SA-TIED

Southern Africa – Towards Inclusive Economic Development

The wage-setting power of firms

Rent-sharing and monopsony in South Africa

Ihsaan Bassier

SA-TIED Working Paper #59 | April 2019

About the programme

Southern Africa –Towards Inclusive Economic Development (SA-TIED)

SA-TIED is a unique collaboration between local and international research institutes and the government of South Africa. Its primary goal is to improve the interface between research and policy by producing cutting-edge research for inclusive growth and economic transformation in the southern African region. It is hoped that the SA-TIED programme will lead to greater institutional and individual capacities, improve database management and data analysis, and provide research outputs that assist in the formulation of evidence-based economic policy.

The collaboration is between the United Nations University World Institute for Development Economics Research (UNU-WIDER), the National Treasury of South Africa, the International Food Policy Research Institute (IFPRI), the Department of Monitoring, Planning, and Evaluation, the Department of Trade and Industry, South African Revenue Services, Trade and Industrial Policy Strategies, and other universities and institutes. It is funded by the National Treasury of South Africa, the Department of Trade and Industry of South Africa, the Delegation of the European Union to South Africa,IFPRI, and UNU-WIDER through the Institute's contributions from Finland, Sweden, and the United Kingdom to its research programme.

Corresponding author: ibassier@umass.edu

UNITED NATIONS
UNIVERSITY
UNU-WIDER

WIDER Working Paper 2019/34

The wage-setting power of firms

Rent-sharing and monopsony in South Africa

Ihsaan Bassier*

April 2019

United Nations University World Institute for Development Economics Research
wider.unu.edu

Abstract: Using administrative tax records from South Africa for the period 2011–14, I find that firm wage premia explain 25 per cent of the total wage variance, 60 per cent of the gender wage gap, and 40 per cent of the gap between workers in the middle and the bottom of the income distribution. Next, I argue that in contrast to the rent-sharing literature, many studies of monopsony fail to use firm-level wage variation. I address this by using the estimated firm wage premia to estimate how wages are related to rent-sharing and monopsony power. I find that the average worker switching from a firm in the 25th percentile to the 75th percentile in profitability is paid 32 per cent more; and that the same switch across the distribution of monopsony, as measured by the Herfindahl–Hirschman Index in industry-by-local labour market hires, decreases wages by 10 per cent on average. When monopsony is measured using a separations approach, I find a labour supply elasticity of 0.75, suggesting substantial wage-setting power. Finally, I provide additional evidence supporting the applicability of the monopsony model in explaining these results. Differences in labour supply elasticities across gender and income groups account remarkably well for the corresponding average gaps in firm wage premia, and unions substantially increase rent-sharing in firms.

Keywords: discrimination, firm effects, inequality, monopsony, rent-sharing
JEL classification: D33, J31, J42, J71

Acknowledgements: Special thanks to Arindrajit Dube, Aroop Chatterjee, Joshua Budlender, Murray Leibbrandt, and Adam Aboobaker for detailed comments. Data access was only possible through support from the South African National Treasury and UNU-WIDER SA-TIED inequality work stream, along with the invaluable help of Amina Ebrahim. Many thanks to the Southern Centre for Inequality Studies (SCIS) for financial and administrative support as well as comments.

---

*University of Massachusetts, Amherst, MA, USA; ibassier@umass.edu

This study has been prepared within the UNU-WIDER project on 'Southern Africa—Towards Inclusive Economic Development (SA-TIED)'.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-668-5

Typescript prepared by Gary Smith.

The 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, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

How important is firm wage-setting power in a developing economy? While the related literatures on firm wage premia, rent-sharing, and monopsony have proliferated,[1] few labour-surplus, low-infrastructure economies have been studied. Moreover, *none* of these studies have connected these three literatures under a coherent empirical framework, which I argue is a natural implication of the leading models of firm wage premia given by Manning (2003) and Card et al. (2018). I study the role of firms in determining wages through rent-sharing and monopsony, including their contribution to inequality in South Africa.

Using administrative tax data from South Africa for the period 2011–14, I begin by estimating firm wage premia based on Abowd et al. (1999) (henceforth AKM), and find they explain 25 per cent of the variance in workers' wages. When decomposed, the sorting of workers into firms explains *60 per cent*[2] of the gender wage gap and 40 per cent of the gap between workers in the middle compared to the bottom of the income distribution.

Next, I use the estimated firm wage premia to find the elasticities with regard to rent-sharing and monopsony. While this is best practice in the rent-sharing literature, this method is absent from the monopsony literature.[3] I show that rent-sharing and monopsony power are overestimated when using workers' wages instead of firm premia, but that the elasticities nevertheless suggest *substantial wage-setting power* of firms in South Africa. The labour supply elasticities, which indicate monopsony power, differ by gender and income group, and explain a large part of the average gap in firm wage premia between these groups. Finally, I give additional evidence that these results are described well by the monopsony model of Manning (2003) where firms compete monopsonistically over workers. This may come as a surprise in a labour-surplus economy and to South African policy makers.

The prominent role of firm wage premia is rapidly gaining recognition. Song et al. (2018) show for the United States that similar workers are paid very differently depending on the firm at which they work, and that sorting between firms has contributed substantially to the rise of inequality. The rent-sharing literature finds that firms with larger profits tend to pay workers more. The growing literature on monopsony suggests that firms are able to mark down wages considerably from marginal productivity, with a labour supply elasticity of 4 suggesting a markdown of 25 per cent. Yet this empirical literature is lacking in two respects which I address in this paper.

First, there are *very few studies on developing countries*. In their review of the rent-sharing literature, Card et al. (2018) list 21 studies, all of which are industrialized economies from Europe or North America. Similarly, in a meta-analysis of monopsony as indicated by labour supply elasticities, Sokolova and Sorensen (2018) find only three studies from developing countries between 1977 and 2018. Just *one* of these, Vick (2017), uses best-practice methods (as defined by the authors) and finds labour supply elasticities of between 1 and 2 in Brazil.[4] Indeed, the only other study of firm wage premia in a developing

[1] Key studies include Song et al. (2018), Card et al. (2018), and Dube et al. (2019)

[2] Subsequent to my last access to the data, I found evidence of outliers that may upwardly bias this estimate. An update will be released with any corrections after my next access to the data.

[3] Key studies using the spatial approach include Benmelech et al. (2018), Rinz (2018), and Azar et al. (2017), and those using the separations approach include Webber (2015), Hirsch et al. (2018), and Depew and Sørensen (2013). All of these use the worker's wage, rather that the firm wage premium.

[4] The other two are Fleisher and Wang (2004), who find the firm labour supply elasticity for China by simply regressing employment on wages, and Ogloblin and Brock (2005), who estimate the gap between productivity and wages by 'stochastic frontier estimation' for Russia.

country uses the same dataset from Brazil: Gerard et al. (2018) find that assortative matching accounts for two-thirds of the under-representation of non-whites in firms.

Developing countries are characterized partly by labour surplus and low infrastructure, relative to industrialized countries. Since the vast majority of workers live in developing economies, it is unclear what the implication is of the above empirical literature. Assume that this literature supports a monopsonistic model of the labour market. On the one hand, the policy implications of a monopsonistic economy are vastly different to those in a standard competitive economy, with an efficiency- and distribution-enhancing role for minimum wages, unions, and job search subsidies, among a range of policy levers (Naidu et al. 2019). On the other hand, *with little empirical evidence*, it is unclear how important we expect firm wage premia to be in developing economies. Costs of migration, poor transport infrastructure, and a limited supply of skills point towards greater search costs and thereby more firm wage dispersion. Yet a labour surplus is strongly at odds with an upwards-sloping labour supply curve, since in a monopsony model firms who want to expand can simply draw on the unemployed rather than compete for workers from other firms. I show that the labour supply elasticity is low, implying a markdown on wages larger than typically estimated for industrialized countries, and further that there *is* wide firm wage dispersion. The recruitment rate from unemployment is relatively large, suggesting that in a low-wage environment unemployed workers still need to be induced to work.

A second lacuna in the empirical literature on firm wage-setting power concerns the connection between the rent-sharing and monopsony literatures. Note that the best-practice empirical approach differs between the two literatures, in that estimates of monopsony from observational data fail to control for unobserved worker heterogeneity. This problem has long been recognized in the rent-sharing literature, where high-productivity workers tend to be in more profitable firms, biasing rent-sharing elasticities upwards. The best practice in the rent-sharing literature on observational data is to use the firm wage premia calculated from an AKM model, instead of workers' wages (Card et al. 2016). *The analogous problem exists for monopsony*. Regarding labour market concentration, low-productivity workers may tend to be in more concentrated labour markets, for example in less dense rural areas. Or, regarding monopsony, more highly paid workers may react more to co-worker effects, leading to a higher separations rate than otherwise and biasing the labour supply elasticity downwards. I contribute by arguing that the best practice in the monopsony literature should be updated, and show that using firm wage premia instead of wages leads to estimates of monopsony power that are about *half as large* for both the labour market concentration and separations approaches. The findings on separations draw on related research by Bassier et al. (2019), who show the importance of using firm-level wage variation for estimating labour supply responses using administrative data on hourly wages from the state of Oregon in the United States.

A further extension regarding the connection between rent-sharing and monopsony is that I evaluate these comparatively. How important is rent-sharing compared to labour market concentration? I show that the elasticity of firm wage premia with regard to rent is much larger in magnitude than the elasticity regarding labour market concentration, though both are substantial. Moreover, workers are strongly sorted by firm rent (workers who are highly paid *anywhere* are concentrated at high-rent firms), whereas sorting by labour market concentration is much less polarized. The result is that labour market concentration contributes much less to inequality. Differences in the estimated labour supply elasticities by gender account for *one-third* of the gender wage gap from sorting (one-fifth of the overall gender wage gap), and the differences in elasticities between the middle and bottom deciles of the income distribution account for *all* of the gap due to sorting (4 per cent of the overall gap). This supports the applicability of the monopsony model in explaining firm wage premia in South Africa.

In terms of the results, I find a rent-sharing elasticity of 0.13, which is high relative to the literature. I measure monopsony in two ways. Using labour market concentration as defined by the Herfindahl–Hirschman Index (HHI) concentration of the industry-by-municipality hires, I find an elasticity of -0.04.

A second, broader measure of monopsony uses separations from the firm as wages vary, and this gives a labour supply elasticity of 0.75. This does not rely on labour market concentration and includes monopsony power due to relocation or transport costs, firm-specific amenities such as co-worker relationships, and inattention to employment alternatives. The estimated elasticity suggests that workers are paid less than *half* of their marginal productivity due to monopsony power. These estimates are robust to a range of controls, alternative samples, and specifications. In particular, I find similar estimates when I restrict the estimated firm wage premia to the sample of switches due to firm closings, which is less likely to be endogenous to the worker–firm match or to co-worker effects.

The importance of firm wage premia, and the related rent-sharing and monopsony elasticities, may come as a surprise in South Africa. The considerable literature on the 'triple crises' of unemployment, poverty, and inequality has been analysed primarily through the lens of *skills shortages*, *returns to education*, and the policy tool of redistributive *social grants* (e.g. Banerjee et al. 2008; Leibbrandt et al. 2010, 2018). While these factors are relevant, roles of substantial wage-setting power of firms should be given more attention. Monopsony power results in depressed labour demand, low wages, and (given differential elasticities) inequality—which are alternative, and as suggested by the results in this paper, *large* contributors to the triple crisis. Policy tools include anti-concentration laws, spatial reallocation, transport subsidies, and increasing workers' bargaining power.

I end with additional evidence on the applicability of a monopsony model to the South African labour market. Workers who move to a different firm experience average real annual wage growth of 5.3 per cent compared to workers who stay at the same firm and experience average growth of 3 per cent, which supports the related job ladder model of job search. The separations specification is graphically well-motivated with a linear relationship between log separations and log firm wage premia, as predicted by the monopsony model. Another prediction that bears out is a positive relationship between firm size and wage premia. Finally, I find that rent-sharing is greater in firms with higher predicted union density, which fits with a model in which firms have bargaining space over the wage paid to workers.

The remainder of the paper is as follows. In the next section, I describe the data. Section 3 details the empirical strategy. Section 4 discusses the first-stage estimates of firm and worker effects, and Section 5 implements the second stage to estimate rent-sharing, labour market concentration monopsony, and separations monopsony. Section 6 provides additional evidence for the consistency of the evidence of a monopsonistic labour market and Section 7 concludes.

2 Data

I use four years of South African administrative tax data between 2011 and 2014, made available through a confidential data-sharing agreement with South Africa's National Treasury and UNU-WIDER. A collection of papers using this dataset appeared in a special issue of the *South African Journal of Economics* in the first quarter of 2018. In my cleaning decisions that follow, I try to address the major data challenges highlighted by these authors and draw on some of their solutions, where appropriate.

I combine two sources of data to form a matched employer–employee panel. The main source is records of individual job certificates submitted by firms on behalf of any employee earning over R2,000 per year (a low threshold of under US$150).[5] The second source of data is a firm-level panel based on

[5] Kerr (2018) uses this dataset in studying job churn over the period 2011–14, and Ebrahim et al. (2017) assess the impact of the Earnings Tax Incentive, a youth wage subsidy recently implemented in South Africa.

corporate income tax data,[6] which gives me the firm-level profits used for estimating rent-sharing elasticities.

Table B1 in the Appendix presents the main cleaning decisions in combining the two data sources. I begin with all job certificates, about 15 million per year. This drops to about 14 million per year when I restrict to workers aged 20–60 years. The matching between the firm-level and worker-level data relies on a correspondence table, which is imperfect and fails to match about one million observations. This may be a source of bias, and its cause is unclear. I convert job-level records to the worker level by selecting only the job for which 'fraction of year employed' is highest.[7] I restrict to private sector firms that have more than 20 workers, following Song et al. (2018). The justification for this is that the empirical strategy relies on estimating firm fixed effects, which may not be estimated well in small firms with idiosyncratic behaviour. Moreover, while cutting out most firms, panel C of Table B1 shows that this restriction maintains over 70 per cent of workers and about 85 per cent of total revenue.

The final dataset consists of 7–8 million workers in each year. Table B1 panel B shows that about 75 per cent of firms in this sample report profit (though fewer in 2014). As a secondary measure of profit, then, I construct gross profits by subtracting intermediate and labour costs from turnover. This variable is similarly populated, and gives similar results. The incompleteness of these variables is a concern, since it may introduce bias—but the direction is ambiguous. Loss-making firms may be drawn disproportionately, since losses are deductible from taxes on the following year's profits. On the other hand, firms that fail have no incentive to report losses or turnover. It is reassuring that the firms who *do* report cover disproportionately more workers (i.e. bigger firms report more)—for example eight million workers have associated firm profits in 2014 out of a total of 11 million workers (73 per cent).

The location, used for the labour concentration estimates, gives the municipality location of the firm, and is reported by establishment from the job certificates. A firm-level location is found by taking the mode of the employee-level location, with 99 per cent of establishment workers on average reporting the same location. While location information was not available for 2011–12, it is near-complete for 2013–14. Lastly, the industry variable is taken from the firm-level data, and supplemented with the industry variable from the job certificates where missing.

To standardize wages across part-time versus full-time workers, I annualize all incomes by multiplying the income by the inverse of the fraction of the year employed. All wages are adjusted for inflation and given in 2016 South African rand. Table 1 gives summary statistics of these workers: the median worker is paid about R66,000 per year, or R5,500 per month (about US$400 per month). The large gap between the 90th and 10th percentile of worker pay reflects the high inequality in South Africa.

How does this compare to survey data? Statistics South Africa, the national statistics agency, conducts several publicly available surveys. I compare to the Income and Expenditure Survey (IES) of 2010/11, which is aimed at providing accurate income data and has a sample of over 90,000 people, and the Quarterly Labour Force Survey (QLFS) for 2010–13, which has a smaller sample size but is conducted quarterly.[8] The IES records 12 million employed, and the QLFS shows 13–14 million employed per year between 2011 and 2014. This implies excellent coverage in the tax data, which record 10–11 million workers employed (see Table B1, panel A) and excludes informally employed workers who are counted

[6] For example, this firm-level panel is used by Kreuser and Newman (2018) in finding total factor productivity trends in manufacturing firms, and by Fedderke et al. (2018) in calculating concentration in the manufacturing product market.

[7] This decision rule follows Webber (2016), but I check its robustness by rerunning all regressions on the job level and for workers who only hold one job.

[8] I use version 3.2 of the Post-Apartheid Labour Market Series, which standardizes the QLFS (Kerr and Wittenberg 2017). Note that the annual tax year corresponds to the previous survey year. For the remainder of the paper where I compare to survey data, I use this dataset.

as employed in survey data. The IES records median and 90th percentile wages of about 80 per cent of those reported in Table 1, and the QLFS similarly records p10, p50, and p90 wages of 70–80 per cent of those reported in Table 1. This is likely due to a combination of measurement error and the distribution being shifted up with the exclusion of informal workers and small firms. Overall, it is reassuring that the employment and wages correspond roughly to survey data.

Table 1: Summary statistics of tax panel data

Tax year	Workers	Annualized Wage			Age	Firms	Firm size
	(freq.)	(p10)	(p50)	(p90)	(mean)	(freq.)	(median)
2011	7,233,802	17,627	66,175	335,165	36.2	44,830	40.4
2012	7,445,292	17,691	66,354	342,866	36.2	46,020	40.6
2013	7,740,862	18,303	65,993	342,522	36.0	47,407	41.2
2014	7,827,953	20,586	67,303	350,480	36.0	48,475	41.1

Notes: wages are annualized by multiplying reported wages by the inverse of the reported fraction of the year employed. The population is restricted to workers at private sector firms that have more than 20 workers.

Source: author's calculations based on South African tax records, 2011–14.

This is a large, rich dataset. There is no sampling error, since I observe records of the entire population. While measurement is still an issue, administrative data are typically much more reliable for wages than surveys, since surveyed wages are typically dependent on respondents remembering exact figures across many months (e.g. including bonuses) in the right definitions (e.g. net or before tax) and being willing to give up socially private information. However, *this dataset is not totally representative for all workers in South Africa*. Informal, unreported work such as domestic workers and informal traders are excluded, and these workers likely have poorer payment and conditions. The imperfect matching of records, firm size restrictions, and incompleteness of profit/turnover are unlikely to be random—though the direction of bias for this selection is unclear. Given all these caveats, a major consolation is that the data still reflect 7–8 million workers per year, which means I observe the actual incomes of most workers in South Africa for four years.

3 Empirical strategy

The two-stage estimation strategy follows Card et al. (2016), who estimate an AKM to find firm fixed effects on wages and regress these firm fixed effects on log value added per worker to find a rent-sharing elasticity. They do so by gender, which allows them to give evidence for differential bargaining power in explaining the gender wage gap. I build on their approach by using the firm fixed effects to estimate, in addition to rent-sharing elasticities, elasticities associated with monopsony power. Further, I disaggregate these rent-sharing and monopsony elasticities by gender and income decile group.

The AKM model imposes an additive structure of 'firm effects' (firm wage premia) and 'worker effects' as follows where the outcome is log annualized wages for individual i, firm j, and year t. I control for a quadratic and cubic in age, as well as year fixed effects:

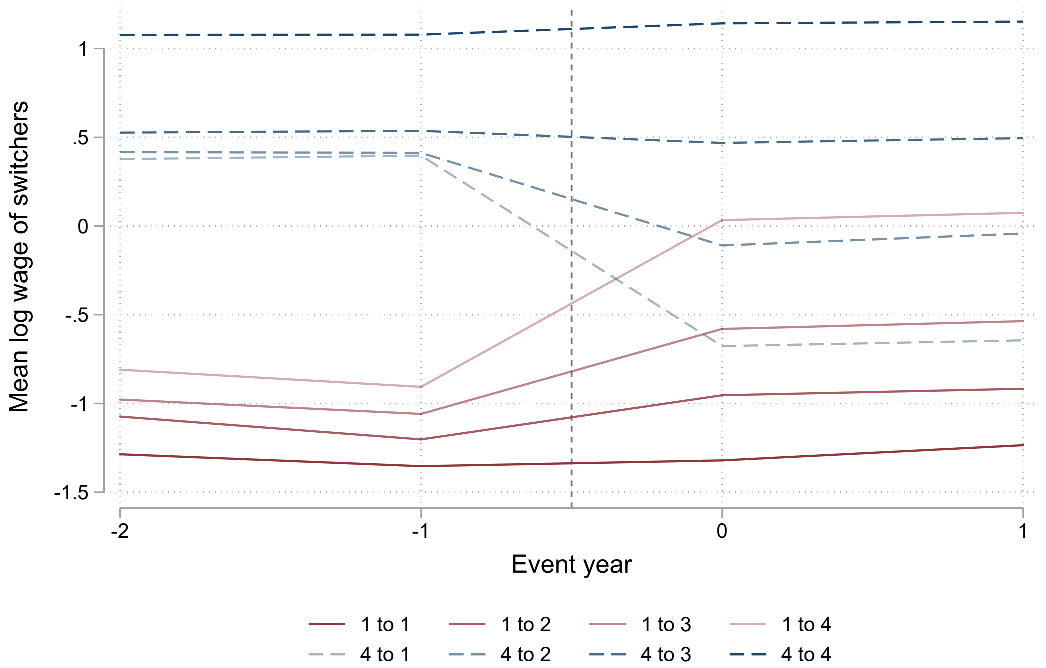
$$\ln(wage_{ijt}) = \alpha + \sum \gamma_i + \sum \delta_j + X_{ijt}\rho + \upsilon_{ijt} \tag{1}$$

Identification of the firm effects relies on movers, workers who switch between firms. Imagine a simple two-firm, two-period case. Hull (2018) shows that assuming parallel trends (the counterfactual wage growth of a mover is that of a stayer) and impersistence (the mover's wage at the new firm is the same as if she had always been there), then the firm-two effect is just a weighted average of wage gain experienced by movers *to* firm two and the wage loss experienced by movers *from* firm two. The additive structure of the AKM supposes that the wage gain and loss are equal.

$$\delta_2 = (E[\Delta Y_i \mid mover_{1,2}] - E[\Delta Y_i \mid stayer_{1,1]}) - (E[\Delta Y_i \mid mover_{2,1}] - E[\Delta Y_i \mid stayer_{2,2}])\omega$$

Figure 1 presents a graphical justification for this structure, following Card et al. (2016). The figure depicts the average wages of workers before and after switching firms, and without imposing any structure. Workers are classified by quartiles of the co-worker wage distribution—that is, leaving out the wage of the worker—to emphasize the firm wage. In the two years before the switch, wages for movers across the distribution are stable, which is consistent with the parallel trends assumption. The slight decrease for quartile 1 workers is concerning as an 'Ashenfelter Dip', that is workers may experience a negative event which registers as below-average wages in $T-1$ and a return to average wages in $T=0$, which then looks like a wage gain on switching. However, the quartile 1 decrease is very small, the subsequent increase is much larger, and the same pattern does not appear for origin quartile 2–4 workers. The stability of wages after the move supports the impersistence assumption. The *symmetric magnitudes* of wage changes for quartile i to j workers compares to quartile j to i workers supports the additive structure of the firm effects on wages.

Figure 1: Wage profiles of workers who switch firms



Notes: only movers are included. Wages of the *full sample* are residualized on year effects. The legend shows the origin quartile to destination quartile switch of the mover. Quartiles are calculated as the mean co-worker quartile in the firm, that is leaving own-wage out of mean firm wage. Event time 0 represents wages at the new firm.

Source: authors' creation based on South African tax records, 2011–14.

Notice the *large magnitude* of the change in wages associated with transitions to different firms, and that workers who are already paid more (less) tend to sort into higher (lower) paying firms. These hint at the results to come.

A second general check on the AKM structure is given in Figure B1 in the Appendix, which shows the mean residuals by deciles of firm and worker effect. The AKM structure does not seem to do too well at the bottom decile of the firm and worker effects distributions, where the residuals are larger. But otherwise the residuals are small (0–2 per cent of wages) in magnitude, especially in comparison to the magnitudes of the worker and firm effects.

The estimated firm effects γ_j are used in the second stage to estimate rent-sharing and monopsony elasticities. Assuming as above that individual wages can be decomposed into an *invariant* worker effect and a firm effect, any *firm-specific* effect regarding rents or monopsony power should reflect as differences in the firm component of the wage. To include the worker effect component is to invite spurious correlation, as has been shown in the rent-sharing literature. For example, more profitable firms tend to employ more workers with higher invariant worker effects. A regression of wages on firm profits yields an upwards-biased coefficient, because wages rise due to selection on worker effects *as well as* rent-sharing. Even individual controls such as for education and experience will fail to control for any number of unknown unobserved worker characteristics that raise wages.

All second-stage regressions are on the firm level, weighted by firm size. The rent-sharing specification is as follows, where profit is measured per employee:

$$\delta_j = \alpha + \varepsilon_{rent} \ln(profit_j) + \nu_j \tag{2}$$

I estimate monopsony effects in two different ways. The first specification follows a labour market concentration approach, and most closely resembles the specification of Azar et al. (2017). The labour concentration approach can theoretically be motivated by a Cournot model in which a small number of firms pay wages below productivity disequilibrium, perhaps maintained by local wage norms, along with large transport and psychological costs to moving to less concentrated areas. I calculate HHI on hires by industry (17 levels) and municipality (about 220 in South Africa), controlling for industry and municipality fixed effects. This allays concerns of different costs of living associated with municipalities (such as rural versus urban) and spurious correlations due to industry clustering (such as concentrated, capital-intensive industries). The identifying variation then comes from, for example, comparing firm effects across a firm with high HHI to a firm with low HHI in the same industry but different municipality, netting out the average firm effect from each municipality:

$$\delta_j = \alpha + \varepsilon_{HHI} ln(HHI_{im}) + \sum industry_i + \sum municipality_m + \nu_j \tag{3}$$

A second approach to estimating monopsony follows Manning (2003), who focuses on how wages vary with separations from the firm (see Sokolova and Sorensen (2018) for a review). The advantage of this approach is that it captures broad aspects of monopsony power; another advantage is that it offers a broad structural model which, I argue later, explains the patterns of firm wage premia from this paper remarkably well. I provide an outline of the related structural model in the Appendix. Note that over the course of this paper, I refer to the following *econometric approach* to finding ε_{LS} as the 'separations' approach and I refer to the related *structural model* based on search theory outlined in the Appendix as the 'dynamic monopsony' model. The intuition is that in a perfectly competitive market, workers should leave low-paying firms immediately for high-paying firms. A higher elasticity of separations from a firm with regard to wages indicates a more competitive market, and less monopsony power. By estimating similar regressions for the probability of separations to employment (ε_{EEsep}), to non-employment (ε_{NEsep}), and the recruitment elasticity from employment ($\varepsilon_{EErecruits}$), weighted by the proportion of hires from employment (ϑ), the direct estimate of the labour supply elasticity to the firm follows (Bachmann et al. 2018). Once again, given that *worker effects are invariant* to the firm a worker is at, a worker is responding to the *firm* effect when making the decision to separate:

$$\ln(Separation) = \alpha + \varepsilon_{sep}\gamma_j + \nu_j \tag{4}$$

$$\varepsilon_{LS} = -(1+\theta)\varepsilon_{EEsep} - (1-\theta)\varepsilon_{NEsep} - \varepsilon_{EErecruits} \tag{5}$$

This yields the classic upwards-sloping labour supply elasticity associated with monopsony in Robinson (1933), without recourse to concentrated labour markets. In a simple model in which firm profits are given by $\pi = pQ(K,L) - wL(w) - rK$, that is a separable production function with an upwards-sloping labour supply curve, then the first-order condition for maximization of profits yields the markdown equation, $\frac{1}{\varepsilon_{LS}} = \frac{pQ'(L)-w}{w}$.

Using the firm component of wages to estimate rent-sharing and monopsony elasticities assumes that firms treat workers homogeneously. I relax this assumption by estimating firm effects separately by gender and income decile group (workers allocated in the base year, with groups defined as deciles 1–4, 5–8, and 10). The rent-sharing and labour market concentration specifications show whether firms discriminate between these groups on average. The labour supply elasticities by group give a broader picture of differential monopsony power as an explanation for the corresponding wage gaps, as done elsewhere regarding gender discrimination (Barth and Dale-Olsen 2009; Ransom and Oaxaca 2005; Vick 2017).

I make numerous robustness checks. For the first-stage AKM, I alternatively restrict to workers who are observed as employed every year (to avoid contamination of firm effects through an unemployment wage penalty or skills training gap), to workers who are recorded as workers for the full year (to avoid spuriously low firm effects based on part-time work), to workers who only held one job at a time (an alternative to the job-to-worker data-cleaning decision), and to the dataset of jobs rather than workers. For the second-stage elasticities, I restrict to the set of firms observed in every year (to avoid selection on successful firms), and to firms in the largest connected network (to ensure comparability in estimated firm effects). A concern is that separations are driven by poor firm matches which would inflate firm wage premia. Beyond the evidence I give in Figures 1 and B1, I identify firm closings and mass layoffs which are much less likely to have endogenous separation decisions, and estimate the firm effects and second-stage elasticities off workers who separate from these firms. The elasticities are remarkably consistent across these variations.

Finally, I use alternative specifications to the firm effects approach. For rent-sharing, I use a first-differences specification. For the labour market concentration monopsony, I use the leave-one-out HHI as in Rinz (2018). And for the separations estimates of monopsony, I use the residuals of profits from time and industry fixed effects to identify shocks to profits. Since the rent-sharing results show pass-through to wages, I leverage changes in profit residuals as an instrument for the change in wages, and use this as the explanatory variable for changes in separations for the firm.

4 First stage: firm and worker effects decomposition

There is substantial sorting of workers who would be paid highly *anywhere* into firms that pay highly to *anyone*, even relative to comparable evidence in other economies. Moreover, this sorting accounts for a large proportion of the wage gap between male and female workers, and the middle compared to the bottom of the income distribution.

Table 2 summarizes the results of the first-stage regression. Song et al. (2018) find that firm effects explain 9 per cent of the male wage variance in the United States for 2007–13, and Card et al. (2016) find they explain 20 per cent of the male variance in Portugal over the period 2002–09. I find that in South Africa, firm fixed effects explain 25 per cent of the total variance. The ‘segregation index’, which indicates sorting by dividing between-firm variance by total variance, is 0.46, compared to 0.25 as reported by Song et al. (2018) in the United States for 2007–13.

Table 2: Variance decomposition of firm and worker effects

Person-year obs.	26,200,000		*Between-firm*	Comp.	Share (%)
WFE (freq)	8,144,556		Var (m_lnWage)	0.84	55
FFE (freq)	59,360		Var (m_WFE)	0.59	39
Adj R^2	0.85		Var (FFE)	0.52	34
	Comp	Share (%)	2*Cov(m_WFE,FFE)	–0.20	–13
Var(lnWage)	1.53		m_Xb terms	–0.07	–4
Var(WFE)	1.09	71	*Within-firm*		
Var(FFE)	0.39	25	Var(diff_lnWage)	0.70	46
Var(Res)	0.15	10	Var(diff_WFE)	0.70	46
2*Cov(WFE,FFE)	0.07	5	Var(Res)	0.13	9
Xb terms	–0.17	–11	2*Cov(diff_WFE,Res)	0.00	0
			diff_Xb terms	–0.14	–9
Cov(lnWage, FFE)	0.41	27	*Segregation index*	0.46	

Notes: These are summary statistics from an AKM, where Xb terms control for year and age (quadratic and cubic). Workers are limited to those at firms in the private sector with more than 20 employees.
Abbreviations: WFE (worker fixed effect), FFE (firm fixed effect), Res (residuals), m_lnWage (mean wage in the firm), m_WFE (mean WFE in the firm), diff_lnWage (lnWage – m_lnWage), diff_WFE (WFE – m_WFE). The segregation index is calculated as var(m_WFE) divided by var(WFE).

Source: author's calculations, based on South African tax records, 2011–14.

Figure 2 plots the distribution of firm and worker effects by decile, as a visualization of the decompositions in Table 3. Workers who would be highly paid anywhere, that is those with a high worker component in wages, are disproportionately located in firms who pay highly to everyone, that is have high firm effects. As firm effects decrease, the composition of worker effects correspondingly reverses so that low-worker-effect workers disproportionately worker at low-paying firms. This demonstrates the inequality *increasing* effect of firm effects: if workers with low and high worker effects were distributed differently across the firm effects, then large firm effects could be inequality *neutral* or even inequality reducing. Note that *some* workers with low worker effects are still in high-paying firms, which is a key requirement of the job ladder model and demonstrates the existence of 'good' and 'bad' jobs available to the same worker.

Figure 2: Distribution of worker and firm fixed effects

Notes: worker and firm effects are results from the AKM regression. Deciles of worker effects are plotted in increasing order by decile of firm effect (dark red is the lowest worker effect, dark blue is the highest). Workers are limited to those at firms in the private sector with more than 20 employees.

Source: author's creation, based on South African tax records, 2011–14.

Table 3 decomposes *average* wages into average worker and firm effects, by group. The average gap between the middle deciles (5–8) and the bottom deciles (1–4) is 45 per cent.[9] There is close to *no difference in worker effects* on average between these groups. In contrast, average differences in firm effects between the groups explains 40 per cent of the wage gap. It seems that invariant worker characteristics such as *ability, experience, and skill are roughly irrelevant* for a worker's location in 80 per cent of the income distribution; rather, a worker's wage turns on whether they find a good or bad job. Further, using the rent-sharing results of the following section, sorting into high-*profit* firms explains 15 per cent of the average difference in firm effects, and I suggest in Section 6 that this supports unionization as an explanation for the large role of good and bad jobs. The wage gap between the top and middle of the income distribution follows a more conventional pattern, with differences in worker effects explaining nearly half of the gap and differences in firm effects explaining less than one-tenth.

Table 3: Mean decomposition of firm and worker effects by gender and decile

Panel A: gender	Men	Women	Gap
Percentage of workers	61%	39%	
Wage	11.27	10.95	0.32
Worker FE	0.052	–0.082	42%
Firm FE	0.073	–0.116	59%
Predicted rent effect	0.042	0.035	2%
Predicted HHI effect	–0.002	–0.002	0%

Panel B: deciles	Dec. 1–4	Dec. 5–8	Dec. 10	Gap (mid–low)	Gap (top–mid)
Percentage of workers	27%	30%	8%		
Wage	10.68	11.12	13.14	0.45	2.02
Worker FE	–0.206	–0.228	1.665	–5%	46%
Firm FE	–0.145	0.031	0.343	40%	8%
Predicted rent effect	–0.028	0.040	0.235	15%	5%
Predicted HHI effect	–0.004	–0.001	0.001	1%	0%

Notes: these are summary statistics of the AKM model. Deciles refer to the 2011 wage distribution and are conditional on employment in 2011. Predicted rent and HHI use the elasticities shown in the main estimates (see Table 4). Workers are limited to those at firms in the private sector with more than 20 employees.

Source: author's calculations, based on South African tax records, 2011–14.

A literature has developed regarding the role of firm wage premia in explaining discrimination. Discrimination is defined by systematically lower wages for equally productive workers of one group compared to another, where these groups are typically defined by race or gender. While in the past legal institutions such as differential rights and job reservations ensured that blacks were disadvantaged compared to whites, or women compared to men, there is a question over why the raw wage gaps (difference in unconditional average wages) have persisted despite formal equality and protections such as equal pay laws.

Pre-market conditions that contribute to lower productivity, that is worse-quality institutions such as schools that are legacies of historical disadvantage, explain a substantial part of the raw wage gap. However, while decreasing slowly over time, the empirical evidence of a wage gap *conditional* on covariates suggests much more is at play. There are three prominent models of such discrimination (Oaxaca 2007). Becker's taste-based model assumes prejudice (by employers, co-workers, or consumers) and statistical discrimination models assume imperfect signals of productivity. Both of these models can be unsatisfactory, the first because it implies that the wage gap should quickly dissipate under perfect competition, and the second because it also has to assume that employers do not learn about workers' productivity over time (Darity and Mason 1998).

[9] This gap exists nearly by construction, since the groups are defined on the base period income distribution. It is the decomposition that concerns us here.

A third possibility is that firms have wage-setting power, which results in discrimination. One mechanism relates to the sorting component: if one group has a lower labour supply elasticity, then average wages would be lower even conditional on productivity. Manning (2003: ch. 7) for the UK, Ransom and Oaxaca (2005) for the United States, and Gerard et al. (2018) for Brazil, among many others in this literature, find a substantial role of firm wage premia in explaining the wage gap.

The gender wage gap in this sample for South Africa is 32 per cent. About 40 per cent of this gap is explained by differences in average worker effects, which includes both average differences in invariant characteristics that increase marginal productivity to the firm (such as skill and experience) as well as average economy-wide discrimination against women (such that women are underpaid no matter which firm they work at). Nearly *60 per cent* of the *unconditional* gender wage gap is explained through differences in sorting. Figure B2 in the Appendix demonstrates this visually. The counterfactual of an even distribution would be a density of 0.5 for each worker–firm decile bar. While there is sorting *within* each gender of workers with high worker effects into high-paying firms, *women* with high worker effects constitute a much smaller portion of high-paying firms than *men* with high worker effects. The reverse is true for sorting into low firm effects.

For comparison, the unconditional gender wage gap for the same period using the QLFS is 34 per cent, which is reassuringly close to the tax data. When decomposed, education and age explain roughly *none* of the difference (-4 *per cent* of the gap) compared to the coefficients. This is consistent with the large firm effects estimated above. Note that differences in average worker effects in the tax data also capture unobserved skill, which is absent from this decomposition using survey data.

These results present strong evidence that substantial firm effects exist. Workers who switch firms experience large increases in wages, while maintaining invariant characteristics such as race, sex, age, and, plausibly, skill. What explains these effects? In the following section, I link these firm effects to rent-sharing and monopsony power.

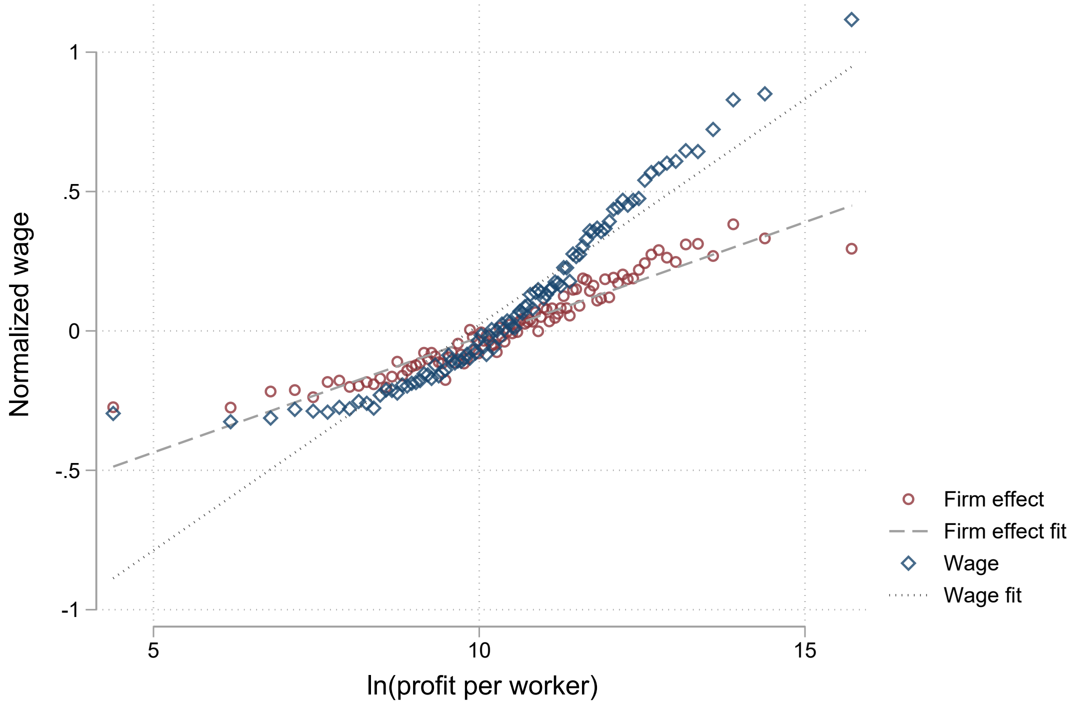
5 Second stage: rent-sharing and monopsony

Firm wage premia *increase strongly with profits*. The associated rent-sharing elasticity is high relative to other industrialized countries, and it explains 20 per cent of the total variance in firm wage premia. Monopsony as measured by labour market concentration decreases workers' wages, but the elasticity is smaller and explains little of the firm wage dispersion. Finally, monopsony as measured broadly by firm labour supply elasticities suggests *substantial wage-setting power* of firms and, as I argue in Section 6, this goes a long way in explaining the results of this paper.

5.1 Rent-sharing

Figure 3 illustrates the association between firm profits per worker and wages (both in log terms). The association for wages is much steeper than for the firm component of wages (firm effects). The upper panel of Figure B4 in the Appendix visualizes the reason for this, in that workers with high wage worker effects tend to be in firms with high firm effects, which upwardly biases the rent-sharing coefficient as argued in Section 3. There is a clear and tight linear relationship between firm effects on wages and the quantiles of firm profitability. Firm profitability ranges from about R4,000 per worker per year in the 10th percentile to over R250,000 in the 90th percentile, with a median profit of about R35,000. For comparison, recall from Table 1 that the median wage is about R65,000 per year.

Figure 3: Rent-sharing in firms: wages and profits



Notes: profit is classified into 100 quantiles, and each dot in the scatterplot gives the mean ln(*wage*) of workers for a quantile of profit. Only one observation per firm is given. Observed wages refer to directly recorded wages for each worker, and are centred around 0 for plotting. Firm fixed effects are the results from the AKM models above. Workers are limited to those at firms in the private sector with more than 20 employees.

Source: author's creation, based on South African tax records, 2011–14.

Table 4 presents estimates of the rent-sharing elasticity, all with p-values below 0.001. Column 1 shows an elasticity of 0.13 using the full sample with firm effects, compared to the range in the literature of 0.05–0.15 reviewed by Card et al. (2018). This implies that moving from a firm at the 25th percentile of firm profits to the 75th percentile is associated with a 32 per cent increase in annualized wages. Column 2 uses the wage instead of firm effects and, as illustrated in Figure 2, the coefficient is substantially (and spuriously) higher. Column 3 restricts the sample to firm closings using firm effects, finding a very similar coefficient to the full sample. Column 4 presents a first-differences specification, which shows a much smaller elasticity. Note that a first-differences approach leverage change in wages of *stayers between years*, and so will fail to pick up longer-term rent-sharing effects, particularly if norms in wage-setting are important.

Tables B4 and B5 in the Appendix present robustness checks, showing that the estimates are stable across winsorizing, controls for worker effects or industry and municipality fixed effects, and a range of alternate samples. An outlier is in column 2 of Table B4, where the coefficient decreases substantially when controlling for assets. This implies that more capital-intensive firms are more profitable and pay more, but conceptually this may be a 'bad control': I would want to include this variation in profits since this ultimately still results in higher wages for workers who switch. The other outlier is the coefficient on firms connected in a network, with a higher coefficient of 0.2, though this is on a smaller population of about 7,500 firms. Using instead the firm effects on the bonus part of the wage recorded, the rent-sharing elasticity is even higher at about 0.5. The R^2 in these regressions is about 0.2, suggesting that the rent-sharing association explains about 20 per cent of the total variance in firm effects.

Table 4: Main estimates of rent-sharing and monopsony

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: rent-sharing						
Profit	0.132	0.208	0.135	0.027		
	(0.021)	(0.021)	(0.025)	(0.006)		
Panel B: monopsony						
HHI	–0.041	–0.086	–0.038		–0.030	
	(0.017)	(0.028)	(0.018)		(0.017)	
Labour supply ε	0.75	0.238	0.722			3.25
	(0.062)	(0.055)	(0.057)			(1.07)
Firm FE	Y					
Wage		Y				
Firm closings			Y			
First diff.				Y		
Leave-one-out					Y	
Profit IV						Y

Notes: all variables are measured in ln units. Monopsony HHI refers to the sum of squared firm share of industry-by-municipality hires (industries at the 17-code level). See text for specification details. Alternative specifications and samples are given in Tables A4–6. Workers are limited to those at firms in the private sector with more than 20 employees. Standard errors are given in parentheses. Source: author's calculations, based on South African tax records, 2011–14.

5.2 Labour concentration monopsony

I measure monopsony power in two ways. I begin with the labour market concentration approach. Table B2 presents the summary statistics on the concentration of employment and of hires. The average industry by municipality HHI in non-metros is substantially higher than in metros, although metros have surprisingly high HHI too. How should the HHI be interpreted? The inverse of the HHI gives the number of equally sized firms that would produce the same HHI. For example, the cities of Tshwane and Johannesburg, which make up the employment hub of South Africa, have an average industry HHI of about 300, which is equivalent to only 30 equally sized firms in each industry. HHI is similar when using the concentration of workers rather than hires, and (as expected) lower when defining labour markets only by municipalities. The top-eight share emphasizes the high concentration in South African labour markets: in the Johannesburg–Tshwane hub, the top eight firms hire about one-third of workers, and in the non-metros about 90 per cent of workers.

As argued in Section 3, a similar bias exists when using wages instead of firm effects as the outcome. The upper panel of Figure B3 demonstrates this, as does the lower panel of Figure B4—although the sorting appears to be less than for profit. Using the firm effects, the estimate given in column 1 of Table 4 is -0.043 (p-value less than 0.05). For a worker switching from the 25th to the 75th percentile in HHI, her wages would decrease by 10 per cent. As a concrete example, a worker switching from the metros of Ekurhuleni (East Rand of Johannesburg) to Mbombela (Nelspruit, a more rural city) would take a decrease in wages of 5 per cent. Using the firm closings (column 3), the estimate is lower, and using the leave-one-out HHI (column 5), the estimate decreases even further to -0.03. The coefficient is much higher using the top-eight share and is relatively stable across most alternative samples—except when restricting to the *network* of firms and firm closings or full-time workers, when the coefficient is insignificant (see Tables B3 and B4).

As a comparison, in a study of job postings online in the United States, Azar et al. (2017) (henceforth AMS) similarly define HHI as the hires concentration in an industry by commuter zone, but use the outcome of wages and find economically large effects. My results show that this is likely upwardly biased. For example, column 2 of Table 3 shows an elasticity that is twice as large in magnitude using wages. Column 4 of Table B3 shows a similar estimate when running the comparable specification of AMS which also utilizes cross-time variation (with time fixed effects). When I run the same AMS specification with controls for firm labour productivity and firm age, the elasticity reduces to -0.046, which is similar to my preferred firm effects estimate and validates my argument that the upward bias is driven by a lack of appropriate controls.

Overall, while labour market concentration does appear to decrease wages, the impact seems to be relatively small. In a study covering the US manufacturing sector over three decades, Benmelech et al. (2018) find similarly significant but economically small effects on wages from labour market concentration. Labour market concentration also does not seem to contribute substantially to inequality, explaining a negligible amount of the firm wage dispersion ($R^2 < 0.05$), explaining very little of the between-group wage gaps (see Table 3), and sorting worker effects along HHI in a much less polarized way than along profits (see the lower panel of Figure B4).

5.3 Separations-based monopsony

Next, consider the separations approach to monopsony. This is a broad measure of monopsony, as argued in Section 3, and gives a structural framework for the wage-setting power of firms. The estimates of firm labour supply elasticities ε_{LS} follows Equation 5 and all have relatively narrow confidence intervals. Note that a large ε_{LS} suggests a *more* competitive labour market and less wage-setting power. Table 4 shows a ε_{LS} of 0.75 using firm effects, and this is similar using firm closings as well as a variety of alternative controls and samples (Tables B3 and B4). This implies that wages are less than half of the marginal product of labour, using the markdown equation in Section 3. These estimates imply substantially more wage-setting power than the literature on industrialized economies, which the meta-analysis of Sokolova and Sorensen (2018) reviews as giving a median ε_{LS} of 2.7.

Column 2 shows a much lower ε_{LS} of 0.24 using the full wage instead of the isolated firm component of wages as above, (spuriously) implying substantially more monopsony power. The lower panel of Figure B3 illustrates this, following the pattern of coefficients for rent-sharing and labour market concentration. Column 6 of Table 3 uses the profit IV explained in Section 3, and finds a much higher ε_{LS} of 3.25. If this is a more persuasive estimate, then firms still have substantial wage-setting power more in line with the literature (implying a markdown of 30 per cent). However, the instrument may be biased by a failed exclusion restriction, since a decrease in profits may prompt separations through a decrease in wages (the desired channel) and *also* layoffs of workers.

Although not reported in Table 4, the $\varepsilon_{LS} = 0.75$ is decomposed into $\varepsilon_{EEsep} = -0.38$ and $\varepsilon_{NEsep} = -0.55$ following Equation 5. In particular, the relatively high elasticity of separations to non-employment suggests that workers are paid low wages on the margin of the reservation wage, in line with qualitative work in South Africa (Zizzamia 2018). It also suggests that off-the-job search is relatively easier than searching while employed.

5.4 Discrimination in rent-sharing and monopsony

How does rent-sharing and labour market concentration monopsony differ across groups? Table 5 shows the results of the primary two-stage specification given in column 1 of Table 4, run separately by gender and by income decile group. For the most part, the differences are not statistically distinguishable (rent-sharing by gender, rent-sharing by middle compared to lower decile group, and all of the labour market concentration comparisons). The assumption of homogeneous firm effects does not seem to be heroic, at least as suggested by these group decompositions. In contrast, differences in ε_{LS} by gender and income decile group suggest that women and lower income deciles face substantially more broadly defined monopsony power, an important result for the next section. This could be due to lower mobility, for example because social conditions force women to prioritize stability, or due to lower search, for example because lower-income workers have less information and are part of less lucrative employment networks or because women have a greater burden of household labour, leaving less time for job search.

Table 5: Distributional estimates of rent-sharing and monopsony

	Gender		Income decile group		
	Female	Male	Dec. 1–4	Dec. 5–8	Dec. 10
Panel A: rent-sharing					
Profit	0.125	0.110	0.105	0.093	0.065
	(0.014)	(0.010)	(0.016)	(0.007)	(0.009)
Panel B: monopsony					
HHI	–0.062	–0.035	–0.017	–0.027	–0.017
	(0.032)	(0.010)	(0.012)	(0.010)	(0.026)
Labour supply ε	0.677	0.757	0.585	0.778	0.807
	(0.052)	(0.06)	(0.042)	(0.058)	(0.091)

Notes: all variables are measured in ln units. The firm effects specification (column 1 of Table 4) is used for all cells. Monopsony HHI refers to the sum of squared firm share of industry-by-municipality hires (industries at the 17-code level). Decile classification refers to decile of the worker in 2011—that is, it is conditional on employment in 2011. Standard errors are given in parentheses. Full specifications are given in text. Workers are limited to those at firms in the private sector with more than 20 employees.

Source: author's calculations, based on South African tax records, 2011–14.

Finally, each of the elasticities in this section can be disaggregated at the industry level, reported in Table B5 in the Appendix. Rent-sharing is stable across industries, though slightly higher in more capital-intensive industries, as discussed earlier (such as mining and electricity-generation) and lower in information and communication services. Monopsony power as indicated by labour market concentration is higher in information and communication services, and finance and insurance, but insignificant in the mining sector, which has a history in South Africa of drawing employment from far regions through migrant labour (Wilson 2001). The separations elasticity ε_{sep} is particularly low for wholesale and retail, which suggests large monopsony power in one of the largest employment sectors in South Africa. ε_{sep} is high in mining, which is a well-unionized sector and hints at the role of bargaining, argued in the next section.

6 Consistency with a monopsony model of the labour market

These results are remarkably consistent with monopsony.

Under competitive conditions, it is difficult to explain why a distribution of wages may be available to the same worker, as argued in this paper. All workers should simply go to the highest-paying firm; and from the firm's perspective, firms should pay workers' next best alternative, which in equilibrium is the same wage. Instead, two prominent models explaining the growing evidence of firm wage premia in the labour market are given by Manning (2003), based on search frictions, and Card et al. (2018), based on firm amenities, where both models rely on an upwards-sloping labour supply curve.

The estimated labour supply elasticities in Tables 4 and 5 support key predictions of these two monopsony models. Most directly, the estimated labour supply elasticity is low, whereas in a competitive market it should be high (infinite, at the limit); moreover, the lower panel of Figure B3 suggests a strongly linear relationship between log separations and log firm effects, as predicted by the monopsony model and which forms the basis of the estimated upwards-sloping labour supply curve. Second, using the markdown equations, the different labour supply elasticities in Table 5 imply a gender wage gap of

6.5 per cent, which is one-third of the observed firm effects gender wage gap; the elasticities also imply a gap between the middle and lower deciles of 17 per cent, which is *all* of the gap due to average firm effects. [10] Third, both models predict rent-sharing, since if more profitable firms are more productive, they have a higher marginal product of labour which implies higher marginal cost of labour—that is, higher wages in the upwards-sloping labour supply curve.[11] Fourth, there is a firm size wage premium as predicted in these models. The results of a specification of firm effect on the log of firm size is presented in the final column of Table 6.

A competing explanation for estimated firm wage premia is a compensating differentials model, which allows for a competitive market with differing wages based on the general amenities offered by the firm. In this case, workers who move firms should not gain wages on average, since firm wage premia are invariant to worker utility. In contrast, Table 6 shows that workers who switch firms (employment-to-employment separations) tend to experience much more wage growth than workers who stay at firms.[12] This is consistent with job search, where workers switch jobs when they find a better offer. This is not to say that compensating differentials have no role, just that this explanation is insufficient.

Table 6: Individual wage growth by separation status

	Workers (freq)	Real wage growth (p50)			
		All	Dec. 1–4	Dec. 5–8	Dec. 10
Stayer	2,700,000	3.0%	3.1%	3.4%	2.3%
E-E separation	1,287,472	5.3%	6.2%	5.6%	3.0%
N-E separation	3,864,942	1.7%	1.3%	1.6%	2.1%

Notes: workers who enter after 2011 are excluded. Stayers are workers who remain at the same firm over 2011–14; E-E separations (employment-to-employment) are workers who remain employed, but at some point change firms; N-E separations (non-employment) are workers who separate into unemployment, where absence of a job certificate is classified as unemployment. Deciles refer to the distribution in 2011. Workers are limited to those at firms in the private sector with more than 20 employees.

Source: author's calculations, based on South African tax records, 2011–14.

Monopsony models of the labour market imply wage-setting power of firms. With relatively strong union institutions in South Africa, we would expect unions to increase bargaining power and thereby wages when there is a gap between the marginal productivity of labour to the firm and the lowest wage accepted by the worker, as in monopsony models. Since unionization is not observed in the tax data, I use the QLFS survey data to find unionization rates by municipality and merge this into the tax data as a proxy for firm union density. Table 7 shows a strong role for unions. Interacting unionization with profit in the specification from Equation 2, I find a large positive elasticity of unionization on firm effects (the level coefficient) and a large positive interaction term on rent-sharing (though both effects are only marginally significant, which is unsurprising given the measurement error due to the proxy). Similar effects are not significant for labour market concentration or labour supply elasticity. If workers are gaining *more* from rent-sharing due to unions, then they are off their supply curves and we do not expect

[10] From $\frac{1}{\varepsilon_{LS}} = \frac{pQ'(L)-w}{w}$, $\ln(w) = \ln(\frac{\varepsilon_{LS}pQ'(L)}{1+\varepsilon_{LS}})$ and the difference in predicted wages from the estimated elasticities gives the explained wage gap.

[11] It would be useful to relate the ε_{rent} and ε_{LS} through calibration. However, this calibration is sensitive to parameter values—see the appendix.

[12] There is a puzzle here for the literature on South African wages: Table 6 suggests that the median real growth rate in wages is high–a weighted average of 2.7 per cent per year. Yet Table 1 shows that median wages were stagnant over 2011–14, as in previous decades in South Africa, growing at about 0.4 per cent per annum. What explains this difference? The proportion of hires from non-employment is extremely high, about 64 per cent, and there is a wage penalty associated with unemployment. Thus, the 0.4 per cent growth rate in the median can be decomposed into a large positive growth rate for those who stay in employment, and a negative growth rate for those who do not (this also includes workers entering the labour force at lower wages than workers who leave, as expected from life cycles).

these interaction terms to be significant. Column 4 shows that more unionized firms tend to be more profitable. Returning to the large role of 'good' versus 'bad' jobs in accounting for the gap between the middle and lower income decile groups, if those in the middle are more unionized, these workers are benefiting from both higher rent-sharing elasticities and sorting into more profitable firms.

Table 7: Firm effects, unionization, and firm size

	(1)	(2)	(3)	(4)	(5)
Profit	0.145				
	(0.028)				
HHI		–0.041			
		(0.016)			
Separations			–0.474		
			(0.037)		
Union	0.237	–0.022	1.13	0.92	
	(0.136)	(0.062)	(0.155)	(0.046)	
Interact	0.173	0.054	–0.104		
	(0.104)	(0.069)	(0.153)		
Size					0.013
					(0.003)
Obs.	38,789	47,504	52,349	156,119	59,653
Union Interact	Y	Y	Y		
Y=Firm FE	Y	Y			Y
Y=Profit				Y	

Notes: all variables are measured in ln units, except for union which is a density. Unionization rate is predicted from the QLFS based on municipalities. See text for primary specifications. Standard errors are given in parentheses. Workers are limited to those at firms in the private sector with more than 20 employees.

Source: author's calculations, based on South African tax records, 2011–14.

It deserves mention that there are other models of the labour market which are consistent with the evidence in this paper. Bowles et al. (2001) briefly discuss the efficiency wage model, where different types of work and capital investment policies may have different associated monitoring technologies, realizing different wages; or a Schumpeterian model of work search, which allows for large jumps in wages when switching firms, as frictions in the market (perhaps long-lasting) are taken advantage of by well-discerning or lucky workers. If there are social returns to production, as in Deming (2017), then the sorting of high worker-effect workers into high firm-effect firms may feed back into increasing profits, which in turn increase the firm effect through rent-sharing. The monopsony model is naturally not a complete explanation.

A final outstanding puzzle is the relationship between the spatial and separations monopsony approaches estimated separately in this paper. As in Naidu et al. (2019), I find that the relationship is consistent but weak. I estimate ε_{LS} for each municipality and regress it on the labour market HHI (for comparability, not an industry level here). I find an elasticity of -0.036, that is an increase in the HHI is associated with a more monopsonistic labour market. This is consistent with the motivating intuition of the monopsony model that spatial immobility due, for example, to transport costs is but one factor that gives firms wage-setting power.

7 Conclusion

The main results of this paper are that firms have a large role in determining wages—large relative to the international literature on firm wage premia and large relative to invariant worker effects. Firm wage premia explain 25 per cent of the total variance in log wages, 60 per cent of the average gender wage gap, and 40 per cent of the average gap between middle and lower income decile workers. Rent-sharing explains one-fifth of this firm wage dispersion, and the average worker switching from the 25th percentile in firm profits to the 75th percentile is paid 32 per cent more. Monopsony as measured by concentration in the labour market does decrease wages, but the effect on wages is smaller and it contributes little to inequality. On the other hand, monopsony more broadly defined by an upwards labour supply elasticity suggests substantially more wage-setting power than found in the literature for industrialized countries.

This paper makes three contributions through these results. First, the use of administrative data from South Africa adds to the very limited empirical evidence of the role of firm wage premia in developing countries. The results suggest that the labour supply curve can be steep, even with a large surplus labour supply, as indicated by South Africa's expanded unemployment rate of over 35 per cent. A cautionary note is that relatively unique features of the South African labour market are likely to increase monopsony power, such as the apartheid legacy, which likely increased search frictions through spatial dislocation, insider bargaining structures, and employment networks (Magruder 2010).

Second, I make a methodological contribution in estimating monopsony power. I argue that estimates of monopsony should follow the best practice in estimates of rent-sharing, which is to utilize variation in the firm component of wages rather than the full wage. I show that failing to do so leads to strongly upwardly biased estimates of monopsony power, for both the labour concentration approach and the separations-based labour supply elasticities. This common framework of firm wage premia further allows a comparison of the impact of rent-sharing and labour market concentration.

Third, I provide additional evidence in support of a monopsony model in explaining these results. Importantly, I show that the wage gap predicted by separately estimated labour supply elasticities accounts for a third of the average gap in firm effects by gender, and all of the average gap in firm effects by income decile group. I also show that unionization tends to increase rent-sharing, which is consistent with bargaining space opened up by monopsonistic wage-setting.

The importance of firm effects is relevant to both the academic literature and policy in South Africa, which typically focus on the supply side to address employment, poverty, and inequality. South Africa's National Development Plan, a central policy document, highlights as its three priorities 'improving the quality of education, skills development and innovation', 'raising employment through faster economic growth', and 'building the capability of the state to play a development, transformative role'. The role of firm wage-setting power is conspicuously missing. The results of this paper suggest that anti-monopsony policy could substantially increase employment and wages, though the general equilibrium counterfactual of no wage-setting power is not straightforward. This includes addressing labour market concentration, but should primarily be aimed at reducing barriers to employment mobility as in the search-based dynamic monopsony model. Sorting of workers into high-wage firms by group, in large part explained through labour supply elasticities, also suggests, for example, that equal pay policies within firms by gender are not as important as discrimination at the point of hire. Although racial classification data were not available to me, given the apartheid legacy I would predict a similarly large role of sorting in explaining the racial wage gap.

References

Abowd, J.M., F. Kramarz, and D.N. Margolis (1999). 'High Wage Workers and High Wage Firms'. *Econometrica* 67(2): 251–333.

Azar J., I. Marinescu, and M.I. Steinbaum (2017). 'Labor Market Concentration'. Technical Report. Cambridge, MA: National Bureau of Economic Research.

Bachmann, R., G. Demir, and H. Frings (2018). 'Labour Market Polarisation and Monopsonistic Competition'. Unpublished paper.

Banerjee, A., S. Galiani, J. Levinsohn, Z. McLaren, and I. Woolard (2008). 'Why has Unemployment Risen in the New South Africa? 1'. *Economics of Transition* 16(4): 715–40.

Barth, E. and H. Dale-Olsen (2009). 'Monopsonistic Discrimination, Worker Turnover, and the Gender Wage Gap'. *Labour Economics* 16(5): 589–97.

Bassier, I., A. Dube, and S. Naidu (2019). 'Monopsony in Movers: The Elasticity of Labor Supply to Firm Wage Policies'. Mimeo.

Benmelech, E., N. Bergman, and H. Kim (2018). 'Strong Employers and Weak Employees: How does Employer Concentration Affect Wages?'. NBER Working Paper 24307. Cambridge, MA: National Bureau of Economic Research.

Bowles, S., H. Gintis, and M. Osborne (2001). 'The Determinants of Earnings: A Behavioral Approach'. *Journal of Economic Literature* 39(4): 1137–76.

Card, D., A.R. Cardoso, and P. Kline (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): 633–86.

Card, D., A.R. Cardoso, J. Heining, and P. Kline (2018). 'Firms and Labor Market Inequality: Evidence and Some Theory'. *Journal of Labor Economics* 36(S1): S13–S70.

Darity, W.A. and P.L. Mason (1998). 'Evidence on Discrimination in Employment: Codes of Color, Codes of Gender'. *Journal of Economic Perspectives* 12(2): 63–90.

Deming, D.J (2017). 'The Growing Importance of Social Skills in the Labor Market'. *The Quarterly Journal of Economics* 132(4): 1593–640.

Depew, B., and T.A. Sørensen (2013). 'The Elasticity of Labor Supply to the Firm Over the Business Cycle'. *Labour Economics* 24: 196–204.

Dube, A., L. Giuliano, and J. Leonard (2019). 'Fairness and Frictions: The Impact of Unequal Raises on Quit Behavior'. *American Economic Review* 109(2): 620–63.

Ebrahim, A., M. Leibbrandt, and V. Ranchhod (2017). 'The Effects of the Employment Tax Incentive on South African Employment'. WIDER Working Paper 2017/5. Helsinki: UNU-WIDER.

Fedderke, J., N. Obikili, and N. Viegi (2018). 'Markups and Concentration in South African Manufacturing Sectors: An Analysis with Administrative Data'. *South African Journal of Economics* 86: 120–40.

Fleisher, B.M. and X. Wang (2004). 'Skill Differentials, Return to Schooling, and Market Segmentation in a Transition Economy: The Case of Mainland China'. *Journal of Development Economics* 73(1): 315–28.

Gerard, F., L. Lagos, E. Severnini, and D. Card (2018). 'Assortative Matching or Exclusionary Hiring? The Impact of Firm Policies on Racial Wage Differences in Brazil'. Technical Report. Cambridge, MA: National Bureau of Economic Research.

Hirsch, B., E.J. Jahn, and C. Schnabel (2018). 'Do Employers Have More Monopsony Power in Slack Labor Markets?'. *ILR Review* 71(3): 676–704.

Hull, P. (2018) 'Estimating Treatment Effects in Mover Designs'. *arXiv preprint arXiv:1804.06721*.

Kerr, A. (2018) 'Job Flows, Worker Flows and Churning in South Africa'. *South African Journal of Economics* 86: 141–66.

Kerr, A., and M. Wittenberg (2017). *A Guide to Version 3.2 of the Post-Apartheid Labour Market Series (PALMS)*. Cape Town: DataFirst Portal, University of Cape Town.

Kreuser, C.F., and C. Newman (2018). 'Total Factor Productivity in South African Manufacturing Firms'. *South African Journal of Economics* 86: 40–78.

Leibbrandt, M., I. Woolard, A. Finn, and J. Argent (2010). 'Trends in South African Income Distribution and Poverty Since the Fall of Apartheid'. OECD Social, Employment and Migration Working Paper 101. Paris: OECD Publishing.

Leibbrandt, M., V. Ranchhod, and P. Green (2018). 'Taking Stock of South African Income Inequality'. WIDER Working Paper 2018/184. Helsinki: UNU-WIDER.

Magruder, J.R (2010). 'Intergenerational Networks, Unemployment, and Persistent Inequality in South Africa'. *American Economic Journal: Applied Economics* 2(1): 62–85.

Manning, A. (2003). *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton, NJ: Princeton University Press.

Naidu, S., E.A. Posner, and E.G. Weyl (2019, forthcoming). 'Antitrust Remedies for Labor Market Power'. *Harvard Law Review*.

Oaxaca, R.L. (2007). 'The Challenge of Measuring Labor Market Discrimination Against Women'. *Swedish Economic Policy Review*, 14(1): 199.

Ogloblin, C., and G. Brock (2005). 'Wage Determination in Urban Russia: Underpayment and the Gender Differential'. *Economic Systems* 29(3): 325–43.

Ransom, M. and R. Oaxaca (2005). 'Sex Differences in Pay in a "New Monopsony" Model of the Labor Market'. Discussion Paper 1870. Bonn: IZA.

Rinz, K. (2018). 'Labor Market Concentration, Earnings Inequality, and Earnings Mobility'. CARRA Working Paper 2018-10. Washington, DC: US Census Bureau.

Robinson, J. (1933). *The Economics of Imperfect Competition*. London: Macmillan.

Sokolova, A., and T. Sorensen (2018). 'Monopsony in Labor Markets: A Meta-Analysis'. Discussion Paper 11966. Bonn: IZA.

Song, J., D.J. Price, F. Guvenen, N. Bloom, and T. Von Wachter (2018). 'Firming Up Inequality'. *The Quarterly Journal of Economics* 134(1): 1–50.

Vick, B. (2017). 'Measuring Links Between Labor Monopsony and the Gender Pay Gap in Brazil'. *IZA Journal of Development and Migration* 7(1): 10.

Webber, D.A. (2015). 'Firm Market Power and the Earnings Distribution'. *Labour Economics* 35: 123–34.

Webber, D.A. (2016). 'Firm-Level Monopsony and the Gender Pay Gap'. *Industrial Relations: A Journal of Economy and Society* 55(2): 323–45.

Wilson, F. (2001). 'Minerals and Migrants: How the Mining Industry has Shaped South Africa'. *Daedalus* 130(1): 99–121.

Zizzamia, R. (2018). 'Is Employment a Panacea for Poverty in South Africa? A Mixed-Methods Investigation'. Working Paper 229. Cape Town: SALDRU.

Appendix A: a review of monopsony models

A.1 Review of the dynamic monopsony model

The key feature of any model of monopsony is that the elasticity of employment with respect to wages is not infinite, as in the competitive model, and this is explained *inter alia* by search costs, heterogeneous preferences for firm-specific amenities (such as co-worker relationships and convenience of location), and inattention to alternatives. The model below follows Manning (2003: ch. 2). I only outline the logic and main equations for the purpose of reference.

Assume each firm pays one wage, the only factor of production is labour, and product per homogeneous worker is p. Denote the distribution of firm wages by F. Employers maximize profits, which is the difference between the wage and the product, multiplied by the number of workers. In equilibrium, the number of workers employed at a firm requires that the number of workers separated must equal the number of workers recruited.

$$\max_{w} \pi = (p-w)N(w;F) \qquad \text{s.t.} \qquad s(w;F)N = R(w;F)$$

How are separations and recruitment determined? There is a job offer rate λ, which assumes that offers are randomly distributed among workers. Separations depend on an exogenous rate δ (souring co-working relationships, misdemeanours at work, etc.) and job offers that pay above the current firm. In terms of recruitment, unemployed workers accept all job offers, and the number of employed workers recruited is proportional to the number of workers paid below a given wage, where this worker–wage distribution is given by $G(w;F)$. Unemployment is simply a function of the ratio between the separation and offer rates, derived from a standard Bellman equation:

$$s(w;F) = \delta + \lambda(1-F(w))$$

$$R(w;F) = \lambda(u + (1-u)G(w;F))$$

$$u = \frac{\delta}{\delta+\lambda}$$

Firms face a trade-off in paying higher wages, which increases N through attracting more recruits and decreasing separations, but which also reduces profits. The equilibrium condition is that firms make equal profits, which determines profit as a function of offers and exogenous separations, in turn reducible to the unemployment rate:

$$\pi^* = \frac{\delta\lambda(p-b)}{(\delta+\lambda)^2} = u(1-u)(p-b)$$

The first result is that firms pay a distribution of wages for the same job, $b \leq w \leq p - u^2(p-b)$. Intuitively, if many firms paid the same wage, a firm could increase wages slightly to attract many more workers and thereby increase profits; a contradiction in equilibrium. Therefore, firms pay a distribution of wages. The model can easily be modified to allow for firms with different products to pay at different parts of the distribution, which is more realistic and I follow this assumption above by relating the distribution of firm wage premia to the rent-sharing elasticities. A feature of this model is the effect of unemployment as a 'labour disciplining' device. As unemployment increases, wages of employed workers decrease. Moreover, a large structural unemployment rate is accommodated as there is no tendency towards full employment in this model:

$$E[w] = \frac{\delta}{\delta+\lambda}b + \frac{\lambda}{\delta+\lambda}p \tag{6}$$

I follow Bachmann et al. (2018) in estimating this econometrically. The outline is as follows. Note that in a steady state, labour is equal to recruitment divided by the separations rate, $L(w) = \frac{R(w)}{s(w)}$. Further, recruitment can be decomposed into hires from employment and unemployment (similarly for separations). Taking the log of both sides and multiplying by w to convert to elasticities, we find Equation 5 as a relationship between ε_{LS} and the component separations and recruitment elasticities. Note that the estimation of ε_{LS} relies on the responsiveness of separations and recruitment to wages, and *not* on the assumptions embedded in the search-based structural model above. In particular, the elasticities could be based on the model proposed by Card et al. (2018). Note also that if recruitments and separations from unemployment are wage inelastic, then the recruitment of one firm is a separation from another and $\varepsilon_{LS} = -2\varepsilon_{sep}$.

A.2 Relating the rent-sharing and labour supply elasticities

It would be useful to compare the estimated ε_{rent} to the predicted ε_{rent} based on the model and the estimated ε_{LS}. However, their relation is dependent on other parameters. In a simple profit-maximizing model with $\pi = pTQ(L) - wL(w)$, where T indicates productivity, the profit-maximizing wage $w = \frac{\varepsilon_{LS}}{1+\varepsilon_{LS}}TP$ yields $\varepsilon_{rent} = \frac{\partial lnw}{\partial lnT} = 1$. Note that this does not imply workers receive the full gain, since workers' base wage is below marginal product.

Introducing a reservation wage b makes ε_{rent} dependent on b. For example, in the amenities model of monopsony (Card et al. 2018), the homogeneous skill model implies $\varepsilon_{rent} = 1 - \frac{1}{\varepsilon_{LS}(R-1)}$, where $R = \frac{v}{b}$ and v is average rent per worker. Using this equation and my estimated $\varepsilon_{rent} = 0.13$ and $\varepsilon_{LS} = 0.75$, I find $R = 2.5$. This is plausible, given that median profit per worker is R35,000 and the 10th percentile wages are about R17,000. However, the predicted ε_{rent} is sensitive to R, which means using an indirect estimate of R to compare the predicted ε_{rent} to the estimated ε_{rent} will be subject to high uncertainty. Note that even in the amenities model, $b \to 0 \Rightarrow \varepsilon_{rent} \to 1$, as in the simply profit-maximizing model above with $b = 0$.

The point estimates of ε_{rent} associated with different ε_{LS} in Table 5 are *not* consistent with the predictions from either of the monopsony models. For the amenities model above, rent-sharing is positively related to the labour supply elasticity. For the dynamic monopsony model, $\varepsilon_{LS} \propto \frac{\lambda}{\delta}$ with Equation 6 implies that $\varepsilon_{rent} = \frac{\partial lnE(w)}{\partial lnp} = \frac{\lambda}{\lambda+\delta b}$, that is $\varepsilon_{rent} \propto \varepsilon_{LS}$. However, in Table 5, higher rent-sharing is associated with *lower* labour supply elasticities. Reassuringly, none of these point estimates are statistically differentiable from the other groups (except decile 10) and the proportional relation for both models is relatively weak.

Another difficulty associated with relating these two elasticities is the modelling of how rents arise. If instead of the higher productivity assumption given by T above, rents arise through profits that are not associated with marginal labour productivity, then we do not expect *any* rent-sharing under monopsony.

Overall, the sensitive dependence on uncertain parameters renders a calibration exercise between ε_{rent} and ε_{LS} less meaningful.

Appendix B: additional tables and figures

Table B1: Summary statistics on data cleaning

Panel A:cleaning of tax panel data

Tax year	Jobs	Age	Matched	Main job	Firm size, private
2011	15,710,824	13,516,248	12,520,350	10,029,076	7,233,802
2012	15,494,214	13,932,034	12,883,484	10,384,594	7,445,292
2013	15,240,497	13,926,262	13,050,878	10,599,202	7,740,862
2014	14,755,060	14,030,797	13,023,084	10,759,572	7,827,953

Panel B: completion of key firm-level data

Tax year	Freq.	Firm size	Industry	Turnover	Profit	Location
2011	44,830	99.9%	99.7%	76.6%	76.9%	0.0%
2012	46,020	99.9%	99.8%	75.8%	75.8%	0.0%
2013	47,407	99.9%	99.7%	73.1%	73.1%	97.7%
2014	48,475	99.6%	99.8%	59.4%	59.3%	99.8%

Panel C: comparison of firm restriction

	All firms			Firm size > 20 and private		
Tax year	Firms (freq)	Workers (total)	Turnover (total, R'bn)	Firms (freq)	Workers (% of all)	Turnover (% of all)
2011	228,734	10,300,000	7,060	44,830	72%	84%
2012	233,407	10,200,000	7,760	46,020	71%	85%
2013	235,604	10,100,000	7,810	47,407	72%	85%
2014	239,266	9,821,087	4,880	48,475	70%	81%

Notes: Wages are annualized. Age is restricted to workers between 20 and 60 years of age. Firm size is calculated by adding the number of workers in each firm, adjusted by the fraction of the year the worker is employed by the firm. Industry, turnover, profit and location are all reported directly by firms at the firm-level. Profit is the primary measure for the profit-wage elasticity, and alternative constructed measures such as gross profits have similar completion. Where incomplete, the industry of the majority of a firm's workers supplements the industry variable. Location is reported at municipal level and exclude the years 2011 and 2012 due to non-reporting.

Source: author's calculations, based on South African tax records, 2011–14.

Table B2: HHI by municipality

	HHI				Top eight
	Industry hires	Industry emp.	Hires	Emp.	Industry hires (%)
Province					
Eastern Cape	4,658	4,126	2,481	2,344	92
Free State	3,439	3,085	1,914	2,141	92
Gauteng	2,049	1,637	790	1,714	78
KwaZulu-Natal	3,232	2,710	1,824	1,486	85
Limpopo	3,930	3,481	2,182	2,404	91
Mpumalanga	3,050	2,644	1,158	1,245	85
Northern Cape	5,408	4,930	3,139	2,748	96
North West	3,510	3,055	1,450	1,468	94
Western Cape	1,779	1,672	733	570	77
Metros					
Buffalo City	546	488	189	150	52
City of Cape Town	199	239	152	174	31
City of Johannesburg	306	361	72	58	36
City of Tshwane	291	202	42	70	36
Ekurhuleni	173	117	109	49	28
Mangaung	869	713	138	131	65
Mbombela	615	620	121	136	54
Nelson Mandela Bay	331	323	93	70	42
eThekwini	345	425	112	64	38

Notes: rural average gives the labour market concentration of each municipality averaged for each province. Industries are defined in 17 levels. Top-eight statistics refer to the share of the eight firms that hire the most workers. Separate branches of the same firm in a local municipality are treated as one firm. HHI is calculated as the sum of the squared share of firms in the labour market, multiplied by 10,000. Hires are calculated as the number of workers at the firm which were not employed by the same firm in the previous year. All statistics are medians.

Source: author's calculations, based on South African tax records, 2011–14.

Table B3: Alternative specifications for rent-sharing and monopsony

	(1)	(2)	(3)	(4)
Panel A: rent-sharing				
	Winsor	*Assets cont.*	*WFE cont.*	*indusXmun*
Profit	0.106	0.046	0.145	0.119
	(0.007)	(0.011)	(0.023)	(0.021)
Panel B: monopsony				
	Winsor	*Emp.*	*Top eight*	*AMS*
HHI	–0.053	–0.022	–0.125	–0.098
	(0.017)	(0.014)	(0.045)	(0.023)
	Firm FE	*WFE*	*First diff.*	
Separations ε	–0.477	–0.486	–0.332	
	(0.03)	(0.031)	(0.038)	

Notes: all variables are measured in ln units. Monopsony HHI refers to the sum of squared firm share of industry-by-municipality hires (industries at the 17-code level). T-stats are given in parentheses. Winsorization is done at 5 per cent tails. AMS refers to the AMS specification. Workers are limited to those at firms in the private sector with more than 20 employees. Full specifications are given in text.

Source: author's calculations, based on South African tax records, 2011–14.

Table B4: Alternative samples for rent-sharing and monopsony

	(1) Close	(2) Layoff	(3) Full-time	(4) Network	(5) Non-annual	(6) Job-level	(7) Bonus
Panel A: rent-sharing							
Profit	0.121	0.130	0.135	0.202	0.143	0.130	0.491
	(0.025)	(0.022)	(0.023)	(0.038)	(0.018)	(0.020)	(0.048)
Panel B: monopsony							
HHI	0.038	–0.040	–0.009	0.027	–0.061	–0.038	-0.067
	(0.084)	(0.020)	(0.029)	(0.053)	(0.017)	(0.018)	(0.172)
ε_{LS}		0.719	0.721	0.718	0.782	0.805	0.234
		(0.058)	(0.053)	(0.114)	(0.051)	(0.066)	(0.039)
Obs	2,132	36,108	37,478	7,481	40,625	40,949	7,524

Notes: all variables are measured in ln units. Observations are given as the number of the firms in the profit regression. Monopsony HHI refers to the sum of squared firm share of industry-by–municipality hires (industries at the 17-code level). Column 1 gives the firm closings estimate, restricted to firms in the same network. Column 2 uses firms with mass layoffs instead of firm closings, where mass layoff is defined as at least 500 employees or one-third of the last-period firm employment. Column 3 is the panel of workers who are reported to have been employed for the full year. Column 4 restricts the regression to the largest connected set of firms. Column 5 uses reported wages instead of annualized wages as the outcome. Column 6 is run on the level of jobs rather than workers. Column 7 uses reported annual bonus as the outcome. Standard errors are given in parentheses. Full specifications are given in the text. Workers are limited to those at firms in the private sector with more than 20 employees.

Source: author's calculations, South African tax records, 2011–14.

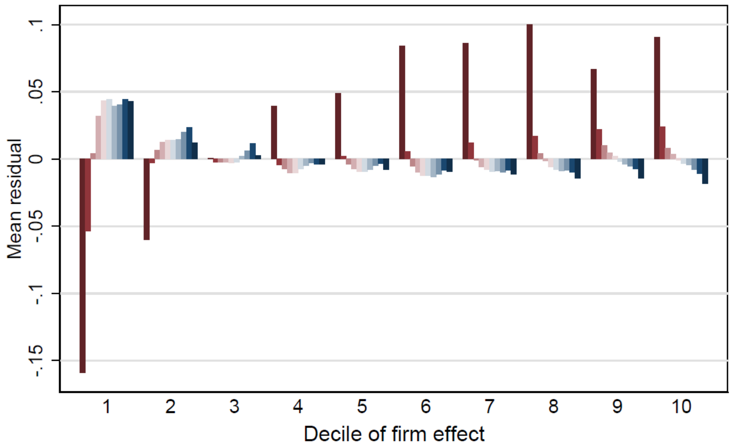
Table B5: Firm effects, rent-sharing, and monopsony by industry

Industry	Firm FE		Rent	Monopsony	
	p50	p90–p25	Profit	HHI	Sep. ε
Agriculture, forestry and fishing	–0.34	0.89	0.094	–0.059	–0.418
			(0.026)	(0.02)	(0.035)
Mining and quarrying	0.71	0.45	0.164	0.010	–0.836
			(0.023)	(0.018)	(0.133)
Manufacturing	0.23	0.81	0.129	–0.046	–0.573
			(0.023)	(0.023)	(0.070)
Electricity, gas, and water	0.87	0.20	0.159	0.044	–1.29
			(0.022)	(0.017)	(0.500)
Construction	0.13	0.83	0.133	–0.031	–0.340
			(0.026)	(0.021)	(0.071)
Wholesale and retail	–0.05	0.60	0.107	–0.022	–0.288
			(0.025)	(0.024)	(0.045)
Transport, storage, and comm.	0.36	0.37	0.140	–0.031	–1.17
			(0.023)	(0.018)	(0.233)
Catering and accommodation	–0.13	0.58	0.111	–0.063	–0.365
			(0.026)	(0.019)	(0.063)
Information and communication	–0.63	2.58	0.062	–0.094	–0.366
			(0.025)	(0.025)	(0.117)
Financing and insurance	–0.12	1.34	0.116	–0.085	–0.463
			(0.027)	(0.018)	(0.052)
Real estate activities	–0.01	0.63	0.116	–0.047	–0.294
			(0.025)	(0.016)	(0.113)
Administrative activities	0.02	0.33	0.155	–0.053	–0.396
			(0.032)	(0.017)	(0.143)
Educational services	–0.64	1.60	0.085	–0.055	–0.198
			(0.026)	(0.033)	(0.023)
Human health and social work	0.42	0.98	0.132	0.039	–0.364
			(0.027)	(0.034)	(0.089)

Notes: workers are limited to those at firms in the private sector with more than 20 employees.

Source: author's calculations, based on South African tax records, 2011–14.

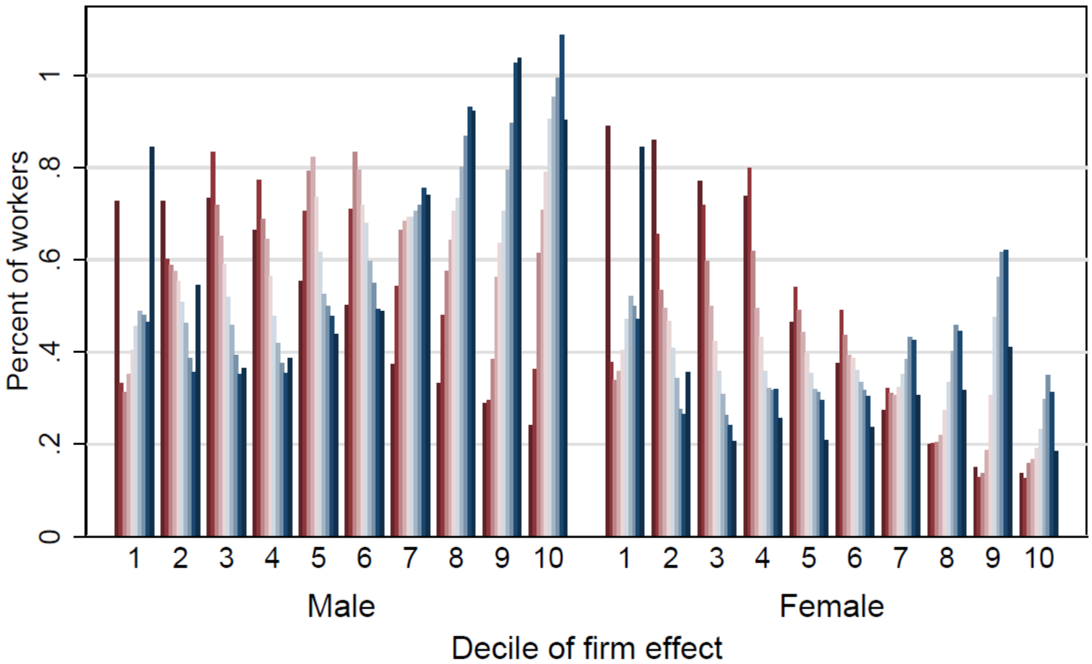
Figure B1: AKM residuals by worker and firm deciles



Notes: residuals are calculated from the AKM regression on worker and firm effects, as well as year dummies and age controls. Deciles of worker effects are plotted in increasing order by decile of firm effect (dark red is the lowest worker effect, dark blue is the highest). Workers are limited to those at firms in the private sector with more than 20 employees.

Source: author's creation based on South African tax records, 2011–14.

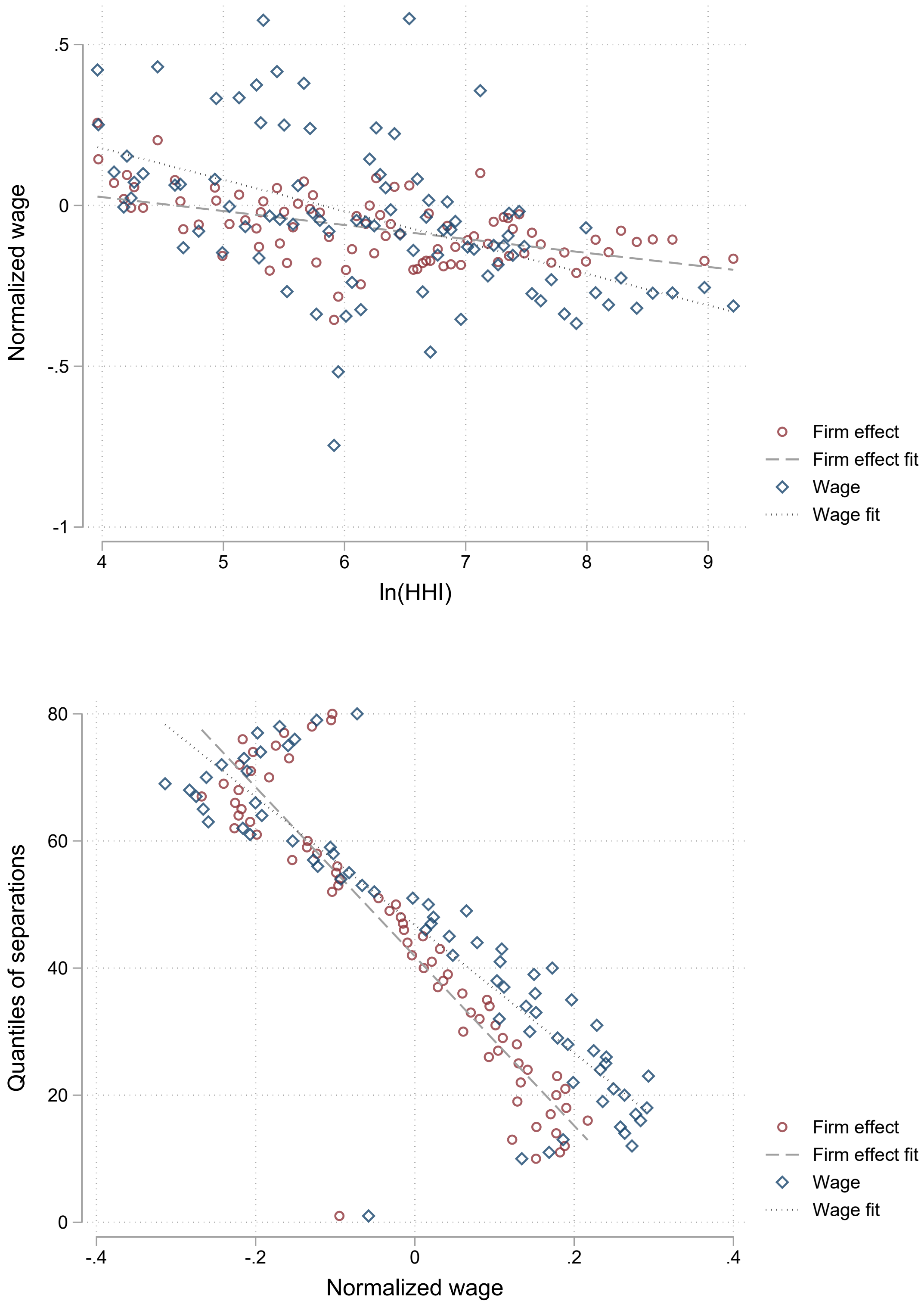
Figure B2: Distribution of worker and firm effects, by sex



Notes: Main worker and firm fixed effects are calculated from the AKM, and then the distribution calculated separately by gender. Deciles of worker effects are plotted in increasing order by decile of firm effect (dark red is the lowest worker effect, dark blue is the highest). Workers are limited to those at firms in the private sector with more than 20 employees.

Source: author's creation based on South African tax records, 2011–14.

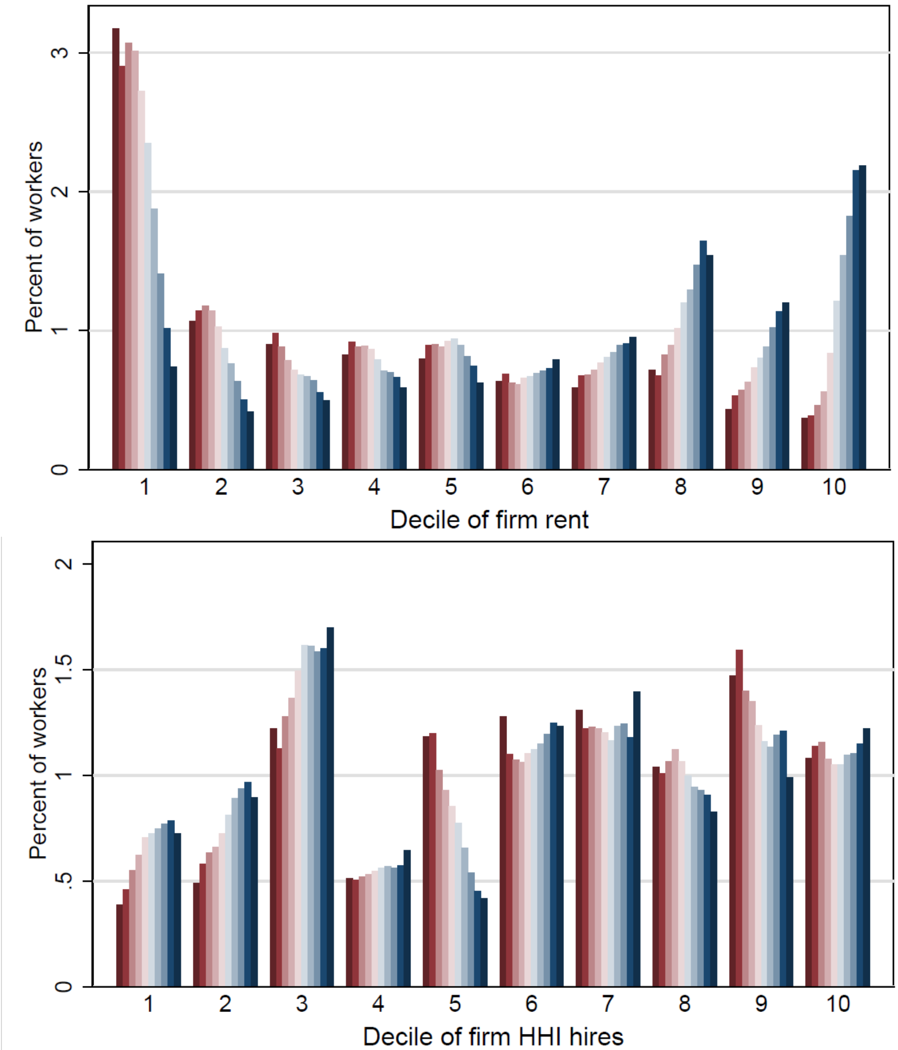
Figure B3: Monopsony scatters: wages by HHI and separations



Notes the *x*-axis shows 100 quantiles of HHI by industry-by-municipality (left) and firm separation rates (right). Each dot in the scatterplot gives the mean ln(wage) of workers for a quantile of profit. Only one observation per firm is given. Observed wages refer to directly recorded wages for each worker, and are centred around 0 for plotting. Firm fixed effects are the results from the AKM models. Workers are limited to those at firms in the private sector with more than 20 employees.

Source: author's creation based on South African tax records, 2011–14.

Figure B4: Distribution of workers among rent-sharing and monopsony effects



Notes: worker and firm effects are results from the AKM regression. Deciles of worker effects are plotted in increasing order by decile of firm effect (dark red is the lowest worker effect, dark blue is the highest). Firm rent is measured as reported profits. Firm HHI hires is the HHI in industry-by-municipality hires, as used in the main regressions of Table 3. Workers are limited to those at firms in the private sector with more than 20 employees.

Source: author's creation based on South African tax records, 2011–14.