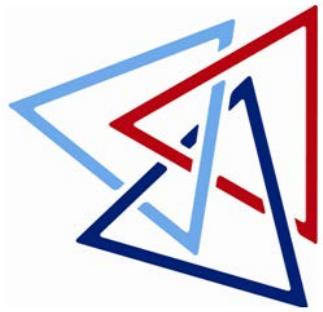


Public sector wages and employment in South Africa

Andrew Kerr and Martin Wittenberg

Abstract

This paper examines the wage structure and size of the public sector over the post-apartheid period. It does so both descriptively, examining the mean and median earnings in the public sector in the post-Apartheid period and the estimates of total public sector employment and wage bill using household survey data from the post-Apartheid Labour Market Series (PALMS). Trends in earnings and employment from the household survey data are also compared with the macro aggregates published by the South African Reserve Bank (SARB). The earnings premium for public sector workers is estimated using Ordinary Least Squares (OLS) Mincerian wage regressions and quantile regression techniques. This paper also investigates the effects of earnings imputation in the Quarterly Labour Force surveys on the estimates of the public sector earnings premium.

The **Research Project on Employment, Income Distribution and Inclusive Growth** is based at SALDRU at the University of Cape Town and supported by the National Treasury. Views expressed in REDI3x3 Working Papers are those of the authors and are not to be attributed to any of these institutions.

Public sector wages and employment in South Africa

Andrew Kerr and Martin Wittenberg

(DataFirst, School of Economics, University of Cape Town)

1. Introduction

The public sector is a significant employer in the South African economy, with an average of nearly 18% of employment in the public sector in the Labour Force Surveys (LFSs) and Quarterly Labour Force Surveys (QLFSs). This means that the wage and employment policies in the public sector have a significant direct impact on inequality in labour earnings and on the labour market more generally.

A key question is how the trends in private sector and public sector wages compare and relate to each other. A number of studies have found large public sector wage premia (Woolard 2002, Kerr and Teal 2015, van der Berg and Burger 2010). Woolard (2002) finds a premium of around 18% for public sector workers over formal private sector workers. Kerr and Teal (2015) show much higher premia than this using the KIDS panel survey. Bhorat et al (2015) is one study suggesting that there is no public sector premium on average, although the authors find positive premia in quantile regressions at higher levels of earnings and negative premia at lower earnings levels. Whether there are discernible patterns in public sector wages over time has not been investigated and we undertake this below.

Another important issue worthy of investigation is how the estimates of the size of the public sector using household survey data, both in the number of people employed and the total wage bill, compare to the macro data compiled by the South African Reserve Bank and how this has changed over time. This is important both in its own right and also as a check on the quality of the household survey data, since a relatively reliable source of external data can be used as a check on the estimates from this household survey data.

The large public sector wage premium has also been found to be robust to controlling for individual fixed effects using panel data (Kerr and Teal 2015). These results are hardly surprising since public sector wage setting is not approximated well by the standard competitive model. One of the key questions, however, is whether the size of the premia have stayed constant or whether there are shifts which seem to reflect changing labour market conditions, if with considerable lags. We use OLS regression in our analysis to explore these premia in the post-Apartheid period.

An additional question is how the wages have evolved at different points in the distribution. Van der Berg and Burger (2010) have suggested that premia for teachers are most pronounced for individuals with less than ten years' education, with hardly any premia for graduates or even negative ones (in the case of teachers). This finding is not in line with the Bhorat et al (2015) paper, since we would expect those with higher levels of education to have higher earnings (Keswell and Poswell, 2004) and thus the public sector premia patterns to be similar for earnings and education. The obvious explanation is that teachers and the rest of the public sector are different. Another possible explanation is the imputation in the QLFS earnings data, which we investigate further below. We use quantile regressions in our analysis below to explore the public sector premia at different points in the earnings distribution.

2. PALMS description

PALMS (Kerr, Wittenberg and Lam 2016) is a stacked cross sectional dataset created by DataFirst. The data in PALMS version 3.1 come from 54 Statistics South Africa surveys between 1994 and 2015 and the 1993 Project for Statistics on Living Standards and Development (PSLSD) run by SALDRU at the University of Cape Town. The Statistics South Africa surveys include the 6 October Household Surveys (1994-1999), the 16 Labour Force Surveys (2000-2007) and 32 Quarterly Labour Force Surveys (2008-2015). These household surveys contain some information on the labour market outcomes of the individuals surveyed that can be used to estimate earnings and employment, although not all contained information on whether the individuals worked in the public sector, an issue we discuss in the following section.

2.1 Measuring the public sector in PALMS

Questions about whether the worker worked in the public sector were asked in the Project for the PSLSD, Labour Force Survey (LFS) and Quarterly Labour Force Survey (QLFS). In the OHSs no question was asked directly but the industry code variable for OHS 1996-1999 contains enough information that public sector employment can be inferred, although not perfectly.

In OHS 1996-1999 we use the industry codes 910 (public administration and defence activities), 911 (central government activities), 912 (regional service council activities), 913 (Local government activities), 914 (provincial administrations), 915 (SA Defence Force), 916 (SA Police Force), 917 (Correctional service) as well as 410 (Electricity, Gas, Steam And Hot Water Supply), 411 (Production, Collection And Distribution Of Electricity) and 711 (Railway Transport). These are all unambiguously public sector employers. Employees in industries 920

(education) and 931 (human health activities) are also mostly going to be public sector employees although private medical personnel and private teachers are also included in this category. Excluding these two industries results in estimates that are too small and so we include them, knowing that we are making errors.

There are likely to be further public employees in post and telecommunications but we have chosen not to code these as public employees since a sizable fraction will be employees in the private sector. Using these codes produces estimates of the size of the public sector that look too small in 1997-1999 and about right in 1996, when compared to the 1993 PSLSD and LFSs from 2000 onwards (though as noted above we are including private medical personnel and teachers in these estimates).

2.2 Measuring earnings in PALMS

PALMS contains earnings data for the 1993 PSLSD, the OHS from 1994 -1999, the LFS from 2000-2007 and the QLFS from 2010-2014. No earnings data was collected in the 2008-2009 QLFS. In OHS 1994-1995 there is no public sector indicator so in the analysis of earnings in the public sector below we exclude OHS 1994-1995.

The main PALMS data file includes a weight variable bracketweight to allow analysis of earnings that takes account of responses in brackets but that conservatively estimates the standard errors on earnings estimates (see Wittenberg 2008 for further details as well as the PALMS guide (Kerr and Wittenberg 2016)).

3. Descriptive analysis

3.1 Employment

To start with we estimate the number of public sector employees between 1993 and 2015 using the household surveys in PALMS. Figure 1 shows that approximately 2.1 million individuals were estimated to be employed in the public sector in 1993. The 1993 figure looks perhaps too large and, bearing in mind the limitations of the public employment indicator in the OHSs, the trend in public employment was flat until about 2007, increased a little up until 2010 and then increased steadily by about 400 000 employees in the 5 years up until 2015, a similar finding to Bhorat et al (2015).

The trend in private sector employment is shown in Figure 2. We exclude self-employed agricultural workers, as Neyens and Wittenberg (2016) suggest, because this series is very

different in the OHSs, LFSs and QLFSs- with the LFS having substantially more measured agricultural self-employment. The growth in private sector employment follows almost the opposite trend to public sector employment. There is strong growth between 1997 and 2008 but the financial crisis results in a strong decrease in employment in 2009 and whilst growth returns after a year this is much slower than in the pre-financial crisis period. This low growth highlights how different the growth in public sector employment has been from private employment- with the strongest growth coming in the last few years.

Figure 3 shows the ratio of public sector employment relative to total employment excluding self-employed agricultural workers. The average over the period is 17.5%, with the proportion decreasing from around 22% in the mid-1990s to less than 16% in both 2006 and 2008, and then increasing steadily to reach just less than 20% by the end of 2015. Causal explanations of these trends are beyond the scope of the paper but a few comments are warranted. The decrease in the 1990s is perhaps a result of the emphasis on fiscal discipline by the first National Unity and ANC governments, which were unpopular with the particularly the ANC's own alliance partners COSATU and the South African Communist Party. The increasing trend since 2008 coincides both with the onset of the financial crisis and Jacob Zuma's presidency since 2009. With the private sector experiencing a large drop in employment following the financial crisis there was pressure for jobs to be created within the public sector.

3.2 Earnings

Figure 4 shows the mean and median real earnings for public sector workers, expressed in 2000 rands, as well as separate series for the mean and median earnings for formal private sector employees. We use the bracketweight variable to account for bracket responses (see Wittenberg, 2014) and use the outlier variable in PALMS to exclude earnings flagged as outliers in PALMS. The mean for public sector workers shows a steady increase until around 2010 where the trend is less clear- the last 3 or 4 quarters in 2014 are around the same values as the early period of the QLFS but 2012 and 2013 are substantially higher. The mean for private formal sector workers is very similar to that for public sector workers although the private sector mean is always less than for public sector workers, generally between 1000 and 1500 less than the public sector mean. The large increase in the means for both public and private sector employees in 2012 that continues in 2013 is suspicious and may be the result of changes in Statistics South Africa's imputation methods. We take up this issue below.

The median earnings in the public sector is substantially less than the public sector mean across all survey waves. The median initially tracks the mean in the public sector, increasing

until 2007 but the start of earnings data in the QLFS 2010 suggests that the median was relatively static for the first 11 QLFSs that measured income and then declined substantially in mid-2013 and 2014. Like the mean, the median for formal private sector workers is always below the median for the public sector. The gap between median public and private earnings is generally higher than for the means, indicating that the public sector has relatively few low earners compared to the private sector and relatively few higher earning. This gap declines due to the sharp decrease in the median in the public sector in 2013/2014 when the private sector median was more stable. The jump up in the public sector median and then the large decline between 2012 and 2014 is of concern. We investigate this further in the analysis below, suggesting that this is likely to be the result of changes in the way Statistics South Africa's earnings imputations were conducted.

3.3 Non-response

In our analysis above we have used bracketweights to correct for those who respond in categorical amounts rather than in rand amounts. Here we briefly discuss how the likelihood of non-response and bracket responses differs across public employment and non-public employment. We can only do this between OHS 97 and LFS 2007:2. For the QLFS the publicly available data does not allow us to identify the bracket responders at all because their earnings are imputed without flags, and the complete non-responses are only identified in the QLFS after 2012 Q2 in the publicly available data. In a section below we explore the effects of imputation in the QLFS more detail, using non-public earnings data from Stats SA.

Table 1 shows the proportion of bracket responses and complete non responses by whether an individual was employed in the public sector or not across the 3 surveys. The OHSs were much more likely to have bracket responses than the LFS, and there are no bracket responses in the publicly available QLFS. There is a small difference between the levels of complete non-response between the public sector and non-public sector. There are larger differences in the likelihood of bracket response- 32% of the employed in the public sector in the LFSs gave a bracket response compared to 24% of the employed outside the public sector, and there was a ten percentage point difference in the OHSs. One immediate explanation for this large difference is that public sector workers are relatively well remunerated and higher earners are more likely to give a bracket response.

4. Comparisons of earnings and employment in the public sector using macro and micro data

In this section we compare estimates of total earnings and an index of employment trends in the public sector using PALMS with the national accounts values of the same total and index. The National Accounts data for South Africa includes total compensation paid to "general government" employees, which has been released by the South African Reserve Bank for 1995-2015 as series KBP6783J. The SARB also produces an employment series for government employees (in index form), series KBP7002J, which goes back even further in time. It is not released with the total number of employees so we cannot compare levels in the two series. However we can compare evolutions over time and how these differ. For the earnings data we can compare the SARB total compensation series to the estimates of total compensation in the public sector from PALMS. Because we cannot compare employment levels in the macro and micro data but only trends, we unfortunately cannot test whether differences in employment are partly or fully responsible for any differences in total compensation that we find. This is simply a result of the limited macro data put out by the SARB, which is due to a poor quality employment level series that is generated from the firm survey data used by the SARB to create their statistics**.** The poor quality of the employment level series is due to changes in the sample frame of firms, for a discussion of this see Bhorat and Oosthuizen (2006) and Yu et al (2017).

4.1 Comparisons of employment estimates in the public sector

We can compare the employment index released by the South African Reserve Bank with an index created using the employment data in PALMS. The SARB figure is indexed so that 2010 is 100 but that is not true for the PALMS index we created. The PALMS series is indexed so that the figure from the first year of data, 1993, is set to the same value as the SARB index value for that year, which is 92.6. If more than one survey per year was conducted then the mean for the year was used. Figure 8 shows the two indices. They tell fairly similar stories about the trends in public sector employment over the last 23 years. Both series suggest that public employment declined early in the period and then rose from about 2003, although the SARB data suggests the decline was stronger and happened later than the data from the household surveys in PALMS show. At the very end of the period the indices diverge somewhat, with the SARB index decreasing and the index derived from PALMS continuing to increase. Further data is needed to understand if this divergence has continued and, if it has, why this may be.

4.2 Comparisons of earnings estimates in the public sector

Using PALMS to estimate total compensation in the public sector using the bracketweight corrects for bracket responses but does not correct for complete refusals. This matters for the estimates of total compensation in the public sector, as we will be implicitly assuming these individuals contribute zero to the total public sector wage bill and thus the estimated wage bill will be too low if we exclude the refusals. For the analysis of the total wage bill we thus use the imputed data released as a separate file with PALMS. But we do not use all 10 versions of the data with earnings imputations included in PALMS, which are included so that analysts can take account of the true uncertainty inherent in any imputation method using multiple imputation methods. Instead, we just use one version. This means that any standard errors we estimate will be too small. We thus do not report standard errors in the analysis of total compensation.

Research comparing micro and macro earnings data has been undertaken by Wittenberg (2014) for all of the employed. In his analysis Wittenberg used the publicly available Quarterly Employment Statistics (QES) reports from Statistics South Africa for the earnings data and the SARB series for employment and compared these with the Quarterly Labour Force Survey (QLFS). That option is not available to us because the QES reports do not break down earnings (or employment for that matter) by public and private sectors[1]. Thus we use the SARB total compensation series, both for general government employees and all employees, and compare this to estimates from PALMS. In PALMS we use each survey to create a monthly total wage bill and then multiply this by 12 to get a yearly figure comparable to the SARB series.

Figure 5 shows the wage bill from the National Accounts and estimated total gross earnings from PALMS for the public sector whilst Figure 6 shows the equivalent figure but for all compensation, not just public sector employees. Figure 7 shows the ratio of compensation in PALMS to the National Accounts, for all employment and then for the public sector only, in each year. If there was more than 1 survey per year in PALMS then we used the mean for each year. For the public sector the figures suggest that the OHSs generally estimated total employment income that was between 65 and 80% of the national accounts figure. This was higher in the LFSs- with these surveys capturing between 70-80% of the national accounts

[1] The QES does distinguish between private firms and public sector entities but these distinctions are not made in the publicly released reports. See the analysis of Kerr et al (2014), who use the QES firm-level data and are able to distinguish between public sector entities and private firms.

figure. It was then much lower in the QLFS- between 55 and 70% of the national accounts value.

For all employment the ratio of total compensation in PALMS to the National Accounts data is generally larger than the same ratio for public sector only. This is surprising for two reasons. Firstly one would expect that the National Accounts do not fully account for earnings in the informal sector and that would lead the ratio to be lower for all employment than for public sector. Secondly the household survey data is likely to miss high earning individuals in the private sector, both employees and the self-employed (Wittenberg, 2014), and thus should underestimate earnings from this group.

The trend over time in the ratio of earnings in PALMS to the National Accounts for all employment is less marked than in the public sector only but, if anything, suggests the opposite to the public sector trend. The OHSs generally seem to have captured more of total compensation in the national accounts, the LFSs a bit less and the 2010-2012 and 2014 QLFSs about the same as the LFS and the 2013 QLFSs a larger fraction than either the OHS or LFS.

Discussion of the causes of these trends is not helpful without further information on how the national accounts are compiled and how this has changed over time as well as further information on how incomes are imputed in the QLFS. Changes in the fractions of total compensation in the national accounts captured in the household surveys could come from changes either in the national accounts or in the surveys. Income imputation by Statistics South Africa in the QLFSs and changes in the way this has been done may explain the decrease in the fraction of public sector compensation captured in the QLFS. We discuss this further below.

Comparing the QLFS 2011 with the QES 2011 Wittenberg (2014) found that the total quarterly wage bill in the QLFS was around R40 billion below that found in the QES, or around 66% of the QES wage bill. This ratio was just less than 80% when comparing PALMS total earnings to the National Accounts total compensation data for 2011. A report on how the national accounts are compiled by the South African Reserve Bank (SARB, 2015) suggests that the QES is used to estimate total compensation but that further adjustments are also made, which may account for the difference between the ratio of PALMS total earning to the QES and the national accounts. Presumably the national accounts include estimates of total earnings in non-VAT registered firms, which are excluded from the QES sample (Kerr et al 2014) and which include informal firms and small formal firms. The document explaining the sources and methods using in compiling the National Accounts (SARB, 2015: 32) mentions the

informal sector only once, noting that "Estimates are extended annually and quarterly to incorporate the activities of the informal sector."

5. Regression analysis

5.1 OLS

We now use Ordinary Least Squares to examine the public sector premium. Gregory and Borland (1999) presented evidence on the size of the public sector premium estimated using similar methods for a number of developed countries. The premium was generally found to be positive but relatively small- with a range of 3-11%. Premia for public sector workers have generally been explained by the bargaining and wage setting arrangements found in the public sector. Fogel and Lewin (1974) argue that in the US a premium can arise because public wages are set by comparisons to private sector firms but small private firms, which pay lower wages on average, are excluded from the comparison groups. Holmlund (1993) models non-cooperative union wage setting with two unions in the public and private sectors and shows this results in a public sector premium because of the externality that increased public sector wages mean higher taxes, which is ignored by the public sector union.

To begin with we present results from an OLS regression of log earnings on a number of individual characteristics. We include as explanatory variables a public sector dummy, age, age squared, population group dummies, a male dummy, 1 digit industry code dummies, 1 digit occupation code dummies, province dummies, marital status dummies and years of education. In Figure 9 we report the coefficients on the public sector dummy from these regressions as well as the 95 percent confidence intervals. The interpretation of the coefficient on a dummy variable with a log dependent variable is that the percentage increase can be calculated as (e^(beta)-1*(100). The estimated premium for the public sector is around 30% in 1993 but then increases to around 75% in the 1997 OHS. The premia in the other two OHSs are also substantially higher than in the 1993 PSLSD. The premia in the LFS are a bit lower than those estimated in the OHSs but increase until 2006, during which they decrease a little. But in all the LFSs the premium is at least 50%. This analysis suggests that public sector workers are paid substantially more than observably similar workers in the private sector. It should be noted that our regressions include all the employed, rather than just formal employees, as in Woolard (2002), or formal non-agricultural workers, as in Bhorat et al (2015). This means that our estimated premia are not comparable to those estimated by these authors. We did this because many surveys (all the QLFSs and some OHSs) do not have direct questions on informality.

We can redo the analysis including dummies for informal self-employment, informal wage employment and formal self-employment from OHS 1997- LFS 2007:2. The public sector dummy is thus interpreted as the premium relative to formal sector wage employees. The coefficients from these dummies are shown in Figure X, along with the coefficients from the previous regressions. They suggest that the premium is reduced somewhat, by an average of .11. But it is still very high, with a mean of .43, or about a 54% premium to working in the public sector.

Returning to the original regressions without informal sector controls, the QLFS public sector premium is around 20-30 percentage points lower than in the OHS or LFS in the first few years income was collected. It then decreases in the last few QLFS for which income data was available at the time of writing. It is not clear why the premium has decreased in the QLFS although it is in line with the drops in the median and mean public sector earnings and the relatively static mean and median within the private sector shown in Figure 4. It seems likely that the public sector premium increased in the late 1990s but given that the major changes in the estimated premium occur with the change from the LFS to the LFQS it would not be correct to claim that the reduction in the public sector premium after 2007 is the real trend. We take this issue up in the following section.

5.2 Imputation in the QLFS

One line of exploration on the effects of changes in survey instrument on the public sector premium and some of the other anomalies noted in the previous sections is the effects of imputation in the QLFS. For the regression analysis above and below we are using the unimputed PALMS data with the bracketweight variable to take account of bracket responses. However in the case of the QLFS this "unimputed" data is not actually unimputed! The reason for this is that Statistics South Africa chose to release imputed data for the QLFS but did not document how this imputation was undertaken nor which individuals with income data were imputed and which were not. This means analysts using the publicly available QLFS data cannot undo the imputation or try to improve them.

There is no analysis that we are aware of that explores the imputations in the QLFS so we briefly describe the imputations and what data has been released by Statistics South Africa before estimating the effects of this imputation on the estimated public sector premium. The QLFS is released every quarter but this release does not include the income data. The QLFS income data is released once a year in the "Labour Market Dynamics" data release.

In 2010 and 2011 the public release data contained no employed individuals with earnings data flagged as being a "refusal". This is shown in Table 2 below. In 2012 Q3 and Q4 there was an increase in the fraction of refusals to around 7-9%, suggesting that complete imputation was no longer being undertaken. These individuals are missing the monthly income variable created by Statistics South Africa. In 2013 and 2014 it was between 9- 11% and again these individuals have missing monthly income.

Table 2 also shows the fraction of refusals to this initial question for unimputed labour market Dynamics data from 2011. DataFirst obtained this data from Statistics South Africa to check how and in what ways the imputations undertaken by Statistics South Africa matter. The proportions of refusals in the unimputed 2011 data is similar to the proportions in the last half of 2012, 2013 and 2014, suggesting that these releases have the correct levels of refusals. However this does not mean there are no imputations in 2013 and 2014. Table 2 shows that these waves contain no individuals who have a bracket response but no rand amount and no individuals who are recorded as having responded as "Don't know". This suggests that those who gave a bracket response or a don't know then had a monthly income amount imputed for them by Statistics South Africa.

This analysis shows that within the publicly released QLFS there are important differences in what earnings data was imputed between the Q2 2012 QLFS and earlier surveys on the one hand and the Q3 2012 QLFS and later surveys on the other. This means that each of these QLFS periods has very different income data to the OHS and LFS, where no imputation was undertaken for refusals, don't knows or bracket responses.

The imputation of earnings may affect the estimated public sector premium in the QLFS. To check this we compare the premia estimated using the QLFS 2011 data public release which, as Table 2 showed, has fully imputed data, with the unimputed QLFS 2011 data DataFirst obtained from Statistics South Africa. We then prepared this data in the same manner as we did for the LFS in the PALMS release and use the bracketweight variable to account for bracket responses but exclude the complete refusals or don't knows.

Another important difference between the LFS and QLFS that may affect the estimation of the public sector premium is the response rates. Table 3 shows the response rates in the LFS and compares to the QLFS 2011 which we calculated using the unimputed 2011 data. The table shows that the proportion of complete refusals in the QLFS was more than double the proportion in the LFS. The proportion of amount refusals - either because of a complete refusal or because of giving a bracket amount - was about 33% higher in QLFS 2011 relative

to the LFS average. The don't know proportion was more than 4 times higher in the QLFS, although off a much smaller base. Thus, although both the unimputed QLFS 2011 and LFS were prepared in similar ways a much larger fraction of employees were excluded in the QLFS analysis below than in the LFS because of the higher non-response rate. The complete non-response rate was about the same in the public and private sector (not shown in the table), although the bracket response rate was higher in the public sector, again suggesting that the bracketweight correction we implement is important.

5.3 The effects of imputation on the public sector premium

The OLS regression results from the imputed and unimputed 2011 QLFS data are shown in Table 4. They clearly show that the imputation undertaken by Statistics South Africa makes a substantial difference to the estimated size of the public sector premia. The estimated public sector dummy coefficients in the imputed data are between 0.29-0.33 whilst with the unimputed data prepared using the same method as the LFS from the earlier period are between 0.48 and 0.56. These are in the same range as those estimated in the older LFS data. Thus rather than a decrease in the public sector premia over time, the analysis of the unimputed QLFS data suggests that the imputation is responsible for the reductions in the premia in the QLFS relative to the LFS and that once calculated in a comparable manner the premia have actually been relatively stable since the early 2000s.

The likely reason for the public sector earnings premium being smaller in the imputed data than the unimputed data is that the imputation routine used by Stats SA does not include whether the individual works in the public sector. We have shown that public sector workers earn more on average. These two facts imply imputed earnings for public sector workers will be too low for public sector workers and too high for non-public sector workers on average.

6. Estimates of the public sector premium using quantile regression

In this section we estimate quantile regressions using 5 quantiles: q=.1, .25, .5, .75 and .9. Figure 10 shows the estimated public sector coefficients from the quantile regressions of these 5 quantiles for the PALMS waves in which there is both income data and a public sector dummy. The results indicate that the PSLSD is different from all the other waves, with very little difference in the estimated public sector premium across the 5 quantiles. The OHS and LFS suggest a broadly similar pattern, with the premium being highest at the lower quantiles and lowest at the higher quantiles. The average premium across the OHS and LFS for the .1 quantile is .59 whilst for .9 it is .29, so that the public sector premium at the 10^{th} percentile is more than double that at the 90^{th} percentile.

The QLFS is again very different, with the premia estimated from q=.1 and q=.25 generally lower than those estimated for the higher quantiles. This is a complete reversal of the ordering of the quantiles compared to the OHS and LFS but is similar to the findings of Bhorat et al (2015) using the 2013 QLFS. Given the effects of imputation on the public sector premium estimated using OLS that we showed above, we estimate the public sector coefficients on the unimputed 2011 QLFS data using quantile regression. Figure 11 shows the estimated public sector coefficients using the unimputed data and the bracketweight variable to account for bracket responses. Surprisingly the public premium is again lowest for q=.1 and highest for q=.5 whilst for q=.9 the premium is 2nd lowest (of 5 quantiles) in 3 of the 4 surveys. It is not clear why even with the unimputed data such a different pattern emerges in the 2011 QLFS compared to the OHS and LFS and further work is required before this can be definitively answered.

For all 5 quantiles the premia are higher in the unimputed 2011 data than the imputed 2011 data. This accords with the OLS results, which showed that the average premium was higher in the unimputed data than in the imputed data. Again, this suggests that Stats SA is excluding public sector employment from the variables used to undertake the imputation routine, leading to underestimation of the premium across the earnings distribution.

7. Conclusion

The public sector is an important and large feature of the South African labour market, with an average of 17.5% of employment occurring in the public sector. In this paper we have used the PALMS data to estimate the trends in employment and earnings in the public sector, to describe median and mean earnings, to compare the trends in employment and earnings in macro and micro data and to estimate public sector earnings premia both in OLS and using quantile regressions, which led to a discussion of the role of imputation in these estimates.

Employment in the public sector was relatively flat until around 2007 when it began to increase, with stronger increases since 2010. This was a very different trend to employment in the private sector, which grew until 2009, when the financial crisis caused a decline, and the subsequent growth has been much smaller than in the pre-2009 period. This meant that the proportion of employment in the public sector grew by 4 percentage points from 16% to 20% between 2009 and 2015, after shrinking from 22% in the early 1990s.

Earnings in the public sector is much higher than in the private sector, with median and mean earnings in the public sector growing by about 33% and 25% respectively between 1997 and

2007. The private sector mean grew by about 25% also whilst the median was virtually unchanged between the start and end of the same period.

We raised some concerns about the large changes particularly in the mean for the public sector and in the public sector premium estimated using OLS in the QLFS. We thus discussed imputation in the QLFS, showing that imputation was used in these surveys when it had not been in the OHS and LFS and that imputation had changed between Q2 and Q3 of the 2012 QLFS. This makes comparisons with the OHS and LFS and between the first and second parts of the QLFS problematic. We also showed that the earnings premium estimated in the QLFS is much more similar to the premia in the LFS when unimputed data is used.

Quantile regression was used to estimate the public sector premia at different points in the earnings distribution. In the OHS and LFS we found positive premia that were highest at the lowest quantiles and lowest at the highest quantiles. In the QLFS this was completely reversed. Whilst this may have been thought to be the result of the imputations, in fact q=.1 still had the lowest premium using the unimputed data.

We have shown in this paper that the household survey data can be used to estimate the size of the public sector, both in terms of wage bill and employees. Public sector wages are much more generous than those in the private sector, although analysis of earnings is complicated by the imputation of earnings in the QLFS. Any future work using earning data from the QLFS should pay much more attention to the impacts of this imputation.

* * *

References

Bhorat, H and Oosthuizen, M, 2006. Evolution of the labour market: 1995–2002. In Bhorat, H and Kanbur, K (Eds), *Poverty and Policy in Post-Apartheid South Africa*, HSRC Press, Pretoria.

Bhorat, H., Naidoo, K., Oosthuizen, M. and Pillay, K, 2015. Demographic, employment and wage trends in South Africa. *WIDER Working Paper* 2015/141.

Fogel, W and Lewin, D (1974). Wage determination in the public sector", *Industrial and Labor Relations Review* 27:410-431.

Gregory and Borland (1999). Recent Developments In Public Sector Labor Markets. Chapter 53 in *Handbook of Labour Economics*, 3573-3630. O Ashenfelter and D Card (eds).

Holmlund, B. (1993). Wage setting in private and public sectors in a model with endogenous government behaviour", *European Journal of Political Economy* 9: 149-162.

Kerr, A. and Teal, F. 2015. The Determinants of Earnings Inequalities: Panel Data Evidence from KwaZulu-Natal, South Africa, *Journal of African Economies*, 24(4), 530-558.

Kerr, A., Wittenberg, M. and Arrow, J., 2014. "Job Creation and Destruction in South Africa," *South African Journal of Economics*, 82(1), pages 1-18.

Kerr, A and Wittenberg, M. 2016. A Guide to version 3.1 of the Post-Apartheid Labour Market Series. Available at https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/434

Kerr, A., Lam, D. and Wittenberg, M.: 2016, Post-Apartheid Labour Market Series [dataset],DataFirst, University of Cape Town. Version 3.1.

Neyens, E. and Wittenberg, M.: 2016, Changes in self-employment in the agricultural sector, South Africa: 1994-2012, *Working Paper*, DataFirst University of Cape Town.

South African Reserve Bank. 2015. South Africa's national accounts 1946–2014 An overview of sources and methods. *Supplement to the South African Reserve Bank Quarterly Bulletin March 2015*. Available at https://www.resbank.co.za/Lists/News%20and%20Publications/Attachments/6650/March%20%E2%80%93%20Supplement%202015.pdf

Van Der Berg, S. and Burger, R. 2010. Teacher pay in South Africa. *Stellenbosch Economic Working Papers: 26/10.*

Wittenberg, M.: 2014, Analysis of employment, real wage, and productivity trends in South Africa since 1994, *Conditions of Work and Employment Series 45*, International Labour Office.

Woolard, I. 2002. "A Comparison of Wage Levels and Wage Inequality in the Public and Private Sectors, 1995 and 2000," *DPRU Working Paper 02062*, Development Policy Research Unit, University of Cape Town.

Yu D, Burger R and Rankin N. 2017. South African employment trends: How reliable are the different data sources? *REDI Working Paper 31*.

Tables

Table 1: Non-response across OHS, LFS and QLFS

	Public sector dummy			
		0	**1**	**All**
OHS	No response	4068	628	4696
		7.68	5.16	7.21
	Bracket response	15804	4860	20664
		29.85	39.92	31.73
	Actual earnings response	33080	6687	39767
		62.47	54.92	61.06
	Total	52952	12175	65127
		100	100	100
LFS	No response	17478	3848	21326
		5.43	6.1	5.54
	Bracket response	78223	19887	98110
		24.32	31.53	25.5
	Actual earnings response	225986	39348	265334
		70.25	62.37	68.96
	Total	321687	63083	384770
		100	100	100
QLFS	No response	16524	4476	21000
		9.59	10.97	9.85
	Actual earnings response	155840	36311	192151
		90.41	89.03	90.15
	Total	172364	40787	213151
		100	100	100

Note: The first figure in is the number of observations, underneath is the proportion.

Table 2: Labour Market Dynamics (QLFS) income question refusals

Income response categories	2010	2011	2012 Q1+Q2	2012 Q3	2012 Q4	2013 LMD	2014 LMD	2011 Unimputed LMD
Income amount	100	100	99.79	92.92	90.48	89	90.49	55.5
Income bracket without amount	0	0	0	0	0	0	0	24.97
Don't know	0	0	0	0	0	0	0	9.1
Refusal	0	0	0.21	7.08	9.52	11	9.51	10.43
Total	100	100	100	100	100	100	100	100

Table 3: Response rates in the LFS and QLFS 2011

		1	2	3	4	1+3+4
		Complete refusal	**Amount refusal**	**Don't know**	**Zero income**	
LFS	LFS 00:1	0.05	0.39	0.04	0.12	0.21
	LFS 00:2	0.03	0.21	0.00	0.09	0.12
	LFS 01:1	0.03	0.28	0.02	0.06	0.11
	LFS 01:2	0.03	0.29	0.03	0.03	0.09
	LFS 02:1	0.03	0.34	0.03	0.07	0.13
	LFS 02:2	0.04	0.34	0.03	0.04	0.11
	LFS 03:1	0.04	0.33	0.02	0.03	0.10
	LFS 03:2	0.05	0.36	0.03	0.03	0.11
	LFS 04:1	0.04	0.35	0.02	0.03	0.09
	LFS 04:2	0.04	0.32	0.03	0.04	0.11
	LFS 05:1	0.04	0.33	0.02	0.06	0.11
	LFS 05:2	0.04	0.31	0.02	0.04	0.10
	LFS 06:1	0.03	0.31	0.01	0.07	0.11
	LFS 06:2	0.03	0.30	0.01	0.05	0.10
	LFS 07:1	0.03	0.28	0.01	0.05	0.10
	LFS 07:2	0.03	0.28	0.01	0.04	0.09
QLFS 2011	QLFS 2011:1	0.11	0.41	0.09	0.00	0.20
	QLFS 2011:2	0.11	0.42	0.10	0.00	0.21
	QLFS 2011:3	0.12	0.44	0.10	0.01	0.22
	QLFS 2011:4	0.11	0.43	0.09	0.01	0.21
LFS average		0.04	0.31	0.02	0.05	0.11
QLFS 2011 average		0.10	0.40	0.09	0.01	0.19

Table 4: OLS earnings functions using Imputed and unimputed 2011 QLFS

	Imputed earnings				Unimputed earnings with bracketweight			
	35	36	37	38	35	36	37	38
VARIABLES								
public3	0.339***	0.297***	0.317***	0.291***	0.549***	0.559***	0.483***	0.486***
	(0.0306)	(0.0317)	(0.0330)	(0.0267)	(0.0458)	(0.0470)	(0.0480)	(0.0448)
age	0.0159***	0.0297***	0.0235***	0.0172***	0.0267***	0.0388***	0.0434***	0.0335***
	(0.00486)	(0.00452)	(0.00518)	(0.00407)	(0.00698)	(0.00607)	(0.00715)	(0.00574)
age2	-0.000101*	-0.000250***	-0.000204***	-0.000119**	-0.000159*	-0.000299***	-0.000392***	-0.000264***
	(5.91e-05)	(5.54e-05)	(6.27e-05)	(4.86e-05)	(8.62e-05)	(7.46e-05)	(8.77e-05)	(6.93e-05)
2.jobindcode	0.625***	0.547***	0.634***	0.717***	0.914***	0.873***	0.962***	1.019***
	(0.0633)	(0.0592)	(0.0527)	(0.0608)	(0.0632)	(0.0853)	(0.0625)	(0.0720)
3.jobindcode	0.215***	0.209***	0.217***	0.186***	0.341***	0.365***	0.358***	0.360***
	(0.0392)	(0.0362)	(0.0363)	(0.0342)	(0.0519)	(0.0758)	(0.0497)	(0.0523)
4.jobindcode	0.117	0.241***	0.245**	0.274***	0.550***	0.261***	0.790***	0.490***
	(0.0911)	(0.0864)	(0.0987)	(0.0819)	(0.170)	(0.0983)	(0.216)	(0.119)
5.jobindcode	0.0544	0.143***	0.125**	0.0786**	0.134**	0.142*	0.227***	0.0997
	(0.0446)	(0.0424)	(0.0487)	(0.0387)	(0.0610)	(0.0744)	(0.0624)	(0.0619)
6.jobindcode	0.0720*	0.0783**	0.115***	0.110***	0.0867*	0.0620	0.141***	0.124**
	(0.0382)	(0.0344)	(0.0350)	(0.0341)	(0.0496)	(0.0681)	(0.0448)	(0.0497)
7.jobindcode	0.165***	0.180***	0.161***	0.163***	0.328***	0.237***	0.298***	0.232***
	(0.0468)	(0.0439)	(0.0482)	(0.0413)	(0.0754)	(0.0751)	(0.0654)	(0.0574)
8.jobindcode	0.214***	0.195***	0.185***	0.171***	0.293***	0.291***	0.256***	0.267***
	(0.0401)	(0.0366)	(0.0383)	(0.0350)	(0.0519)	(0.0708)	(0.0495)	(0.0513)
9.jobindcode	0.0836*	0.116***	0.126***	0.130***	0.133**	0.0623	0.131**	0.177***
	(0.0441)	(0.0414)	(0.0429)	(0.0378)	(0.0586)	(0.0747)	(0.0561)	(0.0598)
10.jobindcode	-0.331***	-0.328***	-0.322***	-0.386***	-0.185***	-0.323***	-0.265***	-0.311***
	(0.0606)	(0.0530)	(0.0493)	(0.0507)	(0.0628)	(0.0746)	(0.0563)	(0.0777)
11.jobindcode	0.939**	0.321	0.937***	0.282***	1.134***	0.729**	0.935***	0.227
	(0.377)	(0.378)	(0.162)	(0.102)	(0.241)	(0.304)	(0.272)	(0.222)
2.marstat	-0.0950**	-0.0924**	-0.122***	-0.0551	-0.207***	-0.136**	-0.243***	-0.129**
	(0.0410)	(0.0405)	(0.0409)	(0.0372)	(0.0530)	(0.0547)	(0.0529)	(0.0500)
3.marstat	-0.0123	-0.104***	-0.0563	-0.0418	-0.00550	-0.163***	-0.0444	-0.0128
	(0.0376)	(0.0378)	(0.0438)	(0.0312)	(0.0528)	(0.0533)	(0.0568)	(0.0924)
4.marstat	0.0653***	-0.0496***	-0.0680***	-0.0662***	-0.117***	-0.102***	-0.129***	-0.0981***
	(0.0183)	(0.0178)	(0.0193)	(0.0168)	(0.0231)	(0.0256)	(0.0255)	(0.0229)
yrseduc	0.0844***	0.0875***	0.0727***	0.0805***	0.0859***	0.0879***	0.0738***	0.0795***
	(0.00325)	(0.00323)	(0.00345)	(0.00295)	(0.00431)	(0.00472)	(0.00451)	(0.00421)
Constant	6.785***	6.517***	6.814***	6.807***	6.571***	6.349***	6.387***	6.558***
	(0.118)	(0.106)	(0.122)	(0.100)	(0.165)	(0.146)	(0.168)	(0.147)
Observations	19,189	19,170	20,072	20,583	11,418	11,237	11,342	11,859
R-squared	0.502	0.505	0.482	0.505	0.538	0.522	0.517	0.518
Standard errors in parentheses								

*** p<0.01, ** p<0.05, * p<0.1
In addition to the variables reported in this table we also controlled for occupation, gender and population group

Figures

Figure 1: Public employment as estimated in PALMS

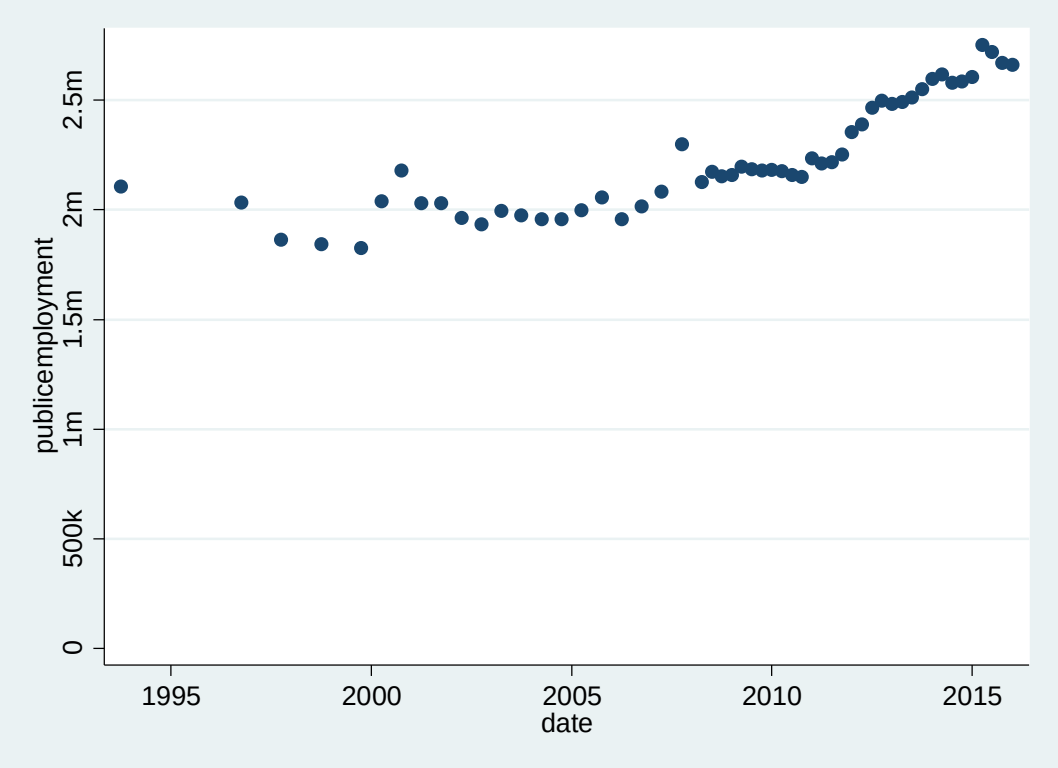


Figure 2: Private employment excluding agricultural self-employment

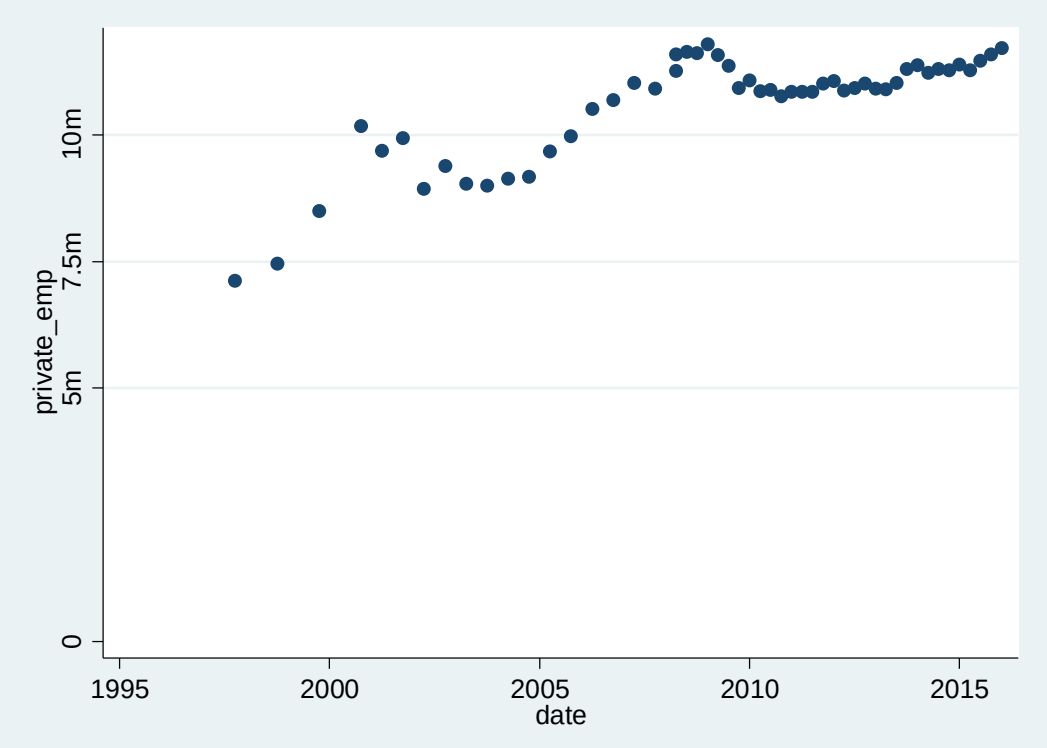


Figure 3: The proportion of public sector workers in total employment, excluding agricultural self-employment

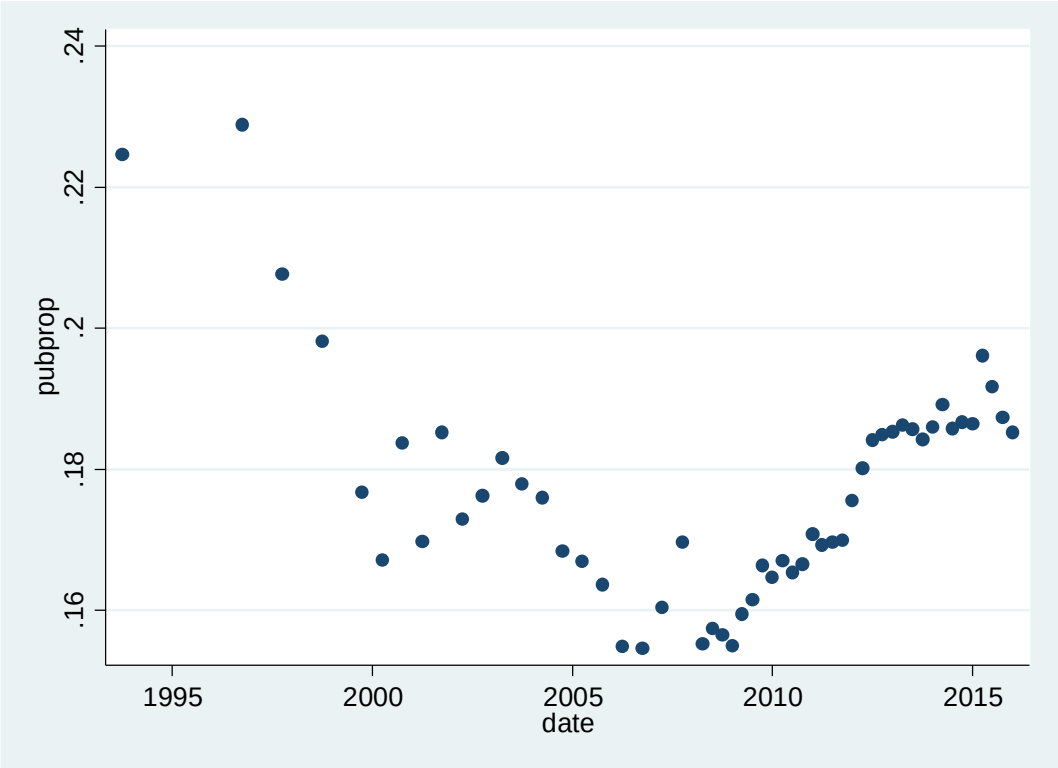


Figure 4: Median and mean earnings for public and formal private sector employees

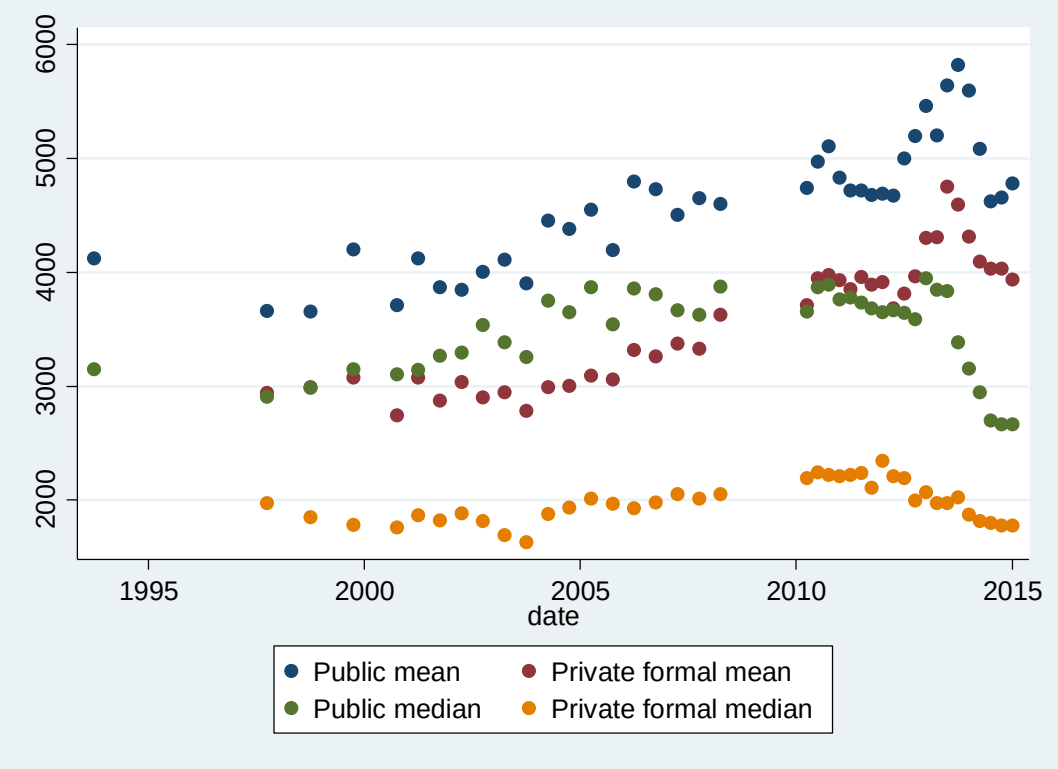


Figure 5: Total compensation in government entities and total earnings for PALMS public sector employees

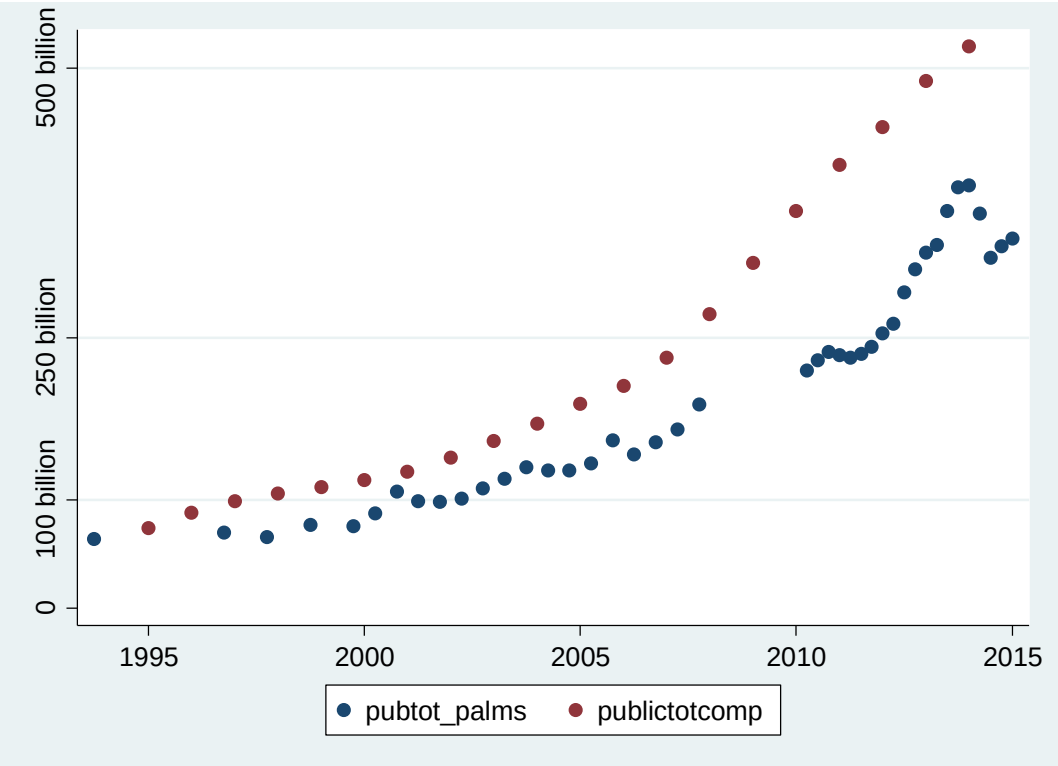


Figure 6: Total compensation in NA and total earnings for all employed in PALMS

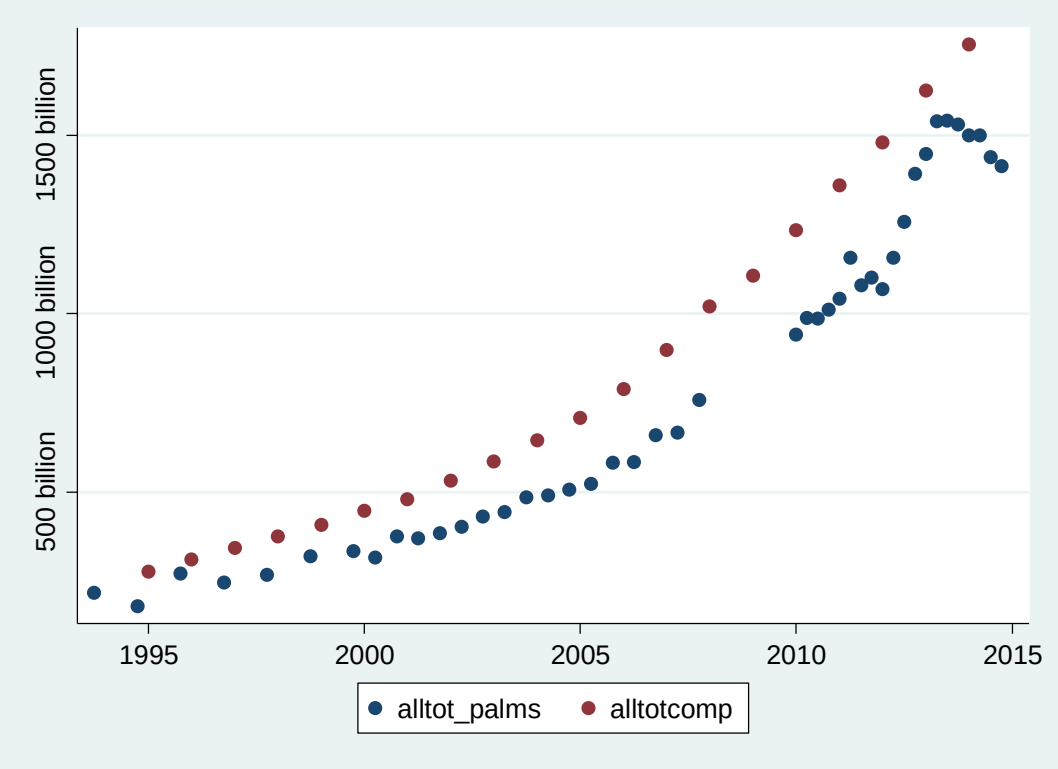


Figure 7: Ratios of total compensation in the NA to total earnings in PALMS in the public sector and in total

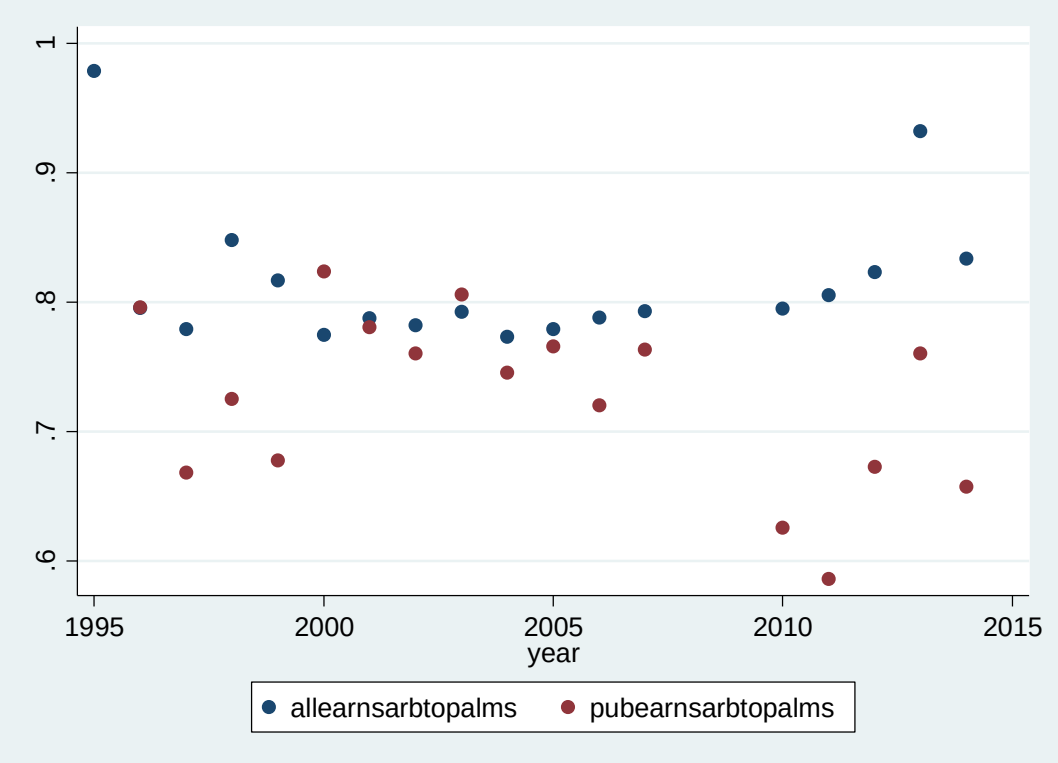


Figure 8: Public Employment in PALMS and the SARB public sector employment series

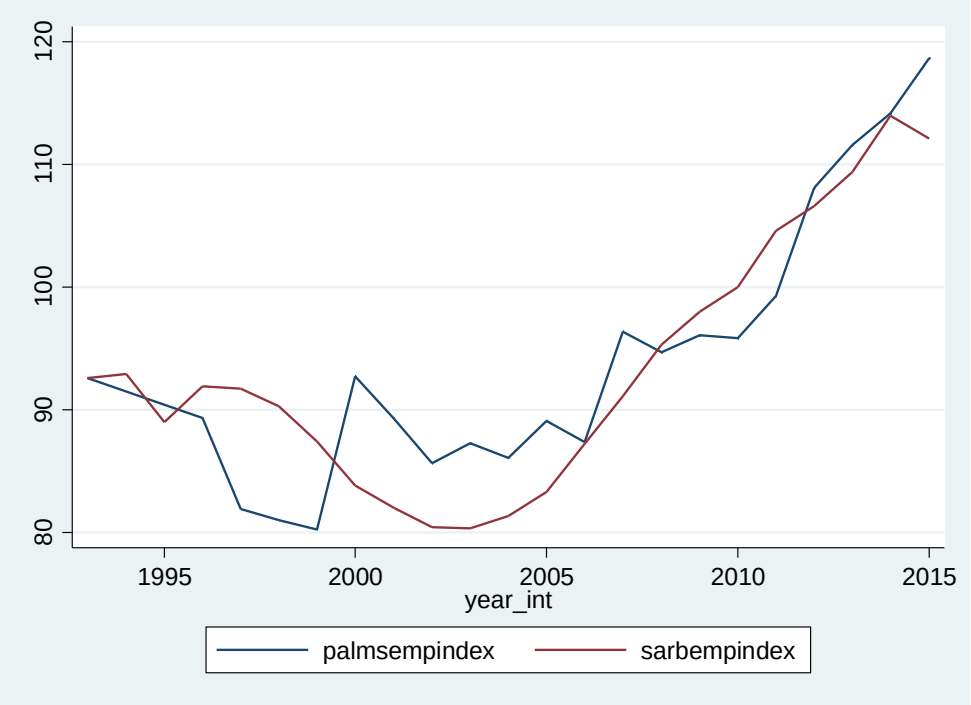
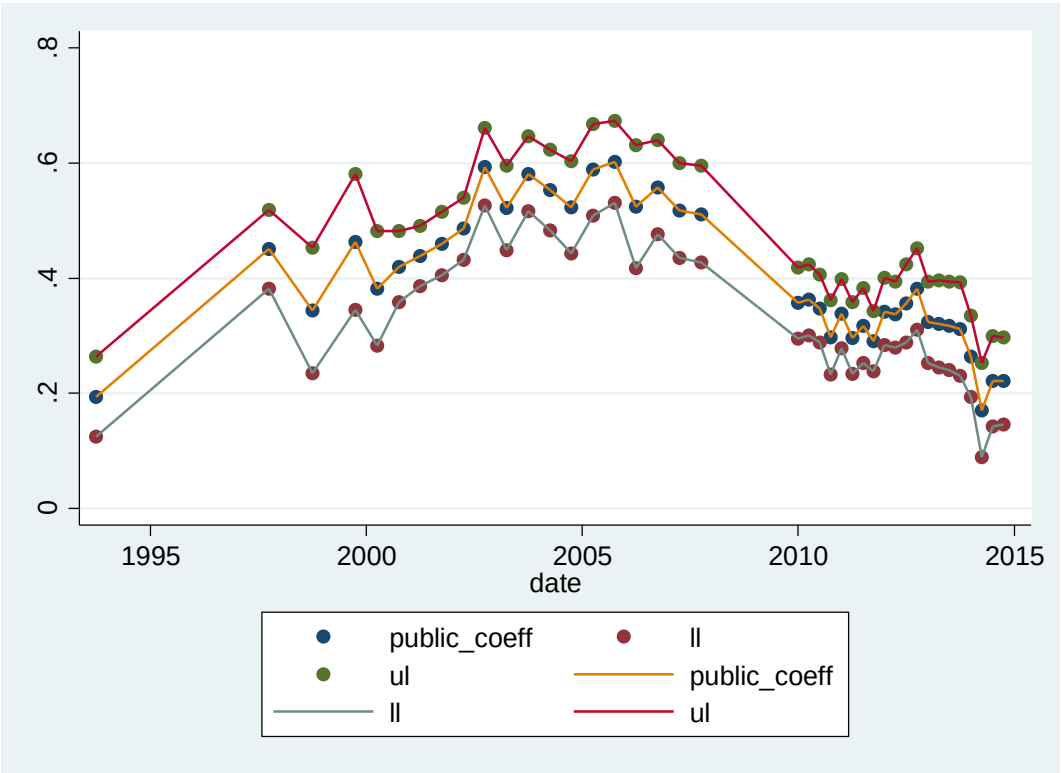


Figure 9: Public sector dummy OLS regression coefficients



Notes: Ul is upper limit and ll is lower limit- these are the 95% confidence intervals for the estimated coefficients. Explanatory variables were public sector dummy, age, age squared, population group dummies, a male dummy, 1 digit

industry code dummies, 1 digit occupation code dummies, province dummies, marital status dummies and years of education

Figure 10: Quantile regression estimates of the public sector dummy coefficient

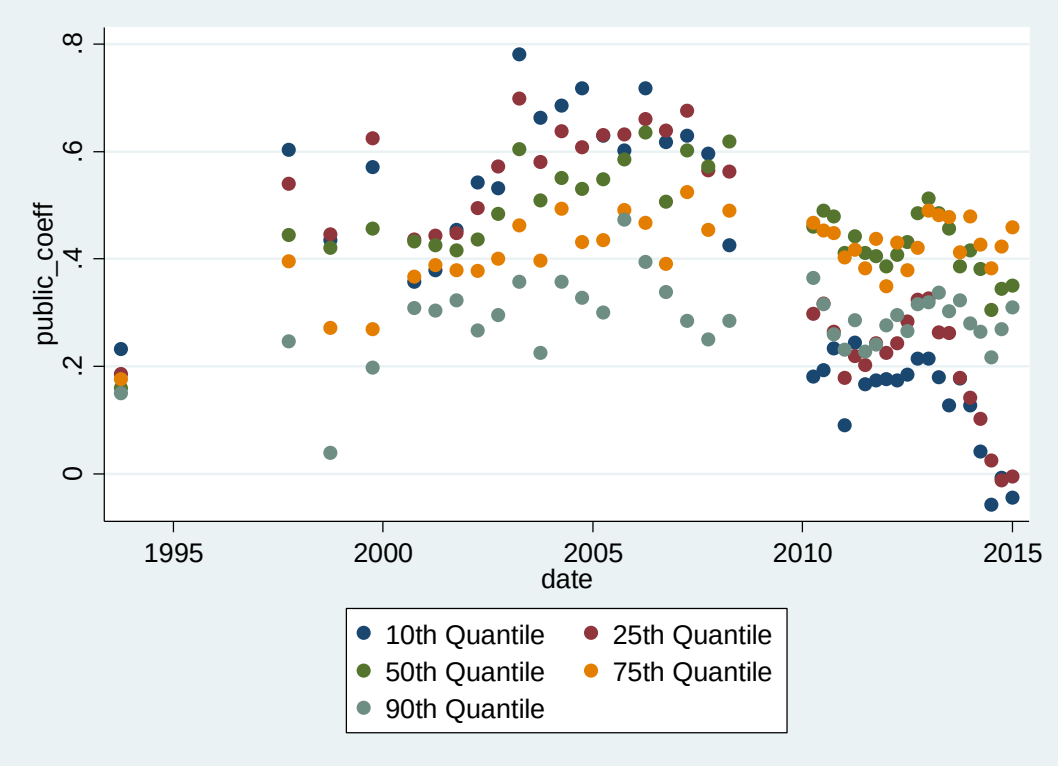
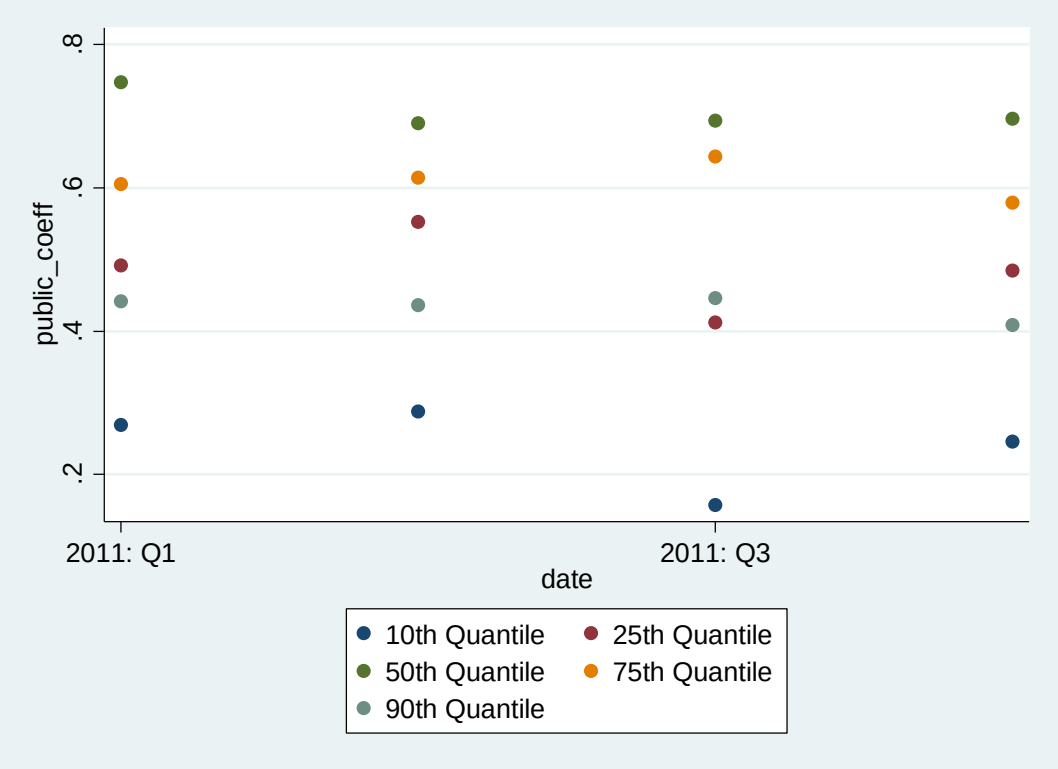


Figure 11: Public sector coefficients from quantile regression using unimputed QLFS 2011 income data


public_coeff
.2
.4
.6
.8
2011: Q1
2011: Q3
date
10th Quantile
25th Quantile
50th Quantile
75th Quantile
90th Quantile

The **Research Project on Employment, Income Distribution and Inclusive Growth (REDI3x3)** is a multi-year collaborative national research initiative. The project seeks to address South Africa's unemployment, inequality and poverty challenges.

It is aimed at deepening understanding of the dynamics of employment, incomes and economic growth trends, in particular by focusing on the interconnections between these three areas.

The project is designed to promote dialogue across disciplines and paradigms and to forge a stronger engagement between research and policy making. By generating an independent, rich and nuanced knowledge base and expert network, it intends to contribute to integrated and consistent policies and development strategies that will address these three critical problem areas effectively.

Collaboration with researchers at universities and research entities and fostering engagement between researchers and policymakers are key objectives of the initiative.

The project is based at SALDRU at the University of Cape Town and supported by the National Treasury.

Tel: (021) 650-5715

www.REDI3x3.org

REDI3X3

Research Project on Employment, Income Distribution & Inclusive Growth

Towards an integrated response to unemployment, inequality and poverty