DataFirst

DataFirst Technical Papers

Changes in self-employment in the agricultural sector, South Africa: 1994-2012

by
Liz Neyens and Martin Wittenberg

Technical Paper Series
Number 36

About the Author(s)
Liz Neyens, Analyst, Analysis Group Inc., Boston
Martin Wittenberg, Director, DataFirst, University of Cape Town

Acknowledgements
This work was begun while Liz Neyens was an intern at DataFirst. We are grateful for the financial support provided by the Jackson Institute for Global Affairs at Yale University which made that visit possible. We would like to thank Andrew Kerr for reading this document and making helpful comments.

Disclaimer
The views presented are those of the authors and do not reflect those of Analysis Group or its clients. Research for this paper was partially undertaken while the first author was a graduate student at Yale University.

This is a joint DataFirst Technical Paper and SALDRU Working Paper.

Recommended citation

Neyens, L., Wittenberg, M. (2016). Changes in self-employment in the agricultural sector, South Africa: 1994-2012. A DataFirst Technical Paper 36. Cape Town: DataFirst, University of Cape Town

---

DataFirst, University of Cape Town, Private Bag, Rondebosch, 7701, Tel: (021) 650 5708,
Email: info@data1st.org/support@data1st.org

Changes in self-employment in the agricultural sector, South Africa: 1994-2012

Liz Neyens and Martin Wittenberg

DataFirst Technical Paper 36
University of Cape Town
August 2016

1. Introduction

While South Africa enjoys a wealth of household and firm data that speaks to the evolution of the labour market since the end of apartheid in 1994, the interpretation of these data is complicated by a variety of measurement and fieldwork changes that have occurred over this time period. These changes have been well documented by Wittenberg (2004, 2014), Casale, Muller, and Posel (2004), and Yu (2007). One of the most dramatic changes that must be considered when examining employment trends over this period is the apparent increase in self-employment that took place with the switch from the October Household Surveys (OHS) to the Labour Force Surveys (LFS). With this change in survey instrument, there was a seeming increase in the number of self-employed agricultural workers from roughly 150 000 in the last wave of the OHS (October 1999) to more than 1.4 million in the first wave of the LFS (February 2000). The number of self-employed agricultural workers (SEAWs) drops somewhat after September 2000 but remains elevated throughout all waves of the LFS compared to previous OHS waves and later Quarterly Labour Force Survey (QLFS) waves. This series, calculated using the Post-Apartheid Labour Market Series (PALMS) which combines all three survey instruments—OHS, LFS, and QLFS—is shown in Figure 1.

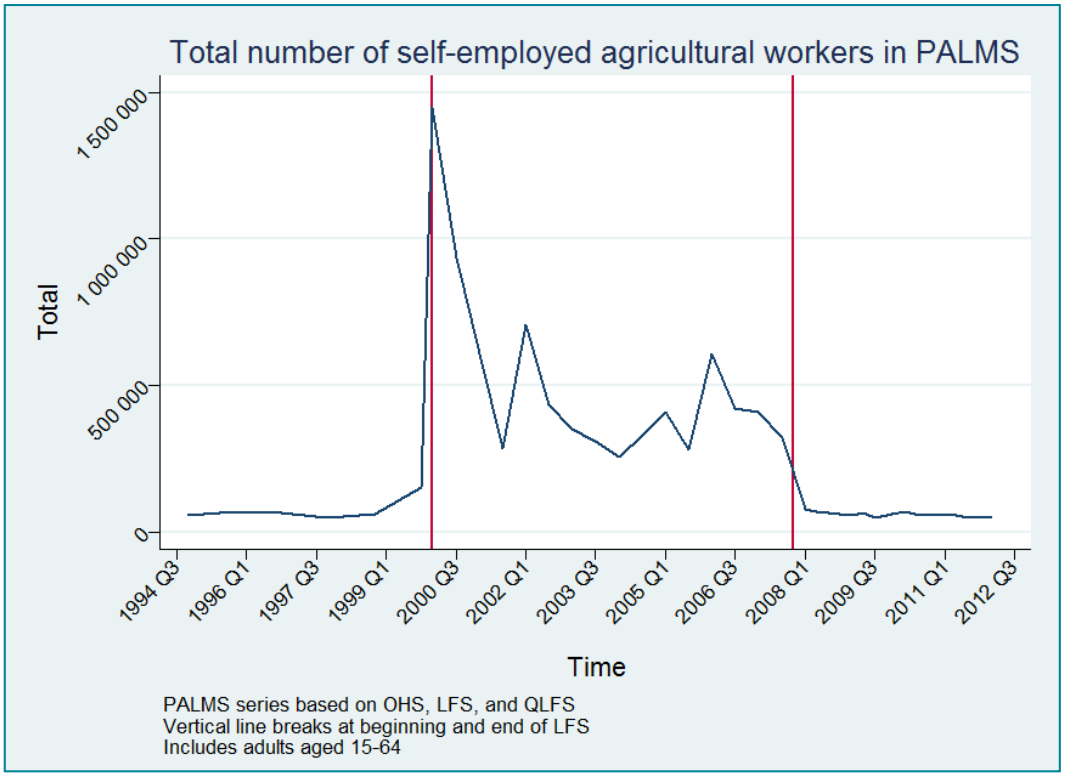

Figure 1: Total number of self-employed agricultural workers in PALMS

Casale, Muller, and Posel (2004) and Wittenberg (2014) attribute this increase in self-employment to a change in the definition of "work" that occurred with the switch from the OHS to the LFS, as well as to an increase in the amount of probing that fieldworkers did regarding informal sector employment. Beginning with the February 2000 LFS, fieldworkers were instructed to classify as employed anyone who was engaged in any informal or small-scale agricultural work, even if for only one hour in the previous week.[1] It is unsurprising, then, that switching to such a broad definition of work would create the appearance of an upward trend in employment between the OHS and the LFS. In order to draw any conclusions about labour market trends in South Africa, it is necessary to account for the fact that some number of individuals enumerated as employed during the LFS waves would likely have been enumerated as unemployed or not economically active during the OHS or QLFS waves.

Adding support to the theory that the jump in the number of SEAWs is largely the result of a changing survey instrument, Wittenberg (2014) demonstrates that the average number of hours worked in the past seven days by SEAWs drops dramatically during the LFS waves compared to the OHS and QLFS; other sectors experience no such drop.[2] The average number of hours worked in the past seven days by self-employed agricultural workers is shown in Figure 2.

[1] See Casale, Muller, and Posel 2004, pp. 980-81, and Wittenberg 2014, p. 26.
[2] See Wittenberg 2014, pp. 32-33.

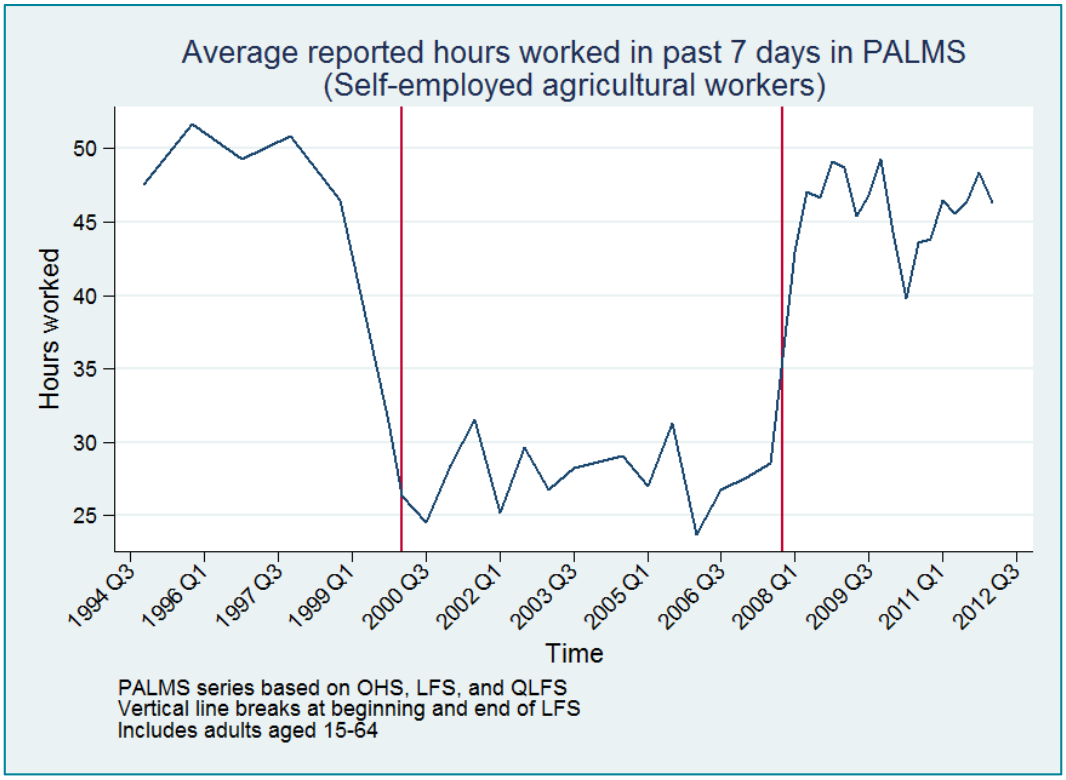

Figure 2: Average hours worked in the past 7 days in PALMS among self-employed agricultural workers

To create a smoothed series of self-employment in the agricultural sector using PALMS, and to estimate what the series might have looked like had the same definition of work been used across all waves of the OHS, LFS, and QLFS surveys, this paper employs a probit model to estimate the probability of an individual being classified as a self-employed agricultural worker during the first two—and most dramatically different—waves of the LFS. This model is then used to estimate the probability of being a self-employed agricultural worker for all out-of-sample individuals—that is, individuals in all waves of the OHS, LFS, and QLFS. These estimated probabilities are then used to simulate one hundred smoothed series of what self-employment totals might have been by taking one hundred draws from the uniform distribution.

2. Descriptive statistics: Who are the new self-employed agricultural workers and where did they come from?

As the magnitude of the jump depicted in Figure 1 suggests, the increase in the number of SEAWs that occurred with the switch from the OHS to the LFS cannot be entirely the result of real change in the size and structure of the labour force. It is therefore important to determine where these new SEAWs came from. That is, how would the SEAWs in the LFS waves been classified had they appeared in the OHS or QLFS waves? It is also important to get a sense of which waves did a better job of estimating the number of SEAWs in South Africa—did the LFS waves overestimate the size of this subpopulation, or did the OHS and QLFS waves underestimate its size?

As can be seen in Figure 3, the switch from the OHS to the LFS not only marks a sharp increase in the number of employed (most of which was self-employed agricultural workers), but also an increase in the number of unemployed (on the narrow definition) and a sharp decrease in the number of individuals who were not economically active. These trends suggest that the types of individuals who were categorized as not economically active in the OHS waves may have been categorized as economically active—either employed or unemployed—beginning with the LFS.

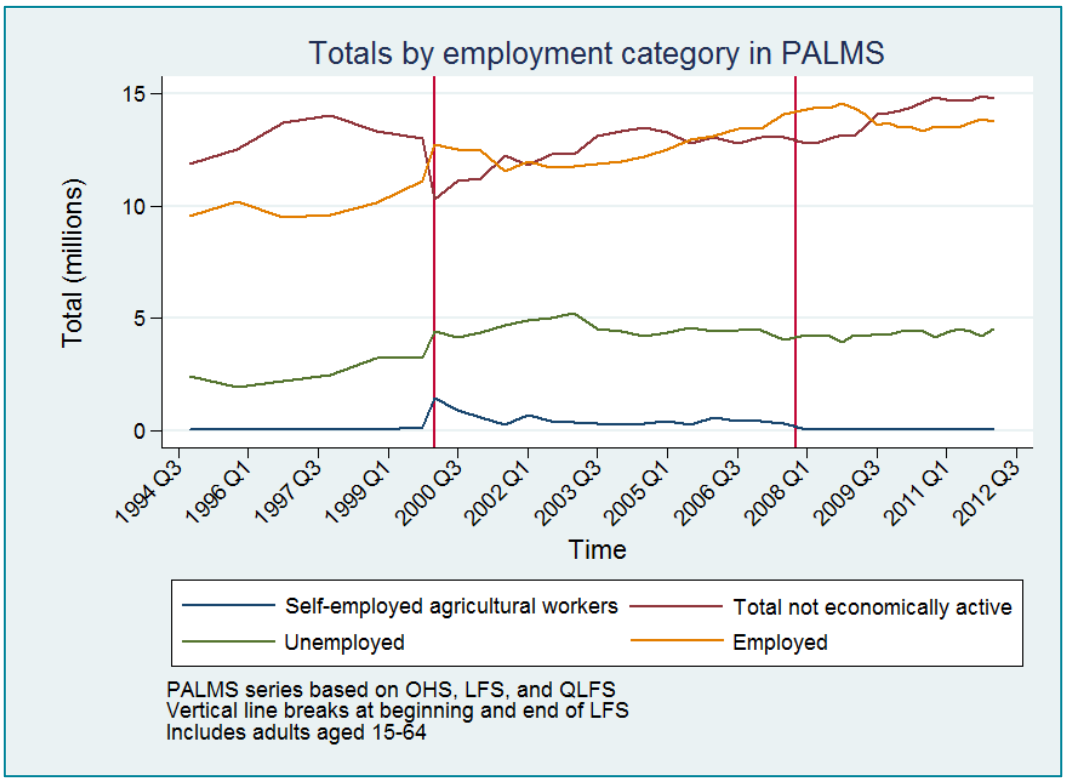

Figure 3: Totals by employment category in PALMS

It is worth noting that the sharp rise in participation and employment in this period has been attributed in part to changing education policies (Burger, van der Berg and von Fintel 2015) but as we will show the shifts are as likely to be due to measurement changes. Indeed, the demographic characteristics of SEAWs in the LFS waves are quite different to those in the OHS and QLFS waves. As can be seen in Table 1, only 42 percent of SEAWs in the OHS and QLFS waves are African, compared to a general population that is 78 percent African during the same waves. In the LFS waves, 87 percent of SEAWs are African. SEAWs in the OHS and QLFS waves are also disproportionately white and male—54 percent of SEAWs are white and 75 percent are male in the OHS and QLFS waves, compared to a general population that is 10 percent white and 48 percent male during the same time period. These characteristics suggest that the OHS and QLFS waves may have done an inadequate job of capturing women and Africans who work as self-employed agricultural workers.

SEAWs in the OHS and QLFS waves are also older and more highly educated than their counterparts in the LFS waves. SEAWs in the OHS and QLFS waves have completed more years of

education on average than the general population, while SEAWs in the LFS waves have completed fewer. SEAWs are older on average than the general population during all surveys, but SEAWs in the OHS and QLFS waves are older on average than SEAWS in the LFS waves.

Table 1: Population characteristics by survey instrument and self-employed agricultural worker status

	Weighted means and proportions (Standard errors)			
	OHS & QLFS Population (1)	LFS Population (2)	OHS & QLFS SEAWs (3)	LFS SEAWs (4)
Proportion African/Black	0.78	0.77	0.42	0.87
	(0.002)	(0.002)	(0.014)	(0.005)
Proportion Coloured	0.09	0.09	0.03	0.01
	(0.001)	(0.002)	(0.004)	(0.001)
Proportion Indian/Asian	0.03	0.03	0.01	0.01
	(0.001)	(0.001)	(0.003)	(0.001)
Proportion White	0.10	0.11	0.54	0.11
	(0.001)	(0.002)	(0.014)	(0.004)
Proportion male	0.48	0.48	0.75	0.48
	(0.001)	(0.001)	(0.010)	(0.005)
Proportion female	0.52	0.52	0.25	0.52
	(0.001)	(0.001)	(0.010)	(0.005)
Mean years of education	9.2	8.8	9.8	6.7
	(0.010)	(0.012)	(0.111)	(0.054)
Mean age	33.6	33.2	44.7	38.6
	(0.018)	(0.022)	(0.259)	(0.184)

The demographic characteristics described in Table 1 are depicted visually in Figures 4, 5, and 6. Figure 4 shows that Africans account for the vast majority of the increase in the number of SEAWs. The increase was more evenly divided between men and women, as Figure 5 shows, but women still account for more of the increase than men during the first two waves of the LFS. As shown in Figure 6, average age and years of education are both lower throughout the LFS.

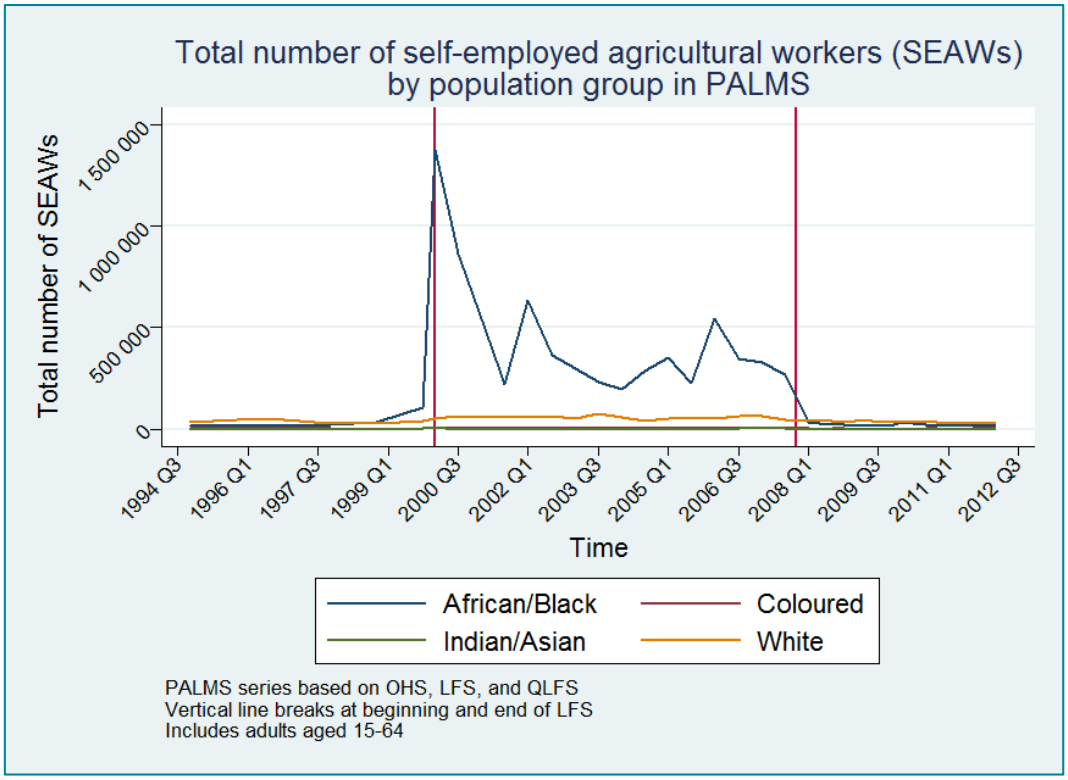

Figure 4: Total number of self-employed agricultural workers (SEAWs) by population group in PALMS

Figure 5: Total number of self-employed agricultural workers (SEAWs) by gender in PALMS

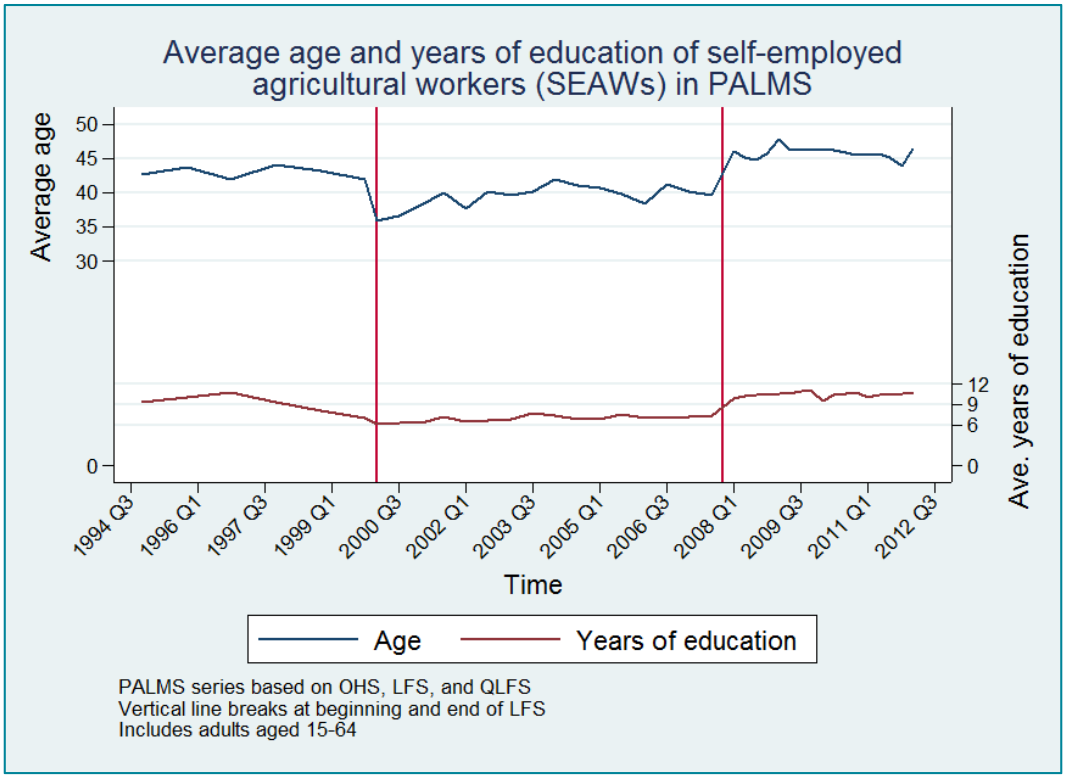

Figure 6: Average age and average years of education of self-employed agricultural workers (SEAWs) in PALMS

3. Estimating the probability of being a self-employed agricultural worker in 2000

As the descriptive statistics described above suggest, Africans and women appear to be more likely to be SEAWs in the first two waves of the LFS, and SEAWs appear to be younger and less educated on average than during the OHS and QLFS. To formally estimate which types of individuals are more likely to be enumerated as self-employed agricultural workers, we use a probit model to estimate the probability of being a self-employed agricultural worker in LFS 2000:1 and LFS 2000:2:

$$Pr(SEAW = 1|X) = \Phi(X \cdot \beta),$$

where SEAW is a binary dummy variable indicating that an individual is a self-employed agricultural worker; Φ is the standard normal cumulative distribution function; X is a vector of covariates that includes gender, population group (race), an interaction term between gender and population group, province, a quartic in age, and a quadratic in years of completed education; and β represents the estimated coefficients.

The dummy variable for self-employed agricultural worker was created from those who reported being both self-employed (*employerAll*==1) and working in the field of agriculture (*jobindcode*==1). These individuals were given a value of 1; individuals whose employment status was missing and individuals who reported being employed in other industries were given a value

of missing (“.”). All other individuals were given a value of 0. We are therefore assuming that the Self-employed Agricultural Workers in other waves would have been classified as either unemployed or not economically active. Only respondents aged 15 to 64 inclusive in LFS 2000:1 and 2000:2 were used in estimating this probit model.

The results of this estimation are shown in Table 2. We then estimate the marginal effects for all covariates, which are shown in Table 3. The marginal effects for continuous variables are calculated holding the variable at the mean, and the marginal effects for factor variables measures the discrete change in probability moving from the reference group.

Table 2: Probit model: Determinants of probability of being a self-employed agricultural worker in LFS 2000:1 or LFS 2000:2

Variable	Coefficient, β (SEs)	Variable	Coefficient, β (SEs)
Female	0.083**_	Years of education	-0.019
	(0.027)		(0.011)
Age	0.072	Years of education^2	-0.001
	(0.088)		(0.001)
Age^2	-0.002	Province	
	(0.004)	Eastern Cape	1.041***
Age^3	0.000		(0.0812)
	(0.000)	Northern Cape	0.447***
Age^4	-0.000		(0.115)
	(0.000)	Free State	0.798***
Population Group			(0.085)
Coloured	-0.942***	KwaZulu-Natal	0.712***
	(0.117)		(0.082)
Indian/Asian	-0.480	North West	0.204*__
	(0.276)		(0.099)
White	0.573***	Gauteng	-0.441***
	(0.070)		(0.121)
Population Group & Gender		Mpumalanga	0.343***
Coloured & Female	-0.342		(0.088)
	(0.204)	Limpopo (Northern Province)	0.786***
Indian/Asian & Female	-0.597		(0.086)
	(0.394)		
White & Female	-1.119***		
	(0.126)		
Observations	56,919		
Pseudo R-squared	0.144		
Wald Chi-squared	1234.91		
Prob > Chi2	0.000		

Notes: Omitted groups: Male, African, African#Female, and Western Cape. Includes individuals of working age (15-64) in waves LFS 2000:1 and LFS 2000:2 of PALMS.
* p<0.05; ** p<0.01; *** p<0.001

As can be seen in Table 3, a one-year increase in age from the mean is associated with an increase of 0.002 in the probability of being a self-employed agricultural worker; a one-year increase in years of education from the mean is associated with a decrease of 0.004. A Coloured individual and an Indian individual of average age and education are 6.8 percentage points and 5.8 percentage points less likely than an African individual to be a SEAW, respectively. There is not a statistically significant difference in the probability of being a SEAW for African and white individuals. Residents of all provinces except Gauteng are more likely than residents of the Western Cape to be SEAWs. Residents of the Eastern Cape have the highest probability of being a SEAW—they are nearly 12 percentage points more likely than residents of the Western Cape to be a self-employed agricultural worker.

Table 3: Marginal effects: Determinants of probability of being a self-employed agricultural worker in LFS 2000:1 or LFS 2000:2

Variable	dy/dx (SEs)	Variable	dy/dx (SEs)
		Province	
Female	0.002	Eastern Cape	0.119***
	(0.003)		(0.006)
		Northern Cape	0.031**_
			(0.009)
Population Group		Free State	0.075***
			(0.007)
Coloured	-0.068***	KwaZulu-Natal	0.062***
	(0.002)		(0.005)
Indian/Asian	-0.058***	North West	0.011*__
	(0.008)		(0.005)
White	0.011	Gauteng	-0.013**_
	(0.007)		(0.004)
		Mpumalanga	0.021***
			(0.005)
		Limpopo (Northern Province)	0.073***
			(0.006)
Observations	56,919		

Notes: Omitted groups: Male, African, and Western Cape. Includes individuals of working age (15-64) in waves LFS 2000:1 and LFS 2000:2 of PALMS. Marginal effects of continuous variables calculated at means. Marginal effects of factor variables calculated as the discrete change in probability moving from the reference group.
* p<0.05; ** p<0.01; *** p<0.001

This model was then used to estimate the probability of being a self-employed agricultural worker for respondents in all waves of PALMS. Individuals' probabilities of being enumerated as a self-employed agricultural worker ranged from 0.00001 to 0.52, and the mean probability was 0.069. The raw weighted probability of being a self-employed agricultural worker in these waves was also around 0.069. This value was used as the cut-off point to divide individuals into those who had a

"low probability" of being a self-employed agricultural worker and those who had a "high probability".

As can be seen in Figure 7, most of the increase in the number of self-employed agricultural workers during the LFS comes from those individuals who have a "high" probability of being a self-employed agricultural worker (i.e., greater than 0.069). And as Figures 8 and 9 show, the changes in the number of people in other employment categories are for the most part much larger among high probability individuals than low probability individuals. The exception, as Figure 9 shows, is the number of unemployed: there is only a moderate increase in the number of unemployed among high probability individuals. Rather, among high probability individuals, there is a large drop in the number of people who are not economically active, accompanied by a large increase in the number of people who are employed. Most of this increase in employment among high probability individuals is accounted for by the increase in self-employed agricultural workers. These trends suggest that "high probability" individuals—those most likely to be enumerated as self-employed agricultural workers—might have been enumerated as not economically active if they had appeared in the OHSs.

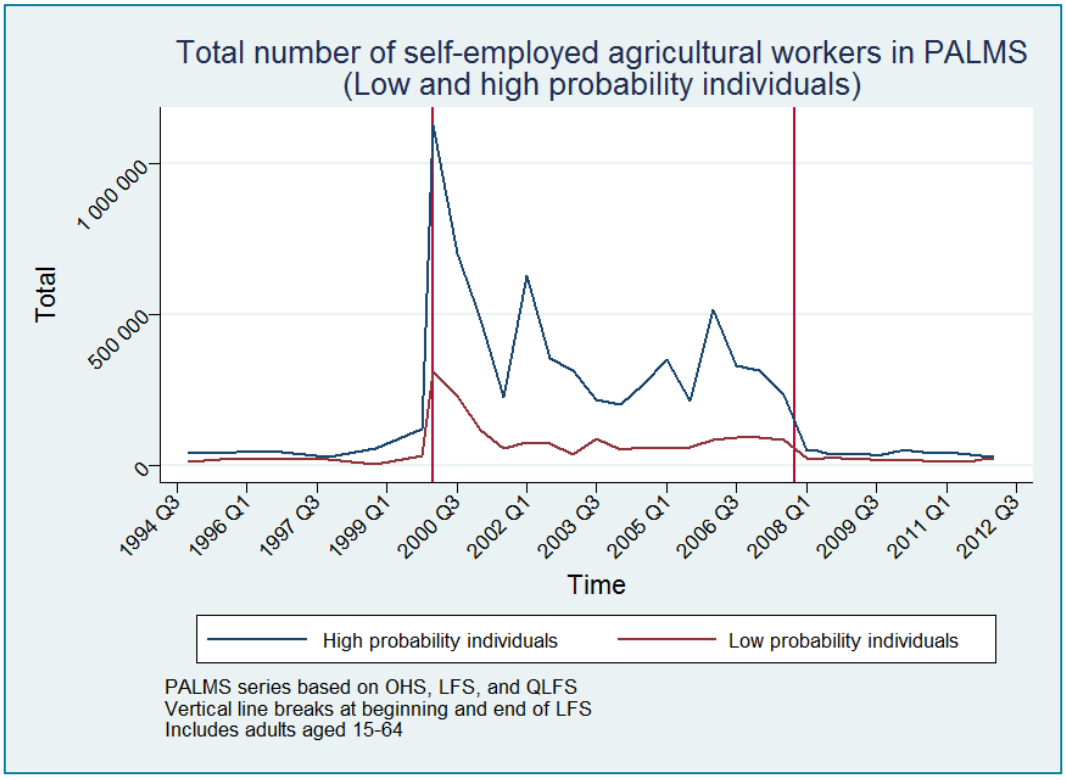

Figure 7: Total number of self-employed agricultural workers in PALMS, by "low probability" and "high probability" individuals

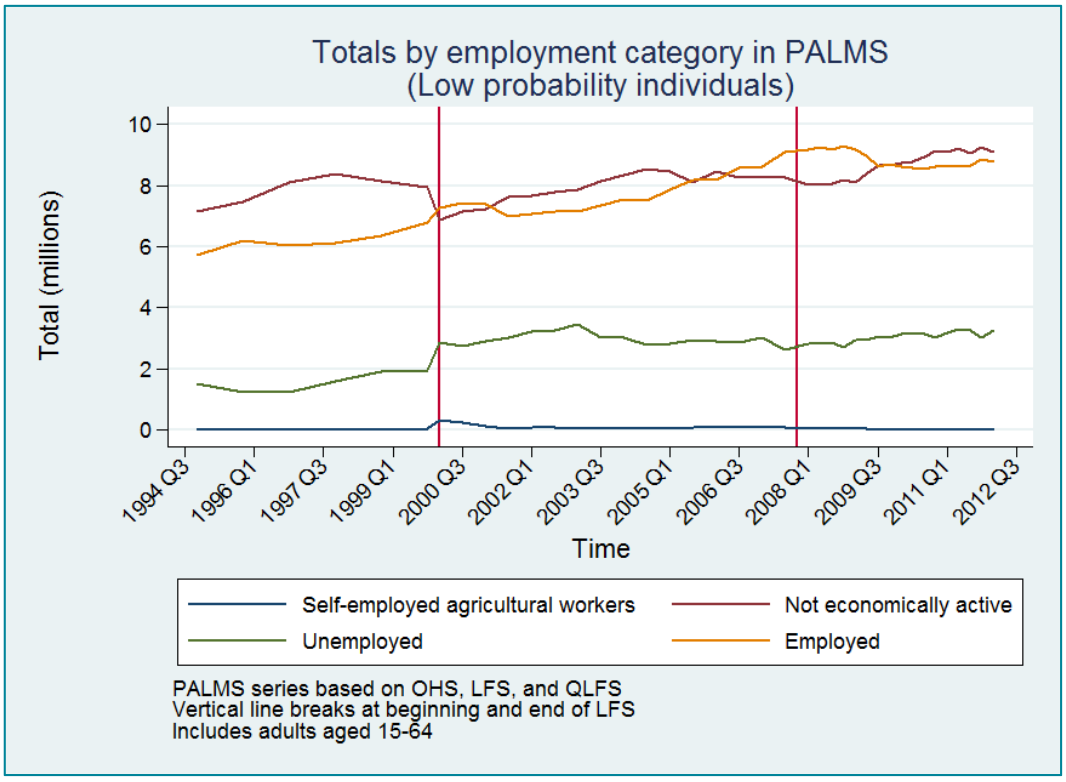

Figure 8: Totals by employment category in PALMS (Individuals who had a "low probability" of being a self-employed agricultural worker in the first two waves of the LFS)

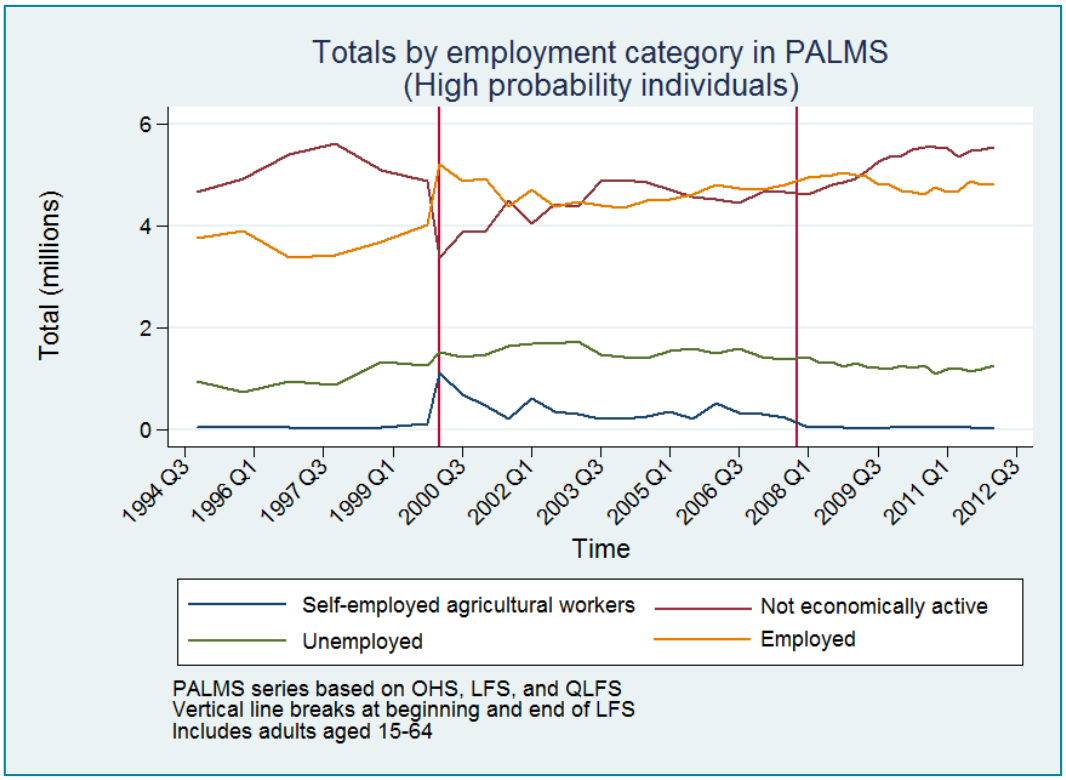

Figure 9: Totals by employment category in PALMS (Individuals who had a "high probability" of being a self-employed agricultural worker in the first two waves of the LFS)

4. Estimating a smoothed series of self-employment

To estimate how many individuals might have been enumerated as self-employed agricultural workers over the entire period (OHS, LFS, and QLFS) had the same definition of work and the same fieldwork standards been used across all survey instruments, we use the probabilities estimated by the probit model to impute who might be a self-employed agricultural worker in each of the surveys. We repeat this procedure one hundred times. The average of the predicted counts is a reasonable estimate of what the true count would have been, if the measurement process had been analogous to that in the (pooled) LFS 2000 surveys. It also assumes that the underlying propensity to become a self-employed agricultural worker, conditional on the measured characteristics, doesn't change markedly.

Our imputation procedure is to take a random draw from the uniform distribution U(0,1) for each individual who is in our target sample (i.e. not economically active, unemployed or working as a self-employed agricultural worker). If the random number generated by a given draw is less than that individual's predicted probability of being a self-employed agricultural worker, that individual is counted as a self-employed agricultural worker for that simulation. The total number of self-employed agricultural workers is calculated for each of the one hundred simulations, and these one hundred simulations are graphed in Figure 10.

In Figure 11 we show the average of the imputed counts, together with a 95% confidence interval for that figure. The confidence interval was constructed using "Rubin's rules" for multiple imputation. The estimate of the variance of $\hat{T}_{MI}$ will be given by

$$\hat{V}_{MI} = \bar{U} + \left(1 + \frac{1}{M}\right) B$$

where $\bar{U} = \sum_{j=1}^{M} \hat{U}_j$ is the average of the "within" dataset estimates of the variance of the estimated total, i.e. $\hat{T}_j$ and

$$B = \sum_{j=1}^{M} \frac{1}{M-1} \left(\hat{T}_j - \hat{T}_{MI}\right)^2$$

is an estimate of the "between" dataset variance (StataCorp 2013, pp.64-65) and M is the number of imputations (in this case 100). The variance of the estimated total within each imputation is estimated correcting for the complex survey design of each survey. The right hand panel of Figure 11 displays the actual counts once more, but this time with a 95% confidence interval again calculated correcting for survey design.

The result of the imputations suggests that there is considerable underenumeration of self-employed agricultural workers in the October Household Surveys and Quarterly Labour Force Surveys. Interestingly enough the estimate from the first wave of the LFS (i.e. LFS 2000:1) is actually higher than the estimate obtained from the average of the imputations, even taking into consideration the confidence bands around that estimate. That suggests that even with the

corrections for multiple imputations and survey design, there is more noise in the estimates than the confidence bands in either the left or right panel suggest.

Nevertheless the left-hand panel of Figure 11 suggests that even if self-employed agricultural workers were under-enumerated in the October Household Surveys and Quarterly Labour Force Surveys, there is little evidence that this would systematically skew the trends in employment. The most parsimonious reading of that graph is that the level of self-employed agricultural workers fluctuated around 1.2 million over the entire period.

It should be noted that this is not a trivial result deriving from the fact that we are using fixed coefficients. Firstly the explanatory variables (in particular education) change over the period. Secondly we are estimating totals, so changes in the overall size of the population of interest would have an impact on the estimates also.

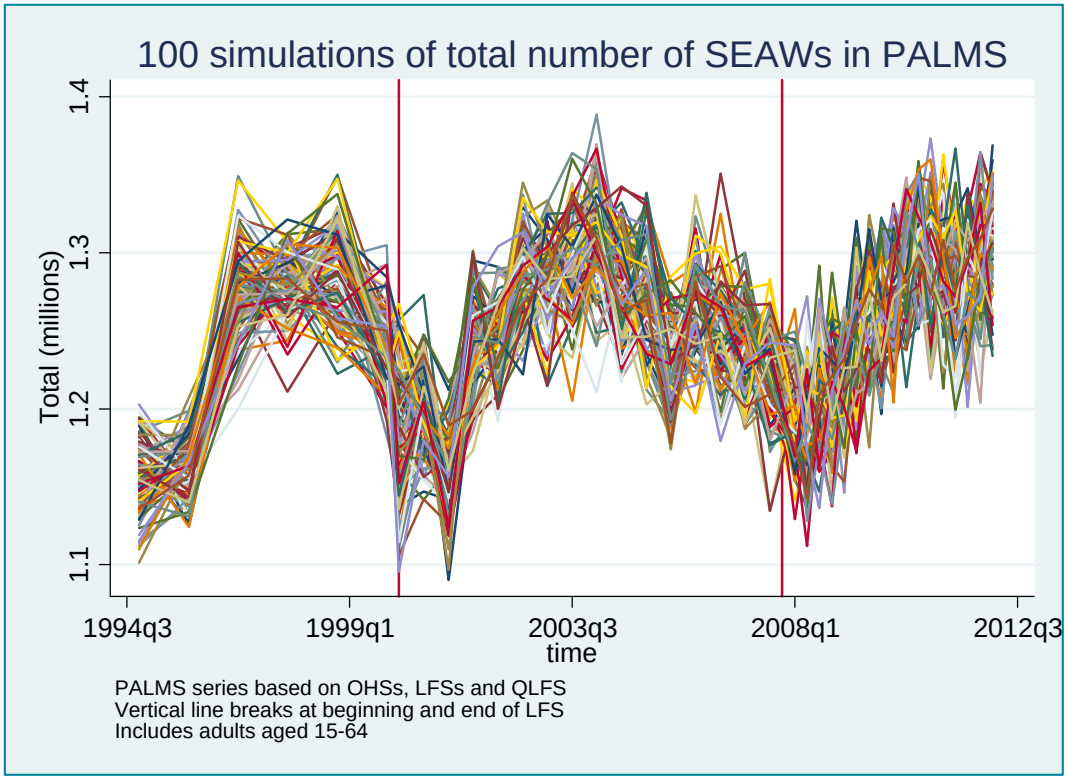

Figure 10: One hundred simulations of the total number of self-employed agricultural workers in PALMS, according to the probit model estimated on LFS 2000:1 and LFS 2000:2

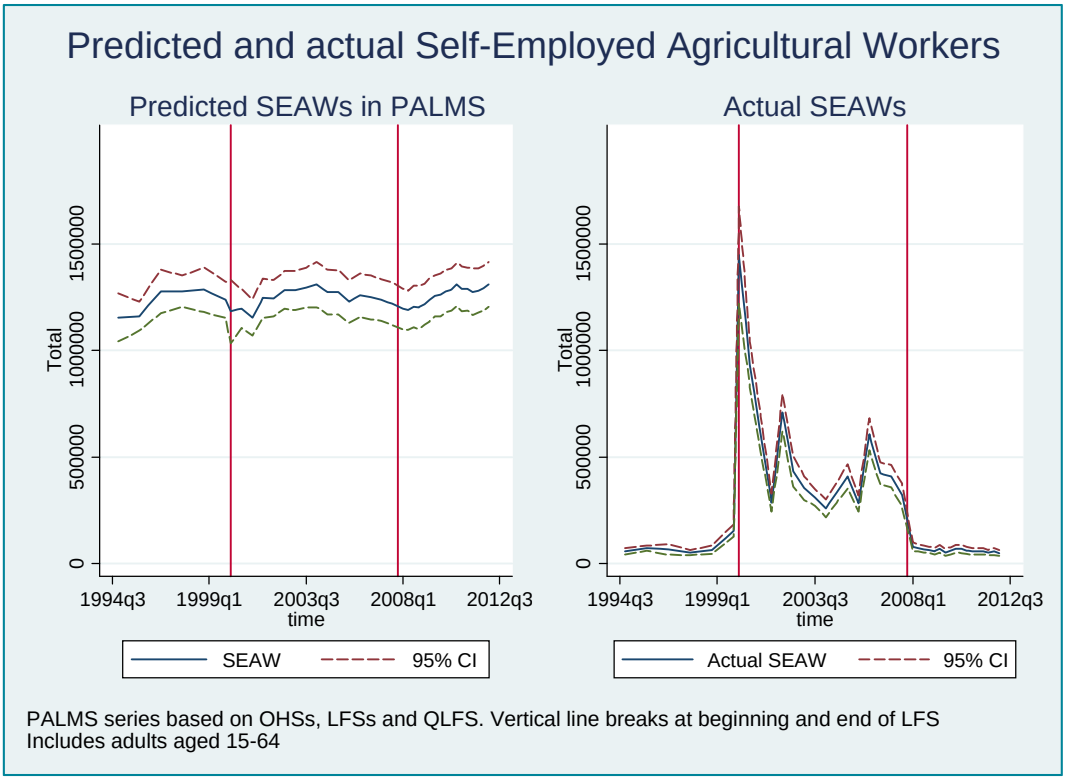

Figure 11: Average of one hundred imputations of the total number of self-employed agricultural workers in PALMS with 95% confidence bands, compared to the actual counts

5. Conclusion

The jump in self-employment in the agricultural sector during the first two waves of the LFS has been shown to be largely the result of a change in the definition of work and of further probing of the informal sector on the part of survey enumerators. The artificial nature of this change raises the question of which survey instruments offer a better reflection of informal work in South Africa. One should take caution drawing conclusions about increases in employment or changes in the number of self-employed workers in post-apartheid South Africa without taking into consideration the measurement changes that have occurred during this period. However, while the large increase in the number of individuals enumerated as SEAWs during the LFS does not necessarily represent an actual increase in employment from previous waves, the LFS may do a better job of measuring informal employment than the OHS and QLFS.

According to the data as they were collected, the number of SEAWs during the OHS and QLFS stayed relatively flat, with a large jump at the beginning of the LFS. According to our imputed series, the actual trend would have also been relatively flat, but at a much higher level, viz. around 1.2 million, with no jump upward at the OHS/LFS boundary.

The demographic characteristics of the self-employed agricultural workers who appear in the various surveys suggest that the OHS and QLFS captured a different subpopulation than did the LFS. In the OHS and QLFS, self-employed agricultural workers are more likely than the general population to be white and male, and they are older and more educated than the general

population. SEAWs in the LFS are more likely to be African than the general population; they are more female and less well educated on average than the general population, and they are younger on average than SEAWs in other surveys. SEAWs in the LFS work far fewer hours per week on average than their counterparts in the OHS and QLFS. These characteristics seem to suggest that the LFS did a better job than other survey instruments of capturing the informal economic activity in agriculture of African individuals, women, and individuals with less education.

The amount of work done by some of the self-employed agricultural workers in the LFS is minimal—one hour of work per week is not very much at all. But despite the informal nature of this work and the small number of hours worked, the LFS does a better job of capturing workers who participate in informal work, and knowing the number of individuals who participate in very small scale subsistence agriculture of this nature is important for further studying this subpopulation. While the definition of work used in the LFS is likely not the ideal definition for calculating the number of people in the labour force who are employed, this definition does allow informal work that otherwise might have been considered home production to be captured as economic activity.

Datasets

DataFirst, Post-Apartheid Labour Market Series [dataset], Version 2.2, Cape Town: DataFirst [producer and distributor], 2013.

References

Burger, Rulof, Servaas van der Berg, and Dieter von Fintel (2015), "The Unintended Consequences of Education Policies on South African Participation and Unemployment", *South African Journal of Economics*, *83*(1), 74-100.

Casale, Daniela, Collete Muller, and Dorrit Posel, (2004), "'Two million net new jobs': A reconsideration of the rise in employment in South Africa, 1995–2003," *South African Journal of Economics*, *72*(5), 978-1002. DOI: 10.1111/j.1813-6982.2004.tb00141.x

StataCorp, (2013), Stata *13 Multiple Imputation Reference Manual*, Stata Press, College Station, TX.

Wittenberg, Martin, (2004), "The mystery of South Africa's ghost workers in 1996: measurement and mismeasurement in the manufacturing census, population census and October Household Surveys." *South African Journal of Economics*,*72*(5), 1003-1022. DOI: 10.1111/j.1813-6982.2004.tb00142.x

Wittenberg, Martin, (2014), "Analysis of employment, real wage, and productivity trends in South Africa since 1994," International Labour Organization, No. 484770.

Yu, Derek, (2007), "The comparability of the Statistics South Africa October Household Surveys and Labour Force Surveys," Stellenbosch Economic Working Papers, No. 17/2007.

About DataFirst

DataFirst is a research data service dedicated to making South African and other African survey and administrative microdata available to researchers and policy analysts.

We promote high quality research by providing the essential research infrastructure for discovering and accessing data and by developing skills among prospective users, particularly in South Africa.

We undertake research on the quality and usability of national data and encourage data usage and data sharing.

DataFirst

www.datafirst.uct.ac.za
Level 3, School of Economics Building, Middle Campus, University of Cape Town
Private Bag, Rondebosch 7701, Cape Town, South Africa
Tel: +27 (0)21 650 5708
info@data1st.org / support@data1st.org