

**Bananas and Petrol: Further Evidence on the Forecasting Accuracy of the ABS
'Headline' and 'Underlying' Rates of Inflation**

Liam J. A. Lenten *

Department of Economics and Finance

La Trobe University

Paper Prepared for Seminar Series, Lund University,

Department of Economics, Sweden, 8 June 2007

Abstract

In the light of the topical nature of 'bananas and petrol' being blamed for driving much of the inflationary pressures in Australia during 2006, the 'headline' and 'underlying' rates of inflation are scrutinised in terms of forecasting accuracy. A general structural time-series modelling strategy is then employed to estimate models for both types of Consumer Price Index (CPI) measures. From this, out-of-sample forecasts are generated from the various models. The underlying forecasts are subsequently adjusted to facilitate comparison to the headline forecasts. Having completed that, the Ashley, Granger and Schmalensee (1980) *AGS* test is performed to find out if there is a statistically significant difference between the root mean square errors (*RMSEs*) of the two models. The results lend weight to the recent findings of Song (2005), insofar that forecasting models using underlying rates are not systematically inferior to those based on the headline rate of inflation. In fact, strong evidence is found that underlying measures produce superior forecasts.

JEL Classification Number: C53, E17

Keywords: Forecasting, Unobserved Components, Inflation, Measurement

*Contact details: Department of Economics and Finance, La Trobe University, Victoria, 3086, AUSTRALIA. Tel: + 61 3 9479 3607, Fax: + 61 3 9479 1654. E-mail: l.lenten@latrobe.edu.au

1. Introduction

In recent years, the move towards inflation targeting by central banks of numerous countries has meant that the inflation rate is watched much more closely by those relevant central banks, the Reserve Bank of Australia (RBA) among them. In their attempts to maintain the inflation rate within their desired bank of between 2 and 3 per cent per annum, they are well aware of the vagaries of the benchmark Consumer Price Index (CPI), often referred to as the ‘headline’ rate of inflation, insofar that it often captures temporary or even one-off effects arising from unusual circumstances. To this end, they (in conjunction with the Australian Bureau of Statistics) produce various alternative measures of inflation that are designed to more accurately reflect the true rate of change of the overall price level. These ‘underlying’ measures are also used in conjunction with the headline rate in matters of interest rate setting by the RBA Board in their monthly meetings.

The compilation and reporting of these underlying measures of inflation took an interesting turn in the second quarter of 2006, when the quarterly headline inflation rate jumped to 1.6 per cent, taking the annual rate to 3.9 per cent. With public and media speculation increasing regarding the possibility of rising interest rates at a politically sensitive time, Prime Minister Howard and Treasurer Costello were quick to blame the CPI ‘aberration’ on significant price hikes in petrol and bananas.¹ Since the former was due to a bout of insurgency activity in the Middle East, and the latter mostly due to the effects of Cyclone Larry in North Queensland, these effects were

¹ Fruit prices comprise 0.95 per cent of the overall index, of which bananas is only one item, however, the price of bananas increased by more than a factor of five in the second quarter of 2006, and in many cases were simply unavailable. The concurrent petrol price hikes were modest in comparison, though this item has more weighting in CPI calculations, at 3.78 per cent. For a full list of CPI weightings, see Australian Bureau of Statistics (2005).

assessed as being temporary. Nevertheless, Howard and Costello were very keen to soothe the concerns of heavily-indebted voters regarding interest rates.

In doing so, they often made reference to underlying inflation. Their argument was that since in the short-to-medium term, prices for bananas and petrol should return to their supposed normal levels, the increase in the CPI should return to the 2-3 per cent band within a few quarters. Meanwhile, the underlying rate (adjusting for these temporary effects), they said, was a more reflective measure of inflation, and should be given more attention. Political science may suggest that this argument is asymmetric in nature, since political leaders would presumably not be nearly as concerned if the headline CPI rate were below the lower bound of the target band.

Nevertheless, should the headline rate be accepted as the 'true' or at least representative rate of inflation (rightly or wrongly), then the simple fact remains that underlying rates involve manipulating (or at least filtering) the composition of components that are used to calculate the headline rate. When it comes to the issue of the manipulation of the series of any economic variable, there is a healthy debate on whether such manipulation distorts the underlying time-series properties of the data. In fact, the debate on other forms of statistical agency manipulation, such as seasonal adjustment (for example) centres on not one, but numerous issues. Among these issues is the question of whether forecasting accuracy is affected adversely by the use of manipulated data to estimate the underlying forecasting model. This issue is just as relevant to the headline versus underlying inflation rate problem as it is for the seasonal adjustment problem.

This empirical study aims to go some way towards addressing the issue raised above. To achieve this aim, this study necessitates the modelling and forecasting of headline inflation and various measures of underlying inflation. The chosen methodology for this purpose is the structural time-series approach of Harvey (1989), which is model-based. Here, out-of-sample forecasts (both one-period ahead and multi-period ahead) are generated from the estimated general model such that the underlying inflation forecasts then have the volatility built back into them (via a simple process) for the purposes of comparability. Following that, numerous forecasting accuracy criteria (in terms of magnitude of errors) are evaluated to determine the relative suitability of headline and underlying inflation data to forecast the underlying model.

This paper proceeds in the following manner: next, a short recount of the literature on headline and underlying inflation, as well as the effect of general transformation of data by statistical agencies on forecasting accuracy, is provided. Following that, the specification of the general model that can then be used for the purpose of model-based seasonal adjustment is presented. This is the structural time series model suggested by Harvey. Subsequently, the results of the relative forecasting accuracy of the series are presented, along with a brief discussion. The paper concludes with some closing remarks.

2. Literature Review on Statistical Manipulation and Forecasting Accuracy

Similarly with respect to other aspects of the debate over manipulation of economic time series, the literature on forecasting possesses no shortage of contributions in relation to the effect of seasonal adjustment. Among the initial contributions were those of Makridakis and Hibon (1979) and Plosser (1979), from which countless other

papers have followed. However, the same stream of research has, by and large, failed to extend to alternative popular methods of other forms of statistical manipulation of time-series data by statistical agencies, such as the Henderson trend methodology used commonly by the ABS.

However, much of the literature pertaining to headline and underlying measures of inflation is quite contemporary, which is unsurprising when it is remembered that inflation targeting is still a relatively recent phenomenon. The early work concentrated mostly on evaluating such measures of underlying inflation. More recent examples along this theme include Vega and Wynne (2003) on Euro-area data, finding that the underlying measures did not result in an improved ability to pick up an impending rise in trend inflation. Dixon and Lim (2004) meanwhile, in an Australian context, concluded that none of the underlying measures were satisfactory, largely on the basis that they improperly excluded useful information.

However, much of the literature is more sanguine about the use of underlying inflation in central bank decision making. Roberts (2005) found the underlying inflation measures to perform reasonably well according to the set criteria, and added that the measures can ‘...add value to the analysis’ (p.28) of trends in inflation. Furthermore, Brischetto and Richards (2006) affirmed this, finding that trimmed mean measures of inflation specifically outperformed both the headline and item exclusion methods, in terms of a signal-to-noise ratio approach.

In recent times, similar exercises in looking at forecasting accuracy have been evident. Stavrev (2006) used a generalised factor dynamic model for Euro-area data.

The forecasts from this method outperformed the trimmed mean, which in turn outperformed the permanent exclusion method. However, all of these individual methods were outperformed by a simple composite average of the three, hence reducing bias without a great loss of efficiency. Meanwhile, Camba-Mendez and Kapetanios (2005) suggest that dynamic factor methods of underlying inflation are superior to traditional methods with respect to forecasting, since they have certain vagaries.

Of most relevance to this study, however, is the application to Australian data by Song (2005), in which a more mechanical procedure than that used here is employed. Song found that the various measures of underlying inflation had very similar forecasting ability to the headline measure, and that there was some (albeit weak) evidence that one of the measures outperformed the headline rate in terms of forecasting. It is this study that the current paper intends to build on, to see whether a model-based approach to forecasting provides similar results to those obtained previously.

3. Specification and Estimation of Harvey's Structural Time Series Model

This section relates primarily to the econometric methodology utilised for estimating the various forecasting models. The structural time-series model of Harvey (1989), based on the Kalman filter, is called 'structural' in this context because each time series is modelled as a set of components that are not observable directly, however, they still do have a direct economic interpretation. These components can then be aggregated additively to reproduce the actual series. Within the representation of Harvey (1989), the time series, y_t , can be expressed by the equation

$$y_t = \mu_t + \phi_t + \gamma_t + \tau x_t + \varepsilon_t \quad (1)$$

whereby μ_t is the trend component, ϕ_t is the cyclical component, γ_t is the seasonal component and ε_t is a white noise random component. Furthermore, x_t is a dummy variable relating to the introduction of the Goods and Services Tax (GST); with a sensitivity parameter, τ , such that

$$x_t = \begin{cases} 0 & \forall t = 1986Q4 - 2000Q2 \text{ inclusive} \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

Equation (1) is restricted by $\text{cov}(\mu_t, \phi_t, \gamma_t) = 0$ and $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$, meaning that the components cannot be correlated and the error term must be normally and independently distributed.

The trend component, representing the long-term movement of the series, is written in its most general form as a stochastic linear process, hence

$$\mu_t = \mu_{t-1} + \lambda_{t-1} + \eta_t \quad (3)$$

$$\lambda_t = \lambda_{t-1} + \zeta_t \quad (4)$$

Equations (3) and (4) are subject to the restrictions: $\eta_t \sim NID(0, \sigma_\eta^2)$ and $\zeta_t \sim NID(0, \sigma_\zeta^2)$. The specification in equation (3) reveals that μ_t (known as the ‘level’) follows a random walk with a drift factor, λ_t (called the ‘slope’), which itself follows a first-order autoregressive process (equation 4).² Within the context of equations (3) and (4), the μ_t process collapses to a random walk with drift factor if $\sigma_\zeta^2 = 0$, and even further to a random walk with no drift if $\lambda_t = 0$ also, and ultimately

² These components correspond to the intercept and slope (respectively) of a conventional regression.

to a deterministic linear trend if $\sigma_\eta^2 = 0$. On the other hand, if $\sigma_\eta^2 = 0$ but $\sigma_\zeta^2 \neq 0$, then μ_t follows a smoothly-changing process.

Also of importance is the form of the cyclical component, which is assumed to follow a stationary linear process, and is represented as thus

$$\phi_t = a \cos \theta_t + b \sin \theta_t \quad (5)$$

For the purposes of allowing the cycle to be stochastic, a and b (the sensitivity parameters) are allowed to change over time. Establishing a recursion for constructing ϕ prior to introducing the stochastic elements ensures that there is no discontinuation of the series. By introducing disturbances and a damping factor, the following is obtained

$$\phi_t = \rho \cos \theta \phi_{t-1} + \rho \sin \theta \phi_{t-1}^* + \omega_t \quad (6)$$

$$\phi_t^* = -\rho \sin \theta \phi_{t-1} + \rho \cos \theta \phi_{t-1}^* + \omega_t^* \quad (7)$$

Within this representation, ϕ_t^* appears by construction, and $\omega_t \sim IID(0, \sigma_\omega^2)$, $\omega_t^* \sim IID(0, \sigma_{\omega^*}^2)$ are requirements of the model. Here, $0 < \rho < 1$ is defined as the damping factor on the amplitude and $0 < \theta < \pi$ is the cycle frequency.

Additionally, of just as much consequence for this study is the form of the seasonal component. Of the number of specifications the seasonal component can take, the one employed in this study is the trigonometric specification (see Harvey, 1989, Chapter 2; Koopman et al., 2000). This specification is chosen because it allows for smooth changes in the seasonals. Hence

$$\gamma_t = \sum_{j=1}^{s/2} \gamma_{j,t} \quad (8)$$

where $\gamma_{j,t}$ is specified by the following equations

$$\gamma_{j,t} = \gamma_{j,t-1} \cos \lambda_j + \gamma_{j,t-1}^* \sin \lambda_j + \kappa_{j,t} \quad (9)$$

$$\gamma_{j,t}^* = -\gamma_{j,t-1} \sin \lambda_j + \gamma_{j,t-1}^* \cos \lambda_j + \kappa_{j,t}^* \quad (10)$$

where $j = 1, \dots, s/2 - 1$, $\lambda_j = 2\pi j / s$ and

$$\gamma_{j,t} = -\gamma_{j,t-1} + \kappa_{j,t}, \quad j = s/2 \quad (11)$$

where $\kappa_{j,t} \sim NID(0, \sigma_{\kappa}^2)$ and $\kappa_{j,t}^* \sim NID(0, \sigma_{\kappa^*}^2)$.

The degree to which the various component evolve over time depends on the values of the variances σ_{η}^2 , σ_{ζ}^2 , σ_{ω}^2 , σ_{κ}^2 and σ_{ε}^2 , which are known as ‘hyperparameters’.

To make numerical estimation easier, it is assumed that $\sigma_{\omega^*}^2 = \sigma_{\omega}^2$ and $\sigma_{\kappa^*}^2 = \sigma_{\kappa}^2$.

These hyperparameters and the components can be estimated via maximum likelihood once the model has been written in a state space representation of equation (1).

In determining the optimal model for the various measures of inflation, a very general methodology is executed – that is, the most general model is estimated for all inflation measures, so that the model is directly comparable. Specifically, the version of the model that is estimated is one that includes a stochastic trend (stochastic level and stochastic slope), a trigonometric seasonal, the maximum three cycles and an irregular component. This combination corresponds to the default settings in the modelling software, *STAMP 6.0* (Koopman, Harvey, Doornik and Shephard, 2000). Since this model produces excellent fit, diagnostic and structural stability results, there is no need to experiment with alternative specifications. Further, imposing a general model on all measures of inflation ensures that there is no intention of implicitly ‘favouring’

some measures over others due to the fitting of a model that is more appropriate for any single measure.

4. Summary of Results

In examining the comparison of ‘headline’ and ‘underlying’ measures of inflation for forecasting accuracy, all of the empirical results presented in this study are produced on the basis of the following four Consumer Price Index (CPI) series. The ‘headline’ rate is the aggregate CPI measure for all 90 expense classes (in 11 different groups) and all cities (henceforth referred to as CPI). The first ‘underlying’ measure is the analogous CPI measure ‘Excluding Volatile Items’ (EVI), whereby the items under scrutiny (namely the group ‘fruit and vegetables’ and expense class ‘automotive fuel’, which are inherently volatile) are removed from the calculations.³ The second underlying measure is the weighted median measure (WMD), whereby the 90 expense classes are ranked from lowest to highest percentage change for that quarter, and the percentage change from the median class (weighted by respective influence of each class on the overall index) chosen. The final measure is the trimmed median measure (TMN), in which the highest and lowest 15 per cent of expense classes are eradicated, and the (unweighted) mean of the remaining 70 per cent of classes calculated. The sample period for all four series extends from 1987Q1-2006Q1, resulting in a total of 77 quarterly observations. The quarter-on-quarter inflation rates are then used to produce a price index, which is assigned an arbitrary value of 1 for 1986Q4, and is retained in levels rather than converted to logs – this means that it is actually the price level that is being forecasted, rather than the inflation rate. All series are obtained electronically from *RBA Bulletin Database* via *DX Database 4.0*.

³ This corresponds to 5.89 per cent of the total CPI weighting.

Figure 1 illustrates the four price level series that are obtained from the headline series and each of the underlying measures, with the legend at the bottom to distinguish the various series. The structural break in both CPI and EVI arising from the GST in 2000Q3 is clearly visible. While it may not be evident from a visual inspection of the price level series, the underlying inflation measures reduce substantially the standard deviation in the quarter-to-quarter inflation rate. Specifically, the EVI, WMD and TMN measures reduce the standard deviation in the CPI series by 19.9, 21.2 and 18.9 per cent, respectively.

Since the model is estimated in levels, it is implied that the components are additive rather than multiplicative. The estimation is based on all four series over the entire sample period less the forecast period. The forecast period extends from 2001Q3 to the end of the sample (2006Q1), resulting in a total of 18 quarterly point forecasts. Two sets of forecasts are then generated for each inflation measure: the first set comprises of one-period ahead forecasts, whereby the model is estimated over a sample period including 2001Q2. The process is then repeated by including one further observation in the estimation sample to generate a forecast for the following quarter. The second set comprises multi-period ahead forecasts, in which the model is estimated up to 2001Q2, and then used to generate dynamic forecasts. Multiple forecast horizons would have been preferable; however, forecasting from prior to 2000Q3 would unfairly discriminate against the CPI and EVI series due to the GST-induced structural break. Furthermore, forecast horizons much after 2001Q3 have limited usefulness due to the small number of point forecasts.

Table 1 outlines the full results for the univariate estimation of equation (1) for all four measures of inflation. The top section displays the final state vector estimates of the various components, along with the standard errors of the level and slope components of both variables. The second section reveals the value of the hyperparameters, in addition to their respective q -ratios. The final section of table 1 shows both the goodness-of-fit and diagnostic results.

For the purpose of the following explanation of the test statistics used in table 1, denote the residuals from equation (1) as ε_t where $t = d + 1, \dots, T$ (T represents the sample size). Given this framework, the goodness-of-fit is measured by the standard unadjusted coefficient of determination, R^2 , which can be calculated as

$$R^2 = 1 - \frac{(T-d)\tilde{\sigma}^2}{\sum_{t=1}^T (\varepsilon_t - \bar{\varepsilon})^2} \quad (12)$$

Also reported is the modified coefficient of determination, R_d^2 , which is calculated as

$$R_d^2 = 1 - \frac{(T-d)\tilde{\sigma}^2}{\sum_{t=2}^T (\Delta x_t - \overline{\Delta x})^2} \quad (13)$$

where $\overline{\Delta x}$ is the sample mean of the first difference of x_t . The reason for the use of R_d^2 is that for data where x_t exhibits trend movements, it is more appropriate to compare the prediction error variance with the variance of x_t , making it preferable to the conventional R^2 (see Koopman, Harvey, Doornik and Shephard, 1999). Also reported is the standard error of the estimated equation, $\tilde{\sigma}$, calculated as the square root of the one-step-ahead prediction error variance.

The diagnostic tests include the Bowman and Shenton (1975) test for normality, N , distributed as $\chi^2(2)$. The N test is based on the joint departures of the third and fourth moments from their predicted values under normality. The third and fourth moments represent measures of skewness (S) and kurtosis (K) respectively (for a normal distribution they have values 0 and 3). The test statistic is calculated as

$$N = \left(\frac{T-d}{6}\right)S^2 + \left(\frac{T-d}{24}\right)((K)^2 - 3)^2 \quad (14)$$

The presence of heteroscedasticity is also tested for, by the $H(h)$ test, which follows an $F(h, h)$ distribution (henceforth denoted as H). H is calculated as the ratio of the squares of the last h residuals to the squares of the first h residuals, where h is the closest integer to $T/3$. Alternatively, H can be expressed more formally as

$$H(h) = \frac{\sum_{t=T-h+1}^T \varepsilon_t^2}{\sum_{t=d+1}^{d+h} \varepsilon_t^2} \quad (15)$$

A high (low) H value implies an increase (decrease) in the variance over time. Also reported is the Ljung and Box (1978) Q statistic for serial correlation based (in this model) on the first 6 autocorrelation coefficients, distributed as $\chi^2(n+1-k)$, where n is the number of autocorrelation coefficients, and k is the number of estimated parameters. Specifically, it is calculated as

$$Q(n, q) = T(T+2) \sum_{j=1}^n \frac{r_j^2}{T-n} \quad (16)$$

where r_j is the autocorrelation coefficient of order j . The most conventional serial correlation test is also reported, the Durbin-Watson (DW) statistic, which can be defined as the ratio of the sum of squared differences in successive residuals to the residual sum of squares, or alternatively

$$DW = \frac{\sum_{t=2}^T (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^T \varepsilon_t^2} \quad (17)$$

Overall, the results indicate a group of four extremely well-specified models. The $\tilde{\sigma}$ (R^2) stat is lowest (highest) for WMD and TMN, however the ordering for the R_d^2 is not the same, with the estimated components explaining 70 and 80 per cent of the variation in the observed series. Furthermore, the model passes all of the diagnostic tests without exception – both of the serial correlation tests, as well as the *NO* and *H* tests, demonstrating model validity.

In order to test for structural breaks, both the predictive failure (*PRF*) and *CUSUM* (*CUS*) statistics are used to determine how well the model predicts out-of-sample. Both test statistics are ultimately calculated from the forecast errors. For a formal representation of these test statistics, assume that there are L out-of-sample forecasts of x_t , denoted by $t = T + 1, \dots, T + \ell$, resulting in the calculation of forecast errors of w_t for each $t = T + 1, \dots, T + \ell$. Within this framework, the *PF* test statistic, distributed as $\chi^2(\ell)$, is calculated as

$$(PF)_t = \sum_{m=1}^{\ell} w_{T+m}^2 \quad (18)$$

whereas the *CUSUM* test statistic, distributed as $t(T - \ell - k)$, is determined by the relation

$$(CUSUM)_t = \ell^{\frac{1}{2}} \sum_{m=1}^{\ell} w_{T+m} \quad (19)$$

The results of the *PF* and *CUSUM* tests for the four models are displayed in table 1. All three models pass the tests for structural stability, a good result considering the difficult nature of the out-of-sample period as discussed previously.

Finally, a comparison a overall model validity can be attained by looking at the Akaike (1973, 1977) information criterion, *AIC* and the Schwarz (1978) Bayesian criterion, *SBC*. The results at the bottom of table 1 indicate that the chosen specification is comparably suitable for each of the four models. However, it must be remembered that the models are not entirely comparable. Nevertheless, this finding is still useful, even if albeit at a casual level.

Figures 2 and 3 reveal the one-period and multi-period ahead forecasts respectively, compared to the actual values, based on both the headline and underlying rates. The raw forecast errors derived from the headline rate (ξ_t^H) are obtained simply by subtracting the raw forecast, \hat{y}_t^H , from the realised value, y_t^H , as in equation (20). However, the raw forecast errors for the trended series (same notation again, but with a *U* superscript instead) need to be adjusted by building back in the volatility that has previously been removed by the various procedures used to calculate the underlying rate. If this correction is not made, then the two sets of forecasts are not comparable directly, because $y_t^H \neq y_t^U$ on the left-hand side of equation (1) in the separate models for each variable. In summary, the two sets of forecasts are derived according to the formulae

$$\xi_t^H = y_t^H - \hat{y}_t^H \quad (20)$$

$$\xi_t^{Us} = y_t^U - \left[\frac{\hat{y}_t^U}{y_t^U / y_t^H} \right] \quad (21)$$

From a quick visual inspection of figures 2 and 3, it does appear to be the case that the underlying rate price level series systematically underperforms the equivalent headline rate data for forecasting purposes, particularly so in the latter case. Nevertheless, we look towards more formal statistical evidence for further indication.

The selected quantitative measures of forecasting accuracy are displayed in table 2. The measures are as follows: (i) mean absolute error (*MAE*); (ii) sum of squared errors (*SSE*); (iii) root mean square error (*RMSE*); (iv) mean absolute percentage error (*MAPE*); and (v) Theil's inequality coefficient (*TIC*), which is simply the quotient of the root mean square error divided by the notional root mean square error from using naïve forecasts. Turning point or directional errors are not reported here, as they do not reveal much.

Starting with the one-period ahead forecasts, the various measures of forecasting accuracy are exposed in the top panel of table 2. It can be gleaned from this panel that the *RMSE* in particular (the benchmark measure) from each of the three underlying measures is quantitatively smaller than those derived from the headline rate. Compared between the underlying measures, the *RMSEs* are roughly comparable, with WMD the smallest quantitatively of the three, followed by TMN then CEV. These findings are reinforced unanimously by all of the other measures of forecasting accuracy. For multi-period ahead forecasts, as exhibited in the bottom panel of table 2, the same findings apply in terms of the underlying measures having quantitatively lower *RMSEs* than the headline rate, and the quantitative ordering of the three underlying measures. Furthermore, the quantitative differences between the four measures are more profound in the multi-step ahead case.

What is required at this stage, however, is a test of whether the differences in these *RMSEs* are statistically significant. In this endeavour, the Ashley *et al.* (1980) *AGS* test is applied to test for the difference of the *RMSEs* between the four models. By taking the models and comparing them in pairings, the *AGS* test requires the estimation of the linear regression

$$D_t = \alpha_0 + \alpha_1(S_t - \bar{S}) + u_t \quad (22)$$

where

$$D_t = \begin{cases} w_{1t} - w_{2t} : \text{if } \overline{w_1}, \overline{w_2} > 0 \\ -w_{1t} - w_{2t} : \text{if } \overline{w_1} < 0 < \overline{w_2} \\ w_{1t} + w_{2t} : \text{if } \overline{w_2} < 0 < \overline{w_1} \\ -w_{1t} + w_{2t} : \text{if } \overline{w_1}, \overline{w_2} < 0 \end{cases} \quad (23)$$

and

$$S_t = \begin{cases} w_{1t} + w_{2t} : \text{if } \overline{w_1}, \overline{w_2} > 0 \\ -w_{1t} + w_{2t} : \text{if } \overline{w_1} < 0 < \overline{w_2} \\ w_{1t} - w_{2t} : \text{if } \overline{w_2} < 0 < \overline{w_1} \\ -w_{1t} - w_{2t} : \text{if } \overline{w_1}, \overline{w_2} < 0 \end{cases} \quad (24)$$

Also, \bar{S} is the mean of S , w_{1t} (w_{2t}) is defined as the out-of-sample error at t of the model with the higher (lower) *RMSE*, and $t = T + 1, \dots, T + \ell$.

The estimates of α_0 and α_1 in equation (22) are used to test the statistical difference between the *RMSEs* of the multivariate and univariate models. If the estimates of α_0 and α_1 are both positive, then significance is determined by the Wald coefficient restriction test, *CR*, of the joint restriction $\alpha_0 = \alpha_1 = 0$, distributed as $\chi^2(s)$. If one of the estimates is significantly negative, however, then the test is inconclusive. Finally, if the estimate is negative but statistically insignificant, the test will still be

conclusive, with significance being determined by the upper-tail of the t -test on the positive coefficient estimate.

The results of the *AGS* test are presented in table 3. Beginning with the one-step ahead forecasts in the top panel, we first compare CPI with CEV, which is the model of the underlying measures with the highest *RMSE*. It can be seen that both coefficients are positive and the *CR* statistic is significant at the 5 per cent level, indicating that the ‘worst’ of the underlying inflation measures still outperforms the headline rate. It seems reasonable to assume, therefore, that this finding also applies for the other two measures. At the other end of the spectrum, we are also interested whether any of the underlying measures provides superior forecasts to the others. Hence, we compare the other two models (with the lowest two *RMSEs*), TMN and WMD. Here, α_0 is found to be significantly negative, meaning that the test is inconclusive here. For this reason, we then compare WMD and CEV (which has a higher *RMSE* than TMN). On this occasion, both coefficients are positive, but *CR* is insignificant, implying that there is no qualitative difference between the two *RMSEs*. Overall, the evidence is not supportive of the proposition that there is a significant difference between the forecast ability of the three underlying measures.

The analogous results for the multi-period ahead forecasts (shown in the bottom panel of table 3) tell a more conclusive story. When comparing CPI and CEV, once again it is shown that both coefficients are positive and *CR* is significant at 5 per cent, demonstrating that even the ‘worst’ underlying measure outperforms the headline rate. This time, the same results apply when comparing WMD with both TMN and

CEV, meaning that the trimmed mean model does actually stand out as a superior forecasting tool, compared with the other underlying rate measures.

5. Conclusion

This study has attempted to cast some assertions regarding the suitability of using underlying measures of inflation compared with the more widely reported headline CPI, for the purposes of forecasting. The focus of this objective has been a comparative one, via the fitting of a general model and the adjustment of the forecast error series wherever necessary, in order to make the two sets of forecasts comparable in a direct sense. The usual quantitative measures of forecasting accuracy are considered here, whereby measured accuracy depends on the deviation of the forecasts from their realised values. Another feature of this paper is the use of both one-period ahead and multi-period ahead forecasts.

Taking the results of Song (2005) as a base, it is possible to conclude that the more flexible structural time-series approach advocated here provides the same conclusions in terms of the null hypothesis that using underlying rates of inflation does not adversely affect forecasting accuracy. However, the results here are more conclusive, insofar that strong evidence is found that, if anything, using underlying measures of the inflation rate actually *improve* forecasting accuracy. The results, however, are very different between the one-step and multi-step forecasts. For the former, in fact, it is very difficult to qualitatively separate between the three underlying measures considered, although for the latter, the evidence is more conclusive, indicating that the weighted mean measure outperforms both the trimmed mean and excluding volatile items measures.

More generally, the findings support the evidence of Moosa and Ripple (2000) and Lenten and Moosa (2007), who both utilise a very similar methodology, insofar that the forecasting accuracy is not affected adversely by the use of data that has been manipulated by statistical agencies using common procedures. Therefore, these results amount to quite significant support of the manipulation of Australian inflation data by the ABS and RBA for reporting purposes, at least of the grounds of forecasting accuracy.

Table 1: Estimation Results

	CPI	CEV	WMD	TMN
Final State Vector Estimates (Standard Errors)				
μ	1.6609* (0.0047)	1.7427* (0.0044)	1.6615* (0.0013)	1.6730* (0.0030)
λ	0.0054* (0.0023)	0.0102* (0.0019)	0.0125* (0.0019)	0.0112* (0.0018)
ϕ	0.0232	0.0007	0.0003	0.0001
ϕ^*	0.0026	0.0011	0.0006	-0.0004
γ_1	-0.0002 (0.0010)	-0.0005 (0.0006)	0.0003 (0.0004)	0.0004 (0.0005)
γ_1^*	0.0004 (0.0030)	-0.0004 (0.0006)	-0.0001 (0.0004)	-0.0004 (0.0005)
γ_2	-0.0003 (0.0030)	-6.02×10^{-6} (0.0004)	-0.0001 (0.0002)	0.0002 (0.0003)
τ	0.0433* (0.0058)	0.0494* (0.0039)	-0.0047 (0.0025)	-0.0027 (0.0026)
Hyperparameters (q -ratios)				
σ_η^2	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	8.70×10^{-5} (0.0702)
σ_ζ^2	1.57×10^{-6} (0.2360)	1.08×10^{-6} (0.4158)	1.08×10^{-6} (0.4158)	0.0012 (1.0000)
σ_ω^2	0.0000 (0.0000)	2.61×10^{-6} (1.0000)	9.40×10^{-7} (0.4238)	0.0003 (0.2229)
σ_κ^2	0.0000 (0.0000)	1.21×10^{-8} (0.0046)	1.08×10^{-9} (0.0005)	8.83×10^{-5} (0.0712)
σ_ε^2	6.63×10^{-6} (1.0000)	1.95×10^{-6} (0.7469)	2.97×10^{-7} (0.1341)	0.0011 (0.9253)
Goodness-of-fit and Diagnostics				
$\tilde{\sigma}$	0.0055	0.0038	0.0027	0.0027
R^2	0.9990	0.9996	0.9998	0.9998
R_s^2	0.7240	0.7834	0.7154	0.7049
DW	1.9718	1.9547	1.9890	1.9732
Q	10.42	6.4405	6.9910	1.3036
NO	2.2747	1.7873	2.1231	2.0089
H	0.4272	0.6942	1.2552	0.7772
PF	34.03*	9.9897	8.5998	12.60
$CUSUM$	0.5942	0.5846	0.0919	0.0626
AIC	-9.8008	-10.54	-11.21	-11.19
BIC	-9.1725	-9.9101	-10.58	-10.57

*Significant at the 5 per cent level. The critical values associated with the 5 per cent level are approximately as follows: $Q \sim \chi^2(6) \approx 12.60$; $N \sim \chi^2(2) \approx 5.99$; $H \sim F(18,18) \approx 2.17$.

Table 2: Measures of Forecasting Accuracy for Attendances Using All Three Forecasting Models

One-Step Ahead				
	CPI	CEV	WMD	TMN
<i>MAE</i>	6.84×10^{-3}	2.69×10^{-3}	1.62×10^{-3}	2.10×10^{-3}
<i>SSE</i>	1.27×10^{-3}	1.59×10^{-4}	7.48×10^{-5}	1.59×10^{-4}
<i>RMSE</i>	8.39×10^{-3}	2.98×10^{-3}	2.04×10^{-3}	2.50×10^{-3}
<i>TIC</i>	1.0317	0.3658	0.2506	0.3079
<i>MAPE</i>	0.3875	0.1461	0.0926	0.1179
Multi-Step Ahead				
	CPI	CEV	WMD	TMN
<i>MAE</i>	0.0992	0.0149	6.67×10^{-3}	8.59×10^{-3}
<i>SSE</i>	0.2096	5.18×10^{-3}	1.27×10^{-3}	5.18×10^{-3}
<i>RMSE</i>	0.1079	0.0170	8.39×10^{-3}	0.0101
<i>TIC</i>	1.3914	0.2188	0.1082	0.1304
<i>MAPE</i>	5.4557	0.7951	0.3702	0.4751

Table 3: Results of the AGS Test for Statistical Significance of the Difference between RMSEs of Alternative Forecasting Models

One-Step Ahead			
	α_0	α_1	CR
CPI/CEV	3.37×10^{-3} (0.2706)	0.6777* (5.1700)	26.81*
TMN/WMD	-0.0993* (-36.86)	-0.1871 (0.2758)	1,525*
CEV/WMD	8.29×10^{-3} (0.8929)	0.4151 (1.5112)	2.5721
Multi-Step Ahead			
	α_0	α_1	CR
CPI/CEV	0.0843* (43.18)	0.7101* (17.98)	2,188*
TMN/WMD	0.0018 (2.7687)	0.6777* (2.3699)	11.60*
CEV/WMD	0.2999* (6.6798)	0.2347* (4.8876)	197.66*

The *t*-statistics are given in parentheses. *Significant at the 5 per cent level. The critical value associated with the 5 per cent level is approximately: $LR \sim \chi^2(2) \approx 5.99$.

Figure 1: Headline and Various Underlying Inflation Measure Price Level Series for the Full Sample

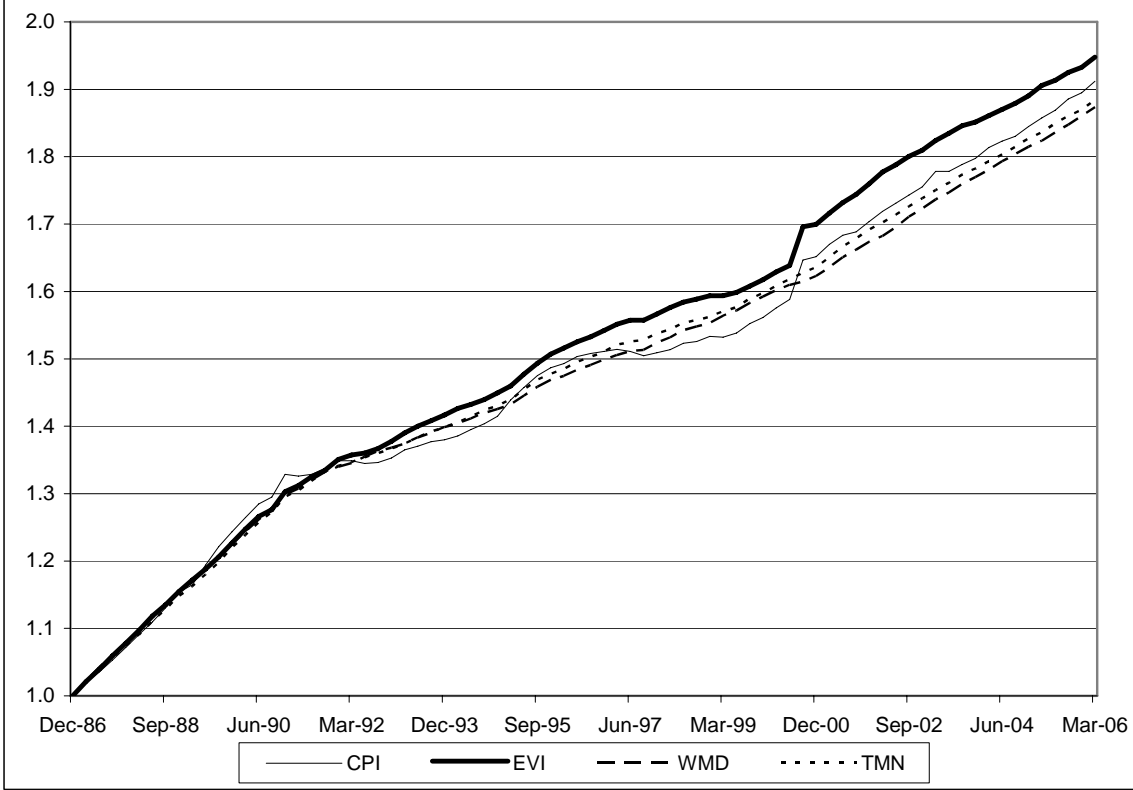


Figure 2: One-Step Ahead Forecasts Using Headline (Solid Line) and Underlying (Dashed Line) Data Versus Actual (Bold Line)

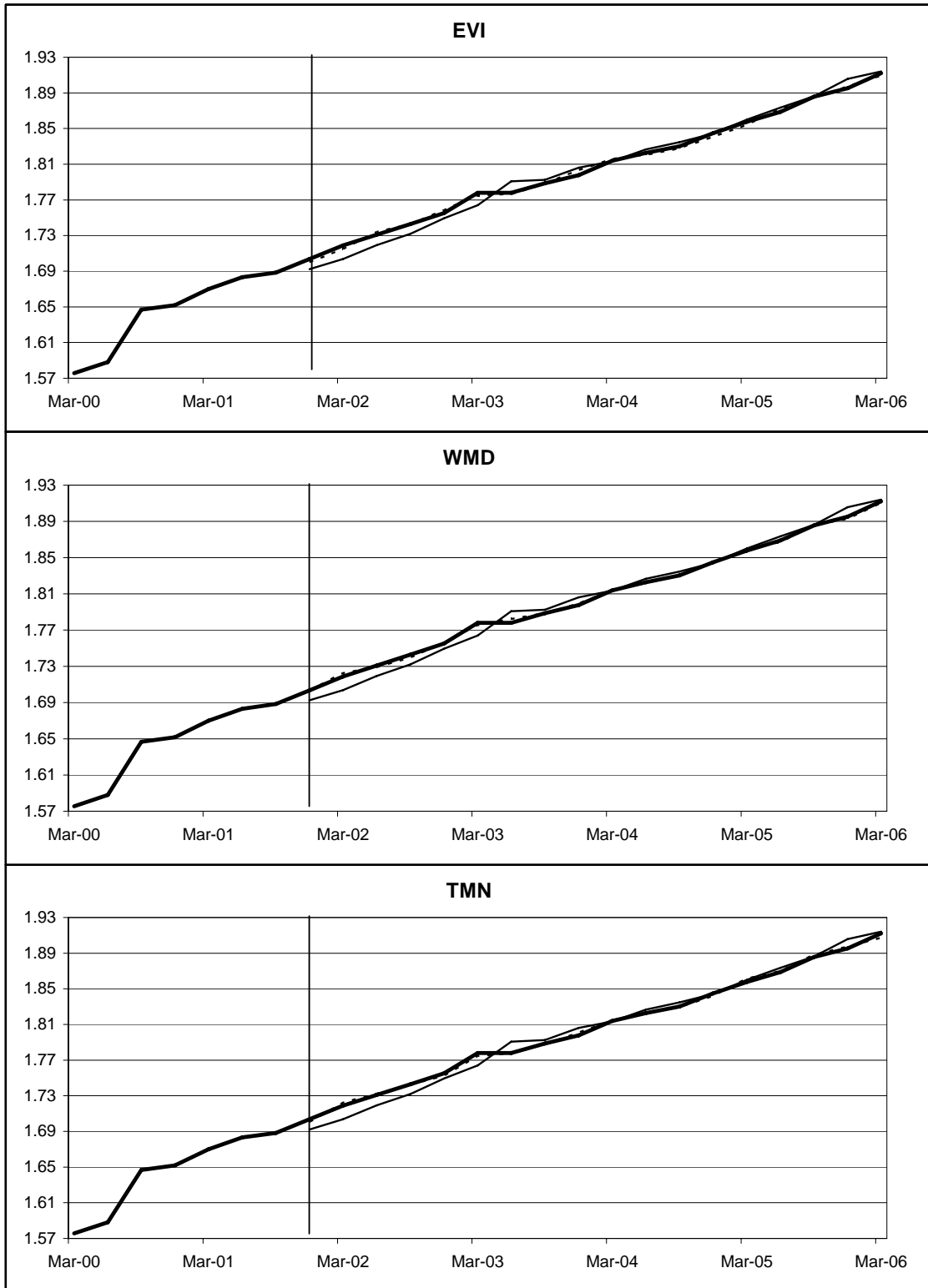
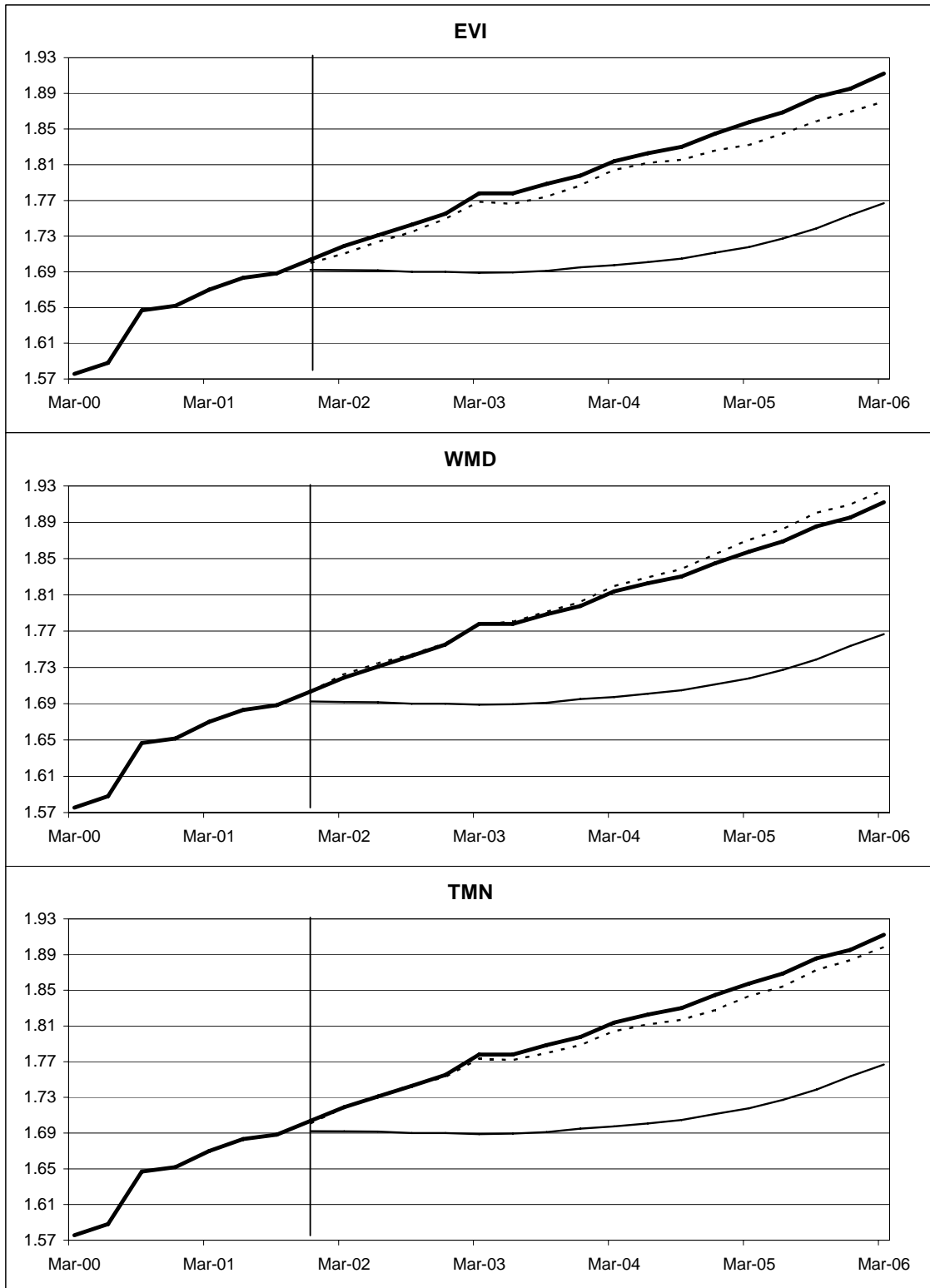


Figure 3: Multi-Step Ahead Forecasts Using Headline (Solid Line) and Underlying (Dashed Line) Data Versus Actual (Bold Line)



References

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle", in *2nd International Symposium on Information Theory*, eds Petrov, B. N. and Csaki, F., Akadémiai Kiadó, Budapest.

Akaike, H. (1977), "On Entropy Maximization Principle", in *Applications of Statistics*, ed. Krishniah, P. R., North Holland, Amsterdam.

Ashley, R., Granger, C. W. J. and Schmalensee, R. (1980), "Advertising and Aggregate Consumption: An Analysis of Causality", *Econometrica*, 48 (5), 1149-1167.

Australian Bureau of Statistics (2005), A Guide to the Consumer Price Index: 15th Series, Catalogue no.:6440.0.

Bowman, K. O. and Shenton, L. R. (1975), "Omnibus Test Contours for Departures from Normality Based on $\sqrt{b_1}$ and b_2 ", *Biometrika*, 62 (2), 243-250.

Brischetto, A. and Richards, A. (2006), The Performance of Trimmed Mean Measures of Underlying Inflation, Reserve Bank of Australia, RBA Research Discussion Papers, 2006-10.

Camba-Mendez, G. and Kapetanios, G. (2005), "Forecasting Euro Area Inflation Using Dynamic Factor Measures of Underlying Inflation", *Journal of Forecasting*, 24 (7), 491-503.

Dixon, R. and Lim, G. C. (2004), "Underlying Inflation in Australia: Are the Existing Measures Satisfactory? " *Economic Record*, 80 (251), 373-86.

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Koopman, S. J., Harvey, A. C., Doornik, J. A. and Shephard, N. G. (1999), *Stamp: Structural Time Series Analyser, Modeller and Predictor*, 2nd edn, Timberlake Consultants Press, London.

Koopman, S. J., Harvey, A. C., Doornik, J. A. and Shephard, N. G. (2000), *Stamp: Structural Time Series Analyser, Modeller and Predictor* [Computer program]. London: Timberlake Consultants Press.

Lenten, L. J. A. and Moosa, I. A. (2007), The Effect of Seasonal Adjustment on Forecasting Accuracy, mimeo, La Trobe University.

Ljung, G. M. and Box, G. E. P. (1978), "On a Measure of Lack of Fit in Time Series Models", *Biometrika*, 65 (2), 297-303.

Makridakis, S. and Hibon, M. (1979), "Accuracy of Forecasting: An Empirical Investigation", *Journal of the Royal Statistical Society - Series A (General)*, 142 (2), 97-125.

Moosa, I. A. and Ripple, R. D. (2000), "The Effect of Seasonal Adjustment on the Accuracy of Forecasting U.S. West Coast Oil Imports", *Journal of Economic Research*, 5 (2), 149-172.

Plosser, C. I. (1979), "Short-Term Forecasting and Seasonal Adjustment", *Journal of the American Statistical Association*, 74 (365), 15-24.

Roberts, I. (2005), Underlying Inflation: Concepts, Measurement and Performance, Reserve Bank of Australia, RBA Research Discussion Papers, 2005-05.

Schwarz, G. (1978), "Estimating the Dimension of a Model", *Annals of Statistics*, 6 (2), 461-464.

Song, L. L. (2005), "Do Underlying Measures of Inflation Outperform Headline Rates? Evidence from Australian Data", *Applied Economics*, 37 (3), 339-345.

Stavrev, E. (2006), Measures of Underlying Inflation in the Euro Area: Assessment and Role for Informing Monetary Policy International Monetary Fund, IMF Working Papers: 06/197.

Vega, J.-L. and Wynne, M. A. (2003), "A First Assessment of Some Measures of Core Inflation for the Euro Area", *German Economic Review*, 4 (3), 269-306.