Southern Africa Labour and Development Research Unit

DataFirst

The measurement of earnings in the post-Apartheid period: An overview

by
Martin Wittenberg and Farah Pirouz

Working Paper Series
Number 108

About the Author(s) and Acknowledgments

Martin Wittenberg - Director, DataFirst and Professor, School of Economics, University of Cape Town
Farah Pirouz - Researcher, DataFirst, University of Cape Town

We acknowledge the support of an infrastructure grant to DataFirst from the Redi3x3 project on "Employment/Unemployment, Income Distribution and Inclusive Growth". The Vice-Chancellor's Strategic Fund of the University of Cape Town paid for the initial construction of the PALMS dataset without which this research would not have been possible.

This is a joint SALDRU and DataFirst Working Paper.

Recommended citation

Wittenberg, M., Pirouz, F. (2013). The measurement of earnings in the post-Apartheid period: An overview. A Southern Africa Labour and Development Research Unit Working Paper Number 108. Cape Town: SALDRU, University of Cape Town

---

ISBN: 978-1-920517-49-6

© Southern Africa Labour and Development Research Unit, UCT, 2013

Working Papers can be downloaded in Adobe Acrobat format from www.saldru.uct.ac.za.
Printed copies of Working Papers are available for R15.00 each plus vat and postage charges.

Orders may be directed to:
The Administrative Officer, SALDRU, University of Cape Town, Private Bag, Rondebosch, 7701,
Tel: (021) 650 5696, Fax: (021) 650 5697, Email: brenda.adams@uct.ac.za

The measurement of earnings in the post-Apartheid period: An overview

Martin Wittenberg and Farah Pirouz*
DataFirst and School of Economics
University of Cape Town

September 2013

1 Introduction

Earnings questions have been asked in South Africa's national surveys annually since 1994. A key question for labour economists has been to track and explain the evolution of earnings over this post-apartheid period. Unfortunately, however, the measurement instrument has changed in ways that make it tricky to simply take the raw figures and compare them even if one restricts the attention to the October Household Surveys and the various Labour Force Surveys. In this paper we analyse some of the changes and indicate where corrections are needed. We implement many of these changes in the second release of PALMS, the Post-Apartheid Labour Market Series (Kerr, Lam and Wittenberg 2013).

The structure of the paper is as follows. In section 2 we review studies done on the earnings variables in the national surveys from Statistics South Africa, particularly those that comment on the comparability of the variables over time. In Section 3 we pay attention in more detail to the evolution of the measurement instrument. We then turn to an analysis of the actual responses in section 4 with a view to pinpointing where the underlying measurements may have changed. The following sections deal with ways of handling bracket information and missing data respectively . In section 8 we look at the impact of these data quality adjustments on the estimation of average real earnings over time.

2 Literature

Several papers have looked at the post-Apartheid wage series and commented on the comparability (or otherwise) of the earnings information. Casale (2004) remarked on a precipitous decline in real earnings over the period 1995 to 2001. She notes that better enumeration of low wage work may explain some of the decline, but argues that

> the fall in informal self-employment earnings between 1995 and 2001 is unlikely to be the result of improved data collection alone, as more and more people crowding

*We acknowledge the support of an infrastructure grant to DataFirst from the Redi3x3 project on "Employment/Unemployment, Income Distribution and Inclusive Growth". The Vice-Chancellor's Strategic Fund of the University of Cape Town paid for the initial construction of the PALMS dataset without which this research would not have been possible.

> into already low income-generating informal activities would be expected to depress average earnings even further. (Casale 2004, p.264)

The most comprehensive review to date is by Burger and Yu (2007). They argue that the bulk of the decline in earnings is driven by the discontinuity between the October Household Surveys (OHSs) and the Labour Force Surveys (LFSs). They also note that outliers contaminate the series, a problem which they argue is particularly noticeable in the September 2000 LFS. In other regards the LFS series is less prone to extreme incomes. On the other hand they suggest that people earning more than R1 million a year affect average earnings in all years prior to 2000, and argue that

> This is due mainly to changes in the earnings intervals that individuals were allowed to specify without revealing their exact incomes, which permitted all workers in 1995 and the self-employed in 1996 to 1998 to answer in higher income brackets than were available to respondents in the subsequent years. (Burger and Yu 2007, p.6)

This comment indicates that the issue of bracket responses and how they are utilised in the calculation of "average earnings" is of some importance. Neither Casale nor Burger and Yu discuss how this was achieved. One of the more popular procedures for doing so is by placing the bracket respondents at the midpoint of the interval. In the case of the open category the typical procedure is to take some multiple of the lower bracket boundary. Both von Fintel (2007, p.297) and Yu (2011, p.14) suggest that the factor that should be used is 1.1. This seems remarkably conservative. Simkins (personal communication) by contrast suggests that the value of the lower bracket boundary should be doubled. The reason is that the distribution in the upper tail is approximately Pareto with a coefficient of approximately two. Other work (Wittenberg 2013) estimates the Pareto parameter at around 1.85, suggesting that the factor should be around 2.17, i.e. Simkins's rule of thumb has much better grounding in the data. We will return to the issue of bracket responses and how to deal with them later. At present it suffices to note that how this information is incorporated into the analysis is of considerable importance and that this is not always appropriately documented in the analyses that have been conducted.

Another strand of the literature deals with changes in the instrument. Daniels (2013) investigates changes in the earnings question addressed to employees in the 1997-1999 October Household Surveys and the 2000-2003 Labour Force Surveys. He focuses particularly on how missing information is dealt with. Unfortunately some rather large changes in the instrument precede this period (as shown below) and one of the biggest breaks between the October Household Surveys (OHSs) and the Labour Force Survey (LFSs) is the collapse of two separate questions (in the OHSs) to just one question in the LFSs.

Yu (2007) provides a detailed discussion of many variables measured in both the OHSs and LFSs, including earnings. He shows that there was an increase in the reporting of zero incomes and unspecified responses with the change to the LFSs. Indeed the February 2000 survey seems particularly anomalous with regard to both (Yu 2007, Table 24, p.27). In a companion piece (Yu 2009) he discusses the relationship between the LFSs and the Quarterly Labour Force Surveys (QLFSs) introduced in 2008. He does not consider the question of earnings, because in the period he reviewed there was no earnings information available for the QLFSs.

Yet other work has considered the evolution of sampling practice and the resultant breaks in the series. Kerr and Wittenberg (2013) argue that the change in survey practice between the 1998 October Household Survey and the 1999 OHS led to better enumeration of backyard shacks, which led to a marked increase in the number of backyard shack dwellers interviewed. As a result some of the early surveys might present a too rosy picture of conditions in South Africa. Indeed 1995, which has been used as the "anchor" for many post-Apartheid analyses (Branson

Table 1: Respondents giving Gross Income or Take-home pay in OHS 1994

Gross/take-home pay? Q3.14	Freq.	Percent
1=Total salary before deductions	10517	34.87
2=Take-home pay	19646	65.13
Total	30163	100

Note: Unweighted counts and percentages of individuals giving nonmissing information

and Wittenberg 2007, Appendix 1, pp.322ff.) appears anomalous on many measures (Branson and Wittenberg 2007, Wittenberg forthcoming). While cognisant of these problems we will not be able to address all of them as fully as they deserve. Indeed the question as to how to deal with changes in the sample coverage is still subject to research.

3 Changes in the Instrument

There are several changes in the survey instrument. As noted above, perhaps the most far-reaching of these is the change from 2 questions in the OHSs to one in the LFS, since this coincides with the shift to more extensive probing of informal sector activity. As Burger and Yu (2007) suggest, this break leads to a downward shift in measured wages. However, there are also some really big changes in the measurement instrument in the first three years, which create additional issues when trying to assess conditions at the point at which Apartheid was being dismantled. Our discussion therefore breaks the series into four parts: the early OHSs 1994-1996, the later OHSs 1997-1999, the LFSs 2000-2007 and the QLFSs since 2008

3.1 The early October Household Surveys: 1994-1996

This period seems to have been a period of experimentation - where the "earnings question" eventually converged on an agreed on template. Each of these surveys has its own idiosyncracies. 1994 allowed both "gross" and "net" wages to be reported, whereas thereafter the employee earnings question was clearly about gross income. In both 1994 and 1995 respondents are allowed to choose an earnings category and are asked **subsequently** whether this income was earned daily, weekly, monthly or annually. The upshot is that there are effectively four **sets** of earnings categories – one for each reporting period. By 1996 this practice had changed and the conversion between reporting periods was done on the instrument, resulting in only one set of categories. 1996 was idiosyncratic in other ways, though, since it was run in conjunction with the Post-enumeration survey for the census. This meant that not only was the sampling done differently, but a smaller sample was extracted and a much simpler instrument used. The upshot was that only categorical income information was collected.

3.1.1 1994

The earnings question asked of employees is shown in Figure 8. The first point to note is that this earnings question does not stipulate whether gross or net income should be reported. Instead a subsequent question (3.14) asks whether the reported number is before or after deductions. Table 1 shows the "raw" responses from that survey. Interestingly around two-thirds of respondents gave a "net" figure when not prompted.

By 1995 (see Figure 10) this level of ambiguity was removed and since then the earnings question has been explicitly about earnings before deductions. What might the impact of this

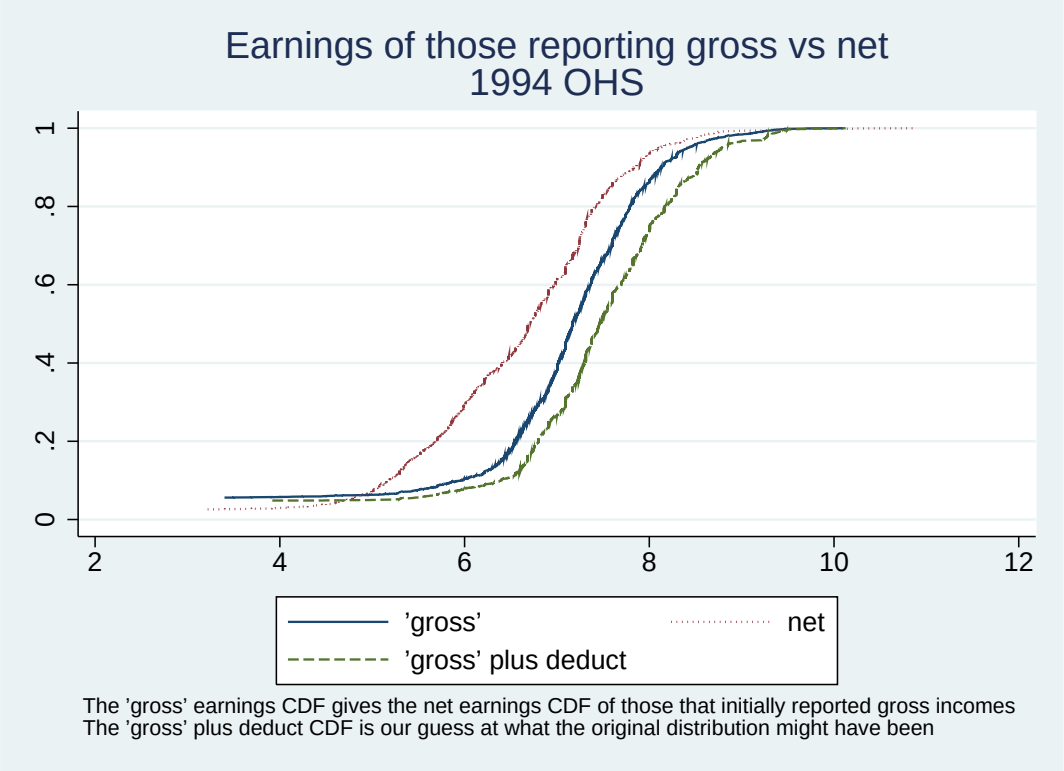


Figure 1: The distribution of earnings of individuals giving gross earnings versus those giving "take-home pay" in the 1994 OHS.

be? If we had the original responses from the 1994 OHS it would be possible to answer this question. Unfortunately (as discussed in more detail below) the responses have been heavily processed and only a "net income per month (calc)" variable released. Nevertheless we see in Figure 1 that if we graph the distributions by whether or not the person initially gave gross figures, we get very dissimilar CDFs, with strong evidence that those giving "gross" figures reported considerably higher incomes. At face value this suggests that individuals who found it easy to recall their gross income tend to be disproportionately higher earners. This may be a reflection of the fact that formal sector workers with formal pay slips may have better knowledge of their gross income. It also suggests that they would have a higher propensity to report gross incomes. The difference between the CDFs is sufficiently marked that one might presume that there was a processing error, i.e. that the reported net income (labelled 'gross' in the figure) was actually not an after-deductions figure. Indeed if we deduct the reported expenses from the "net" figures of those initially reporting gross figures we get a CDF very similar to those giving "take-home pay", at least at higher earnings. At lower earnings, however, the "net" figures really seem post-deductions, because subtracting expenses from these figures results in implausibly low earnings. Indeed for 10% of the figures the deductions would exceed the salary.

This all raises questions not only about the initial responses but precisely what sort of post-data collection processing occurred. Unfortunately with the current data we cannot reconstruct precisely what the information means. One additional problem is that the "deductions" question

Table 2: Responses to the Income bracket question in OHS 1995 by earnings period

Total salary at main job	Number of Responses Earnings period				Percentage of Valid Responses Earnings period			
	Day	Week	Month	Year	Day	Week	Month	Year
1. No income		3	29	1		0.6%	0.6%	0.2%
2. R1 - R999	26	491	1321	2	100.0%	97.2%	25.5%	0.4%
3. R1 000 - R1 249		5	546	1		1.0%	10.5%	0.2%
4. R1 250 - R1 499		2	389	1		0.4%	7.5%	0.2%
5. R1 500 - R1 999		3	547	4		0.6%	10.6%	0.7%
6. R2 000 - R2 499		1	443	5		0.2%	8.5%	0.9%
7. R2 500 - R2 999			478				9.2%	
8. R3 000 - R3 999			546				10.5%	
9. R4 000 - R5 999			538	5			10.4%	0.9%
10. R6 000 - R7 999			192	14			3.7%	2.6%
11. R8 000 - R9 999			70	7			1.4%	1.3%
R10 000 and above			84	506			1.6%	92.7%
Total	26	505	5183	546	100.0%	100.0%	100.0%	100.0%

Note: Unweighted counts and percentages of individuals giving nonmissing responses

in 1994 (question 3.14) is aked only of individuals reporting a "gross" figure. It is therefore not possible to create a gross earnings variable for all respondents in 1994.

To investigate the relationship between deductions and earnings, we ran a regression (of log deductions on log earnings). The estimates for 1994 suggest that deductions increased from around 6% at low incomes to around 38% at incomes of R10 000. In 1995 the deductions question was asked of everyone and it is therefore easier to make sense of the information. Rerunnning the regressions we found a similar relationship with a similar level of fit (R^2 of around .66), with deductions increasing from negligible levels to around 40% with incomes of R10 000.

3.1.2 1995

While the earnings question as a whole is better in 1995, the earnings brackets used in that year are extraordinary, to say the least (see Figure 10 and Figure 11). The lowest category with positive earnings extends in both cases from R1 to R999. Given the shape of South Africa's earnings distribution the bulk of African monthly earnings would be in this category. Even more remarkably, since the earnings period was asked after the bracket question, daily paid workers that ticked this category could be earning anywhere between R1 and R22 000 per month (assuming 22 working days per month). As shown in Table 2 there are over 500 respondents in this category who were paid either daily or weekly.

3.1.3 1996

We noted above that the 1996 survey collected information **only** in brackets. Furthermore the bracket questions are now set up so that the conversion between different reporting periods is made explicitly in the question (see Figures 12 and 13) so that there is only one set of brackets. Moreover the earnings brackets for employees become the standard for all the later OHSs and LFSs. One additional point worth noting is that unlike in the other OHSs, it is impossible to deduce from the released figures whether the respondent worked in terms of a weekly, monthly or annual reporting period.

3.2 The later OHSs: 1997 to 1999

Compared to the earlier period the survey instruments show much fewer changes (see Figure 14). The question that asks for a Rand amount for **employees** is phrased in 1997 and 1998 as :

> What is (the person's) total salary/pay (including overtime and bonus) at the MAIN job (before any deductions)?

In 1999 the question is yet more specific:

> What is ..... 's total salary/pay at the main job? Including, overtime, allowances and bonus, before any tax or deductions. Give amount in figures, without any text

As Daniels (2013) notes, the reporting period also shows some changes. In 1997 and 1998 the question allows daily, weekly and monthly amounts, i.e. the annual category has been dropped (compared to 1994 and 1995), while in 1999 the reporting periods are weekly, monthly and annual. He notes that over this period Statistics South Africa introduce new categories of "missing" - besides "unspecified", there now is also "don't know" (introduced in 1999). Furthermore in the 1999 survey it is possible to pinpoint individuals who respond on behalf of others. It transpires that such individuals are much more likely to give bracket information than Rand amounts (Daniels 2013, Table 4, p.27).

The earnings question for **employers or the self-employed** changes in several respects. The questions in 1997 and 1998 were to all intents and purposes identical to the one for 1996, except that provision was made for giving a Rand amount (see Figure 15). The reporting period in this case is weekly, monthly and annual. In 1999, however, the question asks about "total income" only, dropping all references to "turnover". Concomitantly the question about expenses incurred in obtaining the income was dropped (see Figure 16). Clearly the intention is that business deductions (raw materials and salaries of employees) should be netted out. Of course many self-employed would not be affected by this. Furthermore the earnings brackets now correspond to those for employees, i.e. the top two categories of the "self-employed" scale get lost.

What is the impact of these changes? In 1997 and 1998 it is possible to calculate both the "gross turnover" figure and the residual (i.e. "earnings") after business expenses have been deducted. One big difference is that in around 16% cases in 1997 and 14% in 1998 the deductions exceed the gross turnover figure[1], resulting in negative earnings. From 1999 onwards the earnings question presupposes that only nonnegative earnings are possible[2]. This raises the question what to do with the negative values in earlier years. We have chosen to set them to missing. Figure 2 shows the impact of the expenses on the cumulative distributions of self-employment earnings in 1997 and 1998. It is evident that correcting for expenses makes a difference. It is also noticeable that the distribution function for 1998 is well to the left of that for 1997. Part of the reason for this is that after 1997 the surveys were progressively more successful in finding small scale "informal" activities, significantly lengthening of the left tail of the distribution of self-employment earnings.

Figure 3 looks at the self-employment distributions for 1997, 1998 and 1999 (after deducting business expenses) for those reporting earnings above R2500 per month. The shift in the question between 1998 and 1999 does not seem to have affected the distribution.

3.3 The LFSs: 2000 to 2007

As noted above, the Labour Force Surveys no longer distinguish between those working for themselves (employers/self-employed) and those working for others in the earnings question.

[1] This is after correcting the expenses given by annual paid workers down, for reasons discussed below.

[2] This raises the interesting question why negative earnings should be ruled out *a priori* given that some entrepreneurs do go bankrupt.

Figure 2: The expenses make a noticeable difference to the cumulative distributions

Figure 3: Cumulative distributions for people reporting nominal self-employment earnings above R2500 per month. Business expenses have been deducted in the case of the 1997 and 1998 datasets.

Table 3: Earnings Information by Survey

Survey	Wage Information	Self-employment Income	Both	Total
OHS 1994	29,689	5,508	299	35,496
	83.64	15.52	0.84	100
OHS 1995	26,662	4,675	298	31,635
	84.28	14.78	0.94	100
OHS 1996	12,887	1,031	206	14,124
	91.24	7.3	1.46	100
OHS 1997	23,397	2,411	62	25,870
	90.44	9.32	0.24	100
OHS 1998	14,647	1,611	136	16,394
	89.34	9.83	0.83	100
OHS 1999	20,042	3,054	229	23,325
	85.92	13.09	0.98	100
Total	127,324	18,290	1,230	146,844
	86.71	12.46	0.84	100

Raw frequencies and row percentages are shown in the table

The earnings question (see Figure 17) looks very similar to the OHS 1999 question for employees and indeed comparable to the 1999 OHS question for the self-employed, except that the reference is now to income from the **main job** instead of from all own account business activities. One additional implication of this change is that it becomes impossible to answer both questions. In the previous OHSs there were always a handful of individuals who gave information (whether categorical or value) on both, as shown in Table 3.

3.4 The QLFSs: 2008 onwards

With the introduction of the QLFS, the earnings question was initially dropped. The reason for this was criticism from an IMF delegation that assessed labour market statistics. Their objection (cited in Statistics South Africa's official response) was that

> Data on earnings are collected each survey but considered to be poor quality, especially for the self-employed. Question has relatively high refusal/non-completion rates. Data are not published; hence the importance of the data is not appreciated by the survey officers. (Statistics South Africa 2008, Section 2.3.5, pp.7-8)

The earnings question reappeared in the third quarter of 2009, although the data from 2009 have not been released. The QLFS earnings module shows a reversion to two questions, one for employees and one for employers and own account workers (see Figures 18 and 19 respectively). The same set of brackets were used for both (see Figure 20). Unlike the OHSs, it was impossible to give both a wage and a self-employment income, because there was a prior screening question (Question 5.1) which categorised individuals as either working for someone else or as an employer or own-account worker. Another difference with respect to both the OHS and LFS is that the reporting period now includes the possibility of being paid fortnightly or hourly.

One additional point to note is that the QLFS with earnings information was released separately, under the label of "Labour Market Dynamics in South Africa" study.

4 Analysing the Response Patterns

While changes in the instrument are an obvious source of differences between surveys, there are a number of other important factors. The first of these is changes in coverage, both for the survey as a whole and which people ended up answering the earnings question. The second is field work practice, i.e. the extent to which fieldworkers probed or were able to elicit good responses. The third is respondent behaviour. Fourth is post-fieldwork data processing. In practice very little of how these processes were implemented or changed over time is documented. Much of it has to be deduced from the response patterns observable in the data itself.

4.1 Changes in Coverage

There are several important "breaks" in the coverage of the earnings question. It is well-known that the earliest OHSs did not cover mining employment well. This will obviously impact on the representation of low earning categories of work. Kerr and Wittenberg (2013) argue that the OHSs up to 1998 significantly underenumerated backyard shacks and small households. This led to the underenumeration of certain categories of employed people, probably disproportionately low earners (such as domestic workers).

With the introduction of the LFSs, there was a much greater emphasis on capturing informal activities, including subsistence agriculture (see Figure 4). The coverage of those classified as "employed" therefore increased substantially. Since many of these were low earners, this will have an appreciable impact on the measured earnings distribution. With the introduction of the QLFSs the subsistence agriculture sector seems to disappear again.

There are several more idiosyncratic changes in coverage. The February 2001 LFS shows a much higher level of informal activity than any other survey in the 1994-2011 period. The reason is that the LFS served as screener survey for the Survey of Employers and the Self-employed (SESE). Fieldworkers were apparently paid on a piece-work basis (personal communication, Statistics SA staff) and the SESE questionnaire was administered after the LFS one. They therefore had a direct pecuniary interest in finding as much informal activity as possible.

Less explicable is the spike in subsistence agricultural production in the February and September 2000 LFS (see Yu 2007, Table 14,p.19). This may be due to changed fieldwork instructions, as speculated by Charles Meth (personal communication).

4.2 Brackets and Point Responses

Tables 9, 10 and 11 present the proportion of individuals giving point information (actual Rand amounts) within each of the surveys and separated by employees and the self-employed. Within each survey the probability of giving precise information declines noticeably with income. Nevertheless there are also clear differences between surveys. The response rate jumps between OHS 1999 and the LFSs, suggesting that enumerators were more successful in eliciting precise information, perhaps due to changed enumerator training, fieldwork instructions or fieldwork control. The QLFSs have no "brackets only" responses, indicating that Statistics South Africa imputed Rand amounts for these responses.

4.3 Changes in data cleaning practices

The 1994 data earnings data include imputations for brackets as well as conversion of all figures to a net monthly income figure. The imputation process seems to have been a variant of mean imputation, resulting in significant "spiking" in the distribution as shown in 5. Wittenberg

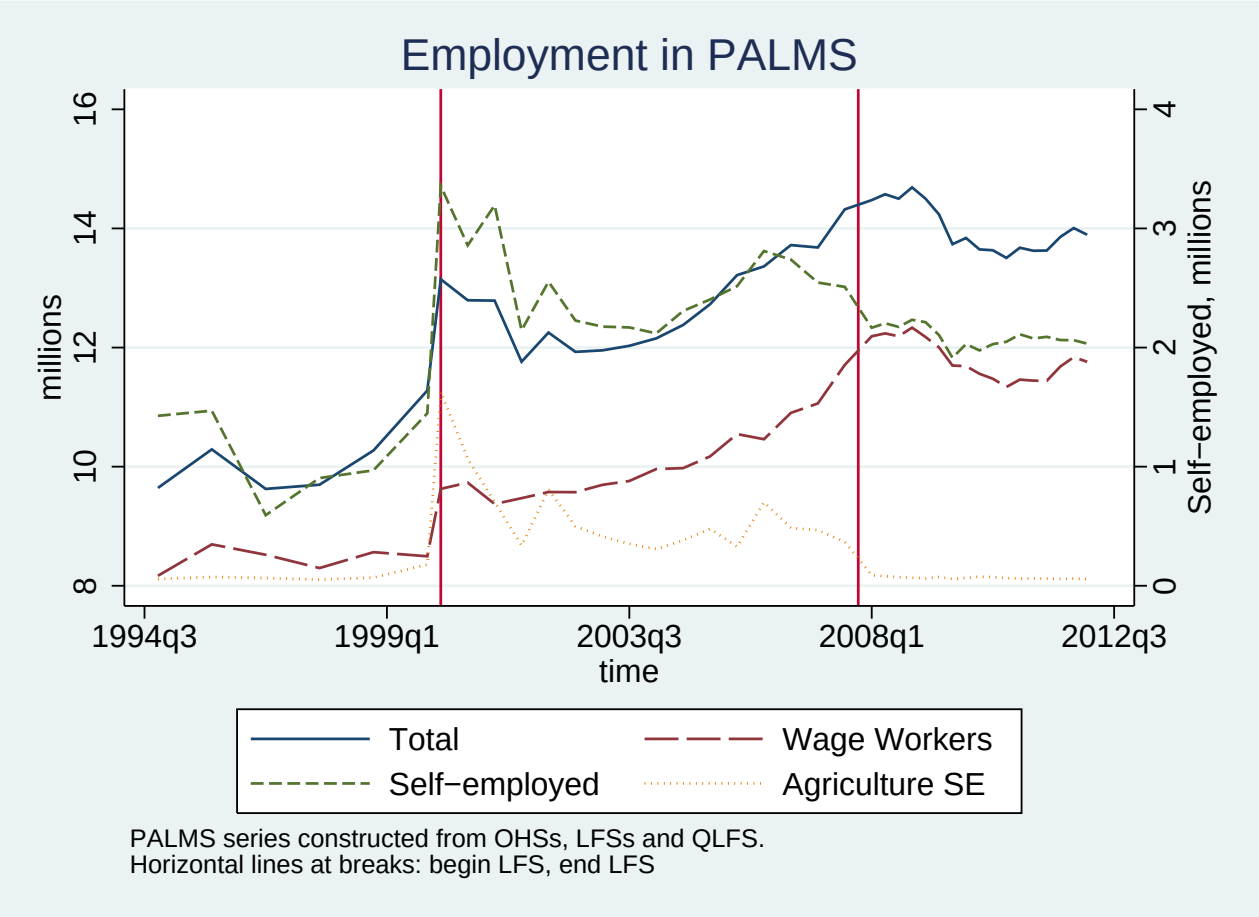


Figure 4: Measured self-employment jumps with the introduction of the LFS. The jump is particularly pronounced in agriculture.

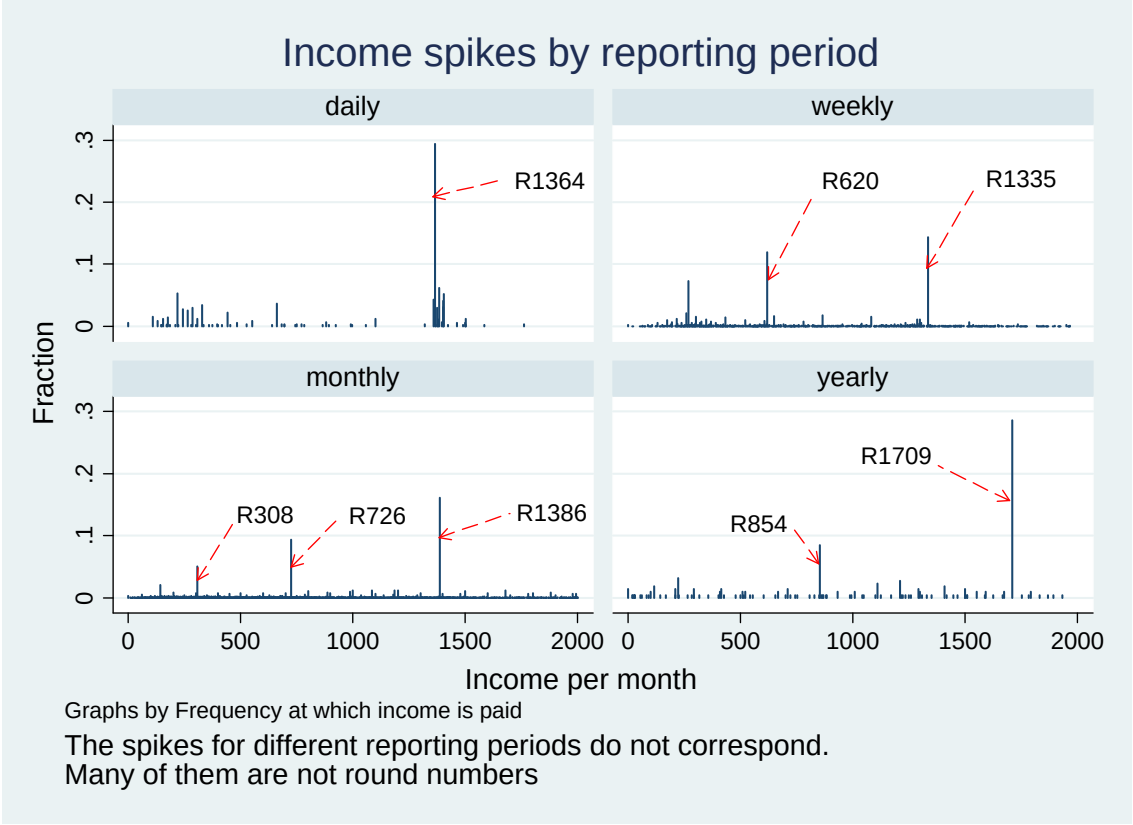


Figure 5: The imputation process in the 1994 OHS led to significant spiking of reported earnings at strange numbers (Wittenberg 2008a, Figure 1)

(2008a) tries to reconstruct in some detail the "spike generating process" in order to separate out genuine point information from bracket responses.

The imputations in the QLFS do not lead to similar spikes, probably because some variant of a hot deck was employed. Nevertheless hot decks can also induce more limited forms of heaping, if the same value is copied to several other observations. Arguably this is what might have happened in the 2011 Q2 version of the QLFS, where the 14 largest monthly earnings figures among the self-employed are R 250 000 (4 observations), R 333 666 (2 cases), R 600 000 (2 times) and R 1 000 000 (6 individuals). The clumping of the millionaires is sufficient to increase the mean income by around R 2 000 in that quarter.

There are other forms of data cleaning that have changed over time. In 1997 everyone who has any form of earnings information is placed into a bracket. This suggests some post-fieldwork imputation since the questionnaire is clear that the categories should be shown only if the respondent refuses to give a Rand amount. Interestingly in both 1997 and 1998 there are many cases where the point information contradicts the bracket (see below). By 1999 this hardly occurs at all.

Table 4: Mismatch between Rand and bracket information in the 1997 OHS

	Recorded bracket information											
Actual	0	1	2	3	4	5	6	7	8	9	10	11
1	0	49	0	0	0	0	0	0	0	0	0	0
2	1	25	1,591	757	1,086	85	52	12	13	2	2	1
3	2	28	130	2,786	208	832	696	22	7	14	6	3
4	3	15	60	49	2,875	76	64	194	48	11	7	7
5	2	4	46	29	26	2,755	142	5	9	27	21	1
6	0	5	44	35	17	60	3,181	136	7	4	9	17
7	2	0	11	37	1	6	34	1,751	68	5	0	0

Actual brackets were constructed from actual Rand amounts, where given, and bracket responses otherwise. Only a subset of the full cross-tabulation is shown.

Table 5: Missing earnings information by survey

Survey	fraction	Survey	fraction
OHS 1994	0.002	LFS 04:2	0.069
OHS 1995	0.014	LFS 05:1	0.058
OHS 1996	0.052	LFS 05:2	0.056
OHS 1997	0.048	LFS 06:1	0.037
OHS 1998	0	LFS 06:2	0.046
OHS 1999	0.079	LFS 07:1	0.041
LFS 00:1	0.090	LFS 07:2	0.049
LFS 00:2	0.038	QLFS 2010:1	0.003
LFS 01:1	0.051	QLFS 2010:2	0.002
LFS 01:2	0.061	QLFS 2010:3	0.002
LFS 02:1	0.061	QLFS 2010:4	0.002
LFS 02:2	0.072	QLFS 2011:1	0.002
LFS 03:1	0.061	QLFS 2011:2	0.002
LFS 03:2	0.079	QLFS 2011:3	0.003
LFS 04:1	0.061	QLFS 2011:4	0.002

4.4 Respondent or fieldworker errors

The mismatch between the Rand information and the brackets is shown in Table 4. It is evident that there are a lot of off-diagonal elements. Given the wording of the question (see Figure 14) it is difficult to know how it was possible for respondents to answer both, let alone produce inconsistent replies. It is evident that there must have been some combination of fieldworker and respondent error. It raises the issue of how to deal with the conflicting information. We have chosen to take the actual Rand amount as authoritative, where it is available.

4.5 Missing values

Daniels (2013) has described the increased sophistication in the instrument in dealing with missing values. There are nonetheless some curious patterns in Table 5. It shows the fraction of respondents giving one of the "nonresponse" codes released with the data ("don't know", "refused" or "unspecified"), i.e. it excludes the "not applicable" (unemployed or out of the labour force).

Table 6: Missing earnings information in the LFSs by category

Survey	valid	DK	refused	unspecified
LFS 00:1	0.910	0.042	0.027	0.021
LFS 00:2	0.962	0.004	0.008	0.026
LFS 01:1	0.949	0.018	0.022	0.011
LFS 01:2	0.939	0.026	0.030	0.005
LFS 02:1	0.939	0.026	0.033	0.002
LFS 02:2	0.928	0.032	0.037	0.003
LFS 03:1	0.939	0.023	0.036	0.002
LFS 03:2	0.921	0.025	0.051	0.002
LFS 04:1	0.939	0.021	0.038	0.003
LFS 04:2	0.931	0.026	0.040	0.004
LFS 05:1	0.942	0.018	0.036	0.004
LFS 05:2	0.944	0.019	0.034	0.003
LFS 06:1	0.963	0.009	0.026	0.002
LFS 06:2	0.954	0.013	0.030	0.003
LFS 07:1	0.959	0.012	0.024	0.004
LFS 07:2	0.951	0.014	0.028	0.007

A startling figure is the zero for 1998. It doesn't mean that there were no refusals – just that they were not recorded as such in the dataset. It is also seems clear that refusals in the QLFS data have been replaced by imputed values. Table 6 provides a more detailed look at the LFS cases. As Daniels (2013) has argued the measurement instrument had stabilised in this period, so changes between surveys are due either to "real" changes (respondent behaviour) or changes in fieldwork or data cleaning practice. The marked reduction in the proportion of "unspecified" responses is most likely a sign of better fieldwork and fieldwork control. Indeed it is difficult to fathom what an "unspecified" response could be given the availability of the other nonresponse codes. The steady reduction in "don't know" and "refused" answers also suggests more stringent fieldwork control. The case of the September 2000 LFS, with hardly any of these responses looks clearly anomalous.

4.6 Zero Earnings

Vermaak (2012) has argued that how one deals with individuals recording zero income can markedly affect one's analysis, particularly if one is investigating the "working poor". There are several mechanisms that might lead to a zero being recorded:

- The worker earns a positive income, but wants to signal that it is a pittance.
- The individual is self-employed and is not valuing consumption from own production or from inventories (in the case of traders). Zero surplus at the end of the period is equated with zero income.
- The individual works in a family enterprise and is not factoring in that they are receiving income in kind.
- The individual is self-employed and in fact made a loss, but can't report this.
- The individual is working as a volunteer (e.g. to gain experience)

- The individual is working on some deferrred compensation scheme. Own-account agriculture which may yield income only at harvest time may be an example of this also.

The majority of these cases are species of measurement error. They are also cases where it is somewhat doubtful whether the "data generating process" is directly comparable to those of individuals reporting positive earnings. Table 7 suggests that zero earners are a big issue only in the LFSs, and then only among the self-employed. Given that the LFSs are marked by a big increase in subsistence agricultural workers, it suggests that the high proportion of zero earners in the LFSs is an artefact of the change in coverage. This is demonstrated in the last column of Table 7 where we show that more than 90% of zero earners in the LFSs come from the agricultural sector. Consequently we have made no attempt to correct for zero earnings.

4.7 Outliers

While zero earners may influence the poverty rates, extremely large values have the potential to shift the mean, as argued by Burger and Yu (2007). Table 8 confirms their point that "millionaires" (in year 2000 Rands) are not evenly distributed over the surveys. The 1999 October Household Survey and the September 2000 Labour Force Survey look clearly anomalous within the OHS and LFS sequence. There also looks to be a break between the LFSs and the QLFSs. Indeed the QLFSs seem to show a markedly higher count of millionaires, induced, perhaps, by the imputation of missing information alluded to in section 4.3. Some of the most extreme values were recorded in the QLFSs – and they tend to occur in clumps. In the third quarter of 2010, for instance, there were three self-employed individuals recording an income of precisely R10 million (nominal) *per month*[3].

The presence or absence of such extreme values has a marked influence on average incomes in those surveys. It is therefore desirable to remove their influence. Three approaches were considered:

1. *Remove the "millionaires"*

   This was the approach taken by Burger and Yu (2007). This obviously lops off the top tail and risks removing "genuine millionaires" along with any bogus ones. It is also rather arbitrary in that there may be bogus observations at lower levels of income.

2. *Outlier detection*

   There are a number of outlier detection algorithms available (see (Billor, Hadi and Velleman 2000) for a review). The BACON algorithm (Billor et al. 2000) has been implemented as a Stata routine (Weber 2010) and was used here for that purpose. The basic problem of every outlier detection algorithm is that the presence of the outliers can contaminate any statistics calculated to detect those outliers. The BACON algorithm begins with a small subset of observations assumed safe from contamination and then incrementally adds (in blocks) observations that are "close" to the existing safe set. The distance measure used is the Mahalanobis distance $\sqrt{(\mathbf{x}_i - \overline{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_i - \overline{\mathbf{x}})}$ where $\overline{\mathbf{x}}$ and $\mathbf{S}$ are the mean and covariance matrix calculated on the "safe" set and $\mathbf{x}_i$ is the vector under consideration. One of the drawbacks of this procedure is that it works better on continuous data. Most of the covariates available are discrete and this blunts the ability of the procedure to find

[3] The three highest values recorded in the dataset are all in the first quarter of 2011, viz. 86 million, 68 million and 60 million (per month). Since these all fall outside the "valid range" noted in the metadata, they are undoubtedly bad data.

Table 7: Proportion of zero earners by survey

Survey	employee	self employed	agriculture[(a)]
OHS 1994	0.039	0.010	0.218
OHS 1995	0.001	0.009	0.149
OHS 1996	0.004	0.007	0.000
OHS 1997	0.002	0.014	0.031
OHS 1998	0.000	0.035	0.091
OHS 1999	0.001	0.028	0.897
LFS 00:1	0.001	0.445	0.991
LFS 00:2	0.001	0.351	0.968
LFS 01:1	0.001	0.187	0.963
LFS 01:2	0.001	0.128	0.887
LFS 02:1	0.001	0.307	0.964
LFS 02:2	0.001	0.203	0.922
LFS 03:1	0.001	0.170	0.965
LFS 03:2	0.001	0.158	0.826
LFS 04:1	0.001	0.127	0.928
LFS 04:2	0.002	0.188	0.920
LFS 05:1	0.001	0.250	0.961
LFS 05:2	0.001	0.164	0.912
LFS 06:1	0.001	0.311	0.966
LFS 06:2	0.001	0.242	0.880
LFS 07:1	0.002	0.257	0.941
LFS 07:2	0.001	0.183	0.855
QLFS 2010:1	0.002	0.006	0.059
QLFS 2010:2	0.001	0.004	0.071
QLFS 2010:3	0.001	0.006	0.056
QLFS 2010:4	0.001	0.001	0.000
QLFS 2011:1	0.002	0.003	0.000
QLFS 2011:2	0.001	0.004	0.000
QLFS 2011:3	0.002	0.003	0.000
QLFS 2011:4	0.001	0.005	0.000

Note:
(a) Proportion of self-employed zero earners in Agriculture

Table 8: Millionaires by survey

Survey	n	proportion	Population total
OHS 1994	6	0.0004	7 394
OHS 1995	17	0.0007	9 524
OHS 1996	N.A.	N.A.	N.A.
OHS 1997	6	0.0003	5 853
OHS 1998	12	0.0012	11 633
OHS 1999	48	0.0036	41 051
LFS 00:1	3	0.0004	3 895
LFS 00:2	23	0.0010	16 804
LFS 01:1	5	0.0002	4 175
LFS 01:2	4	0.0002	2 466
LFS 02:1	1	0.0001	734
LFS 02:2	2	0.0001	2 830
LFS 03:1	1	0.0001	3 072
LFS 03:2	2	0.0001	1 126
LFS 04:1	1	0.0001	2 739
LFS 04:2	1	0.0001	85
LFS 05:1	0	0	0
LFS 05:2	6	0.0003	7 884
LFS 06:1	2	0.0001	3 286
LFS 06:2	3	0.0002	2 564
LFS 07:1	6	0.0003	3 215
LFS 07:2	3	0.0002	4 451
QLFS 2010:1	18	0.0009	15 665
QLFS 2010:2	22	0.0010	17 027
QLFS 2010:3	23	0.0011	19 790
QLFS 2010:4	12	0.0006	9 250
QLFS 2011:1	11	0.0006	9 304
QLFS 2011:2	22	0.0011	18 920
QLFS 2011:3	10	0.0005	6 168
QLFS 2011:4	9	0.0004	3 763

Notes:
Millionaires defined by annual earnings in year 2000 Rands
Proportion based on sample giving Rand information.
Population total estimate includes bracket correction.

anomalous data. It should be noted that these distance measures work better also with symmetrically distributed data, so the earnings data was logged before applying the routine.

The BACON algorithm found only five clearly anomalous observations when using education categories as covariates. The flagged outliers were, indeed, all anomalous – including the three bad data points referred to in footnote 3.

3. *Extreme regression residuals*

   The BACON algorithm treats all variables used to locate outliers symmetrically. Economists tend to think of earnings as the outcome in which education and, perhaps, occupation, are the explanatory variables. Consequently a Mincerian style regression was run with survey specific intercepts, gender, race (both interacted with survey), a quadratic in age, education categories and occupation categories as explanatory variables. One issue that has to be addressed when trying to identify "extreme" residuals, is that points of high leverage will tend to be associated with smaller residuals. The "Studentised residual" corrects for that. It is defined as

$$r_i = \frac{e_i}{\sqrt{s^2_{(i)}\left(1 - h_i\right)}}$$

   where e_i is the "unstandardised" residual, $s^2_{(i)}$ is an estimate of the residual variance with the i-th observation removed and h_i is the leverage (StataCorp 2013a, "regress postestimation" p.1877).

   "Studentized residuals can be interpreted as the t statistic for testing the significance of a dummy variable equal to 1 in the observation in question and 0 elsewhere (Belsley, Kuh, and Welsch 1980). Such a dummy variable would effectively absorb the observation and so remove its influence in determining the other coefficients in the model." (StataCorp 2013a, p.1877)

   Studentised residuals with absolute values greater than five were flagged as extreme. The probability of being flagged if the underlying population is well-behaved is $5.735{*}10^{-7}$. On a sample of around 500 000, one would expect on average to see 0.3 observations this extreme. In practice the procedure flagged 476 observations, including a number deemed to be implausibly low.

4. *Robust regression*

   One problem with using the "Studentised residual" approach is that the presence of a group of outliers will again contaminate the data, making it harder to identify the bad observations. Robust regression attempts to deal with this issue recursively: progressively downweighting observations that appear to be problematic until the regression results and the weights are in agreement. Running this procedure led to over 1000 observations being weighted right down to zero. Every one of the "extreme" values identified through the Studentised residuals was also identified as problematic by the robust regression.

In order to be not overly aggressive in removing high values we opted to use the third approach as the preferred approach for dealing with outliers.

5 Ways of dealing with Coarse Data

Given the fact that the probability of providing bracket-only information increases with income (see Section 4.2) one has to deal explicitly with the bracket responses in order to obtain unbiased estimates. There are many ways in which this can be done.

5.1 Deterministic imputation

The most common procedure (discussed, for instance, by von Fintel (2007)) is to place individuals at either the midpoint or the mean of the observations within the bracket. In the case of the "open" category there is no midpoint, so the procedure is to place observations at some deterministic multiple of the lower bracket boundary, as discussed earlier. Both approaches will lead to artificial spikes in the data and will definitely distort moments other than the mean. They also do not acknowledge that the imputed value is actually the true value with some noise.

5.2 Reweighting

A simple alternative to imputation is to reweight the point information to take into consideration the bracket responses (Wittenberg 2008b). Numerically it will provide identical results to the imputation of means, but it has the advantage of not altering any of the other moments. Among the disadvantages of this approach is that it cannot be used when there is no point information within a bracket and it tends to inflate the standard errors, since the approach only fully counts observations with actual Rand amounts.

5.3 Stochastic Imputation

One way of avoiding putting all imputed values at the same number (e.g. midpoint or mean) is to draw the imputed value from some random distribution, effectively adding some noise to the imputed value. There are again different ways in which this can be done.

5.3.1 Parametric

A distribution can be specified beforehand and the draw taken from there. Woolard and Woolard (2006), for instance, used a uniform distribution within a bracket. Alternatively one could take draws from a truncated log-normal[4]. Obviously the distribution chosen will have an influence on where the observations are placed within the brackets and hence will affect statistics calculated with those values.

5.3.2 Posterior distribution

An additional problem with taking a draw from a distribution whose parameters have been empirically estimated (e.g. the truncated log-normal) is that the process does not acknowledge the uncertainty implicit in the estimation process. One can address this by first taking a draw from the posterior distribution of the estimators (e.g. draw a σ_*^2 from the chi-square distribution[5] and then draw μ_* from a normal distribution centred on $\widehat{\mu}$ with variance $\frac{\sigma_*^2}{n-1}$). The actual imputation would then be drawn from the distribution defined by the parameter μ_* and σ_*^2.

[4] This would require estimating the parameters of the distribution, which could be done using a maximum likelihood approach.

[5] with a non-informative prior we have that

$$\frac{\widehat{\sigma}^2 (n-1)}{\sigma^2} \sim \chi_1^2$$

So the posterior distribution of σ^2 will be given by

$$\frac{\widehat{\sigma}^2 (n-1)}{\chi_1^2}$$

5.3.3 Hot deck

In the case of hot decks, the draw is from the empirical distribution, in this case observations falling into the same bracket. There are different flavours: "deterministic" ones in which the "nearest neighbour" is drawn according to some criterion (typically involving a number of covariates) and more "random" ones in which one value is picked out of a pool (for a fuller discussion see Andridge and Little 2010). The advantage of the hot deck (like reweighting) is that only data that are empirically observed will be utilised. In cases where the "donor pool" is small, it can lead to the same value being copied to a number of other observations.

5.3.4 Predictive mean matching

The key issue for a hot deck is how the "distance measure" is defined. The problem is that the variables on which the "pool" is based need to be correlated with the outcome to be imputed but also related to the nonresponse mechanism. An attractive option is to regress the outcome (in this case earnings) on the available covariates (in this case dummies for earnings brackets, gender, education, and whether it was an employee or somebody self-employed). Distance is then measured by the distance between the predicted outcomes (which is defined even for indiviuals missing the dependent variable). Of course this again treats the estimates as fixed, so it is preferable to take the draw from the posterior distribution of predicted outcomes.

5.4 Multiple imputations

One of the big problems with stochastic imputation is that the resulting value does not signal any of the uncertainties implicit in the process of producing the imputation. The resulting value will be the true value plus some measurement error, but the error has been rendered invisible. The theoretical solution for this problem is to do the imputation multiple times and perform any statistical analyses (e.g. calculation of summary statistics) on all of the resulting datasets (Rubin 1987, StataCorp 2013b). By taking into consideration the differences in estimates between analyses one can obtain appropriate point estimates and standard errors.

5.5 Maximum Likelihood

There are situations in which one can sidestep the issue of imputation altogether. If one is willing to assume a parametric form for the statistical process it is possible to accommodate both coarsened data and the point information within the overall likelihood function. "Interval regression" and "censored regression" models are examples of these, but the approach can be used more generally. These approaches make strong assumptions which may not be all that attractive if there is any doubt about the applicability of the assumed distribution

6 Dealing with Missing Information

All of the approaches outlined in the case of coarsened data can also be used if the data is missing altogether. The "reweighting" approach fails if there are no actual observations to reweight. The 1996 OHS survey records information only in brackets and so the reweighting approach is not really applicable. Instead we took multiple draws from the 1997 real earnings distribution, using predictive mean matching, taking the 1996 covariates into account.

In cases where the bracket information was missing, we (multiply) imputed brackets using an ordered logit model with province, gender, education, race, a quadratic in age and occupation

as explanatory variables. The imputed brackets were then (along with gender and education) to multiply impute Rand amounts, using predictive mean matching.

7 Harmonising the sample weights

One final issue needs to be addressed if one is to obtain a consistent set of estimates over time: the harmonisation of the sample weights released with each survey. Branson and Wittenberg (forthcoming) have argued that there are also breaks in the way in which survey weights were constructed and that this can lead to jumps in a series which are unrelated to "real" changes. The weights that have been released with the PALMS 2 dataset have all been calibrated to the ASSA 2008 demographic model.

8 The impact of data quality adjustments on average real earnings

Figure 6 shows the impact of different types of data quality adjustments on the estimation of average earnings over the period 1994 to 2011. The dotted line indicates that outliers would have moved the mean up substantially in both 1999 and September 2000. The gap between the green dashed line and the other lines indicates the importance of taking appropriate account of the bracket information. It is evident that the multiply imputed sequence (indicated as "imputed" in the figure) produces a slightly higher income than the other two approaches, mainly because it also imputes for observations which were completely missing.

Interestingly the pattern suggested in this Figure does not suggest that there were dramatic movements downwards in average wages. The "dip" at the point at which the LFSs were introduced can be explained in terms of changes in coverage. Overall the picture suggests real earnings growth since 2000. In the period of the QLFSs there are no discernible differences between the series anymore, because missing information and brackets have already been imputed for.

Figure 7 presents the analogous graph in the case of the self-employed. We do not show the line with outliers, since the spikes are so huge that no changes can be discerned in any of the other parts of the graph. Again we observe the importance of dealing with brackets. We again observe that imputing for the missings raises average earnings a bit. In this case we see a big shift in mean earnings precisely at the point where the LFS was introduced. This coincides with the big change in coverage noted in Figure 4. Again the period since 2000 seems to show growth in average real earnings. The curious spike in the QLFS period is due to the clump of "millionaires" referred to earlier. These values were not removed by the outlier detection algorithm and hence the spike appears in the picture.

9 Conclusion

Given the changes documented in this paper one might be inclined to despair of being able to say anything meaningful about the evolution of earnings since 1994. We believe, however, that it is possible to discern some meaningful trends through the noise generated by these measurement changes. Any attempt to grapple with South African labour trends in the post-Apartheid period will at some stage confront the issue of dealing with earnings data. We hope that the PALMS 2 dataset will contribute to a debate that isn't derailed by "dirty" data.

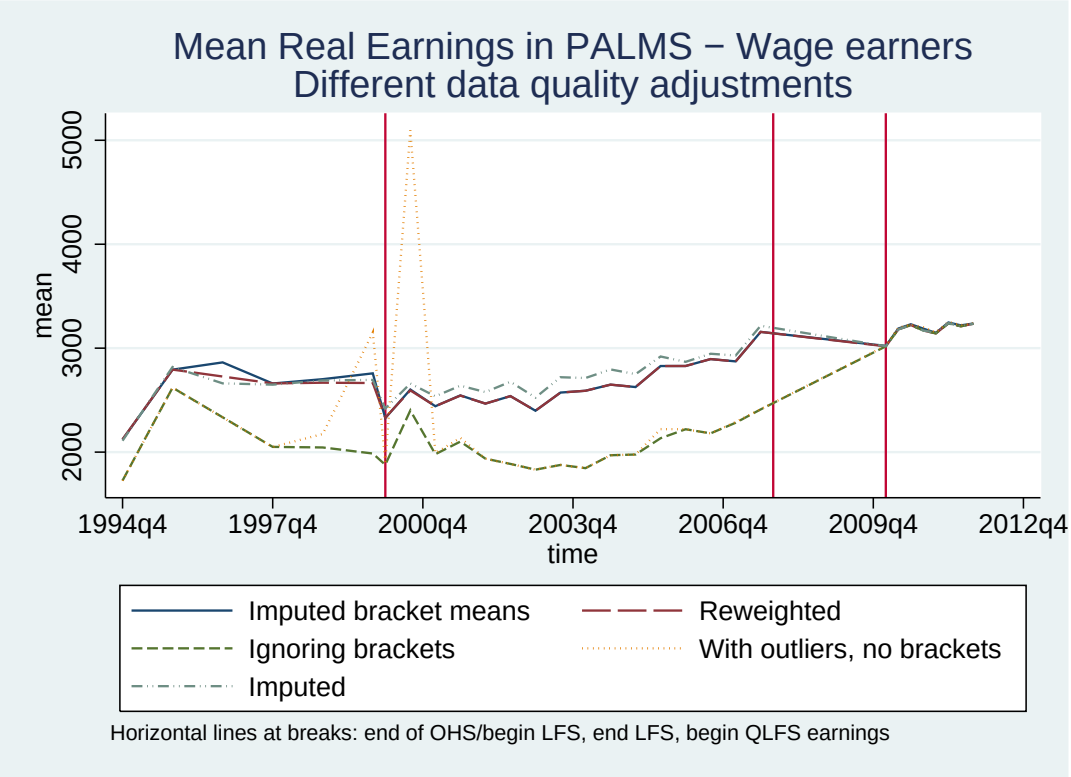


Figure 6: The impact of different data quality adjustments on the estimation of average real earnings among wage employees

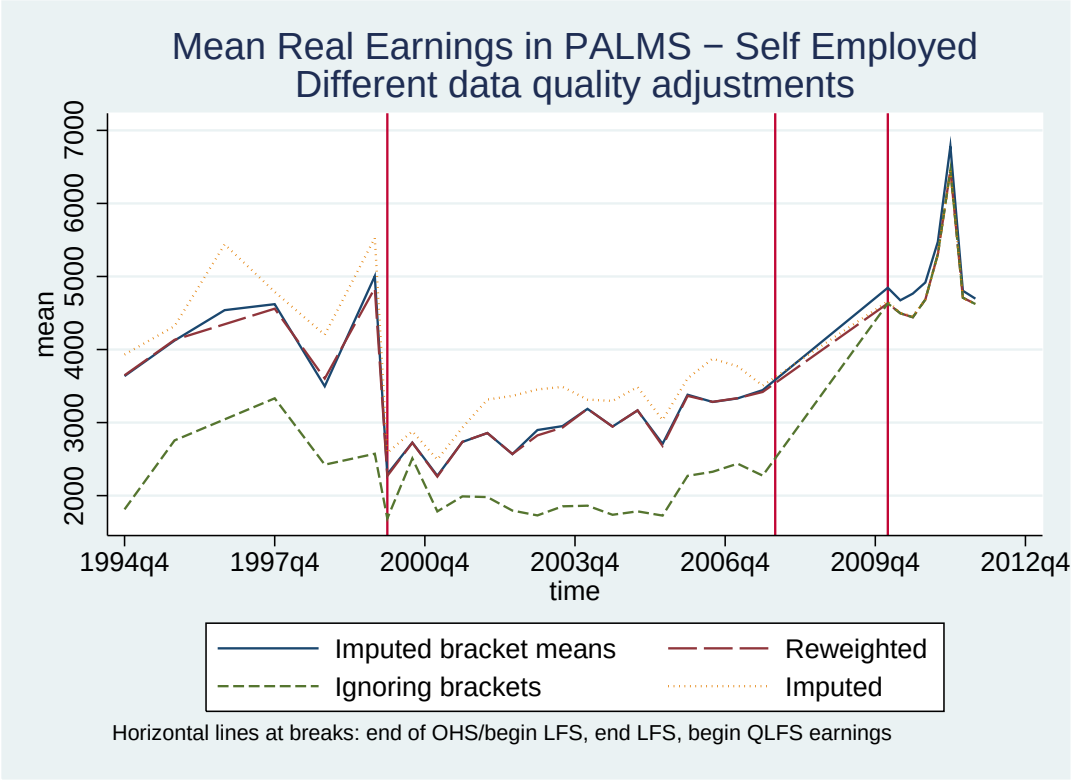


Figure 7: The impact of data quality adjustments on estimated average earnings among the self-employed

References

Andridge, R. R. and Little, R. J. A.: 2010, A review of hot deck imputation for survey non-response, *International Statistical Review* **78**(1), 40–64.

Billor, N., Hadi, A. S. and Velleman, P. F.: 2000, BACON: blocked adaptive computationally efficient outlier nominators, *Computational Statistics and Data Analysis* **34**, 279–298.

Branson, N. and Wittenberg, M.: 2007, The measurement of employment status in South Africa using cohort analysis, 1994-2004, *South African Journal of Economics* **75**(2), 313–326.

Branson, N. and Wittenberg, M.: forthcoming, Reweighting South African national household survey data to create a consistent series over time: A cross-entropy estimation approach, *South African Journal of Economics* .

Burger, R. and Yu, D.: 2007, Wage trends in post-Apartheid South Africa: Constructing an earnings series from household survey data, *Working Paper 07/117*, Development Policy Research Unit, University of Cape Town.

Casale, D.: 2004, What has the feminisation of the labour market 'bought' women in South Africa? Trends in labour force participation, employment and earnings, 1995-2001, *The Journal of Interdisciplinary Economics* **15**, 251–275.

Daniels, R.: 2013, Questionnaire design and response propensities for employee income micro data, *Working Paper 89*, SALDRU, University of Cape Town. Revised version. Retrieved from www.saldru.uct.ac.za on 18 April 2013.

Kerr, A., Lam, D. and Wittenberg, M.: 2013, Post-Apartheid Labour Market Series 1994-2012 [dataset] version 2.1 of the harmonised dataset based on Statistics SA's OHS and LFS surveys, 1994-2012. available at http://www.datafirst.uct.ac.za/catalogue3/index.php/catalog/434.

Kerr, A. and Wittenberg, M.: 2013, Changes in Statistics South Africa's field work and sampling methodology in the 1999 October Household Survey, mimeo, DataFirst, University of Cape Town.

Rubin, D.: 1987, *Multiple Imputation for Nonresponse in Surveys*, Wiley, New York.

StataCorp: 2013a, *Stata 13 Base Reference Manual*, Stata Press, College Station, TX.

StataCorp: 2013b, *Stata 13 Multiple Imputation Reference Manual*, Stata Press, College Station, TX.

Statistics South Africa: 2008, Report on the response by Statistics South Africa to recommendations of the International Monetary Fund on improvements to the labour force survey, mimeo. Available at http://www.statssa.gov.za/qlfs/docs/improvements_to_lfs.pdf, downloaded on 8 June 2013.

Vermaak, C.: 2012, Tracking poverty with coarse data: evidence from South Africa, *Journal of Economic Inequality* **10**, 239–265.

von Fintel, D.: 2007, Dealing with earnings bracket responses in household surveys: How sharp are midpoint imputations?, *South African Journal of Economics* **75**(2), 293–312.

Weber, S.: 2010, st0197. bacon: An effective way to detect outliers in multivariate data using Stata (and Mata), *Stata Journal* **10**(3), 331–338.

Wittenberg, M.: 2008a, Income in the October Household Survey 1994, Working paper, DataFirst, University of Cape Town.

Wittenberg, M.: 2008b, Nonparametric estimation when income is reported in bands and at points, *Working Paper 94*, ERSA. available at http://www.econrsa.org/papers/w_papers/wp94.pdf.

Wittenberg, M.: 2013, Fat tales of South Africa's income distribution, mimeo, School of Economics, University of Cape Town.

Wittenberg, M.: forthcoming, Data issues in South Africa, *in* H. Bhorat, A. Hirsch, R. Kanbur and M. Ncube (eds), *Oxford Companion to the Economics of South Africa*, Oxford University Press.

Woolard, I. and Woolard, C.: 2006, Earnings inequality in south africa 1995–2003, *Employment, Growth and Development Initiative, Occasional Paper 1*, HSRC, Cape Town.

Yu, D.: 2007, The comparability of the Statistics South Africa October Household Surveys and Labour Force Surveys, *Stellenbosch Economic Working Papers 17/07*, University of Stellenbosch, Economics Department. Available from www.ekon.sun.ac.za/wpapers/2007/wp172007/wp-17-2007.pdf, downloaded 8 June 2013.

Yu, D.: 2009, The comparability of Labour Force Survey (LFS) and Quarterly Labour Force Survey (QLFS), *Stellenbosch Economic Working Papers 08/09*, University of Stellenbosch, Economics Department. Available from http://www.ekon.sun.ac.za/wpapers/2009/wp082009/wp-08-2009.pdf, downloaded 8 June 2013.

Yu, D.: 2011, Some factors influencing the comparability and reliability of poverty and inequality estimates across household surveys, Paper presented to MASA conference. Available at http://www.aceconferences.co.za/MASA%20FULL%20PAPERS/Submission%20-%20Derek%20Yu1.pdf.

Section 3 (continued)

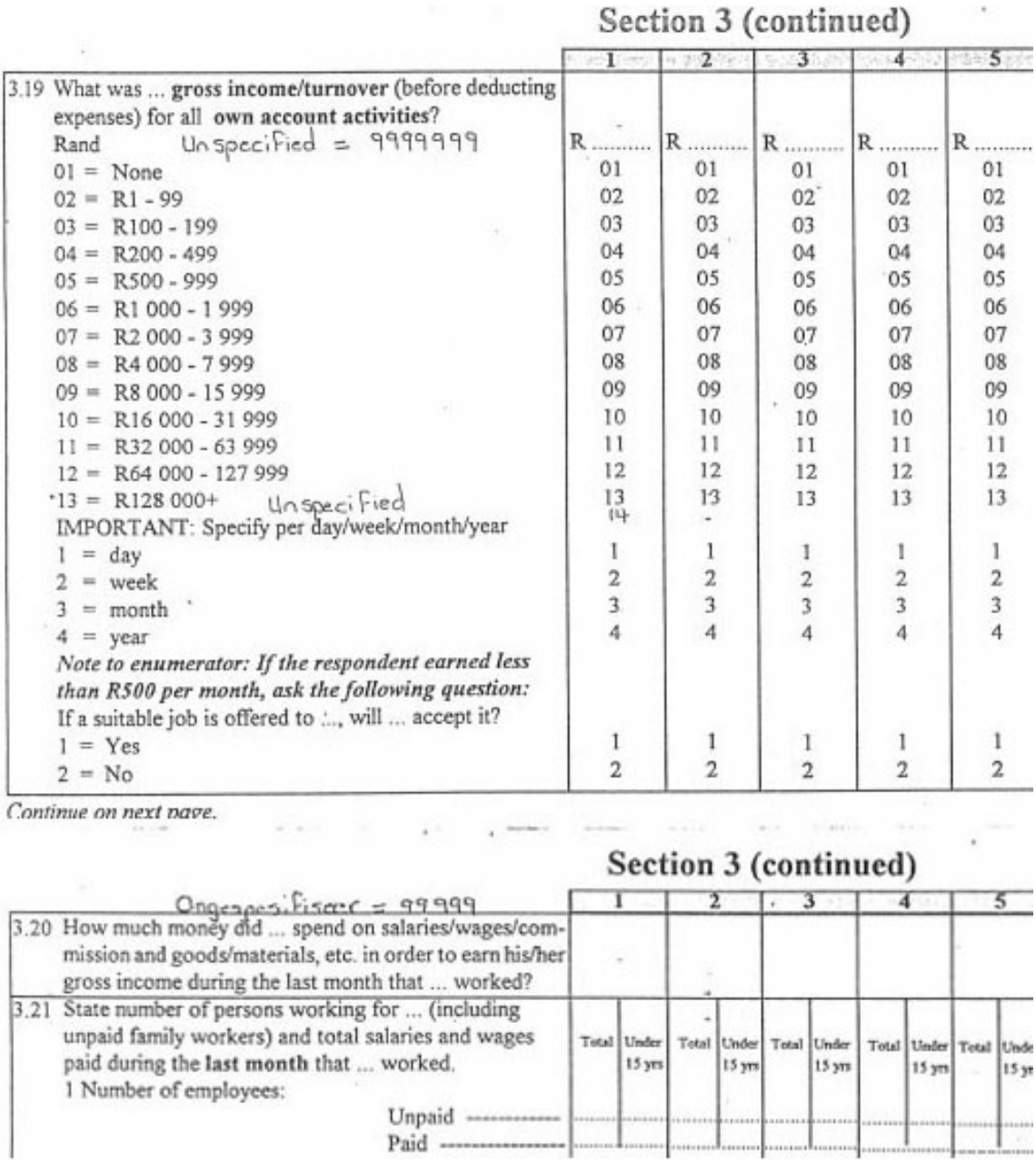
	1	2	3	4	5
3.13 **INCOME from main job:** What is/was ... total salary/pay (including overtime) at main job?					
Rand Unspecified = 9999999	R	R	R	R	R .....
01 = None	01	01	01	01	01
02 = R1 - 99	02	02	02	02	02
03 = R100 - 199	03	03	03	03	03
04 = R200 - 499	04	04	04	04	04
05 = R500 - 999	05	05	05	05	05
06 = R1 000 - 1 999	06	06	06	06	06
07 = R2 000 - 3 999	07	07	07	07	07
08 = R4 000 - 7 999	08	08	08	08	08
09 = R8 000 - 15 999	09	09	09	09	09
10 = R16 000 - 32 999	10	10	10	10	10
11 = R33 000+ Unspecified	11 12	11	11	11	11
IMPORTANT: Specify per day/week/month/year					
1 = day	1	1	1	1	1
2 = week	2	2	2	2	2
3 = month	3	3	3	3	3
4 = year	4	4	4	4	4
What is/was the estimated value ... receives/received in kind for:					
- transport	R.....	R.....	R.....	R.....	R.....
- clothing	R.....	R.....	R.....	R.....	R.....
- food	R.....	R.....	R.....	R.....	R.....
- other Unspecified = 9 999	R.....	R.....	R.....	R.....	R.....
IMPORTANT: Specify per day/week/month/year					
1 = day	1	1	1	1	1
2 = week Calculated per month	2	2	2	2	2
3 = month	3	3	3	3	3
4 = year	4	4	4	4	4

Continue on next page.

Section 3 (continued)

	1	2	3	4	5
3.14 Indicate whether ... income previously stated is/was total salary/pay (before deductions) or is/was 'take-home' pay?					
1 = Total salary/pay before deductions.	1	1	1	1	1
2 = Take-home pay	2	2	2	2	2
If "Total salary/pay", what is the amount that is/was deducted? Unspecified 99999	R	R	R	R	R .....
3.15 In addition to the main occupation, did ... also do any other work for him/herself (part-time or **informal**) for **own account** during the **past year**?					
1 = Yes Go to Question 3.16	1	1	1	1	1
2 = No Go to Question 3.22	2	2	2	2	2
3.16 Describe the work ... does/did for him/herself **(own account)** See code list for occupations and					

Figure 8: The earnings question for employees in the 1994 OHS.



Section 3 (continued)

	1	2	3	4	5
3.19 What was ... **gross income/turnover** (before deducting expenses) for all **own account activities**?					
Rand Unspecified = 9999999	R	R	R	R	R .....
01 = None	01	01	01	01	01
02 = R1 - 99	02	02	02	02	02
03 = R100 - 199	03	03	03	03	03
04 = R200 - 499	04	04	04	04	04
05 = R500 - 999	05	05	05	05	05
06 = R1 000 - 1 999	06	06	06	06	06
07 = R2 000 - 3 999	07	07	07	07	07
08 = R4 000 - 7 999	08	08	08	08	08
09 = R8 000 - 15 999	09	09	09	09	09
10 = R16 000 - 31 999	10	10	10	10	10
11 = R32 000 - 63 999	11	11	11	11	11
12 = R64 000 - 127 999	12	12	12	12	12
13 = R128 000+ Unspecified	13 14	13 -	13	13	13
IMPORTANT: Specify per day/week/month/year					
1 = day	1	1	1	1	1
2 = week	2	2	2	2	2
3 = month	3	3	3	3	3
4 = year	4	4	4	4	4
Note to enumerator: If the respondent earned less than R500 per month, ask the following question:					
If a suitable job is offered to ..., will ... accept it?					
1 = Yes	1	1	1	1	1
2 = No	2	2	2	2	2

Continue on next page.

Section 3 (continued)

Ongespesifiseer = 99999	1		2		3		4		5	
3.20 How much money did ... spend on salaries/wages/commission and goods/materials, etc. in order to earn his/her gross income during the last month that ... worked?										
3.21 State number of persons working for ... (including unpaid family workers) and total salaries and wages paid during the **last month** that ... worked.	Total	Under 15 yrs	Total	Under 15 yrs	Total	Under 15 yrs	Total	Under 15 yrs	Total	Unde 15 yr
1 Number of employees:										
Unpaid ----------										
Paid ----------										

Figure 9: The earnings question for the self-employed/employers in the 1994 OHS.

	1
3.16 **INCOME from main job:** What is/was ... total salary/pay (including overtime) at main job (before deductions)? Rand (unspecified = 9 999 999) 01 = No income 02 = R1 - R999 03 = R1 000 - R1 249 04 = R1 250 - R1 499 05 = R1 500 - R1 999 06 = R2 000 - R2 499 07 = R2 500 - R2 999 08 = R3 000 - R3 999 09 = R4 000 - R5 999 10 = R6 000 - R7 999 11 = R8 000 - R9 999 12 = R10 000 - R12 499 13 = R12 500 - R14 999 14 = R15 000 - R19 999 15 = R20 000 - R24 999 16 = R25 000 - R29 999 17 = R30 000 - R39 999 18 = R40 000 - R59 999 19 = R60 000 - R79 999 20 = R80 000 - R99 999 21 = R100 000 - R124 999 22 = R125 000 - R149 999 23 = R150 000 - R199 999 24 = R200 000 - R249 999 25 = R250 000 - R299 999 26 = R300 000 - R399 999 27 = R400 000 - R499 999 28 = R500 000 - R599 999 29 = R600 000 and over 30 = Unspecified Important, please specify per day/week/month/year. 1 = Day 2 = Week 3 = Month 4 = Year What is/was the estimated value ... receives/received in kind per week for: Transport Food Other	R ..... 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 1 2 3 4 R ..... R ..... R .....

Section 3 (continued) 0317-E

30

	1
3.17 What is/was the total amount deducted, if any? (unspecified = 99 999) 1 = Per day 2 = Per week 3 = Per month 4 = Per year	R 1 2 3 4
3.18 In addition to the main occupation, did ... also do any other work for him-/herself (part-time or informal) for own account during the past year? 1 = Yes *(Go to 3.19)* 2 = No *(Go to 3.25)*	1 2
3.19 Describe the work ... does/did for him-/herself (own account). *Note: Describe type of work in detail.*	

Figure 10: The earnings question for employees in the 1995 OHS.

Section 3 (continued)

0317-E
31

	1
3.22 What was ... **gross income/turnover** (before deducting expences) for all **own account activities**? Rand (unspecified = 9 999 999) 01 = No income 02 = R1 - R999 03 = R1 000 - R1 249 04 = R1 250 - R1 499 05 = R1 500 - R1 999 06 = R2 000 - R2 499 07 = R2 500 - R2 999 08 = R3 000 - R3 999 09 = R4 000 - R5 999 10 = R6 000 - R7 999 11 = R8 000 - R9 999 12 = R10 000 - R12 499 13 = R12 500 - R14 999 14 = R15 000 - R19 999 15 = R20 000 - R24 999 16 = R25 000 - R29 999 17 = R30 000 - R39 999 18 = R40 000 - R59 999 19 = R60 000 - R79 999 20 = R80 000 - R99 999 21 = R100 000 - R124 999 22 = R125 000 - R149 999 23 = R150 000 - R199 999 24 = R200 000 - R249 999 25 = R250 000 - R299 999 26 = R300 000 - R399 999 27 = R400 000 - R499 999 28 = R500 000 - R599 999 29 = R600 000 and over 30 = Unspecified Important, please specify per day/week/month/year. 1 = Day 2 = Week 3 = Month 4 = Year	R 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 1 2 3 4

Section 3 (continued)

0317-E
32

	1	
3.23 How much money did ... spend on salaries/wages/commission and goods/materials, etc. in order to earn his/her gross income during the last **month** that ... worked? (unspecified = 99 999)		
3.24 State number of persons working for ... (including unpaid family workers) and total salaries and wages paid during the last **month** that ... worked.	Total	Under 15 yrs
Number of employees: Unpaid --------_ (unspecified = 99) Paid ----------_ Total salaries/wages/commission paid ----------_ (unspecified = 99 999) ***Note to enumerator:*** ***If one or more of the employees are part of the household, write down their respondent number(s).***		
	R	R .....

Figure 11: The previous question for the self-employed/employers in the 1995 QHS

	1
ASK ALL EMPLOYEES WHO WERE EMPLOYED DURING THE PAST 7 DAYS INCOME FROM MAIN JOB 3.17 What is (the person's) **total** salary/pay **(including overtime and bonus)** at the MAIN job (**before any deductions**) **Note:** Make sure the respondent points at the **correct income column** (weekly, monthly, annually) on the **show card** and circle the applicable code	

Weekly	Monthly	Annually	
1 – None	1 – None	1 – None	01
2 – R1 - R46	2 – R1 - R200	2 – R1 - R2 400	02
3 – R47 - R115	3 – R201 - R500	3 – R2 401 - R6 000	03
4 – R116 - R231	4 – R501 - R1 000	4 – R6 001 - R12 000	04
5 – R232 - R346	5 – R1 001 - R1 500	5 – R12 001 - R18 000	05
6 – R347 - R577	6 – R1 501 - R2 500	6 – R18 001 - R30 000	06
7 – R578 - R808	7 – R2 501 - R3 500	7 – R30 001 - R42 000	07
8 – R809 - R1 039	8 – R3 501 - R4 500	8 – R42 001 - R54 000	08
9 – R1 040 - R1 386	9 – R4 501 - R6 000	9 – R54 001 - R72 000	09
10 – R1 387 - R1 848	10 – R6 001 - R8 000	10 – R72 001 - R96 000	10
11 – R1 849 - R2 540	11 – R8 001 - R11 000	11 – R96 001 - R132 000	11
12 – R2 541 - R3 695	12 – R11 001 - R16 000	12 – R132 001 - R192 000	12
13 – R3 696 - R6 928	13 – R16 001 - R30 000	13 – R192 001 - R360 000	13
14 – R6 929 or more	14 – R30 001 or more	14 – R360 001 or more	14

Figure 12: The earnings question for employed workers in the 1996 OHS.

	1
ASK ALL EMPLOYERS AND SELF-EMPLOYED PERSONS 3.22 For **how many months** during the **past 12 months** has/was (the person's) business/enterprise operated?	.....

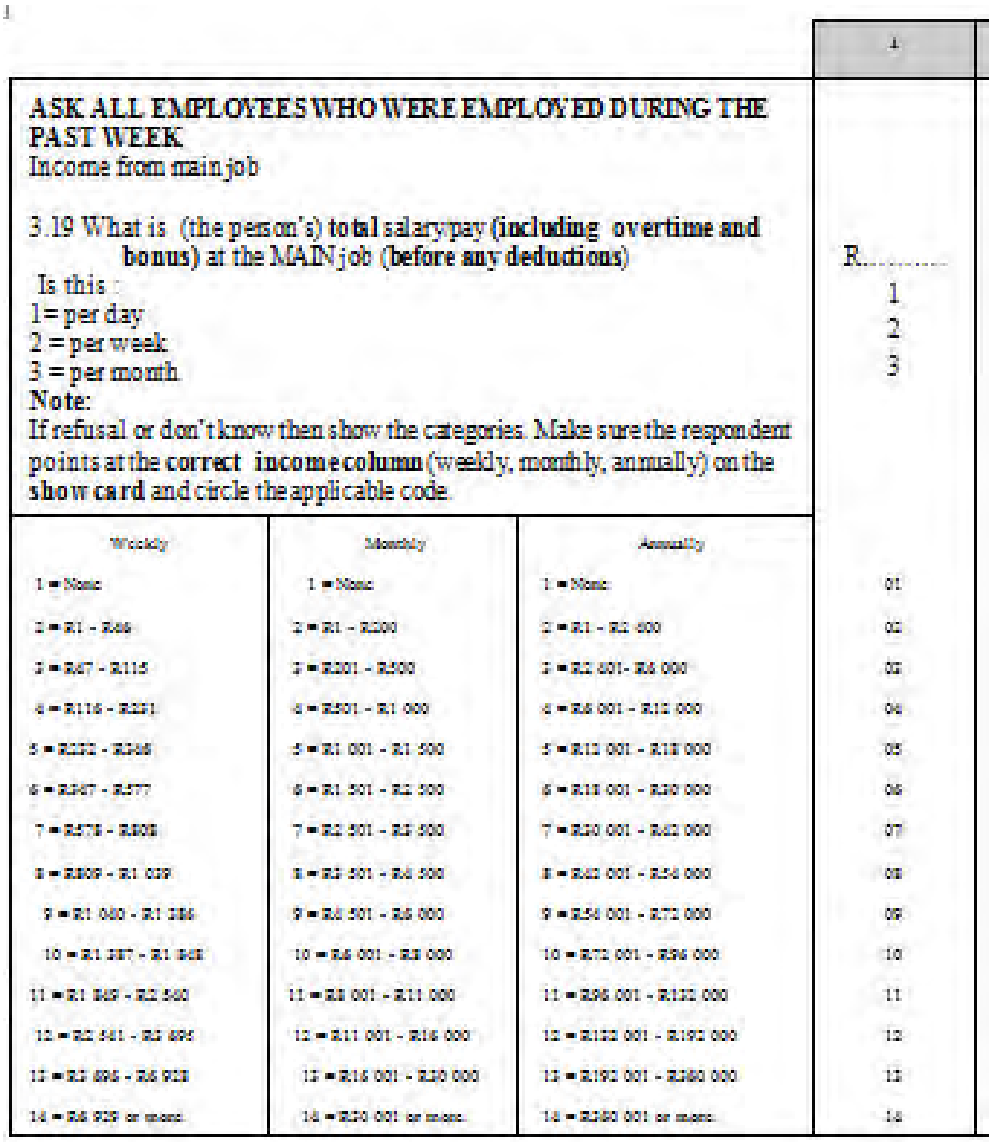
3.23 What is/was (the person's) **total** income/turnover (before deducting expenses) from his/her **own activities/ business** ?

Note:
Make sure the respondent points at the **correct income column** (weekly, monthly, annually) on the **show card** and circle the applicable code

Weekly	Monthly	Annually	
0 1 = None	01 = None	01 = None	01
02 = R1 - R46	02 = R1 - R200	02 = R1 - R2 400	02
03 = R47 - R115	03 = R201 - R500	03 = R2 401 - R6 000	03
04 = R116 - R231	04 = R501 - R1 000	04 = R6 001 - R12 000	04
05 = R232 - R346	05 = R1 001 - R1 500	05 = R12 001 - R18 000	05
06 = R347 - R577	06 = R1 501 - R2 500	06 = R18 001 - R30 000	06
07 = R578 - R808	07 = R2 501 - R3 500	07 = R30 001 - R42 000	07
08 = R809 - R1 039	08 = R3 501 - R4 500	08 = R42 001 - R54 000	08
09 = R1 040 - R1 386	09 = R4 501 - R6 000	09 = R54 001 - R72 000	09
10 = R1 387 - R1 848	10 = R6 001 - R8 000	10 = R72 001 - R96 000	10
11 = R1 849 - R2 540	11 = R8 001 - R11 000	11 = R96 001 - R132 000	11
12 = R2 541 - R3 695	12 = R11 001 - R16 000	12 = R132 001 - R192 000	12
13 = R3 696 - R6 928	13 = R16 001 - R30 000	13 = R192 001 - R360 000	13
14 = R6 929 - R10 393	14 = R30 001 - R45 000	14 = R360 001 - R540 000	14
15 = R10 394 - R13 857	15 = R45 001 - R60 000	15 = R540 001 - R720 000	15
16 = R13 858 or more	16 = R60 001 or more	16 = R720 001 or more	16

	1
3.24 How much money did (the person) spend on the following items in order to earn his/her total income during the **last month** that (the person) worked?	
Goods/materials	R.....
Salaries/Wages/Commissions	R.....
Other expenses	R.....

Figure 13: The earnings question for self-employed workers or employers in the 1996 OHS.



ASK ALL EMPLOYEES WHO WERE EMPLOYED DURING THE PAST WEEK
Income from main job

3.19 What is (the person's) **total** salary/pay **(including overtime and bonus)** at the MAIN job **(before any deductions)**
Is this :
1 = per day
2 = per week
3 = per month
Note:
If refusal or don't know then show the categories. Make sure the respondent points at the **correct income column** (weekly, monthly, annually) on the **show card** and circle the applicable code.

R.....
1
2
3

Weekly	Monthly	Annually	
1 = None	1 = None	1 = None	01
2 = R1 - R46	2 = R1 - R200	2 = R1 - R2 600	02
3 = R47 - R115	3 = R201 - R500	3 = R2 601- R6 000	03
4 = R116 - R231	4 = R501 - R1 000	4 = R6 001 - R12 000	04
5 = R232 - R346	5 = R1 001 - R1 500	5 = R12 001 - R18 000	05
6 = R347 - R577	6 = R1 501 - R2 500	6 = R18 001 - R30 000	06
7 = R578 - R808	7 = R2 501 - R3 500	7 = R30 001 - R42 000	07
8 = R809 - R1 039	8 = R3 501 - R4 500	8 = R42 001 - R54 000	08
9 = R1 040 - R1 386	9 = R4 501 - R6 000	9 = R54 001 - R72 000	09
10 = R1 387 - R1 848	10 = R6 001 - R8 000	10 = R72 001 - R96 000	10
11 = R1 849 - R2 540	11 = R8 001 - R11 000	11 = R96 001 - R132 000	11
12 = R2 541 - R3 695	12 = R11 001 - R16 000	12 = R132 001 - R192 000	12
13 = R3 696 - R6 928	13 = R16 001 - R30 000	13 = R192 001 - R360 000	13
14 = R6 929 or more	14 = R30 001 or more	14 = R360 001 or more	14

Figure 14: The earnings question for employees in the 1997 OHS

	1
ASK ALL EMPLOYERS AND SELF-EMPLOYED PERSONS 3.24 For **how many months** during the **past 12 months** has/was (the person's) business/enterprise operated?	.....
3.25 What is/was (the person's) **total** income/turnover (before deducting expenses) from his/her **own activities/ business** ? Is this: 1= Weekly 2= Monthly 3= Annual **Note**: If refusal or don't know then show the categories. Make sure the respondent points at the **correct income column** (weekly, monthly, annually) on the **show card** and circle the applicable categories.	R..... 1 2 3

Weekly	Monthly	Annually	
0 1 = None	01 = None	01 = None	01
02 = R1 - R46	02 = R1 - R200	02 = R1 - R2 400	02
03 = R47 - R115	03 = R201 - R500	03 = R2 401- R6 000	03
04 = R116 - R231	04 = R501 - R1 000	04 = R6 001 - R12 000	04
05 = R232 - R346	05 = R1 001 - R1 500	05 = R12 001 - R18 000	05
06 = R347 - R577	06 = R1 501 - R2 500	06 = R18 001 - R30 000	06
07 = R578 - R808	07 = R2 501 - R3 500	07 = R30 001 - R42 000	07
08 = R809 - R1 039	08 = R3 501 - R4 500	08 = R42 001 - R54 000	08
09 = R1 040 - R1 386	09 = R4 501 - R6 000	09 = R54 001 - R72 000	09
10 = R1 387 - R1 848	10 = R6 001 - R8 000	10 = R72 001 - R96 000	10
11 = R1 849 - R2 540	11 = R8 001 - R11 000	11 = R96 001 - R132 000	11
12 = R2 541 - R3 695	12 = R11 001 - R16 000	12 = R132 001 - R192 000	12
13 = R3 696 - R6 928	13 = R16 001 - R30 000	13 = R192 001 - R360 000	13
14 = R6 929 - R10 393	14 = R30 001 - R45 000	14 = R360 001 - R540 000	14
15 = R10 394 - R13 857	15 = R45 001 - R60 000	15 = R540 001 - R720 000	15
16 = R13 858 or more	16 = R60 001 or more	16 = R720 001 or more	16

	1
3.26 How much money did (the person) spend on the following items in order to earn his/her total income during the **last month** that (the person) worked?	
Goods/materials	R.....
Salaries/Wages/Commissions	R.....
Other expenses	R.....

Figure 15: The earnings question for the self-employed and employers in the 1997 OHS

			1
3.26 What is/was ……'s total income (before deducting expenses) from his/her own activities/ business?			
Is this:			
1 = Weekly			☐ 1
2 = Monthly			☐ 2
3 = Annually			☐ 3
*If refusal or don't know, then show the categories. Make sure the respondent points at the **correct income column** (weekly, monthly, annually) on the **show card** and mark the applicable categories.*			
WEEKLY	**MONTHLY**	**ANNUALLY**	
01 = NONE	01 = NONE	01 = NONE	☐ 01
02 = R1 - R46	02 = R1 - R200	02 = R1 - R2 400	☐ 02
03 = R47 - R115	03 = R201 - R500	03 = R2 401 - R6 000	☐ 03
04 = R116 - R231	04 = R501 - R1 000	04 = R6 001 - R12 000	☐ 04
05 = R232 - R346	05 = R1 001 - R1 500	05 = R12 001 - R18 000	☐ 05
06 = R347 = R577	06 = R1 501 = R2 500	06 = R18 001 - R30 000	☐ 06
07 = R578 - R808	07 = R2 501 - R3 500	07 = R30 001 - R42 000	☐ 07
08 = R809 - R1 039	08 = R3 501 - R4 500	08 = R42 001 - R54 000	☐ 08
09 = R1 040 - R1 386	09 = R4 501 - R6 000	09 = R54 001 - R72 000	☐ 09
10 = R1 387 - R1 848	10 = R6 001 - R8 000	10 = R72 001 - R96 000	☐ 10
11 = R1 849 - R2 540	11 = R8 001 - R11 000	11 = R96 001 - R132 000	☐ 11
12 = R2 541 - R3 695	12 = R11 001 - R16 000	12 = R132 001 - R192 000	☐ 12
13 = R3 696 - R6 928	13 = R16 001 - R30 000	13 = R192 001 - R360 000	☐ 13
14 = R6 929 OR MORE	14 = R30 001 OR MORE	14 = R360 001 OR MORE	☐ 14
15 = DON'T KNOW	15 = DON'T KNOW	15 = DON'T KNOW	☐ 15
3.27 How many people are/were working for…… (including unpaid and family workers) during the last month that he/she worked?			
Number of employees:		**Unpaid**	
		Paid	

Figure 16: The earnings question for the self-employed/employers in the 1999 OHS.

			01
4.15.a What is ……'s total salary/pay at his/her <u>main</u> job? *Including overtime, allowances and bonus, before any tax or deductions.* *Give amount in whole figures, without any text or decimals* *If refusal or don't know* → ***Go to Q 4.15.c***			
Only if amount given in 4.15.a **4.15.b Is this** 1 = Per week 2 = Per month 3 = Annually			☐ 1 ☐ 2 ☐ 3
Only if refusal or don't know in 4.15.a ***4.15.c*** *Show the categories. Make sure the respondent points at the* ***correct income column*** *(weekly, monthly, annually) on* ***Show card 3*** *and mark the applicable code.*			
Weekly	**Monthly**	**Annually**	
01 = NONE	01 = NONE	01 = NONE	☐ 01
02 = R1 - R46	02 = R1 - R200	02 = R1 - R2 400	☐ 02
03 = R47 - R115	03 = R201 – R500	03 = R2 401 - R6 000	☐ 03
04 = R116 - R231	04 = R501 – R1 000	04 = R6 001 - R12 000	☐ 04
05 = R232 - R346	05 = R1 001 - R1 500	05 = R12 001 - R18 000	☐ 05
06 = R347 = R577	06 = R1 501 = R2 500	06 = R18 001 - R30 000	☐ 06
07 = R578 - R808	07 = R2 501 - R3 500	07 = R30 001 - R42 000	☐ 07
08 = R809 - R1 039	08 = R3 501 - R4 500	08 = R42 001 - R54 000	☐ 08
09 = R1 040 - R1 386	09 = R4 501 - R6 000	09 = R54 001 - R72 000	☐ 09
10 = R1 387 - R1 848	10 = R6 001 - R8 000	10 = R72 001 - R96 000	☐ 10
11 = R1 849 - R2 540	11 = R8 001 - R11 000	11 = R96 001 - R132 000	☐ 11
12 = R2 541 - R3 695	12 = R11 001 - R16 000	12 = R132 001 - R192 000	☐ 12
13 = R3 696 - R6 928	13 = R16 001 - R30 000	13 = R192 001 - R360 000	☐ 13
14 = R6 929 OR MORE	14 = R30 001 OR MORE	14 = R360 001 OR MORE	☐ 14
15 = DON'T KNOW	15 = DON'T KNOW	15 = DON'T KNOW	☐ 15
16 = REFUSE	16 = REFUSE	16 = REFUSE	☐ 16

Figure 17: The earnings question in the LFS 2000 (September).

FOR EMPLOYEES

5.2	**In your main job, what is the easiest way for you to tell us your wages or salary before taxes or any other deductions? Would it be.....** 1 = Monthly? 2 = Weekly? 3 = Fortnightly (Every two weeks)? 4 = Daily? 5 = Hourly? 6 = Annually? 7 = REFUSED/DON'T KNOW → ***Go to Q 5.8***	1 2 3 4 5 6 7
5.3	**Do you usually receive any tips or commission?** 1 = YES 2 = NO	1 2
5.4.a	**What is your *(choose one)* annual/monthly/weekly/daily/hourly wage or salary before deductions?** ***(Include tips and commissions)*** R → ***Go to Q 5.9***	
5.4.b	***If amount not stated*** 1 = DON'T KNOW 2 = REFUSED → ***Go to Q 5.8***	1 2

Figure 18: The earnings question for employees in the 2010 Quarter 3 QLFS.

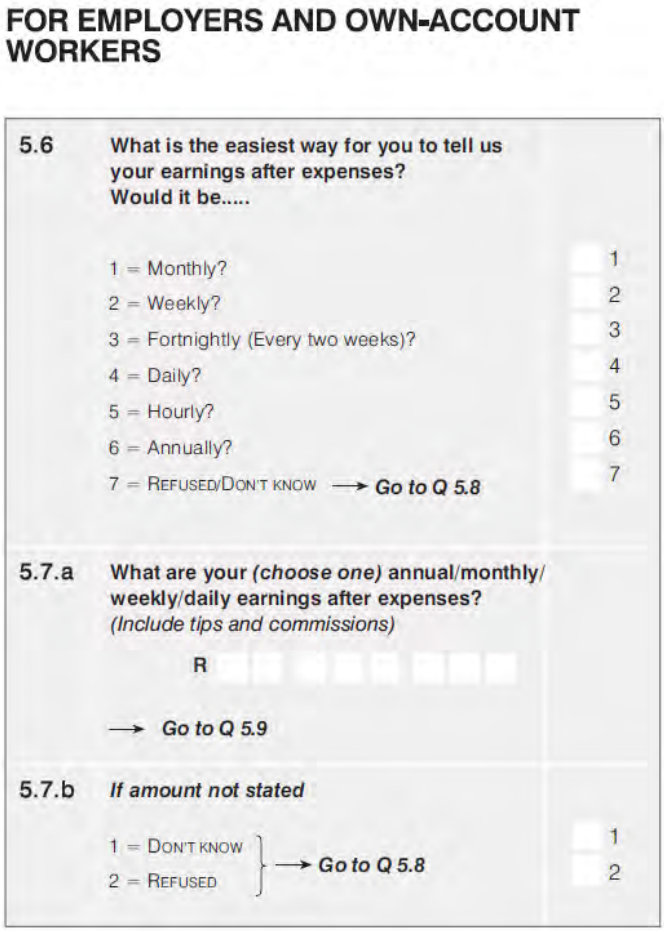


Figure 19: The earnings question for employers and the self-employed in the 2010 Quarter 3 QLFS.

For those who don't know, refused or prefer to provide ranges, please use the Prompt Card. Indicate earnings using the weekly, monthly or annual figures as indicated on the Prompt Card (do not forget to include tips and commission).

5.8	Weekly	Monthly	Annually	
01	None	None	None	01
02	R1 - R46	R1 - R200	R1 - R2 400	02
03	R47 - R115	R201 - R500	R2 401 - R6 000	03
04	R116 - R231	R501 - R1 000	R6 001 - R12 000	04
05	R232 - R346	R1 001 - R1 500	R12 001 - R18 000	05
06	R347 - R577	R1 501 - R2 500	R18 001 - R30 000	06
07	R578 - R808	R2 501 - R3 500	R30 001 - R42 000	07
08	R809 - R1 039	R3 501 - R4 500	R42 001 - R54 000	08
09	R1 040 - R1 386	R4 501 - R6 000	R54 001 - R72 000	09
10	R1 387 - R1 848	R6 001 - R8 000	R72 001 - R96 000	10
11	R1 849 - R2 540	R8 001 - R11 000	R96 001 - R132 000	11
12	R2 541 - R3 695	R11 001 - R16 000	R132 001 - R192 000	12
13	R3 696 - R6 928	R16 001 - R30 000	R192 001 - R360 000	13
14	R6 929 - R8 654	R30 001 - R37 500	R360 001 - R450 000	14
15	R8 655 - R12 500	R37 501 - R54 167	R450 001 - R650 000	15
16	R12 501 - R14 423	R54 168 - R62 500	R650 001 - R750 000	16
17	R14 424 - R16 346	R62 501 - R70 800	R750 001 - R850 000	17
18	R16 347 - R19 231	R70 801 - R83 300	R850 001 - R1 000 000	18
19	R19 232 or more	R83 301 or more	R1 000 001 or more	19
20	Don't know	Don't know	Don't know	20
21	Refuse	Refuse	Refuse	21

Figure 20: The earnings brackets in the 2010 Quarter 3 QLFS.

Table 9: Proportion giving Rand amounts by bracket and earnings period in the 1994 and 1995 OHS

	Employees				Self-employed			
Survey and bracket	1. Per day	2. Per week	3. Per month	4. Per year	1. Per day	2. Per week	3. Per month	4. Per year
OHS 1994								
1	1	1	1	1	0	0	0	0
2	0.69	0.72	0.6		0.51	0.63	0.47	0
3	0.5	0.52	0.63		0.29	0.24	0.48	
4	0.6	0.52	0.7	1	0.42	0.35	0.55	1
5	1	0.53	0.48	1	0.24	0.4	0.45	1
6	0.4	0.52	0.43	0.8	0.67	0.09	0.31	0.27
7		0.52	0.33	0.37	0	0.27	0.24	0.03
8		0.7	0.3	0.35	0	0.84	0.21	0.35
9			0.28	0.08		0.76	0.13	0.74
10			0.35	0.18	0	1	0.11	0.18
11			0.07	0.24			0.2	0.08
12						1	0.09	0.19
13							0	0.13
OHS 1995								
1		0	0	0	0	0	0	0
2	0.79	0.85	0.85	0.85	0.75	0.85	0.83	0.46
3		0.82	0.8	0.78	0	0.62	0.63	0.83
4		0.76	0.76	0.9		0.56	0.64	0.88
5		0.45	0.78	0.67	0	0.81	0.65	0.82
6		0.69	0.76	0.75		0.88	0.63	0.65
7		1	0.73	1	0	0.44	0.64	0.58
8			0.72	1		0	0.53	0.77
9		1	0.69	0.93	0	0.35	0.44	0.71
10			0.69	0.72		0.49	0.51	0.5
11			0.72	0.86		1	0.61	0.62
12			0.64	0.82		0	0.4	0.63
13			0.74	0.73		0	0.43	0.52
14			0.77	0.62			0.41	0.64
15			0.91	0.64			0.21	0.44
16			0.51	0.53			0.42	0.19
17			0	0.61			0.41	0.58
18				0.65			0.31	0.5
19			0.31	0.75			0.32	0.6
20			0	0.63			0.17	0.58
21			0	0.72			0.29	0.32
22				0.67			0.77	0.53
23				0.64			0	0.43
24				0.59			0	0.61
25				0.6			0.32	0.18
26				0.71			1	0.53
27				1				0.41
28							0	0.33
29				0.59			0	0.43

Table 10: Proportion giving Rand amounts by bracket among employees - 1996 to 2011

	Bracket																	
Survey	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
OHS 1996	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
OHS 1997	0	0.92	0.87	0.71	0.61	0.56	0.51	0.46	0.47	0.43	0.3	0.16	0.14	0.14					
OHS 1998	0	0.9	0.82	0.63	0.52	0.47	0.34	0.34	0.4	0.35	0.28	0.25	0.35	0.64					
OHS 1999	0	0.85	0.78	0.62	0.5	0.44	0.36	0.32	0.33	0.24	0.26	0.23	0.33	0.73					
LFS 00:1	0	0.95	0.93	0.84	0.73	0.62	0.54	0.53	0.51	0.45	0.47	0.41	0.37	0.81					
LFS 00:2	0	0.97	0.97	0.93	0.88	0.87	0.85	0.78	0.78	0.77	0.68	0.66	0.7	0.95					
LFS 01:1	0	0.95	0.93	0.86	0.77	0.75	0.66	0.56	0.58	0.51	0.48	0.43	0.36	0.83					
LFS 01:2	0	0.94	0.94	0.86	0.77	0.76	0.66	0.63	0.61	0.6	0.45	0.48	0.38	0.68					
LFS 02:1	0	0.93	0.93	0.87	0.75	0.73	0.61	0.52	0.53	0.52	0.4	0.41	0.4	0.63					
LFS 02:2	0	0.94	0.95	0.87	0.73	0.72	0.6	0.5	0.53	0.48	0.38	0.31	0.25	0.57					
LFS 03:1	0	0.93	0.92	0.87	0.77	0.72	0.62	0.5	0.51	0.46	0.39	0.37	0.32	0.77					
LFS 03:2	0	0.95	0.94	0.87	0.75	0.73	0.58	0.49	0.46	0.44	0.35	0.29	0.3	0.57					
LFS 04:1	0	0.93	0.94	0.9	0.75	0.73	0.59	0.47	0.45	0.45	0.33	0.28	0.28	0.5					
LFS 04:2	0	0.95	0.95	0.91	0.79	0.72	0.62	0.54	0.45	0.46	0.36	0.39	0.31	0.54					
LFS 05:1	0	0.92	0.94	0.9	0.78	0.74	0.66	0.57	0.5	0.47	0.42	0.4	0.27	0.45					
LFS 05:2	0	0.89	0.9	0.89	0.79	0.74	0.62	0.54	0.53	0.51	0.39	0.37	0.31	0.52					
LFS 06:1	0	0.94	0.94	0.9	0.77	0.77	0.69	0.57	0.52	0.53	0.46	0.41	0.42	0.52					
LFS 06:2	0	0.94	0.95	0.92	0.82	0.78	0.67	0.61	0.56	0.51	0.46	0.37	0.36	0.42					
LFS 07:1	0	0.91	0.95	0.92	0.86	0.81	0.75	0.66	0.64	0.6	0.52	0.44	0.36	0.55					
LFS 07:2	0	0.87	0.94	0.92	0.88	0.83	0.73	0.66	0.62	0.58	0.46	0.49	0.43	0.36					
QLFS 2010:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
QLFS 2010:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
QLFS 2010:	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
QLFS 2010:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
QLFS 2011:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
QLFS 2011:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
QLFS 2011:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
QLFS 2011:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 11: Proportion giving Rand amounts by bracket among self-employed - 1996 to 2011

	Bracket																	
Survey	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
OHS 1996	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
OHS 1997	0	0.75	0.69	0.62	0.55	0.58	0.4	0.37	0.55	0.43	0.21	0.31	0.25	0.24	0.15	0.44			
OHS 1998	0	0.71	0.59	0.58	0.43	0.37	0.34	0.23	0.28	0.26	0.23	0.29	0.3	0.17	0.22	0.36			
OHS 1999	0	0.73	0.58	0.47	0.34	0.31	0.25	0.22	0.29	0.14	0.19	0.09	0.15	0.19					
LFS 00:1	0	0.86	0.8	0.72	0.52	0.62	0.43	0.34	0.49	0.58	0.47	0.52	0.28	0.61					
LFS 00:2	0	0.91	0.85	0.88	0.75	0.75	0.81	0.72	0.78	0.7	0.61	0.59	0.6	0.9					
LFS 01:1	0	0.84	0.84	0.78	0.73	0.72	0.64	0.55	0.61	0.45	0.42	0.53	0.49	0.63					
LFS 01:2	0	0.86	0.75	0.79	0.52	0.64	0.65	0.43	0.6	0.42	0.42	0.35	0.27	0.59					
LFS 02:1	0	0.84	0.76	0.73	0.53	0.5	0.47	0.4	0.42	0.42	0.4	0.32	0.4	0.41					
LFS 02:2	0	0.81	0.78	0.7	0.59	0.51	0.57	0.47	0.5	0.28	0.4	0.26	0.33	0.48					
LFS 03:1	0	0.86	0.8	0.71	0.59	0.61	0.52	0.34	0.39	0.36	0.29	0.27	0.26	0.41					
LFS 03:2	0	0.8	0.76	0.66	0.43	0.5	0.52	0.35	0.44	0.3	0.29	0.23	0.3	0.36					
LFS 04:1	0	0.83	0.78	0.71	0.51	0.49	0.45	0.37	0.56	0.35	0.21	0.37	0.23	0.23					
LFS 04:2	0	0.83	0.83	0.79	0.62	0.6	0.48	0.41	0.48	0.33	0.36	0.27	0.21	0.22					
LFS 05:1	0	0.82	0.79	0.72	0.53	0.56	0.48	0.41	0.39	0.42	0.35	0.22	0.2	0.23					
LFS 05:2	0	0.87	0.79	0.77	0.65	0.7	0.59	0.56	0.56	0.46	0.46	0.52	0.34	0.18					
LFS 06:1	0	0.85	0.8	0.81	0.57	0.65	0.45	0.52	0.54	0.33	0.39	0.45	0.48	0.28					
LFS 06:2	0	0.84	0.84	0.82	0.61	0.65	0.66	0.54	0.59	0.34	0.4	0.51	0.49	0.35					
LFS 07:1	0	0.86	0.88	0.83	0.7	0.77	0.61	0.49	0.65	0.47	0.52	0.45	0.59	0.3					
LFS 07:2	0	0.82	0.88	0.87	0.79	0.79	0.72	0.56	0.53	0.39	0.35	0.44	0.4	0.33					
QLFS 2010:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1
QLFS 2010:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1
QLFS 2010:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
QLFS 2010:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
QLFS 2011:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
QLFS 2011:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
QLFS 2011:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
QLFS 2011:	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

southern africa labour and development research unit

The Southern Africa Labour and Development Research Unit (SALDRU) conducts research directed at improving the well-being of South Africa's poor. It was established in 1975. Over the next two decades the unit's research played a central role in documenting the human costs of apartheid. Key projects from this period included the Farm Labour Conference (1976), the Economics of Health Care Conference (1978), and the Second Carnegie Enquiry into Poverty and Development in South Africa (1983-86). At the urging of the African National Congress, from 1992-1994 SALDRU and the World Bank coordinated the Project for Statistics on Living Standards and Development (PSLSD). This project provide baseline data for the implementation of post-apartheid socio-economic policies through South Africa's first non-racial national sample survey.

In the post-apartheid period, SALDRU has continued to gather data and conduct research directed at informing and assessing anti-poverty policy. In line with its historical contribution, SALDRU's researchers continue to conduct research detailing changing patterns of well-being in South Africa and assessing the impact of government policy on the poor. Current research work falls into the following research themes: post-apartheid poverty; employment and migration dynamics; family support structures in an era of rapid social change; public works and public infrastructure programmes, financial strategies of the poor; common property resources and the poor. Key survey projects include the Langeberg Integrated Family Survey (1999), the Khayelitsha/Mitchell's Plain Survey (2000), the ongoing Cape Area Panel Study (2001-) and the Financial Diaries Project.

www.saldru.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 5696
Fax: +27 (0) 21 650 5797
Web: www.saldru.uct.ac.za