UNIVERSITY OF CAPE TOWN

Official Revisions to SA National Accounts Data: Magnitudes and Implications

Corne van Walbeek[1]

Policy Paper Number 10

[1] School of Economics, University of Cape Town

Official Revisions to SA National Accounts Data: Magnitudes and Implications

Corne van Walbeek*

December 2006

Abstract

This paper investigates the bias and dispersion in official revisions of eight national accounting growth rates. The growth in GDP, consumption expenditure and personal disposable income by households has been subject to significant upward revisions and bias, especially after 1994. No significant bias was found in the revisions to the other national accounting aggregates. The official revisions are subject to a high degree of dispersion. Based on the 1984-2003 period, there is a 30 per cent probability that the "final" growth rate in GDE deviates by more than 5 percentage points from the first release growth rate. For most magnitudes, other than exports and imports, the dispersion in South Africa's official revisions is similar to that of a sample of OECD countries. Using two examples, it is shown that the vintage of the data has a profound impact on the magnitude and significance of regression results based on such data.

JEL Classification: E01

Keywords: National accounts, official revisions, bias, dispersion

1 Introduction

The national accounts are an important source of information regarding the health of the economy. They have substantial relevance to policy makers in the public sector and decision makers in the private sector. For example, within the current inflation targeting framework the South African Reserve Bank (SARB) closely monitors the national accounts, and gross domestic product (GDP) and gross domestic expenditure (GDE) in particular, as important inputs in establishing whether there is inflationary pressure in the economy (SARB, 2006). The release of national accounting data is often followed by some discussion in the financial press about the likely changes in interest rates. Also, the stance of fiscal policy is influenced by the perception of the state of the economy, as are investment decisions by businesses.

The statistical authorities are under pressure to provide policy makers and the public with accurate and timely estimates of the national accounts. As may be expected, there is a trade-off between accuracy and timeliness. In South Africa, the first GDP estimates for a particular quarter are calculated by Statistics South Africa according to the production approach, and are released about 50 to 60 days after the end of the relevant quarter. The expenditure components of GDP are independently calculated by the national accounts division of the research department of the

*School of Economics, University of Cape Town

SARB, based on data obtained from Statistics South Africa, the South African Revenue Services, government departments, and own surveys of economic activity. The expenditure components are published in the SARB Quarterly Bulletin a few weeks after the initial GDP release.[1] From a policy and publicity perspective, it is these first releases that attract the most attention.

It is well known that the statistical authorities have incomplete data when they release the first estimates of the national accounts, and that the national accounts are revised as more data become available. Young (1993: 29) and Palis et al. (2004) indicate that revisions can be ascribed to the following factors: (1) replacement of preliminary source data with revised or more comprehensive data; (2) replacement of judgmental projections with source data; (3) changes in the definitions or estimating procedures; (4) updates in and changes of the base year; and (5) changes in the seasonal adjustment factors.

These revisions could have substantial implications for econometricians. For example, most economic analyses that have a long-term focus are performed on data that have gone through a number of revisions. Also, forecasts made with the help of econometric models are typically based on revised data but are evaluated against the first data release (Laubscher, 1992 and Van Walbeek and Van Graan, 2006). It is conceivable that the first releases differ significantly and systematically from revised releases of the data. Should this be the case, a "poor forecast" could in fact itself be a poor first release. Once the official data are revised, the forecasts might be much better, although it would not attract much publicity.[2]

The paper focuses on the magnitudes of official revisions to eight important national accounting aggregates at constant prices: (1) GDP, (2) GDE, (3) final consumption expenditure by households, (4) final consumption expenditure by general government, (5) gross fixed capital formation, (6) exports of goods and services, (7) imports of goods and services, and (8) disposable income of households.[3] The paper has two primary aims. Firstly, it aims to quantify the magnitudes of the revisions and to establish whether the revisions are systematically biased. Secondly, it aims to provide a provisional answer to the following question: how significant are these revisions to econometric analyses?

The paper is structured as follows: section 2 discusses the data; section 3 presents the bias and dispersion measures of the data revisions, together with some international comparisons. Section 4 illustrates the impact of data revisions on the estimated coefficients of two macro-economic relationships, while section 5 concludes the paper.

[1] GDP, calculated according to the production approach, should equal the expenditure on GDP. This seldom happens in practice. On the assumption that production data are more accurate and reliable than the expenditure data, the SARB includes a residual, as a balancing item, in the GDE and the expenditure on GDP. This approach assumes that the GDE components are subject to measurement error. If one believes that the expenditure components are measured accurately, one can add the expenditure components to obtain GDE, with no residual included. To obtain GDP from this GDE value, one would add the trade balance and the balancing residual (i.e. GDP = GDE + X – M + residual). The difference between these two approaches is that the first approach classifies the residual as an "expenditure component", whereas in the second approach it is classified as a "trade component" (i.e. part of X – M). The fact of the matter is that the SARB, and most statistical authorities around the world, use the first approach. According to an official of the SARB, the SARB aims to keep the residual smaller than one per cent of the GDP (Tendani Mantshimuli, personal communication, 2006).

[2] While this is a testable hypothesis, this issue is not addressed in this paper.

[3] The SARB codes these series as KBP6006S, KBP6012S, KBP6007S, KBP6008S, KBP6009S, KBP6013S, KBP6014S and KBP6246S, respectively.

2 The Data

Published copies of the SARB's Quarterly Bulletin are the primary source of the data used in this paper The SARB publishes the real growth rates of selected national accounting magnitudes in the "key information" section of the Quarterly Bulletin.[4] The growth rate is specified as the annualised quarter-on-quarter percentage change in the seasonally adjusted variable at constant prices. Since 1992 the SARB publishes the growth rates for the latest twenty quarters. Between 1983 and 1992 the SARB published the growth rates of the previous six to eight quarters. No growth rates in the national accounting aggregates were published prior to 1983, although in principle it would be possible to calculate these from the data in levels form (which are published in the national accounts section of the SARB's Quarterly Bulletin).

The second source of data is the SARB's website (www.reservebank.co.za). Time series data for the relevant variables, using the same definitions as those obtained from the Quarterly Bulletins (i.e. seasonally adjusted annualised quarter-on-quarter growth rates) were extracted in January 2006. This source is typically used in time series econometric analyses, and includes the latest figures for all the data.[5]

The analysis covers the period 1984Q1 to 2003Q4, i.e. there are 80 quarterly observations. The first, second, fourth and eighth data release are investigated in this paper. As an illustration of the data, different vintages of GDE growth rates for the period 1992Q3 to 1994Q4 are shown in Table 1. The growth rates under discussion are in bold font. In the March 1995 Quarterly Bulletin, the 1994Q4 growth rate (7.1 per cent) was the most recent data and is called the first release. The 1994Q3 growth rate (9.7 per cent) is called the second release, which has been revised from its first release growth rate of 9.2 per cent. The 1994Q1 growth rate (2.1 per cent) is called the fourth release, having been revised from a first release value of 2.5 per cent, to 3.3 per cent in the second release and 0.9 per cent in the third release. The 1993Q1 growth rate of 4.4 per cent is the called the eighth release. The 1993Q1 growth rate was first published as 8.2 per cent, after which it was revised, first upwards and then downwards, in subsequent releases. Revisions in the data were not investigated for more distant releases because the SARB did not publish such data before 1992. In this paper data revisions are consistently defined relative to the first release.

Table 4 also illustrates the effect of a change of base year. In the June 1994 Quarterly Bulletin the base year changed from 1985 to 1990. When the base year changes the statistical authorities change the national accounts retrospectively for at least 20 quarters, to reflect the updated price levels. As a result of the change in the base year, the growth rates change, but this is not necessarily ascribed to inaccurate initial estimates (Johan van den Heever, personal communication, 2006). This issue will be discussed later.

The data were arbitrarily divided into four equal periods of 20 quarters each in order to determine whether the distribution of the revisions have changed over time. The decision was primarily driven by data availability and the desire to have equally sized sub-periods. However, as it happens these four sub-periods correspond with clearly defined phases in the South African economy: the 1984-1988 period was marked by extreme boom and bust volatility, the 1989-1993 period straddled the longest recession in South Africa since the Great Depression,[6] the 1994-1998 period marked a period

[4] For the March 2006 Quarterly Bulletin, this information is published on page S-154.

[5] The SARB data is closely ($r > 0.997$), but not perfectly correlated with the time series data obtained from Global Insight and accessed through Time Series Explorer. While the earlier data is perfectly correlated, (small) differences are typically found for the last two to three years.

[6] According to the SARB Quarterly Bulletin (2006), the downward phase started in March 1989 and lasted till May 1993.

of stabilisation and normalisation, while the 1999-2003 period was characterised by consistent, even accelerating, growth.

3 Bias and Dispersion

3.1 Measures of bias and dispersion

In order to allow for international comparability, this paper evaluates the revisions using measures that are standard in the literature (see Young, 1993, York and Atkinson, 1997 and Palis et al., 2004). If F represents the first release, L represents the latest release, n are the number of observations, and $t = 1, 2, \ldots, n$, then bias is defined as[7]

$$\Sigma \left(L_t - F_t \right) / n \tag{1}$$

A second indicator of bias used in the literature is to establish how many (or what percentage of) revisions are positive and negative respectively. The advantage of this approach is that it is not influenced by one or two very large revisions as is the case with (1).

Dispersion can be defined in absolute terms (e.g. Young, 1993 and York and Atkinson, 1997) or in terms of the standard deviation (e.g. Mankiw and Shapiro, 1986). According to the first approach dispersion is defined as the average of the absolute values of the revisions:

$$\Sigma \left| L_t - F_t \right| / n \tag{2}$$

In order to determine the magnitude of the average deviations relative to the average of the absolute value of the growth rates themselves, Young (1993) defines relative dispersion asIn order to determine the magnitude of the average deviations relative to the average of the absolute value of the growth rates themselves, Young (1993) defines relative dispersion as

$$\left(\Sigma \left| L_t - F_t \right| / n \right) / \left(\Sigma \left| L_t \right| / n \right) \tag{3}$$

The standard deviation is defined as

$$\left[\Sigma \left(d_t - d^* \right)^2 / \left(n - 1 \right) \right]^{1/2} \tag{4}$$

where $d_t = L_t - F_t$ and $d^*$is the sample mean of d_t.

If the deviations are of similar magnitude (i.e. there are no outliers) the absolute measure of dispersion (equation (2)) yields a similar value as the standard deviation (equation (4)). If individual deviations lie disproportionately far from the bulk of the deviations, equation (4) tends to give substantially higher values than equation (2). In discussions about dispersion, this paper uses the absolute average dispersion (equation (2)) as the primary indicator, although other measures are presented as well.

[7] Other than this measure of (absolute) bias, Young (1993:33) defines *relative* bias as $[\Sigma \left(L_t - F_t \right) / n] \; / [\Sigma L_t / n]$, where F and L are defined as above. The problem with relative bias is that it yields very high and even undefined values if the average of the latest available growth rates (i.e. $\Sigma L_t / n$) is close to zero. During the period 1984-1993 this was often the case. Since this measure yields results that are difficult to interpret, it is not used in this paper.

3.2 An example: gross domestic expenditure

The statistics for the various bias and dispersion measures are shown for all eight variables in the Appendix (Tables A.1 through A.8). In order to gain a visual understanding of what such statistics mean, we focus on GDE in the analysis below. GDE was chosen as, relative to the other seven national accounting magnitudes, it experiences an "average" degree of dispersion, and, together with GDP, it is considered a crucial indicator of activity in the economy.

In Fig. 1 the first release estimates and the SARB database growth rates for GDE are shown (the second, fourth and eighth releases are not included for the sake of greater clarity). At first sight the growth rates of the two data vintages look highly correlated. Simple correlation coefficients confirm this: 0.973 between the first and second releases, 0.952 between the first and fourth releases, 0.914 between the first and eighth releases and 0.748 between the first release and the SARB database.

In Fig. 2 the magnitudes of the revisions between first and second releases, as well as between the first release and the SARB database are shown. The absolute average dispersion, defined according to equation (2), between the first and second release is 1.01 percentage points, while the standard deviation is 1.93 percentage points (see Table A.2 in the Appendix). The large difference between the absolute average dispersion and the standard deviation is explained primarily by large deviations between the first and second releases in 1985Q2 (5.6 percentage points), 1986Q1 (-10.7 percentage points) and 1986Q3 (-5.8 percentage points). As more time elapses after the first release, the dispersion of the revisions increases as well. The absolute average dispersion between the first and fourth releases is 1.79 percentage points, between the first and eighth releases 2.88 percentage points and between the first release and the SARB database 4.12 percentage points. The tables in the Appendix show that, nearly without exception, the degree of dispersion increases as more revisions are made, not only for GDE, but for all national accounting aggregates.

Other than the use of better and more accurate source data, the data are revised when there is a change in the base year. In accordance with the 1993 System of National Accounts guidelines, the SARB changes the base year for constant price estimates more or less every five years (Prinsloo, 1999 and Mantshimuli, 2004). For the period under consideration there were five base year changes: in 1985Q3 (from 1975 to 1980), 1988Q1 (1980 to 1985), 1994Q1 (1985 to 1990), 1999Q1 (1990 to 1995) and 2004Q3 (1995 to 2000). Often a rebasing exercise allows the statistical authorities to incorporate new data and data sources and to re-classify certain transactions. Because of the different weights, the growth rates of the national accounts will differ and thus the statistical authorities retrospectively change the constant prices national accounts. Other than these regular and predictable changes, some one-off political events can also have an impact on the national accounts. For example, until 1990 Namibia was included in South Africa's national accounts. With Namibia's independence in 1990 its contribution to the national accounts was retrospectively excluded for the period prior to 1990 (Prinsloo and Van Tongeren, 2002). On the other hand, the "independent states" of Transkei, Bophuthatswana, Venda and Ciskei were never excluded from South Africa's national accounts and so no adjustments were necessary when they were re-incorporated into South Africa in 1994.

For short revision periods (i.e. between the first and second and between the first and fourth releases) the revisions are driven primarily by improved source data, rather than changes in the base year. However, for longer revision periods (i.e. between the first and the eighth releases and between the first release and the SARB database) base year changes are increasingly responsible the dispersion in the revisions. In fact, for GDE, less than 10 per cent of the dispersion of the difference between the first and second releases can be ascribed to changes in the base year. This percentage increases to more than 30 per cent for the dispersion of the differences between the first

and eighth releases and to more than 60 per cent for the dispersion of the differences between the first release and the SARB database.[8] A similar result was found for the other variables.

The change of base years complicates matters for time series researchers. On the one hand, a more recent base year provides a better picture of the performance of the economy, based on its current structure, since a recent base year weighs sectors according to recent prices. On the other hand, the use of a recent base year distorts the reality of the past if there have been sizeable changes in relative prices over time.[9] Econometricians typically use historical data to describe the present or to predict the future. To the extent that the historical data becomes "distorted" by changes in the base year (or for other reasons), such data are likely to have a distorting impact on econometric and other statistical analyses.

The upshot of this discussion is that some data revisions are inevitable as a result of base year changes, while others are the result of inaccurate initial measurements. The net effect of both is to increase the dispersion of the revisions as the revision horizon becomes longer. Whether these changes introduce distortion or removes distortion from the data is a matter for debate which will not be addressed here. Thus, for the remainder of the paper, the effect of base year changes on the revisions is acknowledged, but it remains part of the analysis.

To gain a better understanding of the magnitudes of the revisions, two histograms are presented in Figs. 3 and 4. Since the distributions of the revisions are centred on zero, one can conclude that there is no obvious upward or downward bias in the revisions. A more significant problem seems to be the degree of dispersion in the revisions. In Fig. 3, which indicates the magnitude of the revisions between the first and second releases, 10 of the 80 revisions are greater than 2 percentage points in absolute value. Thus, based on this sample of revisions, there is a 12.5 percent probability that the first GDE figure is revised by 2 percentage points or more in the next quarterly release. In Fig. 4 the dispersion of the revisions between the first release and the SARB database are shown. Based on this sample, there is a 57.5 per cent (46/80) probability that the GDE growth rate published in the SARB database differs by 2 percentage points or more from the first release. More dramatically, there is a 32.5 per cent (26/80) probability that the GDE growth rate published in the SARB database differs by 5 percentage points or more from the first release.[10]

The magnitudes of these differences are large. According to the March 2006 edition of the SARB Quarterly Bulletin, South Africa's GDE grew by 3.9 per cent in 2005Q4. If the quality of

[8] These percentages were calculated as follows: absolute average dispersion measures (see equation (2)) were calculated at the change of a base year. The assumption is that the revision is attributed solely to changes in the base year and that changes in the source data have had no impact on the revision. The measure of dispersion attributable to changes in the base year is then expressed as a percentage of the measure of dispersion for the relevant period.

[9] As a somewhat extreme example to illustrate this point, consider the price of gold. Between 1980 and 1985 the average price of gold decreased from US$ 613 to US$ 317. When the base year was changed from 1980 to 1985 the contribution of gold to South Africa's GDP suddenly became much less important. While the change of base year was necessary because gold had become less important during that five year period, it also diminished the profound impact that gold played in 1980, when its contribution to GDP was very high. As a somewhat extreme example to illustrate this point, consider the price of gold. Between 1980 and 1985 the average price of gold decreased from US$ 613 to US$ 317. When the base year was changed from 1980 to 1985 the contribution of gold to South Africa's GDP suddenly became much less important. While the change of base year was necessary because gold had become less important during that five year period, it also diminished the profound impact that gold played in 1980, when its contribution to GDP was very high.

[10] The high probability of obtaining a deviation greater than 5 percentage points is influenced to a large extent by the extreme volatility in the revisions in the 1984-1993 period and especially in the 1984-1988 sub-period. However, even for the 1994-2003 period, there is a 15 per cent (6/40) probability that the GDE growth rate published in the SARB database differs by 5 percentage points or more from the first release.

the revisions remains the same as in the 1984-2003 period, this means that there is a 30 per cent probability that the "final" GDE growth rate for 2005Q4, as published in the SARB database in some years' time, will be lower than -1.1 per cent or higher than 8.9 per cent.

The previous analysis includes an extremely volatile period in South Africa's economic history, where both the growth rates and the revisions changed dramatically. Is it possible that the dispersion in the revisions is currently much smaller than in the mid-1980s? There is some, but not overwhelming, support for this argument. Considering Table A.2 in the Appendix, the average absolute dispersion of the revisions between the first release and the SARB database decreased from 6.59 percentage points in the 1984-1988 period to 3.84 percentage points in the 1989-1993 period, 3.56 percentage points in the 1994-1998 period and 2.47 percentage points in the 1999-2003 period. It is clear that the 1984-1988 period was subject to particularly large revisions, compared to the subsequent period (see Fig. 2 for a graphical presentation). However, one cannot conclude from this that the dispersion between the first release and the SARB database for the 1999-2003 period is significantly less than in the previous two sub-periods. The reason is that the SARB database growth rates for the 1999-2003 period have not settled, will probably be affected by a base year change, and as such are likely to change in the future.

The revisions between the first and eighth releases are fixed in time, and, although the data may be revised further, this has no influence on the analysis that compares the first release with the eighth release. As such, one can compare the degrees of dispersion between different periods. It turns out that the average absolute deviation between the first and eighth release is 4.63 percentage points for the 1984-1988 period and 2.02 percentage points, 2.54 percentage points and 2.42 percentage points for the three subsequent periods. These results suggest that the dispersion of the revisions has not changed significantly since 1989.

There is no evidence for a systematic upward or downward bias in the GDE revisions. This is confirmed using a standard t-test on the mean deviation, and is true for all sub-periods, as well as the whole period of investigation (see Table A.2). The conclusion is also confirmed by the fact that of the 80 revisions (first release to SARB database), 37 were decreases, while 43 were increases (see Table A.2).

3.3 Bias and dispersion for other national accounting magnitudes

Some summary indicators of bias and dispersion are presented for all eight national accounting aggregates in Table 2. More detailed statistics are presented in the Appendix.

Significant upward bias in the revisions was found for GDP, final consumption expenditure and disposable income by households, especially after 1994. As an example, the average GDP growth rate published in the SARB database for the 1994-1998 and 1999-2003 periods are respectively 1.01 and 1.27 percentage points higher than the growth rates published in the first release. Also, 15 of the 20 revisions between 1994 and 1998 were positive (five were negative), while 18 of the 20 revisions between 1999 and 2003 were positive (two were negative).

In the 2006 budget speech, the Minister of Finance stated that "when the full accounting is done towards the end of this year, we may indeed find that our economy grew 5.5% or 6% last year (2005)", even though the initial GDP growth estimate for 2005 was only 4.9% (Manuel, 2006). Considering the history of GDP revisions the past ten years, this looks like a distinct possibility.

For consumption expenditure by general government, the revisions between the first release and the SARB database had a significantly negative bias for the 1994-1998 period, but a significantly positive bias for the 1999-2003 period.

For the national accounting aggregates, other than those discussed above, there is no significant upward or downward bias. This is true for revisions subsequent to the first release and for comparisons between the first release and the SARB database data.

While systematic bias does not appear to be an important problem, the degree of dispersion is high for some variables. Based on the absolute average measure of dispersion, GDP and final consumption expenditure by household have the lowest dispersion of revisions, followed by disposable income of households and GDE. The absolute average dispersion between the first release and the SARB database for these four variables varies from 1.4 percentage points to 4.3 percentage points for the period 1984-2003. If one excludes the volatile 1984-1993 period, the absolute average dispersion varies from 1.3 percentage points to 3.0 percentage points.

Gross fixed capital formation, consumption expenditure by government, exports and imports are subject to absolute average dispersions (between first release and SARB database) of between 5.6 and 20.3 percentage points for the 1984-2003 period and between 3.5 and 14.8 percentage points for the less volatile 1994-2003 period. An absolute average dispersion of 5 percentage points[11] implies that a 95% confidence band for an initial growth rate of, say 7 per cent [12] could have a lower limit of -3 per cent and an upper limit of 17 per cent. For more dispersed revisions (as is the case for some magnitudes) the confidence bands are even wider.

Every month the balance of trade figures are published by the South African Revenue Services. These balance of trade figures, together with data on services, feed into the import and export figures in the national accounts. As indicated in Table 2, imports and exports, more than any other magnitude, are subject to very large revisions. The financial press is generally very sceptical of the monthly balance of trade figures, emphasising that these figures are provisional, volatile and not very believable.[13] The dispersion measures for revisions to import and export growth rates support this scepticism. Exports, more than any other variable, are subject to large changes in the growth rates when the base year changes. This is due to the fact that a large proportion of South Africa's trade consist of price-volatile primary commodities (Johan van den Heever, personal communication, 2006).

A positive feature from Table 2 is the fact that the degree of dispersion of the revisions has decreased substantially from the 1984-1993 period to the 1994-2003 period.[14] This is especially true in comparisons between the first release and the SARB database. However, the degree of dispersion in the 1994-2003 period is likely to be understated, because, as mentioned previously, not all "final" revisions have been made. Nevertheless, based on a comparison of the first release to the eighth release, seven of the eight magnitudes indicate a decrease in dispersion of the revisions. On average the absolute average dispersion has decreased by about 40 per cent.

[11] The average dispersion is similar to the standard deviation if there are no outliers amongst the revisions. However, if some revisions are somewhat further away from the bulk of the observations, the standard deviation is greater than the absolute average deviation. See the tables in the Appendix for an illustration of the differences between these two measures of dispersion.

[12] For all eight national accounting magnitudes, the dispersion in the revisions in the 1984-1993 period was driven by the highly volatile 1984-1988 period. The dispersion in the revisions in the 1989-1993 period is generally much lower than in the 1984-1988 period and more comparable to the subsequent period.

[13] This comment was regularly made by Michael Coulson on SAfm's Market Update. According to press reports, poor quality trade statistics are blamed on delayed reporting by export and import companies and administrative backlogs within SARS (see Katzenellenbogen, 2001)

[14] For all eight national accounting magnitudes, the dispersion in the revisions in the 1984-1993 period was driven by the highly volatile 1984-1988 period. The dispersion in the revisions in the 1989-1993 period is generally much lower than in the 1984-1988 period and more comparable to the subsequent period.

3.4 International comparisons

National accounting data revisions are not unique to South Africa. They are a fact of life in all countries. While the results discussed in the previous section may be worrying, they need to be compared to other countries. The aim of this section is to compare the dispersion of South Africa's revisions with those of countries for which similar magnitudes have been published.

The results of two studies (York and Atkinson, 1997 and Young, 1993) are presented in Table 3. Both studies concluded that the revisions are generally unbiased (not shown in Table 3). This finding is not generally true for South Africa since revisions in GDP, consumption expenditure and disposable income by households have a significant positive bias, especially after 1994.

The magnitudes and relative rankings of the dispersion of the revisions of GDP and its expenditure components in South Africa compare rather well with that of seven OECD countries. In the seven OECD countries investigated by York and Atkinson, GDP revisions (in four cases) and private consumption expenditure revisions (in three cases) displayed the least dispersion, the same as in South Africa.[15] In OECD countries the two GDP components which displayed the greatest degree of dispersion are imports and exports. This is also the case in South Africa, although the degree of dispersion for South Africa is significantly larger (for both the 1984-1993 and 1994-2003 periods).

In summary, the comparison with other countries shows that: (1) the degree of dispersion of the national accounts revisions are smaller for GDP than for the components of GDP, both in South Africa and a group of seven OECD countries; (2) other than import and export revisions, where the degree of dispersion is much higher for South Africa than for the seven OECD countries, the dispersion of the revisions are similar; and (3) in absolute terms, the dispersion of the revisions are substantial, especially between the first release and the SARB database.

This leads to the following question: do these revisions have a substantial impact on regression results or not? This is addressed in the following section.

4 Do revisions matter when doing economic analyses?

As a measure of the impact of such revisions, this section presents the impact of revisions on two fairly well established relationships in macroeconomics for which appropriate data, within the context of this study, are available. These are the relationships between consumption expenditure and disposable income by households, and between imports and GDE. Formally the two equations are presented as follows:

$$\%\Delta Cons = \alpha_1 + \alpha_2 \%\Delta PDI \tag{5}$$

$$\%\Delta M = \beta_1 + \beta_2 \%\Delta GDE \tag{6}$$

where $\%\Delta Cons$ is the percentage change in consumption expenditure by households, $\%\Delta PDI$ is the percentage change in disposable income by households, $\%\Delta M$ is the percentage change in imports of goods and services, and $\%\Delta GDE$ is the percentage change in GDE. The magnitudes of particular interest are α_2 and β_2, since these represent the income elasticity of consumption and the demand elasticity of imports, respectively.

[15] Private consumption expenditure is called final consumption expenditure by households in South Africa.

One could argue, quite correctly, that these two models are underspecified and would thus yield biased estimates. This criticism is not disputed, but the point of the exercise is not to obtain a "perfect" or even good regression equation. The point is to illustrate the impact of different data vintages on the magnitude and significance of coefficients, which is independent of the model specification. Including additional data from a database (which has presumably gone through a number of revisions as well) in order to obtain a better specified model will add unnecessary complexity.[16]

The results for equation 5 are shown in Table 4 for three periods: 1984-2003, 1984-1993 and 1994-2003. For the full period (1984-2003), the "income elasticity of consumption" (α_2) is calculated as 0.51 if one uses first release data, 0.48 if one uses fourth release data, 0.25 if one uses eighth release data and 0.29 if one uses the SARB database data. Even though the estimated coefficients differ substantially depending on the vintage of the data, they are not statistically different from each other.

A similar pattern is seen for the two sub-periods, namely that the coefficient α_2 tends to decrease as the data becomes more revised.

The Durbin-Watson statistics are also quite sensitive to the data vintage, especially in the 1994-2003 period, ranging from 0.51 for the regression based on fourth release data to 1.25 for the regression based on the SARB database data. For the other periods the differences are less dramatic, but certainly not negligible.

The results for equation (6) are shown in Table 5. The coefficient β_2 can be regarded as the demand elasticity of imports. If one considers the whole period first, β_2 remains fairly constant around a value of 2.5 for regressions based on first, fourth and eighth releases. However, β_2 falls to 1.55 if one uses the SARB database, which is significantly different from the estimates based on earlier releases. A similar pattern is seen for the 1984-1993 sub-period, where β_2 falls from around 2.7, if based on the earlier data releases, to 1.90 if based on the SARB database data. The differences in the coefficients are marginally significant.

In absolute, although not statistical terms, the change in the β_2 coefficients is most dramatic for the 1994-2003 period. As the data becomes more revised, the β_2 coefficients decrease from 2.25 (first release), to 1.95 (fourth release), 1.30 (eighth release) and 0.67 (SARB database). The adjusted R^2 value decreases from 0.257 (first release) to an insignificant 0.035 (SARB database data).

These two examples, especially those for the import function, are quite distressing. As one application, consider a forecaster who aims to predict the growth of imports. Typically the structural model would be based on the SARB or similar database. Based on the results of Table 5, a one percent increase in GDE would be predicted to result in a 0.67 per cent increase in imports (considering the 1994-2003 period). However, according to first release data, the import elasticity of demand is 2.25. In a case like this, the forecasts would systematically under-predict the actual percentage change in imports, since the forecasts are typically compared to the first releases.

Consider a second application. Orphanides (2001) investigates the impact of informational problems associated with the implementation and interpretation of simple monetary rules. He indicates that there is substantial empirical support for the Taylor Rule,[17] but that these analyses

[16] All the data were subject to Dickey-Fuller tests to establish whether they are stationary or not. They were indeed found to be stationary. This being the case, spurious regression is not a problem and it is unnecessary to test for cointegration. The results are shown in Table B.1 in the Appendix.

[17] In its most basic formulation, the Taylor Rule states that the central bank's lending rate is a linear function of the inflation rate and the output gap. If inflation or the output gap (defined as actual less potential output) increases, the central bank will increase the interest rate.

are based nearly exclusively on revised data. When the monetary authorities make their policy decisions the best data they have at their disposal is provisional and uncertain, and could be substantially revised subsequently. Orphanides shows that it is much more difficult to predict the stance of monetary policy using real-time data (i.e. data that were available at the time the decision was made), based on the Taylor Rule than if one uses revised data. He concludes that "interpretation of historical policy based on revised data instead of the data available to policy makers when policy decisions were made appears to be of questionable value" (Orphanides, 2001: 983).

5 Conclusion

This paper has probably raised more questions than answers.

Policy and decision makers require economic data to guide their decisions. The first release of data is usually watched with some degree of expectation since it often has significant bearing on future policy decisions. What this paper has shown is that the first releases of the data, subsequent releases of the data and the "final" data can differ quite substantially. The fact that other countries also experience significant revisions in their data provides some comfort, but does not change the conclusion that the first release of the data is not very robust. As such, initial data releases should always be viewed with great caution.

In fairness to the statistical authorities, the compilation of the national accounts is immensely complex, and in a sense more an art than a science. As an official of the SARB points out, the challenge is to transform very heterogeneous, often incomplete, inconsistent, partly outdated and frequently changing data sets and points into one complete, consistent and up to date picture of the domestic economy (Harald Wagner, personal communication, 2006).

Structural econometric models that aim to understand economic relationships or predict the future are typically based on data that are derived from a database. Other than the most recent data, such data has gone through many revisions. What this analysis has shown, admittedly with only two examples, is that the magnitudes and significance of relationships are strongly affected by the vintage of the data.

This paper does not suggest that first data releases have no value whatsoever. Statistics, even subject to error, are necessary to direct policy decisions. The paper also does not propose that the statistical authorities should delay the first release to improve accuracy at the expense of timeliness.[18]

Even though most users of national accounting data understand that the first release is subject to review and change, first release data is often reported in the press as gospel.[19] Based on the analysis presented here, this paper proposes that the statistical authorities and the financial press should present the first releases of data with more circumspection, emphasising that the data are preliminary and provisional, and subject to revision. To emphasise the provisional nature of the date, the statistical authorities and the financial press might publish informal confidence

[18] As was seen in Fig. 2, the large revisions were not between the first and the second releases, but between the first release and later releases, and between the first release and the SARB database. Should the statistical authorities decide to delay the first release by three months, the release that is currently called the "second release" would become the first release. The magnitudes of the revisions between the second release and the SARB database are still very large.

[19] A perusal of articles in the Business Day illustrates this point. When national accounting figures get released the data are often discussed at length (e.g. contributions by sectors, actual vs. expected growth rates, commentators' views, etc.), but the fact that the data are subject to revision typically receives only cursory mention.

intervals around the initial estimate. For example, a growth rate could be reported as follows: "The provisional growth rate for GDP in the fourth quarter of 2005 is 3.3 per cent, but this is subject to future revisions. Based on previous experience, the actual growth rate is likely to be between 2.3 and 5.0 per cent." The magnitude of the confidence interval should indicate to users how reliable the point estimate is: the greater (smaller) the confidence interval, the less (more) reliable the point estimate is. Also, if users are provided with "lower" and "upper" confidence intervals, this should enhance their decision making, allowing them to do sensitivity analyses.

Buying a second-hand article and using official national accounts statistics may not seem to have much in common. However, they share an important principle: caveat emptor – let the buyer/user beware. That is the main message of this paper.

References

[1] Katzenellenbogen, J. (2001). Global slowdown affecting exports. *Business Day*, 1 June.

[2] Laubscher, P. (1992). The BER's quarterly forecasts: an analysis of its accuracy, 1982-1991, *Studies in Economics and Econometrics,* Vol. 16(3): 63-83.

[3] Mankiw, N.G. and Shapiro, M.D. (1986). News or noise? An analysis of GDP revisions. NBER *Working Paper 1939*, NBER.

[4] Mantshimuli, T.H. (2004). Note on the benchmarking and rebasing of South Africa's national accounts. *South African Reserve Bank Quarterly Bulletin,* Vol. 234: 62-66.

[5] Manuel, T.A. (2006). Budget speech. *Republic of South Africa.*

[6] Orphanides, A. (2001). Monetary policy rules based on real-time data. *American Economic Review*, Vol. 91(4): 964-985.

[7] Palis, R., R., Ramos, R.L.O. and Robitaille, P. (2004). News or noise? An analysis of Brazilian GDP announcements. *International Finance Discussion Paper no 776.*

[8] Prinsloo, J.W. (1999). Revision of South Africa's national accounts. *South African Reserve Bank Quarterly Bulletin*, Vol. 212: 74-79.

[9] Prinsloo, J.W. and van Tongeren, J.W. (2002). Decentralised implementation of the 1993 system of national accounts: South Africa's experience. *South African Reserve Bank Quarterly Bulletin*, Vol. 223: 76-88.

[10] South African Reserve Bank. (2006). Monetary Policy Review May 2006.

[11] Van Walbeek, C. P. and van Graan, M. (2006). A directional analysis of the Bureau of Economic Research's quarterly forecasts. Unpublished document.

[12] York, R. and Atkinson, P. (1997). The reliability of quarterly national accounts in seven major countries: A user's perspective. *Economics Department Working Papers no 171.* Paris: OECD.

[13] Young, A.H. (1993). Reliability and accuracy of the quarterly estimates of GDP. *Survey of Current Business*, Vol. 20: 29-43.

Table 1: Growth rates of real GDE as published in the SARB Quarterly Bulletin

	Date of Quarterly Bulletin and number							
Quarter	June 1993 No. 188	September 1993 No. 189	December 1993 No. 190	March 1994 No. 191	June 1994 No. 192	September 1994 No. 193	December 1994 No. 194	March 1995 No. 195
1993Q1	**8.2**	**8.9**	**9.2**	**3.9**	**4.1**	**4.1**	**4.5**	**4.4**
1993Q2		-11.5	-12.1	-6.2	-3.5	-3.5	-3.0	-3.0
1993Q3			15.7	12.6	10.6	10.6	9.7	9.7
1993Q4				11.3	6.9	6.9	8.6	8.6
1994Q1					**2.5**	**3.3**	**0.9**	**2.1**
1994Q2						5.1	4.9	6.0
1994Q3							**9.2**	**9.7**
1994Q4								**7.1**

Table 2: Summary of bias and volatility of changes between first release and SARB database data

Variable	Bias	Absolute dispersion in percentage points (equation 2)					
		First release vs eighth release			First release vs SARB data		
		1984 - 1993	1994 - 2003	1984 - 2003	1984 - 1993	1994 - 2003	1984 - 2003
Gross domestic product	Significant upward bias, especially after 1994	1.015 (1)	0.769 (2)	0.892 (1)	1.501 (1)	1.297 (1)	1.399 (1)
Gross domestic expenditure	No significant bias (at 10%) in any period	3.289 (3)	2.477 (5)	2.883 (4)	5.216 (3)	3.015 (4)	4.115 (3)
Consumption expenditure by households	Significant upward bias, especially after 1994, but also before 1993	1.800 (2)	0.759 (1)	1.279 (2)	2.324 (2)	1.298 (2)	1.811 (2)
Consumption expenditure by general government	Significant downward bias in 1994-1998; significant upward bias after 1999; no monotone trends	10.818 (8)	3.395 (6)	7.106 (6)	11.603 (6)	4.252 (6)	7.927 (6)
Gross fixed capital formation	No significant bias (at 5%) in any period	4.621 (4)	2.169 (4)	3.395 (5)	7.666 (5)	3.548 (5)	5.607 (5)
Exports of goods and services	No significant bias (at 10%) in any period	7.741 (6)	8.956 (8)	8.349 (7)	25.737 (8)	14.842 (8)	20.289 (8)
Imports of goods and services	No significant bias (at 10%) in any period	9.928 (7)	7.954 (7)	8.941 (8)	12.566 (7)	9.414 (7)	10.990 (7)
Disposable income of households	Significant upward bias after 1994	4.664 (5)	1.095 (3)	2.879 (3)	6.340 (4)	2.267 (3)	4.304 (4)

Note: rank of measure of dispersion, compared to other national accounting aggregates, is shown in brackets

Table 3: International comparison of dispersion of revisions

Country	Period	GDP	Private consumption	Government consumption	Gross fixed capital formation	Exports	Imports
US[1]	1980Q1-1993Q4	1.68	1.76	3.92	5.08	5.48	7.24
Japan[1]	1980Q1-1993Q4	2.00	1.68	3.68	3.04	6.60	5.36
West Germany[1]	1980Q1-1993Q4	2.40	3.28	3.80	4.76	4.40	4.76
France[1]	1980Q1-1993Q4	1.16	1.48	1.60	3.44	6.00	5.76
Italy[1]	1980Q1-1993Q4	1.52	1.52	1.24	3.24	6.20	4.80
UK[1]	1980Q1-1993Q4	3.20	2.48	3.56	7.88	7.32	7.32
Canada[1]	1980Q1-1993Q4	1.56	1.40	3.16	5.20	7.44	8.20
Average	1980Q1-1993Q4	1.93	1.94	2.99	4.66	6.21	6.21
US[2]	1978Q1-1982Q4	1.75	2.12	3.75	3.82	7.07	5.71
US[2]	1983Q1-1991Q4	1.33	1.30	4.89	3.64	5.67	9.61
South Africa	1984Q1-1993Q4	1.50	2.32	11.60	7.67	25.74	12.57
South Africa	1994Q1-2003Q4	1.30	1.30	4.25	3.55	14.84	9.41

Sources:

1. York and Atkinson, 1997. In the original paper the quarter-on-quarter growth rates were not annualised. In order to get an annualised figure, the dispersion figures in the source are multiplied by four.
2. Young, 1993

Table 4: Regression results: $\%\Delta Cons = a_1 + a_2\ \%\Delta PDI$

	Constant		Independent variable				
Data vintage	Coefficient α_1	t-value	Coefficient α_2	t-value	Adjusted R^2	Durbin-Watson	Obs.
1984Q1 – 2003Q4							
1	1.07	2.75	0.51	3.98	0.158	2.55	80
4	1.44	3.63	0.48	4.81	0.219	2.05	80
8	1.41	3.39	0.25	3.28	0.112	1.62	78
SARB	1.86	4.24	0.29	5.09	0.240	1.84	80
1984Q1 – 1993Q4							
1	0.64	0.89	0.42	2.10	0.080	2.62	40
4	1.13	1.52	0.43	2.85	0.155	2.14	40
8	0.52	0.71	0.21	2.03	0.076	1.85	39
SARB	1.08	1.34	0.27	3.32	0.205	1.99	40
1994Q1 – 2003Q4							
1	1.22	4.94	0.66	6.36	0.503	1.07	40
4	1.27	4.15	0.70	6.18	0.489	0.51	40
8	1.59	4.70	0.55	4.92	0.379	0.86	39
SARB	2.17	7.28	0.45	6.42	0.508	1.25	40

Table 5: Regression results: $\%\Delta M = \beta_1 + \beta_2\ \%\Delta GDE$

	Constant		Independent variable				
Data vintage	Coefficient β_1	t-value	Coefficient β_2	t-value	Adjusted R^2	Durbin-Watson	Obs.
1984Q1 – 2003Q4							
1	3.63	1.57	2.45	9.05	0.506	2.69	80
4	3.70	1.59	2.60	9.72	0.542	2.50	80
8	2.86	1.18	2.53	9.71	0.548	2.52	78
SARB	3.67	1.48	1.55	4.94	0.229	2.72	80
1984Q1 – 1993Q4							
1	6.30	1.68	2.51	7.34	0.586	2.74	40
4	7.41	1.92	2.76	7.96	0.615	2.42	40
8	6.26	1.65	2.80	8.68	0.662	2.45	39
SARB	5.61	1.34	1.90	4.22	0.301	2.86	40
1994Q1 – 2003Q4							
1	1.41	0.47	2.25	3.80	0.257	2.69	40
4	1.76	0.64	1.95	3.72	0.248	2.74	40
8	2.82	0.96	1.30	2.57	0.151	2.40	39
SARB	4.43	1.66	0.67	1.56	0.035	2.07	40

Figure 1: First release and SARB database GDE growth rates

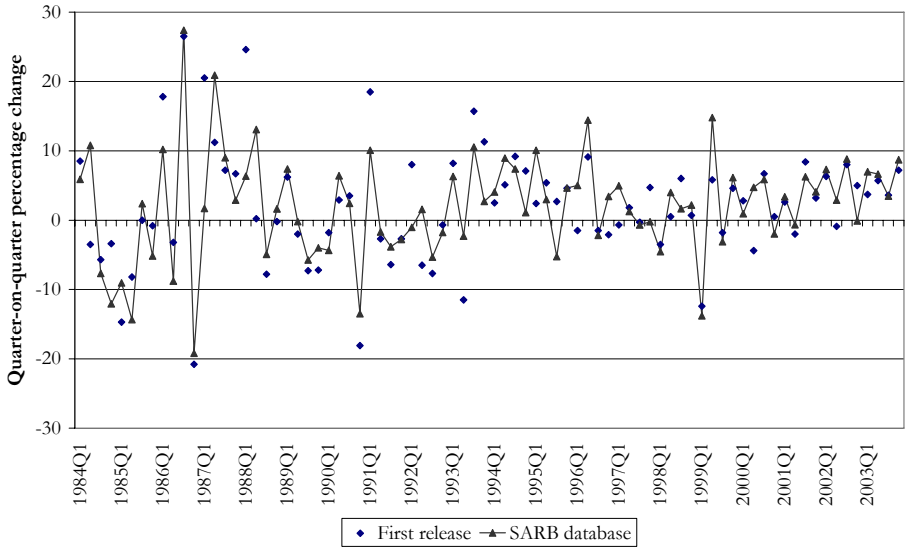


Figure 2: Revisions in the GDE estimates, in percentage points

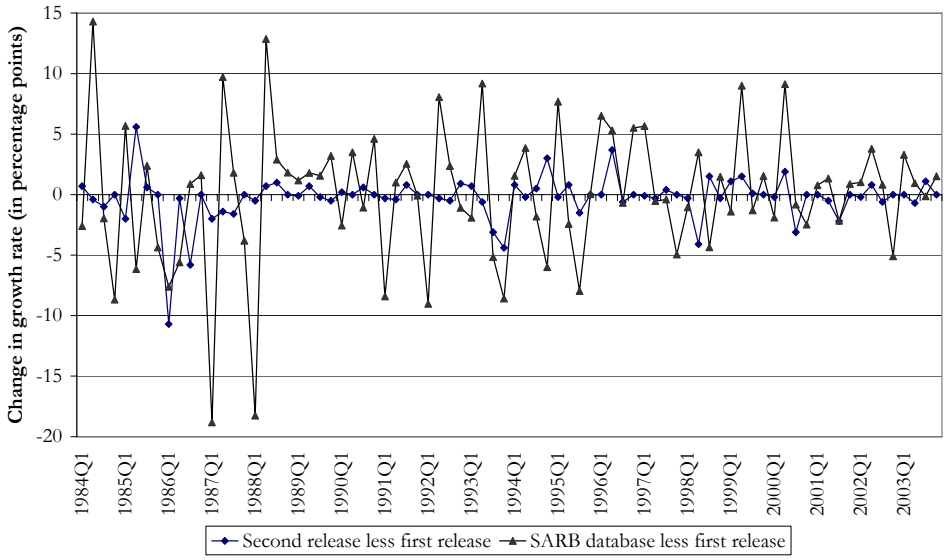


Figure 3: Histogram of revisions between first release and second release of GDE growth rates

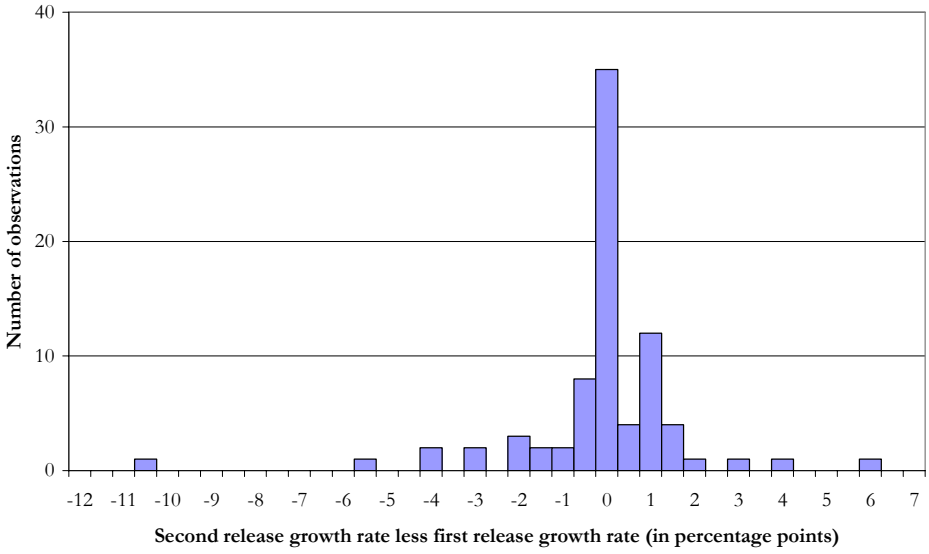
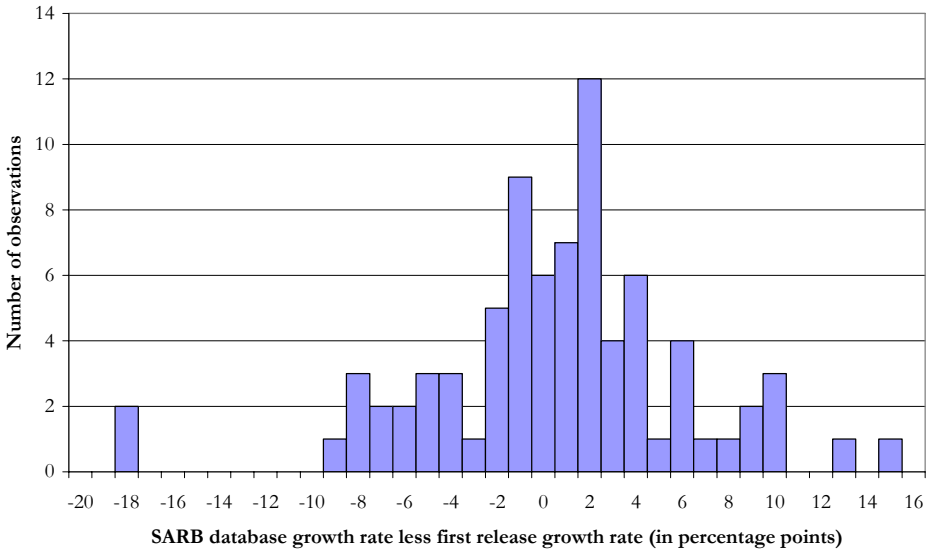


Figure 4: Histogram of revisions between first release and SARB database of GDE growth rates



APPENDIX

Measures of bias and dispersion in the official estimates of eight national accounting aggregates

Table A.1: Gross domestic product (SARB code 6006S)

	84:1 – 88:4	89:1 – 93:4	94:1 – 98:4	99:1 – 03:4	84:1 – 93:4	94:1 – 03:4	84:1 – 03:4
	n = 20	n = 20	n = 20	n = 20	n = 40	n = 40	n = 80
No of increases (cf. to first release)							
Second release	6 (30%)	7 (35%)	5 (25%)	6 (30%)	13 (33%)	11 (28%)	24 (30%)
Fourth release	8 (40%)	8 (40%)	9 (45%)	12 (60%)	16 (40%)	21 (53%)	37 (46%)
Eighth release#	11 (58%)	11 (55%)	14 (70%)	18 (95%)	22 (56%)	32 (82%)	54 (69%)
SARB data	9 (45%)	10 (50%)	15 (75%)	18 (90%)	19 (48%)	33 (83%)	52 (65%)
No. of decreases (cf. to first release)							
Second release	9 (45%)	7 (35%)	6 (30%)	4 (20%)	16 (40%)	10 (25%)	26 (33%)
Fourth release	11 (55%)	8 (40%)	3 (15%)	5 (25%)	19 (48%)	8 (20%)	27 (34%)
Eighth release#	7 (37%)	8 (40%)	4 (20%)	1 (5%)	15 (38%)	5 (13%)	20 (26%)
SARB data	11 (55%)	10 (50%)	5 (25%)	2 (10%)	21 (53%)	7 (18%)	28 (35%)
Bias (average change, cf. to first release)							
Second release	0.075	0.035	0.135	0.070	0.055	0.103	0.079
Fourth release	0.055	-0.065	0.220*	0.305**	-0.005	0.263***	0.129
Eighth release#	0.200	0.190	0.465***	0.921***	0.195	0.687***	0.441***
SARB data	0.310	-0.012	1.006***	1.271***	0.149	1.139***	0.644***
Dispersion (Standard deviation, cf. to first release)							
Second release	1.193	0.460	0.482	0.376	0.893	0.428	0.696
Fourth release	2.728	0.598	0.508	0.648	1.951	0.577	1.435
Eighth release#	1.894	0.780	0.605	0.666	1.416	0.669	1.127
SARB data	2.519	1.545	1.387	1.039	2.069	1.217	1.758
Dispersion (avg. of absolute deviation, cf. to first release)							
Second release	0.765	0.295	0.245	0.200	0.530	0.223	0.376
Fourth release	1.515	0.425	0.310	0.535	0.970	0.423	0.696
Eighth release#	1.463	0.590	0.585	0.963	1.015	0.769	0.892
SARB data	1.749	1.254	1.280	1.313	1.501	1.297	1.399
Relative dispersion (based on avg. of absolute deviation, cf. to first release)							
Second release	22.9	12.5	10.7	9.2	18.6	10.0	14.8
Fourth release	44.8	18.2	13.4	22.2	33.9	17.9	26.7
Eighth release#	49.7	27.6	24.2	31.4	40.2	28.1	33.9
SARB data	48.3	57.9	48.9	38.9	51.9	43.3	47.5

Because one observation is lost at the start and the end of the period under investigation due to data unavailability in the source documents, the analysis is restricted to 78 observations.

* Significantly different from zero at the 10 per cent level

** Significantly different from zero at the 5 per cent level

*** Significantly different from zero at the 1 per cent level

Table A.2: Gross domestic expenditure (SARB code 6012S)

	84:1 – 88:4	89:1 – 93:4	94:1 – 98:4	99:1 – 03:4	84:1 – 93:4	94:1 – 03:4	84:1 – 03:4
	n = 20	n = 20	n = 20	n = 20	n = 40	n = 40	n = 80
No of increases (cf. to first release)							
Second release	5 (25%)	6 (30%)	7 (35%)	6 (30%)	11 (28%)	13 (33%)	24 (30%)
Fourth release	8 (40%)	11 (55%)	11 (55%)	13 (65%)	19 (48%)	24 (60%)	43 (54%)
Eighth release#	9 (47%)	11 (55%)	11 (55%)	11 (58%)	20 (51%)	22 (56%)	42 (54%)
SARB data	10 (50%)	11 (55%)	10 (50%)	12 (60%)	21 (53%)	22 (55%)	43 (54%)
No. of decreases (cf. to first release)							
Second release	10 (50%)	10 (50%)	9 (45%)	7 (35%)	20 (50%)	16 (40%)	36 (45%)
Fourth release	11 (55%)	8 (40%)	8 (40%)	7 (35%)	19 (48%)	15 (38%)	34 (43%)
Eighth release#	9 (47%)	9 (45%)	9 (45%)	8 (42%)	18 (46%)	17 (44%)	35 (45%)
SARB data	10 (50%)	9 (45%)	10 (50%)	8 (40%)	19 (48%)	18 (45%)	37 (46%)
Bias (average change, cf. to first release)							
Second release	-0.855	-0.325	0.155	-0.050	-0.590	0.053	-0.269
Fourth release	-1.025	0.030	0.470	0.245	-0.498	0.358	-0.070
Eighth release#	-0.247	0.240	0.305	0.679	0.003	0.487	0.245
SARB data	-1.196	0.054	0.548	0.937	-0.571	0.742	0.086
Dispersion (Standard deviation, cf. to first release)							
Second release	3.088	1.276	1.555	1.136	2.348	1.348	1.929
Fourth release	3.894	2.614	1.869	1.691	3.317	1.763	2.674
Eighth release#	5.354	2.966	3.326	2.848	4.247	3.067	3.688
SARB data	8.661	5.016	4.423	3.456	7.015	3.923	5.686
Dispersion (avg. of absolute deviation, cf. to first release)							
Second release	1.715	0.715	0.915	0.700	1.215	0.808	1.011
Fourth release	2.905	1.550	1.460	1.255	2.228	1.358	1.793
Eighth release#	4.626	2.020	2.535	2.416	3.289	2.477	2.883
SARB data	6.588	3.844	3.564	2.466	5.216	3.015	4.115
Relative dispersion (based on avg. of absolute deviation, cf. to first release)							
Second release	20.3	10.0	22.0	15.9	15.6	18.9	16.8
Fourth release	26.3	25.1	34.8	29.1	25.9	31.9	27.9
Eighth release#	37.6	34.7	56.5	49.6	36.7	53.0	42.3
SARB data	68.1	81.8	80.1	44.6	72.6	60.4	67.6

Because one observation is lost at the start and the end of the period under investigation due to data unavailability in the source documents, the analysis is restricted to 78 observations.

There was no significant bias on the average change

Table A.3: Consumption expenditure by households (SARB code 6007S)

	84:1 – 88:4	89:1 – 93:4	94:1 – 98:4	99:1 – 03:4	84:1 – 93:4	94:1 – 03:4	84:1 – 03:4
	n = 20	n = 20	n = 20	n = 20	n = 40	n = 40	n = 80
No of increases (cf. to first release)							
Second release	7 (35%)	10 (50%)	9 (45%)	5 (25%)	17 (43%)	14 (35%)	31 (39%)
Fourth release	10 (50%)	16 (80%)	12 (60%)	11 (55%)	26 (65%)	23 (58%)	49 (61%)
Eighth release#	10 (53%)	14 (70%)	15 (75%)	14 (74%)	24 (62%)	29 (74%)	53 (68%)
SARB data	12 (60%)	15 (75%)	16 (80%)	16 (80%)	27 (68%)	32 (80%)	59 (74%)
No. of decreases (cf. to first release)e							
Second release	8 (40%)	5 (25%)	2 (10%)	2 (10%)	13 (33%)	4 (10%)	17 (21%)
Fourth release	7 (35%)	3 (15%)	6 (30%)	3 (15%)	10 (25%)	9 (23%)	19 (24%)
Eighth release#	9 (47%)	6 (30%)	5 (25%)	5 (26%)	15 (38%)	10 (26%)	25 (32%)
SARB data	8 (40%)	5 (25%)	4 (20%)	4 (20%)	13 (33%)	8 (20%)	21 (26%)
Bias (average change, cf. to first release)							
Second release	-0.155	0.155	0.255*	0.125	0.000	0.190**	0.095
Fourth release	0.500	0.475**	0.435**	0.255**	0.488	0.345***	0.416**
Eighth release#	-0.079	0.595**	0.605**	0.500**	0.267	0.554***	0.410*
SARB data	0.653	1.222*	1.398***	0.936***	0.937*	1.167***	1.052***
Dispersion (Standard deviation, cf. to first release)							
Second release	3.248	0.599	0.562	0.404	2.310	0.487	1.662
Fourth release	2.862	0.884	0.824	0.528	2.091	0.689	1.548
Eighth release#	3.527	1.021	0.961	0.850	2.556	0.898	1.909
SARB data	2.856	3.123	1.646	1.071	2.968	1.391	2.306
Dispersion (avg. of absolute deviation, cf. to first release)							
Second release	1.685	0.415	0.315	0.185	1.050	0.250	0.650
Fourth release	1.820	0.815	0.675	0.315	1.318	0.495	0.906
Eighth release#	2.658	0.985	0.875	0.637	1.800	0.759	1.279
SARB data	2.333	2.316	1.576	1.019	2.324	1.298	1.811
Relative dispersion (based on avg. of absolute deviation, cf. to first release)							
Second release	32.0	25.9	12.4	6.4	30.6	9.2	21.1
Fourth release	33.1	46.8	24.8	10.4	36.4	17.2	27.9
Eighth release#	51.8	54.4	29.7	20.3	52.5	25.0	39.6
SARB data	40.8	90.1	42.5	28.3	56.1	35.5	46.4

\# Because one observation is lost at the start and the end of the period under investigation due to data unavailability in the source documents, the analysis is restricted to 78 observations.

* Significantly different from zero at the 10 per cent level

** Significantly different from zero at the 5 per cent level

*** Significantly different from zero at the 1 per cent level

Table A.4: Consumption expenditure by general government (SARB code 6008S)

	84:1 – 88:4	89:1 – 93:4	94:1 – 98:4	99:1 – 03:4	84:1 – 93:4	94:1 – 03:4	84:1 – 03:4
	n = 20	n = 20	n = 20	n = 20	n = 40	n = 40	n = 80
No of increases (cf. to first release)							
Second release	9 (45%)	8 (40%)	6 (30%)	5 (25%)	17 (43%)	11 (28%)	28 (35%)
Fourth release	12 (60%)	11 (55%)	11 (55%)	11 (55%)	23 (58%)	22 (55%)	45 (56%)
Eighth release#	10 (53%)	13 (65%)	9 (45%)	15 (79%)	23 (59%)	24 (62%)	47 (60%)
SARB data	11 (55%)	10 (50%)	4 (20%)	20 (100%)	21 (53%)	24 (60%)	45 (56%)
No. of decreases (cf. to first release)e							
Second release	5 (25%)	7 (35%)	5 (25%)	3 (15%)	12 (30%)	8 (20%)	20 (25%)
Fourth release	7 (35%)	7 (35%)	8 (40%)	5 (25%)	14 (35%)	13 (33%)	27 (34%)
Eighth release#	8 (42%)	7 (35%)	11 (55%)	4 (21%)	15 (38%)	15 (38%)	30 (39%)
SARB data	9 (45%)	9 (45%)	16 (80%)	0 (0%)	18 (45%)	16 (40%)	34 (43%)
Bias (average change, cf. to first release)							
Second release	-2.360	-0.645	0.375	0.015	-1.503	0.195	-0.654
Fourth release	-1.560	-0.495	0.685	0.125	-1.028	0.405	-0.311
Eighth release#	0.732	-0.830	-2.570	1.379***	-0.069	-0.646	-0.358
SARB data	0.643	-1.043	-3.73***	2.685***	-0.200	-0.524	-0.362
Dispersion (Standard deviation, cf. to first release)							
Second release	16.312$	3.957	1.916	0.139	11.748	1.353	8.352
Fourth release	20.332$	4.887	2.644	0.489	14.605	1.898	10.373
Eighth release#	23.816$	10.189	7.875	1.967	17.922	6.070	13.296
SARB data	26.746$$	9.841	5.313	2.293	19.909	5.184	14.456
Dispersion (avg. of absolute deviation, cf. to first release)							
Second release	5.870$	2.255	1.015	0.075	4.063	0.545	2.304
Fourth release	10.350$	3.685	1.895	0.355	7.018	1.125	4.071
Eighth release#	15.121$	6.730	5.020	1.684	10.818	3.395	7.106
SARB data	16.556$$	6.650	5.818	2.685	11.603	4.252	7.927
Relative dispersion (based on avg. of absolute deviation, cf. to first release)							
Second release	37.4	25.8	27.7	2.3	33.3	15.6	29.4
Fourth release	66.1	42.3	43.8	10.2	57.6	28.8	50.6
Eighth release#	95.9	80.6	97.3	53.4	90.4	81.2	88.0
SARB data	101.2	100.7	140.0	60.3	101.0	98.7	100.4

Because one observation is lost at the start and the end of the period under investigation due to data unavailability in the source documents, the analysis is restricted to 78 observations.

*** Significantly different from zero at the 1 per cent level

$ Strongly influenced by 1987Q1: original data = 106.4% growth

$$ Large discrepancies in 1984Q3 (48.8%), 1985Q3 (51.8%) and 1987Q1 (79.4%)

Table A.5: Gross fixed capital formation (SARB code 6009S)

	84:1 – 88:4	89:1 – 93:4	94:1 – 98:4	99:1 – 03:4	84:1 – 93:4	94:1 – 03:4	84:1 – 03:4
	n = 20	n = 20	n = 20	n = 20	n = 40	n = 40	n = 80
No of increases (cf. to first release)							
Second release	11 (55%)	8 (40%)	10 (50%)	4 (20%)	19 (48%)	14 (35%)	33 (41%)
Fourth release	8 (40%)	10 (50%)	14 (70%)	9 (45%)	18 (45%)	23 (58%)	41 (51%)
Eighth release#	9 (47%)	13 (65%)	15 (75%)	7 (37%)	22 (56%)	22 (56%)	44 (56%)
SARB data	10 (50%)	14 (70%)	14 (70%)	6 (30%)	24 (60%)	20 (50%)	44 (55%)
No. of decreases (cf. to first release)e							
Second release	5 (25%)	6 (30%)	4 (20%)	8 (40%)	11 (28%)	12 (30%)	23 (29%)
Fourth release	10 (50%)	8 (40%)	4 (20%)	9 (45%)	18 (45%)	13 (33%)	31 (39%)
Eighth release#	10 (53%)	6 (30%)	5 (25%)	10 (53%)	16 (41%)	15 (38%)	31 (40%)
SARB data	10 (50%)	6 (30%)	6 (30%)	14 (70%)	16 (40%)	20 (50%)	36 (45%)
Bias (average change, cf. to first release)							
Second release	3.256*	0.290	0.985	-0.445*	1.773*	0.270	1.021**
Fourth release	1.105	0.255	1.265*	-0.455	0.680	0.405	0.542
Eighth release#	0.021	1.090	1.160*	-0.305	0.569	0.446	0.508
SARB data	0.042	2.861*	1.579	-0.582	1.451	0.499	0.975
Dispersion (Standard deviation, cf. to first release)							
Second release	7.963	1.325	3.154	1.041	5.831	2.429	4.502
Fourth release	9.663	2.052	3.009	2.116	6.908	2.711	5.216
Eighth release#	8.703	3.258	2.810	2.864	6.440	2.896	4.961
SARB data	13.931	6.184	4.648	3.758	10.734	4.313	8.142
Dispersion (avg. of absolute deviation, cf. to first release)							
Second release	4.195	0.950	1.655	0.685	2.573	1.170	1.871
Fourth release	5.815	1.655	1.945	1.545	3.735	1.745	2.740
Eighth release#	6.495	2.840	2.310	2.021	4.621	2.169	3.395
SARB data	9.784	5.549	4.039	3.058	7.666	3.548	5.607
Relative dispersion (based on avg. of absolute deviation, cf. to first release)							
Second release	47.6	17.6	23.5	10.0	36.2	16.8	26.6
Fourth release	51.9	29.6	26.5	22.7	44.5	24.7	35.4
Eighth release#	61.2	45.3	32.0	31.4	55.1	31.7	44.6
SARB data	88.7	83.0	52.6	43.9	86.6	48.5	69.3

Because one observation is lost at the start and the end of the period under investigation due to data unavailability in the source documents, the analysis is restricted to 78 observations.

* Significantly different from zero at the 10 per cent level

** Significantly different from zero at the 5 per cent level

Table A.6: Exports of goods and services (SARB code 6013S)

	84:1 – 88:4	89:1 – 93:4	94:1 – 98:4	99:1 – 03:4	84:1 – 93:4	94:1 – 03:4	84:1 – 03:4
	n = 20	n = 20	n = 20	n = 20	n = 40	n = 40	n = 80
No of increases (cf. to first release)							
Second release	8 (40%)	8 (40%)	5 (25%)	6 (30%)	16 (40%)	11 (28%)	27 (34%)
Fourth release	10 (50%)	9 (45%)	10 (50%)	11 (55%)	19 (48%)	21 (53%)	40 (50%)
Eighth release#	10 (53%)	12 (60%)	11 (55%)	11 (58%)	22 (56%)	22 (56%)	44 (56%)
SARB data	10 (50%)	12 (60%)	11 (55%)	12 (60%)	22 (55%)	23 (58%)	45 (56%)
No. of decreases (cf. to first release)e							
Second release	5 (25%)	3 (15%)	7 (35%)	3 (15%)	8 (20%)	10 (25%)	18 (23%)
Fourth release	8 (40%)	7 (35%)	8 (40%)	6 (30%)	15 (38%)	14 (35%)	29 (36%)
Eighth release#	9 (47%)	6 (30%)	9 (45%)	8 (42%)	15 (38%)	17 (44%)	32 (41%)
SARB data	10 (50%)	8 (40%)	9 (45%)	8 (40%)	18 (45%)	17 (43%)	35 (44%)
Bias (average change, cf. to first release)							
Second release	0.775	1.010	-0.560	0.430	0.893	-0.065	0.414
Fourth release	-0.500	0.510	-1.110	1.060	0.005	-0.025	-0.010
Eighth release#	-1.526	0.185	1.145	0.853	-0.649	1.141	0.246
SARB data	3.159	5.497	2.685	2.395	4.328	2.540	3.434
Dispersion (Standard deviation, cf. to first release)							
Second release	7.693	2.588	3.778	3.137	5.667	3.464	4.691
Fourth release	11.339	5.644	6.655	5.445	8.856	6.101	7.556
Eighth release#	13.156	6.891	14.977	7.968	10.319	11.929	11.117
SARB data	34.175	30.393	28.956	11.559	31.944	21.762	27.173
Dispersion (avg. of absolute deviation, cf. to first release)							
Second release	4.125	1.500	2.260	1.550	2.813	1.905	2.359
Fourth release	8.580	3.600	4.870	4.220	6.090	4.545	5.318
Eighth release#	10.895	4.745	11.215	6.579	7.741	8.956	8.349
SARB data	29.296	22.177	20.690	8.994	25.737	14.842	20.289
Relative dispersion (based on avg. of absolute deviation, cf. to first release)							
Second release	25.5	9.6	17.2	10.4	17.7	13.6	15.8
Fourth release	53.5	25.8	42.3	29.5	40.6	35.2	38.1
Eighth release#	69.3	37.0	71.5	40.6	54.4	56.2	55.4
SARB data	96.0	82.8	82.3	52.7	89.8	70.4	81.6

Because one observation is lost at the start and the end of the period under investigation due to data unavailability in the source documents, the analysis is restricted to 78 observations.

There was no significant bias on the average change

Table A.7: Imports of goods and services (SARB code 6014S)

	84:1 – 88:4	89:1 – 93:4	94:1 – 98:4	99:1 – 03:4	84:1 – 93:4	94:1 – 03:4	84:1 – 03:4
	n = 20	n = 20	n = 20	n = 20	n = 40	n = 40	n = 80
No of increases (cf. to first release)							
Second release	3 (15%)	3 (15%)	5 (25%)	4 (20%)	6 (15%)	9 (23%)	15 (19%)
Fourth release	10 (50%)	8 (40%)	11 (55%)	9 (45%)	18 (45%)	20 (50%)	38 (48%)
Eighth release#	10 (53%)	11 (55%)	10 (50%)	8 (42%)	21 (54%)	18 (46%)	39 (50%)
SARB data	8 (40%)	12 (60%)	8 (40%)	9 (45%)	20 (50%	17 (43%)	37 (46%)
No. of decreases (cf. to first release)e							
Second release	8 (40%)	7 (35%)	5 (25%)	5 (25%)	15 (38%)	10 (25%)	25 (31%)
Fourth release	8 (40%)	10 (50%)	7 (35%)	9 (45%)	18 (45%)	16 (40%)	34 (43%)
Eighth release#	9 (47%)	8 (40%)	10 (50%)	10 (53%)	17 (44%)	20 (51%)	37 (47%)
SARB data	12 (60%)	8 (40%)	12 (60%)	11 (55%)	20 (50%)	23 (58%)	43 (54%)
Bias (average change, cf. to first release)							
Second release	-3.225	0.170	-0.665	-0.035	-1.528	-0.350	-0.939
Fourth release	-0.960	1.115	-0.160	0.725	0.078	0.283	0.180
Eighth release#	-0.168	0.750	0.285	-1.163	0.303	-0.421	-0.059
SARB data	-4.770	-0.453	-0.415	-0.819	-2.611	-0.617	-1.614
Dispersion (Standard deviation, cf. to first release)							
Second release	14.337$	1.746	3.831	1.717	10.226	2.948	7.501
Fourth release	15.687	6.097	6.782	5.543	11.794	6.130	9.340
Eighth release#	18.955	7.610	12.592	11.716	14.120	12.34	13.038
SARB data	18.594	12.841	13.638	11.959	15.923	12.662	14.329
Dispersion (avg. of absolute deviation, cf. to first release)							
Second release	6.225$	0.580	1.905	0.975	3.403	1.440	2.421
Fourth release	8.710	3.895	4.740	3.325	6.303	4.033	5.168
Eighth release#	14.368	5.710	8.765	7.100	9.928	7.954	8.941
SARB data	15.143	9.990	10.090	8.739	12.566	9.414	10.990
Relative dispersion (based on avg. of absolute deviation, cf. to first release)							
Second release	18.7	2.7	13.9	6.0	12.5	9.6	11.4
Fourth release	24.7	19.8	40.3	20.2	22.9	28.6	24.8
Eighth release#	41.6	31.2	62.5	47.8	37.9	55.2	44.0
SARB data	55.3	61.5	78.1	77.4	57.6	77.7	64.8

Because one observation is lost at the start and the end of the period under investigation due to data unavailability in the source documents, the analysis is restricted to 78 observations.

There was no significant bias on the average change

$ 1986Q1: large change from 35.1 % (first release) to -22.3% (second release). Checked at source.

Table A.8: Disposable income of households (SARB source 6246S)

	84:1 – 88:4	89:1 – 93:4	94:1 – 98:4	99:1 – 03:4	84:1 – 93:4	94:1 – 03:4	84:1 – 03:4
	n = 20	n = 20	n = 20	n = 20	n = 40	n = 40	n = 80
No of increases (cf. to first release)							
Second release	10 (50%)	7 (35%)	8 (40%)	12 (60%)	17 (43%)	20 (50%)	37 (46%)
Fourth release	8 (40%)	8 (40%)	10 (50%)	16 (80%)	16 (40%)	26 (65%)	42 (53%)
Eighth release#	12 (63%)	12 (60%)	13 (65%)	16 (84%)	24 (62%)	29 (74%)	53 (68%)
SARB data	13 (65%)	10 (50%)	12 (60%)	16 (80%)	23 (58%)	28 (70%)	51 (64%)
No. of decreases (cf. to first release)e							
Second release	6 (30%)	11 (55%)	8 (40%)	5 (25%)	17 (43%)	13 (33%)	30 (38%)
Fourth release	12 (60%)	11 (55%)	9 (45%)	3 (15%)	23 (58%)	12 (30%)	35 (44%)
Eighth release#	7 (37%)	8 (40%)	7 (35%)	3 (16%)	15 (38%)	10 (26%)	25 (32%)
SARB data	7 (35%)	10 (50%)	8 (40%)	4 (20%)	17 (43%)	12 (30%)	29 (36%)
Bias (average change, cf. to first release)							
Second release	1.560	-0.275	0.000	0.225*	0.643	0.113	0.378
Fourth release	0.480	-0.510	0.165	0.460***	-0.015	0.313*	0.419
Eighth release#	3.089	0.030	0.645*	0.674***	1.521	0.659***	1.090*
SARB data	4.149	-0.381	1.385*	1.294**	1.884	1.339***	1.612*
Dispersion (Standard deviation, cf. to first release)							
Second release	4.306	1.147	1.325	0.569	3.246	1.013	2.404
Fourth release	5.845	2.132	1.510	0.516	4.371	1.124	3.176
Eighth release#	9.222	3.100	1.610	0.865	6.891	1.285	4.943
SARB data	13.762	3.726	3.298	2.074	10.212	2.719	7.431
Dispersion (avg. of absolute deviation, cf. to first release)							
Second release	3.500	0.695	0.790	0.375	2.098	0.583	1.340
Fourth release	4.240	1.490	1.065	0.500	2.865	0.783	1.824
Eighth release#	7.058	2.390	1.375	0.800	4.664	1.095	2.879
SARB data	10.189	2.491	2.806	1.729	6.340	2.267	4.304
Relative dispersion (based on avg. of absolute deviation, cf. to first release)							
Second release	96.7	41.2	51.6	13.4	79.1	26.9	55.6
Fourth release	83.8	70.0	69.4	16.2	79.7	33.8	61.7
Eighth release#	96.7	113.8	67.6	24.4	100.7	41.5	79.2
SARB data	98.5	90.6	75.6	45.4	96.8	60.3	83.5

Because one observation is lost at the start and the end of the period under investigation due to data unavailability in the source documents, the analysis is restricted to 78 observations.

* Significantly different from zero at the 10 per cent level

** Significantly different from zero at the 5 per cent level

*** Significantly different from zero at the 1 per cent level

Table B.1: Testing for stationarity (see section 4)

Variable	Release	Lags included	Dickey-Fuller statistic	One-sided MacKinnon p-value
%ΔCons	1	5	-3.70	0.006
	4	3	-3.52	0.010
	8	0	-7.77	0.000
	SARB	3	-3.92	0.003
%ΔPDI	1	1	-3.35	0.016
	4	4	-3.43	0.016
	8	1	-7.02	0.000
	SARB	0	-10.23	0.000
%ΔM	1	0	-11.69	0.000
	4	0	-11.84	0.000
	8	0	-12.54	0.000
	SARB	0	-11.89	0.000
%ΔGDE	1	0	-10.53	0.000
	4	0	-10.40	0.000
	8	0	-11.68	0.000
	SARB	0	-10.92	0.000

The number of lags was automatically selected based on the Schwartz Information Criterion