WIDER Working Paper 2024/80

The firm-wage gender gap and formal sector churn over the life cycle

Ihsaan Bassier[1] and Leila Gautham[2]

December 2024

United Nations University World Institute for Development Economics Research

wider.unu.edu

Abstract: We find that women sorting into lower wage firms explains nearly half of the gender wage gap in South Africa, using matched employer-employee panel data covering the universe of formal sector workers. Sorting varies considerably over the life cycle: the firm-wage gender gap is negligible for the youngest workers, grows steeply for 25–35-year-olds (i.e. typical child-rearing years), and narrows for older workers. The increase is driven by those continuously employed---while women are almost as likely as men to switch firms, men are more likely to switch to better-paying firms, consistent with discrimination or non-wage amenities. Churn also contributes to the gap (though is relatively constant), since women enter formal employment at worse-paying firms than men. The importance of these two groups, the continuously employed versus entrants, depends on the size of the formal sector, thus linking the life cycle patterns underlying gender gaps with economic development.

Key words: firm wage premia, gender wage inequality, sorting, worker transitions

JEL classification: J31, J16, J42, J71

Acknowledgements: We are grateful to Joshua Budlender, Nancy Folbre, Peter Howley, Surbhi Kesar, Barbara Petrongolo, and seminar participants at the NXNW Labour Economics Workshop at the University of Manchester for helpful comments, as well as to Noreen Kajugusi for excellent research assistance. We are grateful to Southern Africa – Towards Inclusive Economic Development (SA-TIED) for granting access to unique administrative firm data.

[1] School of Economics, University of Surrey, UK, i.bassier@surrey.ac.uk; [2] Department of Economics, University of Leeds, UK, l.gautham@leeds.ac.uk

This study is published within the UNU-WIDER project Southern Africa—Towards Inclusive Economic Development (SA-TIED).

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9267-543-1

https://doi.org/10.35188/UNU-WIDER/2024/543-1

Typescript prepared by Siméon Rapin.

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland and Sweden, as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

Firm sorting matters a great deal for gender wage inequality. We begin by showing that the sorting of women into lower wage firms accounts for 45 per cent, or nearly half, of the mean gender wage gap for the universe of formal sector South African workers. The dominant role of sorting for gender wage gaps is consistent with findings from other settings (Card et al. 2016; Li et al. 2023). However, the mechanisms underlying sorting are less well understood. In this paper we focus exclusively on why women are concentrated among low-paying employers. In particular, given that frequent job switching dynamically reproduces sorting patterns, we make progress on this literature by focusing on *changes* in sorting over the life cycle with worker transitions. While we find little sorting for the youngest workers, the firm-wage gender gap (the average difference in firm wage premia between men and women across firms) quickly grows over the early life cycle driven by continuously employed women being less likely than men to move to better-paying firms.

We contribute to the small but growing literature on the evolution of firm pay dynamics over the life cycle (Bruns 2019; Casarico and Lattanzio 2024), where firm wage premia are defined following Abowd et al. (1999). Our first contribution is that we find a striking U-shape in the evolution of the firm-wage gender gap, with an increase in the gap beginning in the mid-20s up until the mid-40s, after which women begin to catch up with men. Firm-wage gender gaps diverge at exactly the period when women are likely to co-reside with their young children, as shown by household survey data, consistent with other findings on life cycle gender pay gaps (Goldin et al. 2017; Barth et al. 2021). Although these studies show some evidence on narrowing gaps after cohorts enter their mid-40s, none (to our knowledge) have documented or investigated this convergence later into the life cycle.

Why does the firm-wage gender gap grow over the first part of the life cycle? We observe that changes in the firm-wage gap can occur either by continuously employed workers switching differentially (by gender) to higher-paying firms, or by workers differentially entering or leaving formal employment. Our second contribution is therefore to decompose these gaps in firm pay dynamics over the life cycle by worker transitions. We find that firm-wage gender gaps over the life cycle are due to both women entering (or re-entering) formal employment in relatively disadvantaged positions, and continuously employed women making less advantageous moves to higher-paying firms. However, much of the *increase* in the firm-wage gap for those aged mid-20s to mid-40s is driven by the continuously employed. To our knowledge, our paper is the first to decompose the evolution of gender gaps in firm premia into the role of differential mobility among the continuously employed and differences in sorting upon entry into the formal labour market, an important distinction which links this literature to studies of the motherhood penalty. Studying the role of churn is all the more important in a developing-country

context where, firstly, the formal sector is smaller and so out-of-sector worker transitions are consequently more important, and secondly, movements from formal sector employment to more flexible forms of informality are a major channel through which child penalties operate (Berniell et al. 2021).

We discuss evidence on three mechanisms. Beginning with why women enter formal employment at worse-paying firms, we show that men's disproportionate entry into high value-added firms covered by collective bargaining agreements, and in high-paying industries, explains two thirds of the firm-wage gender gap among labour market entrants (commuting distance explains close to nothing). Secondly, the steep increase in the firm-wage gender gap over the life cycle among the continuously employed coincides with the ages at which women are most likely to co-reside with young children. While women are slightly less likely than men to switch firms, much of the growth in the firm-wage gap among the continuously employed comes from women moving to lower-paying firms, *conditional* on having moved. This rules out differential moving frictions as the driver, and we instead discuss possible demand- and supply-side reasons. Thirdly, to explain the closing of the gender gap in firm wages for older women, we note that there is a mechanical effect at play: groups sorted lower in the firm wage distribution have more potential to move up than groups already towards the top. While this is true throughout the life cycle, child-related constraints may be relieved in later years, which allows women to move up the job ladder.

Our estimate that firm sorting accounts for nearly half the gender pay gap is higher than existing estimates from high-income country settings, but comparable to the few estimates that come from developing country settings. Monopsony power is plausibly higher in developing countries where weak outside options, thin markets, and high labour market frictions confer significant wage-setting power to firms: South Africa, with high levels of informality and unemployment, offers an illustrative example (Bassier 2023). We provide suggestive evidence linking low levels of formality to greater churn and higher firm-wage gender gaps. This helps explain why firms play such an outsized role in shaping gender wage inequality, and our findings signal the importance of employment segregation across firms in such a developing country context.

2 The overall firm-wage gender gap

2.1 Estimation of firm wage premia

We use matched worker- and firm-level data from South African administrative tax records between 2010 and 2018, made available through a data-sharing agreement with South Africa's National Treasury and UNU-WIDER (2023a, 2023b) (see Bassier 2022 for a detailed descrip-

tion of the data, as well as this paper's Appendix B). Our dataset consists of 8–9 million workers aged between 20 and 60 in each year, resulting in a total sample of 78.8 million worker-year observations with about 34.6 million belonging to women and 44.2 million to men. We observe unique identifiers for workers as well as establishments (throughout the paper, we use the word 'firm' interchangeably with establishment, though our data is really about establishments).

Average real annualized wages for women are 15 per cent lower than for male workers. This is larger than, but similar to, the 13.2 per cent gender gap we estimate using the South African Labour Force Survey (LFS) for a comparable set of workers, with differences possibly arising from sampling variability in the household survey, how formal sector employment is defined, and measurement issues with survey earnings (Kerr and Wittenberg 2019).[1] A caveat to our analysis is that our tax data exclude the informal sector, which forms a larger share of female employment and also has lower wages. Consequently, the gender earnings gap across all workers is likely much larger than in our data: the gender gap including both formal and informal sector workers, for example, is 22.5 per cent in the LFS data, considerably larger than the gap within the formal sector.

To compute the contribution of firms to gender wage inequality in South Africa we implement an Abowd et al. (1999) (henceforth AKM) two-way fixed effects model of the form:

$$Y_{ijt} = \theta_i + \psi_j + X_{ijt}\beta + e_{ijt} \tag{1}$$

where Y_{ijt} are annual wages for worker i in firm j at time t; θ_i are individual fixed effects and ψ_j are firm fixed effects.[2] X_{ijt} indicate controls for tax year, a cubic polynomial in age, and the fraction of months in the year employed at the firm. Following Kline et al. (2020), we implement this through a leave-out estimator which adjusts for mobility bias. Due to the computational intensity of this procedure over the full sample, we compute firm fixed effects

[1] This gap is computed using LFS data (2010–2018) for workers between 20 and 60 (Kerr et al. 2023).

[2] Here, we diverge from existing studies on the contribution of firms to gender wage inequality (Card et al. 2016 and, subsequently, Casarico and Lattanzio 2024 among others) in not computing gender-specific firm premia to examine the role of bargaining or gender differences in firm premia for women and men in the *same* firm. Such estimations necessarily rely on arbitrary normalization of male and female firm effects (typically picking the food and accommodation sector as the reference), and as such, the estimation of the role of bargaining is biased if there are gender gaps in firm premia within the reference sector. Since our focus is on sorting across firms, not within firms, we bypass this issue by using homogeneous firm premia (i.e., estimated over the entire sample of women and men). If women are paid lower premia than men in the same firms, our estimates of the firm contribution to gender wage inequality are underestimates (and vice versa).

over a smaller subset of years (2011–2016).[3] We use the term "firm premia" or "firm wages" interchangeably with these AKM firm fixed effects estimated in equation 1. Following the literature, in order to ensure that our firm premia are computed over a sufficiently large group of workers, we restrict our sample to firms with at least 10 workers of each gender in each period.

Our final dataset, consisting of all workers for whom we have firm premia, consists of 62.6 million worker-year observations (27.3 million female and 35.3 million male).[4] Focusing on workers in larger firms as we do results in a slightly smaller gender wage gap (0.122 log points or 11.5 per cent) in our final analysis sample compared to the gap in the full tax sample. Again, this is comparable to the gender gap (12.5 per cent) in the LFS sample closest to our own (formal sector workers in firms with over 20 workers).

To provide evidence on the plausibility of the exogenous mobility assumptions needed to measure AKM firm premia, we reproduce the checks suggested in Card et al. (2016). We plot mean log wages of job changers, by quartile of mean co-worker wage at origin and destination firm (Appendix Figure A1). Employees who move from jobs with low-paid co-workers experience large average wage gains when they move to jobs with highly-paid co-workers, with symmetric wage losses when workers move in the opposite direction, and flat average trends in wages before and after moving.

A possible concern with our estimates of AKM firm premia is that our data track earnings and months worked, but do not record hours worked or the hourly wage. Therefore, in theory, gender differences in firm premia could be driven by a divergence in hours worked. In Figure A2, we use LFS data to plot average hours worked per week by 2-digit industry (the most granular disaggregation possible in the LFS data): average weekly hours across industries, clustered tightly between 40 to 45 hours, show weakly negative relationships with the share of female employment in that industry, and with industry firm wage premia.[5] In the absence of firm-level data on hours worked, this evidence suggests gender differences in firm premia are not being driven by lower hours worked in female-dominated firms.

[3] We therefore exclude firms in 2010 or 2017–2018 that are absent in 2011–2016. However, for 2010 and 2017–2018, we still observe the distribution of workers across firms and therefore retain these years in our overall analysis.

[4] See Appendix Table A1 for summary statistics on the fullsample and the analysis sample.

[5] This is consistent with literature showing that average hours worked at a firm bears little relation to firm wages (Card et al. 2016; Labanca and Pozzoli 2022).

2.2 The gender gap in firm wage premia

Our first and baseline result is that the firm-wage gender gap is 0.055 log points or 45 per cent of the 0.122 log point raw gender wage gap. As we impose a single wage premium per firm (see equation 1), this difference arises purely from female employment being concentrated more in lower premium firms than men. Figure A3 presents this visually, by showing a clear negative correlation of the share of female workers in a firm with both the firm wage premium and the firm's valued added (as a non-parametric proxy). For instance, firms in the bottom decile by value added have, on average, 43 per cent of their workers being women, compared to only 35 per cent at the top decile. Similarly, ranked by firm premia, firms in the bottom decile have a female share of 53 per cent on average, compared to 41 per cent at the top decile.

Our finding that sorting contributes 45 per cent—or nearly half—of the gender wage gap is strikingly high in absolute magnitude, but also in comparison to findings from high-income country settings. Estimates for the contribution of sorting to the gender wage gap range from 15–20 per cent for Portugal (Card et al. 2016); 21–26 per cent for Italy (Casarico and Lattanzio 2024); and 11 per cent for France (Coudin et al. 2018). Corresponding estimates from countries lower in the income ladder are closer in magnitude: 20–29 per cent in Estonia (Masso et al. 2022), 32–36 per cent in Chile (Cruz and Rau 2022), and 45–57 per cent for Brazil (Morchio and Moser 2021).

We note that our estimates of the contribution of sorting do *not* capture gender differences in pay coming from individual time-invariant characteristics; as such, they also exclude labour-market-wide discrimination against women. Instead, they only capture gender differences in pay coming from women being less likely than men to be employed at firms that pay higher wages (to *all* their workers).[6] This sorting could be due to either worker or firm choices (Morchio and Moser 2021). On the worker side, unpaid work responsibilities—and social norms dictating that women shoulder much of this work—could constrain women from moving to higher-paying firms, either due to the reduced intensity of job search or the provision of better non-pay amenities at such firms.

On the firm side, even when firms have the same pay policy for male and female employees, the firm's fraction of women employed will predict firm premia if firms know that women can be paid less. For instance, in a simple wage-posting model with differential firm labour supply elasticities by gender, the firm premium is the fraction-weighted average of the two

[6] As we do not have information on occupation, firm wages may include differences due to occupations. To the extent that occupations are proxied by industry, we show that sorting across firms *within* industries remains substantial. In terms of comparability of our estimates with prior work, some studies either do not include occupational controls in their estimation of AKM firm premia (Card et al. 2016; Coudin et al. 2018) or include coarse controls such as dummies for white-collar and managerial occupations (Casarico and Lattanzio 2024), or include detailed occupation dummies (Morchio and Moser 2021).

markdowns.[7] Employers might also be less willing to hire women, either in the form of taste-based (Becker 1957) or statistical (Arrow 1971) discrimination. Correspondence studies, for instance, find that women—but not men—are subject to hiring discrimination for expected family responsibilities, varying over women's reproductive age (Petit 2007; He et al. 2023). Therefore, both worker and firm preferences are likely to vary over the life cycle as women become mothers in their 20s, and then experience lessening childcare responsibilities in their 40s (see Figure A13). The next section discusses how firm wage premia evolve over the life cycle for women and men.

3 Gender gaps over the life cycle

Figure 1 presents our estimates of gender gaps over the life cycle. Gaps in overall wages between women and men are either zero, or favor women, up until workers cross their mid-20s.[8] After age 28, a 3 per cent gap between women and men expands to 15 per cent by age 33, then expanding more slowly to reach 27 per cent by age 49, after which it declines.[9] The firm-wage gender gap also follows this U-shape, and its contribution to the raw gap in wages holds steady at 40–50 per cent over much of the life cycle. While other studies investigating life cycle gender wage gaps show some evidence on narrowing gaps after cohorts are in their late 40s, they either terminate their analysis at this point (Barth et al. 2021) or do not investigate the convergence that they observe (Casarico and Lattanzio 2024).

As our data spans 9 calendar years, we cannot directly observe the full life cycle. In the pooled sample, differences across age may reflect differences in gender wage gaps across cohorts, rather than life cycle effects. Figure 1 therefore includes interactions of birth-year with female to allow for cohort differences (we cannot include an unrestricted interaction of birth year with gender in addition to tax year and age due to the well-known collinearity). To allow for a more transparent picture of cohort differences, and to ensure that our findings are not driven by this particular restriction on cohort effects, we also reproduce our main results separately for 10-year birth cohorts. In this case, each 10-year cohort traces out a similar pattern of a U-shape in the firm-wage gap (see Figure A5). We also find similar patterns when we do not include cohort controls (Panel A, Figure A4).

[7] Denoting with subscripts male M and female F for labour L separately, firm profits are $\pi = F(L_M, L_F) - w(L_M + L_F)$. Then allowing for different labour supply elasticities ε by gender (but assuming one wage and equal productivity) and female fraction f, firms set $w = \frac{\partial F}{\partial L} \frac{\varepsilon_m(1-f)+\varepsilon_f f}{1+\varepsilon_m(1-f)+\varepsilon_f f}$, which is the usual markdown formula with the markdown as an f-weighted average of the male and female labour supply elasticities.

[8] Figure A4.B shows the evolution of firm wage premia for women and men separately over the life cycle.

[9] We observe a similar U-shape in the gender gap in wages in the labour force survey data: Figure A8.

Coarse (two-digit) industry codes explain nearly half the firm-wage gap, and about a quarter of the gender gap in wages: within 2-digit industries, gender differences in firm premia range from 28 per cent of the overall wage gap at age 30, gradually falling to about 21 per cent by the early 50s (Figure 1). Finely detailed industry (5-digit SIC codes) explain more of the firm-wage gap, but not by much: firm-wage gaps within 5-digit industry range from 23 per cent of the total wage gap at age 30, declining to 15 per cent by age 50. The contribution of sorting across firms therefore remains substantial, even within these fine industry categories.

4 Gender gaps in labour market transitions

In this section, we attempt to shed more light on why the firm-wage gender gap rises over the lifecycle—why do women increasingly sort into worse-paying firms? First, continuously employed women may be less likely to move to higher-paying firms. Second, women might enter (or re-enter) formal employment in more disadvantaged positions. Third, a higher proportion of women may be more likely to leave and re-enter formal employment (which matters if it is easier to move to higher-paying firms while already employed). We decompose the contribution of each these factors to the evolution of the firm-wage gap over the life cycle.

We split all workers at a particular age into stayers (those present in the previous and next period), entrants (those not present in the previous period), and leavers (those present in the previous, but not the next period). We then decompose the firm-wage gap into gaps within each category and gender differences in population shares of each category. Specifically, we decompose the firm-wage gender gap, separately for each age (in years), in the following way:

$$\psi^f - \psi^m = \underbrace{\sum_{k \in s,l,e} p_k^m (\psi_k^f - \psi_k^m)}_{\text{Gender gaps within category}} + \underbrace{\sum_{k \in s,l} (\psi_k^f - \psi_e^f).(p_k^f - p_k^m)}_{\text{Gender gaps in population shares}} \tag{2}$$

where ψ^f and ψ^m are shorthand representations of average AKM firm effects for women and men, respectively (therefore, corresponding to $E[\psi_j|g = F]$ and $E[\psi_j|g = M]$, respectively, following our notation for equation 2); the subscript s, l, and e denote stayers, leavers and entrants, respectively; and p_k^g is the share of each group $k \in s,l,e$ in the total population of workers belonging to gender $g \in f,m$. We compute this decomposition separately at each age. The first term on the right-hand side collects gender gaps among stayers, leavers, and entrants (weighted by male proportions for each category), while the second term is the sum of the gender difference in the proportion of stayers (weighted by female gaps in premia between stayers and entrants) and the gender difference in the proportion of leavers (weighted by female gaps between leavers and entrants).

Figure 2 shows that the U-shape in the firm-wage gender gap is driven primarily by gender differences in mobility within stayers. That is, from their mid-20s till their mid-40s, women who are continuously employed are less likely than their male counterparts to transition to higher-paying firms, resulting in an increasing gap in firm premia within this category. This changes in their late 40s, when they are *more* likely than men to move to higher-premium firms, resulting in a convergence of the firm-wage gender gap.

The contribution of leavers and entrants is substantial but remains relatively steady over the life cycle.[10] The contribution of firm-wage gaps among leavers to the overall firm-wage gap remains at 17–21 per cent over the lifecyle. On the other hand, the contribution of firm-wage gaps among entrants increases up until the mid-40s and then declines. However, the rate of increase is lower than the overall increase in the gender gap, resulting in a declining contribution of entrants (from 35 per cent in the late 20s to 18 per cent in the mid-50s). The contribution of entrants is higher than that of leavers up until the early 50s.[11]

Finally, the contribution of gender differences in the proportions of stayers and leavers is close to zero. This is surprising, as we might expect churn to be higher for women in their childbearing years. However, proportions of stayers, leavers, and entrants are remarkably similar across gender (Figure A7). Instead the drivers seem to be entirely in the gaps rather than proportions: women (re-)enter formal employment at worse-paying firms than their male counterparts, and fail to move to better-paying firms when continuously employed. And while the former explains a substantial part of the level of the gap in firm wage premia, it is the latter that drives the evolution of gender gaps in firm premia.

Our decomposition remains qualitatively similar when we look simply at the decomposition of the growth of firm premia over time, separately for different birth cohorts, as a more transparent (though cumbersome) approach to decomposing patterns of mobility over the life cycle (Figure A9). We also disaggregate entrants into new entrants (those who have not been observed previously in our data) and re-entrants (those whom we observe prior to a spell of non-employment), but both show a similar, stable contribution to firm-wage gender gaps (Figure A10).

[10] Gender gaps in firm premia among leavers and entrants are actually larger than among stayers—see Figure A6—but their contribution is balanced out by their smaller likelihood of workers belonging to these groups, relative to stayers, who dominate the population.

[11] In a dynamic sense, a smaller (weighted) gap among leavers relative to stayers implies that the average firm-wage gap is bound to grow; Arellano-Bover et al. (2024) document the importance of such cohort effects for the gender wage gap in other settings.

5 Discussion

5.1 Firm-wage gender gaps among entrants and leavers

Formality, churn, and gender gaps in firm premia. The importance of the decomposition of the firm-wage gender gap by labour market transitions in Figure 2 is motivated by the relatively low share of the formal sector in the total working-age population in South Africa (29% according to the 2011 Census). The low share of formal employment in South Africa (a characteristic of developing countries: see ILO 2021) in turn implies high churn or transition rates in and out of formal employment. As noted in section 4, firm-wage gender gaps among entrants and leavers are larger than among stayers (see Figure A6) — for example, due to employment gap penalties or re-training costs. If entrants and leavers are a larger proportion of formal employment in settings where the formal sector is relatively small, the firm-wage gender gap would be correspondingly larger.

We provide suggestive evidence in Figure A11: across regions (municipalities), a smaller share of formal employment is significantly associated with a smaller share of the continuously employed (i.e. stayers) in total formal employment. Additionally, low levels of formality are associated with a wider dispersion of firm-wage premia via low labour supply elasticities (see Bassier 2023): if women are positioned lower in the firm-wage hierarchy, a more dispersed distribution of firm premia would worsen the firm-wage gender gap. We observe a corresponding narrowing of the firm-wage gender gap in regions with high levels of formality (Figure A11).[12] While these estimates are subject to many biases stemming from unobserved heterogeneity across regions, they provide suggestive evidence that the characteristically lower levels of formality in lower income countries may help explain the high contribution of firms to gender wage inequality more generally in relation to high income country estimates.

Predictors of the gap in firm premia among entrants. Why do we observe a high firm-wage gender gap among those entering formal employment ('entrants')? In Table 1, we implement a simple Kitagawa-Oaxaca-Blinder-style decomposition of the 13.4 percentage point firm-wage gender gap among entrants.[13] We simply view this exercise as descriptive (not causal) insight into why women enter employment at worse-paying firms. About two thirds of the gap (8.8 percentage points) is explained by compositional differences in covariates.

[12] Note that we do not account for any selection effects that are likely to arise from the expansion of female formal employment associated with rising formality. Patterns are broadly similar when we use formal employment as a proportion of the labour force.

[13] As commuting distance is available only for a subset of the sample, we present estimates with and without commuting distance as a covariate.

Male entrants are slightly older (see Table A2 for covariate means), which may predict differential ability to move up the firm-wage ladder (Haltiwanger et al. 2018), but this explains only 1 per cent of the gap. Secondly, distance from home to workplace may proxy for supply-side constraints: male workers in our population commute longer distances, and women's unpaid responsibilities may prevent them from commuting to higher-paying firms (Barbanchon et al. 2021; Petrongolo and Ronchi 2020). Surprisingly, commuting distance has a (precisely estimated) zero contribution. Thirdly, women are less likely to work in firms with high rents (as captured by log value-added, see Card et al. 2016) or at firms covered by bargaining council agreements (Bassier 2022; Corradini et al. 2022). These firm- or demand-side factors explain about 21 per cent.

Finally, nearly half the gap (44 per cent) is explained by 2-digit industry (Petrongolo and Ronchi 2020; Folbre et al. 2023). Women are disproportionately employed in sectors like education, retail, personal care services, and some manufacturing subsectors (primarily food products and textiles) that are composed of firms that pay lower wages (see Figure A12). On the other hand, men are concentrated in high-paying sectors like mining, utilities, the manufacture of metals, minerals, and electrical machinery, and construction. Notable exceptions to these patterns include healthcare and public administration that have moderately high firm premia but also a high share of female employment.[14] It appears the explicit gender-progressive government hiring policy has played a substantial role in decreasing the firm-wage gender gap: back-of the-envelope calculations, based on female employment shares while keeping firm premia constant, suggest that this helps reduce the gender wage gap by 17 per cent.[15] Overall, our evidence is suggestive of women's sorting into low-paying firms as driven by firm, rather than worker-side, factors.

5.2 The evolution of gaps among stayers

Women might enter the labour market at worse-paying firms than men, but what explains the evolution of the firm-wage gap among stayers? We seek to explain why employed women are first less likely, then more likely, than their male counterparts to switch to better-firms.

Initial rise in gaps. We note that the rise in the firm-wage gap among stayers that begins in women's mid-20s coincides with the period when they are most likely to become mothers. Women in South Africa experience a 60 percentage point increase in the incidence of motherhood over their 20s (see Figure A13 using survey data). Household care responsibilities—as

[14] Public "firms" are administration units, such as local municipalities or government departments.

[15] Public administration constitutes 25.1 per cent of all female employment, but only 15.9 per cent of male employment. Shifting women away from public administration towards construction (a sector with firm premia close to the average of the distribution), until these shares are equalized, while keeping firm premia constant, would increase the wage gap by 2.1 percentage points (or 17 per cent of the overall gap).

proxied by the fraction of women co-residing with biological children under 10—are at their highest around age 30, and then begin to decline, nearly halving by the time women are in their early 40s, and becoming very small by the time women are in their late 40s.

In Figure 3 we decompose the growth in the firm-wage gender gap over the lifecycle for stayers into the contributions of women being less likely to (i) switch firms than men, and (ii) make advantageous moves conditional on switching. We observe that while female stayers are slightly less likely than their male counterparts to switch firms, much of the gender differential in firm-wage growth comes from women having lower firm-wage increases *conditional* on switching.

Under existing models of firm wage premia, one may think of the relationship of child-related constraints to this pattern in the following two ways. Firstly, women's moves to firms paying higher pay premia may be constrained on the supply side, as changes in women's preferences due to household care responsibilities increase the value of non-pay amenities (see Card et al. 2018, i.e. ew classical monopsony models). This would imply women may be more likely to make inter-firm moves which trade lower firm wage premia for higher non-pay amenities like flexibility. Secondly, women's moves to firms paying higher wage premia may be constrained on the demand side, due to hiring discrimination on the part of firms against women in their childbearing years (see Manning 2003, i.e. modern monopsony models). This could take the form of lower offers to childcare constrained women or even a stereotype penalty for women of this age. Unfortunately we do not observe child-related responsibilities in the data so cannot directly test these mechanisms. We can however reject that the gap is driven by differential moving on the supply side, as illustrated in figure 3.

Later convergence of gaps. If the presence of young children constrains women's ability to transition to higher-wage firms in earlier years, it is likely that these constraints ease as women enter their 40s. Instead of a stable gap thereafter, however, we observe a U-shape pattern, i.e., the gap begins to close as employed women in their 40s make relatively more advantageous firm moves than their male counterparts. What we see in Figure 2 is not merely stagnation, but rather convergence in female and male firm premia.

One explanation is that this could partly be a "mechanical" effect: the degree of advantageous moves is partly a function of the ranking of the worker's current firm. If men in their 40s are far more likely than women to be located at high-paying firms, it is possible that there are fewer opportunities for men to move up the ladder of firm wage premia, in comparison to women. In canonical Burdett-Mortenson models, for example, transition rates to higher wage firms are proportional to the proportion of firms above a worker's current firm in the wage distribution.

Indeed, we confirm empirically in our data that lower current firm premia are highly predictive of larger increases in firm premia next period. In order to assess the extent to which current premia would limit (or promote) advantageous moves differently by gender, we use this (homogenous) coefficient from this regression, and plot the predicted change in firm premia based on the firm premium of where one is currently employed. Taking the average separately by gender and age, Figure 4 shows that the predicted increase in firm premia declines over the life cycle (because workers move up the job ladder as they age). Observed changes in firm premia track predicted changes remarkably closely for men, throughout their life cycle. However for women between their mid-20s to their 40s, actual growth in firm premia is lower than predicted—it is only after their mid-40s, that the two trends coincide. This is suggestive of childcare-related constraints preventing women from realizing these "mechanical" moves up the firm job ladder up until their mid-40s, but as these constraints are relieved, the mechanical effect kicks in, and the gender gap closes.

6 Conclusion

We have shown that sorting into lower-wage firms explains nearly half of the gender wage gap in South Africa, a figure that is comparable to estimates from other developing countries. We contribute to the literature on gender sorting across firm wage premia by documenting a U-shape in the firm-wage gender gap over the life cycle, and assessing the contribution of worker transitions in and out of formal employment to the evolution of this sorting gap. The rise in the gap is driven primarily by gender differences in mobility among the continuously employed. We document a striking convergence in gender gaps in firm premia when workers enter their mid- to late-40s, and provide suggestive evidence that this is due to the easing of child-related constraints coupled with women's relatively lower position in the firm-wage hierarchy. Women also enter formal employment at worse-paying firms compared to men, and much of this gap is explained by firm characteristics.

From a policy perspective, our results imply that addressing gender differences in sorting across firms is critical for closing gender wage gaps. In a context with high levels of informality and unemployment, we show that the high dispersion of firm-wage premia matters substantively for gender wage gaps as well, more so than for wealthy country settings that have dominated existing research. Low levels of formality are associated with higher transition rates in and out of formal employment, and higher firm-wage gender gaps. However, the key mechanism driving the growth of the sorting gap over the lifecycle is the inability of employed women to climb up the firm-wage ladder – with the instructive exception of public administration. Focusing on women's inability to enter high-surplus, unionized firms in well-paying industries promises to

be a high-impact avenue for both research and policy aiming to close pay differentials between women and men.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). 'High wage workers and high wage firms'. *Econometrica*, 67(2): 251–333.

Arellano-Bover, J., Bianchi, N., Lattanzio, S., and Paradisi, M. (2024). 'One cohort at a time: A new perspective on the declining gender pay gap'. Available at SSRN. https://doi.org/10.2139/ssrn.4784734

Arrow, K. J. (1971). *Some models of racial discrimination in the labor market.* Santa Monica. Available at: https://www.rand.org/pubs/research_memoranda/RM6253.html (accessed in December 2024)

Barbanchon, T., Rathelot, R., and Roulet, A. (2021). 'Gender differences in job search: Trading off commute against wage'. *The Quarterly Journal of Economics*, 136(1): 381–426.

Barth, E., Kerr, S. P., and Olivetti, C. (2021). 'The dynamics of gender earnings differentials: Evidence from establishment data'. *European Economic Review*, 134(May): 103713. https://doi.org/10.1016/j.euroecorev.2021.103713

Bassier, I. (2022). 'Collective bargaining and spillovers in local labor markets'. CEP Discussion Paper 1895. London: Centre for Economic Performance.

Bassier, I. (2023). 'Firms and inequality when unemployment is high'. *Journal of Development Economics*, 161(March): 103029. https://doi.org/10.1016/j.jdeveco.2022.103029

Baum, C. F., Schaffer, M. E., and Stillman, S. (2002). *IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation.* Statistical Software Components, Boston College Department of Economics. Available at: https://ideas.repec.org/c/boc/bocode/s425401.html

Becker, G. S. (1957). *The economics of discrimination*. University of Chicago press.

Berniell, I., Berniell, L., De la Mata, D., Edo, M., and Marchionni, M. (2021). 'Gender gaps in labor informality: The motherhood effect'. *Journal of Development Economics*, 150(May): 102599. https://doi.org/10.1016/j.jdeveco.2020.102599

Bischof, D. (2017). 'New graphic schemes for Stata: Plotplain and Plottig'. *Stata Journal*, 17(3): 748–59. https://doi.org/10.1177/1536867X1701700313

Bravo, M. C. (2018). *GTOOLS: Stata module to provide a fast implementation of common group commands.* Statistical Software Components, Boston College Department of Economics. Available at: https://ideas.repec.org/c/boc/bocode/s458514.html

Bruns, B. (2019). 'Changes in workplace heterogeneity and how they widen the gender wage gap'. *American Economic Journal: Applied Economics*, 11(2): 74–113.

Budlender, J., and Ebrahim, A. (2020). 'Industry classification in the south african tax microdata'. WIDER Working Paper 2020/99. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/ 2020/856-6

Card, D., Cardoso, A. R., Heining, J., and Kline, P. (2018). 'Firms and labor market inequality: Evidence

and some theory'. *Journal of Labor Economics*, 36(S1): S13–S70.

Card, D., Cardoso, A. R., and Kline, P. (2016). 'Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women'. *The Quarterly Journal of Economics*, 131(2): 633–686.

Casarico, A., and Lattanzio, S. (2024). 'What firms do: Gender inequality in linked employer-employee data'. *Journal of Labor Economics*, 42(2): 325–355. https://doi.org/10.1086/723177

Corradini, V., Lagos, L., and Sharma, G. (2022). 'Collective bargaining for women: How unions can create female-friendly jobs'. IZA Discussion Paper. Bonn: Institute of Labor Economics (IZA).

Correia, S. (2014, July). *REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects.* Statistical Software Components, Boston College Department of Economics. Available at: https://ideas.repec.org/c/boc/bocode/s457874.html

Correia, S. (2018, September). *IVREGHDFE: Stata module for extended instrumental variable regressions with multiple levels of fixed effects.* Statistical Software Components, Boston College Department of Economics. Available at: https://ideas.repec.org/c/boc/bocode/s458530.html

Coudin, E., Maillard, S., and Tô, M. (2018). 'Family, firms and the gender wage gap in france'. IFS Working Paper W18/01. London: Institute of Fiscal Studies (IFS).

Cruz, G., and Rau, T. (2022). 'The effects of equal pay laws on firm pay premiums: Evidence from chile'. *Labour Economics*, 75(April): 102135. https://doi.org/10.1016/j.labeco.2022.102135

Folbre, N., Gautham, L., and Smith, K. (2023). 'Gender inequality, bargaining, and pay in care services in the united states'. *ILR Review*, 76(1): 86–111.

Goldin, C., Kerr, S. P., Olivetti, C., and Barth, E. (2017). 'The expanding gender earnings gap: Evidence from the lehd-2000 census'. *American Economic Review: Papers and Proceedings*, 107(5): 110–114.

Haltiwanger, J., Hyatt, H., and McEntarfer, E. (2018). 'Who moves up the job ladder?' *Journal of Labor Economics*, 36(S1): S301–S336.

He, H., Li, S. X., and Han, Y. (2023). 'Labor market discrimination against family responsibilities: A correspondence study with policy change in china'. *Journal of Labor Economics*, 41(2): 361–387.

Kerr, A., Lam, D., and Wittenberg, M. (2023). 'Post-apartheid labour market series 1993-2019 [dataset]'. Version 3.3.1. Cape Town: DataFirst [producer and distributor], 2019. https://doi.org/10.25828/gtr1-8r20

Kerr, A., and Wittenberg, M. (2019). 'Earnings and employment microdata in south africa'. WIDER Working Paper 2010/47. Helsinki: UNU-WIDER.

Kline, P., Saggio, R., and Sølvsten, M. (2020). 'Leave-out estimation of variance components'. *Econometrica*, 88(5): 1859–1898.

Labanca, C., and Pozzoli, D. (2022). 'Constraints on hours within the firm'. *Journal of Labor Economics*, 40(2): 473–503.

Li, J., Dostie, B., and Simard-Duplain, G. (2023). 'Firm pay policies and the gender earnings gap: The mediating role of marital and family status'. *ILR Review*, 76(1): 160–188.

Manning, A. (2003). *Monopsony in motion: Imperfect competition in labor markets*. Princeton: Prince-

ton University Press.

Masso, J., Meriküll, J., and Vahter, P. (2022). 'The role of firms in the gender wage gap'. *Journal of Comparative Economics*, 50(2): 454–473.

Morchio, I., and Moser, C. (2021). 'The gender pay gap: Micro sources and macro consequences'. CEPR Discussion Paper DP16383. Paris & London: CEPR Press.

National Treasury and UNU-WIDER (2023a). *CIT-IRP5 firm-level panel 2008–2020 [dataset], version 5.0.* Pretoria: South African Revenue Services [producer of the original data], 2020. Pretoria: National Treasury and UNU-WIDER [producer and distributor of the harmonized dataset], 2022.

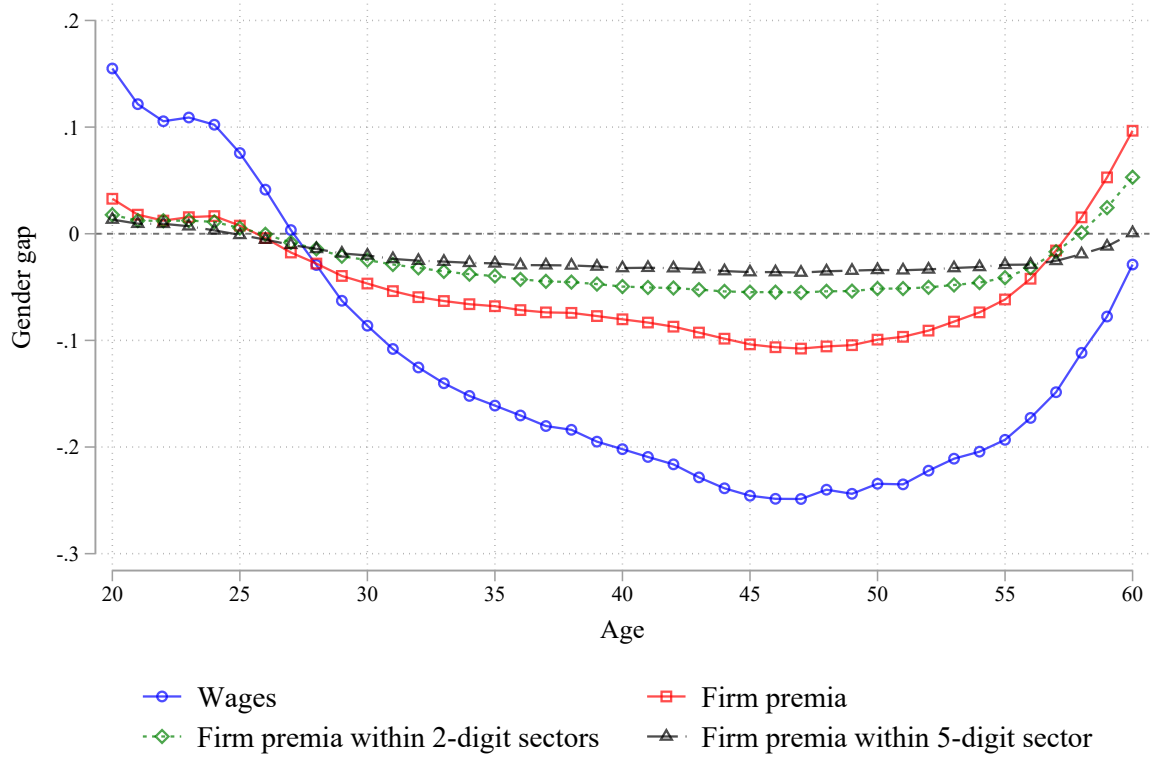
National Treasury and UNU-WIDER (2023b). *IRP5 worker-level data 2009–2021 [dataset], version 5.* Pretoria: South African Revenue Services [producer of the original data], 2021. Pretoria: National Treasury and UNU-WIDER [producer and distributor of the harmonized dataset], 2022.

Petit, P. (2007). 'The effects of age and family constraints on gender hiring discrimination: A field experiment in the french financial sector'. *Labour Economics*, 14(3): 371–391.

Petrongolo, B., and Ronchi, M. (2020). 'Gender gaps and the structure of local labor markets'. *Labour Economics*, 64(June): 101819. https://doi.org/10.1016/j.labeco.2020.101819

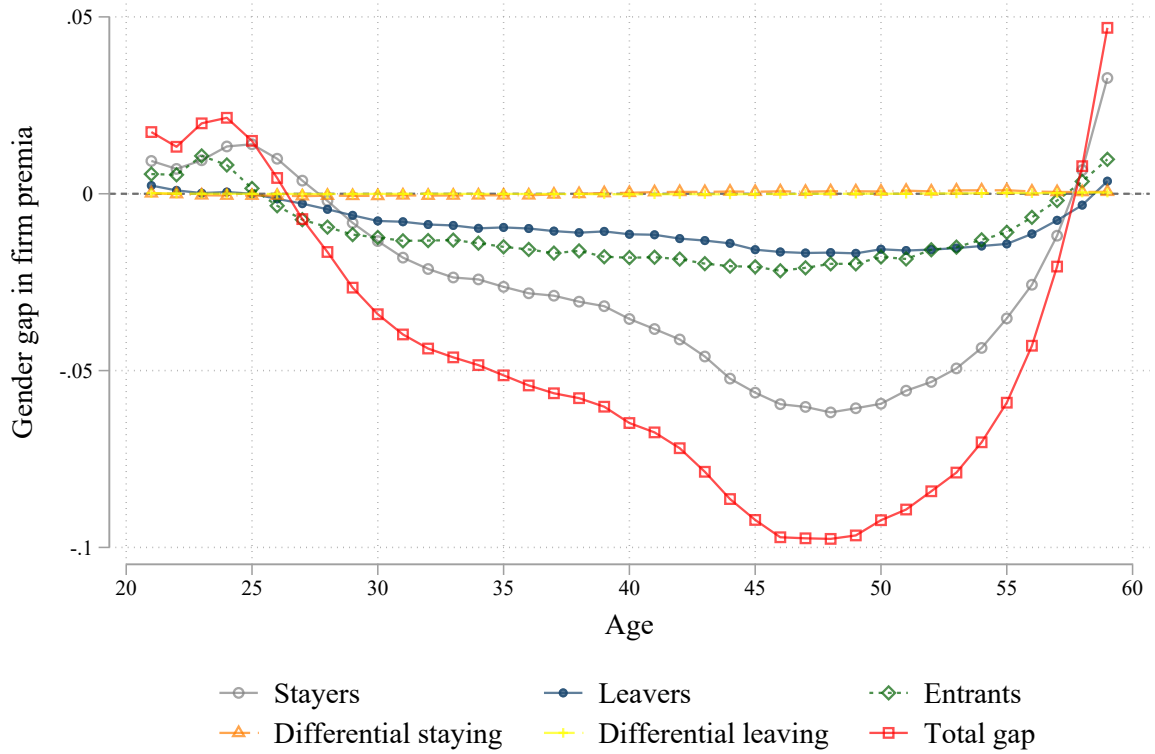
Figures and tables

Figure 1: Gender gap in firm wage over the life cycle



Note: firm premium within sectors is estimated by residualizing on the relevant sector fixed effects.
Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

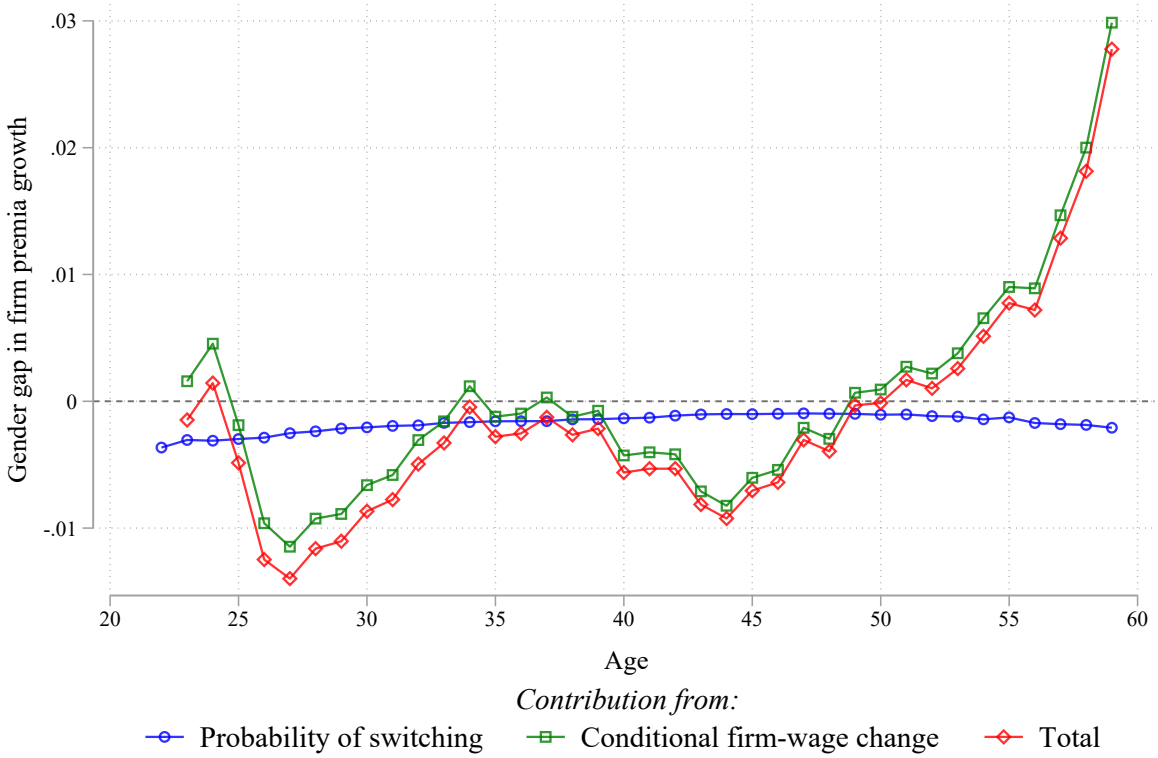
Figure 2: Decomposing the gender gap in firm wage premia over the life cycle



Note: the figure plots the contributions to the gender gap in firm wage premia by category, corresponding to equation 2. Stayers are defined as workers who are continuously employed, with changes in firm wage premia arising from switches across firms. Leavers are workers who are not observed in formal employment the following year, and entrants are workers who are not observed in formal employment the previous year. Differential staying and leaving are probabilities of staying and leaving the workforce. The total gap is the sum of these components, and corresponds to the total cross-sectional firm-wage gender gap.

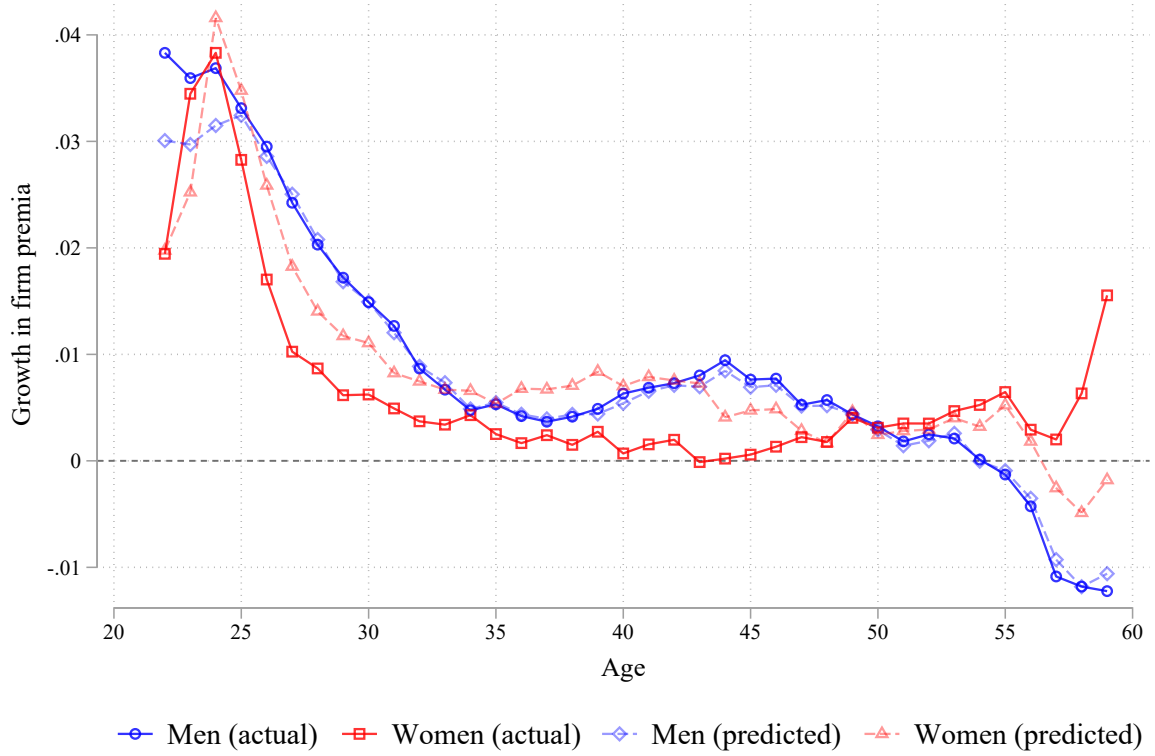
Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

Figure 3: Gender gap in firm premia growth for the continuously employed



Note: the change in the gender gap in firm premia over the lifecycle for stayers ("total") is decomposed into the contribution of female stayers being less likely to switch firms than men ("probability of switching") and to being less likely to make advantageous moves conditional on switching ("conditional firm-wage change"). To construct the total, differences in probability of switching are weighted by female firm-premia growth conditional on switching, while differences in firm premia growth conditional on switching are weighted by male probabilities of switching.
Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

Figure 4: Growth in firm premia for the continuously employed, actual and predicted



Note: the figure shows the average growth in firm wage premia for the continuously employed ("stayers") by gender and age. Actual corresponds to the observed growth, and predicted corresponds to the predicted growth in firm wage premia from a regression of the growth in firm premia and the current level of firm premia.
Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

Table 1: Predictors of the gap in firm premia (entrants)

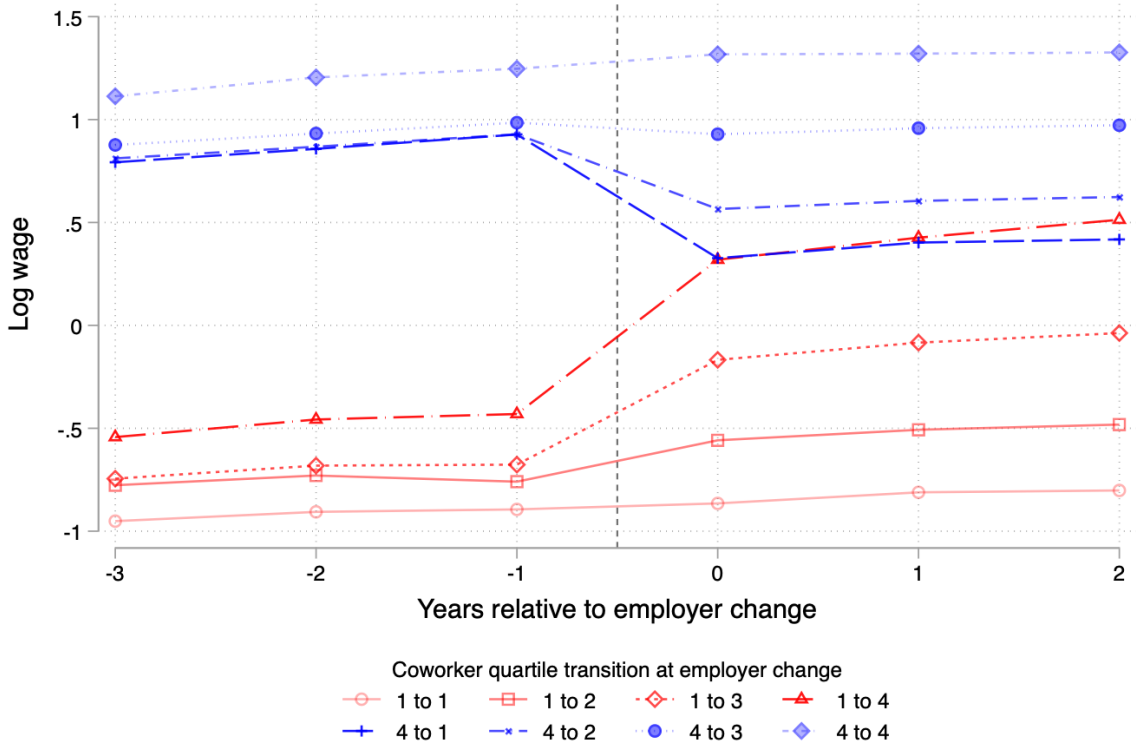
	(1)		(2)	
	Contribution	Standard errors	Contribution	Standard errors
Firm premiums				
Men	-0.530***	(0.001)	-0.670***	(0.000)
Women	-0.664***	(0.001)	-0.810***	(0.000)
Total gap	0.134***	(0.001)	0.139***	(0.000)
Explained	0.088***	(0.001)	0.079***	(0.000)
Unexplained	0.045***	(0.000)	0.060***	(0.000)
Explained				
Worker age	0.002***	(0.000)	0.001***	(0.000)
Commuting distance	0.000***	(0.000)		
Firm value added	0.009***	(0.001)	0.008***	(0.000)
Bargaining council agreement	0.018***	(0.000)	0.005***	(0.000)
Industry	0.059***	(0.001)	0.066***	(0.000)
N	500581		7990484	

Note: the decomposition uses coefficients from the pooled regression of both women and men to weight the contribution of the explained components. As commuting distance is available only for a subset of the sample, we present estimates with and without commuting distance as a covariate (corresponding to specifications (1) and (2), respectively). * p<0.10, ** p<0.05, *** p<0.01.

Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

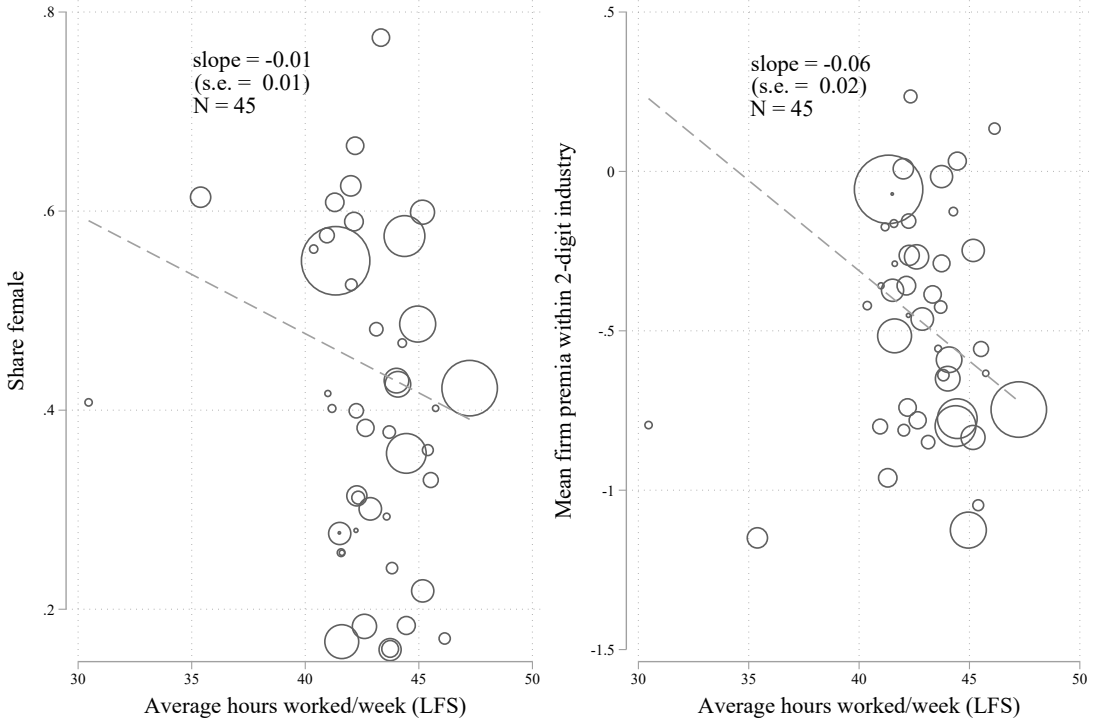
A Appendix: additional figures and tables

Figure A1: Mean log wages of job changers, by quartile of mean co-worker wage at origin and destination firm



Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

Figure A2: Average hours worked per week, by 2-digit industry

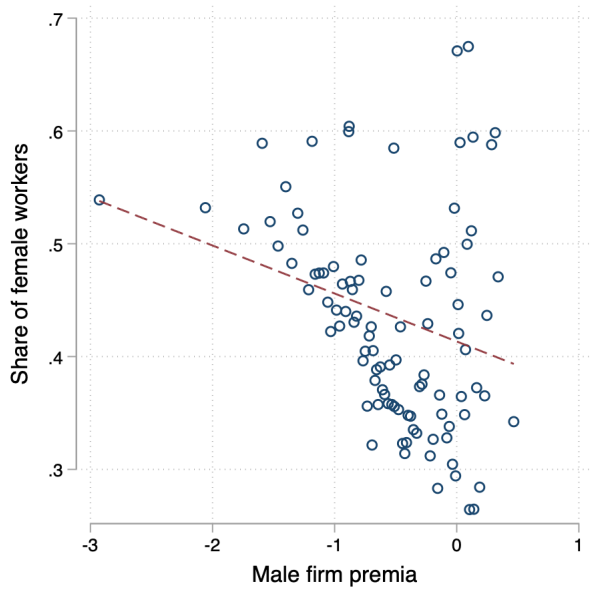


Note: Marker size given by the share of each 2-digit industry in total employment.

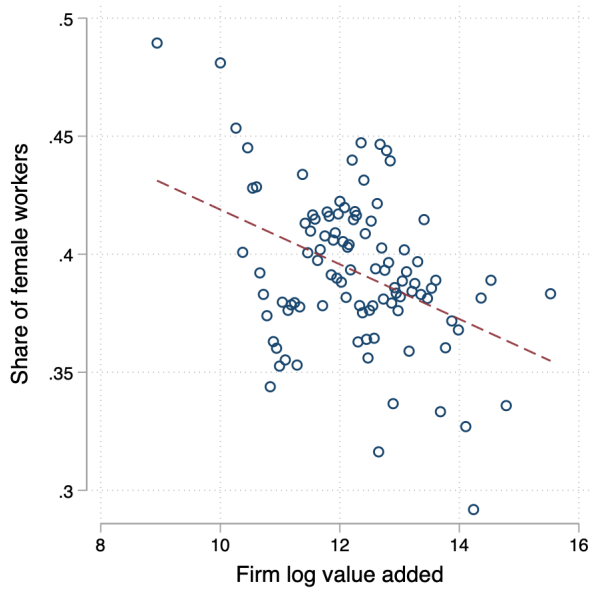
Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b). Data on weekly hours worked from LFS, 2010–2018, from Kerr et al. (2023), with LFS sample restricted to formal sector workers between the ages of 20 and 60 in firms with over 20 workers.

Figure A3: Share of female workers by firm value added and firm premia

(a) **Firm wage premia**



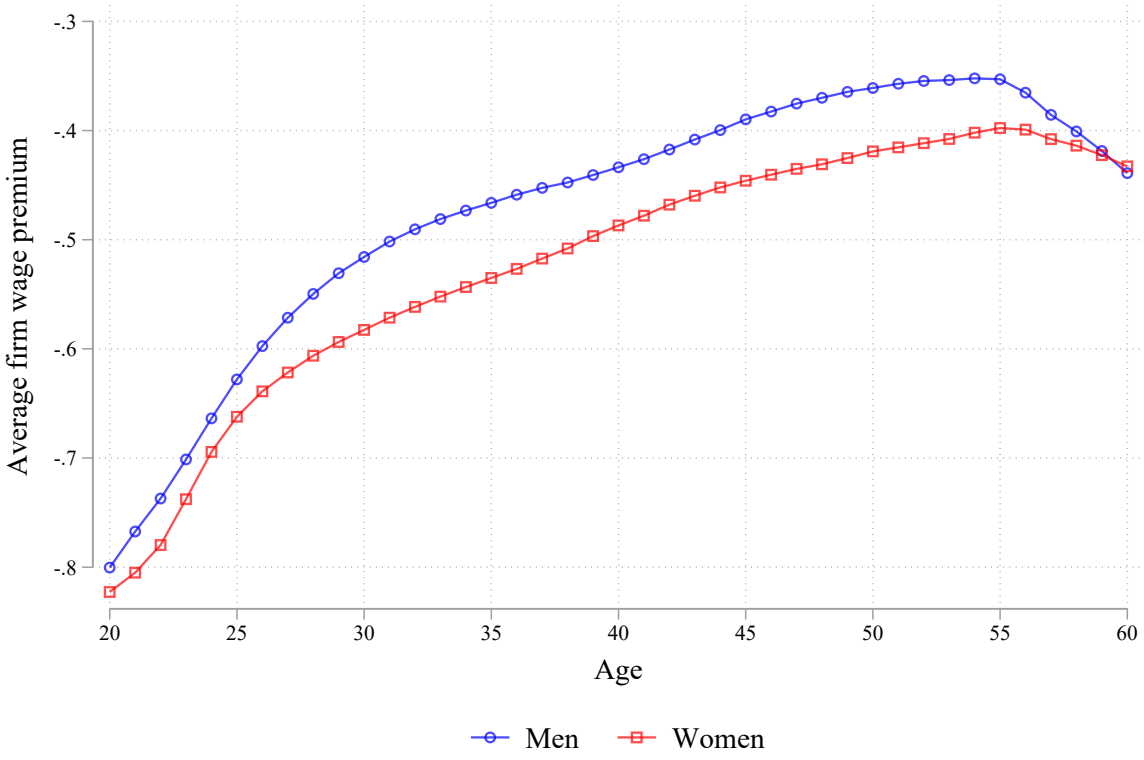
(b) **Value added per worker**



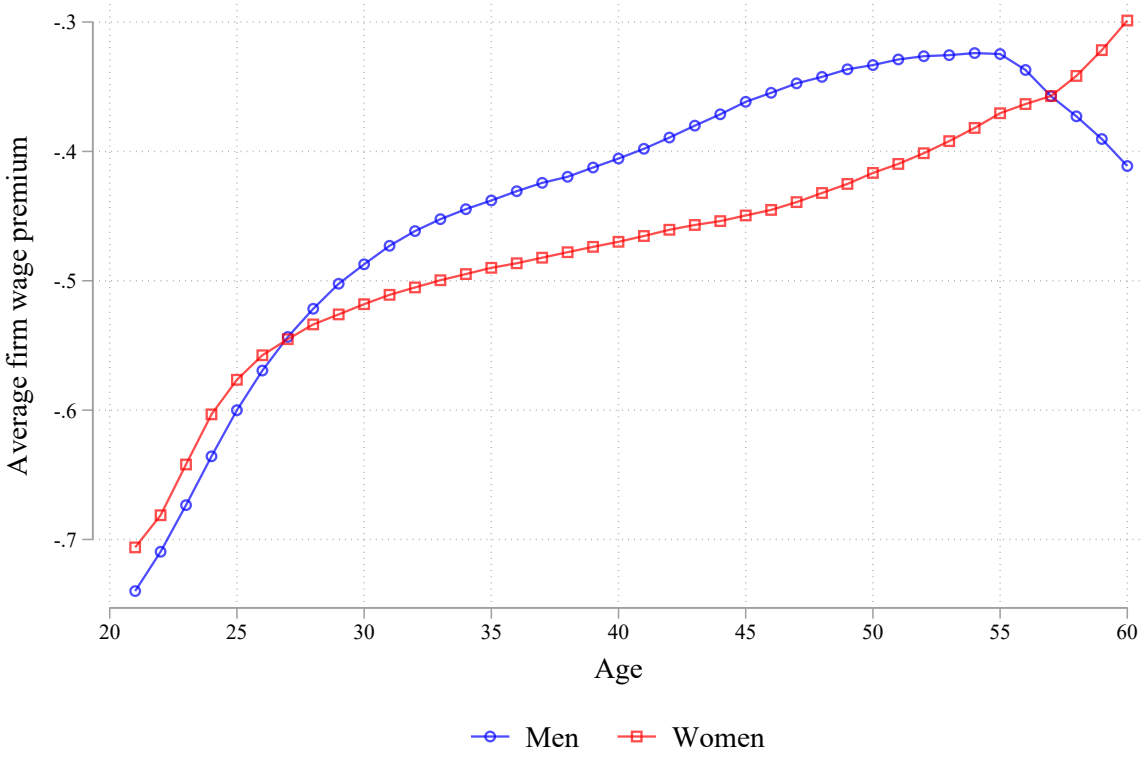
Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

Figure A4: Gender gaps in firm wage premia across age

(a) **Without cohort controls**

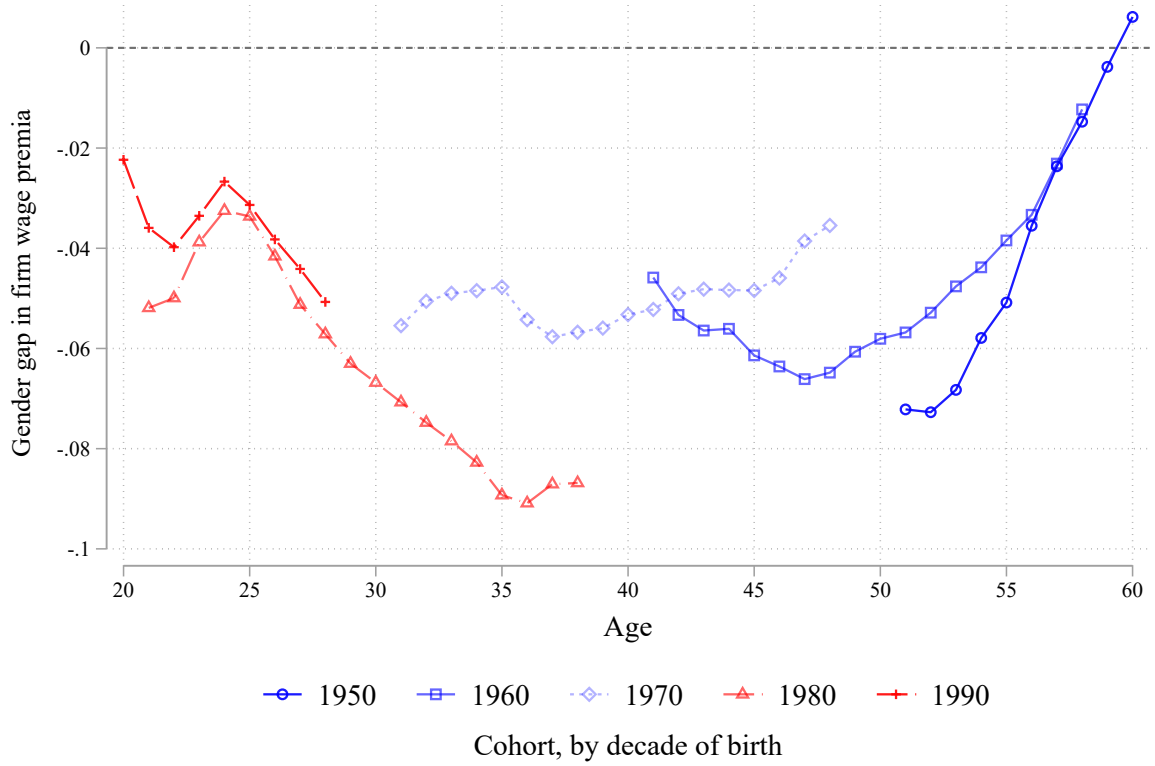


(b) **With cohort controls**



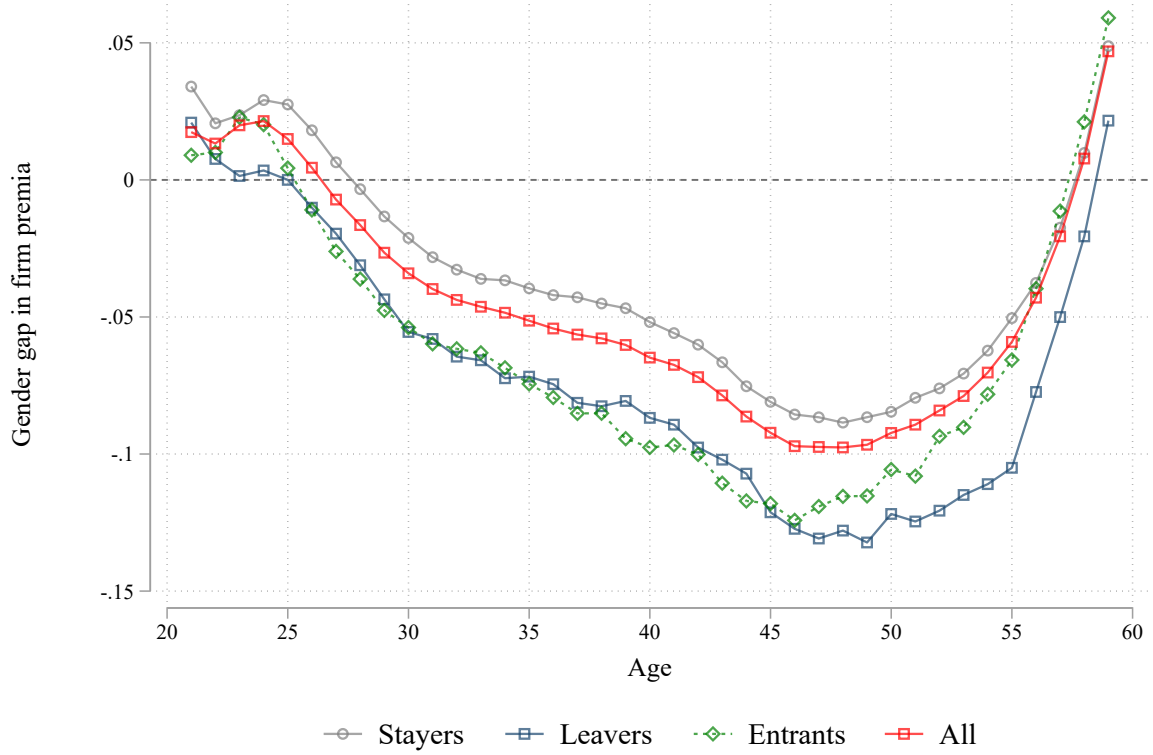
Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

Figure A5: Gender gaps in firm wage premia, by 10-year birth cohorts



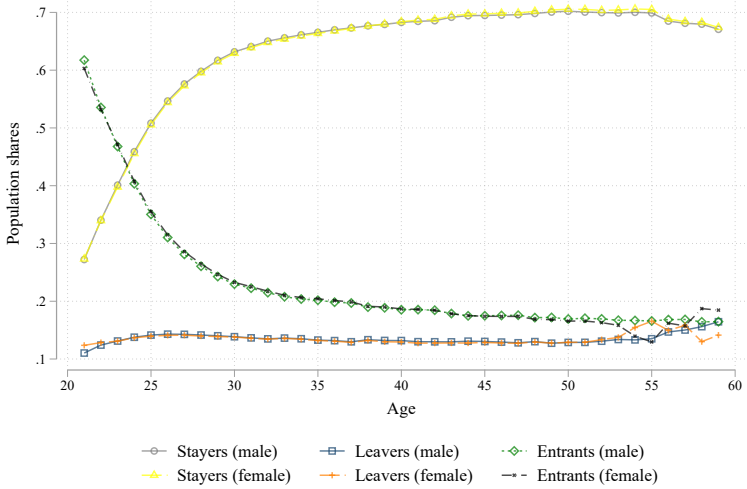
Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

Figure A6: Gender gaps in firm wage premia over the life cycle, by category of worker



Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

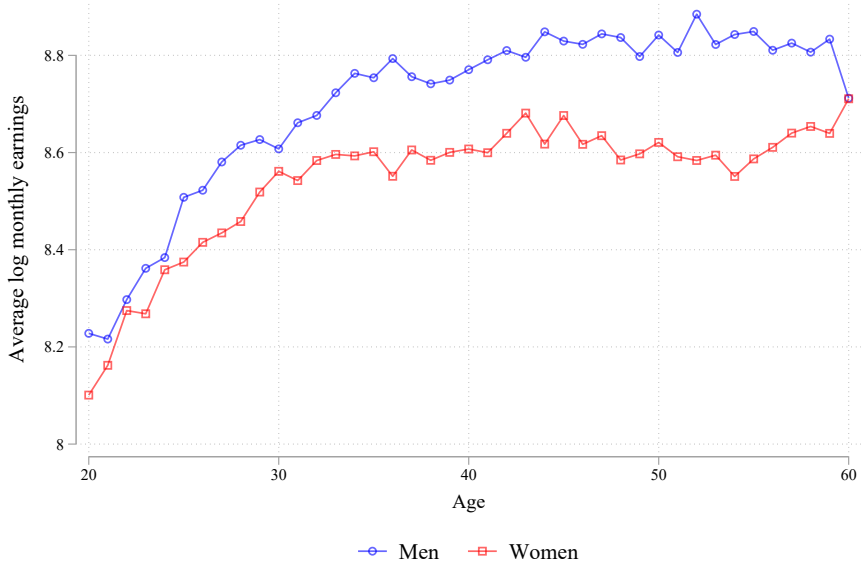
Figure A7: Population proportions of stayers, leavers, and entrants, by gender



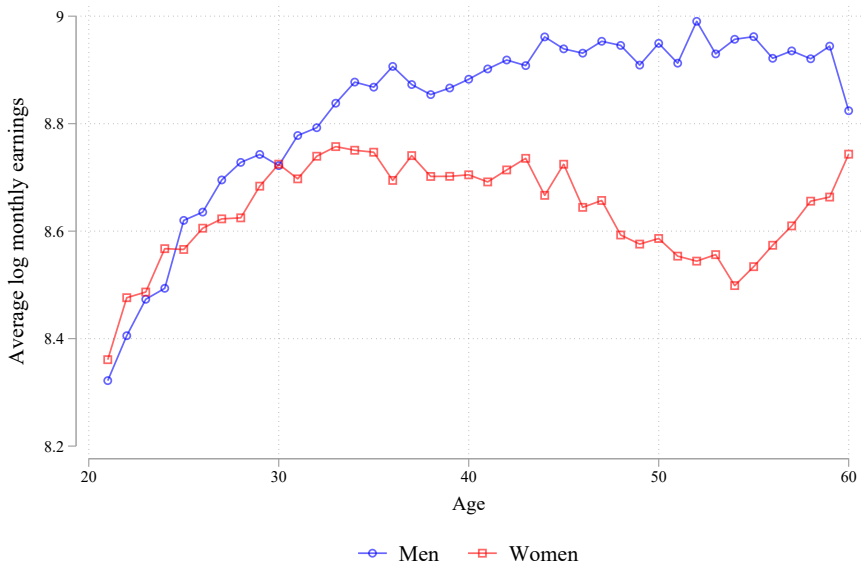
Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

Figure A8: Average log monthly earnings (LFS)

(a) **Without cohort controls**



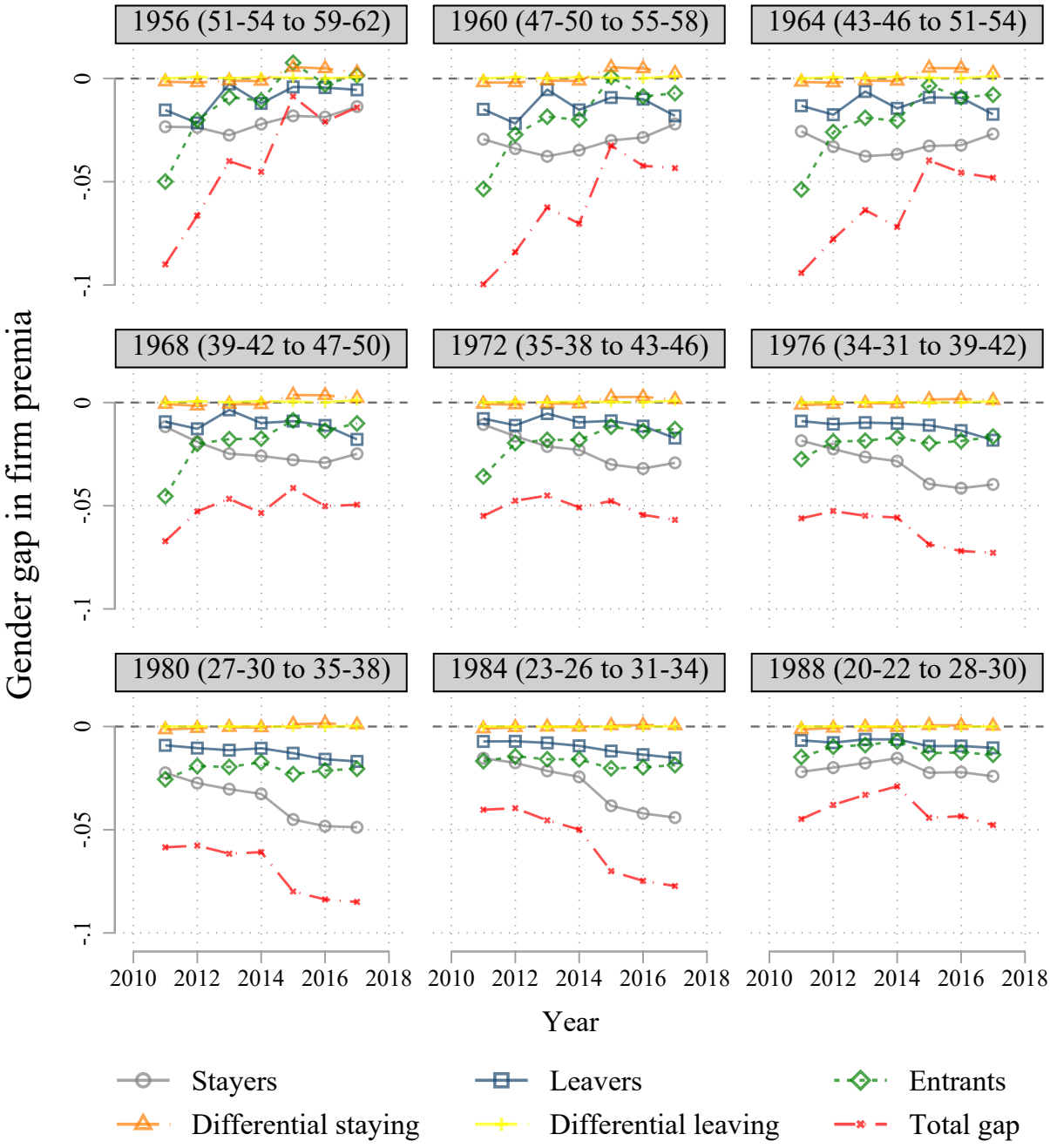
(b) **With cohort controls**



Note: restricted to formal sector workers between the ages of 20 and 60 in firms with over 20 workers. Both panels control for survey year. Cohort controls refer to the interaction of female with dummies for birth year.

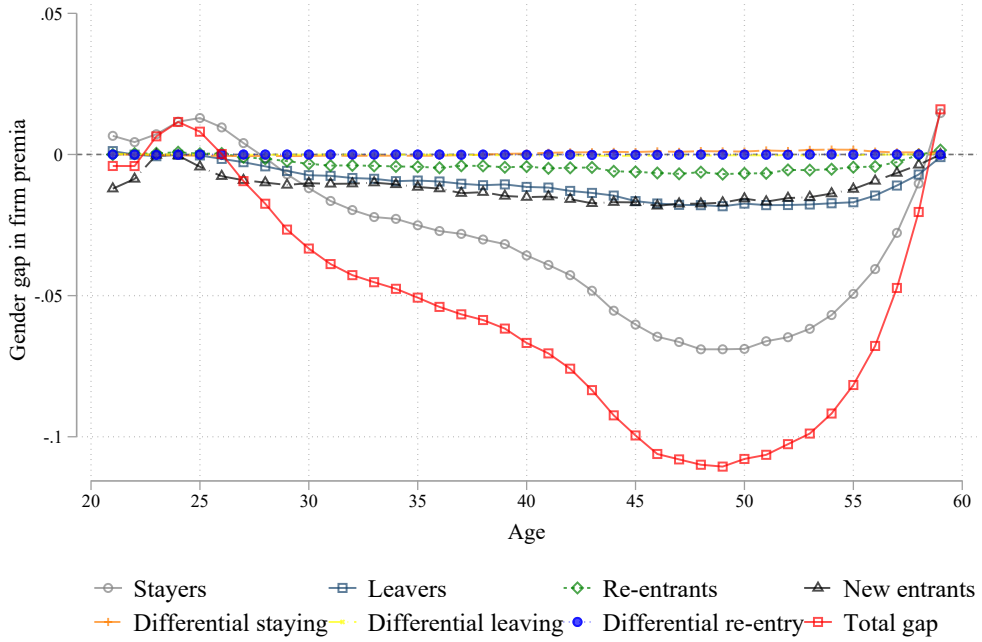
Source: authors' illustration based on LFS, 2010–2018, from Kerr et al. (2023).

Figure A9: Gender gaps in firm wage premia over age, by 4-year birth cohorts



Note: age range corresponding to calendar year range shown in parentheses for each cohort.
Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

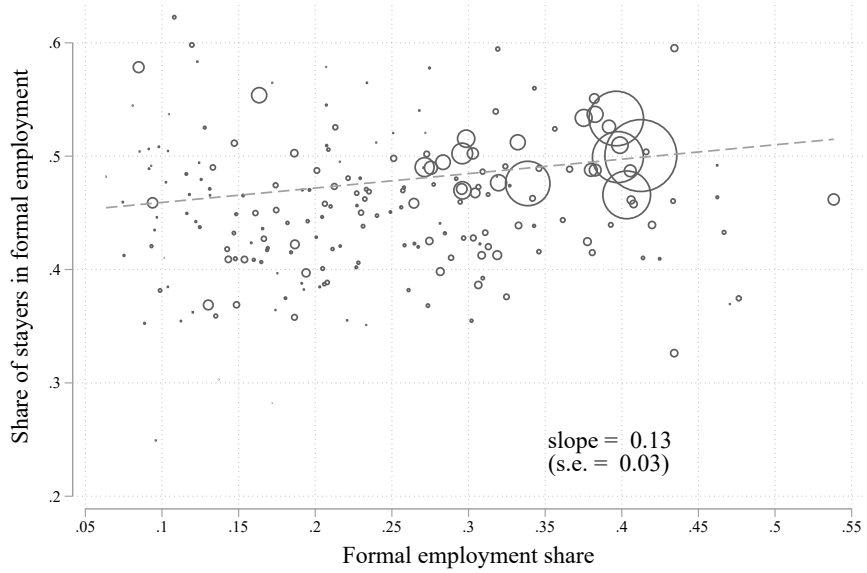
Figure A10: Gender gaps in firm wage premia over the life cycle (new entrants vs. re-entrants)



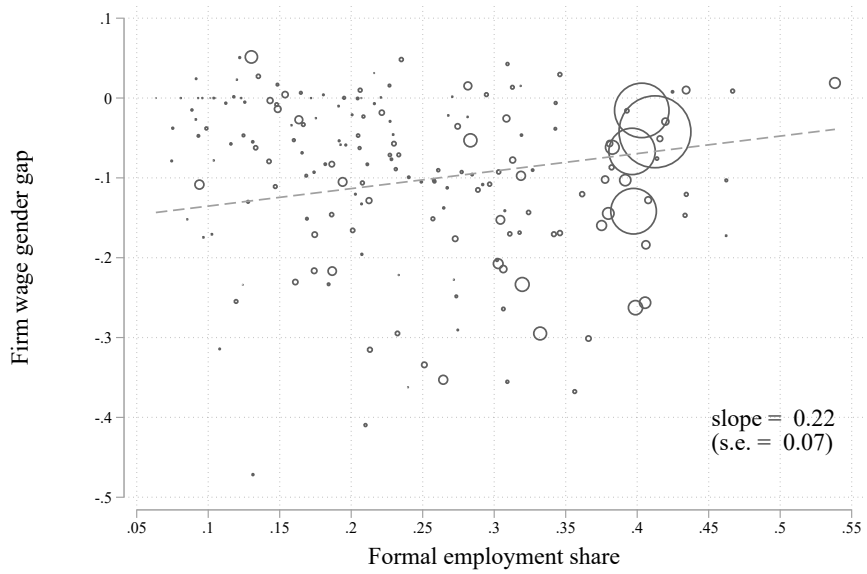
Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

Figure A11: Formality and regional firm-wage dynamics

(a) **Share of stayers among total formal employment**



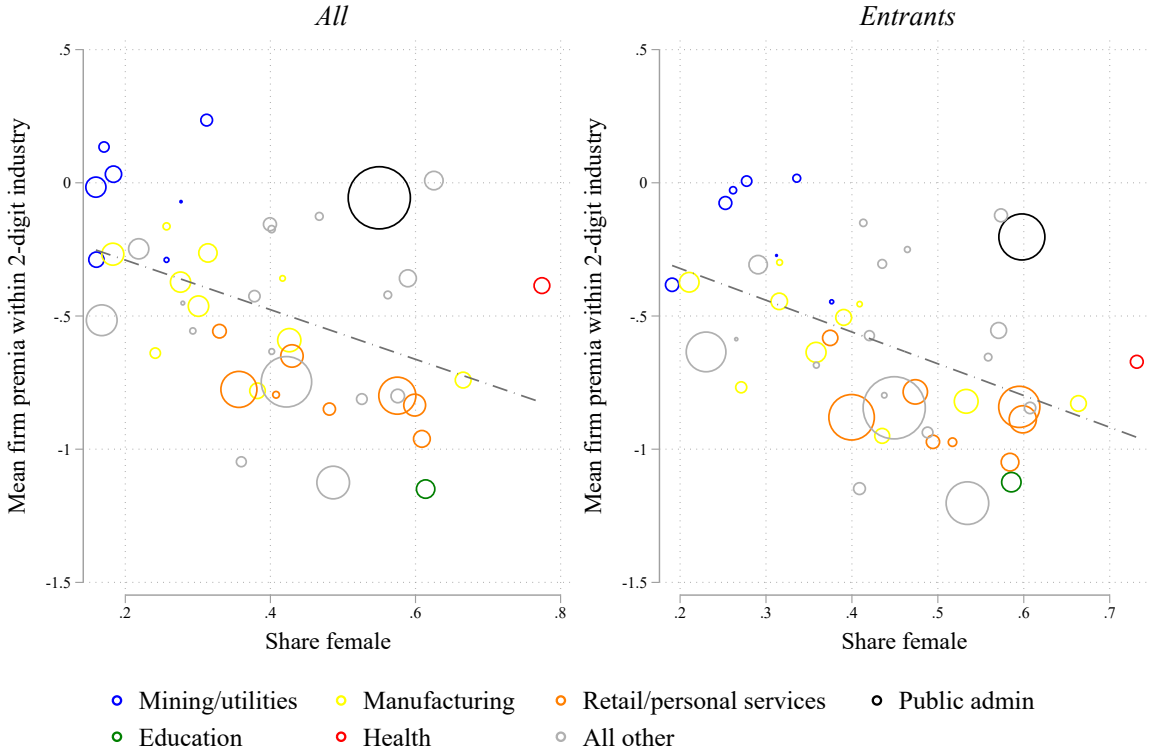
(b) **Firm-wage gender gap**



Note: the formal employment share is measured as the municipal formal employment to population ratio from the 2011 Census. Municipalities (*N=208*) weighted by population size. We drop municipalities with implausibly low firm-wage gender gaps (i.e., female firm-wages higher by 6 log points, *N=18*).

Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

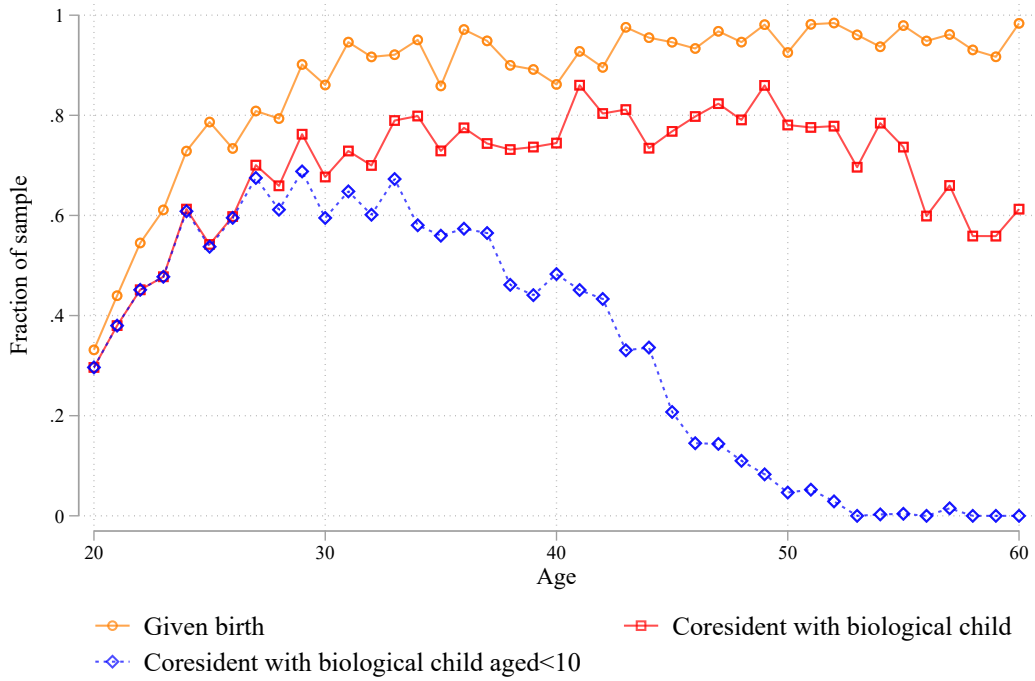
Figure A12: Average firm wage premia and female share, by 2-digit industries



Note: marker size given by the share of each 2-digit industry in total employment.

Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

Figure A13: Incidence of parenthood and co-residence with children among women (NIDS)



Source: authors' illustration based on National Income Dynamics Study 2017, Wave 5.

Table A1: Summary statistics

	Full sample		*Analysis sample*	
	Women	Men	Women	Men
Annual real wages				
Mean	133594	158249	135139	146734
25th percentile	25116	30201	26425	31510
Median	66505	76124	73958	82365
75th percentile	188745	189153	203661	196831
Average firm size	15861	11441	18628	13870
Fraction of year employed	0.81	0.80	0.81	0.79
Person-year observations	34577777	44246505	27312497	35265020

Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

Table A2: Covariate means (entrants)

	Women	Men
Commuting distance (in kilometers)	4.658	4.679
Worker age	33.612	34.020
Firm log value added	11.930	11.972
Bargaining council agreement	0.332	0.375

Note: commuting distance is available only for a subset of the sample (N=500,581).
Source: authors' illustration using South African tax records, 2010–2018 (National Treasury and UNU-WIDER 2023a, 2023b).

B Data appendix

This data appendix is created as per UNU-WIDER requirements for users of the National Treasury Secure Data Facility (NT-SDF). It reports on data directly used for the results presented in this paper and does not include other variables and programmes used in our ongoing research on this topic.

Data access

The data used for this research was accessed from the NT-SDF. Access was provided under a non-disclosure agreement, and our output was checked so that the anonymity of no firm or individual would be compromised. Our results do not represent any official statistics (NT or SARS). Similarly, the views expressed in our research are not necessarily the views of the NT or SARS.

Data used: CIT-IPR5 panel (`citirp5_v5_0`) (National Treasury and UNU-WIDER 2023a) and year-by-year IRP5 job-level data (`v5`) (National Treasury and UNU-WIDER 2023b). Date of first access for this project: 20 October 2020. Last accessed: 19 August 2024.

Software

Our analysis was conducted using Stata 17. User-written programmes and schemes used include `reghdfe` (Correia 2014), `gtools` (Bravo 2018), `ivreg2` (Baum et al. 2002), `ivreghdfe` (Correia 2018) and `plotplain` (Bischof 2017).

Variables

Variables used from the raw IRP5 data include: `taxyear taxrefno payereferenceno dateofbirth gender idno passportno province_geo busprov_geo districtmunicip_geo busdistmuni_geo periodemployedfrom periodemployedto totalperiodsinyearofassessment totalperiodsworked`.

Employment income was created from the following IRP5 amount codes: `amt3601 amt3605 amt3606 amt3607 amt3615 amt3616`. A record of employment-related allowances was created from the following IRP5 amount codes: `amt3701 amt3704 amt3710 amt3711 amt3712 amt3713 amt3715`.

IRP5 employment records were identified by records which had non-zero income or allowances; those with zero or missing income and allowances data are dropped from the analysis.

Variables used from the CIT-IRP5 data include: `taxyear finyear FYE taxrefno g_sales g_cos g_grossprofit g_grossloss k_ppe k_faother comp_prof_sic5_1d comp_prof_sic5_2d`.

Value added was calculated by subtracting cost of sales from gross sales.

The 'composite profit code' industry variables we use were created by Budlender and Ebrahim (2020). We merge in Bargaining Council variables created by Bassier (2022).

Cleaning and sample notes

We restrict our main sample to those aged 20–60 years, and who work for firms with at least 10 workers of each gender classification. We classify a separation of a worker from a firm when they are not observed at the same firm in the following year. Stayers are defined as workers who are continuously employed, with changes in firm wage premia arising from switches across firms. Leavers are workers who are not observed in formal employment the following year, and entrants are workers who are not observed in formal employment the previous year. These notes represent some particularly noteworthy data cleaning and sample construction decisions, but we cannot outline all such decisions here without reproducing our many thousands of lines of code; users are referred to our do-files which are available at the NT-SDF.