UNU WIDER

WIDER Working Paper 2023/131

Wage inequality, firm characteristics, and firm wage premia in South Africa

Shakeba Foster*

October 2023

Abstract: This paper investigates the role of firm characteristics in driving wage inequality and firm wage premia in the South African labour market. The Abowd, Kramarz, and Margolis (AKM) and Kline, Saggio, and Sølvsten (KSS) regression-based decomposition methods are applied to matched employer–employee administrative tax data for the period 2011–19. Additionally, the Theil index is used as a comparative tool for estimating wage inequality, given that the variance of logarithms applied in the regression-based decomposition methods has been established as an imprecise measure of inequality. The results show significantly high dispersion in wages, as estimated by both the AKM and the KSS methods as well as the Theil index, reaffirming the extent of high inequality in the country. Worker and firm characteristics account for 35 per cent and 18 per cent of wage dispersion, respectively, with a positive worker–firm covariance accounting for 11 per cent. Firm size, industry, profits, geographical location, and whether firms are locally or foreign-owned are found to be important in driving firm wage premia.

Key words: wage inequality, firm wage premia, South Africa

JEL classification: D22, J31, J40

Acknowledgements: Special thanks to Paul Segal and Sunil Kumar for their detailed comments and supervision. The author would also like to thank the staff at the National Treasury Secure Data Facility (NT-SDF) for their assistance throughout the data cleaning and analysis process. This paper has benefited from comments by participants at the SA-TIED technical workshop (April 2023), held in Pretoria, South Africa.

Note: As the research is part of the author's PhD thesis, she will hold copyright to facilitate its publication.

---

* King's College London (KCL), London, shakeba.foster@kcl.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 978-92-9267-439-7

https://doi.org/10.35188/UNU-WIDER/2023/439-7

Typescript prepared by Luke Finley.

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 South African economy is characterized by persistently high income inequality. This fact is emphasized in the economic literature assessing income growth (Bassier and Woolard 2021), unemployment (Lilenstein et al. 2018), taxation (Maboshe and Woolard 2018), and the extent of social welfare (World Bank 2021) in South Africa. Wittenberg (2017c), based on an assessment of household survey data, highlights increasing wage inequality post apartheid in the late 1990s, which then settled at very high levels between 2000 and 2011. He notes that while the bottom of the wage distribution has moved towards the middle, the top continues to move away from the median, resulting in higher inequality. Bassier and Woolard (2021), using both tax and household survey data, similarly highlight that while the middle of the income distribution has stagnated, incomes at the top continue to grow.

For South Africans, wage income represents the most significant component of income for all except the very highest earners, whose incomes are driven substantially by earnings from capital investments (Bassier and Woolard 2021). Given that wage inequality has been consistently high and rising (Wittenberg 2017b, c), an assessment of the drivers of wage income will allow for a deeper understanding of what drives income inequality and will aid in directing policy decisions geared towards reducing income inequality in South Africa.

This paper uses tax administrative employer–employee matched data to investigate the role of firm characteristics in driving wage inequality in the South African labour market. It shows that firm heterogeneity within and between firms of different sizes, of varying profitability levels, and with local or foreign parent companies account for non-negligible proportions of overall wage inequality. I also find that the higher wage inequality within industries, provinces, and other firm groupings does not necessarily imply a high firm wage premium.

Research on wage inequality has for a long time been focused on worker characteristics. One reason for this is that researchers, particularly in developing economies, have been limited to assessing income and wage inequality predominantly based on household survey data, which includes little information on the firms that workers are engaged with. Kerr (2021) notes that users of South African labour force survey data face two main issues. First, if individuals give bracket responses, do not know, or opt not to state their income, this data is imputed by Statistics South Africa (Stats SA)—the agency that conducts household surveys in South Africa. As there are no indicators in the data for which incomes are imputed, it is likely that such imputations may bias conclusions on inequality. Secondly, earnings tend to be under-reported in survey data, and high earners are less likely to share their true income information. Thus, compared with administrative tax data, South African labour force survey data have been found to underestimate income, particularly at the top of the distribution, by up to 40 per cent (Wittenberg 2017a). The increasing availability of employer–employee matched data therefore allows for an examination of both worker and firm contributions to overall wage inequality while avoiding these potential drivers of bias.

The literature that focuses on this decomposition of firm and worker contributions to inequality predominantly utilizes the framework introduced by Abowd et al. (1999). The framework (henceforth AKM) rests on the principle that the overall variance in wages can be separated into the variance of average wages between firms and the variance of wages between workers, or within firms. Therefore, in these models the heterogeneity of firms and workers is captured by attributing the variance in overall wage income to firm and worker fixed effects as well as their time-varying characteristics, including the education and skill levels of workers and firms' compensation policies.

Applying the AKM framework to German data, Card et al. (2013) find that the worker component accounts for up to 40 per cent of overall wage inequality. In addition, they find that the firm component has become an increasingly important driver of wage inequality, accounting for up to 25 per cent of the overall variance. They posit that this increasing contribution of the firm component is driven by an increased variability in the wage premia offered by firms. Criscuolo et al. (2021) find a similar 50 per cent contribution of the firm component for nine countries,[1] highlighting that the pass-through of firm productivity differences to the wage premium is one of the main explanations, and that differences in wage premia may be the result of low job mobility in these countries. Song et al. (2019) finds an even larger firm component in the USA, noting that up to two-thirds of overall wage inequality is explained by firm heterogeneity, and further that the rise in this component is attributable mainly to wage increases in larger firms. Similar findings on the importance of firm characteristics are reiterated in Schneck (2021), who finds that the between-firm component explains almost all of the increase in wage inequality in the Netherlands from 2001 to 2016, and Kristal et al. (2020), who finds faster-growing inequality in fringe benefits relative to core wages, explained by greater firm control over benefits.[2] Hence, in more developed countries, heterogeneity between firms is found to be the main driver of increasing wage inequality.

Analyses of firm and worker contributions to wage inequality in developing countries have been limited predominantly to Latin American countries (LACs). Alvarez et al. (2018) uses the AKM decomposition framework to explain the decrease in earnings inequality in Brazil between 1996 and 2012, finding that most of this decrease came from a 'compression in the worker component' as returns to education and experience fell in the Brazilian labour market. Messina and Silva (2021) and Rodríguez-Castelán et al. (2022) similarly assess the determinants of a reduction in wage inequality in a group of LACs, finding that an increase in the wages of the lowest-paid workers and a reduction in the education and experience wage premia explain most of this reduction, reiterating the finding of Contreras and Gallegos (2011) that education is one of the most important determinants of wage inequality in LACs.

For South Africa, individual- and firm-level tax datasets were made available in 2014, and more widely since 2018, through a partnership between UNU-WIDER and the South African National Treasury. This paper utilizes a matched employer–employee dataset created from the available tax data, to investigate the role of firm characteristics in driving wage inequality in South Africa. To my knowledge, three other papers have previously applied these data to similarly investigate the role of firms and workers in South African wage inequality. Bhorat et al. (2017) apply fixed effect models to the data to assess the role of firms and workers in wage inequality. For the period 2011–13, the authors find a worker effect of approximately 61 per cent and a firm effect of at least 13 per cent. Given the computational intensity which these models require, the authors were limited to using random samples of the data in their assessment instead of the full dataset. Kerr (2021), in a more general analysis of earnings inequality, applies a simple variance decomposition to these data for the period 2011–17. He finds that 41–47 per cent of the variance in log earnings is explained by differences within firms, while 53–59 per cent is explained by differences between firms. Bassier (2023) applies the regression-based AKM decomposition method as in Card et al. (2013) and Song et al. (2019) for the period 2011–16. He focuses his second-stage analysis on the role of monopsony in increasing wage inequality. He finds that worker characteristics account for about 37 per cent of overall wage inequality while firm wage premia account for about approximately

[1] Canada, Costa Rica, Finland, France, Germany, Hungary, Japan, Netherlands, and Portugal.

[2] Other authors apply this framework to examine monetary policy shocks (Moser et al. 2021), trade frictions and reforms (Coşar et al. 2016; Krishna et al. 2011), and returns to on-the-job training (Almeida and Faria 2014).

28 per cent of the variance in wages, and that this is predominantly due to productivity differences across firms and a low labour supply elasticity.

Bhorat et al's (2017) initial assessment suggests linkages between wages and firm age, size, industry, productivity, and levels of profitability. They find that average wages are higher in firms with more market power—a finding explored more extensively in Bassier (2023). They also find that average wages are higher in more productive firms and, in line with findings in the developed-country literature, that firms' trading status is an important determinant of wage inequality, as firms involved in international trade tend to pay higher average wages (Asteriou et at. 2014; Jaumotte et al. 2013). Bhorat et al. (2017) also note a U-shaped relationship between average wages and firm size—the highest-paid workers are attached to the smallest and the largest firms. This result contrasts with results from developed countries showing that wages increase with firm size, resulting in higher overall wage inequality (Ohlert 2016). The literature also documents foreign firms as being important drivers of wage inequality, as they attract the highest-skilled workers and offer higher wages than local firms (Almeida 2007; Chatterjee 2016; Saglam and Sayek 2011). The results here emphasize the role of both firm size and firm profitability; they also show the larger impact of foreign firms on firm wage premia.

Building on this initial work, this paper makes two main contributions. First, the paper examines the impact of a range of firm characteristics on firm wage premia. These include firm size, profitability, industry, province, and whether firms' operations are based locally or internationally. While previous papers have considered a subset of these factors or were largely limited to examining the relationship between individual factors and average wages, this paper directly estimates the impact of each of these characteristics on firm wage premia. The paper also shows how wage inequality evolves over time for each group of firms, complementing the analysis of the drivers of firm wage premia. The paper therefore provides a more descriptive analysis of wage inequality across firms in South Africa and highlights which basis of heterogeneity contributes most significantly to firm wage premia and therefore wage inequality.

A second contribution rests on the methods applied. As I explain in more detail in Section 2, there is a longstanding criticism of the variance of logarithms as a measure of inequality. The variance of logarithms violates an important principle of inequality measurement—the Lorenz criterion—and is therefore not the most appropriate or accurate measure of wage or income inequality (Sen 1973). This paper therefore complements the regression-based decomposition analysis with a simple non-regression-based measure—the Theil index—allowing for a more accurate measure of wage inequality and of the dynamics over time. The paper thus includes both an assessment of worker and firm components of wage dispersion using the AKM and KSS methods, and a dynamic assessment of wage inequality over time using the Theil index.

The remainder of this paper is organized as follows. Section 2 presents an overview of the data and some descriptive statistics. Section 3 then outlines the empirical methods. Section 4 discusses the results of the AKM, KSS, and Theil analyses. Section 5 concludes the paper.

2 Data and descriptive statistics

The paper uses matched employer–employee administrative tax data spanning the period 2011 to 2019 (NT and UNU-WIDER 2023a, b). These data are made available jointly by the South African National Treasury and UNU-WIDER. Employee-level data from individual job certificates (IRP5 forms) and firm-level data from submitted corporate income tax (CIT) information, available separately, are merged to create the resulting CIT-IRP5 panel.[3]

The IRP5 data include 16–20 million observations each year. A variety of adjustments and restrictions are applied to the data to ensure consistency and to allow for comparisons with previous work. First, there is a disconnect between the IRP5 (from individual job certificates) and the CIT data (from CIT tax returns). While the IRP5 data include the full amount of income earned by the worker during the tax year, the CIT data hold firm information for the firm's financial year, which is then assigned to a particular tax year based on the firm's financial year-end. Hence, the IRP5 datasets are realigned to ensure income values cover the same period. The data from IRP5 certificates also include information on partnerships, clubs, associations, and other institutions which have a submission requirement (Ebrahim and Axelson 2019). Thus, I limit the sample to data for individuals only. Third, the sample is limited to individuals between the ages of 20 and 60 years. This ensures that the results are capturing only typical working-age individuals. Finally, the sample is restricted to one job per individual; the job from which the worker earns their highest income is regarded as their main job. Information on the resulting number of observations following each adjustment is presented in Appendix Table A1. On average, 8–9 million observations remain per year. These restrictions on individuals, age, and jobs allow for comparisons with Bassier (2023), which applies similar limits to the data.[4]

Following the above restrictions on the individual-level data, this sample of workers is matched with firms in the firm-level panel, based on tax reference numbers which are identical for firms and their respective employees. About one million workers per year are not matched with any firm in the firm-level panel. The final two columns of Table A1 provide details on the resulting number of worker observations following the matching process as well as a count of workers in firms with more than 20 employees. This latter restriction is applicable to the AKM analysis, where identification is achieved through workers who move between firms; analysis of firms with at least 20 workers allows for more accurate measurement of the firm and worker components of wage inequality (Bassier 2023; Song et al. 2019).

Table A2 presents firm-level summary statistics, highlighting the percentage of firms of different sizes which remain post matching. The sample is comprised largely of micro firms, with an overall median firm size of seven employees. When the sample is restricted to firms with more than 20 employees, the median firm size increases to about 43 employees. Notably, only approximately 25 per cent of firms employ more than 20 workers. Table A3 presents worker-level summary statistics on real income post matching.

3 Empirical framework

The estimation of worker and firm components is done using the AKM decomposition framework. The framework rests on the principle that individual wage income can be decomposed into firm and worker effects, and thus the overall variance in wages can be separated into the variance of average wages between firms and the variance of wages between workers, or within firms. In line with previous literature (Bassier 2023; Card et al. 2013; Song et al. 2019), firm and worker effects are estimated using the largest network or connected set of firms. Firms are

[3] Ebrahim and Axelson (2019) was used as a guide to the CIT-IRP5 panel data.

[4] The sample in Bhorat et al. (2017) is individuals aged 15–64 years; Kerr (2021) does not apply an age restriction.

designated as ‘connected’ or in a ‘network’ based on worker transitions between firms. For example, three firms (1, 2, 3) form a network if worker A moves between firm 1 and 2 and worker A (or another worker B) also moves between firms 2 and 3 (Bassier 2023). The sample is limited to firms with more than 20 workers (as in Bassier 2023 and Song et al. 2019) to allow for a more precise identification of firm and worker effects, which derive from workers who transition between firms.[5]

3.1 First stage: worker and firm effects

The following equation is estimated:

$$lny_{ijt} = \gamma_i + \zeta_{J(i.t)} + X_{it}\delta + \varepsilon_{ijt} \tag{1}$$

where lny_{ijt} is the log wages of worker i at firm j at time t, γ_i are worker fixed effects, $\zeta_{J(i.t)}$ are firm fixed effects, and X_{it} includes workers’ age and year fixed effects. Outliers—earnings greater than ZAR1 billion per year—are removed, and wages are then winsorized by 1 per cent at both the upper and the lower bounds. Kerr (2020) presents a more detailed discussion of outliers in the IRP5 data. I also estimate firm and worker components using the KSS estimator (Kline et al. 2020), which corrects for potential limited mobility bias in two-way fixed effects models such as the AKM framework.[6]

3.2 Second stage: what determines the firm effect?

Following the estimation of overall firm and worker fixed effects in the first stage, firm effects are regressed on firm characteristics to determine which characteristics account for the overall impact of firms on wage inequality. The following equation is estimated at the firm level:

$$\zeta_j = Y_j\beta + \eta_j \tag{2}$$

where ζ_j are firm effects, resulting from Equation 1, and Y_j is a vector of firm characteristics including industry, firm size, firm profits, foreign-local status, and the province in a firm is located. Of note, a similar second-stage equation cannot be estimated for workers as the South African Revenue Service – National Treasury (SARS-NT) data include only information on workers’ age and gender. These worker characteristics are therefore controlled for in the first stage.

3.3 Theil index

As noted above, the AKM and similar variance decomposition methods rely on the variance of the logarithm of wage income. And despite the widespread use of these variance decomposition frameworks, it is also recognized that the variance of logarithms violates an important principle of inequality measurement—the Lorenz criterion—and is therefore not an appropriate measure of inequality. Sen (1973), in one of the earliest papers to elaborate on this issue, posits that the difficulty with this variance as a measure of relative inequality is that it is not mean-independent: that is, ‘x may be a more relatively equal distribution in a uniform way than y, but x can still have a higher variance if the mean income is higher for x than y’ (Sen 1973: 1458). He notes further that

[5] On average, only 1% of employees move out of or into firms with fewer than 20 employees (see Table A7 in the Appendix). Appendix Table A6 shows the total percentages of hires and separations between 2011 and 2019.

[6] I would like to thank Ihsaan Bassier for providing the required code to estimate the wage decomposition using the KSS framework.

in using the standard deviation of logarithms as an inequality measure, 'a perverse result is possible in the sense that a transfer from a poorer person to a richer man can, under certain circumstances, reduce the standard deviation of logarithms for relatively high incomes'.

Foster and Ok add to Sen's assessment, noting that the violation of the Lorenz criterion by the variance of logarithms is not limited to changes in income at the top or extreme upper tail of the distribution but involves 'broad based changes in income up and down the income distribution' (Foster and Ok 1999: 901). They find in an earlier work (Foster and Ok 1997) that the probability of violation lies between 8 per cent and 12 per cent, and is therefore non-negligible, when the variance of logarithms is used to examine microdata. Allison (1978) posits similar findings, noting that the range of choice of inequality measures is significantly limited when the measure is required to satisfy the Lorenz criterion. The only measures left to choose from, as explained by Allison, are the Gini index, the coefficient of variation, and the generalized entropy indices, including the Theil index and mean log deviation. And though these violations have been acknowledged in the recent wage inequality literature (Schaefer and Singleton 2020), the suggested methods are scarcely applied.

Therefore, in addition to the AKM and KSS regression-based analyses presented in this paper, I use the Theil index to decompose overall wage inequality into between-firm and within-firm components. Unlike the regression-based models, this approach does not provide an identification of firm and worker effects on wages. Thus, the within-firm component of wage inequality will include wage inequality within firms due to both firm and worker characteristics. The same is true for the between-firm component. Its advantage, however, is that it is a valid measure of inequality, unlike the variance of logarithms as applied in the AKM and KSS models.

The Theil index compares individual wage income to overall mean wage and is calculated as follows:

$$T = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i}{\mu}\ln\left(\frac{y_i}{\mu}\right) \tag{3}$$

where mean income μ is calculated as:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{4}$$

The Theil index can be decomposed into inequality within groups and between groups as follows:

$$T = \sum_{i=1}^{m} s_i T_i + \sum_{i=1}^{m} s_i ln\frac{\bar{y}_i}{\mu} \tag{5}$$

where $s_i = \frac{N_i}{N}\frac{\bar{y}_i}{\mu}$, T_i is the Theil index for each firm, m is the number of firms or groups of workers, $\bar{y}_i$ is average income for each firm, N_i is the number of workers in each firm and, N is the total number of workers.

The first term in Equation 5 estimates the within-firm inequality component or the portion of wage inequality explained by worker characteristics, while the second term estimates the between-firm component or the proportion explained by firm characteristics.

4 Results

This section presents the results of the regression-based decomposition of individual wages. As noted in Section 3, the AKM and KSS regression-based decomposition framework relies on the principle that the overall variance or dispersion in wages can be separated into the variance of average wages between firms—a firm component—the variance of wages between workers or within firms—a worker component—and twice the covariance of both components. In addition to worker and firm fixed effects which generate these respective components, the estimation here includes year fixed effects and worker's age. As highlighted in Section 3, the SARS-NT data do not include other relevant worker characteristics such as levels of education and experience, race, or geographical location. Hence, the estimations include only workers' age. Similarly, important firm characteristics such as wage-setting and hiring policies are not observable and are therefore absorbed by the firm fixed effects.

Also noted in Section 3 is the fact that the AKM and KSS methods do not allow for an analysis of wage inequality over time, given the violation of the Lorenz criterion. Hence, results from Theil estimations of wage inequality are also presented to complement the regression-based results. The section therefore includes discussion both of the magnitude of worker and firm components and what I will refer to as the implied components of wage dispersion, produced by the AKM and KSS estimation methods, and of the dynamics of wage inequality over time, produced by the Theil index.

4.1 First stage: wage dispersion in the AKM model

The first column of Table 1 shows the results of the AKM and KSS decomposition for all firms over the period 2011–19. The estimated dispersion in real wages is 2.19,[7] indicating that wage dispersion in South Africa is on par with inequality in income. South Africa currently holds the status of the most unequal country in the world with regard to income and, in line with previous studies that estimate wage dispersion, this finding highlights wage dispersion in South Africa as among the highest in both the developed and the developing world. This in unsurprising given that wages account for the greatest proportion of income for most of the country's earners (Bassier and Woolard 2021). For Germany, Card et al. (2013) estimate wage dispersion in the range 0.14–0.25 over the 25-year period from 1985 to 2009.[8] Schneck (2021) similarly finds that wage dispersion in the Netherlands ranges from 0.22 to 0.23.[9] The highest dispersion in wages among advanced economies is estimated by Song et al. (2019) for the USA, who find the variance to be 0.708 for the period 1980–86 and 0.924 in the more recent period 2007–13. Among developing countries, Alvarez et al. (2018) finds a similarly mild variance in wages relative to South Africa—ranging from 0.47 to 0.75.

The rest of this section is therefore dedicated to (a) matching these results against estimates of the Theil index—which accurately measures wage inequality and allows for a discussion of changes over time, and (b) identifying how much of this dispersion is attributable to observable firm characteristics, including firm size, industry, geographical location, and foreign versus domestic status, which have been noted elsewhere as important drivers of wage inequality (Asteriou et at

---

[7] This estimate is reached after removing 'impossibly high values' (Kerr 2020)—wages greater than ZAR1 billion per year—and income is winsorized at the 1 per cent and 99 per cent levels.

[8] Card et al. (2013) reports standard deviations in the range 0.370–0.499. These amounts are squared.

[9] Schneck (2021) reports standard deviations in the range 0.464–0.484. These amounts are squared.

2014; Bhorat et al. 2017; Chatterjee 2016; Jaumotte et al. 2013). The size of each variance component is estimated first using the AKM method and second using the KSS method, which corrects for potential limited mobility bias in the AKM analysis. The KSS output is therefore the preferred result.

Table 1: Variance decomposition results—AKM and KSS

	AKM (2011–19)	KSS (2011–19)	AKM (2011–16)	KSS (2011–16)
Var (lnrealwage)	2.19	2.19	2.23	2.23
Var (firm FE)	18.6%	17.7%	19.5%	18.1%
Var (worker FE)	49.8%	34.6%	53.8%	34.5%
2 × covariance (firm FE, worker FE)	9.6%	11.2%	7.8%	10.3%
Other terms (year FE, age, residual)	22%	36.5%	18.9%	37.1%
Observations (millions)	56.1	56.1	35.2	35.2

Note: this table gives results of the AKM decomposition of log wages (lnrealwage) for all connected firms with more than 20 workers; the first row shows the variance of log wages, and each subsequent row shows the variance of individual components; FE = fixed effects.

Source: author's construction based on SARS-NT administrative tax data (NT and UNU-WIDER 2023a, b).

Figure 1 presents the Theil index, as well as the between- and within-firm components of wage inequality for the group of connected firms.[10] The index hovers around 0.86 for the entire period from 2011–19. This result is in line with previous assessments of wage inequality in the South African labour market that highlight rising wage inequality in the late 1990s post apartheid, followed by stabilization at high levels in the two decades which followed (Kerr 2020; Wittenberg 2017b, c). Wage inequality has remained relatively unchanged despite the implementation of a variety of industry-level minimum wage laws, and more recently a National Minimum Wage (NMW).[11] That is, despite positive impacts on the real wages of the lowest paid and the limited disemployment effects of sectoral minimum wages (Bhorat et al. 2021), wage inequality has remained unwaveringly high.[12] Additionally, while the impact of the NMW cannot be fully assessed here given that it was implemented in January 2019, only two months before the end of the sample, the initial assessment in other parts of the literature point to very limited impacts on wage inequality. Bhorat et al. (2021), who investigated the short-term impact of the NMW, explain that 46 per cent of all employees earned wages below the NMW prior to its implementation. However, despite the requirement for wage increases for this significant proportion of workers, they fnid that the impact of the NMW has been relatively muted, with no substantial increase in hourly wages for most workers and 43.5 per cent of workers still reporting wages below the NMW at the

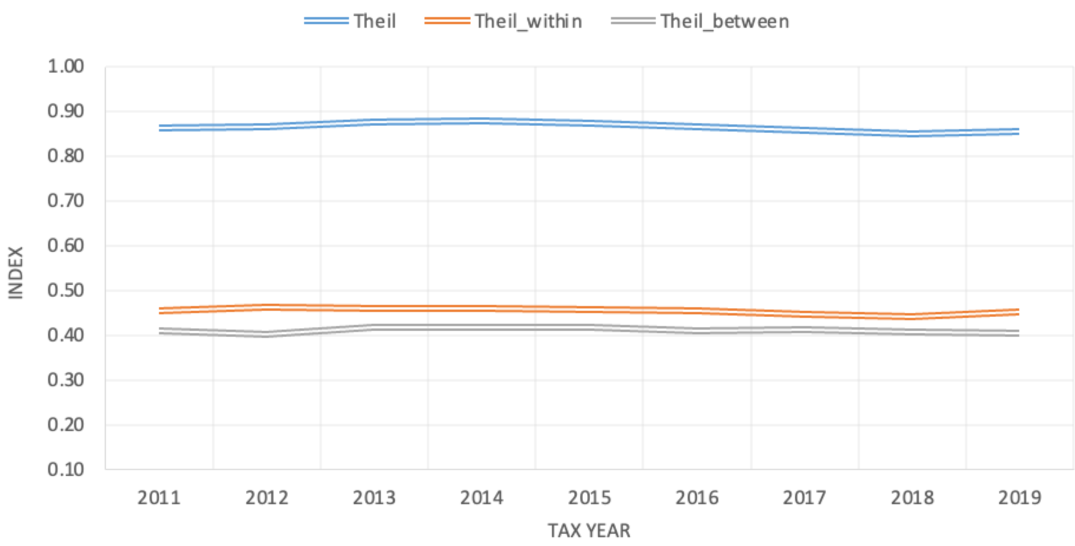
[10] Connected firms are those which workers move between, allowing for identification of the worker and firm components in the AKM and KSS. Estimates of the Theil index for the entire sample of firms shows slightly lower overall wage inequality (average = 0.85) and the same dynamics of both overall inequality and the within and between components.

[11] See Bhorat et al. (2021) for a summary of the sectors in which minimum wages have been implemented in South Africa and the impacts of each.

[12] The only sector that experienced significant disemployment effects subsequent to the implementation of a sectoral minimum wage was the agricultural sector (Bhorat et al. 2021; Piek and von Fintel 2020).

end of 2019. They argue that this minimal wage impact and corresponding absence of employment effects may be attributable to widespread non-compliance with the NMW.

Figure 1: Theil index, between-firm, and within-firm components for connected firms with more than 20 employees, 2011–19



Source: author's illustration based on SARS-NT administrative tax data (NT and UNU-WIDER 2023a, b).

While there are not many comparable estimates of the Theil index of wage inequality for other countries, the available estimates also suggest that South Africa's labour market is characterized by the highest levels of wage inequality in the world. Cardoso (1997) provides early estimates for Portugal of 0.28 for the period 1983–92. She also notes that these estimates are on par with those for the United Kingdom and slightly lower than estimates for the USA, which is the country 'usually taken as the paradigm of an unequal labour market' (Cardoso 1996: 5). More recently, Egger et al. (2020) estimate the Theil index of wage inequality in Germany's export sector as 0.056. Therefore, with an index three times the size of the Portuguese index of wage inequality and 15 times that of one of Germany's most productive sectors, the South Africa labour market certainly stands high in the world ranking of wage inequality.

Firm and worker effects

Also shown in Figure 1 are the between-firm and within-firm components of overall wage inequality. The between-firm component reports the proportion of overall inequality that is attributable to differences in average real wages between firms, while the within-firm component reports the proportion attributable to differences in average wages within firms, or between workers in the same firm. Both components account for almost equal proportions of overall inequality, with the within-firm portion being slightly higher. However, as noted in Section 3, the Theil approach does not provide an identification of firm and worker effects on wages. Thus, the between- and within-firm components might not purely reflect the impact of firms and workers, respectively, on wage inequality. The between-firm component, for example, might include both the impact of differences in firm characteristics and the type of workers that are hired by respective firms. This is evident when compared with the results from the decomposition analysis (Table 1), which does identify firm and worker effects.

The worker component accounts for 34.6 per cent of the overall variance in wages, indicating that worker characteristics are responsible for more than a third of the observed dispersion in wages, in line with the Theil decomposition. This is also in line with previous analysis by Bassier (2023) using data for 2011–16. The finding also corresponds with the emphasis on worker characteristics as the main driver of wage dispersion in the literature that reports on such decomposition analysis for developing countries (Alvarez et al. 2018; Messina 2021; Rodríguez-Castelán et al. 2022).

On the other hand, the firm component identified by the variance decomposition accounts for 18 per cent of the overall variance in wages. Therefore, some of the heterogeneity between firms estimated by the Theil index decomposition is explained by other factors, including the fact that firms do not employ the same types of workers. The covariance of worker and firm fixed effects is positive and accounts for 11 per cent of the overall dispersion, indicating that high-earning workers tend to be clustered in high-wage firms. And if the worker effect reflects skills and education levels, this implies that higher-skilled workers sort to high-wage firms. Card et al. (2013) refers to this as 'assortativeness', which they note accounts for a third of the rise in German wage inequality over the period 1985–2009. Song et al. (2019) note similarly that the composition of workers within firms accounts for increasing proportions of wage inequality. They find that this comes from two effects: (1) high-wage workers are more likely to work in high-wage firms—a sorting effect; and (2) high-wage workers are more likely to work together—a segregation effect. In addition to the firm and worker components, the combination of year fixed effects, age, and the residual account for 36.5 per cent of the overall dispersion in wages.

Of note is the fact that the firm effect estimated here is less than the 28 per cent estimated in Bassier (2023) and closer to the 13 per cent estimated by Bhorat et al. (2017). There is a fundamental difference between the structure of the data used here and that of the data used in Bassier (2023), which accounts for a portion of the difference. As highlighted in Section 2, the disconnect between firms' financial years—which determine the tax year to which CIT returns are attached, and workers' income data as recorded in the IRP5 certificates—requires a realignment of the income data. This realignment is conducted prior to the analysis presented in this paper but was not conducted in Bassier (2023).[13] As shown in Columns 3 and 4 of Table 1, when the estimation is conducted only for the years 2011–16, corresponding with Bassier (2023), the firm component is higher by only 1 per cent. Nonetheless, as emphasized in both of these previous assessments, while the percentage of wage dispersion attributed to South African firms is not as high as that estimated in the developed-country context (Criscuolo et al. 2021; Schneck 2021; Song et al. 2019), firms play a significant role in driving wage dispersion in South Africa.

4.2 Second stage: what explains the firm component?

This section presents the second-stage results based on the estimation of Equation 2, which outlines a regression of the firm fixed effects from the first-stage estimation on individual firm characteristics including firm profits, size, industry, province, and local versus foreign status. The results from Equation 2 therefore provide insight into which of these firm characteristics drive the firm effect or firm wage premia and by extension drive wage inequality. These results are shown in Table 2. Column 1 includes results from the estimation with firm size, industry, province, and local versus foreign status, Column 2 adds interactions between firm size and industry, and Columns 3 and 4 add firm profits.

[13] The data developers at the South African National Treasury and UNU-WIDER are constantly identifying and addressing issues with the tax administration dataset, as these data were not collected for research purposes. Hence, it is possible that at the time of writing, the author was not aware of this issue.

Table 2: Second-stage estimation—firm characteristics

Dependent variable: firm FE	(1)	(2)	(3)	(4)
Log profit per worker			0.052***	0.052***
			(0.001)	(0.001)
Log firm size	0.020***	0.010**	0.066***	0.064***
	(0.003)	(0.004)	(0.004)	(0.005)
Log firm size^2	−0.005***	−0.006***	−0.009***	−0.010***
	(0.000)	(0.000)	(0.000)	(0.000)
Industry:				
Agriculture	−0.479***	−0.273***	−1.073***	−0.922***
	(0.013)	(0.016)	(0.017)	(0.022)
Mining and quarrying	0.267***	−0.369***	−0.363***	−1.015***
	(0.014)	(0.024)	(0.019)	(0.032)
Manufacturing	0.034***	−0.018	−0.560***	−0.570***
	(0.012)	(0.014)	(0.017)	(0.018)
Utilities	0.109***	−0.220***	−0.499***	−0.763***
	(0.017)	(0.042)	(0.021)	(0.051)
Construction	−0.137***	−0.265***	−0.712***	−0.844***
	(0.013)	(0.015)	(0.017)	(0.020)
Wholesale and retail	−0.040***	0.042***	−0.611***	−0.508***
	(0.012)	(0.014)	(0.017)	(0.018)
Transport	0.094***	−0.095***	−0.514***	−0.668***
	(0.013)	(0.019)	(0.018)	(0.024)
Financial services	−0.024*	−0.023	−0.609***	−0.657***
	(0.012)	(0.014)	(0.017)	(0.019)
Community and social services	−0.123***	−0.061***	−0.676***	−0.648***
	(0.012)	(0.015)	(0.017)	(0.020)
Firm size × agriculture		−0.033***		−0.028***
		(0.003)		(0.004)
Firm size × mining and quarrying		0.153***		0.147***
		(0.005)		(0.006)
Firm size × manufacturing		0.028***		0.009***
		(0.003)		(0.003)
Firm size × utilities		0.098***		0.073***
		(0.010)		(0.012)
Firm size × construction		0.046***		0.039***
		(0.003)		(0.004)
Firm size × wholesale and retail		−0.006**		−0.020***
		(0.003)		(0.003)
Firm size × transport		0.060***		0.043***
		(0.004)		(0.005)
Firm size × financial services		0.015***		0.018***
		(0.003)		(0.003)
Foreign firm	0.328***	0.327***	0.228***	0.181***
	(0.004)	(0.015)	(0.005)	(0.018)
Firm size × foreign firm		−0.002		0.009**
		(0.003)		(0.004)
Provinces:				
Eastern Cape	−0.002	−0.005	0.005	0.003
	(0.004)	(0.004)	(0.005)	(0.005)

Northern Cape	0.006	0.007	0.011	0.009
	(0.007)	(0.007)	(0.008)	(0.008)
Free State	0.016***	0.011**	0.017***	0.012**
	(0.006)	(0.006)	(0.006)	(0.006)
KwaZulu-Natal	0.036***	0.035***	0.034***	0.033***
	(0.003)	(0.003)	(0.004)	(0.003)
North West	0.049***	0.045***	0.052***	0.049***
	(0.006)	(0.006)	(0.006)	(0.006)
Gauteng	0.131***	0.130***	0.116***	0.115***
	(0.002)	(0.002)	(0.003)	(0.003)
Mpumalanga	0.092***	0.089***	0.083***	0.080***
	(0.005)	(0.005)	(0.005)	(0.005)
Limpopo	0.020***	0.019***	0.018***	0.015**
	(0.005)	(0.005)	(0.006)	(0.006)
Observations	257,766	257,766	188,193	188,193
Year fixed effects	Yes	Yes	Yes	Yes
R-squared	0.143	0.150	0.179	0.184

Note: this table shows results of Equation 2; base categories: province = Western Cape, industry and firm size interaction = firm size × community and social services; standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Source: author's construction based on SARS-NT administrative tax data (NT and UNU-WIDER 2023a, b).

Firm size and industry

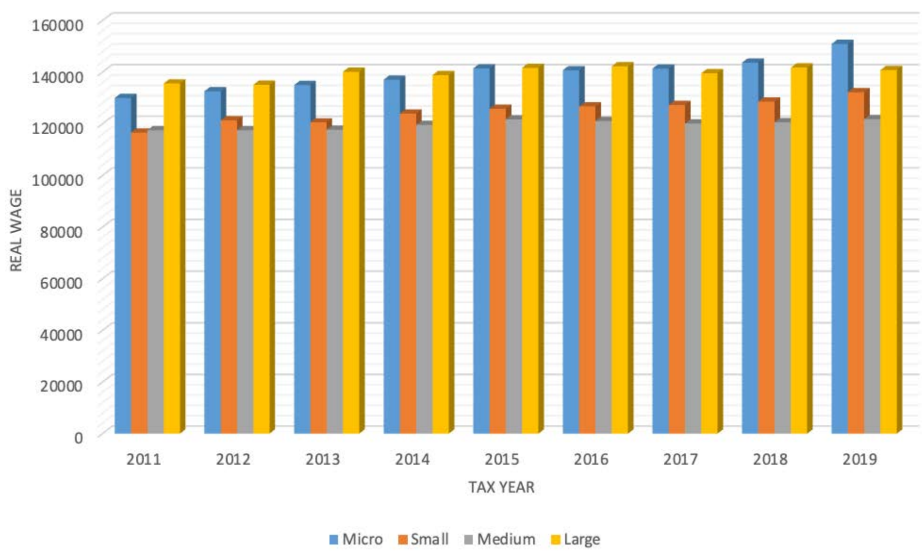
The firm wage premium initially increases with increases in firm size but starts to decrease as firm size reaches a certain threshold. This implies a non-linear relationship between firm size and firm wage premia, and importantly, that the largest South African firms pay less than smaller firms. This is in contrast with evidence from advanced economies that shows firm size as a main driver of firm wage premia (Ohlert 2016; Schaefer and Singleton 2020; Song et al. 2019). It is also different from the available developing-country evidence presented by Alvarez et al. (2018), where larger Brazilian firms pay higher wages.

Bhorat et al. (2017) provide a potential explanation for this surprising outcome in the South African context. They note that the South African labour market is characterized by a U-shaped relationship between firm size and wages. That is, the highest average wages are found among the smallest and the largest firms. The authors further investigate this phenomenon, finding that the small firm wage premium tends to be present in the agricultural industry, driven by 'small-scale, high value and capital-intensive agricultural farms … [with the] low minimum wage to farm workers driving down the average in large firms' (Bhorat et al. 2017: 24–25). Small firm wage premia are also found in the financial services and manufacturing industries, where firms with one to three employees pay higher wages than the largest firms with over 1,000 employees. On the other hand, the large firm wage premium tends to be found in the utilities and mining industries, where the largest firms pay the highest average daily wages. The U-shaped relationship between firm size and wages described by Bhorat et al. (2017) is shown to be true in Figure 2. Figure 3 then presents scatterplots of firm wage premium by firm size in 2011 and 2019, clearly indicating that when other firm characteristics are accounted for, firm size is not a main driver of firm wage premia.

When interactions between firm size and industry are considered in relation to firm wage premia, the results in Column 2 of Table 2 confirm more of the preliminary results outlined in Bhorat et al. (2017). More specifically, while a large firm wage premium exists in most industries, including the utilities and mining industries, it is smaller firms that pay more in the agricultural sector. The wholesale and retail sector similarly exhibits a small firm wage premium while large firm wage premia are found in the financial services and manufacturing industries.

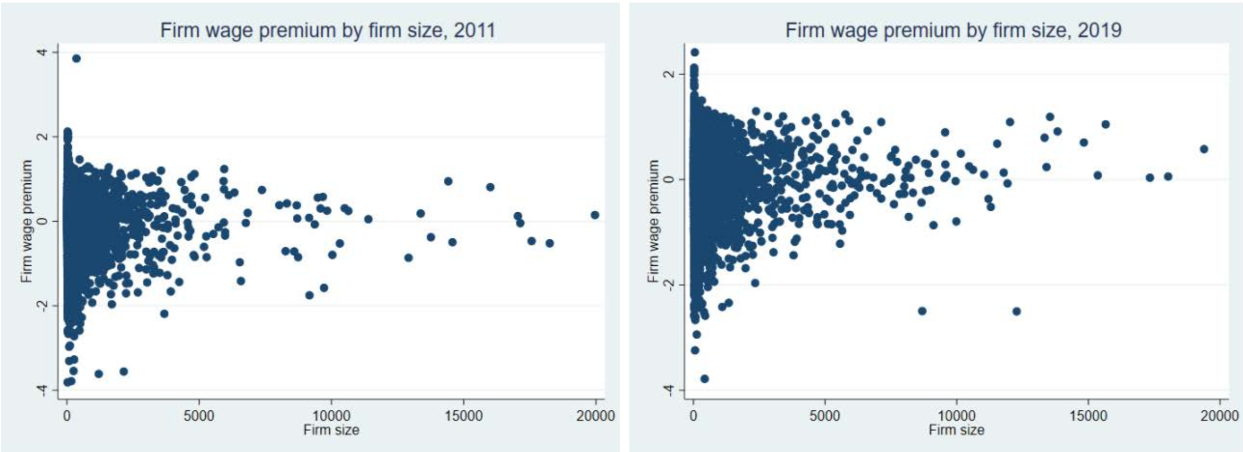
Nonetheless, without considering interactions with firm size (Column 1), on average it is the mining, utilities, manufacturing, and transport sectors that tend to have positive impacts on firm wage premia. These are also the industries with the lowest levels of wage inequality, as estimated by the Theil index and displayed in Figure 4. This indicates that the industries that pay the highest average wages are also those that have the smallest gap between the lowest- and highest-paid workers. Therefore, it is the combined analysis of firm size and the industry that firms operate in that gives us the clearest insights into how these two factors impact on firm wage premia.

Figure 2: Average wages by firm size, 2011–19



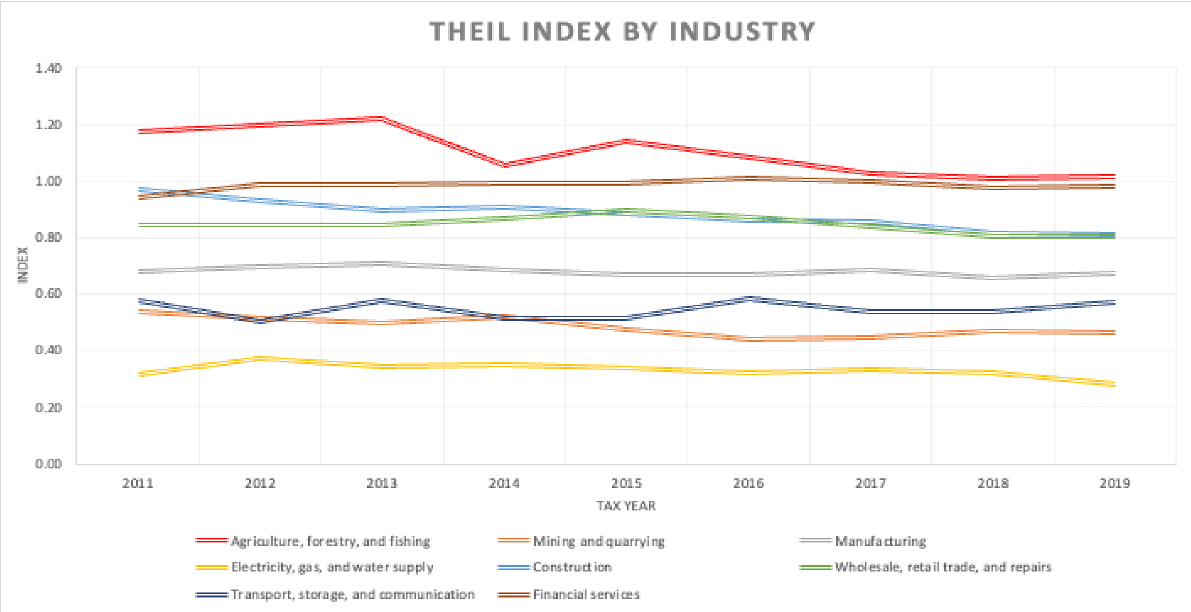
Source: author's illustration based on SARS-NT administrative tax data (NT and UNU-WIDER 2023a, b).

Figure 3: Firm wage premium by firm size, 2011 and 2019



Source: author's illustration based on SARS-NT administrative tax data (NT and UNU-WIDER 2023a, b).

Figure 4: Theil index by industry, 2011–19



Source: author's illustration based on SARS-NT administrative tax data (NT and UNU-WIDER 2023a, b).

The agricultural industry stands out both as one of the only two industries with a small firm wage premium and as the sector with the highest level of and most volatile trend in wage inequality. While wage inequality in all other industries has remained relatively stable over the nine-year period analysed—mirroring the trend in overall wage inequality (Figure 1)—wage inequality in the agricultural sector has declined substantially, first following a 50 per cent increase in the agricultural minimum wage in 2013 (Tan 2021) and further in the 2016 tax year. These declines have brought wage inequality in the agricultural sector in line with that of the financial services sector, which is characterized by the second-highest level of inequality among major industries. At the end of 2014, there remained significant non-compliance with the agricultural sector's prescribed minimum wage (Ranchhod and Bassier 2017), which may help to explain the residual high inequality despite declines following each change. The agricultural sector therefore represents a key area of potential policy focus to address the degree of wage inequality in South Africa.

Local and foreign firms

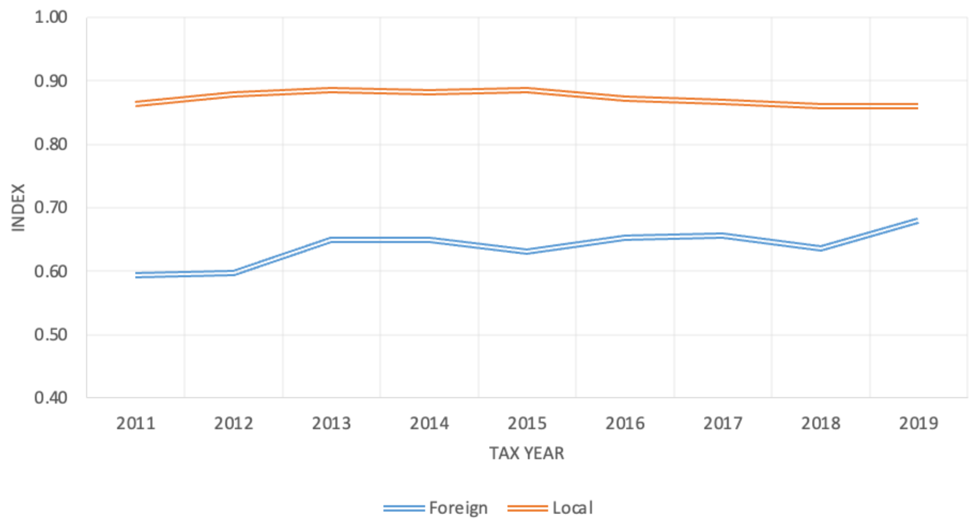
Foreign firms and multinational enterprises (MNEs) are identified in the firm CIT-IRP5 panel using various questions on the tax forms submitted by firms operating in South Africa. Firms are classified first as subsidiaries or associates of foreign companies. Subsidiaries have majority shareholdings (over 50 per cent) held by foreign companies while associates have minority shareholdings (at least 10 per cent) held by foreign companies. (Kilumelume et al. 2021). Subsidiaries whose ultimate holding company is outside South Africa are categorized as *strict* foreign firms, while a *broad* definition of foreign firms is based on the group of subsidiaries plus associate firms. The main analysis here uses the strict definition of foreign firms, as some associate firms are locally controlled despite having part-foreign ownership. This allows for a more precise view of the impact of foreign direct investment (FDI), for example, which is noted as a driver of wage dispersion between foreign and local firms (Chatterjee 2016).

Foreign firms are further classified as MNEs, with either South African or foreign parents. Firms with foreign parent companies are those with ultimate holding companies outside South Africa and are therefore the same as strictly defined foreign firms, while firms with South African parent companies are 'locally held firms with foreign connections'—with at least 10 per cent ownership held outside of South Africa (Kilumelume et al. 2021). Appendix Table A4 lists the number of each type of firm per tax year.

Table 2 also reports the differential impact of strict foreign and local firms on the firm component of wage inequality. The results show that foreign firms have a significantly higher firm wage premium than local firms. This is consistent with assessments in the literature of wage differences between foreign and local firms. Chatterjee (2016) examines wage differences between domestic and foreign firms in the Ghanaian manufacturing sector, finding that foreign firms consistently hire more skilled workers relative to domestic firms, forcing lower-skilled workers to sort to local firms and therefore leading to higher wage inequality between local and foreign firms as foreign firms pay higher average wages (Almeida 2007; Saglam and Sayek 2011).

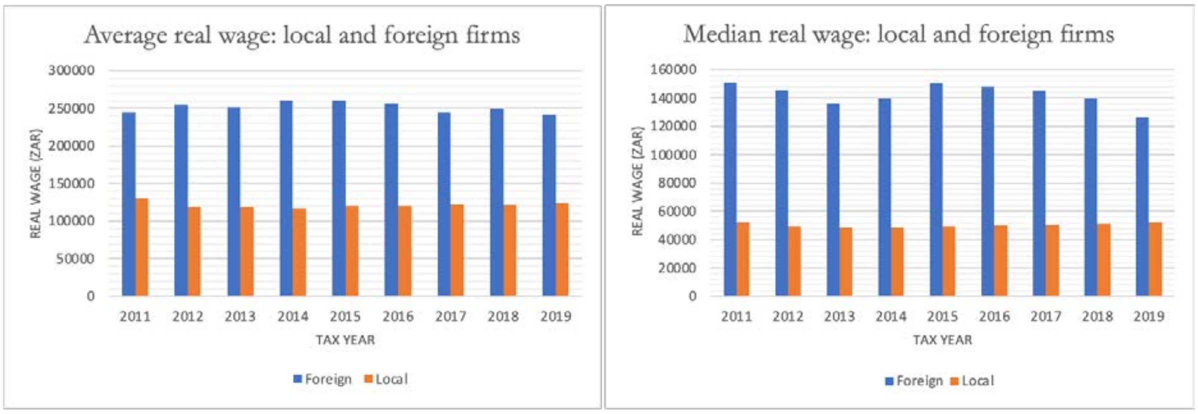
In the South African labour market, despite higher wage inequality within the group of local firms (Figure 5), it is foreign firms that pay higher average wages (Figure 6), in line with predictions in the literature (Almeida 2007; Chatterjee 2016; Saglam and Sayek 2011). While Figure 5 shows high and stable wage inequality in local firms relative to lower but incrementally increasing inequality in foreign firms, Figure 6 shows both higher average and higher median wages in foreign firms over the entire period. The higher wages in foreign firms may be attributable to a sorting effect, where higher-skilled and therefore higher-paid workers are clustered in foreign firms. While the available data include only employees' age and gender, there are no distinct differences in between local and foreign firms average worker age and male/female distribution (see Table A4 in Appendix), indicating that these wage differences may in fact be a result of differences across these two firm types in other worker characteristics, such as experience, race, and education. These factors are noted as being the main drivers of wage inequality in developing countries (Alvarez et al. 2018; Contreras and Gallegos 2011). Higher foreign firm wages may also be indicative of a greater capacity to pay higher wages in foreign-owned firms, given that potential access to more sources of funding, such as FDI boosts to capital from parent companies (Chatterjee 2016).

Figure 5: Theil index—local and foreign firms, 2011–19



Source: author's illustration based on SARS-NT administrative tax data (NT and UNU-WIDER 2023a, b).

Figure 6: Average and median wages—local and foreign firms, 2011–19



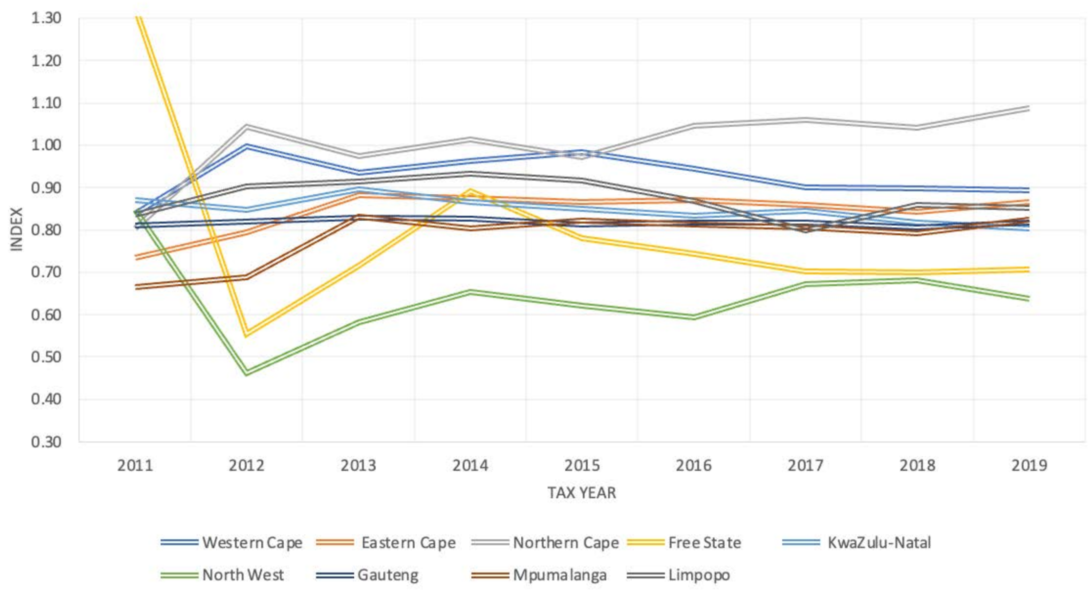
Source: author's illustratuion based on SARS-NT administrative tax data (NT and UNU-WIDER 2023a, b).

Provinces

Across provinces, the firm wage premium varies significantly. First, compared with the Western Cape province, being located in any other province with the exception of the Northern and Eastern Cape has positive and significant impacts on firm wage premia. Among these provinces with a positive effect, operating in Gauteng—the province in which South Africa's capital city and business and administrative hubs are located—has the highest overall impact on the firm wage premium,. Gauteng is also among the provinces with the highest levels of wage inequality, as shown in Figure 7.

The latter outcome may be largely attributed to the migration of high proportions of low-skilled, low-educated individuals into these provinces. During apartheid, laws such as the Native Land Act 1913 prohibited ownership of land by black South Africans, who were forcibly removed to areas known as 'homelands'. These homelands were located predominantly in the Limpopo, Eastern Cape, and KwaZulu-Natal provinces (Burger and Woolard 2005; David et al. 2018). Since the abolishment of the laws in the 1980s and 1990s, prompting free movement across provinces, South Africans have flocked to the metropolitan centres located in Gauteng and Western Cape provinces. Burger and Woolard (2005) note that these two provinces are characterized by the highest urbanization rates and the lowest unemployment rates. Gauteng province had the second-largest absolute and relative increase in population between 1996 and 2001, and 75 per cent of these migrants were black South Africans (Oosthuizen and Naidoo 2004). Moreover, within Gauteng province migrants tend to be less educated and therefore to engage in lower-skilled sectors and occupations, but they are nonetheless more likely to be employed than natives of the provinces to which they have migrated (Oosthuizen and Naidoo 2004). Because the capital city Johannesburg, in Gauteng, is the industrial and technological hub of South Africa, the wage scale tends to be wider—ranging from the lowest-educated migrants in low-skilled jobs to the highest-paid professionals. In addition, the wages of the highest-paid workers in Johannesburg tend to far outweigh the wages of similarly placed workers in other provinces, thus driving firm wage premia.

Figure 7: Theil index by province, 2011–19



Source: author's illustration based on SARS-NT administrative tax data (NT and UNU-WIDER 2023a, b).

Second, the Cape provinces have non-significant impacts on firm wage premia. This is despite the Cape provinces reporting the highest levels of wage inequality in South Africa for the period studied (Figure 7). High wage inequality in the Western Cape is explained by a similar migration phenomenon of low-skilled, low-educated individuals into Cape Town—the province's main metropolitan area and South Africa's tourism enclave. A similarly wide scale from the lowest-paid migrants to the highest-paid professionals and business owners can be found in the Western Cape. The Eastern Cape, on the other hand, houses one of the largest proportions of former homelands and so is still characterized by mostly low-educated black South Africans. Dodd and Nyabvudzi (2014) find that 10.5 per cent of adults over 20 years old in the Eastern Cape have received no education. They find further that among the employed, 11 per cent face food insecurity, indicating that they are employed in low-wage jobs. The province has, however, seen significant growth in high-value manufacturing since the 1990s, introducing a range of higher-paid jobs and thereby widening the wage gap in the province.

In contrast to the findings for industries, these results indicate that higher inequality in provinces is not necessarily correlated with a higher firm wage premium. This is because higher provincial inequality is driven mainly by historical factors which have significantly impacted access to education, training, and jobs for the majority black population located in specific geographical areas, and in turn it has incentivized migration to metropolitan centres, whereas firms in non-metropolitan areas tend to pay higher wages. This further highlights the finding that while firm characteristics have driven wage inequality, worker characteristics, partly determined by these historical factors, still account for almost half of the inequality in wages.

Firm profits

Finally, Columns 3 and 4 of Table 2 show the results of Equation 2 when firm profit per worker is added to the list of predictors. Unsurprisingly, firms with higher profits per worker drive up firm wage premia. Additionally, the inclusion of firm profits leaves the coefficient signs and significance levels for other firm characteristics largely unchanged. This implies first that all of the firm characteristics included here are important in the determination of firm wage premia in South Africa and second, by extension, that each is distinct in its contribution.

These findings on firm profits are in line with those of Bhorat et al. (2017), who report that more profitable South African firms pay higher average wages and that this is evident at all points on the wage distribution. They are also consistent with findings in Bassier (2019) of a 32 per cent increase in wages when the average worker switches from a firm in the 25th percentile for profits to one in the 75th percentile. Bassier (2019: 11) also estimates a rent-sharing elasticity, proxied by firm profits, and finds that 'the firm wage premia increase strongly with profits'. While other firm characteristics are not controlled for in Bassier's estimation, the inclusion of industry and provincial controls confirms that, other factors remaining the same, firm profits have a significant impact on firm wage premia and therefore on overall wage inequality in South Africa.

5 Concluding remarks

This paper investigates the role of firm characteristics in driving wage inequality and firm wage premia in the South African labour market by applying the Abowd, Kramarz and Margolis (AKM) decomposition framework to matched employer–employee administrative tax data for the period 2011–19. Most of the existing literature using matched employer–employee data focuses on developed economies, and the limited developing-economy literature has so far explored only a few firm characteristics. This paper thus expands the scope of the literature by focusing on a developing economy as well as investigating the role of a range of firm characteristics, including firm size, industry, location, foreign versus local status, and profitability. Additionally, given the established criticisms of the variance of logarithms as a measure of inequality, the Theil index is applied as a comparative tool for estimating wage inequality. The paper therefore includes both an assessment of worker and firm components of wage dispersion—using the AKM and KSS regression-based decomposition models—and a dynamic assessment of wage inequality over time using the Theil index.

First, the results show significantly high dispersion in wages, estimated by both the regression decomposition and the Theil methods, reaffirming the extent of high inequality in the country. In line with the results in the literature on income inequality in South Africa, wage inequality has also remained high and stable over the period investigated. The agricultural sector accounts for the highest level of wage inequality among industries, while geographically, the highest level of inequality is found in the Cape provinces. Second, worker and firm characteristics account for 35 per cent and 18 per cent of wage dispersion, respectively, with a positive worker–firm covariance accounting for 11 per cent, indicating that high-earning workers tend to be clustered in high-wage firms. Finally, firm size, industry, profits, geographical location, and whether a firm is locally or foreign-owned are identified as important drivers of firm wage premia. Notably, the industries and geographical locations with the highest levels of wage inequality are not necessarily those with higher firm wage premia, indicating that while the firm wage premium does drive wage inequality, it is not the only factor; there are other factors, such South Africa's history of apartheid and related migration, that drive wage inequality, through their impact on worker characteristics.

References

Abowd, J.M., F. Kramarz, and D.N. Margolis (1999). 'High Wage Workers and High Wage Firms'. *Econometrica*, 67(2): 251–333. https://doi.org/10.1111/1468-0262.00020

Allison, P.D. (1978). 'Measures of Inequality'. *American Sociological Review*, 43(6): 865–80. https://doi.org/10.2307/2094626

Almeida, R. (2007). 'The Labor Market Effects of Foreign Owned Firms'. *Journal of International Economics*, 72(1): 75–96. https://doi.org/10.1016/j.jinteco.2006.10.001

Almeida, R.K., and M. Faria (2014). 'The Wage Returns to On-the-Job Training: Evidence from Matched Employer–Employee Data'. *IZA Journal of Labor & Development*, 3(1): 1–33. https://doi.org/10.1186/2193-9020-3-19

Alvarez, J., F. Benguria, N. Engbom, and C. Moser (2018). 'Firms and the Decline in Earnings Inequality in Brazil'. *American Economic Journal: Macroeconomics*, 10(1): 149–89. https://doi.org/10.1257/mac.20150355

Asteriou, D., S. Dimelis, and A. Moudatsou (2014). 'Globalization and Income Inequality: A Panel Data Econometric Approach for the EU27 Countries'. *Economic Modelling*, 36: 592–99. https://doi.org/10.1016/j.econmod.2013.09.051

Bassier, I. (2019). 'The Wage-Setting Power of Firms: Rent-Sharing and Monopsony in South Africa'. WIDER Working Paper 2019/34. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2019/668-5

Bassier, I. (2023). 'Firms and Inequality when Unemployment Is High'. *Journal of Development Economics*, 161: 103029. https://doi.org/10.1016/j.jdeveco.2022.103029

Bassier, I., and I. Woolard (2021). 'Exclusive Growth? Rapidly Increasing Top Incomes amid Low National Growth in South Africa'. *South African Journal of Economics*, 89(2): 246–73. https://doi.org/10.1111/saje.12274

Bhorat, H., A. Lilenstein, and B. Stanwix (2021). 'The Impact of the National Minimum Wage in South Africa: Early Quantitative Evidence'. Working Paper 202104. Cape Town: Development Policy Research Unit: University of Cape Town, (UCT).

Bhorat, H., M. Oosthuizen, K. Lilenstein, and F. Steenkamp (2017). 'Firm-Level Determinants of Earnings in the Formal Sector of the South African Labour Market'. WIDER Working Paper 2017/25. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2017/249-6

Budlender, J., and A. Ebrahim (2021). 'Estimating Employment Responses to South Africa's Employment Tax Incentive'. WIDER Working Paper 2021/118. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2021/058-0

Burger, R., and I. Woolard (2005). 'The State of the Labour Market in South Africa after the First Decade of Democracy'. *Journal of Vocational Education and Training*, 57(4): 453–76. https://doi.org/10.1080/13636820500200297

Card, D., J. Heining, and P. Kline (2013). 'Workplace Heterogeneity and the Rise of West German Wage Inequality'. *The Quarterly Journal of Economics*, 128(3): 967–1015. https://doi.org/10.1093/qje/qjt006

Cardoso, A.R. (1996). 'Earnings Inequality in Portugal: High and Rising?' European University Institute (EUI) Working Paper ECO 96/1. Florence: EUI.

Cardoso, A.R. (1997). 'Workers or Employers: Who Is Shaping Wage Inequality?' *Oxford Bulletin of Economics and Statistics*, 59(4): 523–47. https://doi.org/10.1111/1468-0084.00081

Chatterjee, S. (2016). 'The Role of the Firm in Worker Wage Dispersion: An Analysis of the Ghanaian Manufacturing Sector'. *IZA Journal of Labor & Development*, 5(1): 1–16. https://doi.org/10.1186/s40175-016-0062-x

Contreras, D., and S. Gallegos (2011). 'Wage Inequality in Latin America: A Decade of Changes'. *Cepal Review*, 103: 27–44. https://doi.org/10.18356/aa36cbb4-en

Coşar, A.K., N. Guner, and J. Tybout (2016). 'Firm Dynamics, Job Turnover, and Wage Distributions in an Open Economy'. *American Economic Review*, 106(3): 625–63. https://doi.org/10.1257/aer.20110457

Criscuolo, C., A. Hijzen, M. Koelle, C. Schwellnus, E. Barth, W.-H. Chen, R. Fabling, P. Fialho, A. Garloff, K. Grabska-Romagosa, R. Kambayashi, V. Lankester, B. Murakozy, O. Nordström Skans, S. Nurmi, B. Stadler, R. Upward and W. Zwysen (2021). 'The Firm-Level Link between Productivity Dispersion and Wage Inequality: A Symptom of Low Job Mobility?' Economics Department Working Paper 1656. Paris: Organisation for Economic Co-operation and Development (OECD).

Csardi, G., and T. Nepusz (2006). 'The igraph Software Package for Complex Network Research'. *InterJournal, Complex Systems*, 1695(5): 1–9.

David, A., N. Guilbert, Y. Hamaguchi, H. Higashi, M. Hino, M. Leibbrandt, and M. Shifa (2018). 'Spatial Poverty and Inequality in South Africa: A Municipality Level Analysis'. SALDRU Working Paper 221. Cape Town: SALDRU (Southern Africa Labour and Development Research Unit), UCT.

Dodd, N.M., and T.G. Nyabvudzi (2014). 'Unemployment, Living Wages and Food Security in Alice, Eastern Cape, South Africa'. *Journal of Human Ecology*, 47(2): 117–23. https://doi.org/10.1080/09709274.2014.11906744

Ebrahim, A., and C. Axelson (2019). 'The Creation of an Individual Panel using Administrative Tax Microdata in South Africa'. WIDER Working Paper 2019/27. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2019/661-6

Ebrahim, A., F. Kreuser, and M. Kilumelume (2021). 'The Guide to the CIT-IRP5 Panel Version 4.0'. WIDER Working Paper 2021/173. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2021/113-6

Egger, H., P. Egger, U. Kreickemeier, and C. Moser (2020). 'The Exporter Wage Premium when Firms and Workers are Heterogeneous'. *European Economic Review*, 130: 103599. https://doi.org/10.1016/j.euroecorev.2020.103599

Foster, J., and E. Ok (1997). 'Lorenz Dominance and the Variance of Logarithms'. Working Paper. New York: CV Starr Center for Applied Economics, New York University.

Foster, J.E., and E.A. Ok (1999). 'Lorenz Dominance and the Variance of Logarithms'. *Econometrica*, 67(4): 901–07. https://doi.org/10.1111/1468-0262.00057

Jaumotte, F., S. Lall and C. Papageorgiou (2013). 'Rising Income Inequality: Technology, or Trade and Financial Globalization?', *IMF Economic Review*, 61: 271–309. https://doi.org/10.1057/imfer.2013.7

Kerr, A. (2020). 'Earnings in the South African Revenue Service IRP5 Data'. WIDER Working Paper 2020/62. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2020/819-1

Kerr, A. (2021). 'Measuring Earnings Inequality in South Africa using Household Survey and Administrative Tax Microdata'. WIDER Working Paper 2021/82. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2021/020-7

Kilumelume, M., H. Reynolds, and A. Ebrahim (2021). 'Identifying Foreign Firms and South African Multinational Enterprises'. Technical Note. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/WTN/2021-1

Kline, P., R. Saggio, and M. Sølvsten (2020). 'Leave-Out Estimation of Variance Components'. *Econometrica*, *88*(5): 1859–98. https://doi.org/10.3982/ECTA16410

Krishna, P., J. Poole, J.P. Poole, and M.Z. Senses (2011). 'Trade Liberalization, Firm Heterogeneity, and Wages: New Evidence from Matched Employer–Employee Data'. World Bank Policy Research Working Paper 5711. Washington, DC: World Bank. https://doi.org/10.1596/1813-9450-5711

Kristal, T., Y. Cohen, and E. Navot (2020). 'Workplace Compensation Practices and the Rise in Benefit Inequality'. *American Sociological Review*, 85(2): 271–97. https://doi.org/10.1177/0003122420912505

Lilenstein, K., I. Woolard, and M. Leibbrandt (2018). 'In-Work Poverty in South Africa: The Impact of Income Sharing in the Presence of High Unemployment'. In I. Marx and H. Lohman (eds), *Handbook of Research on In-Work Poverty*. Cheltenham, UK, and Northampton, MA: Edward Elgar Publishing. https://doi.org/10.4337/9781784715632.00032

Maboshe, M., and I. Woolard (2018). 'Revisiting the Impact of Direct Taxes and Transfers on Poverty and Inequality in South Africa'. WIDER Working Paper 2018/79. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2018/521-3

Messina, J., and J. Silva (2021). 'Twenty Years of Wage Inequality in Latin America'. *The World Bank Economic Review*, 35(1): 117–47. https://doi.org/10.1093/wber/lhz029

Moser, C., F. Saidi, B. Wirth, and S. Wolter (2021). 'Credit Supply, Firms, and Earnings Inequality'. CEPR Discussion Paper 16123. Brussels: CEPR (Centre for Economic Policy Research). https://doi.org/10.2139/ssrn.3597181

NT and UNU-WIDER (2023a). 'CIT-IRP5 Firm-Level Panel 2008–2019 [dataset]. Version 5.0'. Pretoria: South African Revenue Service (producer of the original data), 2019. Pretoria: National Treasury and UNU-WIDER (producer and distributor of the harmonized dataset), 2023.

NT and UNU-WIDER (2023b). 'IRP5 Worker-Level Data 2008–2019 [dataset]. Version 5.0'. Pretoria: South African Revenue Service (producer of the original data), 2019. Pretoria: National Treasury and UNU-WIDER (producer and distributor of the harmonized dataset), 2023.

Ohlert, C. (2016). 'Establishment Heterogeneity, Rent Sharing and the Rise of Wage Inequality in Germany'. *International Journal of Manpower*, 37(2): 210–28. https://doi.org/10.1108/IJM-01-2015-0005

Oosthuizen, M., and P. Naidoo (2004). 'Internal Migration to the Gauteng Province'. Cape Town: Development Policy Research Unit, UCT.

Piek, M., and D. von Fintel (2020). 'Sectoral Minimum Wages in South Africa: Disemployment by Firm Size and Trade Exposure'. *Development Southern Africa*, 37(3): 462–82. https://doi.org/10.1080/0376835X.2019.1702504

Ranchhod, V., and I. Bassier (2017). 'Estimating the Wage and Employment Effects of a Large Increase in South Africa's Agricultural Minimum Wage'. REDI3x3 Working Paper 38. Cape Town: SALDRU, UCT.

Rodríguez-Castelán, C., L.F. López-Calva, N. Lustig, and D. Valderrama (2022). 'Wage Inequality in the Developing World: Evidence from Latin America'. *Review of Development Economics*, 26(4): 1944–70. https://doi.org/10.1111/rode.12912

Saglam, B.B., and S. Sayek (2011). 'MNEs and Wages: The Role of Productivity Spillovers and Imperfect Labor Markets'. *Economic Modelling*, 28(6): 2736–42. https://doi.org/10.1016/j.econmod.2011.06.029

Schaefer, D., and C. Singleton (2020). 'Recent Changes in British Wage Inequality: Evidence from Large Firms and Occupations'. *Scottish Journal of Political Economy*, 67(1): 100–25. https://doi.org/10.1111/sjpe.12225

Schneck, C. (2021). 'Trends in Wage Inequality in the Netherlands'. *De Economist*, 169(3): 253–89. https://doi.org/10.1007/s10645-021-09388-z

Sen, A. (1973). 'Poverty, Inequality and Unemployment: Some Conceptual Issues in Measurement'. *Economic and Political Weekly*, 8(31–32–33): 1457–64.

Song, J., D.J. Price, F. Guvenen, N. Bloom, and T. Von Wachter (2019). 'Firming Up Inequality'. *The Quarterly Journal of Economics*, 134(1): 1–50. https://doi.org/10.1093/qje/qjy025

Tan, B.J. (2021). 'The Minimum Wage and Firm Networks: Evidence from South Africa'. WIDER Working Paper 2021/100. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2021/040-5

Wittenberg, M (2017a). 'Measurement of Earnings: Comparing South African Tax and Survey Data'. SALDRU Working Paper 212. Cape Town: SALDRU, UCT.

Wittenberg, M. (2017b). 'Wages and Wage Inequality in South Africa 1994–2011: Part 1—Wage Measurement and Trends'. *South African Journal of Economics*, 85(2): 279–97. https://doi.org/10.1111/saje.12148

Wittenberg, M. (2017c). 'Wages and Wage Inequality in South Africa 1994–2011: Part 2—Inequality Measurement and Trends'. *South African Journal of Economics*, 85(2): 298–318. https://doi.org/10.1111/saje.12147

World Bank (2021). "South Africa Social Assistance Programs and Systems Review". Policy Brief. Washington, DC: World Bank.

Appendix A: Descriptive statistics

Table A1: Summary statistics—cleaning IRP5

Tax year	Original data	Assignment to correct tax year*	Individuals	Aged 20–60 years	Main job: highest income	Workers: matched	Workers in firms with >20 employees
2011	16,304,393	19,730,082	7,857,265	7,484,449	6,590,357	5,577,887	4,924,547
2012	17,085,352	20,176,322	9,784,269	9,327,983	7,992,458	6,838,445	6,220,095
2013	17,229,167	20,931,052	11,320,385	10,783,019	9,047,889	7,664,147	6,916,245
2014	17,791,554	21,597,739	11,909,177	11,351,185	9,398,277	7,813,002	7,045,994
2015	20,331,932	21,895,225	12,115,489	11,543,395	9,596,521	8,007,201	7,217,008
2016	19,605,155	22,108,709	12,054,698	11,477,895	9,603,716	8,057,876	7,236,587
2017	20,744,947	22,249,397	12,301,276	11,718,526	9,727,326	8,132,430	7,294,800
2018	19,905,482	22,562,591	12,518,464	11,921,972	9,878,168	8,168,674	7,311,602
2019	19,842,744	23,097,416	12,686,355	12,066,438	10,031,193	8,145,648	7,276,442

Note: the table presents the number of worker observations following consecutive limitations on the individual-level sample; * in correcting for incorrectly assigned tax years, the full set of administrative tax data—available from 2008 to 2021—is used; however, the earlier years (2008–10) are excluded from the analysis due to a variety of changes to tax forms prior to 2011, and the later years (2020–21) are also excluded as at the time of writing the UNU-WIDER data team had not concluded populating these data.

Source: author's construction based on SARS_NT administrative tax data (NT and UNU-WIDER 2023a, b).

Table A2: Descriptive statistics—firm-level panel

Tax year	Firms pre-match (freq.)	Firms post-match (freq.)	Firm size (median)	>20 workers (%)	Firm size >20 workers (median)	Dormant (%)	Micro (%)	Small (%)	Medium (%)	Large (%)
2011	757,522	144,667	7	43	24.30	1.19	58.95	30.58	8.41	2.06
2012	824,771	138,034	7	45	26.16	1.64	57.06	30.92	9.54	2.49
2013	848,066	163,975	7	43	25.44	1.61	57.50	31.35	8.92	2.23
2014	876,549	168,157	7	43	25.55	1.69	57.42	31.41	8.94	2.23
2015	904,297	173,761	7	43	25.41	1.72	57.59	31.32	8.89	2.20
2016	997,785	181,852	7	43	24.94	1.83	58.30	30.85	8.74	2.12
2017	1,049,866	186,860	7	43	24.57	1.80	58.94	30.29	8.65	2.13
2018	1,070,575	192,302	7	43	24.22	1.66	59.41	29.98	8.54	2.07
2019	1,006,236	196,931	6	43	23.65	2.27	60.49	29.22	8.28	2.01

Note: firms are designated as micro, small, medium, or large based on number of employees: micro firms are those with less than 10 employees, small firms those with 10–49 employees, medium firms those with 50–249 employees, and large firms those 250 or more employees; dormant firms are so designated by SARS.

Source: author's construction based on SARS_NT administrative tax data (NT and UNU-WIDER 2023a, b).

Table A3: Descriptive statistics—income

Tax year	Workers (freq.)	Age (mean)	Annualized real income (ZAR)				
			p10	p25	p50	p75	p90
2011	5,577,887	35.90	4,954.06	19,910.76	53,444.70	135,602.40	320,937.10
2012	6,838,445	35.80	4,546.39	19,652.06	54,342.09	138,592.80	319,505.30
2013	7,664,147	35.87	4,866.18	20,467.40	55,015.82	140,708.80	329,334.10
2014	7,813,002	35.89	5,336.01	21,550.46	54,912.84	139,231.70	326,875.20
2015	8,007,201	35.94	5,663.88	22,131.28	55,600.88	143,398.80	333,883.60
2016	8,057,876	36.07	5,949.53	22,870.04	56,298.76	145,378.20	332,209.50
2017	8,132,430	36.13	6,094.82	23,073.31	55,909.99	143,710.70	327,957.00
2018	8,168,674	36.24	6,478.34	23,728.81	56,806.97	145,278.00	333,629.90
2019	7,741,598	36.35	6,779.62	24,487.39	56,801.80	144,150.10	331,128.40

Note: income is annualized based on the fraction of the year for which workers are employed in the respective firms; it is further deflated by the consumer price index, base year 2016.

Source: author's construction based on SARS_NT administrative tax data (NT and UNU-WIDER 2023a, b).

Table A4: Foreign firms

Tax year	Foreign firms (broad)	Foreign firms (strict)	MNE (SA parent)	MNE (Foreign parent)
2011	116,919	51,915	8,075	51,915
2012	826,156	537,779	360,847	537,779
2013	1,476,226	788,189	605,462	788,189
2014	1,482,015	806,456	685,645	806,456
2015	1,501,461	827,833	769,865	827,833
2016	1,512,441	863,396	1,163,100	863,396
2017	1,322,229	707,978	1,211,332	707,978
2018	1,392,367	839,048	1,577,992	839,048
2019	1,192,414	726,427	1,597,650	726,427

Source: author's construction based on SARS_NT administrative tax data (NT and UNU-WIDER 2023a, b).

Table A5: Mean age and percentage of males—foreign and local firms

	Foreign firms		Local firms	
Tax year	Mean age	Percentage of males	Mean age	Percentage of males
2011	38.01	67.85	35.87	60.80
2012	37.20	64.91	35.68	59.21
2013	36.98	62.19	35.74	59.38
2014	36.97	61.63	35.76	59.66
2015	36.93	60.71	35.82	59.15
2016	36.88	59.15	35.97	58.75
2017	37.05	60.53	36.04	58.07
2018	37.14	58.98	36.13	57.68
2019	36.88	58.79	36.29	57.14

Source: author's construction based on SARS_NT administrative tax data (NT and UNU-WIDER 2023a, b).

Table A6: Hires and separations

Tax year	% hires	% E-E hires	% N-E hires	% separations	% E-E separations	% N-E separations
2011	0	0	0	36.11	12.32	23.80
2012	47.89	10.04	37.85	32.06	12.65	19.40
2013	39.38	11.29	28.09	31.71	12.44	19.26
2014	33.00	12.21	20.80	30.53	12.94	17.59
2015	32.21	12.62	19.59	31.59	13.32	18.28
2016	32.02	13.23	18.79	31.93	13.43	18.51
2017	32.56	13.30	19.25	31.08	12.65	18.44
2018	31.39	12.59	18.80	31.67	12.41	19.27
2019	31.47	12.44	19.03	0	0	0

Note: the table shows the percentages of hires and separations year-to-year; E-E hires/separations refers to employment-to-employment hires/separations—that is, employees who move between firms; N-E hires/separations refer to hires from non-employment or separations into non-employment.

Source: author's construction based on SARS_NT administrative tax data (NT and UNU-WIDER 2023a, b).

Table A7: Employment-to-employment hires and separations, firms with <20 employees

Tax year	% hires	% E-E hires	% N-E hires	% separations	% E-E separations	% N-E separations
2011	0	0	0	5.15	1.27	3.89
2012	3.80	0.84	2.96	2.94	1.02	1.93
2013	4.42	0.99	3.43	2.99	1.04	1.95
2014	3.34	1.07	2.27	2.86	1.03	1.83
2015	3.19	1.09	2.10	2.89	1.03	1.86
2016	3.35	1.16	2.19	3.07	1.05	2.01
2017	3.29	1.17	2.13	3.05	1.04	2.01
2018	3.35	1.17	2.19	3.30	1.03	2.27
2019	3.56	1.20	2.35	0	0	0

Note: the table shows the percentages of hires and separations year to year for firms with <20 employees; E-E hires/separations refers to employment-to-employment hires/separations—that is, employees who move into or out of firms with <20 employees; N-E hires/separations refer to hires from non-employment or separations into non-employment.

Source: author's construction based on SARS_NT administrative tax data (NT and UNU-WIDER 2023a, b).

Appendix B: Data

This data appendix[14] is created based on the requirements of UNU-WIDER for users of the National Treasury Secure Data Facility (NT-SDF). It provides an overview of the data directly used for the results presented in this paper.

Data access

The data used for this research were accessed from the NT-SDF. Access was provided subject to a non-disclosure agreement and all output was reviewed by NT-SDF staff to ensure that the anonymity of firms and individuals was not compromised. The results in this paper do not represent any official statistics (NT or SARS). Similarly, the views expressed in in this paper are not necessarily the views of the NT or SARS.

Data used: CIT-IPR5 panel (citirp5_v5) and year-by-year IRP5 job-level data (v5). Date of first access for this project: 30 January 2023. Last accessed: 31 August 2023.

Software

The analysis was conducted primarily using Stata 18. User-written programs used for estimations include reghdfe (Correia 2014). Analysis for variance decompositions was also conducted using R version 4.1.2. Packages used include igraph (Csardi and Nepusz 2006) and foreign.

Variables

Variables used from the IRP5 dataset include: taxyear, taxrefno, payereferenceno, UID, dateofbirth, gender, idno, passportno, periodemployedfrom, periodemployedto, totalperiodsinyearofassessment, and totalperiodsworked. Wage income was created from the following IRP5 amount codes: amt3601, amt3605, amt3606, amt3608, amt3615, amt3616, amt3617, amt3701, amt3717, amt3810, and amt3906. Variables used from the CIT-IRP5 data include: taxyear, taxrefno, FID, y_np, y_nl, comp_prof_sic5_1d, ITR14_c_foreign_strict, and ITR14_c_mne_type c_province.

Cleaning and sample notes

The IRP5 data include the full amount of income earned by the worker during the tax year and the CIT data hold firm information for the firm's financial year, which is then assigned to a particular tax year based on the firm's financial year-end. Hence, both datasets are realigned to ensure that income values cover the same period. The IRP5 sample is then limited to data for individuals only, excluding information on partnerships, associations, etc. Third, individuals outside the age range 20–60 years are excluded, to ensure that the results capture only typical working-age individuals. Fourth, the IRP5 sample is restricted to one job per individual; the job from which the worker earns their highest income is regarded as their main job. Finally, entries with 'impossibly high values' (Kerr, 2020)—wages greater than ZAR1 billion per year—are removed and the income is winsorized at the 1 per cent level. CIT entities are limited to those with

[14] The outline of this appendix has been adapted from Budlender and Ebrahim (2021).

at least 20 workers on average over the sample period, in line with restrictions related to wage decomposition regression analyses in the literature.