

# Wage inequality within and between firms in South Africa

Aroop Chatterjee\*      Andrew Kerr†

January 31, 2018

## Abstract

Research on earnings inequality in South Africa has almost entirely used household survey data. This work has shown the earnings inequality is extremely high and has remained high or even increased in the post-Apartheid period. However it does not shed light on some important processes generating inequality, particularly the extent to which inequality in earnings is driven by inequality in average earnings within and between firms. In this paper we use the South African Revenue Service (SARS) tax data for South Africa, to document wage inequality, and to calculate the within and between firm contributions to overall inequality.

One measure for describing earnings inequalities is the variance of the log of earnings. Song et al. (2015) made an important contribution to the study of inequality by documenting, and attempting to explain, changes in the contributions of within and between firm inequality. We follow their procedure in this paper to decompose overall earnings inequality into within and between firm components and thus to measure their relative contributions to overall earnings inequality.

---

\*South African National Treasury

†DataFirst, University of Cape Town, andrew.kerr@uct.ac.za

# 1 Introduction

South Africa is a country with extremely high levels of income inequality. Much of this is due to inequality in earnings for those in the labour market. Leibbrandt et al. (2010) used a variance decomposition to show that 85% of overall income inequality is caused by earnings inequality in the labour market, and that of this, one third is due to the large number of those not working, and two-thirds is due to earnings differences between those in employment. Wittenberg (2017b) has investigated changes in earnings inequality over the post-Apartheid period, finding that, as measured by the Gini coefficient, earnings inequality increased in the 1990s and stabilised at a very high level.

Research on earnings inequality in South Africa has focused mainly on household survey data, which has become ubiquitous since the 1993 PSLSD conducted by SALDRU and the public release of household survey microdata from surveys conducted by Statistics South Africa. However no work on earnings inequality has focused on the role of firms in generating inequality in earnings in South Africa. For example an important question is whether inequality in earnings is the result of large average differences in earnings between firms, so that which firm a worker works for is very important, or because within all firms there is a high degree of inequality between well and poorly paid workers, or a situation in between.

In this paper we provide a first look at the relative importance of within and between firm inequality in contributing to the extremely high levels of inequality in South Africa, using matched firm and worker data provided by the South African Revenue Service (SARS), following methods suggested by Song et al. (2015).

The paper is organized as follows. We first present a review of the literature on inequality in South Africa, as well as research that explores the extent of between and within firm inequality. We then describe the SARS tax data we use in the paper and describe the extent of inequality using Ginis and percentile ratios, before describing and implementing a variance decomposition method that breaks down overall inequality into within and between firm components, explore how the relative contributions of the within and between components have changed, and finally conclude.

## 2 Literature Review

### 2.1 South African Inequality

Leibbrandt et al. (2010) estimated that 85% of overall income inequality is due to earnings inequality in the labour market, and that of this, and two-thirds is due to earnings differences for those in employment. This suggests that inequality in earnings is a very important part of the puzzle in understanding inequality in South Africa. The multiple household-based surveys that have been conducted since 1994 (October Households

surveys followed by Labor Force Surveys and then Quarterly Labor Force Survey) allowed Wittenberg (2017b) to study changes in inequality in earnings. As Wittenberg (2017b) and Wittenberg (2017c) show, numerous methodological questions arise from these surveys but after taking these into consideration inequality as measured by the Gini coefficient remain high. The Gini in earnings increased from 0.46 at the end of 1994 to 0.55 in 1998q4 and then gravitated around this value - making South Africa one of the most unequal country in the world, while some other developing countries, such as Brazil, experienced a downward trend in inequality (Benguria 2015).

More recent work by Wittenberg (2017a) has used SARS tax data to examine inequality. This data was a 20% sample of assessed tax records and is essentially the top approximately 20% of the income distribution. Wittenberg (2017a) used the data to explore the extent of under-reporting in the household surveys conducted by Statistics South Africa, finding that there is fairly substantial under-reporting in the QLFS, particularly close to the top of the earnings distribution, and that the Gini coefficient is probably understated by around 3 percentage points in the household surveys if we believe the tax data.

## **2.2 Inequality within and between firms**

The first empirical work estimating inequality within and between firms was undertaken by Davis and Haltiwanger (1991). The authors used a firm survey to estimate the variance of earnings between firms and an household survey to estimate the total variance of wages, showing that these two pieces of information could then be used in a variance decomposition to estimate the variance of earnings within firms, even though they had no data on individual earnings within firms. This research was motivated by the observation that workers' characteristics did not explain very much of the variance in earnings.

Davis and Haltiwanger (1991) found that around 50% of earnings inequality, as measured by the variance in log earnings, was generated by differences between firms. Chennells and Reenen (1998) found that only about 26% of the overall variation in earnings in the UK was due to differences between firms whilst Cardoso (1999) found that between 66% and 59% of the variance in earnings in Portugal was due to between firm differences. Thus there seem to be large differences across countries in the extent to which between firm differences contribute to overall variance in earnings.

One of the issues in using household survey data to estimate the total variance in earnings is that this earnings may be underestimated due to not capturing the extremely high earners who will almost certainly be missed. This would lead to underestimating total variance and thus overestimating the contribution of within firm differences, which is calculated as a residual in the studies mentioned above. This is one reason that, as matched firm-worker data has been made available, it has been used to directly estimate

the contribution of within firm inequality to overall inequality, rather than estimate it as a residual after estimating overall and between firm inequality, as in Davis and Haltiwanger (1991).

Lazear and Shaw (2009) give estimates from matched firm-worker data in a number of EU countries and the USA. They find that the between-firm contribution to overall variance in log earnings is around 20-40% in the countries they study. Song et al. (2015) used matched firm and worker data from the US to examine the changes in the relative importance in between firm and within firm contributions to overall dispersion in earnings. They find that between firm differences have become larger, but that within firm differences have been stable. We follow the methods of Song et al. (2015) in this paper.

Benguria (2015) explores the evolution of earnings inequalities in Brazil from 1999 to 2013. He shows that both the variance of log wages and the part explained by the difference between firms decrease during the period, but that it was an average of 66%, much higher than for the sample of countries in Lazear and Shaw (2009). The comparison between Brazil and South Africa has been studied by Lam et al. (2015). These two countries were competing for the title of the most unequal country in the world, but Brazil saw its level of inequality decreasing during the past decade, while no such trend was observed in South Africa. This is another reason to examine the structure of inequalities on the firm side in South Africa.

### **2.3 Inequality between firms in South Africa**

The ILO 2016/2017 Global Wage report (ILO 2016) focuses on inequality between and within firms. Most of the analysis of between firm wage inequality is on European countries but there is some analysis for a few developing countries, including South Africa. The firm data that was used for South Africa was the Survey of Employers and Self-employed (SESE) conducted by Statistics South Africa in 2013. This is actually a survey of non-VAT registered firms that are identified through a 30 000 household survey - where the owners of such firms are identified and a follow up survey of them is conducted. These are thus mostly unregistered, informal firms (see Fourie and Kerr (2017) for further details) and are thus relatively small - nearly 80% are own account workers and the total employment in the firms (employees and owners) is around 2.4 million, when total employment was around 15 million in 2013. This is thus not really a useful sample to explore inequality in average earnings between firms. Nevertheless, the ILO report uses this data to estimate various measures of inequality in the average earnings between firms. One important finding is that the P90/P10 ratio of average earnings in these firms was 12 in South Africa, much higher than in Chile (2), Vietnam (8) and a number of developed countries where the ratio was generally between 2 and 5. In the analysis below we compare our findings with those in the ILO report.

### 3 SARS IRP5 tax certificate data

In order to estimate the contributions of within and between firm variation in earnings to overall earnings we use data from the South African Revenue Service (SARS). This is matched firm and worker data collected as part of the administration of taxation in South Africa. The data contains information from the IRP5 certificates issued to individuals. Any individual earnings more than R2000 per year who works in a firm registered for Pay as you earn (PAYE) tax is issued an IRP5 certificate. Kerr (2017) estimated that around 70% of all employed individuals were included in this data for the tax years 2011-2014. In this paper we use data from the 2011-2016 tax years.

The tax certificate includes an anonymised version of the company tax number of the firm (if it has one) and the PAYE number. This allows us to create a matched firm-worker dataset, as undertaken by Kerr (2017). The data includes the annual earnings of the worker, the time period employed in the firm and information on the structure of earnings (including contributions to pensions, health insurance, tax deducted etc). In theory age and sex are recoverable from the ID number but the data we have is anonymised and thus far SARS has not extracted this data for researchers.

The data used in this paper should be distinguished from the data used by Wittenberg (2017a). In his paper Wittenberg (2017a) used a 20% sample of individuals assessed for income tax. Roughly speaking, this is data on individuals who paid income tax and thus covers a smaller and richer group of individuals than the IRP5 data used in this paper, although the top 472 earners were excluded, and instead summary information on these earners was provided. Self-employed individuals were also included in the data used by Wittenberg (2017a), whereas in the IRP5 they would only be included if they paid themselves a salary and thus received an IRP5 tax certificate.

The benefits of using the IRP5 data are that it includes earnings data for all employees, there are employer id numbers, there is no top or bottom coding of income, we have the start and end dates of employment during the tax year and it covers all tax registered firms and all workers in these firms earning more than R2000 per year. One downside is that there is no data on self-employed incomes, unless owners paid themselves a salary and received an IRP5 certificate. The other is that the information on the time period of employment during the tax year has indications that it is measured with error. See Kerr (2017) for further details.

### 4 Methods

In this paper we make two important contributions. Firstly, we document the extent of inequality in the IRP5 data using a number of different earnings inequality measures, showing how these have changed over

the 6 years of data which we have. We document changes in income percentile ratios, including the 90/10, 90/50, 10/50 ratios, as well as the Gini coefficient and the variance of log earnings. We also do this by firm size and industry. We also undertake some comparisons of the income tax data with the Quarterly Labour Force Survey (QLFS)- the labour force household survey conducted by Statistics South Africa, following Wittenberg (2017a), but we are able to do the comparison for a much larger part of the labour income distribution.

We follow Song et al. (2015) in describing the extent of inequality within and between firms and how this has changed over time using three different methods.

Firstly, following Song et al. (2015)  $y_t^{i,j}$  is the earnings of worker  $i$  in firm  $j$  in year  $t$ . We can write  $y_t^{i,j} \equiv \bar{y}_t^j + [y_t^{i,j} - \bar{y}_t^j]$ . In this equation  $\bar{y}_t^j$  is the average earnings paid in firm  $j$ . It is then possible to express the decomposition of variance as

$$var_i(y_t^{i,j}) = var_j(\bar{y}_t^j) + var_i(y_t^{i,j} | i \in j)$$

The second method of describing the changes in inequality, particularly within and between firms is to plot changes in percentiles, changes in average earnings for the firms employing workers in each percentile and finally the relative change in earnings for individuals in each percentile relative to their co-workers.

The final method follows Juhn et al. (1993) in looking at inequality across percentiles.

## 4.1 Robustness checks

We are using a dataset that includes only workers who both are reported to have been working for a firm for at least 3 months, and who must have earned at least the earnings equivalent to 3 months at the minimum wage in domestic work- around R2200 a month or R6600 for 3 months. This is to exclude individuals with very limited attachment to the labour market, and some strange anomalies in the data, such as a number of individuals reported to have earned only R1 in the year. But we will also check to see whether our results are robust to including low earners.

## 5 Conclusion

In this paper we plan to use tax data from the South African Revenue Service. This data, obtained from the IRP5 certificates issued to every employee in a tax registered firm that earned more than R2000 per year, covers around 70% of employment in South Africa and has firm identifiers that have enabled us to create a matched firm-worker dataset. Firstly we will describe the extent of inequality in the data. We will then

use the data to estimate the within and between firm contributions to overall inequality, as measured by the variance of log earnings.

## References

- Benguria, F.**, “Inequality Between and Within Firms: Evidence from Brazil,” 2015.
- Chennells, Lucy and John Van Reenen**, “Establishment Level Earnings, Technology And The Growth Of Inequality: Evidence From Britain,” *Economics of Innovation and New Technology*, 1998, 5 (2-4), 139–164.
- Davis, S. and J. Haltiwanger**, “Wage Dispersion Between and Within U.S. Manufacturing Plants 1963-86,” *Brooking Papers : Microeconomics*, 1991.
- Fourie, Fredrick and Andrew Kerr**, “Informal sector Employment Creation in South Africa: What can the SESE Enterprise Survey tell us,” REDI Working Paper Series 32, REDI, University of Cape Town 2017.
- ILO**, “Global Wage Report 2016/2017, Wage Inequality in the workplace,” Technical Report, ILO 2016.
- Juhn, Chinhui, Kevin M Murphy, and Brooks Pierce**, “Wage Inequality and the Rise in Returns to Skill,” *Journal of Political Economy*, June 1993, 101 (3), 410–442.
- Kerr, Andrew**, “Job Flows, Worker Flows And Churning In South Africa,” *South African Journal of Economics*, 2017.
- Lam, D., A. Finn, and M. Leibbrandt**, “Schooling Inequality, Returns to Schooling and Earnings Inequality,” *Wider Working Paper 050*, 2015.
- Lazear, E. and K. Shaw**, “Wage Structure, Raises and Mobility: An Introduction to International Comparisons of the Structure of Wages Within and Across Firms,” in E. Lazear and K. Shaw, eds., *The Structure of Wages: An International Comparison*, University of Chicago Press, 2009, chapter 1.
- Leibbrandt, M., I. Woolard, A. Finn, and J Argent**, “Trends in South African Income Distribution and Poverty since the Fall of Apartheid,” *OECD Social, Employment and Migration Working Paper*, 2010, 101.
- Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter**, “Firming Up Inequality,” NBER Working Papers 21199, National Bureau of Economic Research, Inc May 2015.
- Wittenberg, Martin**, “Measurement of earnings: Comparing South African tax and survey data,” Technical Report, REDI Working Paper 2017.

- , “Wages and Wage Inequality in South Africa 19942011: Part 1 Wage Measurement and Trends,” *South African Journal of Economics*, June 2017, 85 (2), 279–297.
- , “Wages and Wage Inequality in South Africa 19942011: Part 2 Inequality Measurement and Trends,” *South African Journal of Economics*, June 2017, 85 (2), 298–318.