



University of Cape Town • Private Bag X3
Rondebosch, 7701 • South Africa
Tel: +27 21 650 5708 • Fax: +27 21 650 5709
Email: info@data1st.org
Website: <http://www.datafirst.uct.ac.za>



Science and Nonsense: Further Criticisms of Adcorp

Martin Wittenberg and Andrew Kerr,
School of Economics and DataFirst, University of Cape Town
March 2012

Science and Nonsense: Further criticisms of Adcorp

Martin Wittenberg and Andrew Kerr
School of Economics and DataFirst, University of Cape Town

March 12, 2012

Abstract

In this paper we deal with some of the criticisms levelled at us and DataFirst. We also produce some new and more detailed critiques of the Adcorp methods (as we understand them). In particular we show that the Adcorp Employment Index has to be **less** accurate than the Statistics South Africa employment series for the simple reason that Adcorp actually tries to mimic that series. This runs counter to the many grandiose claims that Adcorp makes for it. The combination of zero detail on what Adcorp does, wildly inflated claims about the power and reputability of the techniques and the neglect of statistical measures of accuracy are all hallmarks of non-science.

1 Introduction

Loane Sharp's response to our document (Sharp 2012) exemplifies many of the problems with Adcorp research that we tried to address in our initial critique (Kerr and Wittenberg 2012). His response focuses at length on the weaknesses in Stats SA's measurements without dealing properly with the very real problems that we pointed out in the Adcorp numbers. Sharp wants to polarise the options: either you believe Statistics South Africa 100% or you have to believe Adcorp. Furthermore he seems to suggest that we have a vested interest in "the unassailability of Stats SA's employment estimates" and that is why we are criticising the Adcorp Employment Index (AEI). Actually he is far off the mark: Wittenberg has pointed out many issues with Statistics South Africa data in peer reviewed research (e.g. Wittenberg and Collinson 2007, Branson and Wittenberg 2007, Wittenberg 2007). These attacks are all smoke-screens to divert attention from the gaping methodological holes in the Adcorp product. Before turning to those (in perhaps more detail than is warranted) we need to address the direct attacks on our integrity.

2 Science and Sales

Sharp accuses us of partisan research because we

“fail to disclose that [our] research outfit, DataFirst, provides training and consulting services to users of Stats SA data, which gives them [i.e. us] a material vested interest in the unassailability of Stats SA's employment estimates” (introduction to Sharp 2012)

The implication is that we sell data and/or data related services. Actually (as a lot of academic users of our web site www.datafirst.uct.ac.za will verify) we give the data away for free. Much of the training that we do – of postgraduate students and academics – in the skills of

survey analysis is done for free too. So who picks the tab up for all of that? Some of it is paid for by the University and some by big donor foundations who are interested in improving statistical skills and statistical literacy in South Africa – something that (as the AEI demonstrates) is in rather short supply in South Africa. Indeed one of our big donor funded programmes (the “Data Quality Project”) was designed specifically to deal with problems in the comparability of post-apartheid survey information, i.e. it was designed to detect and fix (where possible) flaws in Stats SA surveys. We believe that research and constructive engagement with Statistics South Africa is essential to improve our understanding of the trends in the South African economy.

Furthermore much of DataFirst’s work is actually not on Statistics South Africa surveys. Over 70% of downloads of datasets are of surveys conducted by academics that we archive: the Cape Area Panel Survey, the National Income Dynamics Survey and the Project for the Statistics on Living Standards and Development. We are interested in the Statistics South Africa surveys because they are a crucial component of the national statistical system, not because our livelihoods depend on them.

As to our personal situations: Martin Wittenberg is a full-time member of the School of Economics and his time is fully paid for by the University. He also happens to teach the most advanced courses on econometrics and microeconometrics in the School. The DataFirst position is an add-on. Andrew Kerr is a Research Officer employed on a project funded by the Vice Chancellor of the University of Cape Town aimed to try to make as much sense of the post-apartheid labour market information as we can. In short we have a **scientific** interest in Statistics South Africa data and not a **pecuniary** one.

Indeed the people who seem to have a direct pecuniary interest in the matter are Loane Sharp and Adcorp. They sell a product called the “Labour Market Navigator” to subscribers which is described on the Adcorp web site as “the definitive quantitative guide to South African labour market trends”. By calling the integrity of the Adcorp Employment Index into question are we, perhaps, threatening this activity/revenue stream?

And of course, in a Mail and Guardian “comment” in January 2011 Loane Sharp envisaged a “privatised statistics system”:

“The [Adcorp employment] index has shown itself to be reliable, frequent, punctual, apolitical and vastly superior to the stuff Stats SA puts out. Perhaps it is time to privatise South Africa’s national statistics. Without the discipline and rigour of private sector participation in the statistics-generating process, official statistics will gradually lose credibility and eventually, like Stats SA’s employment data, become irrelevant.” (Sharp 2011)

3 Rigour and rot

We believe that there is very little rigour in the research work of Adcorp. Of course we ourselves have been accused of a lack of rigour by Loane Sharp. He specifically accuses us of failing

to cite two ground-breaking research contributions by South Africa’s foremost monetary economist, Prof Brian Kantor, which studies are reference works for estimating unrecorded economic activity and its associated employment levels in South Africa, and which studies we provided to Wittenberg and Kerr and which they suppressed.

So let us examine the two works in question. Gerson and Kantor (1980) is a theoretical piece trying to determine under what circumstances someone is really unemployed or not in the labour

force. We failed to see how this article was connected to the way Adcorp measured employment, particularly since Adcorp does not interview the “unemployed” to determine whether they should more appropriately be classified as “not economically active”. In the second article, Kantor (1989) tries to estimate the value of the informal economy. At no stage does he attempt to estimate any employment numbers. Furthermore Kantor is quite explicit that the technique that he is using is not one of his own devising:

This particular study, in part, replicates the earlier papers, replacing quarterly with annual data, using the methods developed first by Feige for the analysis of unrecorded activity. (Kantor, 1989, p.36)

We cited the Feige paper and documented that this method has been explicitly criticised by the IMF, the OECD and various academic economists. Loane Sharp seems to labour under the illusion that if author A develops a technique, author B uses it and then C criticises A that this does not in any way affect the work of B. Indeed he seems to labour under the misapprehension that unless someone publishes an article directly criticising B’s work, then B’s approach must stand, even if its intellectual underpinnings have been shaken:

It is worth noting that there is nothing innovative or controversial in Adcorp’s use of this procedure. Prof Kantor’s methods and conclusions have never been faintly criticized, let alone refuted, in a recognized peer-reviewed publication and, following universal academic practice, his methodology must stand.

This, of course is just rot. The cash-demand method has been criticised (as we’ve shown). But on top of that in neither of the articles does Kantor provide a formula for converting a (flawed) estimate of the value of the informal sector into actual employment numbers. And actual employment numbers is what Adcorp sells to the public. In fact Adcorp goes far beyond that. Loane Sharp made the astonishing claim

Using well-established statistical techniques in widespread use around the world, Adcorp estimates that Stats SA has under-recorded informal sector employment by 6.19-million persons, suggesting that total employment in South Africa is 19.17-million, not 12.98-million. (Sharp 2011)

It was when we challenged Adcorp to show us this “well-established statistical technique” that Sharp pointed us to the IMF paper. We read it quite carefully and found that it actually criticised the currency demand method that Adcorp uses. So at the end of the day the only support for the Adcorp procedure are two papers, one of which is not even about measurement and the second is not about deriving employment estimates. But making extravagant claims backed up by no evidence seems to be the hallmark of Mr Sharp.

And deflecting attention from the weakness of his own work is the persistent second strand. In his rebuttal (Sharp 2012), he challenges us to publish our criticism in an “accredited, peer-reviewed economics journal” and doubted that we would be able to. We are fairly sure that he is right – but for very different reasons. No accredited, peer-reviewed economics journal would take his employment index seriously¹ so criticisms of it would be uninteresting also. So we would like to issue our own challenge to Loane Sharp in response: **Publish the underpinnings of your research in an accredited, peer-reviewed economics journal, and we will definitely get our critique published likewise.**

¹We did a “Google Scholar” search and couldn’t find any reference to the Adcorp Employment Index or Loane Sharp in anything that looked like an accredited peer-reviewed paper; actually we could find hardly any research papers referencing them at all.

We will now turn to the substantive problems that we have with Adcorp’s research and why we think Adcorp’s work does not even begin to measure up to the standards of science. In brief these criticisms are:

- The Adcorp methodology is not properly documented at all, hence it is not open to peer review let alone replication.
- Adcorp does not seem to understand the standard statistical tools or concepts (such as confidence intervals).
- Even if we take Loane Sharp’s assurances at face value, his estimates have to be worse than Statistics South Africa’s when measured by normal statistical criteria.
- While the errors for the **levels** of employment are likely to be high, they will be truly enormous for **changes** in those levels, something that Sharp does not even begin to understand.
- All of these imply that Adcorp’s estimates of formal employment are of dubious value. But Adcorp’s work on the “unrecorded sector” is much, much worse.

4 Openness and obfuscation

One of the fundamental characteristics of scientific work is that the research is properly documented and is, in principle, replicable by other researchers. In the case of survey research, this implies that the sampling methodology is adequately described, the questionnaire is available for scrutiny and (with appropriate controls for confidentiality) the data made available. A look at the DataFirst website will confirm that the surveys that are on our web portal generally conform to those criteria. Indeed Statistics South Africa’s documentation of its procedures has improved immeasurably since the 1990s. The reason why we can “drill down” and detect more problems is precisely because the data are more open and transparent.

This contrasts in a fundamental way with the Adcorp Employment Index. **None** of the key procedures are documented at all. Our understanding of the Adcorp procedures was all obtained in e-mail exchanges with Loane Sharp. We still don’t fully understand how different numbers are derived, since the key information (e.g. the regression models, estimated coefficients and regression diagnostics) is simply not supplied.

Indeed it is worse than that: some of the public pronouncements are actively designed to obscure an understanding of the Adcorp procedures. For instance in the January 2011 Adcorp Employment Index it is announced that:

From January 2011 onwards, the Adcorp Employment Index includes the unofficial sector. The unofficial sector, which is not recognized by Statistics SA, numbers 6.19 million people according to Adcorp’s estimates. It includes unrecorded and, in some cases, illegal transactions such as employment of unregistered foreigners, evasion of income, payroll and other taxes, and other economic activity in the underground economy. Adcorp will continue to report on unofficial sector employment on a monthly basis.(Adcorp Employment Index, January 2011)

One would be forgiven for assuming that at least since January 2011 the “unofficial sector” is included in the published AEI series. A look at Figure 1 however suggests differently. Unless the “unofficial sector” has remained as a constant proportion of the economy it looks as though the AEI has not been changed at all. Indeed in a televised debate (ABN 2 March 2012) Loane

Sharp suggests that the AEI is equivalent to Stats SA's "formal sector". On Adcorp's website, however, the index is described as follows:

The Adcorp Employment Index was created in 2009, is released every quarter [every month? M.W. and A.K.] and is recognised as the most accurate and holistic barometer of employment trends in South Africa (AEI 2012 "About the index")

Presumably a "holistic" look at employment should include the "unofficial sector"? But does it? We'd like to know.

Indeed the definition of the "unofficial sector" exemplifies the obfuscation that has made our understanding of the AEI so difficult. In his reply to our criticism Sharp explains how the size of this sector is determined:

Specifically, we take the total (recorded plus unrecorded) economy, estimate the labour intensity of the unrecorded economy, estimate the total employment connected with the economy, and deduct Stats SA's estimate of total employment to give us the "employment discrepancy", which numbers around 6.2 million people. (Sharp 2012, section 2)

So let us be clear about this procedure (it took us some time to get this straight): Adcorp uses their currency-demand method to estimate a total "unrecorded sector" employment figure of around 8.3 million. From this they deduct Stats SA's estimate of the informal sector to arrive at a discrepancy of 6.2 million. This is then allocated to something they label the "unofficial sector". So employment in the "unofficial sector" is any employment that is **by definition** not measured by Stats SA - these are people who lie to Stats SA enumerators when they knock on their door or who can never be found by Stats SA enumerators. It is therefore conceptually distinct from the "informal sector". Nevertheless when it suits them, the Adcorp "unofficial sector" morphs into the "informal sector". So, for instance in the September 2011 AEI we read:

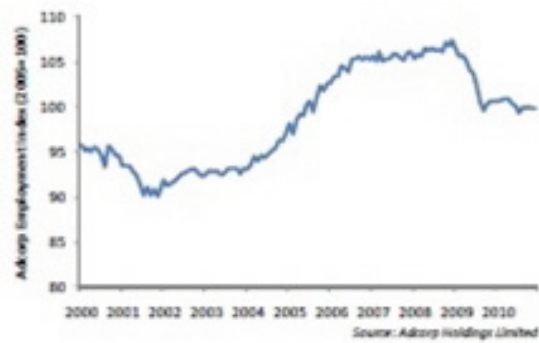
South Africa's informal sector – i.e. the unofficial part of the economy whereby many people are forced to eke out a meagre economic existence through lack of formal job opportunities [sic]. This sector of the economy which evades income taxes and circumvent labour laws, now represents 32.8% of SA'S potential workforce. During September the informal sector grew at an annual rate of 7.7% making it the fastest-growing segment of South African economic activity as it relates to individuals. More than 6.2 million people eke out a living in this sector, unprotected by labour laws and beneath the tax authorities' radar screens, making it the second-largest sector of the labour market after officially recorded employment, which numbers 12.7 million people.

...

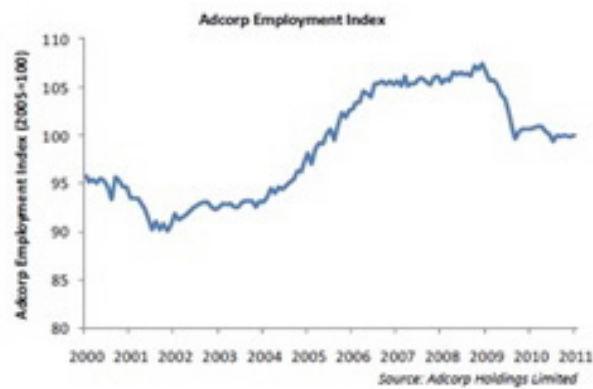
The informal sector possesses several important characteristics:

- Contracts of employment, both written and verbal, are strictly speaking absent
- Employers do not make contributions to medical aids and/or pension funds
- Employers do not make statutory deductions (i.e. payroll taxes such as Unemployment Insurance and Skills Development Levies)
- Employers do not report or pay Pay-As-You-Earn (PAYE) to the South African Revenue Services
- Employees, such as they are, do not have recourse to formal labour dispute resolution mechanisms such as the CCMA and the Labour Courts.

(Adcorp Employment Index, September 2011)



a December 2010



b January 2011

Figure 1: **Spot the difference:** Between December 2010 (top panel) and January 2011 (bottom) the Adcorp Employment estimates went up by 6 million people. Can you see it? No, we can't either.

How can we possibly be sure of all this when all that we supposedly know is that these are the type of employees that will lie about their employment information (claiming to be unemployed when they are, in fact, working) or that manage to evade enumeration?

And why would we think that the “unofficial sector” is all that there is to the informal sector? Stats South Africa already counted 2.16 million in the informal sector in September 2011 (Statistics South Africa, 2011). So the total informal sector count should surely be 8.36 million? The slippage between “unofficial sector” and “informal sector” now seems to work in reverse: those 2.16 million counted by Statistics South Africa as being in the informal sector are, of course, in the “official sector” (they were officially counted, duh). This means that they are actually in the formal sector!

A cryptic argument by Loane Sharp seems to suggest something along these lines:

Stats SA’s QLFS, based on a survey of 30 000 dwellings each quarter, has a correlation of 95.9% with Stats SA’s QES, based on a survey of 20 000 formal business enterprises each quarter showing that the QLFS and QES are measuring the same thing: the formal, established parts of the economy. (Sharp 2011)

So ALL employment measured in the QLFS is formal? Even if the respondents are pretty sure that they are in the informal sector?

Words, in the hands of Loane Sharp seem to mean pretty much what he wants them to mean. But he can’t really have it both ways: he cannot claim (as he did in his response to us, Sharp 2012) that the major disagreement between Statistics South Africa and Adcorp is about whether the informal sector has 2.1 million employees or 6.2 million employees. Either Statistics South Africa measures NO informal sector employment, i.e. the difference is 0 million versus 6.2 million; or Statistics South Africa really does measure 2.1 million in which case, presumably, the Adcorp estimate of informal sector employment should be 8.3 million. We suspect that if Adcorp seriously suggested that the informal sector was that large, their numbers would be laughed out of court immediately.

5 Confidence intervals and confidence tricks

Loane Sharp’s argument about the equality of QLFS employment and QES employment based on the high correlation between them, betrays a profound lack of understanding of statistical concepts and econometric practice. There are many variables that are highly correlated yet that measure distinct things. For instance, height (in centimetres) and weight (in kg) among respondents of the NIDS survey have a correlation coefficient of .72, but nobody would argue that height and weight are in reality the same thing. With time series data correlations have to be treated with even more suspicion. Basically **any** trending time series will be correlated with any other trending time series. In the undergraduate econometrics text books this is discussed as the case of “spurious regressions”. The key question is not how correlated different time series are, but whether they have a common underlying economic process (whether they are “cointegrated” in econometric parlance). It is, of course, plausible that formal and informal employment may share a common trend. Even then it doesn’t prove that they are really the same thing. They just happen to be different outcomes of the same underlying economic process.

Loane Sharp happens to be enamoured of correlations and correlation coefficients. His argument for the validity of the Adcorp regressions are built entirely around the high correlation coefficient between the Statistics SA employment numbers and the Adcorp data:

Adcorp and Stats SA agree about almost everything. The only disagreement concerns informal sector employment. The correlation between Adcorp’s and Stats

SA’s estimates of formal employment is 83%, not only across time, but also across sectors and occupations. (Sharp 2012)

Unless we have more information (preferably see the data) and are reassured that this is not just a case of spurious regression (many examples of which have stratospheric R^2 numbers) we remain sceptical.

While Sharp’s use of correlations is highly questionable, some of the other references to statistical concepts are plain mind-bogglingly bad:

- On the home page of the AEI the accuracy of the Statistics South Africa employment figures is attacked because of their “confidence value”:

“Due to small sample sizes, many of the survey results are unreliable. For example, estimates of total employment have a confidence value of just 34%, and four out of nine provinces have confidence values below 40%.” (Adcorp Employment Index, home page).

Confidence value has a nice sciencey² sounding ring. Except that there is no such concept in statistics. There are confidence intervals (ranges which are constructed in such a way that they should encompass the true population value at least 95% of the time). And of course the CV values that Statistics South Africa reports are coefficients of variation, as Sharp would discover if he actually read the entire documentation of the QLFS rather than simply look at the tables. Reporting CVs and confidence intervals is good statistical practice, something that Adcorp singularly fails to do.

- Adcorp suggests that they do not need to report confidence intervals, since they use “population measures rather than sample survey methodologies”. Really? In statistical parlance “population” means pretty much what the word implies – it includes everyone. The only way that Adcorp could be using a “population measure” is if they were to conduct a census of the South African population every month. At best this use of “population” refers to the Adcorp’s database of transactions. And the moment that you try to extrapolate from that to the South African population there is statistical error of at least two types: sampling error (in the sense that Adcorp’s transactions are a sample from the space of all transactions in the economy) and coverage error (in the sense that there will be some transactions that will **never** go through Adcorp). So the fact that Adcorp doesn’t report estimates of precision (indeed it would be hard to see how they might calculate them) doesn’t mean that there aren’t very real precision issues involved. But this of course is never flagged. Instead their spuriously accurate numbers are contrasted with the openly acknowledged uncertainty in the Stats SA figures.

6 Models and moonshine

In fact the Adcorp figures have to be less reliable than the Statistics South Africa figures for the simple reason that Adcorp is trying to predict those very figures! We can make this point very precisely, but unfortunately it requires a little bit of formal modelling. Let us assume that the true employment series for South Africa is y_t . We know that Statistics South Africa estimates this series, but with error. Let us call this estimated series \hat{y}_t . We have the relationship

$$\hat{y}_t = y_t + e_t \tag{1}$$

where e_t is the measurement error that Statistics South Africa has made.

²A word coined by Goldacre (2009) that ought to be in the dictionary.

Adcorp wants to predict the true employment series using regression. To this end they have a bunch of variables, say $x_{2t}, x_{3t}, \dots, x_{kt}$. We don't know what they are (since none of this is documented) but let us believe that there is a true regression relationship of the sort

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \dots + \beta_k x_{kt} + u_t \quad (2)$$

The u_t term here is the regression error term and stands for all those things that are left out of the equation. It is simply not believable that Adcorp has sufficient data to predict total employment perfectly.

Of course Adcorp doesn't know what total employment is, so Adcorp uses the Statistics South Africa estimate of employment, i.e. \hat{y}_t . Equation 2 can be rewritten (using equation 1) as

$$\hat{y}_t - e_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \dots + \beta_k x_{kt} + u_t$$

i.e.

$$\hat{y}_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \dots + \beta_k x_{kt} + u_t + e_t \quad (3)$$

Equation 3 is the one that Adcorp actually estimates (in so far as we think that they are estimating anything). Note that there are now two sources of error: the original regression error plus Statistics South Africa's measurement error. Unless they magically exactly offset each other this will **increase** the noise in the system. Statistically speaking the variance of $u_t + e_t$ will be the variance of u_t **plus** the variance of e_t , unless they are correlated.

It is worse than this, because Adcorp now uses this regression model to **predict** the figure of \hat{y}_t for the next period. There is now an additional source of noise in the system: the regression coefficients that Adcorp uses to make the prediction are not the true β figures (they are not available for inspection) but the **estimated** coefficients $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$. The formula for the standard error of this prediction is given by

$$\text{var}(u_t + e_t) \left\{ 1 + \mathbf{x}^0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}^{0'} \right\}$$

The **standard error** of this forecast will be much larger than the square root of the variance of $u_t + e_t$. Typically it will be much bigger than the standard deviation of the initial measurement error. In short, by their very procedure, they are directly importing all the measurement error that was in the original Statistics South Africa estimates and making it worse.

By pure mathematical logic the Adcorp estimates cannot be better than the Statistics South Africa estimates unless they have a magical set of data that just so happen to predict the Stats SA measurement error itself (and thus correct for it). All the talk of the supposed accuracy of the Adcorp index is pure moonshine. Publishing a set of numbers without any confidence intervals doesn't suddenly transform the accuracy of those numbers. It just hides the inaccuracy.

7 Errors of differences and indifference to errors

One of the reasons why Adcorp is so dismissive of Statistics South Africa numbers is that the changes in levels are estimated imprecisely:

by its statisticians' own estimation, Stats SA's statistical procedure is theoretically only capable of stating that employment "increased", "decreased" or "didn't change" with any usable degree of confidence. (Sharp "Questions that StatsSA must answer", letter to **Business Day**, 21/2/2012)

The statistical reason for this is straightforward. Taking the notation of the previous section, the period-on-period change in measured employment is

$$\begin{aligned}\widehat{y}_t - \widehat{y}_{t-1} &= (y_t + e_t) - (y_{t-1} + e_{t-1}) \\ &= (y_t - y_{t-1}) + (e_t - e_{t-1})\end{aligned}$$

So the measured change in employment $\widehat{y}_t - \widehat{y}_{t-1}$ now incorporates two sources of error e_t and e_{t-1} . Again the variance of this measurement error will be the sum of the two variances, unless they are correlated. What compounds this problem is that typically the levels are quite similar, so the true difference $y_t - y_{t-1}$ is small relative to the size of the error. To make this more precise, let us assume that the **level** of formal employment is estimated with a standard error of about 100 000 on an estimate of total employment of 13 million, i.e. a CV of just under 1%. If the errors between two periods were independent of each other, then the standard error for the change would be now around 140 000³. So unless the period-on-period change is greater than 280 000 (twice the standard error) it would be statistically insignificant⁴ by standard criteria. In practice the errors for changes are not quite as imprecise⁵, but the point remains: even when levels are measured pretty accurately, it is difficult to be as accurate about period-on-period changes, i.e. we can typically at best hope to give the direction of the change, and then only if the changes are large enough. To point out this uncertainty is just being honest.

Of course this same statistical relation will apply to the Adcorp data, but those numbers are not reported with equal honesty. In fact (as pointed out above) the errors in the Adcorp series have to be larger in levels and therefore also larger in first differences!

But the problem for Adcorp is likely to be even bigger than that. Remember that Adcorp gets their estimates off a regression of Stats SA employment levels on some figures emanating from the Adcorp database of transactions. Unless these two series are “cointegrated” (i.e. have a common underlying process) the differences in the trends will go into the error term and dominate the real change. There is a simple diagnostic test: regress the change in employment from the Stats SA employment series on the changes in the explanatory (“X”) variables from the Adcorp database. Only if the R^2 from that regression is reasonable would we have any confidence that changes in the Adcorp values are truly predictive of changes in Stats SA numbers. We would not be surprised if the actual R^2 is far below the 83% that Loane Sharp trumpets.

And it will be even worse when it comes to estimating changes in employment in the “unofficial sector”. We know that this is estimated by applying some labour intensity to the difference between cash and gross domestic expenditure. Let us say, for argument’s sake, that

$$\widehat{us}_t = c(m_t - gde_t)$$

where us stands for “unofficial sector”, m is the supply of cash and gde is gross domestic expenditure. c is the coefficient that translates the value of the “unrecorded economy” into people. We know that the coefficient c is essentially fixed by an ad-hoc assumption on the part of Adcorp. It therefore does not change from period to period. Consequently the change in Adcorp’s estimate of us_t is driven entirely by changes in m and changes in gde :

$$\widehat{us}_t - \widehat{us}_{t-1} = c(m_t - m_{t-1}) - c(gde_t - gde_{t-1})$$

It would be hard to believe that either m or gde is measured 100% accurately. Even if these errors in levels are small, we again get bitten by the fact that they will start to dominate the changes.

³ $var(e_t - e_{t-1}) = var(e_t) + var(e_{t-1}) = 100\,000^2 + 100\,000^2$. The standard error is the square root of that which is $100\,000 * \sqrt{2}$.

⁴which simply means that we can’t be sure that the true change is different from zero.

⁵this is because of the rotating panel nature of the data, so that the errors are correlated which will reduce their variance.

And of course there are now four sources of error: the errors in m_t , m_{t-1} , gde_t and gde_{t-1} . It would require heroic levels of faith to believe that the Adcorp estimates of the “changes in informal sector employment” are anything but pure noise. When viewed from this perspective, the following statement is almost comic in the way it takes its calculations seriously

During September the informal sector grew at an annual rate of 7.7% making it the fastest-growing segment of South African economic activity as it relates to individuals. (AEI, September 2011)

8 Gold standard and garbage

Loane Sharp took exception to our characterisation of survey evidence as the “gold standard” for measuring employment. He points to the fact that survey measurements are subject to errors of various kinds: sampling errors and non-sampling ones (such as lying to the enumerator or coverage errors). Those points are all valid. By “gold standard” we did not mean to imply that surveys will invariably lead to “precise, accurate, reliable and unassailable results”. There are badly run surveys and much better ones. And even the best of surveys still suffer from errors. There is in the entire history of statistics no error-free survey. By “gold standard” we mean that there are no other methods (that we know of) that will lead to more accurate results for this particular type of measurement.

This incidentally does not mean that other methods cannot be used to “sense-check” the results. We ourselves have used other data (such as anthropometric measurements) to throw a different light on the underlying measurement issues. Such sense-checks can help to interrogate and fine-tune the more direct measures. But they are no substitute for good, direct measures.

But the enterprise of Adcorp is not to fine-tune and improve direct measures. Statements such as the one by Sharp quoted earlier: “the index has shown itself to be reliable, frequent, punctual, apolitical and vastly superior to the stuff Stats SA puts out” is not about how to improve the direct measures – it is to substitute those with some extremely crude proxies. And what makes that enterprise non-science is that all the standards of typical scientific work go out of the window: the procedures are not properly documented, no attempt is made to interrogate the errors in those proxies and the accuracy of those instruments is so limited that they are unlikely to provide accurate data at the frequency that Adcorp wants to provide it.

And of course, as we pointed out in our original critique, there is little support for the idea that the difference between cash and gross domestic expenditure can be converted into people, or that we would know that these people are all in the informal sector and haven’t been counted somewhere else before⁶.

At the end of the day Adcorp produces a bunch of assertions with no science to back it up. They do, however, sprinkle their discourse liberally with sciencey expressions, e.g. that they used “Simultaneous equations modelling (used to disentangle the demand for and supply of cash and determine the excess demand for cash in the economy)”. These may sound impressive, but with no detailed information are pure verbiage. Indeed given the weakness of the statistical work referred to earlier we have to ask whether these models are estimated in a way that would survive scrutiny. This mixture of zero detail, grandiose claims, no attention to the standard errors of the estimates are all the hallmarks of “bad science” (Goldacre 2009). Sharp takes umbrage at our characterisation of this enterprise as “humbug”. He claims that this is an unwarranted *ad hominem* attack. To us it seems a perfectly fair way to assess the quality of the work that we can see.

⁶A lot of “illegal” activities are carried out behind the front of legal ones. Such double-shifts do not increase the employment level. And in what sense would we want to think of the burglar as being employed?

9 References

ABN (2 March 2012) "Adcorp Employment Index under Academic Criticism", podcast available at <http://www.abndigital.com/page/multimedia/video/power-lunch/1191620-Adcorp-Employment-Index-under-Academic-Criticism>

Adcorp Employment Index (December 2010), "Adcorp Employment Index, December 2010. Release date: Tuesday, 10 January 2011", downloaded from Adcorp website <http://www.adcorp.co.za/Industry/Pages/Adcorp'sEmploymentIndex.aspx> on 8 March 2012.

Adcorp Employment Index (January 2011), "Adcorp Employment Index, January 2011. Release date: Thursday, 10 February 2011", downloaded from Adcorp website <http://www.adcorp.co.za/Industry/Pages/Adcorp'sEmploymentIndex.aspx> on 8 March 2012.

Adcorp Employment Index (September 2011), "Adcorp Employment Index, September 2011. Release date: Monday, 17 October 2011", downloaded from Adcorp website <http://www.adcorp.co.za/Industry/Pages/Adcorp'sEmploymentIndex.aspx> on 8 March 2012.

Adcorp Employment Index (2012) "About the Index", available at <http://www.adcorp.co.za/Industry/Pages/AbouttheIndex.aspx>, downloaded on 8 March 2012.

Branson, Nicola and Martin Wittenberg (2007) "The measurement of employment status using cohort analysis, 1994-2004", **South African Journal of Economics**, 75(2):313-326

Gerson, Jos and Brian Kantor (1980) "An Analysis of Black Unemployment", **Studies in Economics and Econometrics**, pp.81-93, available at <http://www.zaeconomist.com/research/1980.pdf>

Goldacre, Ben (2009) **Bad Science**, London: Fourth Estate.

Kantor, Brian (1989) "Estimating the value of unrecorded economic activity in South Africa", **Journal for Studies in Economics and Econometrics**, 13(1):33-41.

Kerr, Andrew and Martin Wittenberg (2012) "Criticisms of the Adcorp employment index", mimeo, available at <http://www.datafirst.uct.ac.za/home/index.php?/Download-document/16-Criticisms-of-the-Adcorp-Employment-Index>

Sharp, Loane (2011) "Trickery in employment figures", **Mail and Guardian**, 28 January 2011, available at <http://mg.co.za/article/2011-01-28-trickery-in-employment-figures>

Sharp, Loane (2012) "Adcorp stands by its employment estimates", mimeo available at <http://www.politicsweb.co.za/politicsweb/view/politicsweb/en/page71619?oid=284298&sn=Detail&pid=71616>

Statistics South Africa (2011), "Statistical Release P0211: Quarterly Labour Force Survey, Quarter 3, 2011". downloaded from <http://www.statssa.gov.za/publications/P0211/P02113rdQuarter2011.pdf>.

Wittenberg, Martin (2007) "Dissecting post-apartheid labour market developments: Decomposing a discrete choice model while dealing with unobservables", **ERSA working paper 46**, available at http://www.econrsa.org/papers/w_papers/wp46.pdf

Wittenberg, Martin and Mark Collinson (2007) "Household Transitions in Rural South Africa, 1996-2003", **Scandinavian Journal of Public Health**, 35 (suppl69): 130-137