FORECASTING MACROECONOMIC VARIABLES IN SOUTH AFRICA: PARAMETRIC V.S NON-PARAMETRIC METHODS

Isaac K. Kanyama[ǂ] and Bonolo M. Thobejane*

*Department of Economics and Econometrics, University of Johannesburg
[ǂ]Lecturer, Department of Economics and Econometrics, University of Johannesburg

Abstract. Most literature that investigates the forecasting ability of econometric techniques in providing projection about macroeconomic variables in South Africa are based on parametric models. The aim of this research paper is to provide a modern econometric tool such as the non-parametric model and compare its forecasting performance with that of a parametric model thus contributing to the existing literature. We will evaluate whether the non-parametric model can outperform the univariate time-series model in forecasting inflation and GDP growth rate. We compare the out-of-sample forecast of the two models based on the root mean squared error (RMSE) and decide on which method provides the best forecasting accuracy which policy makers can rely on in forecasting inflation and GDP growth rate.

JEL Classification: E31, C14, C53
Key words: forecasting, macroeconomic variables, non-parametric model.

1. INTRODUCTION

Forecasting macroeconomic variables has become a second nature to policy makers, government and monetary authorities. Projection about macro-economic variables provides a clear picture of how the state of the economy is likely to perform in the future. It is then crucial for researchers to investigate which models are appropriate and relevant enough to perform an accurate projection of macroeconomic variables, which policy makers, government and monetary authorities can use to predict, in our case inflation and GDP growth rate.

Inflation and GDP growth rate are two of the major macroeconomic variables that are closely monitored not only by the government but also the average citizens. They serve as an indication of how the economy is performing and how it is performing in comparison to the countries trading partners. When investors decide on which country to invest their capital not only do they consider the liquidity of the stock and bond market, interest rate, exchange rate stability and political stability but they also keep a close eye on inflation which gives an idea about the stability of the macroeconomic environment in which investors intent to place their capital and GDP growth rate which signifies the performance of the country's economy. Hence we decided on evaluating the forecasting of inflation and GDP growth rate as our key macroeconomic variables.

Of uttermost importance is the fact that forecasting macroeconomic variables provides a clear picture of what the state of the economy will be in future. Having the relevant models to forecast these macroeconomic variables is significant for policy makers and government in allocating resources efficiently. According to Croushore (1996) forecast of inflation and possibly any macroeconomic variables based on rational expectation should meet two criteria. First they must be unbiased. The deviation of actual inflation from predicted inflation on average should equal zero. Secondly they must be efficient; forecasters must make the best use of all the relevant and available information.

It is evident from the existing literature that not enough papers have considered a comparison of parametric and non-parametric model in forecasting. Although existing research has been conducted in forecasting inflation solely, only few papers have considered both inflation and GDP growth. Hence the aim of this paper is to assess the forecasting performance of parametric model and the non-parametric model in forecasting the key macroeconomic variables and thus contributing to the existing literature.

The rest of the paper is structured as follows: Section 2 looks at the existing literature done on this topic; Section 3 describes the methodology and the data followed by the model specification in Section 4. Section 5 provides the empirical results and lastly Section 6 concludes with the overall findings of the paper.

2. LITERATURE REVIEW

An extensive number of research has been conducted not only on forecasting macroeconomic variables but also investigating the performance of alternative model through out-of-sample forecasting. The following models have been considered in forecasting inflation: univariate time series model (ARIMA), Phillips curve, interest rate model and the naïve model. Hafer and Hein (1990) and Alles and Horton (2000) have evaluated the forecasting performance of univariate time series and interest rate models in forecasting inflation. In addition to these models Alles and Horton (2000) also used the error correction model and public survey of inflation forecast. Both papers found that the univariate time series model outperforms or equally performs well as the respective models.

A recent paper by Lee (2012) looked at the predictive performance of the univariate time series model, Phillips curve and the naïve model in forecasting inflation for countries that have adopted inflation targeting. The paper found that the univariate time series model yields superior forecast than the Phillips curve and the naïve model. A wide number of literature [see Hafer and Hein (1990), Ang, Bekaert and Wei (2007)] have also shown that the univariate time-series model (ARIMA) model tends to outperform and/or equally perform the same as the interest rate model or Phillips curve in forecasting inflation in the USA.

Forecasting of macroeconomic variables with non-parametric methods mainly use Artificial Neural Networks (ANNs). Moshiri and Cameron (2000), Kamruzzaman and Sarker (2003), Binner *et al*. (2005) and Duzgun (2010) compared ANNs to ARIMA models in forecasting inflation. This strand of research finds that the sophisticated ANNs model tends to outperform ARIMA model and performs equally as the ARIMA model in some other cases. However, He *et al.* (2012) show that ARIMA model are suitable for forecasting inflation in the USA and that ANNs model sometimes cannot improve the forecasting results.

Although there has been an increasing interest in forecasting and assessing macroeconomic variables in South Africa, little has been done on comparing econometric techniques specifically parametric and non-parametric models in order to give an indication on their suitability in making predictions. Woglom (2005) looks at the determinants of inflation forecasts and finds output gap, short term interest rate and import price inflation providing useful information in explaining inflation. On the other hand, Dave *et al*. (2009) forecast the growth rate of output, inflation and nominal short term interest rate in South Africa using the New-Keynesian Dynamic Stochastic General Equilibrium (NKDSGE) model. The authors compare the

NKDSGE model with the classical and Bayesian VAR model and find the former model provides the best forecasting accuracy than the later models.

However, fewer studies on South Africa have considered alternative models of forecasting in South Africa. Gupta and Kabundi (2010) considered five alternative models namely, the small open economy new Keynesian dynamic stochastic general equilibrium (SOENKDSGE) model, small-scale classical and Bayesian vector autoregressive (VAR) model and large-scale dynamic factor models (DFMs) and Bayesian VAR model (BVAR) in predicting per capita growth rate, CPI inflation, money market rate and growth rate of the nominal effective exchange rate. Their forecasting results indicated that the large-scale BVAR model outperformed the remaining models.

To the best of our knowledge, the only paper that uses parametric and non-parametric models with a case study on South Africa is Bonga-Bonga and Mwamba (2011). This paper investigates the predictability of stock market returns in South Africa using a generalized autoregressive conditional heteroskedastic mean (GARCH-M) model and the conditional heteroskedastic non-linear autoregressive (NAR) model. Although this paper is not directly related to our analysis, it is interesting to note that the its overall conclusion points to the fact that the non-parametric model out-performs the parametric model for the one-day-ahead forecast while the parametric model does better for the 90-days horizon foresting. Closely related to our analysis is the paper by Ogwang (1993), who investigates the forecasting performance of ARIMA model versus the Kernal regression in predicting inflation in Canada. The author finds the non-parametric model providing the best forecasting accuracy than the ARIMA model.

3. METHODOLOGY AND THE DATA

This paper assesses the forecasting performance of a parametric model and a non-parametric mode in forecasting inflation and GDP growth rates in South Africa. We use quarterly data sourced from Easy Data (Quantec) database. The annual percentage growth of inflation and GDP growth where calculated as [(CPIt/$CPI_{t-4)}$ -1]*100 and [(GDP_t/$GDP_{t-4)}$ -1]*100 respectively. Where CPI_t is the headline consumer price index and GDP_t is the gross domestic product at constant prices. The sample period for both our key variables is 1960:01- 2012:04 and this was selected in order to capture the different phases of the business cycle and of course, the different behaviour of inflation and GDP growth. The out-of-sample forecasting period is 2000:01 – 2012:04 with only 52 observations. Using the root mean squared error and other statistics we identify which of the two methods best forecast inflation and GDP growth rate.

When selecting an appropriate model for a given set of observation {X_t, t=1,2,3,…,n} it is important that the researcher examine the type of data they are dealing with. This means checking if the series is non-stationary and whether it exhibits a trend and/or seasonality. If the series shows signs of non-stationarity, trend or seasonality then appropriate transformation of the series is required in order to avoid obtaining spurious results that will lead to misleading conclusion.

The graphical analysis of inflation rate indicates that the series is non-stationary. Figure 1 illustrates the plot of inflation rate for the sample period 1990:01-2012:04.It is evident from figure 1 that the series is non-stationary since it tends to wonder up and down with no distinct pattern and hence transformation of the series is required. After first-differencing the inflation rate series became mean reverting.

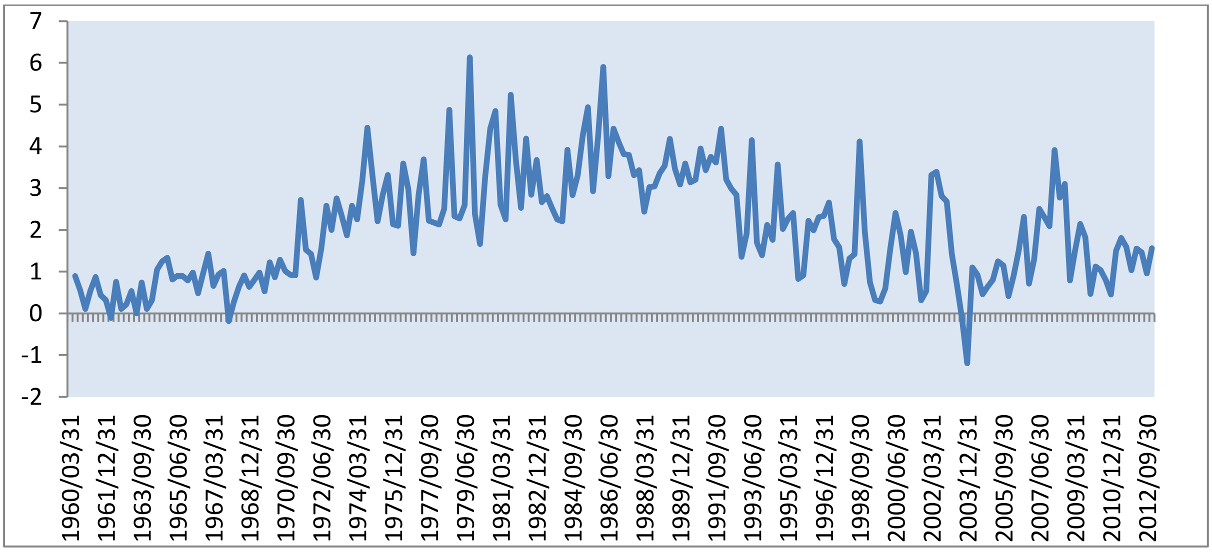

Figure 1: Trends in the South African headline inflation
Source: EasyData (Quantec) database

On the other hand a graphical analysis of the GDP growth on figure two for the sample period 1960:01-2012:04 shows that the series exhibit seasonality. The plot shows that there is a decrease in GDP growth rate for the first quarter compared to the previous year and there are spikes in every second quarter and also that there is a decrease in growth rate to a positive figure during quarter two to quarter three in every year. Given that GDP growth exhibit seasonality we desesonalized the series. Therefore our empirical estimation of parametric and non-parametric model will be based on transformed series of both our key macroeconomic variables.

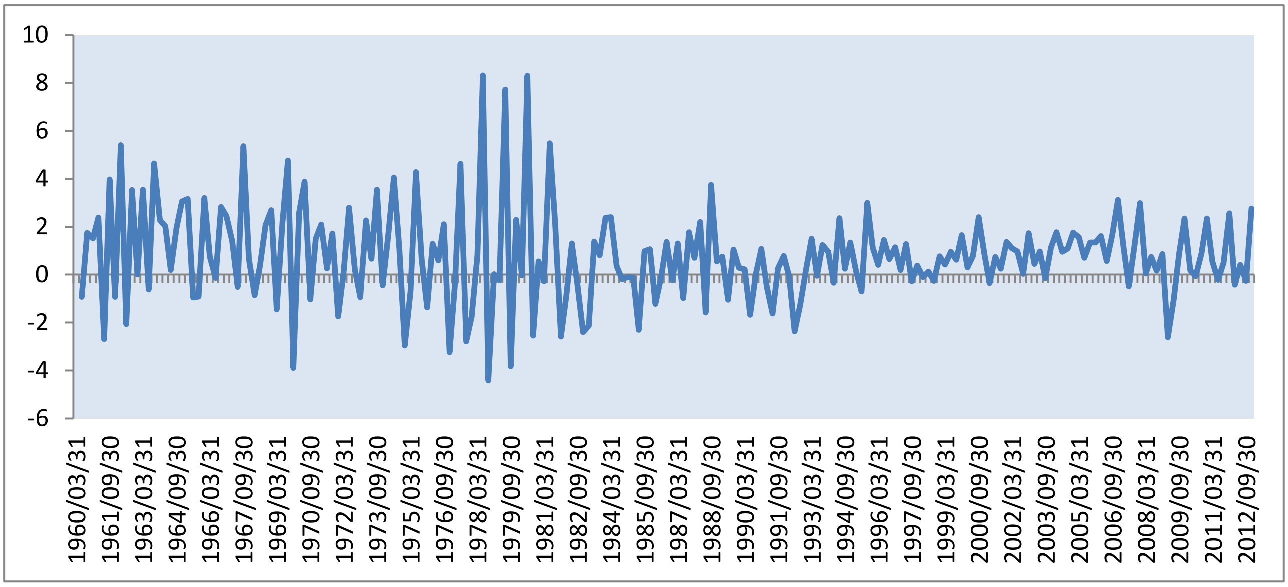

Figure 2: Trends in the South African real GDP growth
Source: EasyData (Quantec) database

4. Model specification

4.1 The parametric model

The ARIMA model has been used extensively in time series analysis ever since the publication of *Time Series Analysis: Forecasting and Control* by Box and Jenkins. The popularity of this model also known as the Box- Jenkins (BJ) methodology is based on the philosophy let the data speak for itself. Stevenson and Mcgarth (2003) considered the model as atheoretical, implying that it ignores all other potential theories except the ones that are related to the variable under study.

The generalized ARIMA model with *p,d,q* process has the following specification:

$$y_t = \delta + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_n y_{t-p} + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \ldots + \beta_q \varepsilon_{t-q} \qquad (1)$$

Where δ and ε denotes the constant and error term respectively. The lagged autoregressive (AR) process are symbolised by *p* and that of a moving average (MA) process are symbolised by *q* and of course the order of integration is *d.*

The following test statistics (Akaike's Information Criterion (AIC), Schwartz Bayesian Criteria (SBC) the modified portmanteau test by Ljung and Box, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE)) will be used to select our best model for forecasting, Stevenson and McGarth (2003):

$$AIC = Tln(RSS) + 2n \qquad (2)$$

$$BIC = Tln(RSS) + nln(T) \qquad (3)$$

$$Q = T(T+2)\sum_{j=1}^{h} \rho_w^2\ (j)/(T-j) \qquad (4)$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{j=1}^{T}(y_j - \hat{y}_j})^2 \qquad (5)$$

$$MAE = \frac{1}{T}\sum_{j=1}^{T}\ |y_j - \hat{y}_j| \qquad (6)$$

4.2 The non-parametric model

One of the drawbacks of a parametric model is that the data is assumed to be normally distributed and the functional form underlying the model is known. An important implication of this fact is that if prior to the estimation and forecasting of the model an error is made about the functional form underlying the model, then it is evident that projection made about macroeconomic variables will be misleading and unreliable. Hence a non-parametric model is explored, which does not require the specification of the functional form

The general specification of a non-parametric model:

$$y_t = m(y_{t-1,}y_{t-2,}\dots y_{t-n}) + \varepsilon_t \qquad t = 1,2,3,\dots,n$$

Where m (·) is the conditional mean and u_t is a white noise sequence. Note that the function m (·) is unknown whereas in the parametric model functional form m (·) is known.

5. EMPIRICAL RESULTS

After a thorough investigation of identifying the order of *p,d q,* for the ARIMA model, a number of ARIMA models where estimated. According to Newold (1983) “Timeseries model building is not an attempt to make the data fit a particular model, but rather to make a model that fits the data.” In order to select a model that fits the data well, the residuals obtained from the estimated model should be white noise. This leads us to diagnostic checking in which, sample autocorrelation function (ACF) and partial autocorrelation function (PACF) are used to test whether the residual are random. In addition the Akaike’s Information Criterion (AIC), Schwartz Bayesian Criteria (SBC) and the modified portmanteau test by Ljung and Box (1978) are used.

Table 1a and 1b illustrates an analysis of residuals obtained from possible ARIMA models and the kernel regression non-parametric model for inflation and GDP growth respectively using the above mentioned test statistics. These models have coefficients that are statistically significant at 5 percent level of significance. A model with the lowest AIC, SBC and whose residuals exhibit no serial autocorrelation will be deemed to be competitive.

Table 1a.
Analysis of residuals from ARIMA Models using inflation rate series for the sample period 1960:01-1999:04

Model	AIC	SBC	Ljung and Box Q	P-value
ARIMA (||2||,1,1)	2.736	2.791	26.22	0.09477**
ARIMA (0,1,3)	2.728	2.802	21.63	0.1994**
ARIMA (3,1,||2||)	2.757	2.850	21.07	0.1759**
ARIMA (3,1,0)	2.780	2.8554	33.28	0.0104*
Kernal regression			24.95	0.0707**

AIC denotes the Akaike's Information Criterion, SBC denotes Schwartz Bayesian Criterion and the Q statistics is the modified portmanteau test with H_0: residuals exhibit no serial correlation.
** Do not reject the null hypothesis at 5 percent level of significance.
* Reject the null hypothesis at 5 percent level of significance.

Table 1b.
Analysis of residuals from ARIMA Models using GDP growth series for the sample period 1960:01-1999:04

Model	AIC	SBC	Ljung and Box Q	P-value
ARIMA (3,0,3)	3.632	3.762	16.26	0.2977**
ARIMA (1,0,||4||))	3.723	3.778	71.66	0.0000*
ARIMA (||1, 8,11||,0,||4||)				
ARIMA (||4||,0,||4||))	3.635	3.691	25.36	0.1153**
Kernal regression			16.35	0.4289**

AIC denotes the Akaike's Information Criterion, SBC denotes Schwartz Bayesian Criterion and the Q statistics is the modified portmanteau test with H_0: residuals exhibit no serial correlation.
** Do not reject the null hypothesis at 5 percent level of significance.
* Reject the null hypothesis at 5 percent level of significance.

From table1a when comparing the potential ARIMA model it is clear that ARIMA (0,1,3) has the lowest AIC criterion but in terms of SBC criterion the ARIMA(||2||,1,1) seem to have the lowest value. Stevenson and McGarth (2003) used only the SBC criterion when selecting model while Tse (1997) used solely the AIC criterion. However in this paper we will consider both criterions. Using the Q statistics under the null hypothesis of residuals not exhibiting serial correlation we do not reject the hypothesis at 5 percent level of significance for both models including the kernel regression non-parametric model. Hence ARIMA (0,1,3) and ARIMA(||2||,1,1) are two competing models and the one with the lowest root mean squared error will be compared against the non-parametric model.

On the other hand under GDP growth, ARIMA (3,0,3) and ARIMA (||4||,0,||4||) where found to have the lowest AIC and BIC respectively. Both models including the kernel regression non-parametric model also showed to have residuals that are not serially correlated and thus we will identity of the three models which one has the lowest root mean squared error. Such model will then provide the best forecasting accuracy of GDP growth.

Now that we have identified the potential ARIMA models (i.e. models in which all coefficients are significant and have residuals that are white noise) for forecasting our key macroeconomic variables we are one step closer in discovering which methods whether non-parametric or parametric provide the best forecasting accuracy which policy makers can rely on..

First we perform the dynamic forecast in which previously forecasted values of the lagged dependent variables are used in forming forecasts of the current value and next we consider a one-step ahead static forecasting. ARIMA (||2||,1,1) model proved to have the lowest RMSE and MAE. However when compared to the Kernal regression non-parametric model the ARIMA (||2||,1,1) fails to outperform the former model. For dynamic forecasting the RMSE and MAE for the kernel regression are 0.903814 and 0.70644 respectively and 0.983696 and 0.742370 respectively for the ARIMA (||2||,1,1).

Table 2a

Forecasting summary statistics of out-sample inflation forecasts for the period 2000 Q1-2012 Q2 (Dynamic forecasting)

Model	RMSE	MAE
ARIMA (||2||,1,1)	0.983696	0.742370
ARIMA (0,1,3)	1.012938	0.755882
Kernal regression	0.903814	0.70644

Note: RMSE denotes the root mean squared. MAE is the mean absolute error. These results were generated using the E-views package

Table 2

Forecasting summary statistics of out-sample inflation forecasts for the period 2000 Q1-2012 Q2 (static forecasting)

Model	RMSE	MAE
ARIMA (||2||,1,1)		
ARIMA (0,1,3)		
Kernal regression		

Note: RMSE denotes the root mean squared. MAE is the mean absolute error. These results were generated using the E-views package

Table 3a illustrates the results for out-of-sample forecast for GDP growth rate. As mentioned earlier we consider both dynamic and static forecasting performance and recalling that the GDP growth series was stationary and deseasonalized the order of integration therefore becomes zero (*d*=0).

Table 3a

Forecasting summary statistics of out-sample GDP growth forecasts for the period 2000:01-2012:04(Dynamic forecasting)

Model	RMSE	MAE
ARIMA (3,0,3)	1.031872	0.769339
ARIMA (||4||,0||4||)	1.057553	0.791573
Kernal regression	0.915038	0.690113

Note: RMSE denotes the root mean squared. MAE is the mean absolute error. These results were generated using the E-views package.

In terms of dynamic forecasting an ARIMA (3,0,3) proved to have the lowest RMSE and MAE of 1.031872 and 0.769339 respectively .

Table 3a

Forecasting summary statistics of out-sample GDP growth forecasts for the period 2000:01-2012:04(static forecasting)

Model	RMSE	MAE
ARIMA (3,0,3)		
ARIMA (||4||,0||4||)		
Kernal regression		

Note: RMSE denotes the root mean squared. MAE is the mean absolute error. These results were generated using the E-views package.

6. CONCLUSION

This paper attempts to provide reliable econometric tool that can be used by monetary authorities and policy makers when making projections about the future values of inflation and GDP growth in South Africa. The paper compares forecasting ability of parametric and non-parametric model. Both models were estimated using quarterly data for the sample period 1960:01-1999:04 using out-of-sample forecast period of 2000:01-2012:04.

<u>**REFERENCES**</u>

Alles L, Horton D.1999. An evaluation of the alternative methods of forecasting Australian inflation. The Australian Economic Review **32** (2) 237-248.

Ang A, Bekaert G & Wei M. 2007. Do macro variables, asset markets or surveys forecast inflation better? Journal of Monetary Economics **54** 1163-1212.

Binner JM., Bissoondeeal RK., Elger T., Gazely AM. & Mullineux AW. A comparison of linear forecasting models and neural networks: an application to Euro inflation and Euro Divisia. Journal of Applied Economics **37**:665-680

Bonga-Bonga L. & Mwamba M. 2011. The predictability of stock markert returns in South Africa: Parametric VS. Non-parametric methods. South Africa Journal of Economics **79**(3) 301-311.

Croushore D.1996. Inflation forecasts: How good are they? Business Review (May) 15-25.

Duzgun R. 2010. Generalized Regression Neural Networks for Inflation Forecasting. International Research Jourmal of Finance and Economics **51** 59-70.

Hafter RW. & Hein SE.1990. Forecasting inflation using interest rate and time-series models: Some international evidence. Journal of Business **63** 1-17.

Gupta R, Kabundi A. 2010. Forecasting macroeconomic variables in a small open economy: A comparison between small- and large-scale models. Journal of Forecasting **29** 168-185.

He Q., Shen H. & Tong Z. 2012. Investigation of inflation Forecasting. Applied Mathematics & Information Science **6** (3): 649-655.

Kamruzzaman J. & Sarker RA. 2003. Forecasting of currency exchange rates using ANN:A case study. IEEE Int. Conf. Neural Networks & Signal Processing 14-17

Lee U. 2012. Forecasting inflation for inflation-targeted countries: A comparison of the predictive performance of alternative inflation forecasting models. Journal of Business & Economic Studies **18** (1) 75-95.

Liu GD.,Gupta R. & Schaling E. 2009. A New-Keynesian DSGE for forecasting the South African Economy. Journal of Forecasting **28** 387-404.

Ljung GM. & Box GEP. 1978. On the measure of lack of fit in time series models. Biometrika **65** 297-303.

Marcellino M. 2008. A linear benchmark for forecasting GDP growth and inflation? Journal of Forecasting **27**: 305-340.

Moshiri S. & Cameron N. 2000. Neural Networks Versus Econometric models in Forecasting Inflation. Journal of Forecasting **19**: 201-217.

Newbold P.1983. ARIMA model building and the timeseries analysis approach to forecasting. Journal of Forecasting **2**: 23-35

Ogwang T. 1993. A comparison of the Kernel and the ARIMA estimates of inflationary expectations: Some evidence from Canada. Empirical Economics **18**: 281-288.

Stevenson S. & Mcgarth O. 2003. A comparison of alternative rental forecasting models: empirical tests on the London office market. Journal of Property Research **23** (3) 235-260.

Stock JH. & Watson MW. 2007. Why has U.S. inflation became harder to forecast? Journal of Money, Credit and Banking **39** (1) 3-33.

Tse RYC. 1997. An application of the ARIMA model to real estate prices in Hong Kong. Journal of property finance **8**:281-99.

Woglom G. 2005. Forecasting South African Inflation. South African Journal of Economics **73** (2) 302-320.