinity at work: The experiences of men in female dominated occupations. *Work, employment and society*, *18*(2), pp.349-368.

Tophoven, S., du Prel, J.B., Peter, R. and Kretschmer, V., 2015. Working in gender-dominated occupations and depressive symptoms: findings from the two age cohorts of the lidA study. *Journal for Labour Market Research*, *48*(3), pp.247-262.

Torre, M., 2018. Stopgappers? The occupational trajectories of men in female-dominated occupations. *Work and Occupations*, *45*(3), pp.283-312.

United Nations Centre Against Apartheid, 1978. The Effects of Apartheid on the Status of Women in South Africa. *The Black Scholar*, *10*(1), pp.11-20.

Van der Westhuizen, C., Goga, S. and Oosthuizen, M., 2007. Women in the South African Labour Market, 1995-2005.

Zhang, J., Han, J., Liu, P.W. and Zhao, Y., 2008. Trends in the gender earnings differential in urban China, 1988–2004. *ILR Review*, *61*(2), pp.224-243.

Appendix

Section A: Table of Key Variables' Definitions

Variable	Definition
household size	Number of household members
AGE	Age of an individual
Age squared	Age of an individual squared
Married	Dummy variable of marital status; "1" is married, "0" is unmarried
No education	0 if Individual is educated, 0 otherwise
Primary education	1 if Individual's highest level of education is primary, 0 otherwise
Secondary education	1 if Individual's highest level of education is secondary, 0 otherwise
Diploma certificate	1 if Individual's highest level of education is diploma/certificate, 0 otherwise
Higher levels	1 if Individual's highest level of education is above diploma/certificate, 0 otherwise
Urban Areas	1 if Individual's residential area is an urban area, 0 otherwise
Traditional	1 if Individual's residential area is a traditional rural area, 0 otherwise
Farms	1 if Individual's residential area is a rural farm, 0 otherwise
Mining Areas	1 if Individual's residential area is a mining area, 0 otherwise
Presence of kids below 7yrs	1 if an individual resides with children below 7 years, 0 otherwise
Presence of kids between 7&14yrs	1 if an individual resides with children between 7&14yrs, 0 otherwise
maternity	1 if an individual's job offers maternity/paternity leave, 0 otherwise
Average hours worked per week	Hours worked per week
Wages per hour	Hourly wages

Section B: Summary Statistics

Table B1: Full Table of Summary Statistics by broad occupation group

Variable	Number of Observations	Mean/proportion	Std. Dev.	Min	Max
Female-Dominated Occupations					
Household size	9,318	7.801889	6.342961	1	57
Age	9,318	38.86435	10.53056	18	59
Age squared	9,318	1621.318	837.8412	324	3481

No education	566	.0608501	-	0	1
Primary Education	1,218	.1306074	-	0	1
Secondary Education	6,315	.6777205	-	0	1
Diploma certificate	1,103	.118373	-	0	1
Higher levels	116	.012449	-	0	1
Urban Areas	7,478	.8025	-	0	1
Traditional	1,339	.1437	-	0	1
Farms	289	.0310	-	0	1
Mining Areas	212	.0228	-	0	1
Households with no kids	9,318	.5639622	.4959186	0	1
Households with kids below 7 years old	9,318	.2863275	.4520686	0	1
Households with kids between 7 & 14 years old	9,318	.3077914	.4616045	0	1
maternity	9,318	.4767117	.4994842	0	1
Average hours worked per week	9,318	40.00011	12.58092	1	84
Wages per hour	9,318	43.65071	64.53526	5.021605	966.6589
Log of Wages per hour	9,318	3.23617	.9654075	1.61375	6.873846
Male-dominated Occupations					
Household size	11,745	6.714432	6.097515	1	64
Age	11,745	37.08753	10.11956	18	59
Age squared	11,745	1477.881	790.6948	324	3481
No education	829	.0705583	-	0	1
Primary Education	2,183	.1858663	-	0	1
Secondary Education	8,031	.6837803	-	0	1

Diploma certificate	553	.047169	-	0	1
Higher levels	149	.0126862	-	0	1
Urban Areas	7,827	.6664	-	0	1
Traditional	1,759	.1498	-	0	1
Farms	1,504	.1281	-	0	1
Mining Areas	655	.0558	-	0	1
Households with no kids	11,745	.6790123	.4668759	0	1
Households with kids below 7 years old	11,745	.2236696	.4167209	0	1
Households with kids between 7 & 14 years old	11,745	.2104725	.4076616	0	1
maternity	11,745	.3589613	.4797162	0	1
Average hours worked per week	11,745	43.72771	11.8785	1	84
Wages per hour	11,745	39.58796	62.03844	5.000984	977.3953
Log of Wages per hour	11,745	3.152735	.917017	1.609635	6.884891
Mixed Occupations					
Household size	9,917	7.262378	5.89175	1	55
Age	9,917	37.38913	9.964678	18	59
Age squared	9,917	1497.232	783.1958	324	3481
No education	1,143	0.114349	-	0	1
Primary Education	430	.0433599	-	0	1
Secondary Education	6,507	.656146	-	0	1

Diploma certificate	925	.0932742	-	0	1
Higher levels	921	.0928708	-	0	1
Urban Areas	8,378	.8448	-	0	1
Traditional	1,186	.1196	-	0	1
Farms	181	.0183	-	0	1
Mining Areas	172	.0173	-	0	1
Households with no kids	9,917	.6232732	.4845899	0	1
Households with kids below 7 years old	9,917	.2429162	.4288665	0	1
Households with kids between 7 & 14 years old	9,917	.2499748	.43302	0	1
maternity	9,917	.5679137	.4953912	0	1
Average hours worked per week	9,917	45.71645	12.61239	1	84
Wages per hour	9,917	66.03255	95.90459	5.012305	992.2997
Log of Wages per hour	9,917	3.542748	1.096821	1.611896	6.900025

Table B2: Summary Statistics for all occupations combined

Variable	**Number of Observations**	**Mean**	**Std. Dev.**	**Min**	**Max**
Household size	30,980	7.215072	6.118427	1	64
Age	30,980	37.92842	10.46242	18	59
Age squared	30,980	1526.724	803.9922	324	3481
No education	2,560	.0826425	-	0	1
Primary Education	3,776	.1218752	-	0	1
Secondary Education	20,833	.6724504	-	0	1
Diploma certificate	2,634	.0850202	-	0	1

Higher levels	1,177z	.0380117	-	0	1
Urban Areas	22,442	.7244	-	0	1
Traditional	4,715	.1522	-	0	1
Farms	2,580	.0833	-	0	1
Mining Areas	1,242	.0401	-	0	1
Households with no kids	11,554	.627691	.4834279	0	1
Households with kids below 7 years old	30,980	.2467065	.4311014	0	1
Households with kids between 7 & 14 years old	30,980	.2506266	.4333808	0	1
maternity	30,980	.4597712	.498387	0	1
Average hours worked per week	30,980	43.245	12.51803	1	84
Wages per hour	30,980	49.28851	75.87187	5.000984	992.2997
Log of Wages per hour	30,980	3.303974	1.006321	1.609635	6.900025

Section C: Selected Logit Regressions and Corresponding Balancing Tests

Female-dominated Versus Male-dominated

a. Pooled Data

Logistic regression

Number of obs = 20,381
LR chi2(14) = 1983.64
Prob > chi2 = 0.0000
Pseudo R2 = 0.0709

Log likelihood = -13003.223

treat	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
household size	0.005	0.003	1.740	0.082	-0.001	0.011
AGE	0.017	0.002	8.220	0.000	0.013	0.021
Age squared	0.000	0.000	5.650	0.000	0.000	0.000
married	-0.398	0.032	-12.290	0.000	-0.462	-0.335
primary education	-0.004	0.092	-0.040	0.967	-0.184	0.176
secondary education	0.257	0.088	2.920	0.004	0.084	0.429
diploma certificate	1.008	0.102	9.860	0.000	0.808	1.208
higher levels	0.181	0.154	1.170	0.242	-0.122	0.484
Traditional	-0.243	0.042	-5.760	0.000	-0.325	-0.160
Farms	-1.377	0.068	-20.110	0.000	-1.511	-1.243
Mining Areas	-1.089	0.083	-13.050	0.000	-1.252	-0.925

presence of kids below 7yrs	0.412	0.050	8.270	0.000	0.314	0.509
presence of kids between 7&14yrs	0.518	0.044	11.840	0.000	0.432	0.604
maternity	0.352	0.032	11.070	0.000	0.290	0.415
_cons	-1.379	0.114	-12.110	0.000	-1.602	-1.156

Treatment assignment	Off Support	On Support	Total
Untreated	0	11,349	11,349
Treated	32	9,000	9,032
Total	32	20,349	20,381

Balancing Test Results

Variable	Treated	Control	%bias	t	p>t	V(C)
household size	7.812	7.873	-1.000	-0.630	0.530	0.90*
AGE	38.916	38.824	. 9	0.590	0.553	1.040
Age squared	1646.100	1650.400	-0.500	-0.330	0.739	1.010
married	0.442	0.447	-1.100	-0.740	0.458	.
primary education	0.134	0.128	1.7	1.250	0.212	.
secondary education	0.705	0.709	-0.900	-0.590	0.556	.
diploma certificate	0.121	0.123	-1.000	-0.550	0.585	.
higher levels	0.013	0.012	1	0.710	0.480	.
Traditional	0.144	0.134	3	2.030	0.042	.
Farms	0.032	0.035	-1.000	-1.020	0.309	.
Mining Areas	0.023	0.022	.5	0.450	0.654	.
presence of kids below 7yrs	0.128	0.122	2	1.300	0.195	.
presence of kids between 7&14yrs	0.309	0.315	-1.400	-0.870	0.385	.
maternity	0.471	0.469	.6	0.390	0.698	.

* if variance ratio outside [0.96; 1.04]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.001	12.72	0.549	1.2	1.0	5.3	1.05	33

* if B>25%, R outside [0.5; 2]

b. Male only

Logistic regression

Number of obs = 12,139
LR chi2(14) = 955.13
Prob > chi2 = 0.0000
Pseudo R2 = 0.0649

Log likelihood = -6882.0707

treat	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
household size	0.030	0.004	6.830	0.000	0.021	0.039
Q14AGE	0.000	0.003	0.010	0.989	-0.006	0.006
Age squared	0.000	0.000	1.910	0.057	-0.000	0.000
married	0.163	0.047	3.430	0.001	0.070	0.256
primary education	-0.175	0.132	-1.320	0.185	-0.434	0.084
secondary education	0.085	0.125	0.680	0.498	-0.161	0.331
diploma certificate	0.777	0.145	5.370	0.000	0.493	1.061
higher levels	0.082	0.228	0.360	0.720	-0.366	0.530
Traditional	-0.221	0.061	-3.650	0.000	-0.340	-0.102
Farms	-1.371	0.104	-13.240	0.000	-1.574	-1.168
Mining areas	-1.246	0.119	-10.480	0.000	-1.479	-1.013
Presence of kids below 7yrs	0.166	0.072	2.300	0.022	0.024	0.308
Presence of kids between 7&14yrs	0.248	0.066	3.730	0.000	0.118	0.378
maternity	0.418	0.044	9.440	0.000	0.331	0.504
_cons	-1.411	0.160	-8.810	0.000	-1.725	-1.097

Treatment assignment	Off Support	On Support	Total
Untreated	0	8,562	8,562
Treated	2	3,575	3,577
Total	2	12,137	12,139

Variable	Treated	Control	%bias	t	p>t	V(C)
household size	7.711	7.989	-4.700	-1.770	0.077	0.78*
AGE	37.671	37.401	2.6	1.110	0.267	1.08*
Age squared	1561	1539.200	2.6	1.080	0.279	1.11*
married	0.535	0.541	-1.300	-0.560	0.577	.
primary education	0.127	0.121	1.6	0.750	0.451	.
secondary education	0.723	0.744	-4.700	-2.020	0.043	.
diploma certificate	0.111	0.097	5.2	1.920	0.055	.
higher levels	0.012	0.013	-0.500	-0.210	0.830	.
Traditional	0.132	0.126	1.7	0.740	0.459	.
Farms	0.032	0.035	-1.100	-0.690	0.489	.
Mining areas	0.025	0.019	2.9	1.740	0.082	.
Presence of kids below 7yrs	0.124	0.133	-2.600	-1.020	0.306	.
Presence of kids	0.265	0.267	-0.600	-0.230	0.820	.

between 7&14yrs						
maternity	0.472	0.456	3.4	1.390	0.165	.

* if variance ratio outside [0.94; 1.07]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.002	16.17	0.303	2.5	2.6	9.5	1.05	100

* if B>25%, R outside [0.5; 2]

c. Female only

Logistic regression

Number of obs = 8,205
LR chi2(14) = 862.03
Prob > chi2 = 0.0000
Log likelihood = -4845.2868
Pseudo R2 = 0.0817

treat	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
household size	-0.029	0.004	-6.580	0.000	-0.038	-0.020
AGE	0.028	0.003	8.230	0.000	0.021	0.034
Age squared	0.000	0.000	0.240	0.808	-0.000	0.000
married	-0.774	0.053	-14.740	0.000	-0.877	-0.671
primary education	0.180	0.147	1.230	0.219	-0.107	0.468
secondary education	0.242	0.140	1.730	0.084	-0.032	0.516
diploma certificate	0.773	0.162	4.760	0.000	0.455	1.091
higher levels	-0.303	0.225	-1.340	0.179	-0.744	0.139
Traditional	-0.173	0.069	-2.500	0.012	-0.309	-0.038
Farms	-1.726	0.103	-16.780	0.000	-1.927	-1.524
Mining areas	-0.242	0.163	-1.490	0.137	-0.562	0.077
presence of kids below 7yrs	-0.009	0.079	-0.120	0.905	-0.164	0.145
presence of kids between 7&14yrs	0.026	0.066	0.390	0.697	-0.103	0.154
maternity	0.310	0.053	5.870	0.000	0.206	0.413
_cons	-0.093	0.187	-0.500	0.618	-0.461	0.274

Treatment assignment	Off Support	On Support	Total
Untreated	0	2,815	2,815
Treated	9	5,381	5,390
Total	9	8,196	8,205

Balancing Test Results

Variable	Treated	Control	%bias	t	p>t	V(C)
household size	7.825	7.939	-1.700	-0.940	0.345	1.13*
AGE	39.672	39.428	2.4	1.210	0.226	0.960
Age squared	1686.800	1653	3.9	2.010	0.045	0.90*

married	0.374	0.386	-2.600	-1.360	0.175	.
primary education	0.129	0.132	-0.900	-0.460	0.644	.
secondary education	0.696	0.680	3.5	1.800	0.072	.
diploma certificate	0.132	0.139	-2.400	-1.100	0.272	.
higher levels	0.014	0.014	.1	0.040	0.966	.
Traditional	0.153	0.156	-0.900	-0.490	0.625	.
Farms	0.030	0.034	-1.400	-1.150	0.251	.
Mining areas	0.020	0.022	-1.200	-0.620	0.535	.
presence of kids below 7yrs	0.125	0.126	-0.300	-0.140	0.891	.
presence of kids between 7&14yrs	0.334	0.331	.8	0.400	0.689	.
maternity	0.481	0.474	1.5	0.750	0.451	.

* if variance ratio outside [0.95; 1.05]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.001	18.03	0.206	1.7	1.4	8.2	1.02	67

* if B>25%, R outside [0.5; 2]

d. Public Sector

Logistic regression

Number of obs = 2,995
LR chi2(14) = 324.12
Prob > chi2 = 0.0000
Log likelihood = -1768.7126
Pseudo R2 = 0.0839

treat	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
household size	-0.006	0.008	-0.800	0.421	-0.021	0.009
AGE	0.016	0.006	2.860	0.004	0.005	0.028
Age squared	0.000	0.000	1.530	0.127	-0.000	0.000
married	-0.440	0.089	-4.940	0.000	-0.614	-0.265
primary education	0.614	0.305	2.010	0.044	0.016	1.213
secondary education	1.201	0.291	4.130	0.000	0.631	1.771
diploma certificate	2.357	0.306	7.700	0.000	1.757	2.957
higher levels	1.847	0.387	4.770	0.000	1.089	2.606
Traditional	-0.235	0.100	-2.340	0.019	-0.432	-0.038
Farms	-0.868	0.234	-3.710	0.000	-1.326	-0.410
Mining areas	-0.696	0.248	-2.800	0.005	-1.182	-0.209
presence of kids below 7yrs	0.352	0.134	2.620	0.009	0.088	0.615
presence of kids between 7&14yrs	0.373	0.112	3.320	0.001	0.153	0.593
maternity	0.415	0.090	4.600	0.000	0.238	0.592

_cons	-1.643	0.361	-4.560	0.000	-2.350	-0.937

Treatment assignment	Off Support	On Support	Total
Untreated	0	1,035	1,035
Treated	35	1,925	1,960
Total	35	2,960	2,995

Balancing Test Results

Variable	Treated	Control	%bias	t	p>t	V(C)
household size	7.990	8.038	-0.700	-0.230	0.817	0.930
AGE	40.834	40.216	5.8	1.820	0.069	0.90*
Age squared	1790	1743.600	5.6	1.710	0.087	0.90*
married	0.491	0.503	-2.400	-0.740	0.459	.
primary education	0.072	0.076	-1.500	-0.550	0.580	.
secondary education	0.578	0.564	2.9	0.880	0.379	.
diploma certificate	0.308	0.322	-3.500	-0.940	0.349	.
higher levels	0.033	0.024	5.5	1.650	0.099	.
Traditional	0.212	0.179	7.9	2.600	0.009	.
Farms	0.018	0.021	-1.200	-0.530	0.599	.
Mining Areas	0.020	0.015	2.7	1.040	0.297	.
presence of kids below 7yrs	0.134	0.132	.7	0.210	0.831	.
presence of kids between 7&14yrs	0.339	0.356	-3.700	-1.130	0.257	.
maternity	0.730	0.737	-1.500	-0.510	0.610	.

* if variance ratio outside [0.91; 1.09]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.004	21.42	0.091	3.3	2.8	14.9	1.06	67

* if B>25%, R outside [0.5; 2]

e. Non-public Sector

Logistic regression

	Number of obs	=	17,492
	LR chi2(14)	=	1325.76
	Prob > chi2	=	0.0000
Log likelihood = -11136.928	Pseudo R2	=	0.0562

treat	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
household size	0.006	0.003	1.730	0.084	-0.001	0.012
AGE	0.013	0.002	5.620	0.000	0.008	0.017
Age squared	0.000	0.000	4.340	0.000	0.000	0.000
married	-0.361	0.035	-10.360	0.000	-0.429	-0.292
primary education	-0.054	0.093	-0.580	0.562	-0.237	0.129

secondary education	0.124	0.089	1.390	0.165	-0.051	0.298
diploma certificate	0.409	0.110	3.710	0.000	0.193	0.625
higher levels	-0.622	0.190	-3.270	0.001	-0.995	-0.249
Traditional	-0.309	0.047	-6.560	0.000	-0.402	-0.217
Farms	-1.383	0.073	-19.020	0.000	-1.526	-1.241
Mining Areas	-1.233	0.094	-13.140	0.000	-1.417	-1.049
presence of kids below 7yrs	0.321	0.054	5.970	0.000	0.216	0.426
presence of kids between 7&14yrs	0.510	0.048	10.670	0.000	0.416	0.604
maternity	0.297	0.035	8.550	0.000	0.229	0.366
_cons	-1.106	0.117	-9.420	0.000	-1.336	-0.876

Treatment assignment	Off Support	On Support	Total
Untreated	0	10,426	10,46
Treated	6	7,060	7,066
Total	6	17,486	17,492

Balancing Test Results

Variable	Treated	Control	%bias	t	p>t	V(C)
household size	7.716	7.748	-0.500	-0.290	0.770	0.91*
AGE	38.504	38.524	-0.200	-0.110	0.914	1.010
Age squared	1564.100	1582	-2.200	-1.290	0.199	0.990
married	0.424	0.435	-2.000	-1.210	0.227	.
primary education	.15	0.148	.6	0.380	0.705	.
secondary education	0.738	0.739	-0.300	-0.170	0.863	.
diploma certificate	0.071	0.072	-0.500	-0.260	0.794	.
higher levels	0.007	0.007	.5	0.360	0.721	.
Traditional	0.130	0.129	.3	0.200	0.841	.
Farms	0.036	0.038	-0.700	-0.580	0.560	.
Mining Areas	0.022	0.022	.1	0.090	0.931	.
presence of kids below 7yrs	0.125	0.110	4.6	2.760	0.006	.
presence of kids between 7&14yrs	0.297	0.313	-3.800	-2.130	0.033	.
maternity	0.404	0.398	1.2	0.720	0.471	.

* if variance ratio outside [0.95; 1.05]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.001	16.46	0.286	1.3	0.6	6.8	1.14	33

* if B>25%, R outside [0.5; 2]

f. Male and Public sector

Logistic regression

Number of obs = 1,359
LR chi2(14) = 149.49
Prob > chi2 = 0.0000
Log likelihood = -864.94624 Pseudo R2 = 0.0795

treat	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
household size	0.025	0.013	1.940	0.052	-0.000	0.050
AGE	0.010	0.009	1.110	0.269	-0.007	0.027
Age squared	0.000	0.000	0.690	0.489	-0.000	0.000
married	-0.183	0.136	-1.350	0.177	-0.450	0.083
primary education	0.595	0.474	1.260	0.209	-0.333	1.523
secondary education	1.215	0.456	2.660	0.008	0.320	2.109
diploma certificate	2.449	0.479	5.120	0.000	1.511	3.387
higher levels	2.315	0.589	3.930	0.000	1.160	3.470
Traditional	0.034	0.148	0.230	0.817	-0.256	0.325
Farms	-0.963	0.336	-2.860	0.004	-1.622	-0.304
Mining Areas	-0.346	0.345	-1.000	0.315	-1.022	0.330
Presence of kids below 7yrs	0.182	0.201	0.910	0.365	-0.212	0.577
Presence of kids between 7&14yrs	0.400	0.179	2.230	0.026	0.049	0.751
maternity	0.124	0.130	0.950	0.340	-0.131	0.379
_cons	-2.004	0.556	-3.610	0.000	-3.094	-0.915

Treatment assignment	Off Support	On Support	Total
Untreated	0	640	640
Treated	15	704	719
Total	15	1,344	1,359

Balancing Test Results

Variable	Treated	Control	%bias	t	p>t	V(C)
household size	7.918	7.830	1.5	0.280	0.781	1.33*
AGE	40.054	40.270	-2.100	-0.390	0.693	0.970
Age squared	1818.500	1862.600	-5.100	-0.940	0.348	0.920
married	0.563	0.568	-1.100	-0.210	0.830	.
primary education	0.081	0.070	3.3	0.760	0.450	.
secondary education	0.609	0.618	-1.800	-0.330	0.743	.
diploma certificate	0.263	0.263	0	0.000	1.000	.
higher levels	0.037	0.034	1.8	0.290	0.774	.
Traditional	0.207	0.161	11.400	2.240	0.025	.
Farms	0.018	0.017	.7	0.200	0.840	.

Mining Areas	0.024	0.031	-4.200	-0.810	0.417	.
Presence of kids below 7yrs	0.132	0.148	.5.100	-0.880	0.378	.
Presence of kids between 7&14yrs	0.304	0.303	.2	0.030	0.977	.
maternity	0.692	0.702	-2.200	-0.430	0.664	.

* if variance ratio outside [0.86; 1.16]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.005	9.66	0.786	2.9	1.9	16.6	1.19	33

* if B>25%, R outside [0.5; 2]

g. Female and Public Sector

Logistic regression

Number of obs = 1,627
LR chi2(14) = 271.19
Prob > chi2 = 0.0000
Log likelihood = -788.27481
Pseudo R2 = 0.1468

treat	**Coef.**	**Std.Err.**	**z**	**P>z**	**[95%Conf.**	**Interval]**
household size	-0.015	0.011	-1.350	0.176	-0.037	0.007
AGE	0.043	0.009	4.930	0.000	0.026	0.059
Age squared	-0.000	0.000	-0.950	0.341	-0.000	0.000
married	-0.848	0.138	-6.140	0.000	-1.118	-0.577
primary education	0.629	0.476	1.320	0.187	-0.305	1.563
secondary education	1.284	0.459	2.800	0.005	0.385	2.184
diploma certificate	2.429	0.488	4.980	0.000	1.474	3.385
higher levels	1.286	0.552	2.330	0.020	0.204	2.369
traditional	-0.332	0.150	-2.210	0.027	-0.627	-0.038
farms	-0.478	0.342	-1.400	0.162	-1.148	0.192
Mining Areas	0.220	0.489	0.450	0.652	-0.738	1.178
presence of kids below 7yrs	0.110	0.199	0.560	0.579	-0.279	0.500
presence of kids between 7&14yrs	0.156	0.165	0.950	0.344	-0.168	0.481
maternity	1.017	0.133	7.650	0.000	0.756	1.278
_cons	-1.927	0.580	-3.320	0.001	-3.064	-0.789

Treatment assignment	Off Support	On Support	Total
Untreated	0	415	415
Treated	31	1,181	1,212
Total	31	1,596	1,627

Balancing Test Results

Variable	Treated	Control	%bias	t	p>t	V(C)
household size	8.167	7.716	6.9	1.720	0.085	1.020
AGE	40.768	38.927	18.000	4.380	0.000	0.88*
Age squared	1809.100	1655.600	18.300	4.570	0.000	0.970
married	0.406	0.428	-4.300	-1.040	0.297	.
primary education	0.058	0.049	2.9	0.920	0.360	.
secondary education	0.571	0.530	8.4	1.970	0.049	.
diploma certificate	0.327	0.384	-14.300	-2.880	0.004	.
higher levels	0.036	0.029	4	1.040	0.297	.
traditional	0.224	0.171	12.200	3.240	0.001	.
farms	0.019	0.017	1.4	0.460	0.644	.
mining areas	0.019	0.025	-4.200	-0.910	0.365	.
presence of kids below 7yrs	0.127	0.117	2.9	0.750	0.451	.
presence of kids between 7&14yrs	0.377	0.334	8.6	2.150	0.032	.
maternity	0.728	0.722	1.3	0.320	0.747	.

* if variance ratio outside [0.89; 1.12]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.015	49.07	0.000	7.7	5.6	29.0*	0.83	33

* if B>25%, R outside [0.5; 2]

h. Male and Non-public Sector

Logistic regression

Number of obs = 10,877
LR chi2(14) = 689.15
Prob > chi2 = 0.0000
Pseudo R2 = 0.0547

Log likelihood = -5954.5121

treat	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
household size	0.032	0.005	6.770	0.000	0.023	0.042
AGE	-0.006	0.003	-1.750	0.080	-0.012	0.001
Age squared	0.000	0.000	2.980	0.003	0.000	0.000
married	0.173	0.051	3.390	0.001	0.073	0.273
primary education	-0.005	0.141	-0.040	0.970	-0.282	0.271
secondary education	0.097	0.135	0.720	0.474	-0.168	0.361
diploma certificate	0.594	0.162	3.670	0.000	0.277	0.911
higher levels	-0.206	0.276	-0.750	0.455	-0.746	0.335
Traditional	-0.390	0.069	-5.670	0.000	-0.525	-0.255
Farms	-1.327	0.106	-12.490	0.000	-1.536	-1.119

Mining areas	-1.243	0.127	-9.780	0.000	-1.492	-0.994
presence of kids below 7yrs	0.230	0.076	3.030	0.002	0.081	0.379
presence of kids between 7&14yrs	0.235	0.072	3.270	0.001	0.094	0.375
maternity	0.333	0.049	6.880	0.000	0.238	0.429
_cons	-1.359	0.170	-7.990	0.000	-1.692	-1.025

Treatment assignment	Off Support	On Support	Total
Untreated	0	7,985	7,985
Treated	1	2,891	2,892
Total	1	10,876	10,877

Balancing Test Results

Variable	Treated	Control	%bias	t	p>t	V(C)
household size	7.537	7.722	-3.200	-1.100	0.273	0.76*
AGE	37.078	37.217	-1.400	-0.510	0.608	1.050
Age squared	1521.600	1524.300	-0.300	-0.120	0.907	1.12*
married	0.528	0.519	1.8	0.700	0.485	.
primary education	0.151	0.157	-1.500	-0.580	0.560	.
secondary education	0.735	0.734	.2	0.070	0.941	.
diploma certificate	0.078	0.069	3.8	1.290	0.199	.
higher levels	0.008	0.010	-1.500	-0.560	0.577	.
Traditional	0.115	0.114	.2	0.080	0.934	.
Farms	0.038	0.036	.8	0.420	0.675	.
Mining Areas	0.027	0.023	1.9	0.980	0.329	.
presence of kids below 7yrs	0.135	0.136	-0.300	-0.120	0.908	.
presence of kids between 7&14yrs	0.252	0.251	.3	0.090	0.928	.
maternity	0.411	0.392	3.9	1.460	0.144	.

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.001	8.12	0.883	1.5	1.4	7.5	1.11	67

* if B>25%, R outside [0.5; 2]

i. Female and Non-public Sector

Logistic regression

Number of obs = 6,581
LR chi2(14) = 704.94
Prob > chi2 = 0.0000
Pseudo R2 = 0.0823

Log likelihood = -3931.2102

treat	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
household size	-0.017	0.005	-3.470	0.001	-0.027	-0.008
AGE	0.024	0.004	6.330	0.000	0.016	0.031
Age squared	0.000	0.000	0.740	0.460	-0.000	0.000
married	-0.775	0.058	-13.420	0.000	-0.889	-0.662
primary education	0.169	0.157	1.080	0.282	-0.138	0.475
secondary education	0.129	0.149	0.870	0.386	-0.163	0.422
diploma certificate	0.205	0.183	1.120	0.264	-0.154	0.564
higher levels	-1.640	0.323	-5.080	0.000	-2.273	-1.007
Traditional	-0.376	0.080	-4.670	0.000	-0.533	-0.218
Farms	-1.892	0.109	-17.310	0.000	-2.106	-1.678
Mining Area	-0.592	0.185	-3.200	0.001	-0.953	-0.230
_PRESENCE OF KIDS BELOW 7YRS	-0.216	0.085	-2.550	0.011	-0.382	-0.050
_PRESENCE OF KIDS BETWEN 7 TO 14YRS	-0.034	0.074	-0.470	0.642	-0.180	0.111
maternity	0.052	0.058	0.890	0.375	-0.062	0.166
_cons	0.259	0.204	1.270	0.204	-0.141	0.659

Treatment assignment	Off Support	On Support	Total
Untreated	0	2,341	2,341
Treated	13	4,227	4,240
Total	13	6,568	6,581

Variable	Treated	Control	%bias	t	p>t	V(C)
household size	7.947	8.009	-1.000	-0.450	0.653	1.19*
AGE	39.251	39.059	1.9	0.850	0.395	1.010
Age squared	1658.200	1632.100	3.1	1.400	0.161	0.980
married	0.372	0.387	-2.900	-1.360	0.173	.
primary education	0.149	0.141	2.1	0.990	0.323	.
secondary education	0.744	0.742	.4	0.200	0.841	.
diploma certificate	0.068	0.073	-2.000	-0.890	0.371	.
higher levels	0.004	0.004	-0.200	-0.180	0.861	.
Traditional	0.130	0.131	-0.100	-0.050	0.956	.
Farms	0.035	0.037	-0.900	-0.630	0.527	.
Mining Area	0.017	0.018	-0.800	-0.400	0.693	.
PRESENCE OF KIDS BELOW 7YRS	0.127	0.131	-1.100	-0.540	0.587	.
PRESENCE OF	0.330	0.320	2.1	0.990	0.323	.

KIDS BETWEN 7 TO 14YRS						
maternity	0.395	0.405	-2.100	-0.960	0.339	.

* if variance ratio outside [0.94; 1.06]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.001	10.03	0.760	1.5	1.5	6.9	1.11	33

* if B>25%, R outside [0.5; 2]

Female-dominated versus Mixed Occupations

a. Pooled

Logistic regression

Number of obs = 17,924
LR chi2(14) = 1384.48
Prob > chi2 = 0.0000
Log likelihood = -11731.242
Pseudo R2 = 0.0557

treat	Coef.	Std.Err.	z	P>z	95%Conf.	Interval]
household size	0.008	0.003	2.540	0.011	0.002	0.014
AGE	0.009	0.002	4.260	0.000	0.005	0.013
Age squared	0.000	0.000	5.290	0.000	0.000	0.000
married	-0.205	0.033	-6.190	0.000	-0.270	-0.140
primary education	-0.246	0.151	-1.630	0.103	-0.542	0.050
secondary education	-1.077	0.143	-7.530	0.000	-1.358	-0.797
diploma certificate	-0.822	0.150	-5.490	0.000	-1.115	-0.528
higher levlels	-3.049	0.172	-17.740	0.000	-3.386	-2.712
Traditional	0.033	0.046	0.710	0.478	-0.058	0.123
Farms	0.170	0.103	1.650	0.099	-0.032	0.372
Mining Areas	0.131	0.109	1.200	0.229	-0.083	0.344
presence of kids below 7yrs	0.126	0.051	2.470	0.013	0.026	0.226
presence of kids between 7&14yrs	0.252	0.044	5.770	0.000	0.167	0.338
maternity	-0.087	0.033	-2.670	0.008	-0.151	-0.023
_cons	0.484	0.161	3.010	0.003	0.168	0.799

Treatment assignment	Off Support	On Support	Total
Untreated	0	8,896	8,896
Treated	6	9,022	9,028
Total	6	17,918	17,924

Balancing Test Results

Variable	Treated	Control	%bias	t	p>t	V(C)
household size	7.900	8.010	-1.800	-1.120	0.261	0.990
AGE	38.899	38.792	1	0.690	0.490	1.05*
Age squared	1653.400	1655.900	-0.300	-0.200	0.845	1.000
married	0.444	0.455	-2.200	-1.510	0.132	.
primary education	0.129	0.128	.4	0.220	0.824	.
secondary education	0.703	0.702	.2	0.140	0.889	.
diploma certificate	0.128	0.130	-0.600	-0.360	0.718	.
higher levlels	0.014	0.014	0.000	0.000	1.000	.
Traditional	0.144	0.143	.1	0.090	0.926	.
Farms	0.031	0.030	.5	0.330	0.739	.
Mining Areas	0.021	0.021	0	0.010	0.991	.
presence of kids below 7yrs	0.125	0.126	-0.500	-0.340	0.733	.
presence of kids between 7&14yrs	0.310	0.320	-2.200	-1.450	0.146	.
maternity	0.484	0.488	-0.800	-0.550	0.583	.

* if variance ratio outside [0.96; 1.04]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.000	6.64	0.948	0.8	0.5	3.8	1.07	33

* if B>25%, R outside [0.5; 2]

b. Male Only

Logistic regression

Number of obs = 8,504
LR chi2(14) = 605.89
Prob > chi2 = 0.0000
Log likelihood = -5477.7972
Pseudo R2 = 0.0524

treat	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
household size	0.017	0.005	3.520	0.000	0.007	0.026
AGE	0.000	0.003	0.010	0.994	-0.006	0.006
Age squared	0.000	0.000	2.890	0.004	0.000	0.000
married	0.082	0.051	1.600	0.110	-0.019	0.183
primary education	-0.133	0.206	-0.650	0.517	-0.537	0.270
secondary education	-1.010	0.195	-5.190	0.000	-1.391	-0.628
diploma certificate	-0.746	0.205	-3.630	0.000	-1.149	-0.344
higher levels	-3.038	0.251	-12.110	0.000	-3.530	-2.547
Traditional	0.080	0.069	1.150	0.250	-0.056	0.216
Farms	0.368	0.147	2.500	0.012	0.079	0.656
Mining Areas	0.308	0.150	2.060	0.040	0.015	0.601
presence of kids	0.160	0.079	2.030	0.042	0.006	0.313

below 7yrs						
presence of kids between 7&14yrs	0.255	0.069	3.670	0.000	0.118	0.391
maternity	-0.130	0.048	-2.730	0.006	-0.223	-0.037
_cons	0.282	0.221	1.280	0.202	-0.151	0.715

Treatment assignment	Off Support	On Support	Total
Untreated	0	4,946	4,946
Treated	3	3,555	3,558
Total	3	8,501	8,504

Balancing Test Results

Variable	**Treated**	**Control**	**%bias**	**t**	**p>t**	**V(C)**
household size	7.662	7.673	-0.200	-0.070	0.943	0.84*
AGE	37.601	37.365	2.3	0.980	0.330	1.10*
Age squared	1565.300	1563.200	.3	0.110	0.915	1.13*
married	0.533	0.544	-2.100	-0.870	0.385	.
primary education	0.132	0.129	.9	0.320	0.752	.
secondary education	0.716	0.718	-0.500	-0.210	0.833	.
diploma certificate	0.115	0.116	-0.300	-0.110	0.911	.
higher levels	0.012	0.012	0	0.000	1.000	.
Traditional	0.132	0.132	-0.100	-0.040	0.972	.
Farms	0.033	0.030	2	0.750	0.454	.
Mining Areas	0.027	0.030	-2.300	-0.890	0.373	.
presence of kids below 7yrs	0.121	0.122	-0.200	-0.070	0.942	.
presence of kids between 7&14yrs	0.270	0.270	0.000	-0.010	0.989	.
maternity	0.471	0.459	2.5	1.070	0.285	.

* if variance ratio outside [0.94; 1.07]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.001	5.89	0.969	1.0	0.4	5.8	1.12	100

* if B>25%, R outside [0.5; 2]

c. Female Only

Logistic regression

Number of obs = 9,480
LR chi2(14) = 923.74
Prob > chi2 = 0.0000
Log likelihood = -5983.1978
Pseudo R2 = 0.0717

treat	**Coef.**	**Std.Err.**	**z**	**P>z**	**[95%Conf.**	**Interval]**

household size	-0.007	0.004	-1.760	0.079	-0.015	0.001
AGE	0.015	0.003	4.820	0.000	0.009	0.021
Age squared	0.000	0.000	2.030	0.043	0.000	0.000
married	-0.328	0.046	-7.100	0.000	-0.419	-0.238
primary education	-0.135	0.222	-0.610	0.542	-0.570	0.299
secondary education	-1.072	0.209	-5.120	0.000	-1.483	-0.662
diploma certificate	-0.967	0.217	-4.450	0.000	-1.393	-0.541
higher levels	-3.386	0.247	-13.720	0.000	-3.869	-2.902
Traditional	-0.016	0.063	-0.250	0.804	-0.139	0.108
Farms	0.085	0.148	0.580	0.563	-0.204	0.374
Mining Areas	0.183	0.172	1.060	0.288	-0.155	0.521
presence of kids below 7yrs	-0.097	0.069	-1.410	0.158	-0.232	0.038
presence of kids between 7&14yrs	0.082	0.059	1.400	0.161	-0.033	0.197
maternity	-0.076	0.046	-1.660	0.097	-0.166	0.014
_cons	0.930	0.236	3.940	0.000	0.467	1.393

Treatment assignment	Off Support	On Support	Total
Untreated	0	3,969	3,969
Treated	2	5,509	5,511
Total	2	9,478	9,480

Balancing Test Results

Variable	Treated	Control	%bias	t	p>t	V(C)
household size	7.882	7.711	2.7	1.420	0.156	1.14*
AGE	39.638	39.648	-0.100	-0.050	0.959	1.040
Age squared	1692.200	1690.100	.3	0.130	0.898	1.000
married	0.374	0.370	.7	0.360	0.715	.
primary education	0.132	0.128	1.4	0.610	0.542	.
secondary education	0.699	0.698	.2	0.100	0.917	.
diploma certificate	0.130	0.132	-0.700	-0.340	0.735	.
higher levels	0.012	0.012	0	0.000	1.000	.
Traditional	0.152	0.145	1.9	0.980	0.329	.
Farms	0.032	0.028	2.5	1.240	0.216	.
Mining Areas	0.018	0.015	2.6	1.360	0.174	.
presence of kids below 7yrs	0.126	0.124	.7	0.370	0.708	.
presence of kids between 7&14yrs	0.330	0.325	.9	0.480	0.633	.
maternity	0.480	0.490	-2.000	-1.060	0.290	.

* if variance ratio outside [0.95; 1.05]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.001	9.77	0.779	1.2	0.8	6.0	1.20	33

* if B>25%, R outside [0.5; 2]

d. Public Sector

Logistic regression

Number of obs = 3,608
LR chi2(14) = 497.94
Prob > chi2 = 0.0000
Log likelihood = -2237.6944
Pseudo R2 = 0.1001

treat	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
household size	-0.005	0.007	-0.750	0.452	-0.019	0.008
AGE	0.016	0.005	3.000	0.003	0.006	0.026
Age squared	0.000	0.000	2.000	0.046	0.000	0.000
married	-0.420	0.077	-5.430	0.000	-0.571	-0.268
primary education	-0.559	0.651	-0.860	0.391	-1.834	0.717
secondary education	-1.386	0.630	-2.200	0.028	-2.620	-0.151
diploma certificate	-0.529	0.634	-0.840	0.403	-1.772	0.713
higher levels	-3.156	0.643	-4.910	0.000	-4.416	-1.897
Traditional	0.123	0.092	1.330	0.183	-0.058	0.303
Farms	0.505	0.311	1.620	0.104	-0.105	1.115
Mining Areas	0.477	0.278	1.720	0.086	-0.068	1.021
presence of kids below 7yrs	0.051	0.116	0.440	0.658	-0.176	0.279
presence of kids between 7&14yrs	0.150	0.098	1.530	0.125	-0.042	0.341
maternity	-0.138	0.086	-1.610	0.107	-0.306	0.030
_cons	0.901	0.662	1.360	0.174	-0.397	2.198

Treatment assignment	Off Support	On Support	Total
Untreated	0	1,644	1,644
Treated	17	1,947	1,964
Total	17	3,591	3,608

Balancing Test Results

Variable	Treated	Control	%bias	t	p>t	V(C)
household size	8.157	8.219	-1.000	-0.300	0.764	0.920
AGE	41.078	40.630	4.5	1.390	0.164	1.10*
Age squared	1836.500	1808.300	3.5	1.100	0.272	1.10*
married	0.474	0.478	-0.900	-0.290	0.773	.
primary education	0.067	0.071	-1.600	-0.420	0.675	.
secondary education	0.582	0.573	2	0.620	0.538	.

diploma certificate	0.311	0.314	-0.800	-0.210	0.836	.
higher levels	0.034	0.034	0	0.000	1.000	.
Traditional	0.213	0.190	5.8	1.770	0.077	.
Farms	0.017	0.020	-2.000	-0.560	0.578	.
Mining Areas	0.020	0.021	-1.100	-0.320	0.752	.
presence of kids below 7yrs	0.126	0.126	-0.200	-0.050	0.957	.
presence of kids between 7&14yrs	0.359	0.355	.8	0.240	0.812	.
maternity	0.735	0.705	7.1	2.120	0.034	.

* if variance ratio outside [0.91; 1.09]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.002	12.22	0.589	2.2	1.4	11.2	0.92	67

* if B>25%, R outside [0.5; 2]

e. Non-Public Sector

Logistic regression

Number of obs = 14,169
LR chi2(14) = 1137.13
Prob > chi2 = 0.0000
Pseudo R2 = 0.0579

Log likelihood = -9250.6218

treat	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
household size	0.000	0.003	0.050	0.963	-0.007	0.007
AGE	0.011	0.002	4.780	0.000	0.007	0.016
Age squared	0.000	0.000	2.500	0.012	0.000	0.000
married	-0.202	0.037	-5.430	0.000	-0.274	-0.129
primary education	-0.389	0.158	-2.460	0.014	-0.699	-0.079
secondary education	-1.223	0.150	-8.130	0.000	-1.518	-0.928
diploma certificate	-1.481	0.161	-9.190	0.000	-1.797	-1.165
higher levels	-3.716	0.219	-16.940	0.000	-4.145	-3.286
Traditional	-0.115	0.055	-2.110	0.034	-0.222	-0.008
Farms	0.063	0.109	0.580	0.560	-0.150	0.277
Mining Areas	0.137	0.121	1.130	0.257	-0.100	0.374
presence of kids below 7yrs	0.203	0.058	3.510	0.000	0.090	0.317
presence of kids between 7&14yrs	0.296	0.050	5.930	0.000	0.198	0.394
maternity	-0.164	0.036	-4.500	0.000	-0.236	-0.093
_cons	0.704	0.171	4.110	0.000	0.368	1.040

Treatment assignment	Off Support	On Support	Total
Untreated	0	7,204	7,204
Treated	15	6,950	6,965
Total	15	14,154	14,169

Balancing Test Results

Variable	Treated	Control	%bias	t	p>t	V(C)
household size	7.738	7.647	1.4	0.830	0.409	0.94*
AGE	38.618	38.576	.4	0.230	0.817	1.020
Age squared	1568.600	1583.200	-1.800	-1.040	0.297	1.000
married	0.429	0.434	-0.900	-0.530	0.596	.
primary education	0.154	0.152	.6	0.310	0.759	.
secondary education	0.739	0.736	.5	0.320	0.750	.
diploma certificate	0.070	0.073	-0.800	-0.490	0.621	.
higher levels	0.006	0.006	0	0.000	1.000	.
Traditional	0.125	0.116	2.6	1.510	0.131	.
Farms	0.035	0.031	2.6	1.450	0.146	.
Mining Areas	0.022	0.024	-1.200	-0.680	0.495	.
presence of kids below 7yrs	0.123	0.125	-0.500	-0.280	0.777	.
presence of kids between 7&14yrs	0.294	0.288	1.2	0.690	0.490	.
maternity	0.403	0.402	.2	0.120	0.904	.

* if variance ratio outside [0.95; 1.05]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.001	10.53	0.722	1.1	0.8	5.5	1.13	33

* if B>25%, R outside [0.5; 2]

f. Male and Public sector

Logistic regression

Number of obs = 1,542
LR chi2(14) = 230.41
Prob > chi2 = 0.0000
Log likelihood = -953.05674
Pseudo R2 = 0.1078

treat	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
household size	0.020	0.012	1.710	0.088	-0.003	0.043
AGE	0.012	0.008	1.570	0.117	-0.003	0.028
Age squared	0.000	0.000	0.830	0.404	-0.000	0.000
married	-0.191	0.133	-1.430	0.153	-0.452	0.071
primary education	-0.858	1.099	-0.780	0.435	-3.012	1.297
secondary education	-2.065	1.063	-1.940	0.052	-4.149	0.019
diploma certificate	-1.222	1.068	-1.140	0.253	-3.316	0.872
higher levels	-3.873	1.086	-3.570	0.000	-6.001	-1.744
Traditional	0.174	0.148	1.170	0.240	-0.117	0.465
Farms	0.419	0.565	0.740	0.458	-0.688	1.526
Mining Areas	0.804	0.422	1.910	0.057	-0.022	1.631

presence of kids below 7yrs	0.099	0.191	0.520	0.605	-0.275	0.473
presence of kids between 7&14yrs	0.152	0.162	0.940	0.347	-0.165	0.470
maternity	-0.640	0.138	-4.650	0.000	-0.911	-0.370
_cons	1.662	1.104	1.510	0.132	-0.501	3.826

Treatment assignment	Off Support	On Support	Total
Untreated	0	792	792
Treated	6	744	750
Total	6	1,536	1,542

Balancing Test Results

Variable	**Treated**	**Control**	**%bias**	**t**	**p>t**	**V(C)**
household size	8.171	8.510	-5.700	-1.030	0.304	0.990
AGE	40.269	39.599	6.8	1.290	0.197	1.20*
Age squared	1829.400	1794.300	4.3	0.830	0.406	1.26*
married	0.575	0.624	-9.800	-1.910	0.057	.
primary education	0.074	0.062	5.7	0.930	0.354	.
secondary education	0.610	0.601	1.8	0.340	0.730	.
diploma certificate	0.280	0.302	-5.700	-0.970	0.332	.
higher levels	0.028	0.028	0	0.000	1.000	.
Traditional	0.184	0.183	.4	0.070	0.947	.
Farms	0.013	0.017	-3.800	-0.630	0.529	.
Mining Areas	0.024	0.022	1.5	0.260	0.796	.
presence of kids below 7yrs	0.117	0.126	-2.700	-0.520	0.606	.
presence of kids between 7&14yrs	0.331	0.339	-1.800	-0.330	0.742	.
maternity	0.724	0.706	4.4	0.780	0.438	.

* if variance ratio outside [0.87; 1.15]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.006	12.17	0.593	3.9	4.1	18.1	0.96	67

* if B>25%, R outside [0.5; 2]

g. Female and Public sector

Logistic regression

Number of obs = 2,048
LR chi2(14) = 348.31
Prob > chi2 = 0.0000
Pseudo R2 = 0.1259

Log likelihood = -1209.1967

treat	**Coef.**	**Std.Err.**	**z**	**P>z**	**[95%Conf.**	**Interval]**
household size	-0.001	0.009	-0.160	0.874	-0.019	0.017

AGE	0.024	0.007	3.480	0.001	0.010	0.038
Age squared	0.000	0.000	0.090	0.930	-0.000	0.000
married	-0.489	0.103	-4.750	0.000	-0.691	-0.287
primary education	-0.382	0.691	-0.550	0.580	-1.736	0.972
secondary education	-0.958	0.660	-1.450	0.146	-2.251	0.335
diploma certificate	-0.144	0.667	-0.220	0.829	-1.453	1.164
higher levels	-3.074	0.682	-4.510	0.000	-4.410	-1.737
Traditional	0.138	0.123	1.130	0.259	-0.102	0.379
Farms	0.394	0.377	1.050	0.296	-0.344	1.132
Mining Areas	0.326	0.428	0.760	0.447	-0.513	1.164
presence of kids below 7yrs	-0.076	0.153	-0.500	0.620	-0.376	0.224
presence of kids between 7&14yrs	-0.082	0.130	-0.630	0.530	-0.336	0.173
maternity	0.144	0.118	1.220	0.223	-0.087	0.374
_cons	0.493	0.707	0.700	0.485	-0.892	1.879

Treatment assignment	Off Support	On Support	Total
Untreated	0	832	832
Treated	3	1,213	1,216
Total	3	2,045	2,048

Balancing Test Results

Variable	**Treated**	**Control**	**%bias**	**t**	**p>t**	**V(C)**
household size	8.166	8.035	2.1	0.510	0.609	1.020
AGE	41.239	40.634	6	1.500	0.135	1.080
Age squared	1830.400	1802.700	3.3	0.820	0.414	1.010
married	0.394	0.395	-0.200	-0.060	0.953	.
primary education	0.062	0.064	-1.200	-0.260	0.797	.
secondary education	0.547	0.546	.3	0.070	0.946	.
diploma certificate	0.348	0.347	.2	0.030	0.973	.
higher levels	0.033	0.033	0	0.000	1.000	.
Traditional	0.233	0.210	5.7	1.380	0.167	.
Farms	0.021	0.021	.4	0.090	0.929	.
Mining Areas	0.015	0.018	-3.100	-0.690	0.491	.
presence of kids below 7yrs	0.138	0.140	-0.800	-0.200	0.843	.
presence of kids between 7&14yrs	0.361	0.358	.6	0.150	0.883	.
maternity	0.747	0.723	5.4	1.320	0.187	.

* if variance ratio outside [0.89; 1.12]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.002	6.93	0.937	2.1	1.0	10.7	1.15	0

* if B>25%, R outside [0.5; 2]

h. Male and Non-public sector

Logistic regression

Number of obs = 6,849
LR chi2(14) = 471.45
Prob > chi2 = 0.0000
Log likelihood = -4392.6438
Pseudo R2 = 0.0509

treat	**Coef.**	**Std.Err.**	**z**	**P>z**	**[95%Conf.**	**Interval]**
household size	0.019	0.005	3.500	0.000	0.008	0.030
AGE	0.001	0.003	0.270	0.788	-0.006	0.008
Age squared	0.000	0.000	1.440	0.149	-0.000	0.000
married	0.104	0.057	1.820	0.068	-0.008	0.215
primary education	-0.181	0.228	-0.800	0.426	-0.627	0.265
secondary education	-1.090	0.216	-5.040	0.000	-1.514	-0.666
diploma certificate	-1.114	0.232	-4.810	0.000	-1.568	-0.659
higher levels	-3.167	0.302	-10.470	0.000	-3.759	-2.574
Traditional	0.048	0.080	0.590	0.552	-0.109	0.205
Farms	0.404	0.153	2.630	0.009	0.103	0.704
Mining Areas	0.321	0.163	1.970	0.048	0.002	0.641
presence of kids below 7yrs	0.289	0.086	3.340	0.001	0.120	0.458
presence of kids between 7&14yrs	0.291	0.078	3.710	0.000	0.137	0.445
maternity	-0.188	0.053	-3.520	0.000	-0.293	-0.083
_cons	0.328	0.245	1.340	0.181	-0.153	0.808

Treatment assignment	Off Support	On Support	Total
Untreated	0	4,061	4,061
Treated	4	2,784	2,788
Total	4	6,845	6,849

Balancing Test Results

Variable	**Treated**	**Control**	**%bias**	**t**	**p>t**	**V(C)**
household size	7.641	7.546	1.6	0.560	0.575	0.85*
AGE	36.968	36.702	2.7	0.980	0.325	1.10*
Age squared	1489.800	1487.100	.3	0.120	0.906	1.08*
married	0.519	0.517	.4	0.160	0.872	.
primary education	0.147	0.149	-0.500	-0.170	0.865	.
secondary	0.740	0.737	.8	0.270	0.784	.

education						
diploma certificate	0.079	0.082	-1.000	-0.370	0.712	.
higher levels	0.009	0.009	0	0.000	1.000	.
Traditional	0.123	0.122	.2	0.080	0.935	.
Farms	0.038	0.036	1.1	0.350	0.723	.
Mining Areas	0.028	0.031	-2.000	-0.670	0.504	.
presence of kids below 7yrs	0.133	0.133	.1	0.040	0.968	.
presence of kids between 7&14yrs	0.254	.25	1	0.370	0.711	.
maternity	0.394	0.377	3.3	1.270	0.205	.

* if variance ratio outside [0.93; 1.08]

Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.001	4.31	0.993	1.1	0.9	5.6	1.05	100

* if B>25%, R outside [0.5; 2]

i. Female and Non-public sector

Logistic regression

Number of obs = 7,362
LR chi2(14) = 680.47
Prob > chi2 = 0.0000
Log likelihood = -4669.3509
Pseudo R2 = 0.0679

treat	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
household size	-0.010	0.005	-2.220	0.026	-0.019	-0.001
AGE	0.017	0.004	4.640	0.000	0.010	0.024
Age squared	-0.000	0.000	-0.290	0.770	-0.000	0.000
married	-0.308	0.052	-5.880	0.000	-0.411	-0.206
primary education	-0.507	0.237	-2.140	0.032	-0.971	-0.044
secondary education	-1.343	0.225	-5.950	0.000	-1.785	-0.901
diploma certificate	-1.694	0.239	-7.090	0.000	-2.162	-1.226
higher levels	-3.989	0.314	-12.720	0.000	-4.603	-3.374
Traditional	-0.126	0.077	-1.640	0.101	-0.276	0.025
Farms	-0.009	0.161	-0.060	0.956	-0.325	0.307
Mining Areas	0.055	0.191	0.290	0.771	-0.318	0.429
presence of kids below 7yrs	0.013	0.078	0.170	0.866	-0.139	0.166
presence of kids between 7&14yrs	0.129	0.067	1.930	0.053	-0.002	0.260
maternity	-0.225	0.051	-4.380	0.000	-0.326	-0.125
_cons	1.333	0.257	5.180	0.000	0.829	1.838

Treatment assignment	Off Support	On Support	Total
Untreated	3,096	3,096	3,096

Treated	4,266	4,266	4,266
Total	7,362	7,362	7,362

Balancing Test Results

Variable	Treated	Control	%bias	t	p>t	V(C)
household size	7.744	7.505	3.7	1.760	0.079	1.16*
AGE	39.291	39.192	1	0.440	0.663	1.010
Age squared	1654.700	1651.300	.4	0.180	0.856	0.970
married	0.370	0.364	1.3	0.600	0.552	.
primary education	0.155	0.150	1.7	0.650	0.517	.
secondary education	0.731	0.726	1.2	0.540	0.592	.
diploma certificate	0.072	0.079	-2.300	-1.150	0.251	.
higher levels	0.005	0.005	0	0.000	1.000	.
Traditional	0.128	0.111	5.2	2.420	0.016	.
Farms	0.033	0.032	.4	0.150	0.879	.
Mining Areas	0.018	0.015	1.9	0.890	0.373	.
presence of kids below 7yrs	0.130	0.125	1.4	0.650	0.516	.
presence of kids between 7&14yrs	0.321	0.310	2.4	1.110	0.269	.
maternity	0.398	0.405	-1.400	-0.670	0.501	.

* if variance ratio outside [0.94; 1.06]

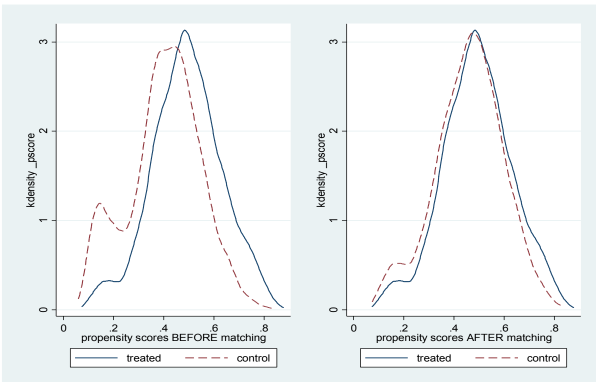
Ps R2	LR chi2	p>chi2	MeanBias	MedBias	B	R	%Var
0.001	12.75	0.547	1.7	1.4	7.7	1.01	33

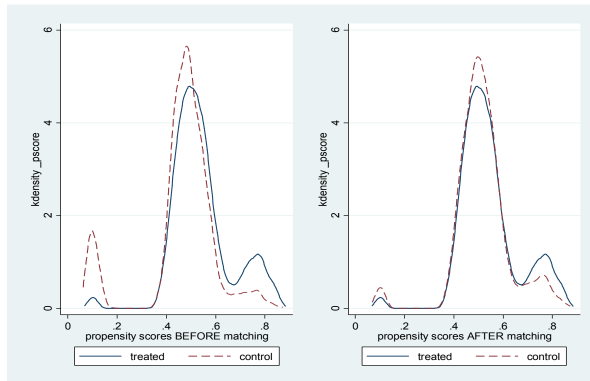
* if B>25%, R outside [0.5; 2]

Section D: Common Support Graphs

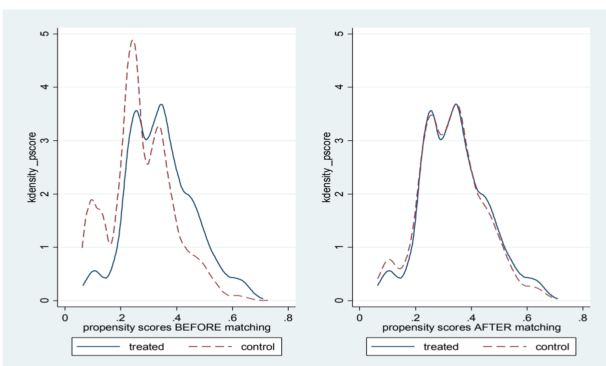
Male-Dominated Versus Female Dominated | **Mixed Occupations Versus Female-Dominated**

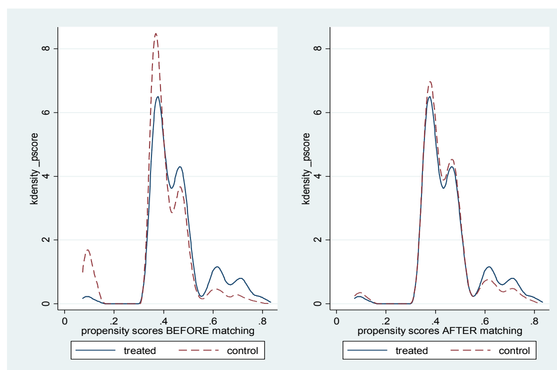
Pooled Data



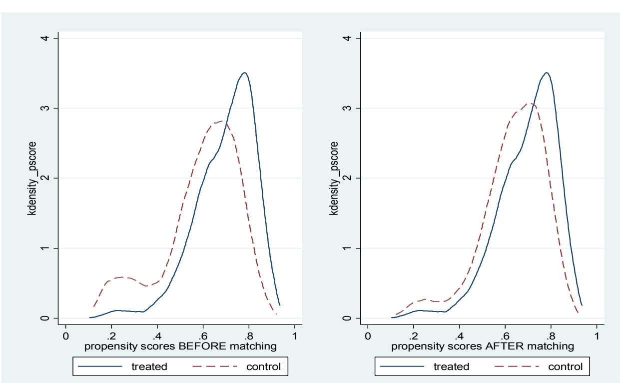


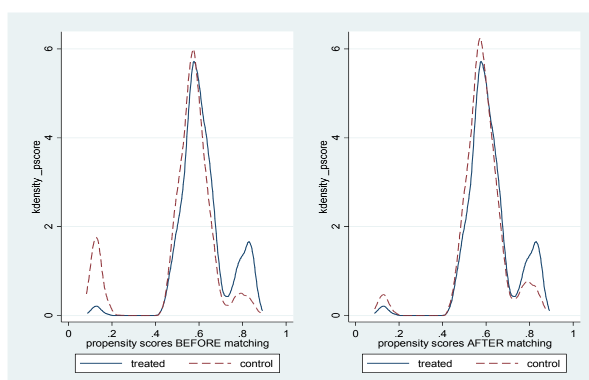
Male Only



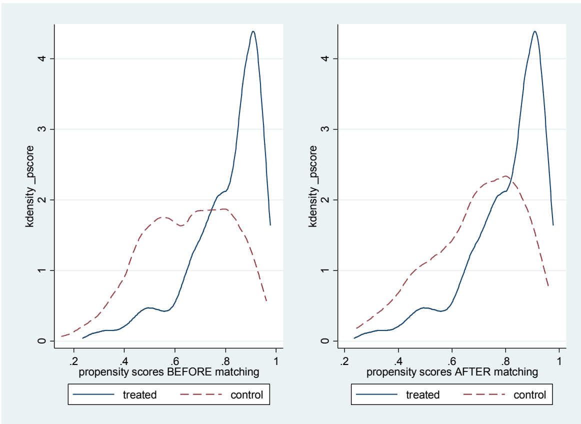


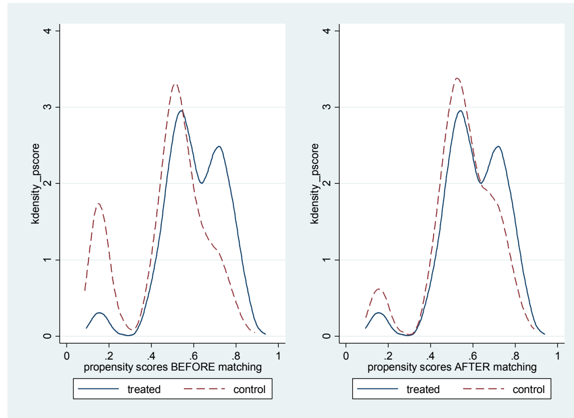
Female Only



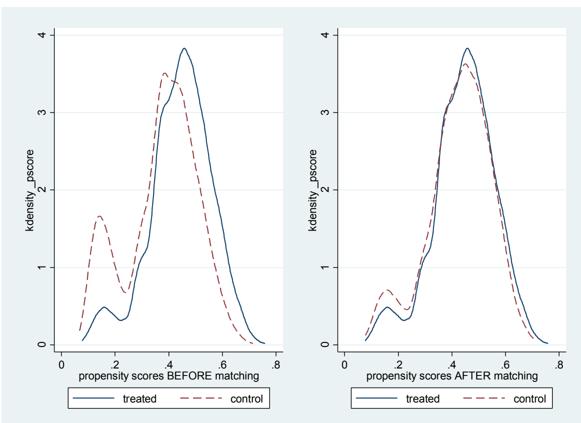


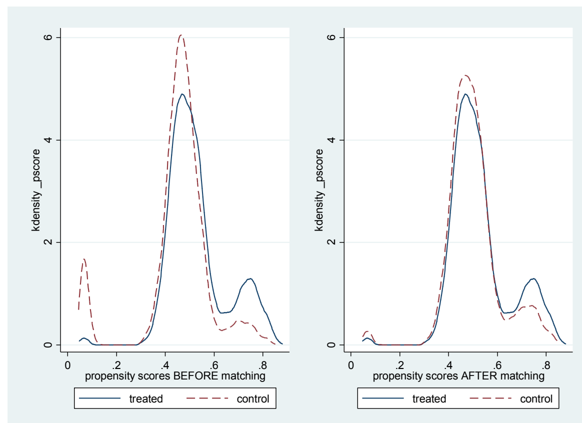
Public Sector



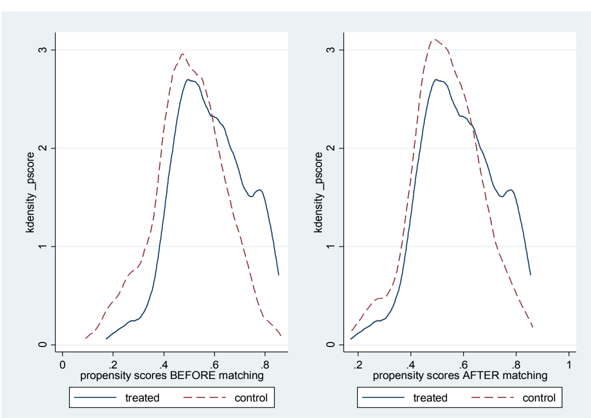


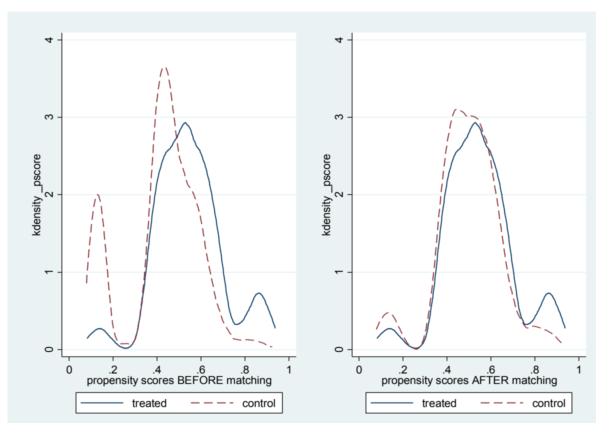
Non-public Sector



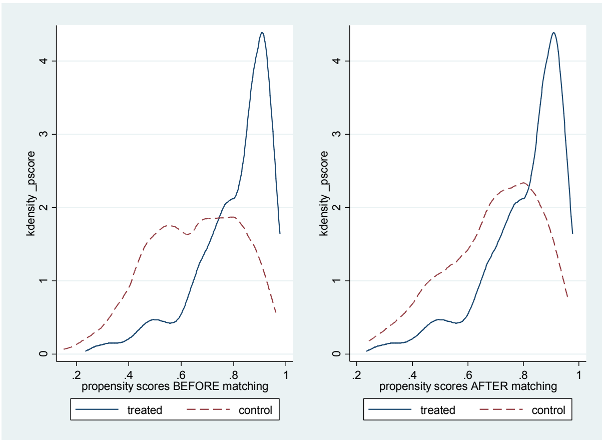


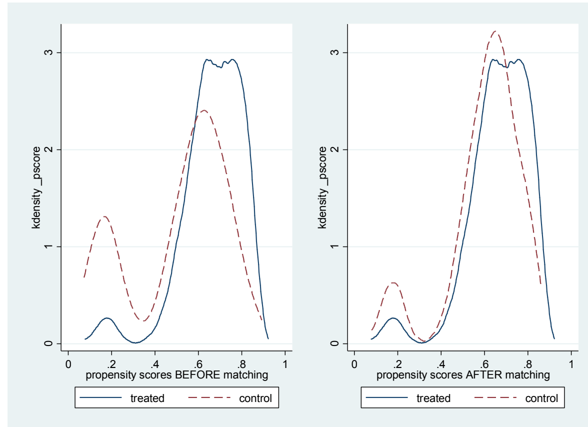
Male and Public Sector



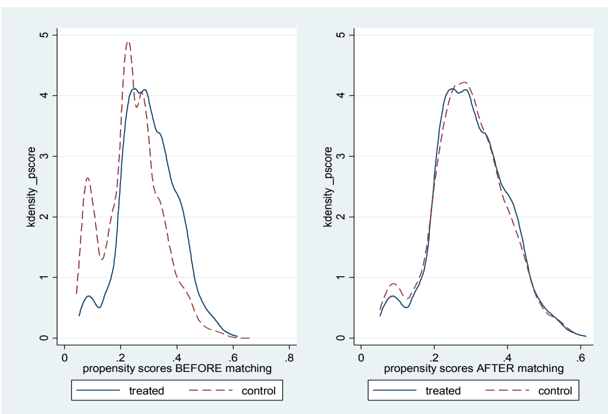


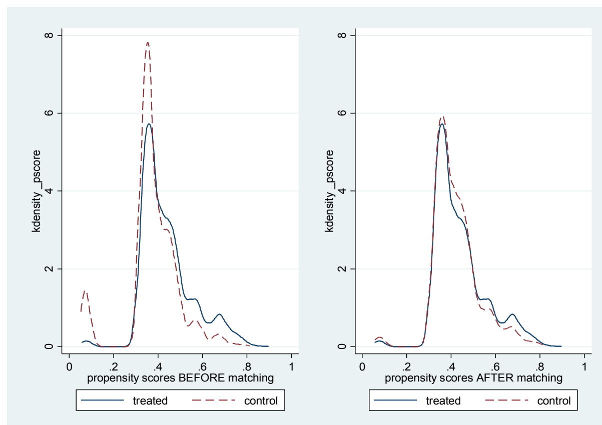
Female and Public Sector





Male and Non-Public Sector





Female and Non-Public Sector