Article:

Bassier, I. and Gautham, L. orcid.org/0000-0001-9098-7272 (2025) The firm-pay gender gap and formal sector churn over the life cycle. Journal of Development Economics, 176. 103498. ISSN 0304-3878

https://doi.org/10.1016/j.jdeveco.2025.103498

The firm-pay gender gap and formal sector churn over the life cycle[*]

Ihsaan Bassier[†] and Leila Gautham[‡]

March 11, 2025

Abstract

We find that women sorting into lower paying firms explains nearly half of the gender pay gap in South Africa. Using matched employer-employee panel data covering the universe of formal workers, we show sorting varies considerably over the life cycle: the firm-pay 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-pay amenities. Churn also contributes to the gap (though is relatively constant), since women enter formal employment at worse-paying firms than men. The relative importance of 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.

JEL Codes: J31, J16, J42, J71.
Keywords: firm pay premia, gender pay inequality, sorting, worker transitions.

---

[*]Ihsaan Bassier and Leila Gautham contributed equally to this publication. We thank 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. We are especially grateful to Noreen Kajugusi for excellent research assistance. Access to the data was granted through the UNU-WIDER SA-TIED programme at the Secure Data Facility in South Africa. This work was supported by a grant from the UNU-WIDER SA-TIED programme, as well as funds for research assistance from the University of Leeds.

[†]Economics, University of Surrey; and Southern Africa Labour and Development Research Unit, University of Cape Town. Email: i.bassier@surrey.ac.uk

[‡]Department of Economics, University of Leeds. Email: l.gautham@leeds.ac.uk

1 Introduction

Firm sorting matters a great deal for gender pay inequality. We begin by showing that the sorting of women into lower paying firms accounts for 45 percent, or nearly half, of the mean gender pay gap for the universe of formal sector South African workers. The dominant role of sorting for gender pay gaps is consistent with findings from other settings (Card, Cardoso, and Kline, 2016; J. Li, Dostie, and Simard-Duplain, 2023). However, the mechanisms underlying sorting are less well understood. In this paper we focus 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-pay gender gap (the average difference in firm pay 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. The sorting of women into low-paying firms may be the result of either (or both) employer discrimination or worker choice, and we discuss evidence on the relative roles of these processes.

We contribute to the small but growing literature on the evolution of firm pay dynamics over the life cycle (Bruns, 2019; Casarico and Lattanzio, 2022), where firm pay premia are defined following Abowd, Kramarz, and Margolis (1999). Our first contribution is that we find a striking U-shape in the evolution of the firm-pay 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-pay 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, S. P. Kerr, and Olivetti, 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-pay gender gap grow over the first part of the life cycle? We observe that changes in the firm-pay 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 transition type. We find that firm-pay 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-pay 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 by transition type, 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 out-of-sector worker transitions are more important (Donovan, Lu, and Schoellman, 2023), 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 explains two-thirds of the firm-pay gender gap among labour market entrants (commuting distance explains close to nothing). Secondly, much of the steep initial increase in the firm-pay gender gap over the life cycle is driven by women moving to lower-paying firms relative to men, *conditional* on having moved. This rules out differential moving frictions as the driver, and we instead discuss possible demand- and supply-side reasons. Thirdly, a mechanical effect helps explain the closing of the gender gap in firm pay for older women: groups sorted lower in the firm pay 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 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 higher firm-pay gender gaps, helping explain why firms play such an outsized role in shaping gender pay inequality.

Important limitations of our analysis include the absence of informal sector firms or workers in our administrative data; instead, we have supplemented our analysis with a comparison of informal sector gender wage gaps from the South African Labour Force Survey. We are also not able to conclusively arbitrate between supply- vs. demand-side explanations for the evolution of gender gap in sorting across firms. Finally, we focus on workers in larger firms given the difficulty in credibly estimating firm premia for very small firms.

2 The overall firm-pay gender gap

2.1 Estimation of firm pay 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 (see our Data Appendix for a detailed description of the data, sample, and cleaning procedures). Our dataset consists of 8–9 million workers aged between 20 and 60 in each year, resulting in a total sample of 75.3 million worker-year observations with about 32.8 million belonging to women and 42.5 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 earnings for women are 15 percent lower than for male workers. This is larger than, but similar to, the 13.2 percent 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 (A. Kerr and Wittenberg, 2019).[1] A caveat to our analysis is that our tax data exclude the informal sector. The formal sector constitutes 60 percent of the employed, and 45 percent of the labour force for individuals aged 20–60; this share is 46 percent for men, and 43 percent for women (LFS, 2010–2018). Informal wage employment, self-employment, and the unemployed are 19 percent, 11 percent, and 26 percent of the labour force.[2] The gender pay gap within the informal sector follows a different life cycle pattern and is larger than the gap in the formal sector, with women's wages lower than men's wages by 30 percent, on average (Figure B1). Consequently, the gender earnings gap across all workers is likely larger than in our formal sector sample.

To compute the contribution of firms to gender pay inequality in South Africa we implement an Abowd, Kramarz, and Margolis (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 logged annual earnings for worker i in firm j at time t; θ_i are individual fixed effects and ψ_j are firm fixed effects. 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, Saggio, and Sølvsten (2020), we implement this through a leave-out estimator which adjusts for mobility bias.[3] Due to the computational intensity of this procedure over the full sample,

[1] This gap is computed using LFS data (2010–2018) for workers between 20 and 60 (A. Kerr, Lam, and Wittenberg, 2019).

[2] Labour force participation rates are 76 percent for men, and 60 percent for women, and have seen modest increases for both men and women over the sample period.

[3] Firm fixed effects are identified based on workers who move between firms: therefore, firms with few movers will have more measurement error, which the Kline, Saggio, and Sølvsten (2020) estimator corrects for. Our results are less sensitive to this problem, however, since we consider the difference in average firm

we compute firm fixed effects over a smaller subset of years (2011-2016).[4] 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).[5] Focusing on workers in larger firms results in a slightly smaller gender pay gap (11.5 percent) in our final analysis sample compared to the gap in the full tax sample. Again, this is comparable to the gender gap (12.5 percent) in the LFS sample closest to our own (formal sector workers in firms with over 20 workers).

We use the term "firm premia" or "firm pay" interchangeably with the AKM firm fixed effects estimated in equation 1. In this paper we focus on estimates using homogeneous firm premia rather than gender-specific firm premia, but also briefly discuss our results using gender-specific firm premia.

The validity of the estimated firm premia hinge on the assumption that workers' moves across firms are not driven by unobserved factors that also affect wages (e.g., unobserved match effects). To provide evidence on this, we reproduce standard checks in the literature (Card, Cardoso, and Kline, 2016). We plot mean log pay of job changers, by quartile of mean co-worker pay at origin and destination firm (Figure C3). Employees who move from jobs with low-paid co-workers experience large average pay gains when they move to jobs with highly-paid co-workers, with symmetric pay losses when workers move in the opposite direction, and flat average trends in pay 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

premia across gender (not its variance). Our findings are robust to using a simple AKM estimator (i.e., without correcting for mobility bias) (Figure C1), as well as for alternative estimators to correct for mobility bias such as Bonhomme et al. (2023) (Figure C2).

[4]See Data Appendix for details. Our main results are unchanged when we use firms from the full set of years 2010–2018, but without having applied the KSS leave-out correction (see Figure C1).

[5]See Table C1 for summary statistics on the full sample and the analysis sample.

theory, gender differences in firm premia could be driven by a divergence in hours worked. However, in the LFS data, the gender gap in annual earnings and hourly wages evolve roughly in parallel (Figure B3). Moreover, average hours worked (within 2-digit industries) are negatively correlated with both mean firm premia and the share of female employment in that industry (Figure C4 (a)).[6] There is also no statistically significant relationship between mean gender gaps in firm premia and hours worked at the industry level (Figure C4 (b)). This evidence suggests gender differences in firm premia are not primarily driven by lower hours worked in female-dominated firms.

2.2 The gender gap in firm pay premia

Our first and baseline result is that the firm-pay gender gap is 0.055 log points or 45 percent of the 0.122 log point raw gender pay gap. We define the firm-pay gender gap as the average difference in firm pay premia between men and women across firms, i.e., the difference between $E[\psi_j|g = F]$ and $E[\psi_j|g = M]$, with ψ_j defined in equation 1, and M and F denoted women and men, respectively. As we impose a single pay premium per firm (see equation 1), this difference arises purely from female employment being concentrated more in lower premium firms than men. As a non-parametric proxy, there is a clear negative correlation of the share of female workers in a firm with both firm pay premia and the firm's valued added (Figure C5). For instance, firms in the bottom decile by value added have, on average, 43 percent of their workers being women, compared to only 35 percent in the top decile. Similarly, ranked by firm premia, firms in the bottom decile have a female share of 53 percent on average, compared to 41 percent in the top decile.

When we estimate the contribution of sorting using gender-specific firm premia: i.e. using male and female firm premia to weight gender differences in sorting, it is 44 and 52 percent respectively. Comparable estimates for the contribution of sorting to the gender pay gap range from 15 and 20 percent for Portugal (Card, Cardoso, and Kline, 2016); 21

[6] This is consistent with literature showing that average hours worked at a firm bears little relation to firm pay (Card, Cardoso, and Kline, 2016; Labanca and Pozzoli, 2022).

and 26 percent for Italy (Casarico and Lattanzio, 2022); and 11 percent for France (Coudin, Maillard, and Tô, 2018). Our finding that sorting contributes 45 percent—or nearly half—of the gender pay gap is therefore strikingly high in absolute magnitude, but also in comparison to findings from high-income country settings. Corresponding estimates from countries lower in the income ladder are closer in magnitude: 20 and 29 percent in Estonia (Masso, Meriküll, and Vahter, 2022), 32 and 36 percent in Chile (Cruz and Rau, 2022), and 45 and 57 percent for Brazil (Morchio and Moser, 2021).

In this paper, we focus on the contribution of sorting rather than bargaining (i.e., within-firm gender differences in pay). However, the contribution of the latter to the overall gap appears very small in our context. Gender differences in bargaining, i.e., the mean difference in female and male premia are -0.00081 log points (when weighted by the female employment distribution) and -0.00990 log points (when weighted by the male employment distribution). In obtaining these, we follow the literature (Card, Cardoso, and Kline 2016; Casarico and Lattanzio 2022; Morchio and Moser 2021) in normalizing male and female firm premia using firms in the the restaurant and fast-food industry. Therefore, in what follows, we focus exclusively on gender differences in sorting.

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 more (to *all* their workers).[7] 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 mov-

[7]As we do not have information on occupation, firm pay 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, Cardoso, and Kline, 2016; Coudin, Maillard, and Tô, 2018) or include coarse controls such as dummies for white-collar and managerial occupations (Casarico and Lattanzio, 2022), or include detailed occupation dummies (Morchio and Moser, 2021).

ing 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 resulting two gender-specific markdowns.[8] 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, S. X. Li, and Han, 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 (Figure B4). The next section discusses how firm pay 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 pay between women and men are either zero, or favor women, up until workers cross their mid-20s. After age 28, a 3 percent gap between women and men expands to 15 percent by age 33, then expands more slowly to reach 27 percent by age 49, after which it declines.[9] The gender gap in firm premia also follows this U-shape, and its contribution to the raw gap in pay holds steady at 40-50 percent over much of the life cycle. While other studies investigating life cycle gender pay gaps show some evidence on narrowing gaps after cohorts are in their late

[8] 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.

[9] We observe a similar U-shape in the gender gap in pay in the labour force survey data (Figure B2).

40s, they either terminate their analysis at this point (Barth, S. P. Kerr, and Olivetti, 2021) or do not investigate the convergence that they observe (Casarico and Lattanzio, 2022).

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

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 pay 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). We find similar patterns when we do not include cohort controls (Figure C6). To allow for a more transparent picture of cohort differences, we also reproduce our main results separately for 10- and 4-year birth cohorts, wherein each cohort traces out a similar pattern of a U-shape in the firm-pay gap (Figure C7).

4 Gender gaps in labour market transitions

In this section, we attempt to shed more light on why the firm-pay gender gap rises over the life cycle—why do women increasingly sort into relatively 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-pay gap over the life cycle.

We split all workers at each age into stayers (those employed in the previous and next period), entrants (those not employed in the previous period), and leavers (those employed in the previous, but not the next period). We then decompose the firm-pay gap into gaps within each category and gender differences in population shares of each category, as follows:

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

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 and leavers (weighted by average firm premia of stayers and leavers, respectively, relative to female entrants).[10]

Figure 2 shows that the U-shape in the firm-pay 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-pay gender gap.

The contribution of leavers and entrants is substantial but remains relatively steady over the life cycle. Gender gaps in firm premia among leavers and entrants are actually larger than

[10]As female and male proportions of stayers, leavers, and entrants are close to identical, the type of weighting and choice of reference group does not matter much, i.e., patterns are nearly the same when we apply alternative weighting and reference group choices.

among stayers (Figure C8), but their contribution is balanced out by the smaller proportion of this group relative to stayers. Importantly, the contribution of firm-pay gaps among leavers to the overall firm-pay gap remains steady at 17-21 percent over the life cycle. The contribution of firm-pay gaps among entrants increases initially, but the rate of increase is lower than the overall increase in the gender gap, resulting in a declining contribution of entrants to the gap (from 35 percent in the late 20s to 18 percent in the mid-50s). Dynamically, since the contribution of entrants is higher than that of leavers until the early 50s, the average firm-pay gap is bound to grow over this range, analogous to cohort effects documented elsewhere (Arellano-Bover et al., 2024). 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.[11]

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 C10). 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-pay gender gaps (Figure C11).

The drivers thus 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 pay premia, it is the latter that drives the evolution of gender gaps in firm premia.

[11]See Figure C9. This is consistent with labour force survey data showing no decline in formal sector participation among women during childbearing years.

5 Discussion

5.1 Firm-pay gender gaps among entrants

Formality, churn, and gender gaps in firm premia. The importance of the decomposition of the firm-pay 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. The low share of formal employment in South Africa (a characteristic of developing countries) in turn implies high churn or transition rates in and out of formal employment: Donovan, Lu, and Schoellman (2023) show this generally, and in South Africa the share of formal sector workers who are continuously employed never rises above 70 percent (Figure C9). As noted in Section 4, firm-pay gender gaps among entrants are larger than among stayers — for example, due to employment gap penalties or re-training costs. If entrants are a larger proportion of formal employment in settings where the formal sector is relatively small, the firm-pay gender gap would be correspondingly larger.

Figure 3 shows how these outcomes vary across regions. Panel (a) shows the formal employment share is significantly positively associated with the share of the continuously employed (i.e. stayers) in total formal employment. Given that low levels of formality may be associated with a wider dispersion of firm-pay premia via low labour supply elasticities (Bassier, 2023), a more dispersed distribution of firm premia would additionally worsen the firm-pay gender gap if women are positioned lower in the firm-pay hierarchy. Panel (b) shows a corresponding narrowing of the firm-pay gender gap in regions with high levels of formality.[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 sorting across firms to gender pay inequality in comparison to high income country estimates.

[12] We do not account for any selection effects 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.

Predictors of the gap in firm premia among entrants. Why do we observe a high firm-pay 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-pay gender gap among entrants. We 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 observed covariates, discussed below.

Male entrants are slightly older (Table C2 gives covariate means), which may predict differential ability to move up the firm-pay ladder (Haltiwanger, Hyatt, and McEntarfer, 2018), but this explains only 1 percent 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, Rathelot, and Roulet, 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, Cardoso, and Kline, 2016) or at firms covered by bargaining council agreements (Bassier 2022; Corradini, Lagos, and Sharma 2022). These firm- or demand-side factors explain about 21 percent.

Finally, nearly half the gap (44 percent) is explained by 2-digit industry (Petrongolo and Ronchi, 2020; Folbre, Gautham, and Smith, 2023). Women are disproportionately employed in lower paying sectors like education, retail, personal care services, and food or textile manufacturing subsectors (Figure C12). 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.[13] It appears the explicit gender-progressive government hiring policy has played

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

a substantial role in decreasing the firm-pay gender gap: back-of the-envelope calculations, based on female employment shares while keeping firm premia constant, suggest that this helps reduce the gender pay gap by 17 percent.[14] Overall, our evidence is suggestive of female entrants 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 U-shape evolution of the firm-pay 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-pay 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 (Figure B4). Household care responsibilities—as 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 4 panel (a), we decompose the growth in the firm-pay gender gap over the life cycle 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-pay growth comes from women having lower firm-pay increases *conditional* on switching.

Under existing models of firm pay premia, one may think of the relationship between child-related constraints and this pattern in the following two ways. Women's moves to higher

[14]Public administration constitutes 25.1 percent of all female employment, but only 15.9 percent 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 pay gap by 2.1 percentage points (or 17 percent of the overall gap).

paying firms 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, Cardoso, Heining, et al., 2018, i.e. new classical monopsony models). This would imply women may be more likely to make inter-firm moves which trade lower firm pay premia for higher non-pay amenities like flexibility. Alternatively, women's moves to firms paying higher pay 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 4 (a).

Later convergence of gaps. If the presence of young children constrains women's ability to transition to higher-paying 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.

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 pay premia. In canonical Burdett-Mortenson models, for example, transition rates to higher pay firms increase with the proportion of firms above a worker's current firm in the pay 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 (homogeneous) 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 panel (b) 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.

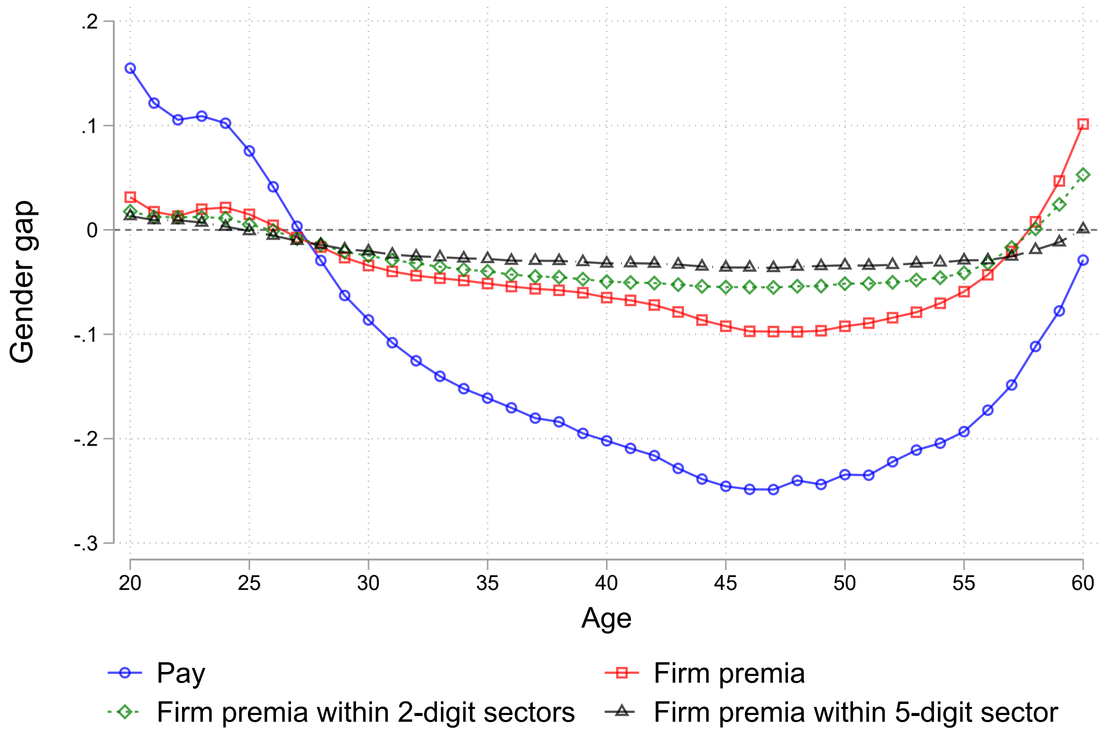
6 Conclusion

We have shown that sorting into lower-paying firms explains nearly half of the gender pay 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 pay premia by documenting a U-shape in the firm-pay 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-pay 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 pay gaps. We suggest that for developing countries, the gap is partly due to higher transition rates in and out of formal employment, along with bigger pay gaps for women entrants. However, the key mechanism driving the growth of the sorting gap over the life cycle is the inability of employed women to climb up the firm-pay 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.

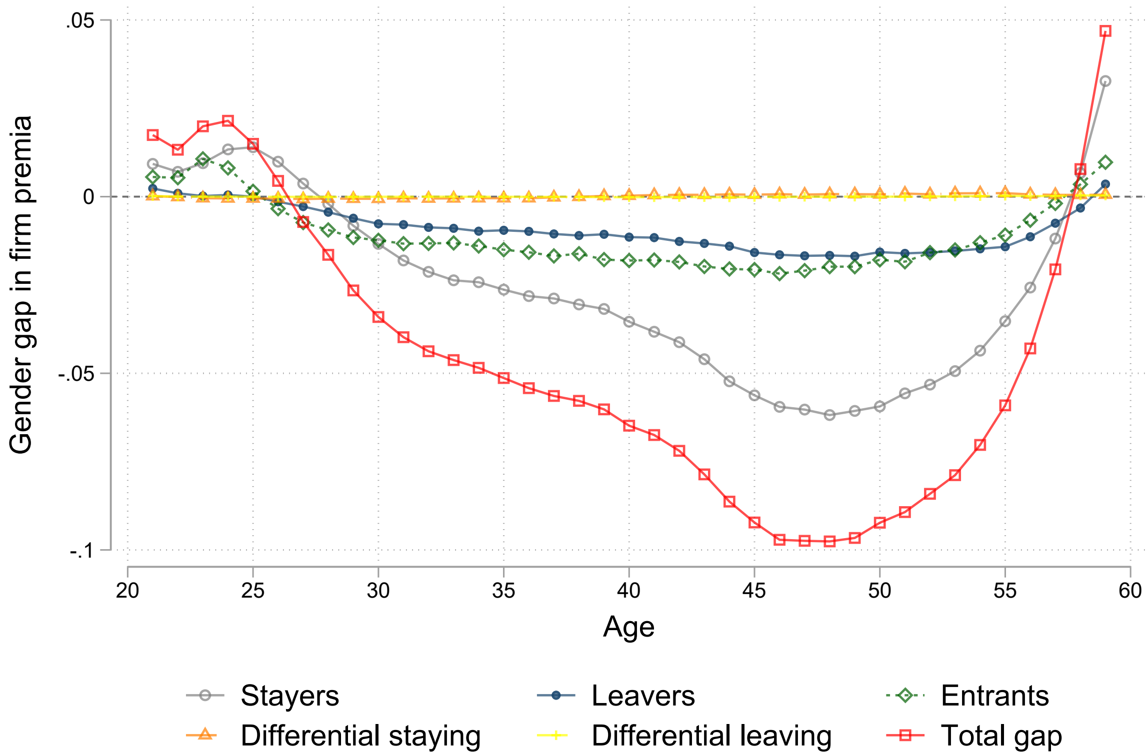
Figures

Figure 1: **Gender gap in firm pay over the life cycle**



Notes. Each data point represents the average gap between the pay of men and women working in the formal sector sample at each age. Pay represents annual earnings, firm premia are estimated based on Kline, Saggio, and Sølvsten (2020), and firm premia within sectors are estimated by residualizing on the relevant sector fixed effects. See text for sample details. *Source*: Own calculations, South African tax records, 2010–2018.

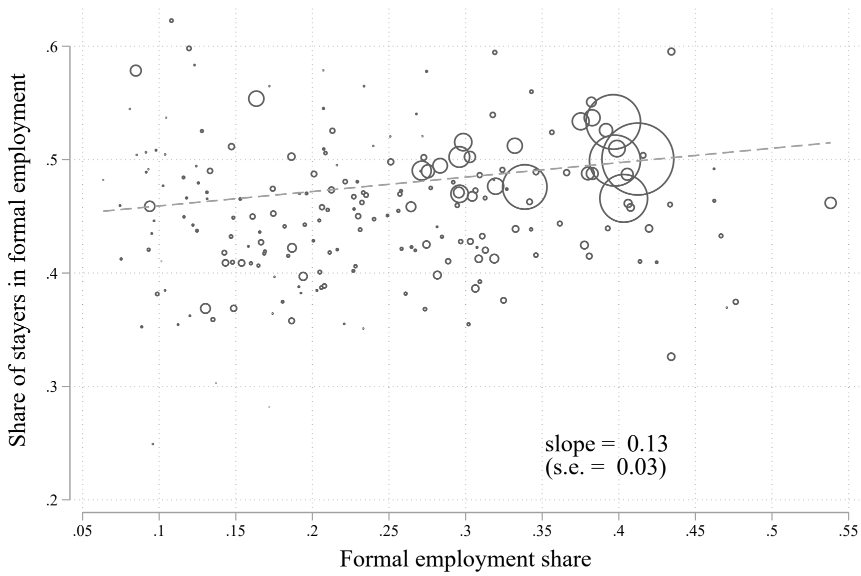
Figure 2: **Decomposing the gender gap in firm pay premia over the life cycle**



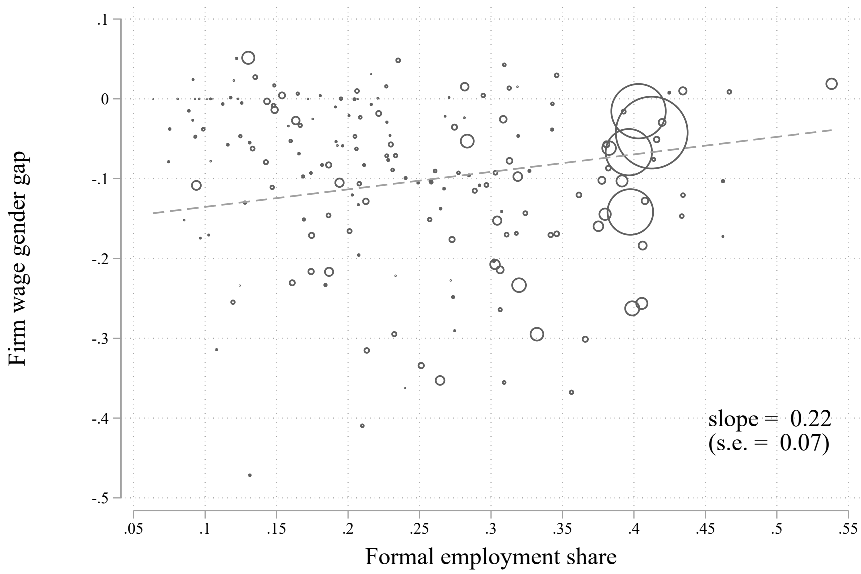
Notes. The figure plots the contributions to the gender gap in firm pay premia by category, corresponding to equation 2. Stayers are defined as workers who are continuously employed, with changes in firm pay premia arising from switches across firms. Leavers are workers who are not observed in formal 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 weighted sum of these components (see text for details on weighting), and corresponds to the total cross-sectional firm-pay gender gap. *Source*: Own calculations, South African tax records, 2010–2018.

Figure 3: **Formal employment share and regional firm-pay dynamics**

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



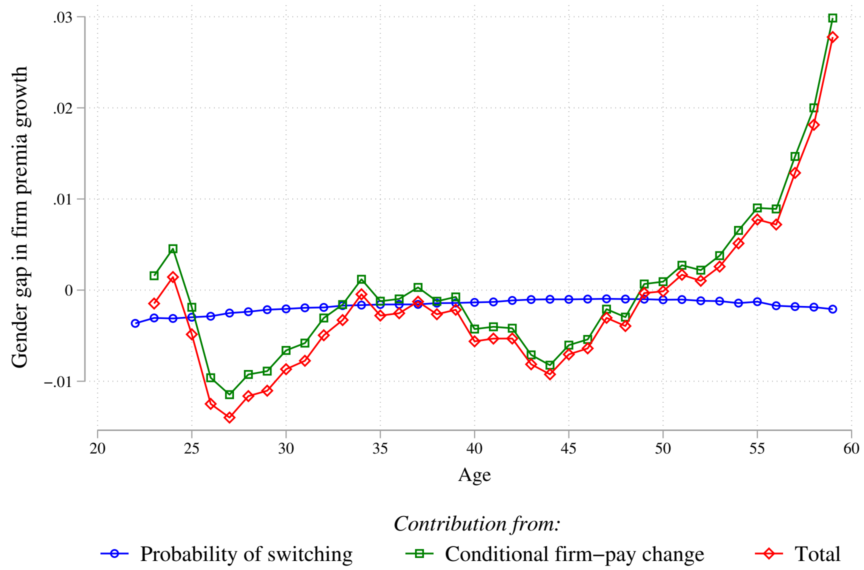
(b) **Firm-pay gender gap**



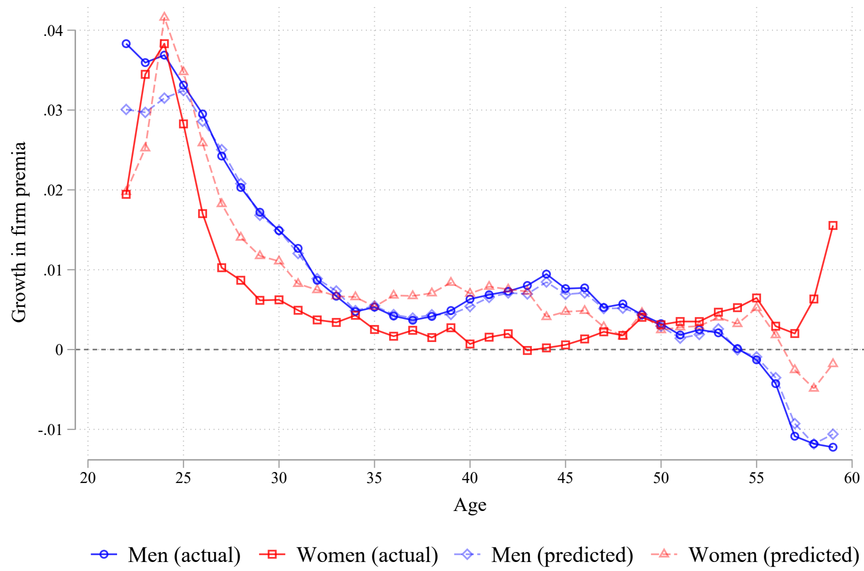
Notes. The scatterplots show the formal employment share across regions against (a) the share of stayers (i.e., workers who are continuously employed) in formal employment, and (b) the mean firm-pay gender gap. The formal employment share is measured at the municipal regional level (N=208), as the formal employment to population ratio from the 2011 Census. The share of stayers and mean firm-pay gender gap are obtained from tax records. Municipalities are weighted by population size. We drop municipalities with implausibly low firm-pay gender gaps (i.e., female firm-pay higher by 6 log points, *N=18*). *Source*: Own calculations, South African tax records, 2010–2018.

Figure 4: **Mechanisms of firm-pay changes among the continuously employed**

(a) **Gender gap contributions from switching vs. firm-pay changes**



(b) **Actual vs. predicted firm-pay growth based on firm-pay ranking**



Notes. Both panels are limited to stayers, i.e. continuously employed workers. Panel (a) decomposes the change in the gender gap in firm premia over the life cycle ("total") into the contribution of women being less likely to switch firms than men ("probability of switching") and being less likely to make advantageous moves conditional on switching ("conditional firm-pay change"). Panel (b) shows the average growth in firm pay premia by gender and age. "Actual" (in darker shade) corresponds to the observed growth. "Predicted" (in lighter shade) corresponds to the predicted growth in firm pay premia, based on a simple regression of the growth in firm premia on the current level of firm premia for all workers, then predicted separately by worker gender and age. *Source*: Own calculations, South African tax records, 2010–2018.

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	

Notes. The decomposition uses coefficients from the pooled regression of both women and men to weight the contribution of the explained components. Levels of firm premia cannot be meaningfully interpreted as have estimated firm premia without specifying or restricting the reference firm. The average gender gap in firm premia and the contribution of sorting to the pay gap are unaffected by the choice of reference firm or normalization. Covariates included are age, commuting distance, firm value added, whether covered by bargaining council agreement, and 2-digit industry (a list limited by information on worker and firm characteristics available in our administrative data). 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: Own calculations, South African tax records, 2010–2018.

References

Abowd, John M, Francis Kramarz, and David N Margolis (1999). "High wage workers and high wage firms". In: *Econometrica* 67.2, pp. 251–333.

Arellano-Bover, Jaime et al. (2024). "One Cohort at a Time: A New Perspective on the Declining Gender Pay Gap". In: *Available at SSRN*.

Arrow, Kenneth J (1971). *Some models of racial discrimination in the labor market*. Tech. rep. Rand Corporation, Santa Monica.

Barbanchon, Thomas, Roland Rathelot, and Alexandra Roulet (2021). "Gender differences in job search: Trading off commute against wage". In: *The Quarterly Journal of Economics* 136.1, pp. 381–426.

Barth, Erling, Sari Pekkala Kerr, and Claudia Olivetti (2021). "The dynamics of gender earnings differentials: Evidence from establishment data". In: *European Economic Review* 134, p. 103713.

Bassier, Ihsaan (2022). *Collective bargaining and spillovers in local labor markets*. Tech. rep. CEP Discussion Paper No. 1895.

— (2023). "Firms and inequality when unemployment is high". In: *Journal of Development Economics* 161, p. 103029.

Bassier, Ihsaan et al. (2021). "Locked down and locked out: Repurposing social assistance as emergency relief to informal workers". In: *World Development* 139, p. 105271.

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

Berniell, Inés et al. (2021). "Gender gaps in labor informality: The motherhood effect". In: *Journal of Development Economics* 150, p. 102599.

Bonhomme, Stéphane et al. (2023). "How Much Should We Trust Estimates of Firm Effects and Worker Sorting?" In: *Journal of Labor Economics* 41.2, pp. 291–322. DOI: 10.1086/720009. eprint: https://doi.org/10.1086/720009. URL: https://doi.org/10.1086/720009.

Bruns, Benjamin (2019). “Changes in workplace heterogeneity and how they widen the gender wage gap”. In: *American Economic Journal: Applied Economics* 11.2, pp. 74–113.

Budlender, Joshua and Amina Ebrahim (2020). *Industry classification in the South African tax microdata*. Tech. rep. United Nations University World Institute for Development Economics Research.

Card, David, Ana Rute Cardoso, Jörg Heining, et al. (2018). “Firms and labor market inequality: Evidence and some theory”. In: *Journal of Labor Economics* 36.S1, S13–S70.

Card, David, Ana Rute Cardoso, and Patrick Kline (2016). “Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women”. In: *The Quarterly Journal of Economics* 131.2, pp. 633–686.

Casarico, Alessandra and Salvatore Lattanzio (2022). “What firms do: Gender inequality in linked employer-employee data”. In: *Journal of Labor Economics*.

Corradini, Viola, Lorenzo Lagos, and Garima Sharma (2022). “Collective Bargaining for Women: How Unions Can Create Female-Friendly Jobs”. In.

Coudin, Elise, Sophie Maillard, and Maxime Tô (2018). *Family, firms and the gender wage gap in France*. Tech. rep. IFS Working Papers.

Cruz, Gabriel and Tomás Rau (2022). “The effects of equal pay laws on firm pay premiums: Evidence from Chile”. In: *Labour economics* 75, p. 102135.

Donovan, Kevin, Will Jianyu Lu, and Todd Schoellman (2023). “Labor market dynamics and development”. In: *The Quarterly Journal of Economics* 138.4, pp. 2287–2325.

Folbre, Nancy, Leila Gautham, and Kristin Smith (2023). “Gender inequality, bargaining, and pay in care services in the United States”. In: *ILR Review* 76.1, pp. 86–111.

Goldin, Claudia et al. (2017). “The expanding gender earnings gap: Evidence from the LEHD-2000 Census”. In: *American Economic Review: Papers and Proceedings* 107.5, pp. 110–114.

Haltiwanger, John, Henry Hyatt, and Erika McEntarfer (2018). “Who moves up the job ladder?” In: *Journal of Labor Economics* 36.S1, S301–S336.

He, Haoran, Sherry Xin Li, and Yuling Han (2023). “Labor market discrimination against family responsibilities: A correspondence study with policy change in China”. In: *Journal of Labor Economics* 41.2, pp. 361–387.

Kerr, Andrew, David Lam, and Martin Wittenberg (2019). “Post-Apartheid Labour Market Series 1993-2019”. In: *[dataset]. Version* 3.

Kerr, Andrew and Martin Wittenberg (2019). *Earnings and employment microdata in South Africa*. WIDER Working Paper 2010/47. UNU-WIDER.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten (2020). “Leave-out estimation of variance components”. In: *Econometrica* 88.5, pp. 1859–1898.

Labanca, Claudio and Dario Pozzoli (2022). “Constraints on hours within the firm”. In: *Journal of Labor Economics* 40.2, pp. 473–503.

Li, Jiang, Benoit Dostie, and Gaëlle Simard-Duplain (2023). “Firm Pay Policies and the Gender Earnings Gap: The Mediating Role of Marital and Family Status”. In: *ILR Review* 76.1, pp. 160–188.

Manning, Alan (2003). *Monopsony in motion: Imperfect competition in labor markets*. Princeton: Princeton University Press.

Masso, Jaan, Jaanika Meriküll, and Priit Vahter (2022). “The role of firms in the gender wage gap”. In: *Journal of Comparative Economics* 50.2, pp. 454–473.

Morchio, Iacopo and Christian Moser (2021). “The Gender Pay Gap: Micro Sources and Macro Consequences”. In.

Petit, Pascale (2007). “The effects of age and family constraints on gender hiring discrimination: A field experiment in the French financial sector”. In: *Labour Economics* 14.3, pp. 371–391.

Petrongolo, Barbara and Maddalena Ronchi (2020). “Gender gaps and the structure of local labor markets”. In: *Labour Economics* 64, p. 101819.

A Data Appendix

A.1 South African Tax Records

A.1.1 Access

The tax 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: February 28, 2025.

A.1.2 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, and 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, and 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, and 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).

A.1.3 Cleaning and sample notes

We begin with a sample of 75.3 million worker-year observations observed between 2010 and 2018, restricted to those aged 20-60 years. For our main results, where we apply the Kline, Saggio, and Sølvsten (2020) (KSS) leave-out estimator, we compute firm premia for a smaller subset of years (2011–2016) due to the computational intensity of this procedure. That is, we estimate firm premia (i.e., AKM firm fixed effects using the KSS leave-out estimator) for every firm that we observe in 2011–2016. Therefore, we have firm premia even for workers in 2010 or 2017–2018, provided that the firms that they are employed in were observed in 2011–2016. In practice, we drop 1.13 million worker-year observations, and compute firm premia for the remaining sample of 74.2 million worker-year observations. In order to ensure that our firm premia are computed over a sufficiently large group of workers, we further restrict our sample to firms with at least 20 worker-year observations in each period, with a resulting sample of 63.4 million worker-year observations. We apply the restriction that at least 10 workers of either gender are in each period, which drops 0.8 million worker-year observations, with a final sample of 62.6 million worker-year observations.

We classify a separation of a worker from a firm when they are not observed at the same firm in the following year, since we observe the universe of formal sector firms. Stayers are defined as workers who are continuously employed, with changes in firm pay 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 but partial data cleaning and sample construction decisions; users are referred to our do-files which are available on request.

A.2 South African Labour Force Survey (LFS)

The LFS is a household-based sample survey conducted by Statistics South Africa (Stats SA), collecting data on the labour market activities of individuals aged 15 years or older. We obtain LFS data for 2010-2018 from the Post-Apartheid Labour Market Series, a publicly-available compilation of nationally representative South African household surveys (A. Kerr, Lam, and Wittenberg 2019).[15] We restrict the sample to individuals aged 20-60. Formal sector workers are those who report that their employers contribute to the Unemployment Insurance Fund (UIF) on their behalf or who are in the public sector (public sector employers do not make UIF contributions), who have a written contract, and who are not domestic workers. All other wage workers are defined as informal sector workers. The formal sector definition follows A. Kerr and Wittenberg (2019). To create a population of employees comparable to those in the tax data it is more conservative than others definitions which simply require a written contract (see, for e.g., Bassier et al. 2021). When comparing LFS statistics against our tax data, we further restrict the formal sector to workers in firms with greater than 20 workers, in order to parallel the restriction that we impose on the tax data. We use survey weights derived by DataFirst from Stats SA mid-year population estimates for 2018 for all estimates.

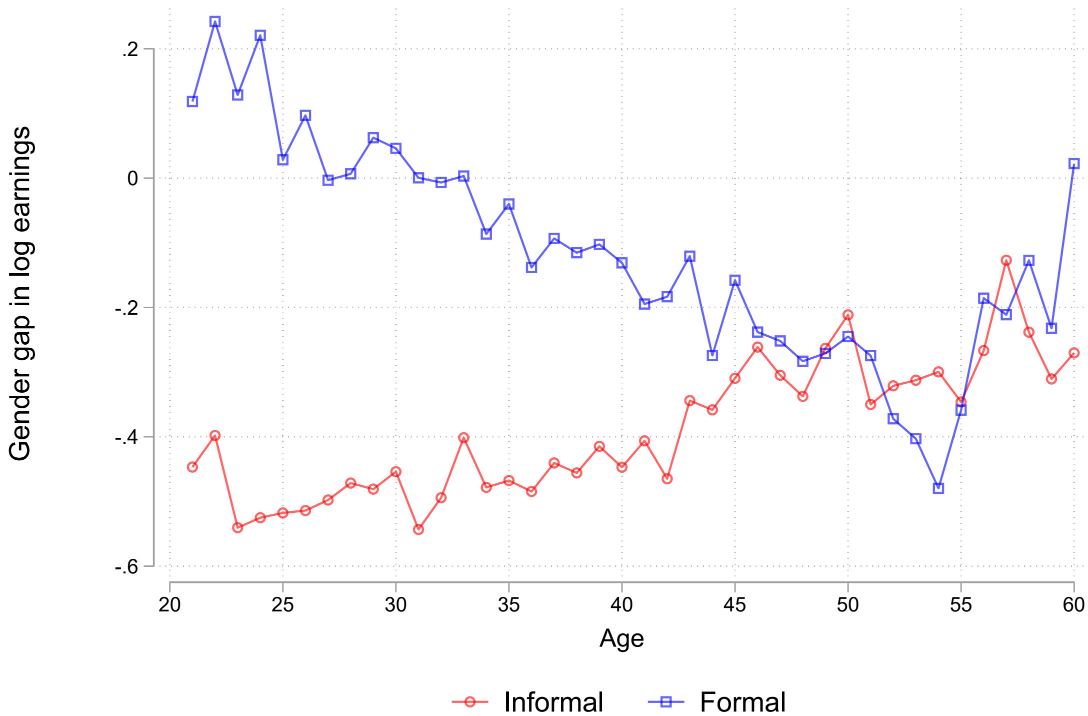
A.3 National Income Dynamics Study (NIDS)

The National Income Dynamics Study (NIDS) is a panel survey, conducted by the Southern Africa Labour and Development Research Unit (SALDRU), that began in 2008 and surveyed the same participants every roughly two years. We use the most recent wave from 2017 (Wave 5), obtained from the NIDS repository (http://www.nids.uct.ac.za/). We restrict the sample to women aged between 20 and 60.

[15] The dataset and the code used to create the data are publicly available from DataFirst, a data repository at the University of Cape Town (www.doi.org/10.25828/gtr1-8r20).

B Appendix: Evidence from representative household surveys

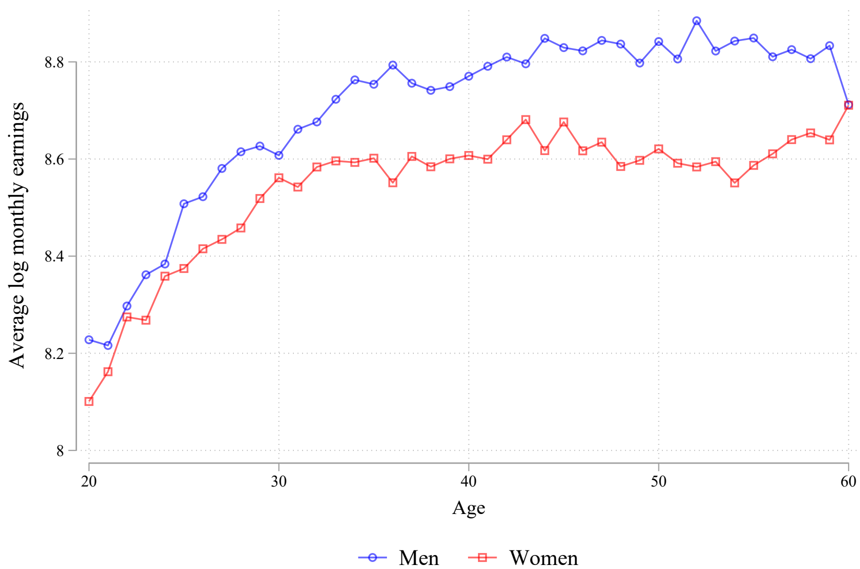
Figure B1: **Formal vs informal gender gap in earnings, by age (LFS)**



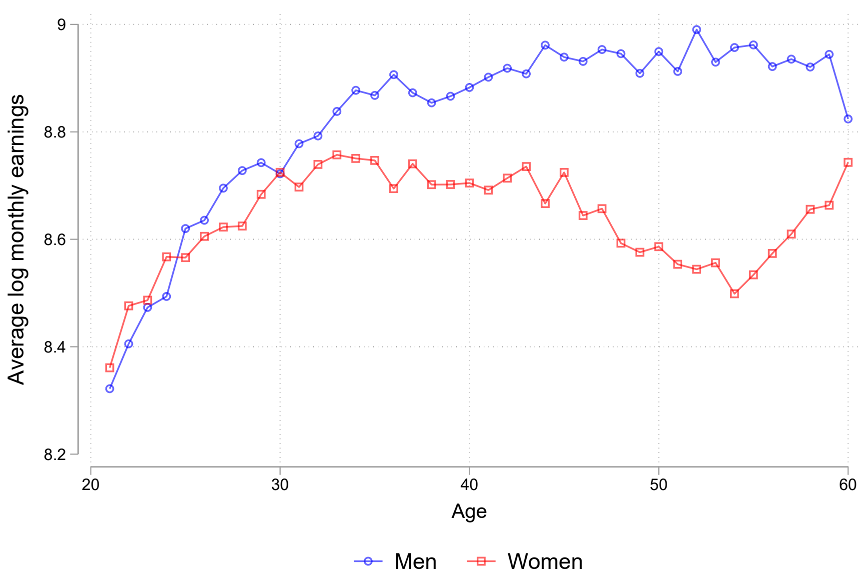
Notes. The y-axis shows the gender gap in average log monthly earnings. The sample is from representative household survey data, restricted to employed people with non-missing earnings between the ages of 20 and 60. Formal sector workers are employees who report that their employers contribute to the Unemployment Insurance Fund (UIF) on their behalf or who are in the public sector, who have a written contract and who are not domestic workers (see Data Appendix for further details). All other wage and salary workers are classified as informal. *Source*: Own calculations, South African Labour Force Survey, 2010–2018.

Figure B2: **Formal worker average earnings by gender (LFS)**

(a) **Without cohort controls**



(b) **With cohort controls**

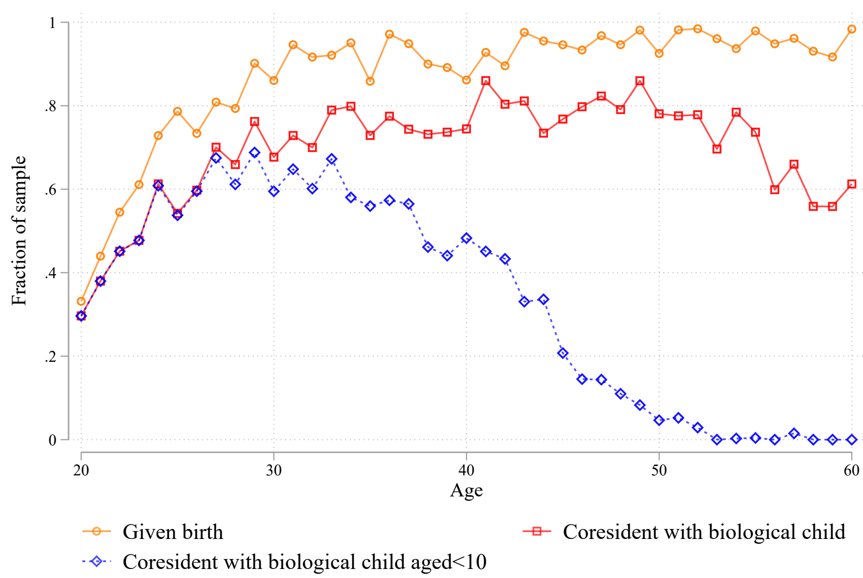


Notes. The y-axis shows average log monthly earnings by gender at each age. The sample is from representative household survey data, restricted to formal sector workers between the ages of 20 and 60 in firms with over 20 workers (see Data Appendix for further details). Panel (b) includes cohort controls: interaction of female with dummies for birth year. *Source*: Own calculations, LFS, 2010–2018.

Figure B3: **Formal worker gender earnings and wage gaps, by age (LFS)**

Notes. The sample is from representative household survey data, restricted to formal sector workers between the ages of 20 and 60 in firms with over 20 workers (see Data Appendix for further details). *Source*: Own calculations, LFS, 2010–2018.

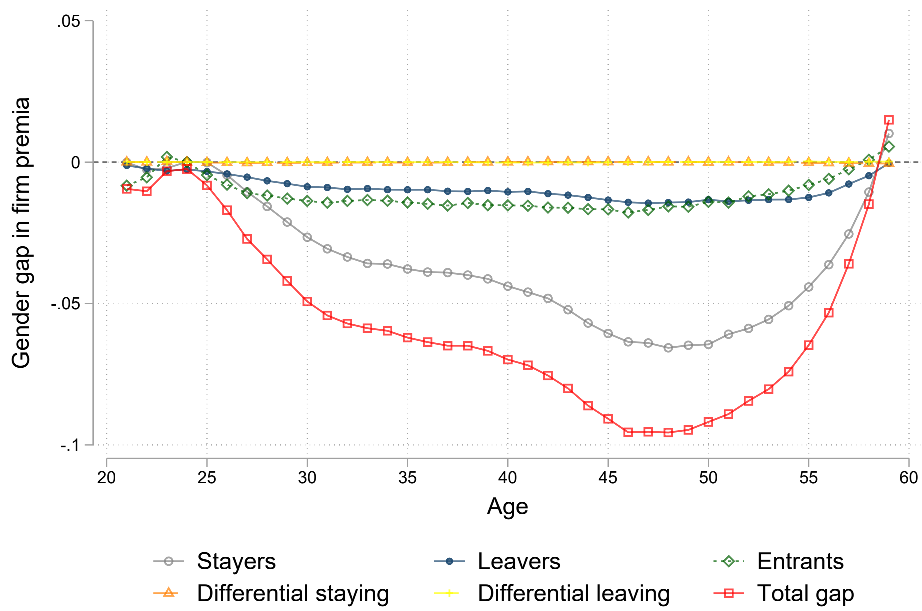
Figure B4: **Co-residence with children among women, by age (NIDS)**



Notes. The sample is from representative household survey data, restricted to women between the ages of 20 and 60. Each data point shows the fraction of the sample at any given that has ever given birth, is co-resident with a biological child, or is co-resident with a biological child under 10. *Source*: Own calculations, National Income Dynamics Study 2017, Wave 5.

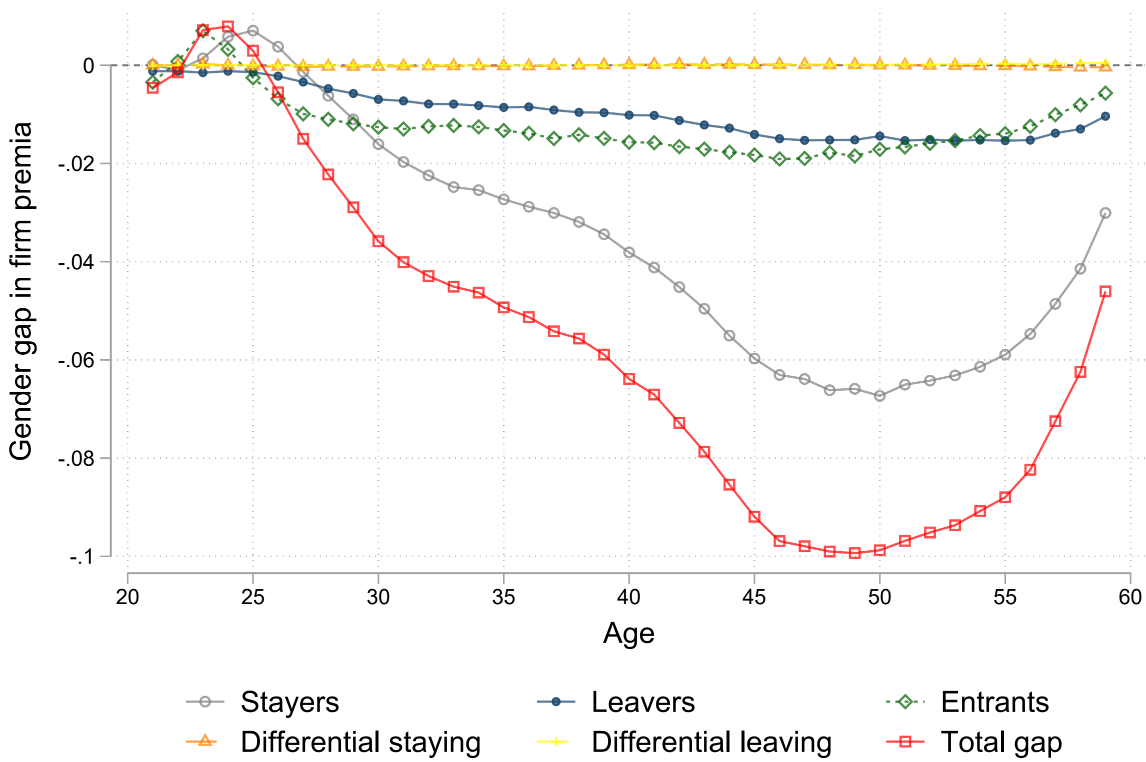
C Appendix: Additional tables and figures (tax data)

Figure C1: **Robustness on firm-pay gender gap decomposition: AKM firm premia for 2010–2018**



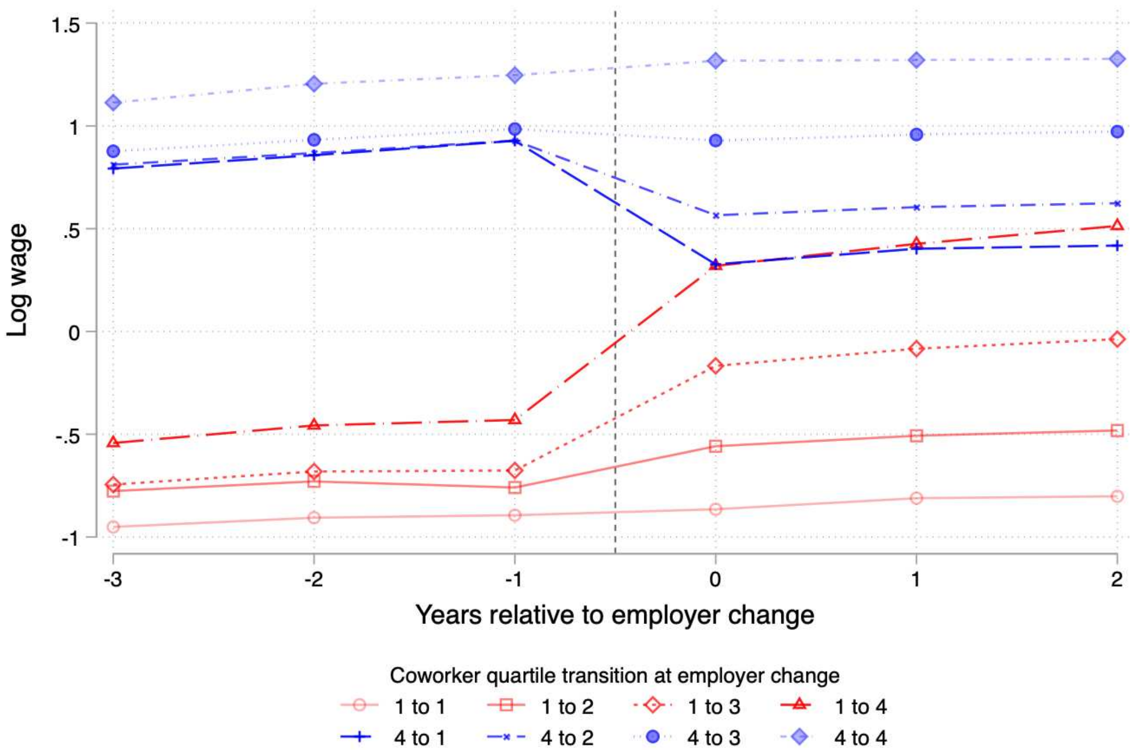
Notes. This figure is identical in construction and interpretation to Figure 2 in the main text, with the following two differences. Firstly, firm premia are computed by applying the AKM estimation (Abowd, Kramarz, and Margolis (1999) i.e. not the modified KSS leave-out estimation – see text for details). Secondly, the firm premia estimation sample uses the full number of years (2010–2018, rather than 2011-2016) and is restricted to firms with more than 120 workers over the full period, and there are no restrictions on the minimum number of workers by gender as in the main text. The figure plots the contributions to the gender gap in firm pay premia by category, corresponding to equation 2 in the main text. Stayers are defined as workers who are continuously employed, with changes in firm pay 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 weighted sum of these components (see text for details on weighting), and corresponds to the total cross-sectional firm-pay gender gap. *Source*: South African tax records, 2010–2018. *Source*: Own calculations, South African tax records, 2010–2018.

Figure C2: **Robustness on firm-pay gender gap decomposition: BLM correction**



Notes. This figure is identical in construction and interpretation to Figure 2 in the main text, with the following difference: firm premia are computed by applying the Bonhomme et al. (2023) estimation (i.e. not the modified KSS leave-out estimation – see text for details). The figure plots the contributions to the gender gap in firm pay premia by category, corresponding to equation 2 in the main text. Stayers are defined as workers who are continuously employed, with changes in firm pay 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 weighted sum of these components (see text for details on weighting), and corresponds to the total cross-sectional firm-pay gender gap. *Source*: South African tax records, 2010–2018. *Source*: Own calculations, South African tax records, 2010–2018.

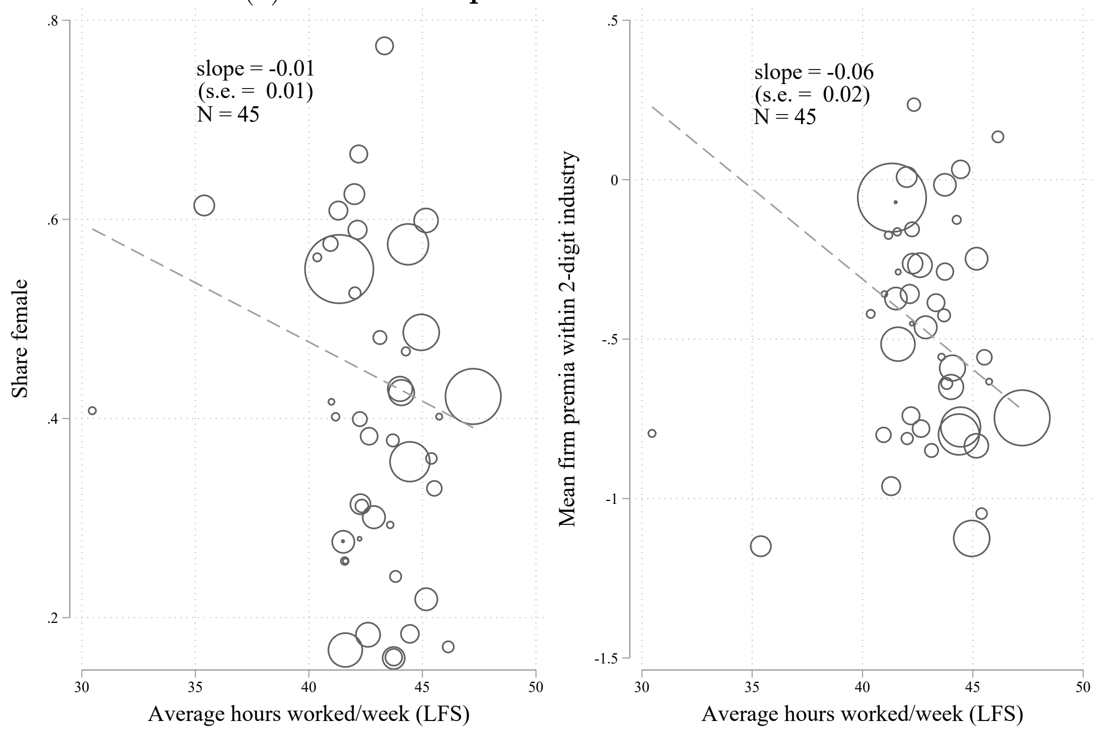
Figure C3: **Mean log pay of job changers, by quartile of mean co-worker pay at origin and destination firm**



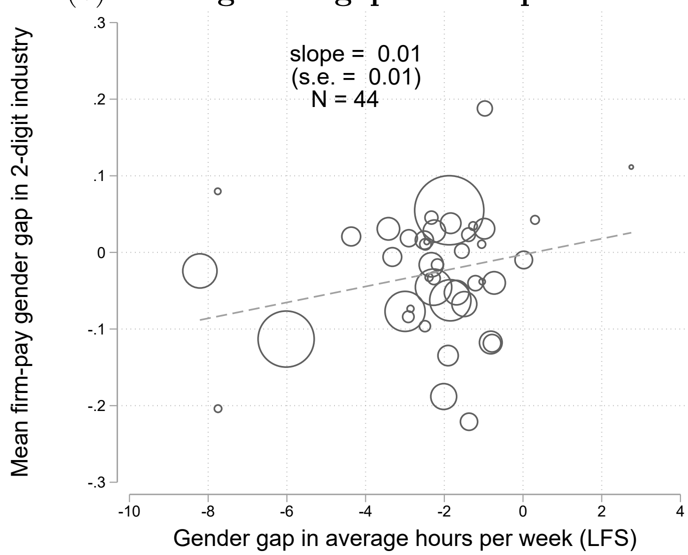
Notes. This figure reproduces checks from Card, Cardoso, and Kline (2016) that workers' moves across firms are not driven by unobserved factors that also affect pay. Plotted are mean log pay of job changers, by quartile of mean co-worker pay at origin and destination firm. Blue (red) lines represent workers initially at top (bottom) quartile of mean co-worker pay. *Source*: Own calculations, South African tax records, 2010–2018.

Figure C4: **Gender gaps in firm premia and hours, by 2-digit industry**

(a) **Mean firm premia and share female**

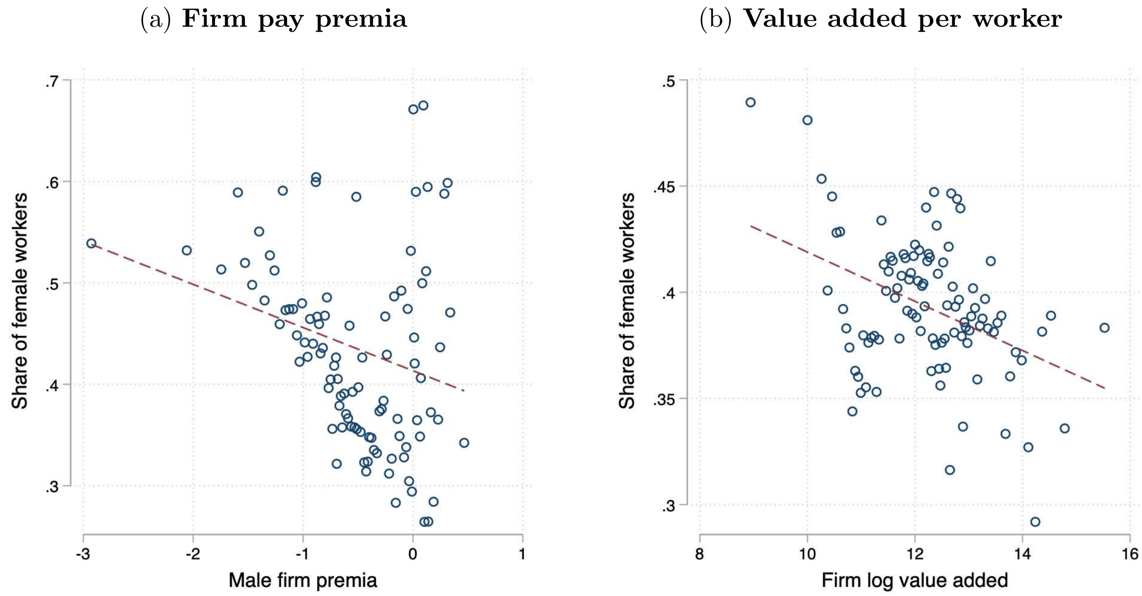


(b) **Mean gender gap in firm premia**



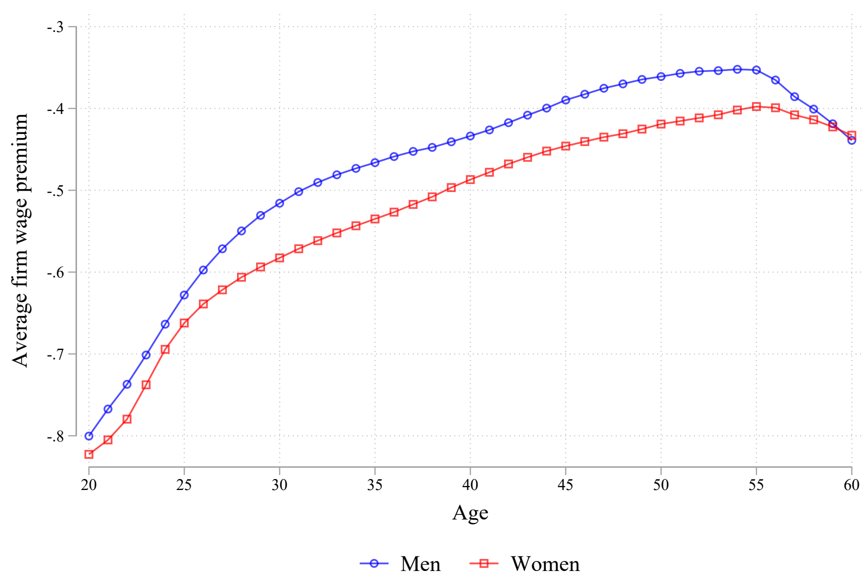
Notes. Panel (a) plots the relationship between average hours worked in 2-digit industry (from LFS data) against share female and mean firm premia within 2-digit industry (tax data). Panel (b) plots the relationship between average gender gap in hours worked in 2-digit industry (from LFS data) against mean gender gap in firm premia within 2-digit industry (tax data). Marker size given by the share of each 2-digit industry in total employment. *Source*: Own calculations, South African tax records, 2010–2018. Data on weekly hours worked from LFS, 2010–2018, with LFS sample restricted to formal sector workers between the ages of 20 and 60 in firms with over 20 workers.

Figure C5: **Share of female workers by firm value added and firm premia**



Notes. Binned scatterplot (100 bins) of share of female workers within firm against firm premia estimated based on Kline, Saggio, and Sølvsten (2020) (panel a) and firm value added per workers (panel b). Value added was calculated by subtracting cost of sales from gross sales. *Source*: Own calculations, South African tax records, 2010–2018.

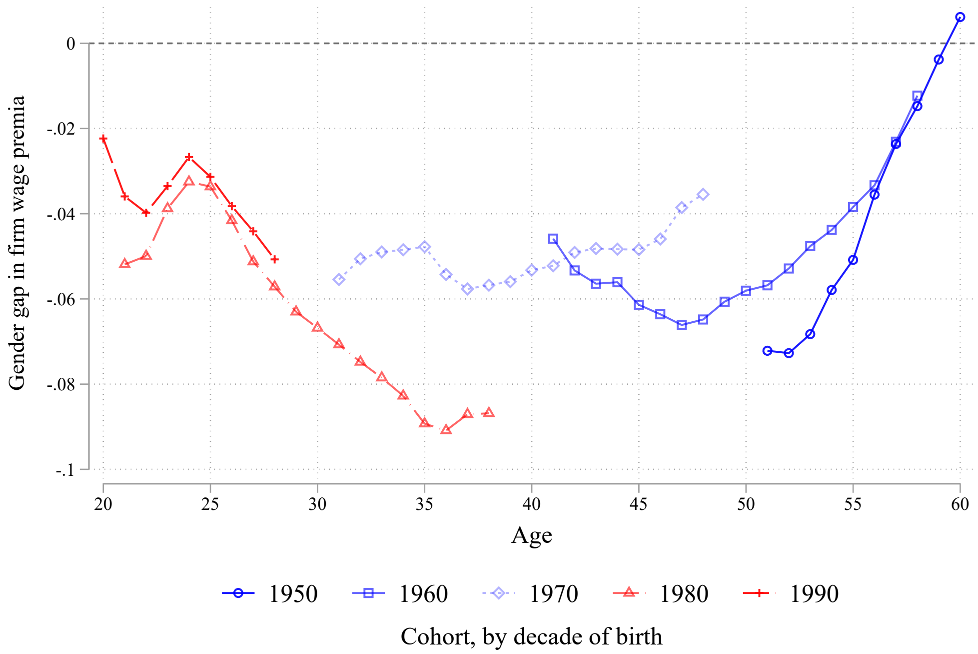
Figure C6: **Average firm pay premia by gender and age: No cohort controls**



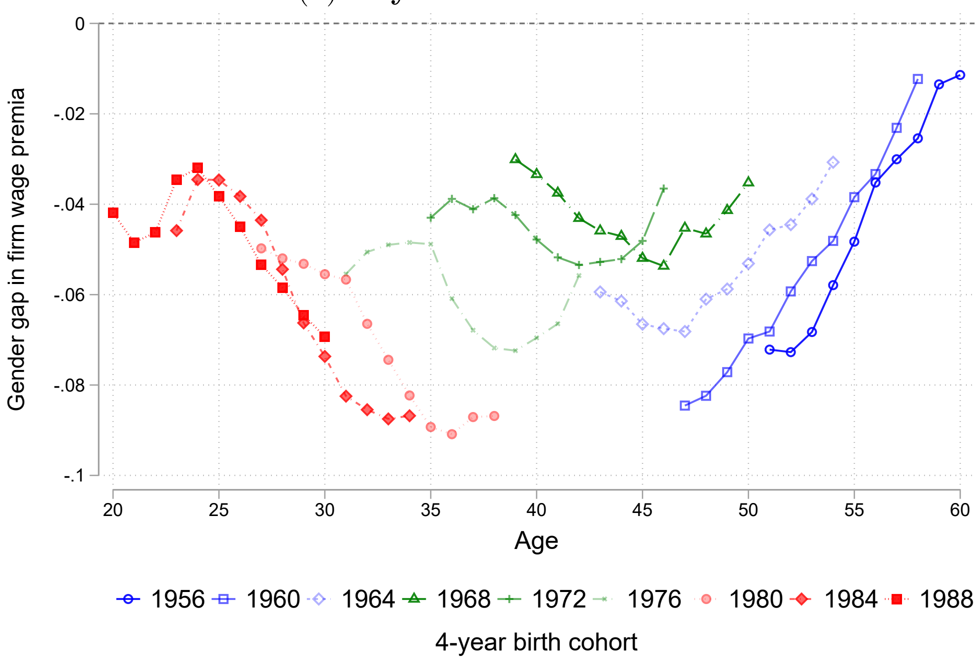
Notes. Each data point represents the average firm premia of men and women working in the formal sector sample at each age. Firm premia are estimated based on Kline, Saggio, and Sølvsten (2020). We do not account for cohort differences in any way. *Source*: Own calculations, South African tax records, 2010–2018.

Figure C7: **Gender gaps in firm pay premia, by age and birth cohort**

(a). **10-year birth cohorts**

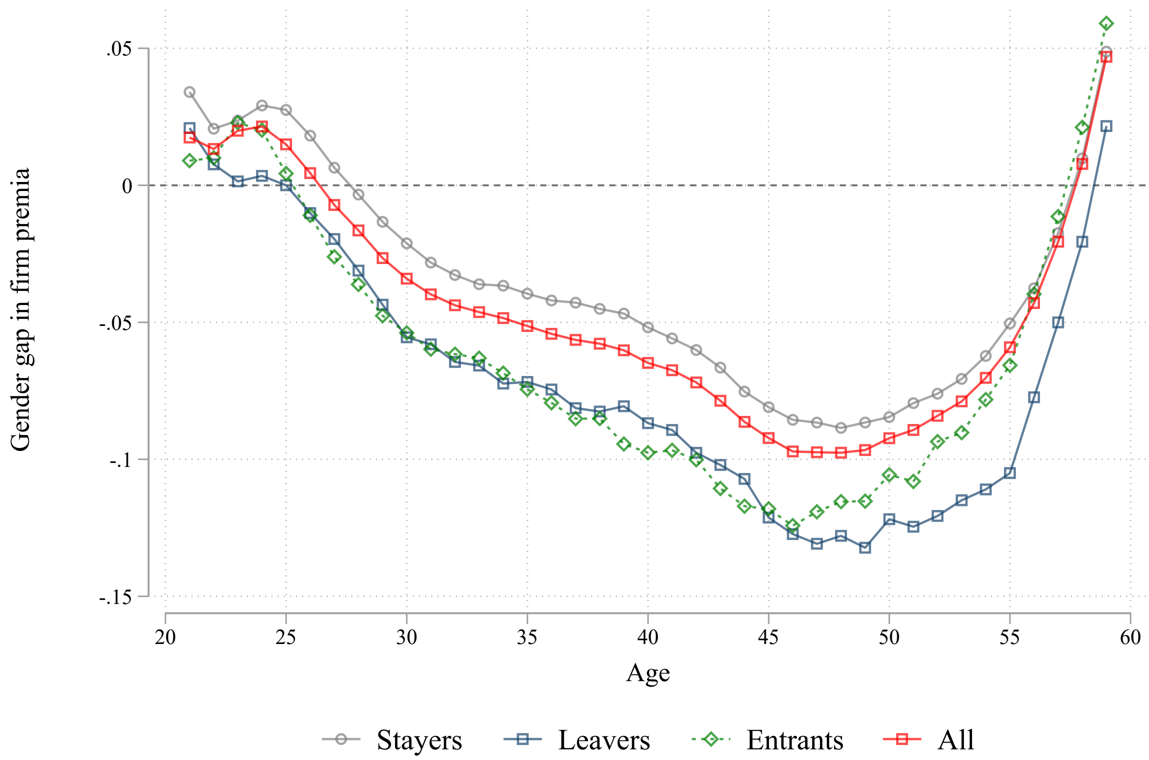


(b). **4-year birth cohorts**



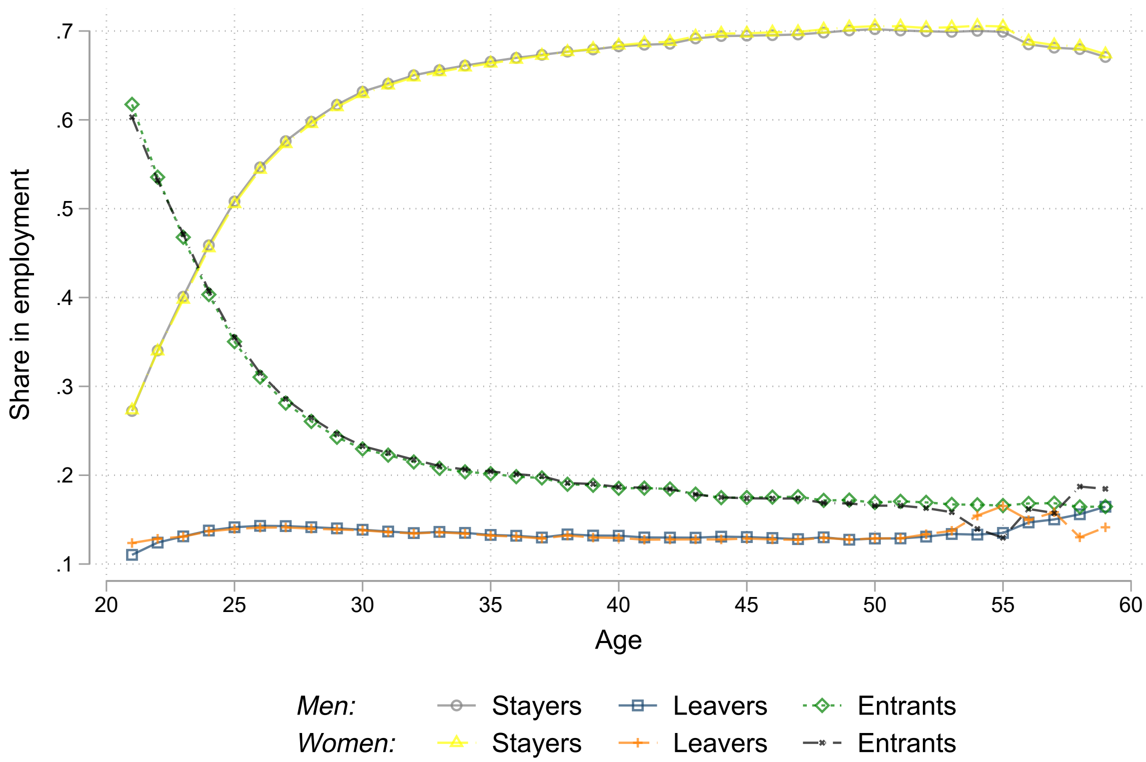
Notes. Each data point represents the gap in average firm premia between women and men working in the formal sector sample at each age. Instead of including our controls for cohort differences (interactions of birth-year with female) we compute gender gaps by age separately for 10-year (panel a) or 4-year (panel b) birth cohorts. *Source*: Own calculations, South African tax records, 2010–2018.

Figure C8: **The level of gender gaps in firm pay premia by category of worker, over the life cycle**



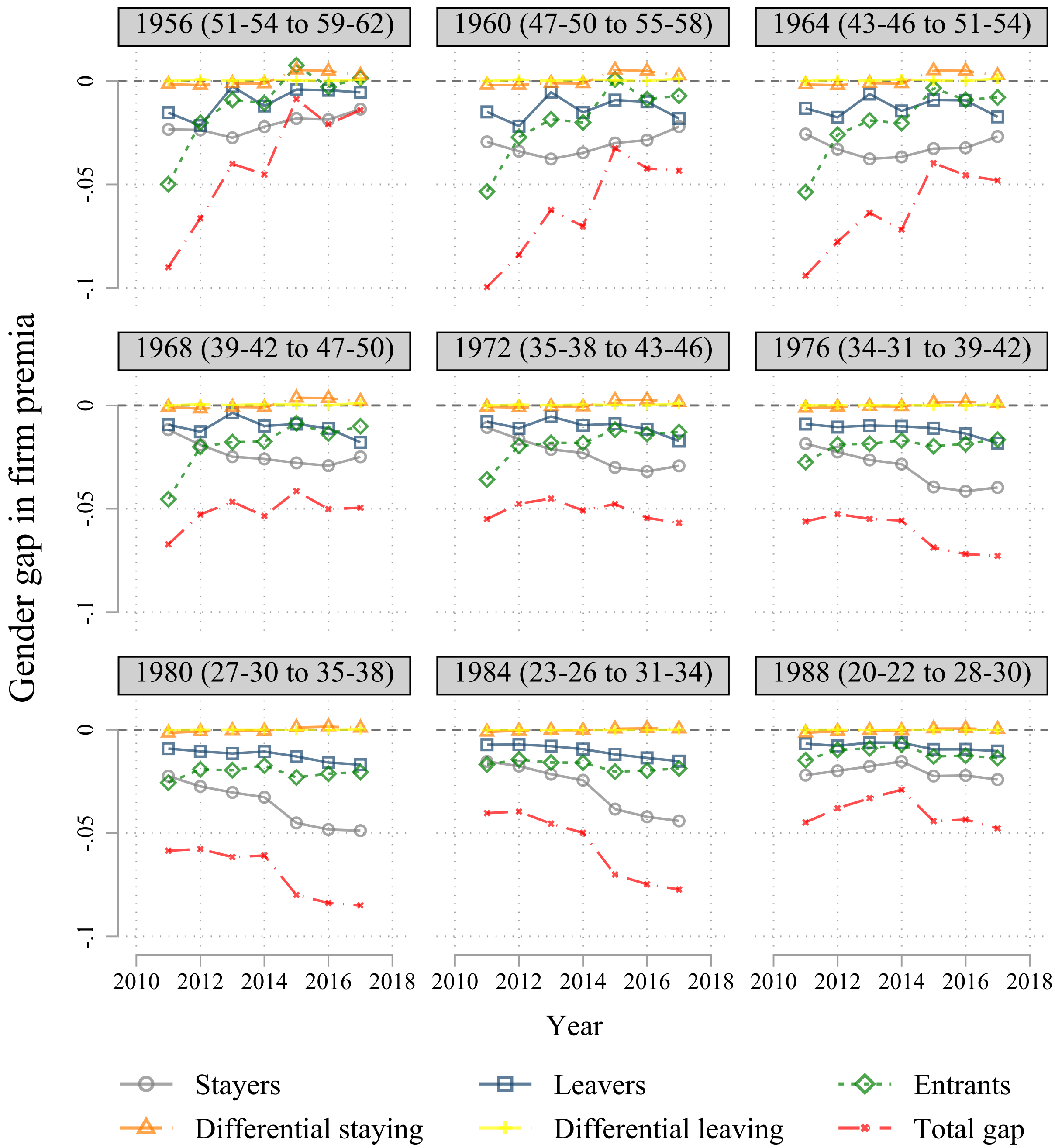
Notes. Each data point represents the average gap in firm premia between women and men within a particular subgroup of formal sector workers (i.e., stayers, leavers, or entrants) at each age. Note that while the main text figure 2 shows the contribution of each category to the overall gender gap, this figure shows the level of the gender gap within each category. Stayers are defined as workers who are continuously employed, with changes in firm pay 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. *Source*: Own calculations, South African tax records, 2010–2018.

Figure C9: **Proportions of stayers, leavers, and entrants, by gender**



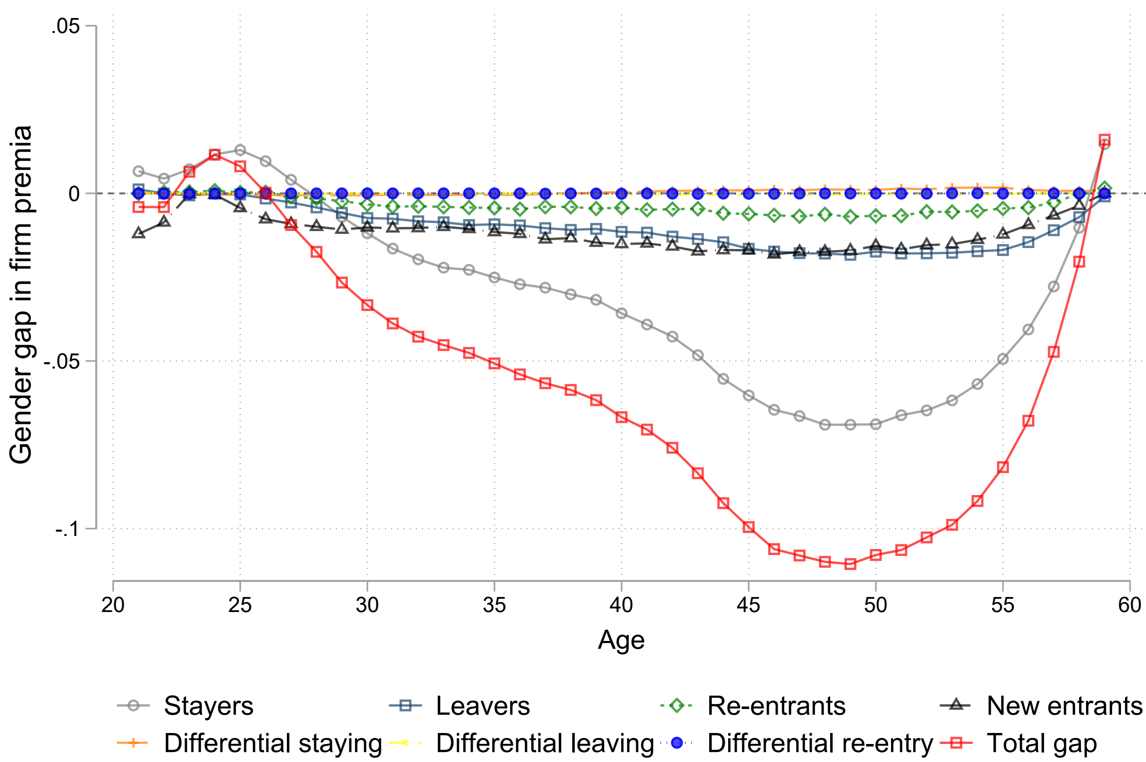
Notes. Each data point represents the proportion of female or male formal sector workers belonging to a particular subgroup (i.e., stayers, leavers, or entrants) at each age. Stayers are defined as workers who are continuously employed, with changes in firm pay premia arising from switches across firms. 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. *Source*: Own calculations, South African tax records, 2010–2018.

Figure C10: **Decomposing the firm-pay gender gap, by 4-year birth cohorts**



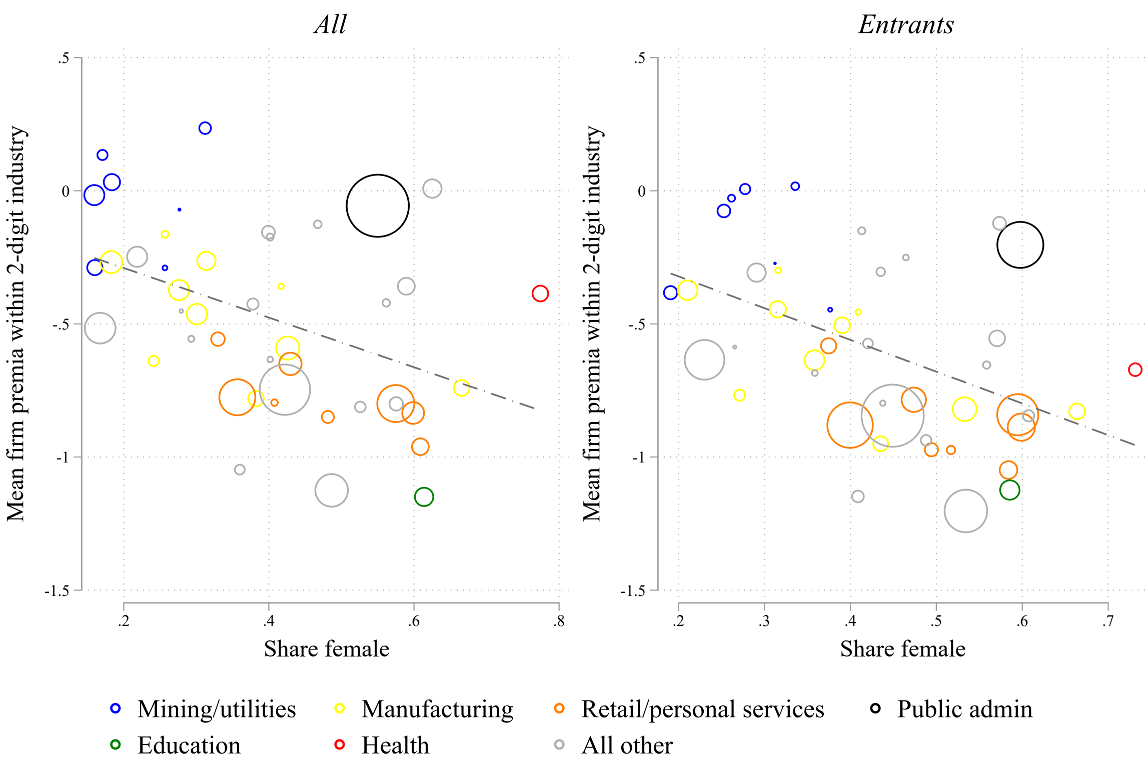
Notes. Each data point represents the gap in average firm premia between women and men working in the formal sector sample at each calendar year. Instead of including our controls for cohort differences (interactions of birth-year with female) we compute gender gaps by age separately for 4-year birth cohorts. Age range corresponding to calendar year range shown in parentheses for each cohort (e.g., ages 20-30 for cohort born between 1988 and 1992). *Source*: Own calculations, South African tax records, 2010–2018.

Figure C11: **Decomposing the firm-pay gender gap (new entrants vs. re-entrants)**



Notes. This figure is identical in construction and interpretation to Figure 2 in the main text, with the following difference: we split 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). As in Figure 2, the contribution of gender gaps in firm premia within groups are weighted by male proportions, and gender differences in proportions are weighted by group-specific average female firm premia relative to entrants. *Source*: Own calculations, South African tax records, 2010–2018.

Figure C12: **Average firm pay premia and female share, by 2-digit industries**



Notes. Scatterplot of mean firm premia within 2-digit industry against share of women in formal sector workforce of that industry. Marker size given by the share of each 2-digit industry in total employment. *Source*: Own calculations, South African tax records, 2010–2018.

Table C1: **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	32847995	42481100	27312497	35265020

Notes. The full sample includes all workers aged between 20 and 60 in our 2010-2018 tax records. The analysis sample includes only workers in firms with at least 10 workers of each gender in each period (see cleaning notes in Data Appendix for more details). *Source*: Own calculations, South African tax records, 2010–2018.

Table C2: **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
Person-year observations	27312497	35265020

Notes. This table shows means for predictors included in Table 1: age, commuting distance, firm value added, and whether covered by bargaining council agreement. Commuting distance is available only for a subset of the population (N=500,581). *Source*: Own calculations, South African tax records, 2010–2018.