Estimating Job and Worker Flows using South African Revenue Service IRP5 data

Andrew Kerr and Martin Wittenberg, DataFirst, University of Cape Town[1].
21 August 2015

Paper for the Economics Society of South Africa conference.

WORK IN PROGRESS- PLEASE DO NOT CITE

Introduction

 In this paper we describe the extent of gross worker flows and gross job flows using South African Revenue Service firm level data, shedding light on labour reallocation and the policy environment in which South African firms operate. This research is important because although South Africa's unemployment problem is well known, labour demand and labour reallocation are not well understood at all. The main reason for this is because to undertake analysis of these issues researchers requires access to firm level panel data and this type of data has been extremely scarce up until this point in South Africa.

The South African Revenue Services matched firm and worker data allows for the possibility for estimating job and worker flows. These data are a census of all tax paying firms and all workers in these firms earning more than R2000 per year. South Africa has a relatively small informal economy relative to many other developing countries (Magruder, 2012) and as a result the tax data cover firms accounting for around 80% of all employment in the country, a relatively high proportion for a developing country.

 There are a number of important policy relevant questions that can be answered by an analysis of job and worker flows. Understanding which types of firms are creating employment and which are not helps to shed light on what regulations are working and which are not, and why it may be that South Africa has far fewer small firms than comparator economies (Magruder, 2012). Low levels of job and worker flows can indicate a rigid labour market (Haltiwanger et al 2008), so estimating how high these flows are deepens our understanding of whether labour market regulations are constraining firm growth and employment. The OECD (2010) notes that labour reallocation is a key

driver of productivity growth- whilst Hsieh and Klenow (2009) estimate that correcting misallocation of labour and capital could raise TFP by 30-50% in Chinese manufacturing firms. One of the ways to achieve this is to have labour market institutions that allow labour to be reallocated to more productive firms. This will show up as larger rates of job and worker reallocation.

Literature Review

To understand labour reallocation we need to understand both the search and matching of firms and workers (worker flows) as well as the job creation and destruction patterns of firms (job flows) (Burgess et al 2000). Job creation and destruction help us to understand which firms are growing and which are not, as well as which industries or types of firms have higher levels of reallocation (Davis et al 1996). Job reallocation statistics can also help policy makers to know if the firms they hope will be drivers of employment actually have any prospects of doing so given their past net job creation rates.

Job creation and destruction rates are only a partial picture of labour reallocation, however, since a firm that hires 10 new employees and simultaneously fires another 10 has no measured job reallocation. A fuller picture emerges from studies of both worker and job flows, illustrated in the pioneering work of Burgess et al (2000). This work and subsequent studies have shown that worker reallocation over and above job reallocation is very high in a number of countries.

Job and Worker Flows

Below we set out more fully the concepts of worker and job flows and then discuss how they can be estimated using the South African Revenue Service firm level data.

Job Flows

As discussed by Burgess et al (2000), job flows measure the gross creation and destruction of jobs, or the increasing or decreasing size of firms:

$$JF_{it} = E_{it} - E_{it-1} = H_{it} - S_{it}.$$

Job reallocation is defined as the absolute value of job flows and job creation in a firm is a positive job flow whilst job destruction in a firm is a negative job flow.

Job flows have been studied in detail using firm level panel data for nearly 30 years. The key piece of research is the work of Davis et al (1996) for US manufacturing. The authors showed a number of important results, such as that small firms did not have higher net job creation rates than large firms, though small firms did have much higher rates of job reallocation- many small firms are entering or growing at a particular point in time but many are also shrinking or exiting. Another key result was that a substantial amount of job destruction and creation came from big contractions or expansions,

such as the deaths and births of firms. The research programme focused on job creation and destruction has subsequently expanded dramatically. A recent OECD study (Criscuolo et al, 2014) describes comparative work on 18 countries, outlining the role of small firms in job creation, the impact of the financial crisis on entry, exit and job destruction as well as a number of other policy relevant issues.

There are now also studies in African countries exploring job creation and destruction, including in Ivory Coast (Klapper and Richmond 2011), Ghana (Sandefur 2010) and Ethiopia (Shiferaw 2009). The most recent African study uses South African data (Kerr et al 2014). The authors use the Quarterly Employment Statistics firm survey undertaken by Statistics South Africa to look at job creation and destruction in South African firms. This work produced descriptive evidence on job reallocation at the firm level, showing for example that small firms had the lowest net job creation rates, which were actually negative, whilst the largest firms had the highest net job creation rates, which were positive. Very interestingly, this result does not hold for other countries, for example Davis et al (1996) show that there is no relationship between firm size and net job creation rates in US manufacturing.

Worker Flows

We can distinguish job flows from worker flows, where worker flows are defined as the sum of hires and separations:

$WF_{it}=H_{it}+S_{it}.$

Job flows are also defined as hires minus separations. It is then possible to link job flows, worker flows and job reallocation so that

$WF_{it} = JR_{it} + CF_{it},$

where CF_{it} is excess worker flows or churning, a residual component of worker flows above those resulting from job reallocation.

Huber and Smeral (2006) note that two different measuring concepts are mainly applied to measure worker flows. With reallocation measures worker flows can be defined as "the number of persons whose place of employment differs between t-1 and t" (Huber and Smeral, 2006:1689). Turnover measures define worker flows as "the number of accessions plus the number of separations that occur from t-1 to [t]." The key difference between the two measuring concepts is how short term spells are treated. Using turnover measures short term spells that start and end between t-1 and t are not counted but they are counted using reallocation measures.

The SARS data potentially allows for the use of both measuring concepts because each tax certificate issued records the date employed from and employed to in the tax year. However in this paper we use a turnover measure- asking how many workers have changed employers between the start of one tax year and the next. Thus we will undercount job flows by missing spells of employment that last less than one year. We hope to improve on this is future versions of the paper both by using a turnover definition with quarterly data and a reallocation definition.

The international literature suggests that focusing on job flows masks large worker flows (Kaplan et al 2007). High turnover of workers implies increased risks to these workers and suggests that policy makers should be concerned about better understanding worker flows at the firm level. In addition, understanding the correlates of worker and job flows can suggest specific policy reforms to aid employment growth, a key issue in South Africa.

As well as documenting worker flows, which has not been done before in South Africa, we can also use the SARS firm data as a check on the job flow estimates of Kerr et al (2014), who noted two important limitations of the Statistics South Africa Quarterly Employment Statistics (QES) survey data they used. The first of these was that births were not well captured in the QES data. This was due to the nature of the QES panel- once a sample was selected no new firms entered the sample and the same sample was used for up to four years. The second limitation was that small firm job reallocation was not measured whenever a new sample was drawn for the panel survey. These limitations meant some uncertainty about the conclusion of the authors that small firms had the lowest rates of net job creation.

SARS IRP5 Data

In the process of collecting income taxes from workers SARS requires all firms registered for PAYE (pay as you earn) tax to issue tax certificates to all employees, the IRP5 certificate. Any firm that has employees is required by law to register for PAYE. Since 2009 firms have been required to issue IRP5 certificates to all employees earning more than R2000 per annum. These certificates contain data on the dates during which the person was employed in the tax year, the source and amount of income earned and the firm's PAYE reference number. This means that it is possible to construct data on worker and job flows from the IRP5 certificates for each firm over the six years from 2009 to 2014. These flows will include almost all workers in tax registered firms except those earning less than R2000 per annum.

The firm identifier in the IRP5 data is the PAYE (pay as you earn) reference number. In discussions with SARS it was noted that firms may shift employees between different payrolls, as some firms

have multiple payrolls and thus multiple PAYE reference numbers. We thus aggregated up to the company income tax number and all our analysis is done using this firm identifier. Thus our analysis is not plant-level analysis but enterprise-level analysis.

To create a worker identifier we have used an anonymised South African identity number provided with the data by SARS. Unfortunately in the latest version of the data we had access to this anonymised id number was not present in 2010. We have thus used data from tax years 2011-2014 in this version of the paper. In future versions we will incorporate 2009 and 2010 data into the analysis when we are able to access an improved version of the 2010 data.

The work we can do using the SARS data estimating worker flows cannot be undertaken with other sources of data in South Africa. Statistics South Africa surveys firms and workers separately but there is no way of linking them. In linking worker movement to firms using the SARS IRP5 data we are able to conduct policy relevant research we could not do using any other source of data.

Data Challenges

There has been no academic analysis of the IRP5 data thus far. Using a new source of data therefore brings both benefits and challenges. The data that is a produced is the by-product of revenue collection by SARS rather than being produced for analysis by researchers, although SARS have undertaken in-house research on the data. There are thus several data issues that we have encountered during our analysis that should be discussed.

The first is that in tax years 2010, 2011 and 2012 SARS amalgamated income source codes on the IRP5 forms for retirement income and income from employment into the 3601 code that in other years is only for employment income. This means that we cannot immediately distinguish between employees earning income and pensioners receiving a pension from a pension fund in these years. In internal work SARS has estimated ratios of pensioners to working people in other years where it is possible to distinguish between them and then applied these ratios to the total number of pensioners plus employed to obtain an estimate of total employment.

It is not possible to undertake this correction when using firm level data. Thus we have tried to identify pension funds by identifying firms that issued very large numbers of tax certificates that could be either for employment income or pension income in 2012 but who issue very few or no employment income tax certificates in 2013 when pension and employment income source codes were separated. 23 "firms" issuing around 1.3 million certificates in 2012 and 1.1 million certificates in 2011 (individuals can have multiple certificates or sources of pension income) fit

this description and we have excluded these records as probable pension income and the firms as probable pension funds. This fix needs some refinement as we have not yet been able to check whether these "firms" also have very large increases between 2009 and 2010 which is the first year in which source codes for pension and income were merged.

Using the South African identity number as the individual identifier is problematic because foreigners without South African identity numbers are excluded. We are hopeful that in future data updates an anonymised version of the passport number will be provided for those without South African ID numbers that will allow the incorporation of foreigners into the analysis below. Table 1 shows that there are about 300 000 - 350 000 certificates with no identity numbers in each year- this is less than 1% of the total certificates we analyse in each wave. Most of these are likely to be foreigners but they are not yet incorporated into our analysis. If firms hiring and separation patterns or workers' movements between firms vary by foreigners and locals our results may change slightly when foreign workers are included- but any changes are likely to be small given the tiny of fraction foreign workers make up in the total.

Data Preparation

Our data that we use to estimate worker and job flows is constructed using the following method. We use the IRP5 data - all tax certificates (for incomes of more than R2000 per year) issued by all employing firms registered for PAYE. We keep only those certificates with 3601 source codes- this code indicates income from employment in 2013 and 2014 and either income from employment or taxable pension income in 2011 and 2012. We match company income tax numbers with PAYE reference numbers (payroll numbers) and conduct our analysis on CIT numbers not payrolls- thus aggregating payrolls where there is more than one payroll per CIT number. We thus have between roughly 210 000 and 230 000 employers in each tax year. We drop 1.3 and 1.1 million certificates in 2012 and 2011 respectively that look like they are actually incomes from pensions not employment. These certificates were issued by 23 "firms" that we have identified as probable pension funds. We keep only individuals reporting working in the first week of the tax year and compare their employers between year t-1 and year t. We drop those with no identity numbers, who are likely to be foreigners. We also drop those identity numbers with more than 5 different certificates in the first week of the tax year- this is about one twentieth of one percent of total records in the four years of data we use. We also drop duplicates (but keep one instance) for individuals reporting multiple certificates for the same firm in the same year.

We construct measures of employment, hires and separations in a worker-level data set. We then collapse the data by firm identifier to calculate the number of hires and separations by firm. We thus conduct our *worker flow* analysis on a firm-level data set created from a worker-level dataset. We can calculate *job flows* using changes in the total employment variable created when we collapse the data by firm identifier. Our data preparation leaves us with a dataset of around 36 million certificates and then 290 000 employers over the four tax years 2011-2014. Table 2 shows the panel of firms that we use to describe job and worker flows. Of the 290 000 firms observed at least once in the four years around 54% are observed in every year. Around 22000 are only observed in 2014- these are assumed to be firm births. Just under 20 000 were born in 2012 and continued employing through 2014.

Analysis

Worker Flows

The last column of table 3 shows that we estimate that worker flows, the sum of hires and separations, constituted around 52-54% of average employment in period t and t-1 over the 3 years we could estimate worker flows. This suggests that, as in many other countries, worker flows are very large and are a pervasive part of the South African labour market. This result is consistent with the conclusions of Banerjee et al (2008, pg 730) using the LFS panel for South Africa, who note that "Most importantly, we are struck by just how much churning there appears to be in the labour market".

As in much previous research, we find that flows rates are declining in average firm size in each of the 3 years of data that we can measure these flow rates. Worker flows are just below 70 percent in firms of less than twenty employees but only around 35 percent for firms with more than 5000 workers.

The extent of worker and job flows can be an indirect test of labour market rigidity. If turnover costs are high due to strong labour regulation then this will influence how firms choose to hire and fire. High turnover costs, for example a requirement for expensive legal processes before a worker can be fired, are likely to lower worker flows. Thus our initial estimates of worker flows shed light on the extent to which South African firms are constrained by labour legislation, which is a common concern raised by the business community. They suggest that rigidities may not be as much of a concern as some researchers have previously though.

Job Flows

The SARS data can also be used to measure job flows, or job reallocation, as defined above. Table 4 shows job creation and destruction rates by firm size and for all firms in the final column. Job creation is between 11 and 14% in the 3 years of data whilst job destruction is slightly lower- around 10%. This means that job reallocation is between 21-24% over the 3 year period between 2012 and 2014 that we can estimate it for. Kerr et al (2014) found that yearly job reallocation rates were on average 20% using Stats SA data from 2005-2011. They also noted that there were some features of the data they used that suggested their estimates of job reallocation rates were an underestimate by a few percentage points. Our analysis of the SARS data confirms this - although the analysis is over a different time period.

Like worker flows, job reallocation- the sum of the absolute value of job creation and destruction is higher in smaller firms. Table 4 shows that job reallocation rates are about 45% in firms with fewer than 20 employees but only 10.5% in the largest firms.

Births were not well captured in the Stats SA QES firm data as a result of a panel being sampled but new firms not entering the panel by design until several years later when the next panel was sampled. Table 5 shows the contribution of births and deaths to job creation and destruction. Deaths contribute about 34% of total job destruction whilst births contribute about 21% of total job creation. Kerr et al (2014) found that deaths contributed 26% of total job destruction on average between 2005 and 2011. Births were not well captured in the QES and, despite some imputation of birth-related job creation, Kerr et al (2014) found births accounted for only an average of 10 percent of total job creation, which they acknowledged was an underestimate.

The contributions to job reallocation are much higher for smaller firms. Table 5 shows that deaths contribute around 48% of job destruction in firms of size less than 20 whilst births contribute around 39% of job creation in these firms. There is not a monotonic decline in the contribution of births and deaths to job creation and destruction however. If there is any pattern it appears medium sized firms have the lowest percentage contribution to job flows by births and deaths.

Conclusions

Regulations that stifle the reallocation of labour and capital can have important effects on productivity (Hsieh and Klenow, 2009). Labour market regulation can lower job and worker reallocation and thus contribute to decreasing productivity in the economy. Our work thus far on the SARS IRP5 data has yielded estimates of worker flows from 2012 to 2014 that are fairly high.

Our work has also led to improved estimates of job reallocation. We have updated the work of Kerr et al (2014) on the Stats SA Quarterly Employment Statistics firm survey data. We find job reallocation rates of between 22-25%, with smaller firms have much higher rates of job reallocation. We also showed that the conclusions of Kerr et al (2014) were correct- actual job reallocation rates from the SARS IRP5 were a few percentage points higher than those estimated with the QES due to new firms not being sampled and changeovers between new samples leading to missing reallocation in small firms.

Further work needs to be done to expand this initial picture of job and worker flows using SARS IRP5 tax data. This work is ongoing and thus this paper is still a work in progress.

References

Banerjee, A., Galiani, S., Levinsohn, J., McLaren, Z., & Woolard, I. (2008). Why has unemployment risen in the new South Africa? *Economics of Transition*, *16*(4), 715-740.

Bassanini, Andrea, Andrea Garnero, Pascal Marianna and Sebastien Martin, 2010. Institutional Determinants of Worker Flows A Cross-Country/Cross industry Approach. *OECD Social, Employment And Migration Papers*, No. 107.

Burgess, Simon & Lane, Julia & Stevens, David, 2000. Job Flows, Worker Flows, and Churning, *Journal of Labor Economics*, University of Chicago Press, vol. 18(3), 473-502.

Criscuolo, C., P. N. Gal and C. Menon (2014), "The Dynamics of Employment Growth: New Evidence from 18 Countries", *OECD Science, Technology and Industry Policy Papers*, No. 14, OECD Publishing.

Davis, Stephen, John Haltiwanger, and Scott Schuh, *Job Creation and Destruction*, Cambridge, MA: The MIT Press, 1996.

Haltiwanger, John & Stefano Scarpetta & Helena Schweiger, 2008. Assessing Job Flows Across Countries: The Role of Industry, Firm Size and Regulations, *NBER Working Paper 13920*.

Hsieh, CT and Klenow, P, 2009. "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*. 124(4), pages 1403-1448, November.

Kaplan, David & Gabriel Martínez González and Raymond Robertson, 2007. Mexican Employment Dynamics: Evidence from Matched Firm-Worker Data. *World Bank Policy Research Working Paper*.

Kerr, Andrew & Wittenberg, Martin & Arrow, Jairo, 2014, Job Creation and Destruction in South Africa, *South African Journal of Economics*. 82(1), pages 1–18.

Klapper, Leora and Christine Richmond, Patterns of business creation, survival and growth: Evidence from Africa," *Labour Economics*, 2011, 18 (S1), S32-S44.

Magruder, Jeremy, 2012. High Unemployment Yet Few Small Firms: The Role of Centralized Bargaining in South Africa, *American Economic Journal: Applied Economics*, 4(3), 138-66, July.

OECD (2010)

Sandefur, Justin, On the Evolution of the Firm Size Distribution in an African Economy," *CSAE Working Paper Series 2010-05*, Centre for the Study of African Economies, University of Oxford 2010.

Shiferaw, Admasu and Arjun S. Bedi, The Dynamics of Job Creation and Job Destruction: Is Sub-Saharan Africa Different*?," IZA Discussion Papers 4623*, Institute for the Study of Labor (IZA) December 2009.

Huber, Peter & Kristin Smeral (2006) Measuring worker flows, *Applied Economics*, 38:14, 1689-1695.

Geeta Kingdon, Justin Sandefur, Francis Teal, 2006. Labour Market Flexibility, Wages and Incomes in Sub-Saharan Africa in the 1990s. *African Development Review* 18:3 392-427.

Tables

Table 1: Number of certificates with missing identity numbers per year

```
    year |      Freq.     Percent       Cum.
---------+---------------------------------
    2011 |    319,274       24.23      24.23
    2012 |    311,897       23.67      47.91
    2013 |    335,554       25.47      73.38
    2014 |    350,705       26.62     100.00
---------+---------------------------------
   Total |  1,317,430      100.00
```

Table 2: Firm yearly panel description 2011-2014

```
  Freq.    Percent    Cum. |  Pattern
--------------------------+---------
 158354     53.89   53.89 |  1111
  22179      7.55   61.44 |  ...1
  19886      6.77   68.20 |  .111
  18864      6.42   74.62 |  1...
  17499      5.96   80.58 |  ..11
  16776      5.71   86.29 |  111.
  15978      5.44   91.73 |  11..
   4436      1.51   93.24 |  .1..
   4388      1.49   94.73 |  1.11
  15490      5.27  100.00 |  (other patterns)
--------------------------+---------
 293850    100.00         |  XXXX
```

Notes: In Pattern a 1 means the firm is present and an employer and a . indicates the firm was not an employer in a particular year. The 4 years covered are 2011-2014.

Table 3: Worker Flows by average firm size

	0-19	20-49	50- 99	100-249	250-499	500-999	1000-4999	5000+	All firms
2012	71.7%	65.7%	66.5%	65.8%	63.6%	62.3%	52.6%	36.9%	54.1%
2013	68.2%	64.1%	65.7%	67.3%	62.9%	62.8%	55.2%	33.4%	52.7%
2014	67.7%	62.6%	64.4%	66.2%	61.7%	63.1%	52.8%	35.6%	52.9%

Table 4: Job reallocation by firm size

	Firm size								
Job destruction	0-19	20-49	50- 99	100-249	250-499	500-999	1000-4999	5000+	All firms
2012	-23.4%	-16.4%	-15.4%	-13.1%	-12.1%	-11.4%	-7.3%	-4.4%	-10.7%
2013	-21.3%	-14.3%	-12.5%	-12.0%	-10.8%	-8.5%	-10.4%	-3.0%	-9.6%
2014	-20.7%	-14.5%	-12.8%	-13.7%	-11.6%	-11.5%	-8.9%	-5.7%	-10.8%
Job Creation									
2012	23.9%	18.9%	18.0%	16.5%	15.8%	12.7%	13.2%	8.5%	14.1%
2013	22.6%	18.9%	18.6%	17.8%	14.6%	14.6%	10.3%	6.1%	12.8%

	0-19	20-49	50-99	100-249	250-499	500-999	1000-4999	5000+	All firms
2014	22.9%	17.2%	17.2%	15.8%	13.5%	12.8%	9.3%	3.8%	11.4%
Job reallocation									
2012	47.3%	35.3%	33.4%	29.7%	27.9%	24.1%	20.5%	12.9%	24.8%
2013	43.9%	33.2%	31.1%	29.8%	25.5%	23.1%	20.7%	9.1%	22.4%
2014	43.6%	31.6%	30.0%	29.5%	25.1%	24.2%	18.1%	9.5%	22.1%

Table 5: Job Destruction and Creation from firm deaths and births

				Firm size					
% Job destruction from deaths	0-19	20-49	50- 99	100- 249	250- 499	500- 999	1000- 4999	5000+	All firms
2012	42.1%	29.4%	28.8%	25.5%	34.7%	38.6%	33.7%	0.0%	29.5%
2013	44.2%	34.0%	32.5%	32.4%	31.9%	27.1%	27.7%	31.3%	34.4%
2014	56.2%	43.5%	39.4%	35.8%	36.1%	39.8%	30.9%	13.3%	37.2%
% Job Creation from births									
2012	48.2%	34.3%	28.5%	23.3%	19.6%	17.0%	11.0%	27.6%	28.5%
2013	35.6%	21.9%	16.7%	16.0%	8.4%	6.8%	10.8%	3.1%	16.7%
2014	33.4%	23.8%	19.7%	19.0%	14.9%	7.7%	10.2%	0.0%	18.6%