UNU WIDER

WIDER Working Paper 2023/132

Rent sharing, wage floors, and development

Joshua Budlender[1] and Ihsaan Bassier[2]

November 2023

Abstract: Faced with more favourable demand conditions, many firms raise wages. However, we show that firms with labour market power, lower productivity, and binding wage floors will absorb these positive revenue productivity shocks as excess profits instead of increasing wages or employment. Our prediction follows from a simple but novel theoretical insight under a standard framework of monopsonistic competition, and we empirically test this theory in South Africa using firm-level administrative data. We first explain how firm wage-setting behaviour changes at a productivity threshold directly related to the wage floor and then show how the predicted wage, employment, and profit patterns are evident in the cross-section of firms covered by collective bargaining agreements. We then replicate and extend a leading method of identifying rent-sharing elasticities, but estimated separately by firm revenue productivity bins. As predicted by the theory, we find that firms below the threshold increase wages and employment less, and profits more, in response to revenue productivity shocks, and that there is a break at the threshold where wage floors bind. The study complicates the conclusions emerging from the literature on firm rent-sharing, and forms part of an explanation for 'stalled' development and 'jobless growth'.

Key words: rent-sharing, monopsony, minimum wage, firm productivity, firm employment decision

JEL classification: J38, J42, M5, O32

Acknowledgements: For comments and suggestions on earlier versions of this work, we owe particular thanks to Arindrajit Dube and thanks to Daniele Girardi, Matt Woerman, Ayanda Hlatshwayo, C. Friedrich Kreuser, Murray Leibbrandt, Carol Newman, and John Rand, as well as participants in the UNU-WIDER Work in Progress meeting, the 2023 ESSA Conference, and the UMass Labor Group. For assistance with data access, we thank the team at the NT-SDF, and in this regard we owe special thanks to Michael Kilumelume for sharing his insights and work on the data. We gratefully acknowledge support from UNU-WIDER as part of the SA-TIED programme, and forthcoming fundin32zg from the Structural Transformation and Economic Growth (STEG) research initiative funded by the FCDO.

1 Department of Economics, University of Massachusetts, Amherst, USA; Southern Africa Labour and Development Research Unit (SALDRU), University of Cape Town (UCT), South Africa; jbudlender@umass.edu; 2 Centre for Economic Performance, London School of Economics & Political Science, UK; Southern Africa Labour and Development Research Unit (SALDRU), University of Cape Town (UCT), South Africa; i.bassier@lse.ac.uk

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

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9267-440-3

https://doi.org/10.35188/UNU-WIDER/2023/440-3

Typescript prepared by Siméon Rapin.

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

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

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

The conclusion emerging from a burgeoning literature on rent-sharing is that firms faced with more favourable demand conditions tend to raise wages. But are all firms equally likely to share rents? Or will some respond by taking higher profits instead? And what are the implications for firms' employment responses? We consider these questions in a setting where some firms are bound by wage floors. We first examine the question by drawing out a novel insight from a standard model of monopsonistic competition, which we show implies differential rent-sharing and concomitantly differential employment responses. We then test these predictions using South African administrative data.

We show that in a standard setup of monopsonistic competition, firms with labour market power, lower productivity, and binding wage floors will absorb revenue productivity increases as excess profits, instead of increasing wages and employment. This stands in contrast to the usual view where revenue productivity increases lead to higher wages—per the emerging literature on rent-sharing (e.g., Card et al. 2018; Lamadon et al. 2022)—and increased employment—implicit in both monopsonistic explanations for rent-sharing (Manning 2003; Lamadon et al. 2022) and in classic models of the development process (e.g. Lewis 1954).[1] Instead, we identify a range of firms where revenue productivity increases only increase profits, corresponding to a subset of firms which do not share rents and to 'stalled' development or 'jobless growth' such that increasing labour is not absorbed into the wage sector.

While unevenness in rent-sharing is of general relevance, the developmental implications may be particularly important for middle-income countries seeking to develop through increasing firm productivity while protecting workers with high minimum wages (e.g. Brazil and South Africa), as well as low-income countries with relatively high wage floors associated with subsistence or efficiency wages. It also has implications for common approaches used to estimate labour supply elasticities.

Our prediction follows from a simple but novel theoretical insight building on the work of Dickens et al. (1999) and Manning (2003), who study minimum wages in monopsonistic settings. A large number of studies show that monopsony is a pervasive feature of labour markets (Caldwell and Oehlsen 2018; Dube et al. 2019; Bassier et al. 2022) and may be worse in developing economies (Vick 2017; Sokolova and Sorensen 2021; Chau et al. 2022; Bassier 2023). While many papers seek to understand the direct employment effects of minimum wages under monopsony (Butcher et al. 2012; Brochu et al. 2023; Dustmann et al. 2022; Engbom and Moser 2022), we ask a different question: what are the wage, employment, and profit responses of firms that increase their revenue productivity in monopsonistic labour markets with binding wage floors?

The modern rent-sharing literature suggests that, per a monopsonistic understanding of the labour market, firms increase wages in response to favourable demand conditions because this is the prerequisite for firm expansion. However, in the presence of a binding minimum wage, firms pay above the unconstrained monopsonistic wage and employ workers either along their labour demand curve (demand-constrained firms, receiving no markdown) or their labour supply curve (supply-constrained firms, receiving a reduced markdown). In the former, very low productivity case, firms do not need to increase wages to expand their size, and wages do not respond to revenue productivity shocks while employment responds strongly. We focus on the supply-constrained (lower/mid-productivity) firms, which due to the minimum wage employ more workers than in the unconstrained monopsony case, and which in order to expand employment further must increase the wage above the minimum. We show that, up to

[1] This assumption is also generally common across the modern labour economics literature, where though firm heterogeneity and processes of creative destruction provide avenues for technological advancement to have ambiguous effects on overall employment, at the firm level it is assumed that productivity increases generally lead to increased employment (Cahuc et al. 2014; Autor et al. 2020).

a threshold, these firms do *not* increase wages (and therefore do not increase employment) in response to favourable revenue productivity shocks, because they instead absorb the shock as excess profits until they have restored their markdown to the unconstrained monopsonistic level. In essence, this region acts as the transition from the no-markdown case to the monopsony-markdown case. For higher productivity firms, where the minimum wage does not bind, the usual monopsonistic rent-sharing dynamics apply. The range of firm productivities which fall in the supply-constrained region is larger for more monopsonistic labour markets and for higher minimum wages.

A secondary contribution and implication of this analysis is that one should be cautious about estimation of labour supply elasticities in the context of high minimum wages. Several prominent papers use a revenue productivity shock to identify wage and employment effects, thereby backing out a labour supply elasticity (e.g., Goolsbee and Syverson 2019, Kline et al. 2019, Saez et al. 2019, Amodio and De Roux 2022, Lamadon et al. 2022, Garin and Silvério 2023 and Kroft et al. 2023).[2] When the productivity shock takes place across the demand- or supply-constrained regions, the estimate will not identify the desired unconstrained labour supply elasticity.

The key empirical prediction of the model is heterogeneous causal wage, employment, and profit responses to revenue productivity shocks along the firm productivity distribution, which we test using South African administrative tax data.

We first show that firms bound by the country's various collectively bargained or government-mandated minimum wage regimes do generally exhibit the predicted patterns in the cross-section. After estimating firm-specific productivity using the leading approach in the industrial organization literature (Ackerberg et al. 2015), we find that there is a discernible productivity threshold before which wages are flatter with respect to productivity and close to the minimum wage, and after which the cross-sectional wage–productivity relationship is steeper. This threshold divides supply-constrained and -unconstrained firms, and importantly we find that the productivity–employment and productivity–profit relationships change at the same point, in the manner predicted by the theory (steeper for the employment response and flatter for the profit share response).

We then use these cross-sectionally identified productivity thresholds (for each minimum wage regime) to test our main prediction, which is that the wages, employment, and profit share of firms on either side of these thresholds differentially *respond* to revenue productivity shocks. Our main specification replicates and extends a leading approach in the rent-sharing literature—the Lamadon et al. (2022) 'internal instruments' method—which entails constructing a stacked event study where firm-specific treatment is defined as an unusually large observed increase in firm value added. We estimate heterogeneous wage, employment, and profit share responses by firm productivity bin, after recentring firm productivity around the (minimum wage regime-specific) productivity threshold estimated in the cross-sectional exercise. Our results, which pass falsification tests, strongly support our theoretical predictions: compared to responses in the unconstrained region, and adjusting for the size of the shock to value added, in the constrained region the wage response (the rent-sharing elasticity) is 29% lower, the employment response is 27% lower, and the profit share response is almost two times higher. These differences are statistically significant at the 5% level, and estimates by bin support the prediction that the break in response size occurs around the productivity threshold.

Interestingly, Card et al. (2016) find a strikingly similar piecewise linear relationship, between firm surplus or productivity (they use value added per worker) and firm wage policies (they use AKM firm wage premia) in the cross-section using Portuguese data. Though they do not explore this result further,

[2] This issue also applies to strategies which use minimum wage changes to estimate the firm labour supply elasticity, e.g. Staiger et al. (2010).

our framework resolves what is otherwise a puzzling pattern and suggests our results may be applicable beyond South Africa.

The core ideas of the model are introduced in Section 2, while the data and South Africa minimum wage institutions are discussed in in Section 3. The cross-sectional descriptive evidence is presented in Section 4, and the main empirical exercise—replicating and extending Lamadon et al. (2022)—in Section 5. Section 6 concludes.

2 Theoretical prediction

2.1 Model

Below we briefly recapitulate a simplified version of the Dickens et al. (1999) model of firm responses to minimum wages under monopsony. We then outline our main argument and illustrate it graphically. We show that this insight is retained when many of the simplifications are relaxed (using the more general model presented in Manning (2003: 338–45)), and highlight the key implications for our purposes. The model assumes there are many firms. Lower case letters denote logs.

Simple model

The marginal revenue product of labour (MRPL) of firm *i* is a simple downwards sloping labour demand curve:

$$\mathrm{mrpl}_i = a_i - \eta n_i \,, \tag{1}$$

where a_i is a demand or productivity shifter, and n_i is employment. The elasticity of the labour demand curve under perfect competition would be $1/\eta$. This can be motivated by a production function such as $Y_i = \frac{1}{1-\eta} A_i N_i^{1-\eta}$, where additional factors such as capital can be log-additively included.

The model assumes a firm-facing labour supply curve $w_i = \varepsilon n_i$, where w_i is the firm wage. The firm-facing labour supply elasticity $1/\varepsilon$ is constant across firms and is finite. Such an upwards sloping labour supply curve implies a marginal cost of labour greater than the wage for firm *i*:[3]

$$\mathrm{mcl}_i = \ln(1+\varepsilon) + w_i = \ln(1+\varepsilon) + \varepsilon n_i \,. \tag{2}$$

This model setup represents a very basic and general monopsonistic form, and remains agnostic as to the source of monopsony power (e.g. search frictions or amenities). Setting marginal product equal to marginal cost, the unconstrained employment and wage for firm *i* are:

$$n_i^* = \frac{1}{\varepsilon + \eta} \left(a_i - \ln(1+\varepsilon)\right) \tag{3}$$

$$w_i^* = \varepsilon n_i = \frac{\varepsilon}{\varepsilon + \eta} \left(a_i - \ln(1+\varepsilon)\right) . \tag{4}$$

When the minimum wage w_m is not binding, that is, $w_\mathrm{m} \leq w_i^*$, Equations 3 and 4 hold and $w_i = w_i^*$ and $n_i = n_i^*$. These are *unconstrained* firms. When the unconstrained wage is lower than the minimum wage ($w_\mathrm{m} > w_i^*$) and the latter binds, firms must pay a wage equal to the minimum wage ($w_i = w_\mathrm{m}$) and they attract the number of workers supplied at that wage, $n_i = (1/\varepsilon)\, w_\mathrm{m}$. These are *supply-constrained* firms.

3 $\frac{\partial WL}{\partial L} = \frac{\partial W}{\partial L} L + W = \varepsilon W + W$, which in log terms is $\mathrm{mcl}_i = \ln((\varepsilon+1)W) = \ln(\varepsilon+1) + w$.

However, this condition for supply-constrained firms only applies as long as the marginal revenue product of labour for the firm is above the minimum wage. If not, so that $w_{\mathrm{m}} > \mathrm{mrpl}_i$, then the firm is *demand-constrained* and it reduces employment until $\mathrm{mrpl}_i = w_{\mathrm{m}}$. Firms must still pay the minimum wage, but the new employment level is now governed by firm labour demand constraint, so that $n_i = (1/\eta)(a_i - w_{\mathrm{m}})$.

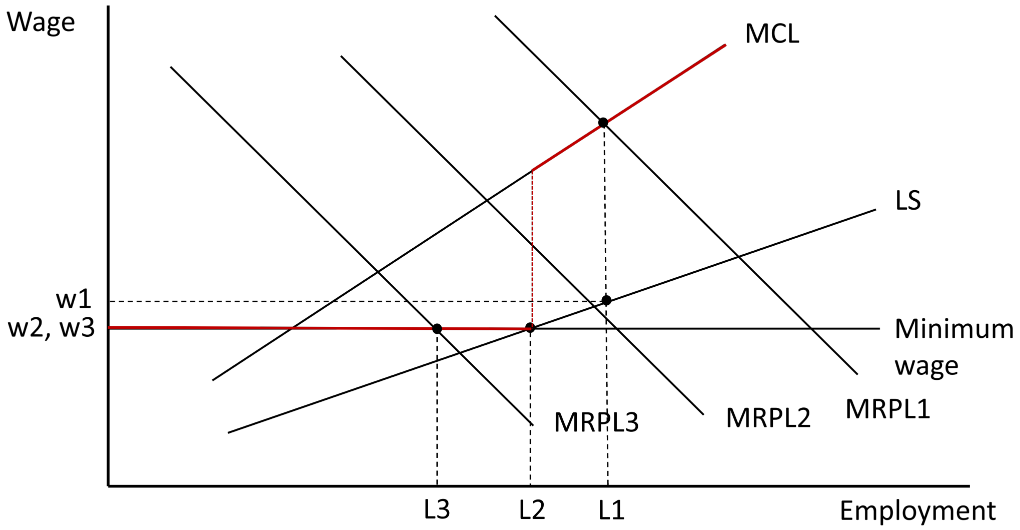
Note that in this simplified model, the only firm-specific factor which determines the unconstrained wage in Equation 4, and therefore whether a firm is unconstrained, supply-constrained, or demand-constrained, is its productivity a_i. The firm's productivity therefore determines which of these *three qualitatively distinct regimes* the firm falls under.

For the unconstrained and demand-constrained firms, we have the well-known result that firm employment increases with productivity (a_i), and for unconstrained firms the wage also increases with a_i. For supply-constrained firms on the other hand, our main insight is that a range of productivity increases does not increase the firm's wage or employment. This is because the unconstrained wage in this range remains lower than the minimum wage ($w_{\mathrm{m}} > w_i^*$), meaning that until the unconstrained wage reaches the minimum, the actual wage remains at the minimum. This supply-bound firm cannot attract additional workers without increasing the wage, so employment remains at the same level, with instead the additional productivity per worker accruing to profits.

Graphical representation

These qualitatively distinct regimes are illustrated graphically in Figure 1. If there is no minimum wage, the usual monopsony set-up that the marginal cost of labour (MCL) is steeper than the firm-facing labour supply curve (LS) applies.[4] For firms sufficiently productive to be unconstrained by the minimum wage (so on the marginal revenue productivity of labour curve MRPL1), the quantity of labour employed is determined by the intersection of the MCL and MRPL curves (L1), and the wage is marked down such that the wage is on the supply curve for that quantity of labour (w1).

Figure 1: Three revenue productivity regimes in the presence of a minimum wage



Source: authors' construction; adapted from Figure 12.6 in Manning (2003: 343).

[4] The marginal cost of labour is steeper than the supply curve because for a monopsonist to hire an additional worker they must increase the wage, which also applies to the wages of already-employed workers.

However, in the presence of a binding minimum wage, the 'effective' marginal cost of labour changes to the discontinuous red line shown in Figure 1. Wages now cannot be below the minimum wage, so the marginal cost of labour when the LS curve is below the minimum wage is simply the minimum wage itself. Of particular interest to us is the discontinuity in the effective MCL at L2, where firms switch from being minimum wage-constrained to unconstrained. Employment is set where the MRPL curve intersects with the effective MCL curve, which for firms with productivity MRPL2 occurs in the region of the effective MCL discontinuity. It is easy to see that for local shifts in MRPL2, the intersection remains in the discontinuity region, and subsequently that these shifts in productivity do not change firm employment (or the wage, which is marked down to the minimum wage level). Instead, local shifts in MRPL in these region are reflected as changes in the size of the markdown from the marginal revenue productivity of labour to the (minimum) wage. These are the supply-constrained firms. Because the minimum wage is above their optimal monopsony wage (on the LS curve), they do not change the wage in response to local productivity shifts, and thus their quantity of labour stays fixed along the supply curve.

For low productivity firms with MRPL at MRPL3—demand-constrained firms—there is no markdown. Shifts in MRPL do affect employment but the wage stays at the minimum wage level.

Full model

The main additions in the full model as presented by Manning (2003: 338–45) are incorporating the average market-level wage as a determinant of aggregate labour supply, and allowing for a firm-specific labour supply shifter b_i (e.g. disamenities) in addition to the revenue productivity shifter a_i. The upwards-sloping firm-facing labour supply curve indicates that labour supply is proportional to a firm's wage premium above the market wage. While in this sub-section we draw out a few key features of the full model, we recapitulate it in full in Appendix B.

The firm's unconstrained wage becomes

$$w_i^* = \frac{\eta\theta w - \varepsilon \ln(1+\varepsilon)}{\eta+\varepsilon} + v_i\,, \tag{5}$$

where w is the (log of the) average market wage, θ is an aggregate labour supply coefficient, and v_i is the firm-specific component of the firm's unconstrained wage, given by

$$v_i = \frac{\varepsilon a_i + \eta b_i}{\eta+\varepsilon}\,, \tag{6}$$

where b_i is the firm-specific supply shifter.

In much the same way that a_i determines whether a firm is unconstrained, supply-constrained, or demand-constrained for a given minimum wage in Equation 4, v_i performs this role for the fuller model in Equation 5. Firms with v_i above some threshold v^* will have $w_i^* \geq w_{\mathrm{m}}$ and will be unconstrained, firms with v_i below v^* but above another threshold v_1^* will be supply-constrained, and firms with v_i below v_1^* will be demand-constrained.

It is useful to focus on the productivity component a_i of v_i and think of v_i as an 'adjusted firm productivity' term. This allows depiction of the three regimes in Figure 1, under the special case where firms only differ in a_i, the positions of their MRPL curves.

Depending on the value of v_i, the three cases are:

1. **Unconstrained (i.e. higher productivity, MRPL1)**: $v_i \geq v^*$

   The optimal monopsony wage is above the minimum wage. These firms are not affected directly by the minimum wage (but may be affected indirectly through spillovers). Employment is given by Equation B2 (in Appendix B) and the wage by Equation 5. Both increase when the MRPL1 curve shifts right (an increase in a_i).

2. **Supply-constrained (i.e mid/lower productivity, MRPL2)**: $v^* > v_i \geq v_1^*$

   The optimal (unconstrained) monopsony wage would be below the minimum wage, and the minimum wage intersects with the firm-facing supply curve. The wage is set at the minimum wage, and employment is along the labour supply curve at the point where it intersects the minimum wage, given by Equation B5. Shifts in MRPL2 (changes in a_i) which keep the firm in this region do not affect wages or employment.

3. **Demand-constrained (i.e. very low productivity, MRPL3)**: $v_i < v_1^*$

   The optimal (unconstrained) monopsony wage would be below the minimum wage, and the minimum wage intersects with the firm MRPL curve. The wage is set at the minimum wage and employment is on the labour demand curve, given by Equation B6. Rightwards shifts of the MRPL3 curve (increases in a_i) increase employment but do not increase wages unless the change is large enough to induce $v_i > v^*$.

Following Manning (2003), we provide expressions for v^* and v_1^* in Equations B3 and B4, respectively, but for our purposes here it is sufficient to note that these thresholds are increasing in the minimum wage, and that the range of the supply-constrained region is increasing with ε (more monopsony):

$$v^* - v_1^* = \frac{\varepsilon \ln(1+\varepsilon)}{\eta + \varepsilon}.$$

It is the supply-constrained region which we are most interested in, where local changes in productivity a_i—which do not cause $v_i \geq v^*$ or $v_i < v_1^*$—<u>do not induce changes in the firm wage</u> (set at the minimum wage), <u>nor the firm size</u> (set where the minimum wage intersects the supply curve). Instead, in this region revenue productivity increases are absorbed by increases in the markdown and profits, until the markdown and profits are at the levels associated with the unconstrained monopsonistic equilibrium.

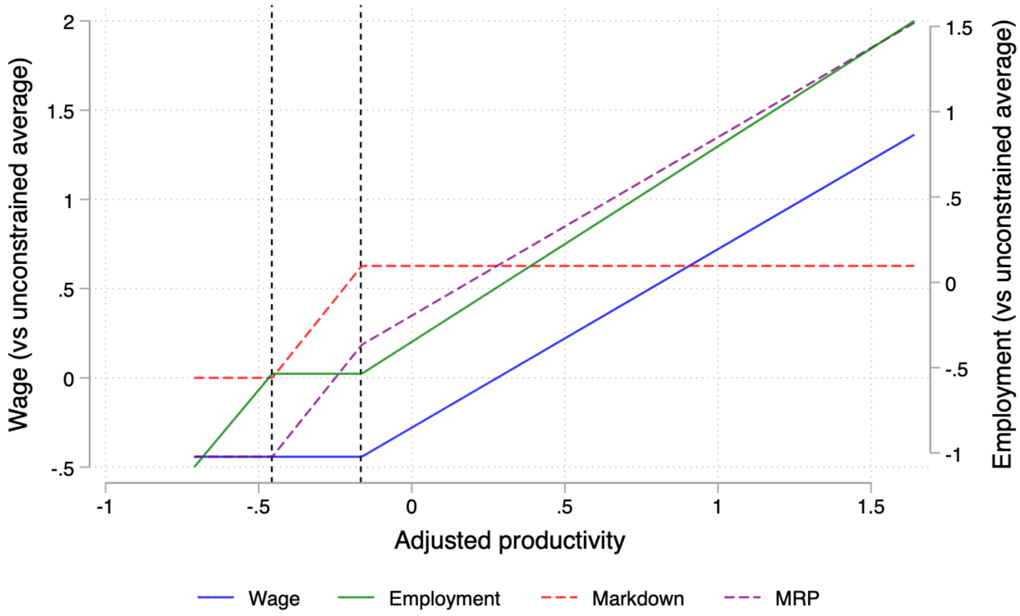
2.2 Simulations

In Figure 2 we present simulations which show the firm wage, employment, and markdown implications of the model for a range of firm productivities.[5] The horizontal axis is a firm's v_i or 'adjusted productivity' with firm-specific supply shifters (b_i from Equation 6) set to zero, so that location along the axis is determined by the firm-specific productivity shifter a_i. This divides firms into very low productivity (demand-constrained), mid/lower productivity (supply-constrained), and higher productivity (unconstrained). In reality, fixed costs and negative stochastic shocks mean that some firms with very low productivity draws will either not be observed in the empirical data or will be short-lived (Olley and Pakes 1996; De Loecker and Syverson 2021); for these illustrative purposes we therefore somewhat arbitrarily truncate the left tail of the v_i distribution at 1.5 standard deviations from the mean.[6]

[5] For purposes of illustration, we ignore firm-specific supply shifters, we impose that MRPL shifters a_i follow a standard normal distribution, we set the market labour supply elasticity to 1.25 and the firm-facing labour supply elasticity to 1. The simulations are based on 1,000 observations, each representing a firm. Wages and employment are normalized by comparison to the average wage and employment under the perfectly competitive case—that is, no monopsony nor minimum wage. The minimum wage is set at -0.5 log units, that is about 70% of the median wage.

[6] Appendix Figure A1 shows the untruncated figure with a long left tail.

Figure 2: Model simulation



Notes: Simulated based on Dickens' model. Productivity is normally distributed.
18.5% of firms are supply constrained (b/n lines), 8.7% are demand-constrained (left of lines).

Note: simulations of wage, markdown, marginal revenue productivity, and firm employment along firm adjusted productivity distribution for 1,000 simulated firms. MRPL shifters a_i follow a standard normal distribution; firm-specific supply shifters b_i set to 0. Market labour supply elasticity is set to 1.25 and firm-facing labour supply elasticity to 1. Wages and employment are normalized by comparison to the average wage and employment under the perfectly competitive case with no monopsony nor minimum wage. The minimum wage is set at -0.5 log units; approximately 70% of the median wage. The left tail of the adjusted productivity distribution is truncated at 1.5 standard deviations from the mean in order to account for firm survival in the observed data.
Source: authors' construction.

The right-most region, after the second vertical line (indicating v^*), represents the unconstrained case: wages and employment are increasing in productivity. Wages are marked down relative to MRPL as in the standard monopsony optimization, and the markdown level is constant.

The left-most region, before the first horizontal line (indicating v_1^*), represents the demand-constrained firms: wages are constrained to equal the minimum wage, though firms do employ more workers as their productivity increases. MRPL is equal to the minimum wage, and there is no markdown.

The middle region between the vertical lines is our subject of interest: these are supply-constrained firms. Firms in this region keep wages fixed at the minimum wage and do not increase employment as productivity increases. Instead, increased productivity is absorbed in higher markdowns, until the markdown is at the optimal level for an unconstrained monopsonist (at v^*). After this point, as firms move to the unconstrained region, they begin again to increase wages and employment with productivity increases, keeping the standard monopsony markdown.

In cases with less monopsony and/or lower minimum wages (Appendix Figure A1), the supply-constrained region shrinks and moves down the productivity distribution, capturing fewer firms. In our baseline specification in Figure 2, 18.5% of firms are supply-constrained and 8.7% are demand-constrained, suggesting that the mechanisms we discuss affect about one in four firms. This compares favourably with our empirical results discussed in Section 4.

3 Data and institutional context

3.1 Data

For our empirical analysis we use the administrative National Treasury-South African Revenue Services (NT-SARS) tax data held at the National Treasury Secure Data Facility (NT-SDF) in Pretoria. This is restricted-access data which can only be accessed in person at the NT-SDF for approved projects.[7] The data we use consist of annual firm balance sheet information from Company Income Tax returns ('ITR14' forms; National Treasury and UNU-WIDER 2023a) which include information such as sales, costs, profits, and industry; and linked worker-level annual payroll data ('IRP5' forms; National Treasury and UNU-WIDER 2023b) which can be used to construct firm-level employment, (approximate) monthly wages, and firm geographic location. The data constitute a panel covering the universe of formal-sector firms in South Africa, and while each dataset covers different periods they all reliably cover at least the period from the 2010 to 2019 tax years (approximately 2009–18).

3.2 Minimum wage institutions

Prior to January 2019, a multilayered wage legislation system operated in South Africa, where minimum wages were set by the government for selected broad industry-locations ('Sectoral Determinations', SDs), or by publicly-recognized Bargaining Councils (BCs) consisting of employers and employees at the sub-industry-location level. Minimum wages can vary substantially by these sectors, and we therefore examine firms separately by their BC or SD. A national minimum wage (NMW) was introduced in January 2019 which partially supersedes the BC and SD system, but its introduction was outside the period of our data so we ignore it.[8]

BCs cover industry-regions, and are constituted by trade union and employer representatives who negotiate industry-region minimum wages.[9] This is a set-up common to a variety of European countries (Bhuller et al. 2022; Jäger et al. 2022), but unlike in some of these countries BC agreements are routinely extended to include non-unionized workers (Bassier 2022). We identify BC firms in the SARS-NT data by matching firms according to their industry and location, using the Bassier (2022) dataset of BC agreements. There are 39 private sector BCs; after restricting for key missing variables, we identify 30 in the data, which cover approximately 26% of the (formal sector) workers. This Bassier (2022) dataset also provides a minimum wage associated with each BC for each year, but it is highly approximate: BC agreements typically specify multiple occupation-specific wages, but because occupations are not observed in the NT-SARS data the lowest BC-specified wage is taken to be the BC minimum.

SDs are government-set wage minima (and conditions of employment) for sectors not fully covered by BCs, often because they are understood as 'hard to organize'. There are 11 SDs, eight of which set minimum wages for formal sector workers (Bassier 2022). SDs are defined more expansively than BCs and sometimes overlap with BCs; in these cases the BC minimum wages apply. While SDs, like BCs, may set occupation- and location-specific wages, there is usually less heterogeneity in minimum wages

[7] This research was approved under the auspices of the SA-TIED programme workstream 1.

[8] Our theoretical mechanism applies regardless of whether minimum wages are set sectorally or by a NMW, though changing the level of the wage floor applicable to each firm would mean that which firms are demand- or supply-constrained would change. The 2019 NMW was generally implemented as a minimum floor: sectoral minima higher than the NMW remained in effect. Insofar as this means the introduction of a NMW weakly increased the level of firm-specific minimum wages, our model suggests that this would increase the proportion of supply-constrained firms (see Section 2.2). Empirically, the last two months of the 2019 tax year overlap with the period of the national minimum wage, but any dynamics in these months are likely to be irrelevant for our results given our empirical design.

[9] Supplementary establishment-level wages can then also be negotiated above these minima (Bassier 2022).

than in BCs. We identify SD firms in the NT-SARS data by matching their industry and location to a dataset we create from promulgated government regulations. We identify the eight formal sector SDs, which exclusively cover about 32% of (formal sector) workers in the data. These are predominantly workers at the lower end of the wage distribution, unlike BCs which have coverage concentrated in the upper half of the wage distribution (Bassier 2022). We also include minimum wages from these regulations, but these are approximate for the same reason as the BC minima. The BC and SD monthly minima in the 2018 tax year are shown in Table A1.

4 Cross-sectional evidence

The prediction we test in this descriptive exercise is the cross-sectional relationship between each firm variable (wage, firm size, and markdown) and firm productivity. Embedded in this test is the existence of qualitatively distinct productivity regions.

4.1 Production function estimation

We first have to estimate firm-specific productivity, and in doing so we draw from a substantial industrial organization literature concerned with production function estimation (Olley and Pakes 1996; Levinsohn and Petrin 2003; Ackerberg et al. 2015; De Loecker and Syverson 2021). Recognizing issues with OLS estimation of productivity such as simultaneity/transmission bias and selection/survival bias, we estimate productivity using the proxy variable/control function method of Ackerberg et al. (2015) (ACF) with materials as the proxy variable, probably the leading approach in the literature (De Loecker and Syverson 2021; Yeh et al. 2022). Cognisant of the Gandhi et al. (2020) critique of attempts to estimate gross output production functions using proxy variable methods, we specify a value-added production function with a flexible translog form:

$$y_{it} = \beta_l l_{it} + \beta_{ll} l_{it}^2 + \beta_k k_{it} + \beta_{kk} k_{it}^2 + \beta_{lk} l_{it} k_{it} + \omega_{it} + \varepsilon_{it} \tag{7}$$

where y_{it} is value added for firm i in period t, l_{it} is firm employment, and k_{it} is firm capital stock, all in logs, while $\omega_{it} + \varepsilon_{it}$ is the productivity residual made up of productivity shocks which are observed or predictable for the firm at time t (ω_{it}) and those which are not (ε_{it}).

4.2 Cross-sectional test

The markdown is of course also unobserved, and so we use the gross profit share as a proxy. This is defined as gross profits over gross profits plus the firm wagebill, and so is equivalent to one minus the labour share as it is defined in Gouin-Bonenfant (2018). At various points we (imprecisely) refer to this as the 'capital share', borrowing from the macroeconomics literature which divides income into labour and capital income.

Our first step is to identify a kink ('knot') in the wage-productivity curve. The pattern we look for is as follows: wages (firm medians) are a piece-wise linear continuous function of productivity, defined over two intervals, and containing a discontinuity in its derivative (a 'knot') at the boundary between the intervals. We identify the knot in the observed distribution by running two OLS regressions, one to the left and another to the right, for each productivity threshold, and selecting the threshold which maximizes the R-squared.[10] According to the model, this is the productivity threshold v^* where firms move from being supply-constrained to unconstrained.

[10] This procedure is analogous to that used by Card et al. (2016) to identify a similar kink in the distribution of AKM firm wage premia against firm log value added.

With this 'wage knot' identified, the prediction from this paper is that at the same productivity threshold, there is also a discontinuity in the derivatives in firm size as a function of productivity, and the markdown (capital share) as a function of productivity. We separately regress each of these variables on productivity to the left and right of the 'wage knot' productivity threshold, which allows us to examine whether there is indeed a change in the slope as we expect.

Due to the different minimum wages which operate in each BC/SD, the above exercise is implemented separately for firms in each BC and SD. We estimate productivity for each firm in 'pre-period' windows for each 'event' (see Section 5), and then the knot-finding exercise above is implemented only for the years after this pre-period.[11]

With the BC- and SD-specific kink-points and cross-sectional patterns having been separately estimated, we then pool the results across all the different BCs and SDs. In order to account for the different minimum wages and other market-level characteristics of each BC and SD, which will necessarily lead to different v^* wage-kink productivity thresholds, before pooling we recentre productivity in each BC and SD around the estimated wage-kink productivity threshold $\hat{v}^*$ in that BC/SD, so that recentred productivity above 0 indicates an unconstrained firm and below 0 indicates a constrained firm. We then re-implement the knot-finding algorithm on this recentred productivity measure, combining all BCs and SDs into one sample.[12]

Finally, though we do not focus on the demand-constrained region, we do try to isolate it from the supply-constrained region by identifying a kink on the employment-productivity curve to the left of the wage threshold (i.e. $\hat{v}_1^*$), for each BC/SD and in the pooled aggregate. In practice, this point is not well identified, potentially because our sample likely has relatively few very low productivity firms. As discussed in Section 2.2, firms with very low productivity draws will be unobserved or under-represented in actual firm data due to fixed costs and endogenous exit (De Loecker and Syverson 2021). This issue is exacerbated in our analysis of Section 5 which of necessity requires a balanced panel. Additionally, an existing literature suggests that Bargaining Councils contain more productive firms, as larger more productive firms endogenously bargain for minima which smaller unproductive firms cannot sustain (Moll 1996; Magruder 2012).

We also do not a priori expect to identify a demand-constrained region in our more powerful pooled cross-sectional exercise, where BCs and SDs are recentred around their estimated wage-kink $\hat{v}^*$, because there is little reason to expect a similar productivity range between the employment-kink v_1^* and the wage-kink v^* across BCs or SDs. In order to illustrate this point, and to more generally facilitate comparison between our pooled empirical results and the theoretical predictions, we present a simulation which analogously pools the simulation of Section 2.2 conducted for 40 different industries with randomly varying labour supply elasticities and minimum wage levels.

4.3 Results

Figure 3 shows results pooled across all BCs and SDs, as well as the pooled simulation discussed above. Appendix Figure A2 presents the results separately for a variety of BCs and SDs.

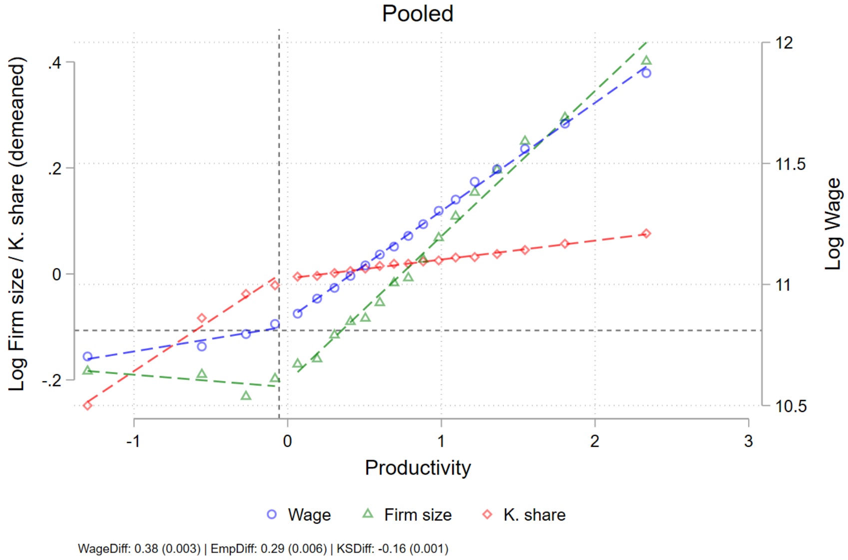
---

[11] We define firm productivity as the average of firm-year-specific productivity in the pre-period, winsorized at the 1% tails. De Loecker and Syverson (2021) note, in a related context, that this averaging may reduce mis-specification error.

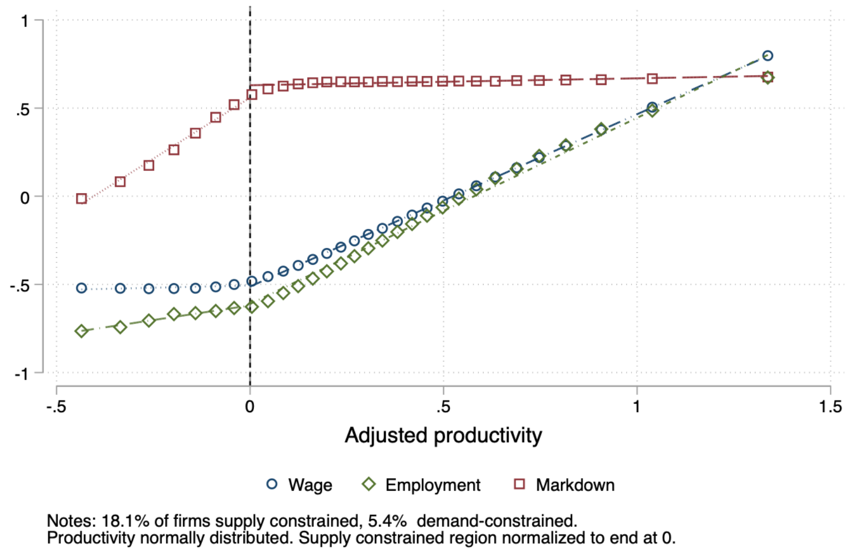
[12] In a few cases our knot-finding algorithm does not identify a plausible interior wage-kink $\hat{v}^*$ and instead identifies a kink at extreme wage values; we trim the estimated wage-kinks $\hat{v}^*$ at the 1^{st} and 99^{th} percentile of the pooled firm distribution.

Figure 3: Pooled cross-sectional results

(a) Pooled empirical results



(b) Pooled simulation



Note: Panel (a) shows firm median wage, employment, and profit share by 20 recentred firm productivity bins (productivity estimated using the ACF method of Section 4; ventiles), for pooled BCs and SDs. Underlying firm productivity is recentred around the estimated wage-knot $\hat{v}^*$ estimated for each BC/SD. The algorithm outlined in Section 4 is used to fit underlying firm median wages as a piece-wise continuous linear function of productivity. Analogous linear fits of employment and profit share are then plotted on either side of the identified wage knot. The line is the value of the aggregate wage-knot $\hat{v}^*$ identified by the algorithm. The horizontal line is the average minimum wage across firms. Panel (b) shows results from pooled model simulations, where the simulation of Section 2.2 is run separately for 40 different industries with randomly varying LSEs and minimum wages, and then industries are pooled after recentring their adjusted productivity v_i around the wage-kink threshold v^*.

Source: authors' construction using National Treasury and UNU-WIDER (2023a, 2023b).

The vertical dotted line indicates the productivity threshold corresponding to the estimated wage-kink productivity threshold $\hat{v}^*$.[13] A horizontal dotted line indicates the minimum wage associated with the general worker (lowest paid) occupation in that BC/SD (or, in the pooled case, the average minimum wage across all the firms). Blue markers show the bin-specific averages of firm median wages; green markers firm employment, and red markers the capital share.

The pooled aggregate case clearly exhibits the predicted features of the model. For wages and firm size, the slope is flatter before the wage-kink threshold, and then more steeply increasing after the threshold. For the profit share, the slope is increasing more to the left of the wage-kink threshold, and flatter to the right. The differences in the slopes around the wage-kink threshold are statistically significant at the 5% level. Reassuringly, the wage-kink is found to be close to 0 on the re-centred productivity threshold, and the average minimum wage appears to correspond quite closely to the firm median wage at $\hat{v}^*$, where firms move from being constrained to unconstrained.

While the same patterns hold for most of the individual BCs and SDs in Appendix Figure A2, they are sometimes noisy or simply not evident for particular BCs or SDs. We view this as unsurprising, given that we are testing a strong prediction that only arises under specific conditions of high minimum wages and significant monopsony, and which in any case may not be detectable given the unavoidably approximate nature of our productivity estimation routine and varying BC/SD sample size.

In Figure 3, 20.2% of observations fall in the constrained region. However, because this is a cross-sectional pooled exercise, firms which appear in the panel for more years are over-represented. If we weight each firm equally, 24% of firms are found in the constrained region. This suggests that the mechanism we discuss affects about one in four firms in South Africa and compares favourably to our simulations in Section 2.2, where the baseline simulation suggested 18.5% of firms are supply-constrained and 8.7% demand-constrained.

5 Heterogeneous responses to shocks

While this descriptive cross-sectional evidence is encouraging, the ideal evidence for our theoretical predictions is heterogeneity in *responses* to revenue productivity increases, for firms along the productivity distribution. To this end, we replicate and extend the 'internal instrument' approach to identifying rent-sharing elasticities from Lamadon et al. (2022) (LMS).

5.1 Method

The core of the LMS 'internal instrument' method is a firm-level event study analysis where treatment is defined as an above-median increase in value added between periods -1 and 0, with some additional specification and variable- and sample-definition restrictions. In the original paper, Lamadon et al. (2022) focus on the effects of these value-added shocks on earnings. We extend this firstly by also examining effects on employment and the capital share, but most importantly by examining heterogeneous responses along the firm productivity distribution. Here we draw from an unpublished presentation by de Frahan et al. (2022), who themselves extend the Lamadon et al. (2022) method to examine heterogeneous effects on employment as well as earnings, but along the firm size distribution.[14] We then further extend this analysis by estimating elasticities with respect to changes in value added rather than semi-elasticities

[13] When there are two vertical lines in Appendix Figure A2, the right-most line is the estimated wage-kink threshold $\hat{v}^*$ and the left-most line is the estimated demand-constraint productivity threshold $\hat{v}_1^*$.

[14] While not a focus of this paper, we note that we have been able to replicate the de Frahan et al. (2022) findings in our data, finding qualitatively similar results.

with respect to the binary treatment, which account for the differently sized value-added responses (to the binary treatment) in each of the heterogeneity regions.

The sample, key variables, and events are defined as follows:

1. Identify 'stayers' in the worker data who remain employed at the same firm for eight consecutive years, separately defining stayers for the event periods 2010–17, 2011–18, 2012–19, and 2013–20, which cover the usable period of the employment data. Drop their records in the first and last years of this tenure (when they may have entered or separated), and only keep workers who are full-time employed over this 6-year period at their firm. Count the number of stayers in each firm for each event period and create year-specific firm-level statistics for stayers' wages (e.g., mean wage, median wage).

2. For each event, only keep firms which have at least N stayers.

   - Lamadon et al. (2022) use a 10-stayer minimum as their baseline, but this is overly restrictive when it comes to the South African firm-size distribution and a labour market context defined by high churn (Kerr 2018). However, in order to identify effects on wages of stayers, one does need to restrict to firms with at least one stayer. In our baseline specification we use a 2-stayer minimum, to mitigate measurement error in 1-stayer firms where the one stayer may be an owner or otherwise unrepresentative of employer/employee dynamics. In Appendix Figure A5 we show that our main results are not sensitive to the number of stayers.

3. Only keep firms which have province and 1-digit industry information, using the interaction of these variables to create 81 labour markets.

   - We have implemented a version which uses the 2-digit industry and 'district' geography variable, which creates 2,600 labour market interactions; our main result is qualitatively unchanged.

4. Over the 6-year period for each event, we treat the first three periods as the pre-period and the latter three as the post. Treatment is defined as an above-median increase in firm value added between periods -1 and 0 for each event, where the median increase is weighted by firm size. Events are stacked (Cengiz et al. 2019). Period -2 is used as the omitted reference period to allow for some mean reversion dynamics in period -1. For the same reason, periods 1 and 2 are considered the post-periods of interest, rather than period 0. Period -3 is used to assess pre-period parallel trend violations.

5.2 Aggregate results

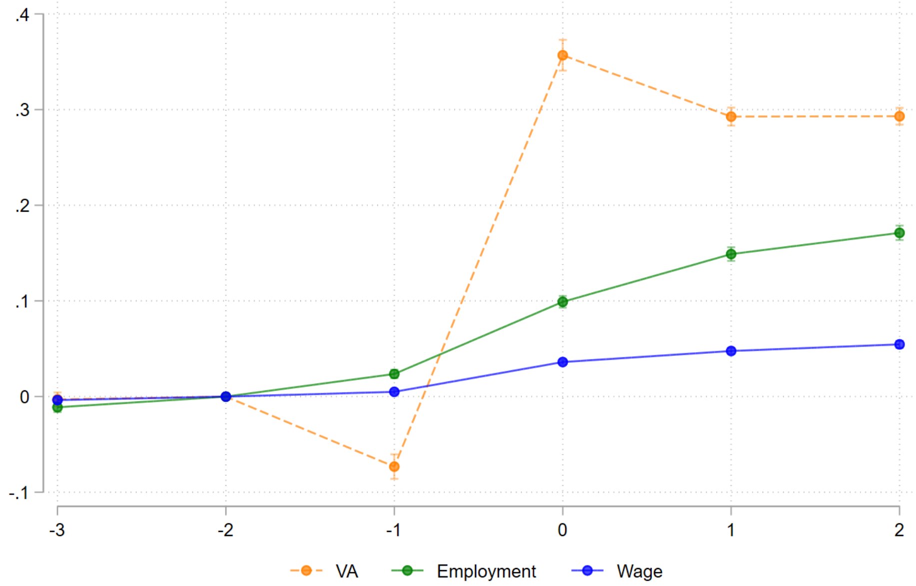
The aggregate event study regression is:

$$y_{i,t,m,e} = \lambda_{i,e} + \delta_t \times \gamma_{m,e} + \sum_{s=-3,s\neq-2}^{2} \beta_s \times \mathbb{1}[t=s] \times D_{i,e} + \varepsilon_{i,t,m,e} \tag{8}$$

where $y_{i,t,m,e}$ is the log of the outcome for firm i at time t in labour market m for event e, $\lambda_{i,e}$ is firm-event fixed effect, $\gamma_{m,e}$ is a market-event fixed effect which is interacted with the time fixed effect δ_t to control for market-event-time fixed effects, $D_{i,e}$ is the treatment variable and the β_s are the coefficients of interest, relative to period -2. Standard errors are clustered at the market-event level.

The aggregate event-study result for our preferred specification (see below) is shown in Figure 4, with the first-stage value-added (VA) response in orange, employment in green, and the median wage of stayers in blue.[15] This closely replicates the figure in the de Frahan et al. (2022) presentation, which also suggests some mean reversion in VA and other dynamics in period -1, but which are small relative to the size of the post-period effects (and seems to dissipate between period 0 and period 1 in any case).

Figure 4: LMS aggregate results



Note: figure shows LMS-style event study where treatment is above-median increase in firm value added between periods -1 and 0. Estimates are normalized relative to period -2. Orange line shows response of log value added, green log employment, and blue the log of median wage of firm stayers (incumbents). Various sample restrictions are discussed in Section 5. 95% confidence intervals are shown with vertical bars.
Source: authors' construction using National Treasury and UNU-WIDER (2023a, 2023b).

5.3 Heterogeneous responses

Following the pooling procedure used for our cross-sectional result (Section 4), we use the recentred productivity terms estimated in Section 4 and previously used to construct the pooled Figure 3: that is estimated firm-specific productivity from equation 7 minus the BC- or SD-specific $\hat{\nu}^*$ wage-kink productivity threshold from the knot-finding exercise. We then estimate heterogeneous responses along this recentred productivity distribution. Recall that the 'wage-kink' indicates the productivity threshold at which firms switch from being supply-constrained to unconstrained.

The theoretical prediction when plotting marginal effects is smaller effects on wages and employment in the supply-constrained region, followed by larger wage and employment effects in the unconstrained region. The prediction is the opposite for the profit share: larger effects in the supply-constrained region and smaller effects in the unconstrained region. Empirically, the difference between the regions may be somewhat attenuated; apart from the issues of combining different labour markets discussed above, or any other dynamics we exclude from the highly stylized model, simple measurement error in our firm productivity measure and wage-kink-finding algorithm would attenuate observed differences between the region, as some unconstrained firms will be observed in the constrained region and vice versa.

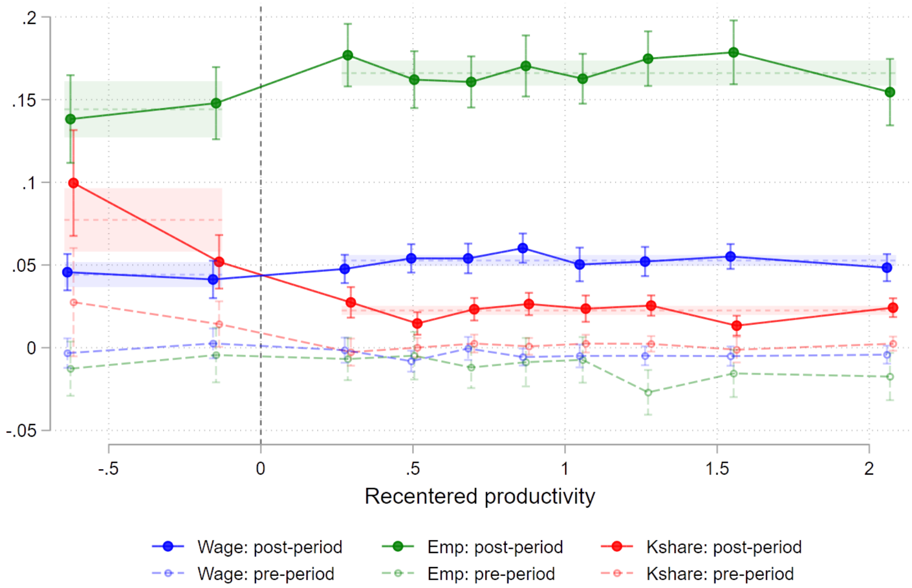
Our results are shown in Figure 5, which presents heterogeneous treatment effects across recentred productivity bins, as well as aggregate responses on either side of the dashed line where recentred pro-

[15] Effects on employment necessarily cannot make the same 'stayers' restriction when defining firm size as we must allow changes in firm size, but the same criteria are used to choose which firms qualify for the analysis.

ductivity=0 ($\hat{v}^*$). Due to the measurement error discussed above, we judge our estimated $\hat{v}^*$ threshold as approximate and drop firms close to the threshold.[16] Treatment effects are the average across post-periods 1 and 2; the dashed lines reflect pre-trend tests for each bin (the coefficients for period -3), and confidence intervals are at the 95% level. We divide firms into ten approximately equally-sized bins.[17]

It is clear from visual inspection that the wage and employment responses are higher in the unconstrained region, while the profit share is lower, and that these responses differentially break around the threshold $\hat{v}^*$. While the 95% confidence intervals for the wage and employment responses overlap slightly when comparing constrained to unconstrained firms, the difference between the estimates is still statistically significant at the 5% level.[18] The statistical significance of the break in the profit share response is visually apparent. Pre-trend coefficients are generally not statistically significantly different from zero.

Figure 5: LMS semi-elasticities by recentred productivity



Note: figure shows estimates from LMS-style event studies when estimated by productivity bin. The horizontal axis is firm productivity (estimated using the ACF method of Section 4), recentred around the BC- or SD-specific cross-sectional estimate of the productivity wage-kink $\hat{v}^*$. Ten approximately equally-sized productivity bins (deciles) are created. The solid lines and points show the average treatment effect across post-periods 1 and 2. The dashed lines and hollow points show effects estimated for pre-period -3. 95% confidence intervals are shown with vertical bars. The horizontal dashed lines with attendant shaded regions (95% confidence intervals) show applicable post-period treatment effects estimated across the productivity bins, separately below and above the wage-kink value $\hat{v}^*$ where recentred productivity equals zero. Red is for firm profit share, green firm employment, and blue the median wage of firm stayers (incumbents).The sample is restricted to firms which have at least two stayers over the event-study period and drops the most productive 5% of constrained firms and least productive 5% of unconstrained firms around the recentred productivity threshold. For all outcomes, responses above and below the $\hat{v}^*$ threshold are statistically significantly different from each other at the 5% significance level (difference calculated using the Delta method). Source: authors' construction using National Treasury and UNU-WIDER (2023a, 2023b).

[16] In our baseline specification we drop the 5% of constrained firms with the highest recentred productivity values, and the 5% of unconstrained firms with the lowest such values. Our results are not sensitive to this trimming procedure; see Appendix Figures A3 and A4.

[17] We require that there be at least two bins on either side of $\hat{v}^*$, as the value of the bins is in seeing the shape of the marginal response against productivity, which means in practice that the bins in the constrained region are smaller than those in the unconstrained region; 12% of firms fall in the constrained region in this baseline specification.

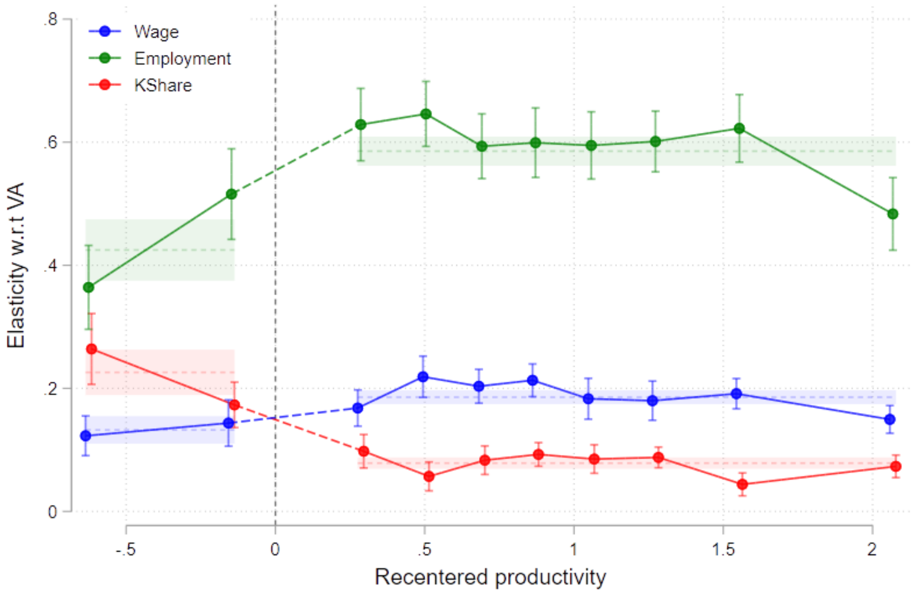
[18] Approximated using the Delta method.

While Figure 5 shows that firm responses to a binary value-added shock are heterogeneous in the manner predicted by the theory, and that pre-trends are approximately flat, we also need to account for differently sized responses in value added to the binary treatment in the different regions, in order to generate truly comparable elasticities in the different productivity regimes. This also creates more interpretable measures; for example the elasticity of the wage to value added is the familiar rent-sharing elasticity estimated in much of the prior literature.

We find that while the change in value added is similar in the constrained and unconstrained regions in our baseline specification, it is slightly larger in the constrained region (0.33 vs 0.28 log points). This has the effect of relatively decreasing elasticities in that constrained region and increasing them in the unconstrained region.

Figure 6, which we consider our main result, shows the wage, employment, and profit share responses as elasticities with respect to the induced changes in value added.[19] It clearly shows the pattern predicted by the theory, with statistically significant differences between constrained and unconstrained firms in their wage, employment, and profit share responses, which break around the estimated wage-kink $\hat{\nu}^*$. Appendix Figure A5 shows robustness to the number of stayers, while Appendix Figure A4 checks sensitivity to the proportion of firms dropped around the estimated threshold.

Figure 6: LMS elasticities by recentred productivity, with respect to value added



Note: figure shows outcome estimates from LMS-style event studies when estimated by productivity bin and divided by the effect on value added. The horizontal axis is firm productivity (estimated using the ACF method of Section 4), recentred around the BC- or SD-specific cross-sectional estimate of the productivity wage-kink $\hat{\nu}^*$. Ten approximately equally sized productivity bins (deciles) are created. The horizontal lines show the average treatment effect across post-periods 1 and 2. 95% confidence intervals are shown with vertical bars and shaded regions. Red is for firm profit share, green firm employment, and blue the median wage of firm stayers (incumbents).The sample is restricted to firms which have at least two stayers over the event-study period and drops the top 5% of constrained firms and bottom 5% of unconstrained firms around the recentred productivity threshold. Elasticities are estimated by regressing the pre-post change in the outcome on the pre-post change in value added, with the change in value added instrumented by the binary treatment variable (together with fixed effects discussed in Section 5.
Source: authors' construction using National Treasury and UNU-WIDER (2023a, 2023b).

[19] We estimate these coefficients with regressions of the log change between the post-period (periods 1 and 2) and period -2 on the equivalent change in value added, with the change in value added instrumented by the binary treatment. We include fixed effects analogous to those in Equation 8 and again cluster at the market-event level.

In terms of magnitudes, the elasticity of the wage with respect to value added (the rent-sharing elasticity) is 0.132 (0.0113) in the constrained region and 0.186 (0.0057) in the unconstrained region; 0.425 (0.0253) and 0.586 (0.012) for the employment elasticity; and 0.226 (0.0187) and 0.078 (0.0046) for the profit share elasticity.[20] The rent-sharing elasticities are very similar to what has been found in the existing literature.[21] As is visually apparent from Figure 6, the differences between the estimates above and below the productivity threshold $\hat{\nu}^*$ are all statistically significant. They imply that in the constrained region, relative to the unconstrained region, the rent-sharing elasticity is 29% lower, the employment elasticity is 27% lower, and the profit share elasticity is 190% higher.

We view the findings in Figure 6, which seem robust, as good evidence for our main theoretical predictions.

6 Conclusion

This paper investigates how monopsonistic firms adjust wages and employment in response to marginal revenue productivity shocks such as more favourable demand conditions. While a large prior literature has emphasized a rent-sharing response—that firms increase wages when faced with such a shock, which the monopsony literature understands as the pre-requisite for expansion in production and employment—we ask whether all firms are equally likely to respond in this way, and in particular consider the case where some firms are bound by a minimum wage. We first show that, theoretically, we should expect heterogeneous responses: lower-productivity, supply-constrained minimum-wage-bound firms will absorb revenue productivity shocks as excess profits per worker instead of increasing wages and employment, unlike demand-constrained or unconstrained firms. We then test this prediction in South African administrative data, finding support for the theoretical prediction both in the cross-section and when replicating and extending a leading approach to estimating rent-sharing elasticities from shocks to value added. The results complicate and enrich the emerging conclusion from the rent-sharing literature—that firms do share rents with workers—by suggesting that this depends on how the firm judges its current level of rents relative to what it expects to receive at its unconstrained equilibrium. While minimum wages may have a variety of positive effects in monopsonistic settings, a firm that is compelled to accept a lower markdown due to a binding minimum wage will not stop trying to increase that markdown when the opportunity arises. This has implications for the common developing-country complaint of 'stalled' development or 'jobless growth'. If labour regulations or other labour supply constraints (e.g. subsistence or efficiency wages) bind and firms earn profits below their desired level, they may choose to respond to market and productivity expansions by simply absorbing these gains as windfall profits rather than expanding production and increasing employment, at least up to some threshold. These firm-side responses seem especially important for understanding labour and development dynamics in developing countries.

References

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). 'Identification properties of recent production function estimators'. *Econometrica*, 83(6): 2411–51. https://doi.org/10.3982/ECTA13408

Amodio, F., and De Roux, N. (2022). *Labor market power in developing countries: Evidence from Colombian plants* [Mimeo].

[20] Standard errors are shown in parentheses.

[21] For instance between 0.137 and 0.156 in Card et al. (2016); between 0.13 and 0.19 in Lamadon et al. (2022); and between 0.14 and 0.17 in Bassier (2023).

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). 'The fall of the labor share and the rise of superstar firms'. *The Quarterly Journal of Economics*, 135(2): 645–709. https://doi.org/10.1093/qje/qjaa004

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

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

Bassier, I., Dube, A., and Naidu, S. (2022). 'Monopsony in movers: The elasticity of labor supply to firm wage policies'. *Journal of Human Resources*, 57(S): S50–S86. https://doi.org/10.3368/jhr.monopsony.0319-10111R1

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

Bhuller, M., Moene, K. O., Mogstad, M., and Vestad, O. L. (2022). 'Facts and fantasies about wage setting and collective bargaining'. *Journal of Economic Perspectives*, 36(4): 29–52. https://doi.org/10.1257/jep.36.4.29

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

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

Brochu, P., Green, D. A., Lemieux, T., and Townsend, J. (2023). 'The minimum wage, turnover, and the shape of the wage distribution'. IZA Discussion Paper 16514. Bonn: IZA Institute of Labor Economics.

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

Butcher, T., Dickens, R., and Manning, A. (2012). 'Minimum wages and wage inequality: some theory and an application to the uk'. CEP Discussion Paper 1177. London: Centre for Economic Performance.

Cahuc, P., Carcillo, S., and Zylberberg, A. (2014). *Labor economics*. MIT press.

Caldwell, S., and Oehlsen, E. (2018). *Monopsony and the gender wage gap: Experimental evidence from the gig economy.* Working paper.

Card, D., Cardoso, A. R., Heining, J., and Kline, P. (2018). 'Firms and labor market inequality: Evidence and some theory'. *Journal of Labor Economics*, 36(S1): S13–S70. https://doi.org/10.1086/694153

Card, D., Cardoso, A. R., and Kline, P. (2016). 'Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women'. *The Quarterly Journal of Economics*, 131(2): 633–86. https://doi.org/10.1093/qje/qjv038

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). 'The effect of minimum wages on low-wage jobs'. *The Quarterly Journal of Economics*, 134(3): 1405–54. https://doi.org/10.1093/qje/qjz014

Chau, N., Kanbur, R., and Soundararajan, V. (2022). 'Employer power and employment in developing countries'. Applied Economics and Policy Working Paper 2022-07. Ithaca: Cornell SC Johnson College of Business.

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

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

de Frahan, I., Lamadon, T., Meling, T., and Mogstad, M. (2022). *Why do larger firms have lower labor shares?* NBER Labor Studies Program Meeting, Spring 2022.

De Loecker, J., and Syverson, C. (2021). 'An industrial organization perspective on productivity'. In *Handbook of industrial organization* (Vol. 4, pp. 141–223). Amsterdam: Elsevier. https://doi.org/10.1016/bs.hesind.2021.11.003

Dickens, R., Machin, S., and Manning, A. (1999). 'The effects of minimum wages on employment: Theory and evidence from Britain'. *Journal of Labor Economics*, 17(1): 1–22. https://doi.org/10.1086/209911

Dube, A., Giuliano, L., and Leonard, J. (2019). 'Fairness and frictions: The impact of unequal raises on quit behavior'. *American Economic Review*, 109(2): 620–63. https://doi.org/10.1257/aer.20160232

Dustmann, C., Lindner, A., Schönberg, U., Umkehrer, M., and Vom Berge, P. (2022). 'Reallocation effects of the minimum wage'. *The Quarterly Journal of Economics*, 137(1): 267–328. https://doi.org/10.1093/qje/qjab028

Engbom, N., and Moser, C. (2022). 'Earnings inequality and the minimum wage: Evidence from Brazil'. *American Economic Review*, 112(12): 3803–47. https://doi.org/10.1257/aer.20181506

Gandhi, A., Navarro, S., and Rivers, D. A. (2020). 'On the identification of gross output production functions'. *Journal of Political Economy*, 128(8): 2973–3016. https://doi.org/10.1086/707736

Garin, A., and Silvério, F. (2023). 'How responsive are wages to firm-specific changes in labor demand? Evidence from idiosyncratic export demand shocks'. *Review of Economic Studies*, corrected proof(-): rdad069. https://doi.org/10.1093/restud/rdad069

Goolsbee, A., and Syverson, C. (2019). 'Monopsony power in higher education: A tale of two tracks'. NBER Working Paper. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w26070

Gouin-Bonenfant, E. (2018). *Productivity dispersion, between-firm competition, and the labor share.* Unpublished paper (JMP).

Jäger, S., Noy, S., and Schoefer, B. (2022). 'The german model of industrial relations: balancing flexibility and collective action'. *Journal of Economic Perspectives*, 36(4): 53–80. https://doi.org/10.1257/jep.36.4.53

Kerr, A. (2018). 'Job flows, worker flows and churning in South Africa'. *South African Journal of Economics*, 86(S1): 141–66. https://doi.org/10.1111/saje.12168

Kline, P., Petkova, N., Williams, H., and Zidar, O. (2019). 'Who profits from patents? rent-sharing at innovative firms'. *The Quarterly Journal of Economics*, 134(3): 1343–404. https://doi.org/10.1093/qje/qjz011

Kroft, K., Luo, Y., Mogstad, M., and Setzler, B. (2023). 'Imperfect competition and rents in labor and product markets: The case of the construction industry'. Revision of NBER Working Paper 27325. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w27325

Lamadon, T., Mogstad, M., and Setzler, B. (2022). 'Imperfect competition, compensating differentials, and rent sharing in the us labor market'. *American Economic Review*, 112(1): 169–212. https://doi.org/10.1257/aer.20190790

Levinsohn, J., and Petrin, A. (2003). 'Estimating production functions using inputs to control for unobservables'. *The Review of Economic Studies*, 70(2): 317–41. https://doi.org/10.1111/1467-937X.00246

Lewis, W. A. (1954). 'Economic development with unlimited supplies of labour'. *The Manchester School*, 22(2): 139–91. https://doi.org/10.1111/j.1467-9957.1954.tb00021.x

Lunt, M. (n.d.). *Piecewise linear ('Hockey-Stick') regression.* Available at: https://personalpages.manchester.ac.uk/staff/mark.lunt/loghockey.html (last accessed 9 September 2023)

Magruder, J. R. (2012). 'High unemployment yet few small firms: The role of centralized bargaining in south africa'. *American Economic Journal: Applied Economics*, 4(3): 138–66. https://doi.org/10.1257/app.4.3.138

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

Moll, P. (1996). 'Compulsory centralization of collective bargaining in South Africa'. *The American Economic Review*, 86(2): 326–29.

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

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

Olley, G., and Pakes, A. (1996). 'The dynamics of productivity in the telecommunications equipment industry'. *Econometrica*, 64(6): 1263–97. https://doi.org/10.2307/2171831

Rovigatti, G., and Mollisi, V. (2016, September). *PRODEST: Stata module for production function estimation based on the control function approach.* Statistical Software Components, Boston College Department of Economics. Available at: https://ideas.repec.org/c/boc/bocode/s458239.html

Saez, E., Schoefer, B., and Seim, D. (2019). 'Payroll taxes, firm behavior, and rent sharing: Evidence from a young workers' tax cut in sweden'. *American Economic Review*, 109(5): 1717–63. https://doi.org/10.1257/aer.20171937

Sokolova, A., and Sorensen, T. (2021). 'Monopsony in labor markets: A meta-analysis'. *ILR Review*, 74(1): 27–55. https://doi.org/10.1177/0019793920965562

Staiger, D. O., Spetz, J., and Phibbs, C. S. (2010). 'Is there monopsony in the labor market? Evidence from a natural experiment'. *Journal of Labor Economics*, 28(2): 211–36. https://doi.org/10.1086/652734

Vick, B. (2017). 'Measuring links between labor monopsony and the gender pay gap in Brazil'. *IZA Journal of Development and Migration*, 7(1): 10. https://doi.org/10.1186/s40176-017-0099-x

Yeh, C., Macaluso, C., and Hershbein, B. (2022). 'Monopsony in the US labor market'. *American Economic Review*, 112(7): 2099–138. https://doi.org/10.1257/aer.20200025

Appendix

Appendix A: additional tables and figures

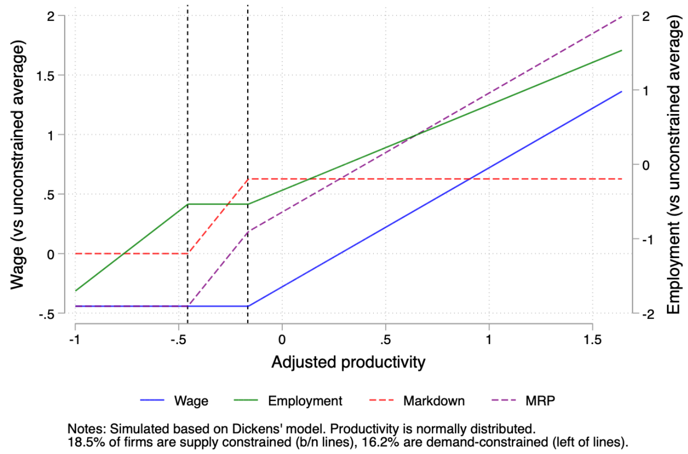
Table A1: Bargaining Council (BC) and Sectoral Determination (SD) minimum wages

Name	Type	Minimum wage (monthly ZAR)
Building (Bloemfontein)	BC	4174
Building (Boland)	BC	3575
Building (Cape)	BC	4810
Building (EC)	BC	6237
Building (Kimberly)	BC	6237
Chemical	BC	6255
Civil engineering	BC	6237
Clothing manufacturing	BC	4476
Electrical	BC	3995
Fishing	BC	3553
Food and restaurant	BC	3087
Furniture (KZN)	BC	2526
Furniture (WC)	BC	2713
Furniture (national)	BC	2714
Hairdressing	BC	2796
Laundry (Cape)	BC	3735
Leather	BC	4963
Meat trade	BC	3281
MEIBC	BC	7550
Motor industry	BC	3812
Restaurant catering	BC	3420
Road freight and logistics	BC	5066
Road passenger	BC	6071
Textile	BC	5546
Transnet	BC	7702
Tyre	BC	11402
Wood and paper	BC	5799
Contract cleaning	SD	3126
Private security	SD	3192
Farm worker	SD	2998
Forestry	SD	2998
Hospitality	SD	3169
Wholesale and retail	SD	3184

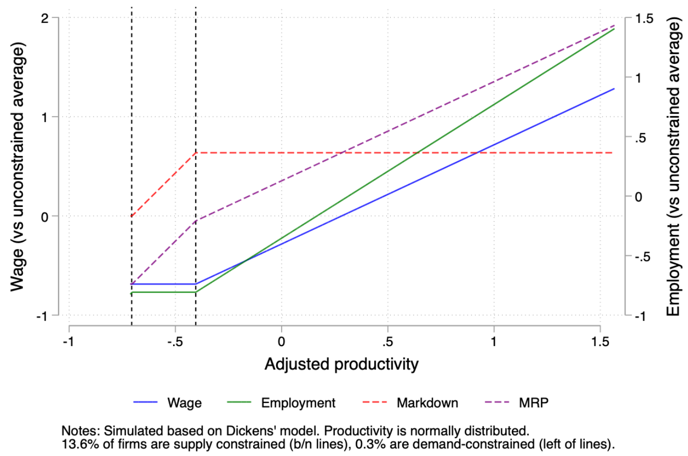
Note: table shows monthly minimum wages associated with each Bargaining Council (BC) and Sectoral determination (SD) in the 2018 taxyear, in nominal Rands. The names are abbreviated. MEIBC refers to the Metal and Engineering Industries Bargaining Council.
Source: authors' construction using data described in Section 3.2.

Figure A1: Additional simulations

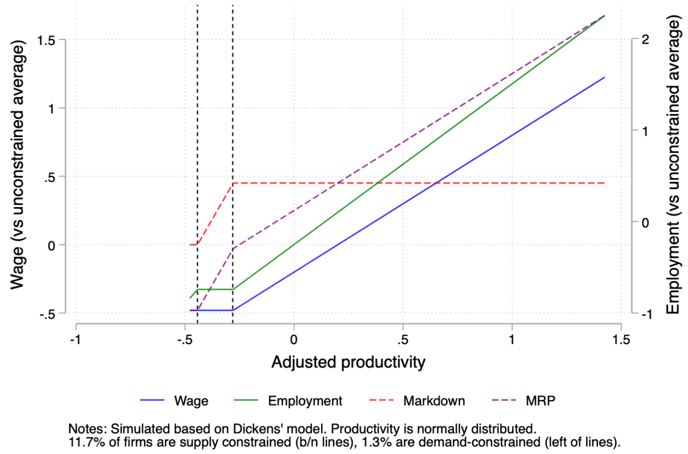
(a) No fixed cost truncation



(b) Lower minimum wage

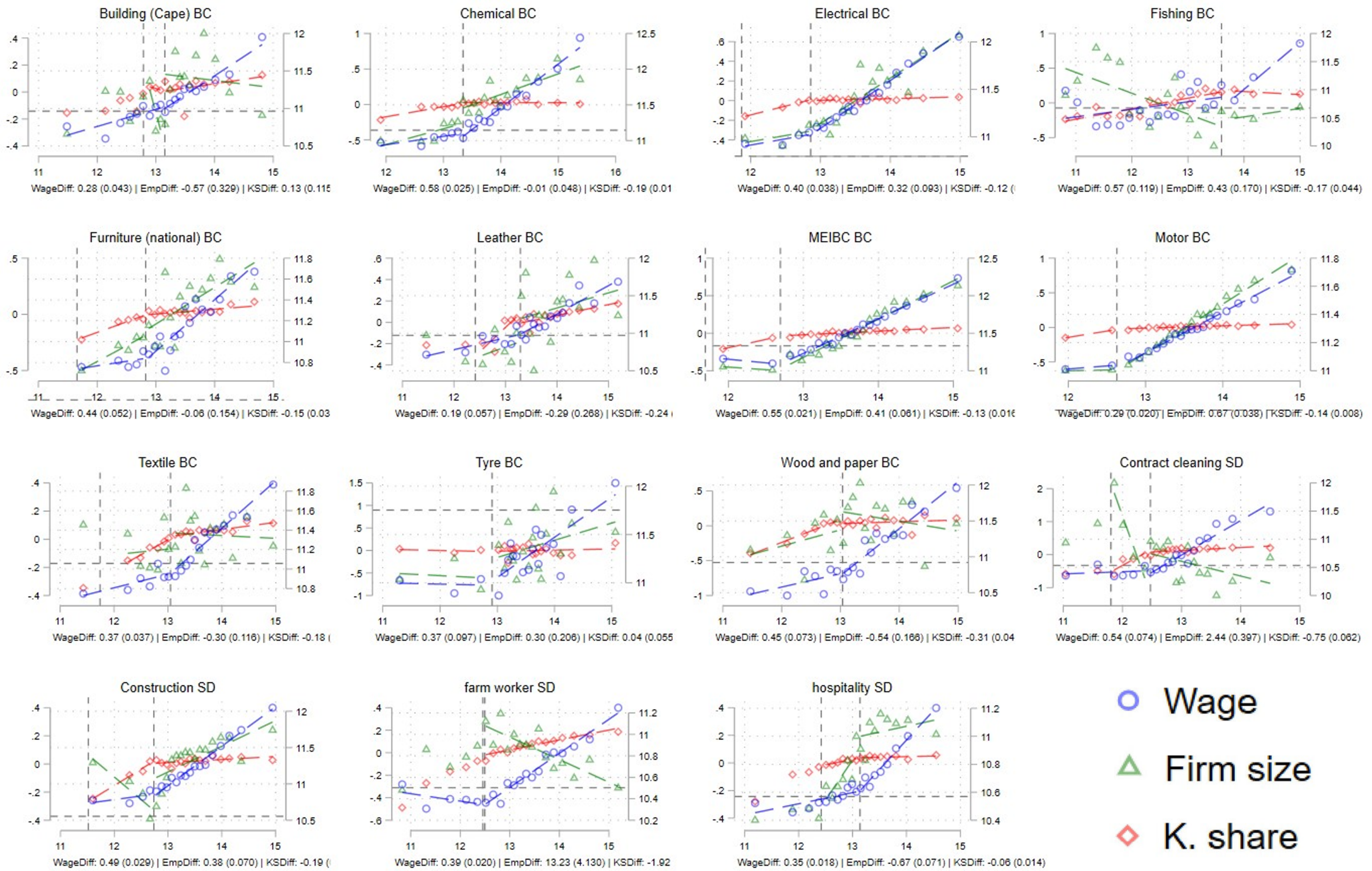


(c) High minimum wage, low labour supply elasticity



Note: simulations are as in Figure 2, except there is no truncation of low productivity firms in Panel (a), the minimum wage is lower (-0.75 log points) in Panel (b), and the firm-facing labour supply elasticity is set to 2 in Panel (c).
Source: authors' construction.

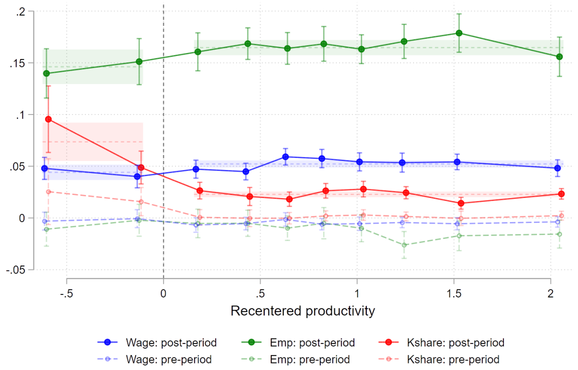
Figure A2: Cross-sectional case studies



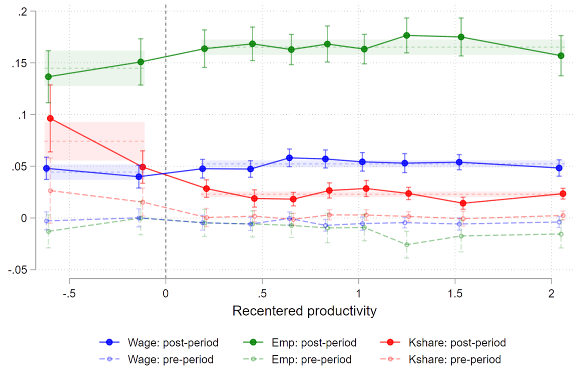
Source: figure shows firm median wage, employment, and profit share by 20 firm productivity bins (productivity estimated using the ACF method of Section 4; ventiles), for selected BCs and SDs. The algorithm outlined in Section 4 is used to fit underlying firm median wages as a piece-wise continuous linear function of productivity. Analogous linear fits of employment and profit share are then plotted on either side of the identified wage knot. The right-most vertical line is the estimated value of the wage-knot $\hat{v}^*$. If there is a second left-most line, this identifies the estimated demand-constrained value $\hat{v}_1^*$. The horizontal line is the BC/SD minimum wage.
Source: authors' construction using National Treasury and UNU-WIDER (2023a, 2023b).

Figure A3: LMS semi-elasticities: trimming around the recentred productivity threshold

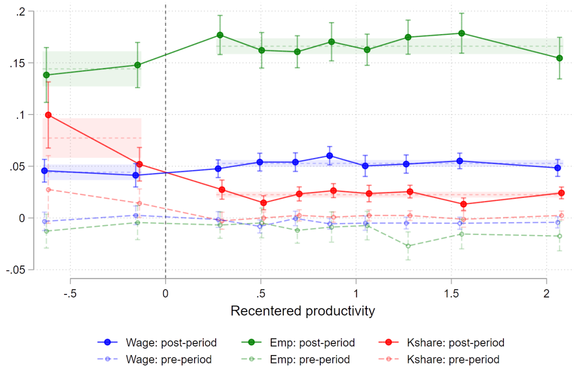
(a) No trimming



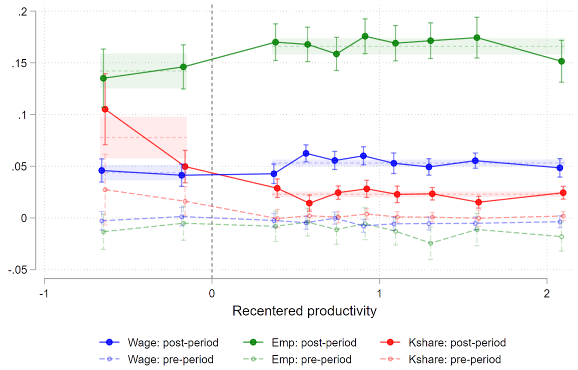
(b) Trimming at 1%



(c) Trimming at 5%



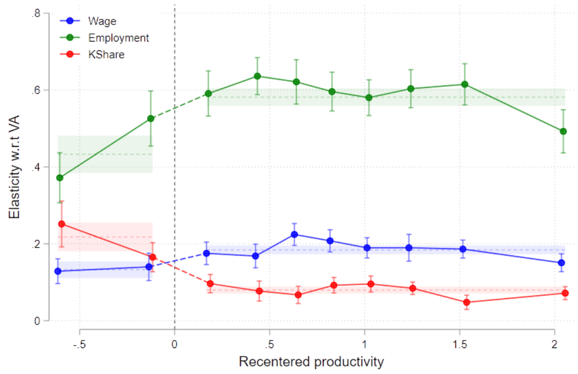
(d) Trimming at 10%



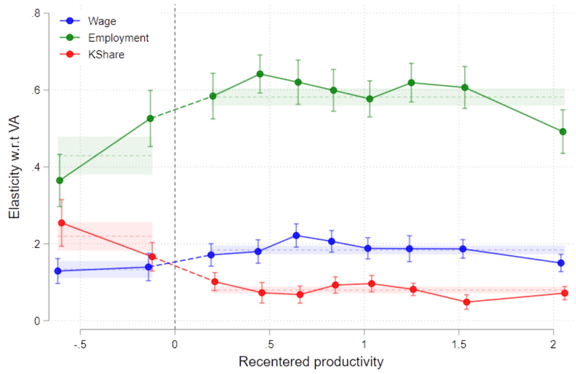
Note: figure shows LMS-style event studies by bin, as more fully described in Figure 5. Each panel shows results for different choices of dropping firms close to the productivity threshold, where 'Trimming at x%' means dropping the most productive x% of constrained firms and least productive x% of unconstrained firms. For all outcomes and specifications, responses above and below the $\hat{v}^*$ threshold are statistically significantly different from each other at the 5% significance level (difference calculated using the Delta method).
Source: authors' construction using National Treasury and UNU-WIDER (2023a, 2023b).

Figure A4: LMS elasticities with respect to value added: trimming around the recentred productivity threshold

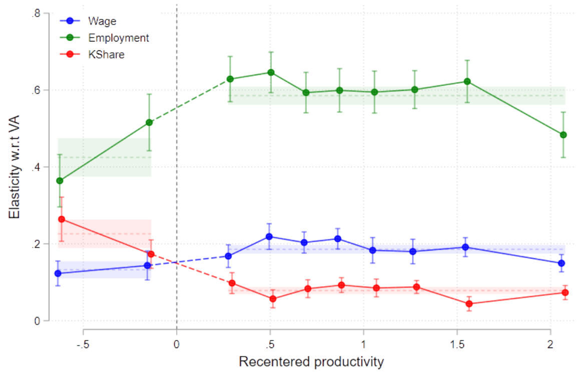
(a) No trimming



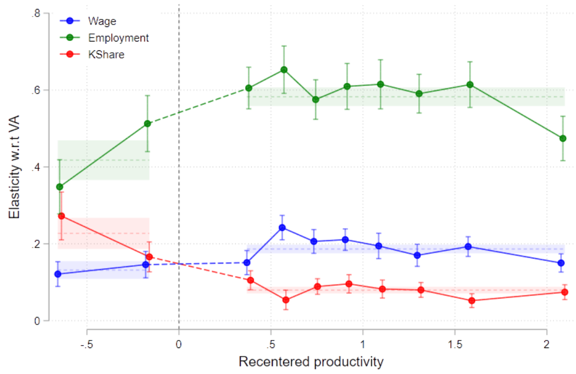
(b) Trimming at 1%



(c) Trimming at 5%



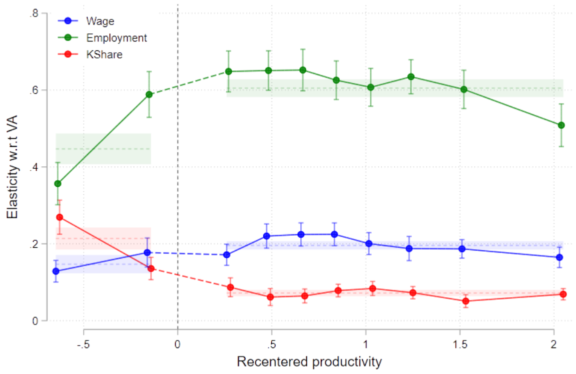
(d) Trimming at 10%



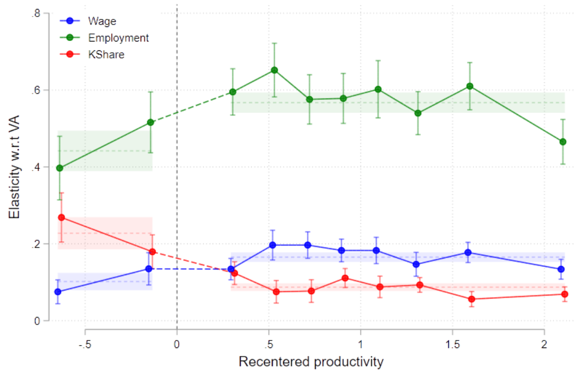
Note: figure shows outcome estimates from LMS-style event studies when estimated by productivity bin and divided by the effect on value added, as more fully described in Figure 6. Each panel shows results for different choices of dropping firms close to the productivity threshold, where 'Trimming at x%' means dropping the most productive x% of constrained firms and least productive x% of unconstrained firms.
Source: authors' construction using National Treasury and UNU-WIDER (2023a, 2023b).

Figure A5: LMS elasticities with respect to value added: various 'stayer' sample restrictions

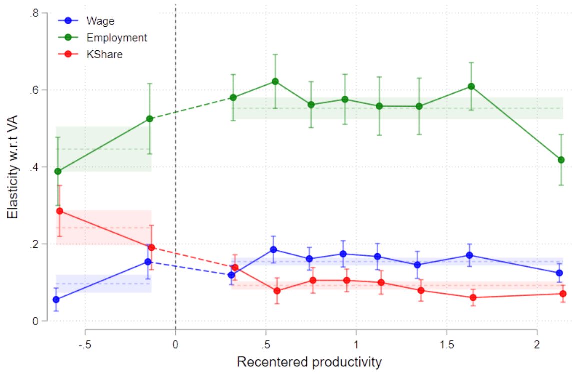
(a) ≥ 1 stayer



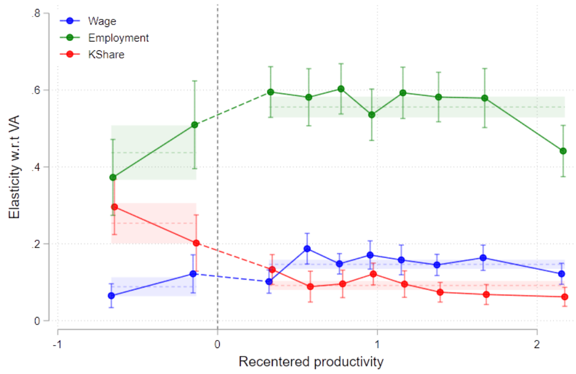
(b) ≥ 3 stayers



(c) ≥ 4 stayers



(d) ≥ 5 stayers



Note: figure shows outcome estimates from LMS-style event studies when estimated by productivity bin and divided by the effect on value added, as more fully described in Figure 6. Each panel shows results when the sample is restricted to have at least 1, 3, 4, or 5 stayers over the event-study period.
Source: authors' construction using National Treasury and UNU-WIDER (2023a, 2023b).

Appendix B: full model detail

The full model in Manning (2003: 338–45) is different from the simplified model in Section 2.1, mainly in that Manning (2003) incorporates the average market wage as a determinant of aggregate labour supply and a firm-specific supply-shifter b_i (e.g. disamenities), so that the firm-specific labour supply depends on the firm wage premium relative to the market wage and the firm-specific disamenity. The model below is essentially a stripped-down re-presentation of Manning (2003).

Specifically, retain Equation 1 for the demand for labour, but now model the share of total employment (N) supplied to firm i (N_i) as a function of its own wage (W_i) relative to an average market-level wage index (W) and the value of its disamenity (B_i): $\frac{N_i}{N} = \left(\frac{W_i}{B_i W}\right)^{1/\varepsilon}$. If one then models the labour supply to the whole market as $N = N_0 W^{\varphi}$ and takes logs (again denoting logs of variables as lower case letters), the labour supply to the individual employer is

$$w_i = (1 - \varepsilon\varphi)w + \varepsilon(n_i - n_0) + b_i\,,$$

or, subsuming n_0 into b_i and defining the coefficient on the average wage as θ,

$$w_i = \theta w + \varepsilon n_i + b_i\,. \tag{B1}$$

The marginal cost of labour in the absence of the minimum wage is then

$$\mathrm{mcl}_i = \ln(1+\varepsilon) + w_i = \ln(1+\varepsilon) + \varepsilon n_i + \theta w + b_i\,,$$

which diverges from the simplified Equation 2 in its two additional terms reflecting the influence of the average wage and the firm-specific disamenity. Equating the expression for the MRPL in Equation 1 to the MCL above, and substituting in Equation B1, the firm's unconstrained wage is given by Equations 5 and 6, while the unconstrained employment level is

$$n_i^* = \frac{-\theta w - \ln(1+\varepsilon) + a_i - b_i}{\eta + \varepsilon}\,. \tag{B2}$$

With the introduction of a minimum wage w_{m}, the discussion of the fuller model in Section 2.1 explains how the value of a firm's 'adjusted productivity' term v_i relative to the thresholds v^* and v_1^* determines which of the qualitative distinct demand-constrained, supply-constrained, or unconstrained regions it falls into. Expressions for these threshold values can be derived by noting that v^* is the value of v_i where the unconstrained wage w_i^* is greater than or equal to the minimum wage w_{m}, so that, from Equation 5,

$$v^* = w_{\mathrm{m}} - \frac{\eta\theta w - \varepsilon \ln(1+\varepsilon)}{\eta + \varepsilon}\,. \tag{B3}$$

For firms which have $v_i < v^*$, for some it will be optimal to accept all workers forthcoming at the minimum wage w_{m}; these are supply-constrained firms. However for other firms with even lower v_i, it is not profitable to employ all the workers forthcoming at the minimum wage w_{m}; these are the demand-constrained firms. To find the threshold value of v_i which delineates these sets of firms, v_1^*, note that these firms set their wage at w_{m} but choose employment less than the potential supply at that wage so that $\mathrm{mrpl}_i = w_{\mathrm{m}}$. From Equations 1 and B1, we can resolve that

$$v_1^* = w_{\mathrm{m}} - \frac{\theta\eta w}{\eta + \varepsilon}\,. \tag{B4}$$

In order to find the equilibrium level of employment for supply-constrained firms, one can substitute $w_i = w_{\mathrm{m}}$ into the labour supply, Equation B1, which Manning (2003) shows can be expressed as:

$$n_i^{\mathrm{sc}} = n(w, a_i, b_i) + \frac{1}{\varepsilon}(v^* - v_i), \tag{B5}$$

where $n(w, a_i, b_i)$ is the unconstrained employment level given in Equation B2. For our purposes it is useful to note that a_i does not enter Equation B1, and therefore does not enter the expression for n_i^{sc}, which reflects our main insight that equilibrium employment for supply-constrained firms is unaffected by local shifts in (revenue-) productivity, and in the special case where $b_i = 0$ all supply-constrained firms will have the same employment level, corresponding to the labour supplied at the minimum wage.[22]

To find the equilibrium employment for demand-constrained firms, again use that they will choose employment such that $\mathrm{mrpl}_i = w_{\mathrm{m}}$, and some rearranging leads to

$$n_i^{\mathrm{dc}} = n(w, a_i, b_i) + \frac{\ln(1+\varepsilon)}{\eta+\varepsilon} - \frac{1}{\eta}(v_1^* - v_i). \tag{B6}$$

[22] While v_i appears in Equation B5, expanding and simplifying necessarily causes the a_i term to drop out, and it resolves to $n_i^{\mathrm{sc}} = (w_{\mathrm{m}} - \theta w - b_i)/\varepsilon$.

Appendix C: data appendix

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

Data access

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

Data used: CIT-IPR5 panel (`citirp5_v5_0`) (National Treasury and UNU-WIDER 2023a) and year-by-year IRP5 job-level data (`v5`) (National Treasury and UNU-WIDER 2023b). Date of first access for this project: 6 January 2023. Last accessed: 9 October 2023.

Software

Our analysis was conducted using Stata 17. User-written programmes and schemes used include `reghdfe` (Correia 2014), `gtools` (Bravo 2018), `prodest` (Rovigatti and Mollisi 2016), `loghockey` (Lunt n.d.), `ivreg2` (Baum et al. 2002), `ivreghdfe` (Correia 2018) and `plotplain` (Bischof 2017).

Variables

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

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

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

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

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

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

Cleaning and sample notes

Our analysis is conducted at the CIT level; PAYE entities without CIT tax reference numbers are excluded from the sample. CIT entities not matched to PAYE entities are also excluded, as this is primarily an employment analysis. When matching firm ITR14/IT14 balance sheet information to the IRP5 data, we match on the basis of firm financial years which best overlap with the tax years used in the IRP5 data. Full-time equivalent employment is calculated using the 'periods worked' variables. We use the unbalanced firm-level panel, which is only balanced after creating stacked events. When industry or location data is missing for a particular firm, we iteratively forwards and backwards impute these variables. These notes represent some particularly noteworthy data cleaning and sample construction decisions, but we cannot outline all such decisions here without reproducing our many thousands of lines of code; users are referred to our do-files which are available at the NT-SDF.