CSAE WPS/2004-09

Construction of CPIX Data for Forecasting and Modelling in South Africa

JANINE ARON
Centre for the Study of African Economies,
Department of Economics, University of Oxford

&

JOHN MUELLBAUER
Nuffield College, University of Oxford

Revised version,
14 July 2004

Abstract: South Africa adopted inflation targeting in 2000, targeting the consumer price index (CPI) excluding mortgage interest cost (or CPIX), for "metropolitan and urban areas". Yet there is no clear technical account of the methodology of construction of CPI and CPIX by *Statistics South Africa*, as published by reputable government statistical agencies in other countries. This paper has two goals. First, we aim to enhance transparency by explaining the CPI methodology (as we understand it), and to encourage publication of an official technical handbook. We also raise various technical issues concerning CPI construction. Second, we produce estimates of CPIX ("metropolitan areas") back to 1970, on a consistent methodology, using monthly price indices, the appropriate weights, and linking correctly when rebasing. While the CPIX ("metropolitan and urban areas") measure only became relevant to monetary policy setting and wage contracts from 2000, and is published monthly only from 1997, a far longer time series is required for the forecasting and modelling exercises of the South African Reserve Bank (SARB), National Treasury and others. Our measure differs in some years from that published by *Statistics South Africa* (published monthly only back to 1994).

JEL codes: E31, E4

1. Introduction

South Africa adopted inflation targeting in 2000, with the target specified as an average rate of increase in the overall consumer price index (CPI) excluding the mortgage interest cost (the CPIX) of 3-6 percent per annum.[1] Amongst other requirements, the shift to inflation targeting demands good forecasting models of inflation and clarity on the mechanisms of monetary transmission (Leiderman and Svensson, 1995), both of which need reliable historical time series of price index data.

This paper is concerned with the lack of a clear technical account of the methodology of construction of both CPI and the targeted index, CPIX, by the official statistical agency, *Statistics South Africa* (which we abbreviate to *Statistics SA*) - by contrast with reputable official statistical agencies in other countries.[2] We aim to enhance transparency by explaining aspects of the methodology as we understand it, and thereby also to encourage the official publication of a technical handbook of CPI methodology. The issuance of index-linked bonds in principle legally requires such transparency.[3] We draw on non-technical descriptions available on *Statistics SA's* website, correspondence with *Statistics SA,* and on a report by Haglund (2000)[4], who was brought in as an outside expert to examine *Statistics SA's* methodology.

While the CPIX measure only became relevant to monetary policy setting and wage contracts from February, 2000, a far longer time series is required for rigorous forecasting and modelling exercises carried out regularly by the South African Reserve Bank (SARB) and the National Treasury. For instance, inflation and output forecasts are crucial inputs into the regular meetings of the Monetary Policy Committee of the SARB. The official statistical agency, *Statistics SA,* publishes monthly data for CPI ("metropolitan and urban areas") and for CPIX ("metropolitan and urban areas") only back to 1997.[5] We document the public

---

[1] The target announced in February, 2000, was to be reached within two years by 2002. This was revised in October, 2001 to 3-5 percent in 2004 and 2005; in October 2002, to 3-6 percent for 2004; and in February, 2003, to 3-6 percent for 2005. In November, 2003, the target definition also altered from a fixed target of an annual average rate over a calendar year to a continuous target, of 3-6 percent beyond 2006.

[2] No technical bulletin corresponding to the *Handbook of Methods*, Chapter 17 *(*Bureau of Labor Statistics, U.S., April 1997) is produced by *Statistics SA*, or otherwise obtainable from other government agencies. The only publicly available information on methodology is the descriptive (non-technical) account contained within the monthly bulletins (in the explanatory notes).

[3] We are grateful to Paul Collier for this point.

[4] This report is now available on our website with the kind permission of Mats Haglund, *Statistics Sweden*.

[5] The headline CPI and its components are defined using surveys of "metropolitan areas", covering 40 percent of household expenditure. The new targeted CPIX measure covers "metropolitan and urban areas", extending the coverage to over 80 percent of expenditure.

availability of CPI and CPIX series and their sub-components under various definitions and from various sources in Table 1. Weights of components for different definitions of CPI are given in Tables 2 and 3.

We produce monthly estimates of CPIX for "metropolitan areas" back to 1970, on a consistent methodology.[6] The earlier data are constructed by subtracting the mortgage interest component from monthly headline CPI ("metropolitan areas"), using the appropriate weights and linking correctly when rebasing and reweighting.

Our estimates differ from monthly data on *Statistics SA's* website (published only back to 1994 for the "metropolitan areas" index). Our estimates also differ somewhat from the internal SARB quarterly approximation for historical CPIX for "metropolitan areas", used by SARB modellers and the International Monetary Fund (e.g. Bhundia, 2002). Given that this measure plays an important part in the SARB's macroeconomic model, guiding policy, any biases in the measure are unfortunate.

In Section 2, we give an account of the methodology of construction of headline CPI. In Section 3, we relate this to the construction of CPIX, and present a consistent method using monthly prices indices for constructing historical data for CPIX ("metropolitan areas"). Conclusions are drawn in Section 4.

2.   The methodology of construction of headline CPI[7]

The following explanation draws on Haglund (2000), which is available on the *Statistics SA* website, and on many communications with Haglund.

The CPI is a Laspèyres-type index, using a fixed basket of goods to weight prices. Historical and current weights for CPI ("metropolitan") and CPI ("metropolitan and urban") indices are given in Tables 2 and 3. However, the CPI for South Africa is not a strict Laspèyres index, but a chained Laspèyres index. The implications of this are discussed below.

2.1   Data Sources for Prices and Weights

---

[6] To be more precise, we use the closest feasible approximation to the consistent methodology given data that are in the public domain.

[7] Our monthly estimates for CPIX ("metropolitan") back to 1970 are available to download from our website.

The main price data come from the "Survey of Consumer Prices", a monthly survey covering a sample of retailers operating in the South African economy. They are combined with price data obtained directly from insurance companies, electricity companies and others to obtain prices for the CPI. The weighting system for the CPI is calculated from the "Survey of Income and Expenditure of Households", last conducted in October 2000. The information obtained through this survey was re-weighted according to the 1996 Population Census figures in order to represent all households in South Africa. *Statistics SA* conducts a "Survey of Income and Expenditure of Households" every five years, covering a sample of 30,000 households. In 2000, the survey collected information on approximately 1,000 different goods and services groups. *Statistics SA* makes a further breakdown of these groups using supplementary sources. This process leads to approximately 1,500 groups on which the current calculation of the CPI is based.

2.2 Laspèyres index

For each period, a strict Laspèyres index compares the cost of a basket of goods at prices of period, t, with the cost of the same basket at the price of the base period, 0. The reference basket is defined as the goods bought at base period prices by the average of households in the sample. It takes the form

$$CPI_{[0,t]} = \frac{\sum_i p_{i,t} q_{i,0}}{\sum_i p_{i,0} q_{i,0}} = \frac{\sum_i \frac{p_{i,t}}{p_{i,0}} p_{i,0} q_{i,0}}{\sum_i p_{i,0} q_{i,0}} = \sum_i w_{i,0} (\frac{p_{i,t}}{p_{i,0}}) \times 100 \qquad (1)$$

where, for base period, 0, and reference period, t: $p_{i,0}$ is the price for the base period of the ith type of good, $i=1,...,n$; and $q_{i,0}$ is the quantity weight derived from the base period consumer expenditure survey. In the notation above, the CPI subscript [0,t] means prices at time t relative to prices on a base of 100 in t=0. Note that $w_{i,0} = p_{i,0} q_{i,0} / \sum_i p_{i,0} q_{i,0}$ is the survey expenditure share.[8]

---

[8] The same formula applies when *i* refers to a group of goods, when $(p_{i,t} / p_{i,0}) \times 100$ is a group price index, for example, for meat, made up of different types and cuts of meat.

In the methodology used by the *Bureau of Labor Statistics* (BLS) in the U.S., for the consumer price index measure, CPI-U, the base periods always exactly coincide with the period for which the weights (derived from the Consumer Expenditure Survey) are calculated, according to a strict Laspèyres index[9]. For example, from January, 2002 onwards, the BLS applied new 1999-2000 weights with a base period of 1999-2000 (i.e. when CPI-U and its sub-indices are set equal to 100). So from January 2002,

$$CPI_{[99-00,t]} = \sum_i w_{i,99-00} \frac{p_{i,t}}{p_{i,99-00}} \times 100 \qquad (2)$$

Thus, the subscript means prices at time t relative to prices on a base of 100 in 1999-2000.

A basic principle of statistical agencies is *not* to revise historical CPI data in percentage change form (in order not to alter historical inflation figures). Thus, to bring the historical figures for the previous weighting period (1993-95) onto the 1999-2000 base, the BLS applied the following conversion factor to $CPI_{[93-95,t]}$, up to December, 2001:

$$\frac{100}{CPI_{[93-95,99-00]}} \qquad (3)$$

It is possible, with the BLS method, to obtain small "spikes" (or outliers) in the month when the new weights are introduced, as items subject to large relative price change also might change substantially in quantities, hence weights, too (e.g., computers).[10]

2.3   Level factors

In South Africa, these "spikes" are avoided by the use of "level factors" (in the terminology of *Statistics SA*). It is not clear exactly what form these "level factors" take in the

---

[9] See *Handbook of Methods*, Chapter 17 *(*Bureau of Labor Statistics, U.S., April 1997), Part II.
[10] One way of explaining the possibility of spikes in the monthly change between December, 2001 and January, 2002 is as follows: the January index could have been computed on exactly the same basis as the December index (using the 1993-1995 weights, but scaling by the conversion factor in equation (3) so that the base is 100 in 1999-2000). Then the change in the index from December to January would have depended only on the change in prices of the components over the month. However, the January index actually uses using the 1999-2000 weights as in equation (2). So part of the change between December and January index values is due to the weights changing and part is due to price changes in the components, and this can generate a spike.

construction of the South African CPI. We have not been able to get clarity from *Statistics SA* on this point. Mats Haglund, who recently acted as a consultant to *Statistics SA* (Haglund, 2000), suggests there are two variants, which we call "level factors" A and B. The first of these he discussed in Haglund (2000), and the second in a recent communication.

2.3.1 Level factor A

The level of the new index is adjusted to the old by means of a "level factor", where splicing occurs, in the first month in which the new weights appear[11], to average prices for the base year. The historical weights are shown in Tables 2 and 3.

For example, to apply 2000 weights (first applied in January, 2002), two computations are done. First, an aggregate index, *I*, is calculated for monthly data from January, 2002 onwards, with average prices for 2000 as base, and with 2000 weights[12]:

$$I_{[2000,t]} = \sum_i w_{i,2000} \frac{P_{i,t}}{P_{i,2000}} \times 100 \tag{4}$$

where t refers to a month after January, 2002. Then the following "level factor", *L*, is computed with 1995 weights as

$$L_{Jan.2002} = \frac{\left(\dfrac{\sum_i w_{i,1995} \dfrac{P_{i,Jan.2002}}{P_{i,1995}}}{\sum_i w_{i,1995} \dfrac{P_{i,2000}}{P_{i,1995}}} \right)}{\sum_i w_{i,2000} \dfrac{P_{i,Jan.2002}}{P_{i,2000}}} = \frac{\left[\dfrac{I_{[1995,Jan.2002]}}{I_{[1995,2000]}} \right]}{I_{[2000,Jan.2002]}} \tag{5}$$

The product of equations (4) and (5) yields the level-adjusted index, the official CPI on a base of 100 in 2000, using 2000 weights from January, 2002 onwards. Equivalently, the index is

---

[11] The weights are derived from consumer expenditure surveys in 2000, 1995, 1990, 1985, 1975 and earlier. Given processing delays, the 1958 weights were applied from *May, 1970* to December, 1977; the 1975 weights from *January, 1978* to October, 1987; the 1985 weights from *November, 1987* to July, 1991; the 1990 weights from *August 1991* to December, 1996; the 1995 weights from *January, 1997* to December, 2001; and the 2000 weights from *January, 2002*. We have italicised the first month of application, because it is used in the formula.

[12] Note that the nomenclature "*I*" is used to denote a price index that has not been level-adjusted, and hence is not the official CPI or CPIX aggregate price index.

$$CPI_{[2000,t]} = L_{Jan.2002} \times I_{[2000,t]} \qquad (6)$$

For months prior to January, 2002, the index would be re-based:

$$\frac{I_{[1995,t]}}{\frac{1}{12} \sum_{t=Jan.2000}^{Dec.2000} I_{[1995,t]}} \qquad (7)$$

The resulting index takes the average value of 100 in the year 2000. Thus, as pointed out by Haglund (2000), the links are of the Laspèyres-type rather than strictly Laspèyres indices, since the weight reference period (2000) precedes the period from which the new weights are operational (January, 2002).

The level-adjusted index is equivalent to a computation of a chain index. It links an index given by equation (6) (with 1995 weights and comparison period January, 2002 relative to 1995 prices), the whole rebased to equal 100 in 2000, with a second index given by equation (5), with 2000 weights, relative to 2000 prices and comparison period the current month.

The basic logic of the procedure is simple if one considers the CPI in percentage or log changes. Then, only equation (4) – and its equivalent for each period over which the weights are held constant – will be relevant, since the "level factors" are constant for each of these periods and hence will be eliminated on taking changes. Thus, for February, 2002 to the present,

$$\Delta \log CPI_t = \Delta \log(\sum_i w_{i,2000} \frac{p_{i,t}}{p_{i,2000}}) \qquad (8)$$

For February, 1997 to January, 2002,

$$\Delta \log CPI_t = \Delta \log(\sum_i w_{i,1995} \frac{p_{i,t}}{p_{i,1995}}) \qquad (9)$$

For September, 1991 to January, 1997,

$$\Delta \log CPI_t = \Delta \log(\sum_i w_{i,1990} \frac{p_{i,t}}{p_{i,1990}}) \tag{10}$$

and analogously, for earlier periods.

Since

$$\log \frac{CPI_t}{CPI_{t_0}} = \sum_{s=t_0+1}^{t} \Delta \log CPI_s \tag{11}$$

it is then easy to construct the official CPI relative to any base month by chain-linking the percentage changes back to the base month. This is equivalent to splicing (with "level factors") fixed weight indices based on equation (5), in January, 2002, January, 1997, August, 1991, November, 1987, January, 1978 and April, 1970.

2.3.2   Level factor B

Another variant on the use of a level factor has been suggested by Haglund (communication). The index as from January, 2002, could be computed according to

$$CPI_{[2000,t]} = \sum_i w_{i,2000} \frac{P_{i,t}}{P_{i,2000}} \times 100 \tag{12}$$

and the series prior to January, 2002, could be re-calculated according to

$$\frac{1}{L_{Jan.2002}} \times \frac{I_{[1995,t]}^{Dec.2000}}{\frac{1}{12} \sum_{t=Jan.2000}^{Dec.2000} I_{[1995,t]}} \tag{13}$$

With this variant, equations (8) to (11) hold, as before.

3. Constructing historical data for CPIX ("metropolitan")

CPIX is defined as the headline CPI, excluding interest rates on mortgage bonds (the mortgage cost component of a homebuyer's cost of housing). Clearly, an understanding of the techniques used by *Statistics SA* in constructing CPI is a pre-requisite for the correct construction of historical monthly CPIX data. We assume that Level factor A (Section 2) is used by *Statistics SA*, and we show how to construct CPIX from headline CPI, employing the appropriate weights for each period.

3.1 Construction of CPIX data

The official CPI can be thought of as having just two components, CPIX, and the mortgage cost component, PM. The former can be constructed by "subtracting" the latter from the CPI. We first explain the principle using the level-*unadjusted* aggregate price indices. In practice, the historical CPIX data have to be constructed from the official level-adjusted indices. This means we have to translate expressions for the level-unadjusted series into their equivalents in level-adjusted series.

For the period after January, 2002, the "raw" index (level-unadjusted), *I*, used in the construction of CPI, can be expressed relative to base period prices at 2000, following the logic of equation (4), as

$$I_{[2000,t]} = (1 - w^{MB}_{2000})IX_{[2000,t]} + w^{MB}_{2000} \frac{PM_t}{PM_{2000}} \qquad (14)$$

where w^{MB} is the weight in 2000 of the mortgage cost component, PM; and PM_{2000} is the average value of PM between January and December of 2000. The "raw" index (level-unadjusted), *IX,* used in the construction of CPIX, is defined as follows,

$$IX_{[2000,t]} = \sum_i w^*_{i,2000} \frac{P_{i,t}}{P_{i,2000}} \times 100 \qquad (15)$$

where the summation now *excludes* the mortgage cost component, PM. The definition of *IX* is analogous to that of *I*, as defined in equation (4). The weights $w_{i,2000}^{*}$ are related to the weights $w_{i,2000}$ by the expression $w_{i,2000} = w_{i,2000}^{*}(1 - w_{2000}^{MB})$, for all components i, except the mortgage cost element.

Rearranging equation (14) yields an expression for the *IX* index (i.e. the index from which CPIX will be constructed by application of the appropriate level factor) relative to its base year 2000, for the period after January, 2002,

$$IX_{[2000,t]} = \frac{I_{[2000,t]} - w_{2000}^{MB}\dfrac{PM_t}{PM_{2000}}}{(1 - w_{2000}^{MB})} \qquad (16)$$

Now we have to translate equation (16) into the level-adjusted equivalent. Since $IX = \dfrac{CPIX}{L^{IX}}$ and $I = \dfrac{CPI}{L}$, equation (16) translates into

$$CPIX_{[2000,t]} = L_{Jan.2002}^{IX} \times \left[\frac{CPI_{[2000,t]} \big/ L_{Jan.2002} - w_{2000}^{MB}\left(\dfrac{PM_t}{PM_{2000}}\right)}{(1 - w_{2000}^{MB})} \right] \qquad (16^*)$$

$$= L_{Jan.2002}^{IX} \times IX_{[2000,t]}$$

where the respective level factors for January, 2002 are denoted $L_{Jan.2002}$ and $L_{Jan.2002}^{IX}$. The mortgage bond component, PM, is not level-adjusted, as it is a unit value (see above).

An analogous procedure is followed for all other periods where weights are held constant. The price indices for each separate period of weighting need to be chain-linked together in the *month* when the weights are first applied (see methodology in Section 2). To apply equation (16*) to construct historical CPIX data back to 1970, requires the historical level factors, *L*, for CPI ("metropolitan") used in May, 1970 (1958 weights); January, 1978 (1975 weights); November, 1987 (1985 weights); August 1991 (1990 weights); January, 1997 (1995 weights); and January, 2002 (2000 weights) – five numbers in all. The chain-

linked index then needs to be multiplied by, $L^{IX}_{Jan.2002}$, the level factor for CPIX ("metropolitan") used in January, 2002.

3.2 Data issues

The series for the mortgage cost component is, in principle, available from *Statistics SA*, as it is used in the computation of the housing cost component of total CPI. Obviously this interest rate series is not seasonally adjusted. While the weight (see Table 2) reflects interest as well as capital repayments, the index component reflects interest costs only and is compiled from a survey of the main financial institutions in the mortgage market[13]. By definition, the series is a unit value index, and is therefore not level-adjusted (equivalently, the level factor is equal to 1). The series is shown in Figure 1.

This series is a sub-component of "housing", one of the ten components of the total CPI, and is only available back to 1997 (Table 1). However, from 1997, this series differs little from the predominant monthly percentage rate on new mortgage loans from banks to "dwelling units"[14], as published from 1965 by the South African Reserve Bank. We therefore use the SARB's published series before 1997, splicing to the mortgage cost sub-component of the housing component[15], in January 1997.

Clearly it is preferable to do the construction with seasonally unadjusted data. This is because the seasonal patterns in headline CPI reflect, in part, movements in the mortgage interest component. Hence, subtraction of a seasonally unadjusted mortgage interest component could induce a spurious seasonal pattern in the resulting CPIX. We thus construct CPIX ("metropolitan areas") using seasonally unadjusted series.

As noted above, the implementation of equation (16*) requires historical level factors, L, for CPI ("metropolitan"), and $L^{IX}_{Jan.2002}$, the level factor for CPIX ("metropolitan") used in January, 2002. Unfortunately, *Statistics SA* has no record of the level factors it used before January, 2002 (communication); and has not made available to us the level factors for CPI ("metropolitan") and CPIX ("metropolitan") used in January, 2002 (2000 weights).

---

[13] The key financial institutions are surveyed in the first week of each month, and a short-term interest rate applying to all mortgages at each bank is obtained. A weighted arithmetic average interest rate is computed from these individual rates with weights reflecting the shares of outstanding mortgages at the respective banks.
[14] The South African Reserve Bank code for this variable is code KBP2011M.
[15] The *Statistics SA* code for this series is VPID3111001, and was kindly provided by *Statistics SA*.

In practice, knowledge of the historical level factors is not essential to obtain adequate historical estimates of CPIX, while the current level factors can be approximately deduced.

We first demonstrate that if the mortgage cost component is a unit value (so level adjustment does not apply), we have a functional relationship between the level factors for CPI and CPIX. By analogy with equation (5), the level factor for CPIX in January, 2002, $L^{IX}_{Jan.2002}$, is defined as

$$L^{IX}_{Jan.2002} = \frac{\left[\dfrac{IX_{[1995,Jan.2002]}}{IX_{[1995,2000]}}\right]}{IX_{[2000,Jan.2002]}} \tag{17}$$

Substituting equation (16) into equation (17), yields an expression for $L^{IX}_{Jan.2002}$ in terms of L and other factors:

$$L^{IX}_{Jan.2002} = \frac{\left[\dfrac{CPI_{[1995,\,Jan.2002]}/L_{Jan.1997} \quad -w^{MB}_{1995}\dfrac{PM_{Jan.2002}}{PM_{1995}}}{CPI_{[1995,\,2000]}/L_{Jan.1997} \quad -w^{MB}_{1995}\dfrac{PM_{2000}}{PM_{1995}}}\right]}{\left[\dfrac{CPI_{[2000,\,Jan.2002]}/L_{Jan.2002} \quad -w^{MB}_{2000}\dfrac{PM_{Jan.2002}}{PM_{2000}}}{1-w^{MB}_{2000}}\right]} \tag{18}$$

where $L_{Jan.1997}$ and $L_{Jan.2002}$ are the level factors for CPI in January, 1997 and January, 2002, respectively. Equation (18) thus gives the level factor required to construct CPIX, using equation (16*).

Next we approximate $L_{Jan.2002}$, and hence $L^{IX}_{Jan.2002}$. Choosing a grid of values for L in the range suggested by *Statistics SA* (0.98 to 1.02)[16], we can calculate the corresponding values of $L^{IX}_{Jan.2002}$ (Appendix 1). These values for $L^{IX}_{Jan.2002}$ are used to compute alternative estimates of CPIX. The best approximation for the level factor for CPI is achieved by

minimising the deviation between published CPIX and our estimates of CPIX based on alternative level factors. By comparing our estimates of CPIX[17] with the published value of CPIX, over the January-February, 2002 threshold (when the weights change), we deduce that *Statistics SA's* level factor for CPI must have been around 0.98. However, given rounding errors, we cannot be very precise.

Finally, we demonstrate mathematically and by numerical simulation in Appendix 1 that $L^{IX}_{Jan.2002}$ is insensitive to alternative assumptions about the historical level factor for January, 1997. However, from equation (16*) and its equivalents for earlier periods, it can be seen that historical values of L are relevant for estimating a historical series for CPIX. To gauge the importance of knowledge of the historical level factors for CPI, we analyse the sensitivity of our estimates using a grid of values for the January, 1997 level factor for CPI. We discover that variations over the range 0.98 to 1.02, makes a maximum difference to the 12-month inflation rate of one quarter of one percent, and mostly a smaller difference (see Figure 5). This suggests that assuming CPI level factors of 1 for earlier years will make hardly any difference to the estimates of annual inflation rates, compared to the estimates that would have resulted if we had known the true level factors all the way back. In the long run, it will make a slight difference to 10 or 20-year comparisons of the price level, but such differences are irrelevant to econometric modelling.

The computer code for our computations is shown in Appendix 2.

3.3   Comparison of our constructed series with official data

We now compare our constructed series for CPIX ("metropolitan areas") back to 1970 with the official monthly series, available from *Statistics SA (*back to 1994, only). We also compare our constructed series for CPIX ("metropolitan and urban areas"), using the same technique, with published data from *Statistics SA* back to 1997. In each case the comparison is made using seasonally unadjusted series to avoid the problems mentioned above. Comparisons are made in Figures 3 and 4.

For the period since 1997, the pattern of difference from our constructed measure for CPIX ("metropolitan areas") is similar to the corresponding differences between the measures for CPIX ("metropolitan and urban areas"). Our estimates of the annual CPIX

---

[16] Communication (*Statistics SA*).

("metropolitan") inflation rates between October, 2001 and June, 2002 are between 0.7 and 1 percentage points below the official ones, for reasons that are not clear.[18] Otherwise, from January, 1998 onwards, our estimates are usually quite close to the official ones.[19]

For the period 1994 to 1997, however, and especially in 1994-5, there are sizeable differences between our measure and the CPIX ("metropolitan areas") from *Statistics SA*. Figure 4 shows the annual percentage rates of change of CPIX ("metropolitan areas") from both sources. For example, in January 1995, *Statistics SA's* measure of CPIX ("metropolitan areas") shows an 8.3 percent annual increase, while the mortgage bond interest rate published by the SARB rose by 6.6 percent. This is plainly inconsistent with the annual rise in CPI ("metropolitan areas") of 9.7 percent - since the rise in CPI should not exceed the rises in both its constituents. By contrast, our measure of CPIX ("metropolitan areas") gives an annual increase of 9.9 percent, 1.6 percentage points higher than the CPIX from *Statistics SA*.[20]

We have had sight also of SARB's internal approximation for CPIX ("metropolitan areas"), which uses quarterly data (seasonally adjusted). This measure also deviates from ours. Since this measure plays an important part in SARB's macroeconomic model, which guides policy making, any biases in the measure are unfortunate.

5.   Conclusion

In the absence of an official technical handbook on CPI methodology, this paper has contributed to enhanced transparency by explaining the methodology of CPI construction in South Africa. We use this methodology to present a consistent method using monthly price indices for constructing historical data for CPIX ("metropolitan areas") back to 1970. A long time series in CPIX is required for rigorous forecasting and modelling exercises, such as are carried out by the South African Reserve Bank (SARB) and the National Treasury. Currently

---

[17] To be precise, we compared the average of three-monthly changes in CPIX for February, March and April, 2002.

[18] Our estimates of the annual CPIX ("metropolitan and urban") inflation rates between October, 2001 and June, 2002 are between 0.7 and 0.8 percentage points below the official ones.

[19] However, there appears to be an error in *Statistics SA*'s CPIX ("metropolitan") in the month of April, 1998, of around 0.8 percent, which does not appear in the CPIX ("metropolitan and urban") series.

[20] It should be noted that this would have had no policy significance at the time. However, it does affect current econometric modelling using historical series of CPIX.

historically constructed CPIX data are published by *Statistics SA* only back to 1994, and there are some troubling differences when compared with our constructed series, discussed above.

The importance of a high capacity and well-resourced national statistical agency cannot be overemphasised. Such agencies are responsible for the production of long series of data, which are required by a wide range of econometric models. Economic policy increasingly relies on the predictions and the revelations of such models. This is especially the case for monetary policy under inflation targeting. Yet, statistical agencies are often amongst the first victims of public budgetary cuts, and in many countries, operate with insufficient resources.

Time series models need decades of reliable data to achieve reliable results. Model outcomes are seriously hampered by poor data. There is unfortunately hysteresis, in that past data errors cannot easily be rectified, and can plague modelling attempts for the future. In the view of most technical experts using these data to model the economy, a high-capacity statistical agency is crucial to the making of sound economic policy for future growth, both now and in the years ahead. In fact, this is an important investment in future growth.

Statistics SA has been under-resourced in recent years, and probably for a generation. One symptom of this has been the errors that led to the revision of the consumer price index announced in April 2003, when annual CPI was revised down by as much as 1.9 percent (Stopford, 2003). It was immediately clear that the Reserve Bank had held interest rates too high over the period, resulting in an avoidable loss of output (see Aron and Muellbauer (2002) for evidence on the interest rate effects on output). Wage settlements based on CPI or CPIX were higher than warranted, with an impact on inflation even after the correction. The costs to the economy in output foregone - both because interest rates were held too high, and because of the output loss necessary to offset the inflation induced by excessive wage claims - are likely to have been many times the entire annual budget of *Statistics SA*. Even to the government alone, which paid excessive indexation payments to owners of indexed bonds, and issued new debt at higher yields than those that would have prevailed if the CPI had been measured correctly, the costs of this mistake are likely to have been substantial.

It is possible that the lack of transparent methods of CPI construction in a published technical handbook on CPI methodology may have contributed to the oversight highlighted above. There are several areas that a future handbook on CPI methodology could usefully address. One is to give a clear account of the logic for measuring costs of homeownership. Indeed, the current method of doing so needs to be reconsidered, as it has been in a number of countries where mortgage costs have, in the past, played a role (e.g. in the U.K., Australia

and New Zealand). The current treatment of home-owners' costs can be criticized from several points of view. It measures only the cost of borrowing a given sum of money, not any increase in the price of housing that the given sum of money can buy, as noted by Haglund (2000). Note that the average level of nominal interest rates in the recent past is no higher than it was in the 1980s, while the price of housing has risen along with that of other goods. Hence this treatment of home-owners' costs in the CPI will result in the housing component of the CPI and hence the total CPI increasing *less* in the long-run than CPIX. And it neglects the fact that, in the context of an increasingly liberal mortgage market in South Africa, an increase in nominal interest rates caused by a rise in general inflation, may not have the same cash flow implications as was once the case. Households with significant net equity can now easily refinance and hence stabilize the real cash flow burden of their mortgage debt.

A future handbook on CPI methodology in South Africa should also explain in some detail how the methodology takes into account issues of quality correction and the treatment of new goods, discount outlets and substitutes, as highlighted by the Boskin Commission in the U.S. (Boskin et al, 1996). Boskin argued that then existing methodology in the U.S. had resulted in inflation being overstated by between 0.8 and 1.6 percent (see also Gordon, 2000). If biases of this order of magnitude existed in South Africa, major implications would follow for the measurement of inflation and growth, monetary policy and the attitude of foreign investors to South Africa.[21] In a parallel paper on modelling the ten components of the CPI such as food, vehicles and others, our models point to the possibility that quality adjustment may not have been adequate historically (Aron, Muellbauer and Pretorius, 2004). We would argue that the brief examination of the issues conducted by Haglund (2000), though very useful, is insufficient. By contrast, at the Bureau of Labour Statistics (U.S.), at the European Central Bank and Eurostat, and at the Bank of England and the U.K. Office of National Statistics, amongst very many other reputable institutions, these issues are of ongoing long-term concern, and the subject of intense internal scrutiny and research programmes, together with interaction with academia.

It is notable that no electronic data are currently available for the sub-components of the ten components of the aggregate CPI prior to 1997 (e.g. the mortgage interest component), despite their having been used to construct the CPI in the past (see Table 1 on limited historical data availability). This puts the SARB in a difficult position of having to

---

[21] It should be noted however, that Robert Gordon, a member of the Boskin Commission, and whose book on durable prices, Gordon (1990), was a major influence on the report, has recently argued, Gordon (2003), that the

make crude approximations for historical series where they are used. Such approximations will be used for example in the key monthly forecasting model of components of the CPI, used by the Monetary Policy Committee in its policy deliberations. We regard it as very important that sufficient resources are made available to *Statistics SA* to extract these data, and allow the publication of continuous and reliable historical series for all the main components and sub-components of the targeted CPI and CPIX variables.

It is also relevant to note potential problems arising from seasonal adjustment procedures. We are informed that *Statistics SA* apply seasonal adjustment procedures at the aggregate level to CPI and to CPIX, rather than constructing these from seasonally adjusted components. Though the mortgage cost component should be non-seasonal, seasonal adjustment procedures applied to CPI based on the X-12 or X-11 software will reflect the accidental correlation of movements in mortgage costs with the seasons. This software relies on backward- and forward-looking moving averages of seasonal deviations to calculate seasonal adjustment factors. Applying these procedures to price indices or other current price data in the level (as occurs at *Statistics SA*) rather than the log form, in an economy with historically high inflation rates, is also problematic.[22]

As a final comment on CPI data construction, we regret the unacceptable methodology whereby rebasing the CPI series allows *Statistics SA* (and its predecessor) to throw away decimal points. All the CPI sub-component data suffer from serious rounding errors in the earlier years. Anyone examining data in the 1970s will note that often the same figure applies for much of a whole year on the current base, whereas there was obviously considerable statistical variation at the time. To illustrate, on the current base of 100 in the year 2000, in August and September, 1979, the food price index was 7.6, and in October and November, 1979, it was 7.7. The lack of a second decimal place reduces the accuracy of the data compared with current data by the order of 12-fold. Throwing away this information impoverishes models attempting to capture structural features of the economy, and this practice should cease.

A clear account of the CPI methodologies used in South Africa would have the benefit of allowing the many research questions that follow from these debates to be opened

---

U.S. price indices for shelter or housing costs, had substantially understated inflation, potentially reducing previously claimed overall biases in the CPI.

[22] In the window over which X-11 or X-12 computes its seasonal adjustments, suppose the index is of the order of 50 for early observations and 100 for late observations. If deviations in percentage terms are constant over time, early deviations in each window will appear spuriously small compared with late deviations. This will tend to generate significant data revisions solely due to the seasonal adjustment procedure. Computing seasonal factors in log form makes more sense.

up, not only for further work within *Statistics SA,* but also in the universities, where they could provide fertile and highly policy-relevant research topics.

REFERENCES

Aron, J., J. Muellbauer and C. Pretorius (S.A. Reserve Bank). 2004. "A Framework for Forecasting the Components of the Consumer Price Index: application to South Africa." CSAE Working Paper Series 2004-07, Department of Economics, Oxford University.

Aron, J. and J. Muellbauer. 2002. 'Interest rate effects on output: Evidence from a GDP forecasting model for South Africa', *IMF Staff Papers* 49 (November, *IMF Annual Research Conference*): 185-213 (also http://www.cepr.org/pubs/new-dps/dplist.asp?authorid=131312).

Bhundia, A. 2002. "An empirical investigation of exchange rate pass-through in South Africa." IMF Working Paper 02-165, International Monetary Fund.

Boskin, M.J., E. Dulberger, R. Gordon, Z.Griliches and D. Jorgenson. 1996. "Toward a more accurate measure of the cost of living." Final Report to the Senate Finance Committee, December 4.

Haglund, Mats. 2000. "The South African CPI - the reliability of present practices and some potential for improvements." Mission Report RSASTAT 2000:2, Statistics Sweden, 2000.

Gordon, R. 1990. *The Measurement of Durable Goods Prices*, Chicago, University of Chicago Press.

Gordon, R. 2000. "The Boskin Commission Report and its aftermath." Working Paper 7759, National Bureau of Economic Research.

Gordon, R. 2003. "A Century of Housing Shelter Prices: How Big is the CPI Bias?" Presented at CRIW Conference in Memory of Zvi Griliches, Hard-to-Measure Goods and Services, Bethesda MD, September 19-20, 2003

Leiderman, L. and L. E. O. Svensson. 1995. *Inflation Targets.* London: Centre for Economic Policy Research.

Stopford, J. 2003. "Inaccurate Statistics a Threat to the Economy."Business Day (Johannesburg), South Africa, April 23.

Figure 1: Mortgage interest rate and the mortgage interest component.
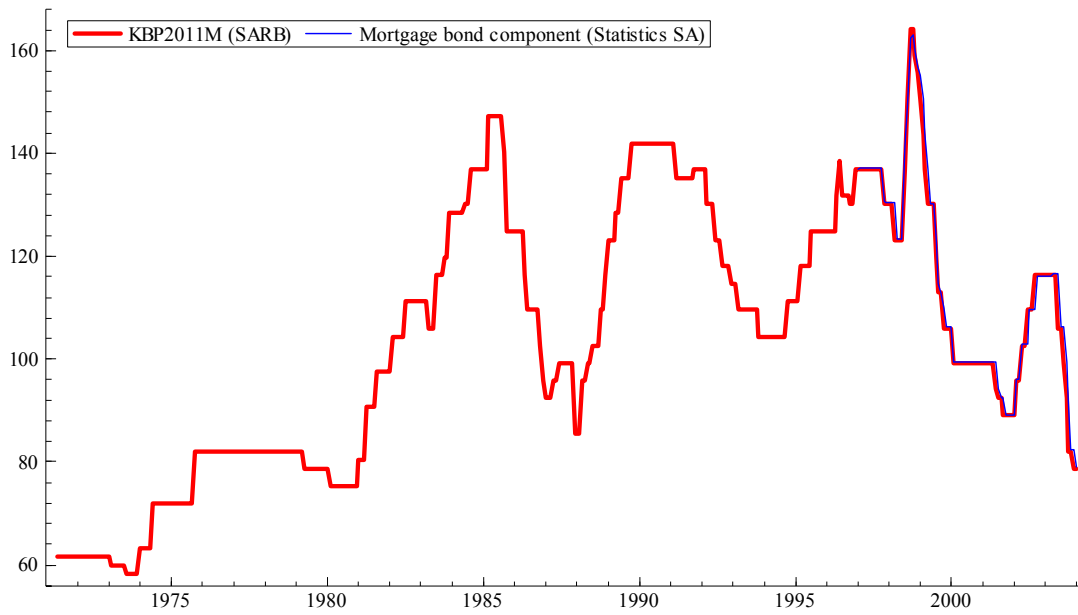
Source: SARB (mortgage interest rate code is KBP2011M); and the mortgage interest component uses data from *Statistics SA* (from 1997 onwards, only).

Figure 2: Annual percentage changes in *seasonally unadjusted* CPI ("metropolitan areas") and *seasonally unadjusted* CPIX ("metropolitan areas")
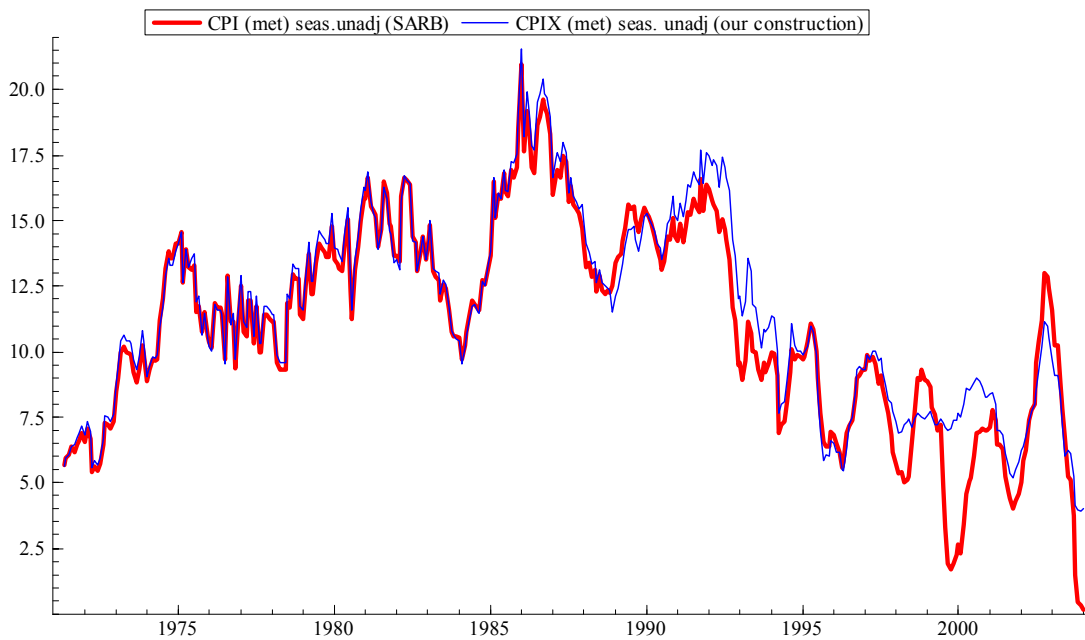
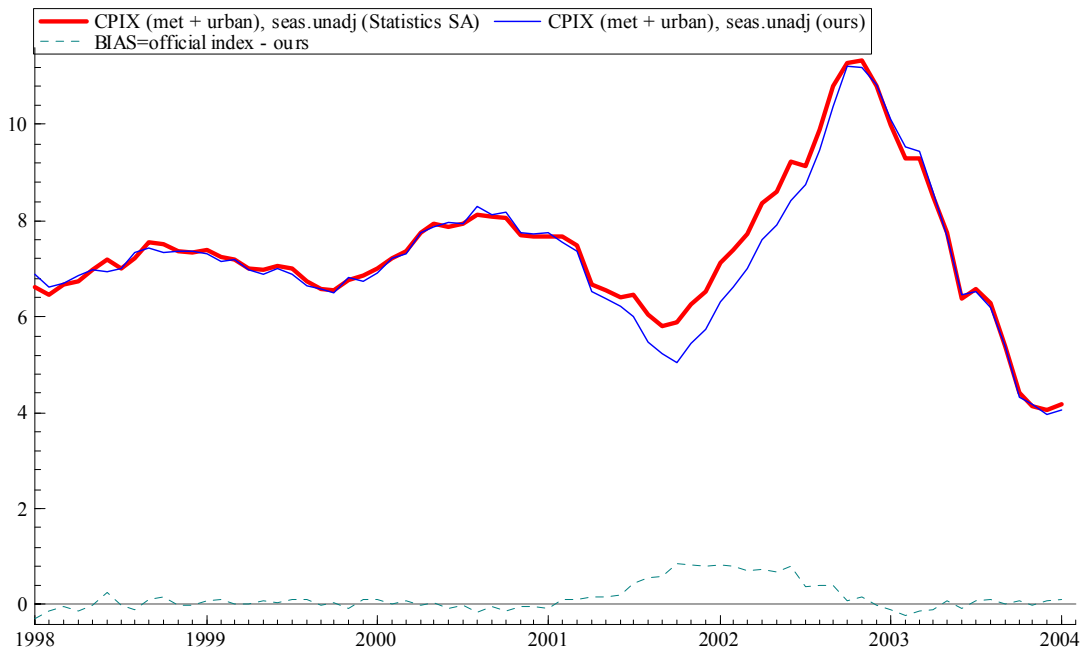**Figure 3 : CPIX ("metropolitan and urban areas"),** *seasonally unadjusted***, annual percentage change.**

Figure 4: CPIX ("metropolitan areas"), *seasonally unadjusted***, annual percentage change.**
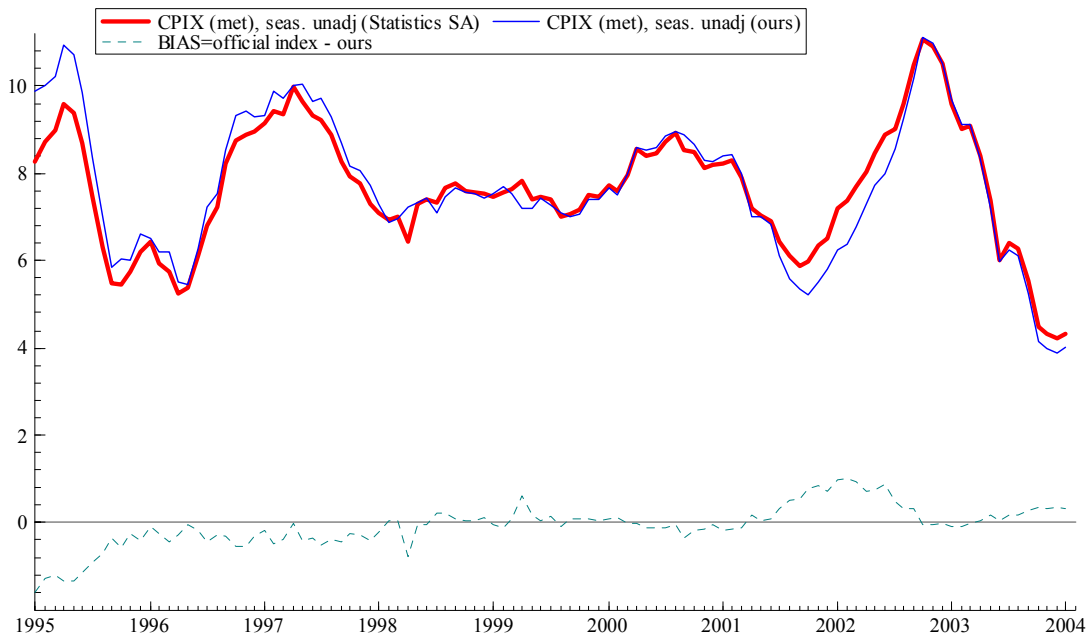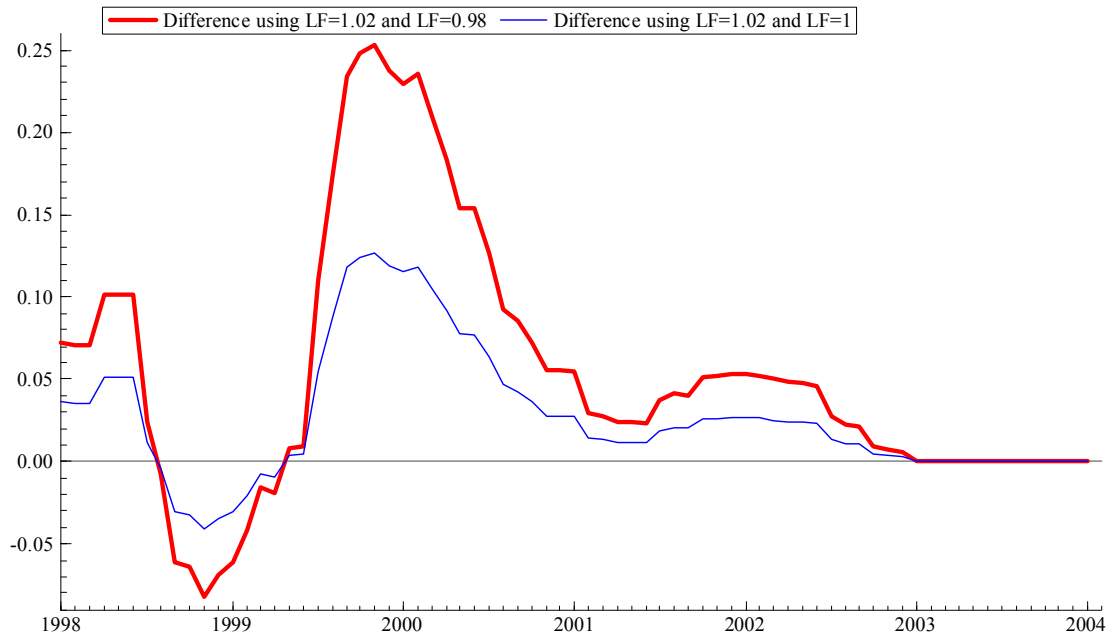
Figure 5: CPIX ("metropolitan areas"), *seasonally unadjusted***, annual percentage change for varying historical level factors for 1997**

Note: Difference between 12-month inflation rates using Level factor =1.02 and using Level factor =0.98; and difference between 12-month inflation rates using Level factor =1.02 and Level factor =1.

Table 1: CPI, CPIX: public availability of (electronic, downloadable) time series data

Coverage	*"Metropolitan areas"*		*"Metropolitan and Urban areas"*	
Source	*SARB*	*Statistics SA*	*SARB*	*Statistics SA*
CPI (Seasonally *adjusted*)	From 1960, monthly. (From 1923 in supplement below.)	From 1986, monthly.	From 1986, monthly.	From 1997, monthly.
CPI (Seasonally *unadjusted*)	NA	From 1970, monthly.	NA	From 1997, monthly.
10 CPI components (Seasonally *adjusted*)	From 1960, monthly.	Not as time series on website.	From 1986, monthly.	Not as time series on website.
10 CPI components (Seasonally *unadjusted*)	NA	Not as time series on website.	NA	Not as time series on website.
Sub-components of these 10 components (Seasonally *unadjusted*) (e.g. mortgage component index)	NA	Not as time series on website. *Apparently only available from 1997, monthly – see text.*	NA	Not as time series on website. (Presumably available from 1997, monthly.)
CPIX (Seasonally *adjusted*)	NA	Not as time series on website.	From 1986, monthly.	Not as time series on website.
CPIX (Seasonally *unadjusted*)	NA	From 1994, monthly.	NA	From 1997 monthly.

Source: South African Reserve Bank Quarterly Bulletins and website, *Statistics SA* website

1. Note that the SARB acquires the price data from *Statistics SA* (and, previously, its predecessor).
2. Until 1994, the SARB did its own seasonal adjustment of the 10 CPI components. From 1994, it received the 10 CPI components seasonally adjusted from *Statistics SA* back to 1994. From September, 2003, it received the 10 CPI components seasonally adjusted from *Statistics SA* back to 1986. Data earlier than this have been seasonally adjusted by the SARB.
3. "Labour, price and other selected economic indicators of South Africa 1923-93." Supplement to the SARB Quarterly Bulletin, September, 1994.

Table 2: Weights for the consumer price index (" metropolitan areas")

Date of Expenditure survey		1958	1975	1985	1990	1995	2000
Weights applied		Apr.70-Dec.77	Jan.78-Oct.87	Nov.87–Jul.91	Aug.91-Dec.96	Jan.97-Dec.01	Jan.02-
Services	Housing	21.6	19.5	22.5	20.5	26.0	24.3
	Mortgage cost	3.61	3.4	9.47	11.51	12.91	11.43
	Transport	4.9	3.7	5.9	4.3	4.3	3.4
	Other	7.1	9.7	11.1	17.3	14.7	15.2
	Total	33.6	32.9	39.5	42.1	45.0	42.9
Goods	Food	23.9	25.5	23.2	19.3	18.8	22.1
	Furniture & equipment	7.8	6.0	4.7	5.5	3.9	2.5
	Clothing & footwear	9.6	8.8	6.0	7.0	4.8	3.2
	Vehicles	6.7	5.6	5.5	5.5	5.3	6.0
	Transport goods	5.0	5.6	5.9	4.6	5.2	5.5
	Beverages & tobacco	4.1	3.8	2.3	2.2	2.1	2.5
	Other	9.3	11.8	12.9	13.8	14.9	15.3
	Total	66.4	67.1	60.5	57.9	55.0	57.1
Total		100.0	100	100.0	100.0	100.0	100.0

Source of figures: South African Reserve Bank, *Statistics SA*

Table 3: Weights for the consumer price index ("metropolitan and urban areas")

Date of Expenditure survey		1995	2000
Weights applied		Jan.97-Dec.01	Jan.02-
Services	Housing	24.3	22.3
	Mortgage cost	11.13	10.32
	Transport	3.6	3.5
	Other	14.7	14.8
	Total	42.7	40.6
Goods	Food	20.3	24.2
	Furniture & equipment	4.3	2.8
	Clothing & footwear	5.1	3.6
	Vehicles	5.3	5.1
	Transport goods	5.2	5.2
	Beverages & tobacco	2.2	2.7
	Other	14.9	15.8
	Total	57.3	59.4
Total		100.0	100.0

Source of figures: South African Reserve Bank, *Statistics SA*

APPENDIX 1: The relationship between level factors for CPI and CPIX

Equation (18) in the text shows that the level factor for CPIX in January 2002, $L_{Jan.2002}^{IX}$, is a function of the level factor for CPI in January, 2002, $L_{Jan.2002}$, as well as the level factor in January 1997, $L_{Jan.1997}$. We now show mathematically that $L_{Jan.2002}^{IX}$ is insensitive to alternative assumptions about $L_{Jan.1997}$. Let us write equation (18) as

$$L_{Jan.2002}^{IX} = \frac{\dfrac{L_{Jan.2002}}{CPI_{[2000,\,Jan.2002]}}\dfrac{CPI_{[1995,\,Jan.2002]}}{CPI_{[1995,\,2000]}}\left(\dfrac{1\text{-}wZ_1/I_1}{1-wZ_2/I_2}\right)}{\left(\dfrac{1\text{-}w^*L_{Jan.2002}Z_3/CPI_{[2000,\,Jan.2002]}}{1\text{-}w^*}\right)} \tag{19}$$

where $w = w_{1995}^{MB}$, $w^* = w_{2000}^{MB}$,

$Z_1 = \dfrac{PM_{Jan.2002}}{PM_{1995}}$, $Z_2 = \dfrac{PM_{2000}}{PM_{2000}}$, $Z_3 = \dfrac{PM_{Jan.2002}}{PM_{2000}}$,

$I_1 = CPI_{[1995,\,Jan.2002]}/L_{Jan.1997}$,

$I_2 = CPI_{[1995,\,2000]}/L_{Jan.1997}$.

Note further that the official level-adjusted CPI has the property that
$\dfrac{CPI_{[1995,\,Jan.2002]}}{CPI_{[2000,\,Jan.2002],}CPI_{[1995,\,2000]}} = 1$.

Also $Z_3 = Z_1/Z_2$, and $Z_3/CPI_{[2000,\,Jan.2002]} = \dfrac{Z_1/I_1}{Z_2/I_2}$.

Thus, equation (19) can also be written as

$$L^{IX}_{Jan.2002} = \cfrac{L_{Jan.2002}\left(\cfrac{1-wZ_1/I_1}{1-wZ_2/I_2}\right)}{\left(\cfrac{1-w^*L_{Jan.2002}\left(\cfrac{Z_1/I_1}{Z_2/I_2}\right)}{1-w^*}\right)} \qquad (20)$$

Only the part of the numerator, $\left(\dfrac{1-wZ_1/I_1}{1-wZ_2/I_2}\right)$, depends on the January, 1997 level factor,

$L_{Jan.1997}$, via the I_1 and I_2 terms. The weight, w, is around 0.12 (see Table 2), making the numerator, and hence $L^{IX}_{Jan.2002}$, relatively insensitive to variations in $L_{Jan.1997}$.

Equation (20) shows that $L^{IX}_{Jan.2002}$ is approximately proportional to $L_{Jan.2002}$. Thus, $L^{IX}_{Jan.2002}$ is sensitive to alternative assumptions about $L_{Jan.2002}$; but it is insensitive to alternative assumptions about $L_{Jan.1997}$. We demonstrate this in the grid below (we are calculating LIX as in Appendix 2, for different values of $L_{Jan.2002}$ and $L_{Jan.1997}$). Column 3 shows that whether $L_{Jan.1997}$ is 0.98 or 1.02 makes hardly any difference to $L^{IX}_{Jan.2002}$. However, variations in $L^{IX}_{Jan.2002}$ are approximately proportional to variations in $L_{Jan.2002}$.

Table: Empirical grid for level factors

$L_{Jan.2002}$	$L_{Jan.1997}$	$L^{IX}_{Jan.2002}$
0.98	0.98	0.96985
0.99		0.98075
1.00		0.99168
1.01		1.00263
1.02		1.01360
0.98	1.02	0.97048
0.99		0.98139
1.00		0.99233
1.01		1.00329
1.02		1.01427

APPENDIX 2: Computer code for constructing CPIX ("metropolitan").

This is the computer code (TSP4.5) for constructing CPIX ("metropolitan").

```
?APPLY EQUATIONS (16*) AND (18) FROM THE TEXT ABOVE,

?DEFINITIONS:

?MIC=MORTGAGE INTEREST COMPONENT FROM STATISTICS SA
?CPI=HEADLINE CPI, SEASONALLY UNADJUSTED

?MSD IS A ROUTINE CREATING MEANS
?CPIBAV75 IS CPI BASE 1975, AND SIMILARLY FOR OTHER BASE YEARS
?MICBAV70 IS MIC BASE 1975, AND SIMILARLY FOR OTHER BASE YEARS
?WEIGHTS ARE APPLIED FOR THE MORTGAGE INTEREST COMPONENT FROM TABLE 1 ABOVE

?ASSUME HISTORICAL LEVEL FACTORS FOR CPI ARE 1 (SEE TEXT)
?ASSUME THE LEVEL FACTOR FOR CPI IN JANUARY, 2002 IS 0.98 (SEE TEXT)
?SEE TEXT TO EXPLAIN APPROPRIATE LINKAGE OF YEARS (FOLLOWS HAGLUND (2000))

?1958 weights were applied from April, 1970 to December, 1977
?(NEAREST WE CAN GET IS 1960 FOR CPI AND 1965 FOR THE MORTGAGE INTEREST
?RATE, SO WE USE 1965 VALUES FOR EACH)
FREQ M; SMPL 1965:1 1965:12;
LIST HCOMP1  MIC CPI; DOT HCOMP1; MSD .; SET .BAV65 =@MEAN; ENDDOT;
FREQ M; SMPL 1970:5 1978:1;
L1970=1; ?Level Factor for CPI in May, 1970
CPIX1=(CPI/(CPIBAV65*L1978)-(3.61/100)*MIC/MICBAV65)/(1-(3.61/100));
?uses equivalent of equation (16*) defined for this period

?1975 weights were applied from January, 1978 to October, 1987
FREQ M; SMPL 1975:1 1975:12;
LIST HCOMP2  MIC CPI; DOT HCOMP2; MSD .; SET .BAV75 =@MEAN; ENDDOT;
FREQ M; SMPL 1978:1 1987:11;
L1978=1; ?Level Factor for CPI in May, 1978
CPIX2=(CPI/(CPIBAV75*L1978)-(3.4/100)*MIC/MICBAV75)/(1-(3.4/100));

?1985 weights were applied from November, 1987 to July, 1991
FREQ M; SMPL 1985:1 1985:12;
LIST HCOMP3  MIC CPI; DOT HCOMP3; MSD .; SET .BAV85 =@MEAN; ENDDOT;
FREQ M; SMPL 1987:11 1991:8;
L1987=1; ?Level Factor for CPI in January, 1987
CPIX3=(CPI/(CPIBAV85*L1987)-(9.47/100)*MIC/MICBAV85)/(1-(9.47/100));

?1990 weights were applied from August 1991 to December, 1996
FREQ M; SMPL 1990:1 1990:12;
LIST HCOMP4  MIC CPI; DOT HCOMP4; MSD .; SET .BAV90 =@MEAN; ENDDOT;
FREQ M; SMPL 1991:8 1997:1;
L1991=1; ?Level Factor for CPI in August, 1991
CPIX4=(CPI/(CPIBAV90*L1991)-(11.51/100)*MIC/MICBAV90)/(1-(11.51/100));

?1995 weights were applied from January, 1997 to December, 2001
FREQ M; SMPL 1995:1 1995:12;
LIST HCOMP5  MIC CPI; DOT HCOMP5; MSD .; SET .BAV95 =@MEAN; ENDDOT;
FREQ M; SMPL 1997:1 2002:1;
L1997=1; ?Level Factor for CPI in January, 1997
CPIX5=(CPI/(CPIBAV95*L1997)-(12.91/100)*MIC/MICBAV95)/(1-(12.91/100));

?2000 weights were applied from January, 2002
```

```
FREQ M; SMPL 2000:1 2000:12;
LIST HCOMP6  MIC CPI; DOT HCOMP6; MSD .; SET .BAV00 =@MEAN; ENDDOT;
FREQ M; SMPL 2002:1 2004:1;
```
L2002=0.98; ?Level Factor for CPI in January, 2002
```
CPIX6=(CPI/(CPIBAV00*L2002)-(11.43/100)*MIC/MICBAV00)/(1-(11.43/100)));
```

?SPLICING FACTORS
```
FREQ M; SMPL 1978:1 1978:1;   SET  SP21=CPIX2/CPIX1;
FREQ M; SMPL 1987:11 1987:11; SET SP32=CPIX3/CPIX2;
FREQ M; SMPL 1991:8 1991:8;   SET  SP43=CPIX4/CPIX3;
FREQ M; SMPL 1997:1 1997:1;   SET  SP54=CPIX5/CPIX4;
FREQ M; SMPL 2002:1 2002:1;   SET  SP65=CPIX6/CPIX5;
```

?CHAIN CPIXmC ("monthly, constructed")
```
FREQ M; SMPL 1970:5 1977:12; CPIXmC=CPIX1*SP21*SP32*SP43*SP54*SP65;
FREQ M; SMPL 1978:1 1987:10; CPIXmC=CPIX2*SP32*SP43*SP54*SP65;
FREQ M; SMPL 1987:11 1991:7; CPIXmC=CPIX3*SP43*SP54*SP65;
FREQ M; SMPL 1991:8 1996:12; CPIXmC=CPIX4*SP54*SP65;
FREQ M; SMPL 1997:1 2001:12; CPIXmC=CPIX5*SP65;
FREQ M; SMPL 2002:1 2003:6;  CPIXmC=CPIX6;
```

?APPLY THE LEVEL FACTOR seas unadj data
```
FREQ m; SMPL 2002:1 2002:1; msd CPIX5; set CPIX5JAN02 =@mean;
FREQ m; SMPL 2000:1 2000:12; msd CPIX5; set CPIX5AV00 =@mean;
FREQ m; SMPL 2002:1 2002:1; msd CPIX6; set CPIX6JAN02 =@mean;
```
SET LIX=(CPIX5JAN02/CPIX5AV00)/CPIX6JAN02;
```
FREQ m; SMPL 1970:5 2004:1; CPIXMCL=CPIXMC*LIX*100;
```