Eastern Mediterranean University

Department of Economics
Discussion Paper Series

Forecasting Core Inflation: The Case of South Africa

Franz Ruch, Mehmet Balcilar, Mampho P. Modise, Rangan Gupta

Discussion Paper 15-08

July 2015

Department of Economics
Eastern Mediterranean University
Famagusta, North Cyprus

Forecasting Core Inflation: The Case of South Africa

Franz Ruch* Mehmet Balcilar † Mampho P. Modise ‡ Rangan Gupta §

June 17, 2015

Abstract

Forecasting and estimating core inflation has recently gained attention, especially for inflation targeting countries, following research showing that targeting headline inflation may not be optimal; a Central Bank can miss the signal due to the noise. Despite its importance there is sparse literature on estimating and forecasting core inflation in South Africa, with the focus still on measuring core inflation. This paper emphasises predicting core inflation using large time-varying parameter vector autoregressive models (TVP-VARs), factor augmented VAR, and structural break models using quarterly data from 1981Q1 to 2013Q4. We use mean squared forecast errors (MSFE) and predictive likelihoods to evaluate the forecasts. In general, we find that (i) small TVP-VARs consistently outperform all other models; (ii) models where the errors are heteroscedastic do better than models with homoscedastic errors; (iii) models assuming that the forgetting factor remains 0.99 throughout the forecast period outperforms models that allow for the forgetting factors to change with time; and (iv) allowing for structural break does not improve the predictability of core inflation. Overall, our results imply that additional information on the growth rate of the economy and interest rate is sufficient to forecast core inflation accurately, but the relationship between these three variables needs to be modelled in a time-varying (nonlinear) fashion.

JEL Classification: C22, C32, E27, E31.

Keywords: Second-round effects, Inflation, relative price shocks, Bayesian estimationCore inflation, forecasting, small- and large-scale vector autoregressive models, constant and time-varying parameters

*South African Reserve Bank, PO Box 3070, Pretoria, South Africa, 0001. Email: franz.ruch@resbank.co.za.The views expressed are those of the author(s) and do not necessarily represent those of the South African Reserve Bank or Reserve Bank policy. While every precaution is taken to ensure the accuracy of information, the South African Reserve Bank shall not be liable to any person for inaccurate information or opinions contained herein.

†Department of Economics, Eastern Mediterranean University, Famagusta, Northern Cyprus, via Mersin 10, Turkey, and Department of Economics, University of Pretoria, Pretoria, 0002, South Africa. Email: mehmet@mbalcilar.net

‡National Treasury, 40 Church Square, Pretoria, 0002, South Africa. Email: mamphomodise@yahoo.com.

§Corresponding author. Department of Economics, University of Pretoria, Pretoria, 0002, South Africa. Email: rangan.gupta@up.ac.za.

Contents

List of Tables

List of Figures

1 Introduction

Like many countries targeting inflation, the South African Reserve Bank uses headline inflation as its operational target. Headline inflation however can be volatile, making it difficult to distinguish increases in generalised price versus temporary shocks. The volatility of headline inflation results from sharp changes in the price of a small number of goods and services, and may be unrelated to the performance of economic activity or underlying trends in price setting behaviour, obscuring trends in underlying inflation (Ranchhod et al., 2013). Core inflation measures therefore attempt to examine the component of inflation that is related to broad trends in economic conditions and pricing behaviour, and which are likely to be more persistence (Ranchhod et al., 2013). Bryan and Cecchetti (1994) further describe core inflation as a process that should be highly persistence, forward looking and strongly linked to monetary policy dynamics. The importance of core inflation has been recognised in South Africa in recent years (see for example Blignaut et al., 2009; Rangasamy, 2009; Ruch and Bester, 2013; and Du Plessis et al., 2015). Du Plessis et al. (2015) argue that for inflation targeting countries, core inflation has a direct effect on the policy-decision making process.

While there is vast domestic and international literature on measuring core inflation, there is limited literature on forecasting core inflation and include Bryan and Cecchetti (1993), Sun (2004), Morana (2007), Camba-Mendez and Kapetanios (2005), and Kapetanios (2004). Sun (2004) proposes an approach to forecast Thailand core inflation. He combines a short-term model which attempts to filter the forecasting power of a variety of monthly indicators purely on goodness-of-fit criteria, with an equilibrium-correction model that pins down the convergence of core inflation to its longer-run structural determinants. Morana (2007) uses principal components frequency domain approach which is suited to estimate systems of fractionally cointegrated processes to estimate and forecast core inflation for the euro area. Bryan and Cecchetti (1993), Camba-Mendez and Kapetanios (2005), and Kapetanios (2004) propose the use of large datasets using factor models in modelling core inflation. Specifically, Camba-Mendez and Kapetanios (2005) estimate factors from datasets of disaggregated price indices for European countries. They then assess the forecasting ability of these factor estimates against other measures of underlying inflation built from more traditional methods. In South Africa, the literature tends to focus on forecasting headline inflation (for detailed literature reviews, see Woglom, 2005; Bahramian et al., 2014; and Gupta et al., 2015) and constructing measures of core inflation (Blignaut et al., 2009; Rangasamy, 2009; Ruch and Bester, 2013; and Du Plessis et al., 2015). It is against this background that our paper focuses on forecasting core inflation for the South African economy by employing a large number of econometric models.

The first contribution of this paper is that we employ methods for forecasting core inflation in large TVP-VARs developed by Koop and Korobilis (2013). These models use forgetting factors for computational feasibility. Second, and in addition to the models in Koop and Korobilis (2013), we also assess the performance power of factor augmented VARs. As stated in Camba-Mendez and Kapetanios (2005), dynamic factor models tend to perform well in comparison to traditional measures. Third, the paper adds value by also considering structural break models. Du Plessis et al. (2015) states that South African core inflation data has recently been subjected to a structural break given changes in the basket of goods and services and the methodology used in constructing this index. To deal with the structural break dilemma, we combine the Pesaran et al. (2006) (PPT) and the Koop and Potter (2007) (KP) methods. The basic idea of the methodology is to use the PPT prior for the break process and the KP prior in conditional mean and variance. We follow Koop and Korobilis (2013), Koop and Korobilis (2012), and Stock and Watson (1999) in the selection of data which is motivated by a basic New Keynesian model with a generalised Phillips curve. We use quarterly data starting from 1981Q1 to 2013Q4 for 21 variables which include activity variables, labour market variables, financial variables and other prices. To the best of our knowledge, this is the first paper to formally forecast core inflation in South Africa. The only other relevant paper is that of Gupta et al. (2015), where the authors used the latent state-information recovered from a Dynamic Stochastic General Equilibrium (DSGE) model to forecast core inflation, which was, however, not modelled within the DSGE model explicitly. This meant that, it was not possible to identify the variables that could help forecast core inflation. Allowing for large number of predictors, in line with the empirical literature on forecasting inflation and or core inflation based on a New-Keynesian Phillips curve, we are clearly able to determine, which variables contain predictive information for core inflation.

The rest of the paper is as follows: section 2 discusses the methodology followed by section 3 discussing the data. Section 4 follows with the discussion of the results before concluding in section 5.

2 Methodology

2.1 Large TVP-VARs

In this paper we follow the specification in Koop and Korobilis (2013) and specify the TVP-VAR as:

$$y_t = z_t \beta_t + \varepsilon_t \tag{1}$$

and

$$\beta_{t+1} = \beta_t + \mu_t \tag{2}$$

Where ε_t is an independently and identically distributed (i.i.d.) error with $N(0, \Sigma_t)$ and μ_t i.i.d. $N(0, Q_t)$. ε_t and μ_t are independent of one another for all s and t. y_t for $t = 1,,T$ is an $M \times 1$ vector containing observations on M time series variables and Z_t is a $M \times k$ matrix defined so that each TVP-VAR equation contains an intercept and p lags of each of the M variables such that $k = M(1 + pM)$. Following Koop and Korobilis (2013), and also in Fagin (1964), Jazwinski (2007), and Raftery et al. (2005), we use forgetting factors instead of standard Bayesian statistical inference since the latter tends to work well only with small TVP-VARs. Forgetting factors allow the Kalman filter to be run only k times providing an accurate approximation of the likelihood function as the state vector becomes independent across models (for further details on formulating the Kalman filter see, amongst others, (Koop and Korobilis, 2013) as well as (Frühwirth-Schnatter, 2006)). In estimating a TVP-VAR using forgetting factors, let $y^s = (y_1, \ldots, y_s)'$ denote observations through time s. The standard Kalman filter states that:

$$\beta_{t-1}|y^{t-1} \sim N(\beta_{t-1|t-1}, V_{t-1|t-1}) \tag{3}$$

The formulae for $\beta_{t-1|t-1}$ and $V_{t-1|t-1}$ are given in Frühwirth-Schnatter (2006). Further,

$$\beta_t|y^{t-1} \sim N(\beta_{t|t-1}, V_{t|t-1}) \tag{4}$$

where

$$V_{t|t-1} = V_{t-1|t-1} + Q_t \tag{5}$$

To estimate using the forgetting factor, we replace the above equation by:

$$V_{t|t-1} = \frac{1}{\lambda} V_{t-1|t-1} \tag{6}$$

λ is the forgetting factor and is restricted to the interval $0 < \lambda \leq 1$. Equation 6 implies that observations j periods in the past have weight ^{j} in the filter estimate of β_t. This controls the degree of time-variation of the coefficients. Equations 5 and 6 also imply that $Q_t = (\lambda^{-1} - 1)V_{t-1|t-1}$ from which it can be observed that the constant coefficient case arises if $\lambda = 1$. Raftery et al. (2010) set $\lambda = 0.99$ while Koop and Korobilis (2012) use $[0.8, 0.95, 0.99]$. In this paper we show results for $\lambda = 0.99$ as well as follow the approach in Koop and Korobilis (2013) of estimating λ at each point in time.

We also use a decay factor, κ, to simplify the implementation of multivariate stochastic volatility in ε_t. Exponential Weighted Moving Average (EWMA) is used to estimate Σ_t following RiskMetrics (1996):

$$\hat{\Sigma}_t = \kappa \hat{\Sigma}_{t-1} + (1-\kappa)\hat{\varepsilon}_t \hat{\varepsilon}_t' \tag{7}$$

where $\hat{\varepsilon}_t = y_t - \beta_{t|t} Z_t$ is estimated by the kalman filter. We set the decay factor equal to 0.96.

Although TVP-VARs work relatively well for modelling gradual evolution of coefficients, they tend to work poorly for more sudden changes of coefficients. Allowing for switches between entirely different models can accommodate more abrupt breaks. We use methods developed in Raftery et al. (2010), Koop and Korobilis (2012), and Koop and Korobilis (2013) for doing dynamic model averaging (DMA) which can also be used for dynamic model selection (DMS). DMA refers to the averaging of a large set of (j) models, weighted based on their predictive content, to forecast at a specific point in time; i.e. calculating the likelihood function for $j = 1, ..., J$ and averaging these likelihoods to generate a forecast. This produces a probability $\pi_{t|t-1,j}$ with $j = 1, ..., J$. $\pi_{t|t-1,j}$ vary over time and the forecasting model can switch over time. Once the $\pi_{t|t-1,j}$ for $j = 1, ..., J$ are obtained, they can either be used to do model selection or model averaging. DMS refers to when the single best model, which can change overtime, given selection over a large number of predictors is used to forecast at each point in time; i.e. selecting the model with the highest likelihood. The advantage of this approach is that optimal values for λ, κ

and the VAR shrinkage parameter can be selected in a time-varying manner.

To construct a dynamic model selection, we follow the basic algorithm in Raftery et al. (2010), Koop and Korobilis (2012), and Koop and Korobilis (2013). Given the initial condition: $\pi_{0|0,j}$ for $j = 1,...,J$, the model prediction equation using the forgetting factor approach can be derived as follows:

$$\pi_{t|t-1,j} = \frac{\pi^{\alpha}_{t-1|t-1,j}}{\sum_{l=1} j\pi^{\alpha}_{t-1|t-1,l}} \tag{8}$$

and a model updating equation of:

$$\pi_{t|t-1,j} = \frac{\pi_{t-1|t-1,j} p_j(y_t|y^{t-1})}{\sum_{l=1} j\pi_{t|t-1,l} p_l(y_t|y^{t-1}} \tag{9}$$

Where $p_j(y_t|y^{t-1})$ is the predictive likelihood, measuring the forecast performance. $\pi_{t|t-1,j}$ can be written as follows:

$$\pi_{t|t-1,j} \propto \prod_{i=1} t-1[p_j(y_{t-i}|y^{t-i-1})]^{\alpha} \tag{10}$$

The above equation can be interpreted as follows: if $\alpha = 0.99$ then the forecast performance five years ago receives 80 per cent as much weight as forecast performance last period, but if $\alpha = 0.95$ then the weight for the forecast performance five years ago will only be 35 per cent. $\alpha = 1$ corresponds to conventional model averaging using maximum likelihood.

Since we are estimating large VARs and time-varying VARs, there is a need to also define models as arising from different priors as opposed to only using values for the forgetting and decay factors ((Bańbura et al., 2010) as well as (Koop and Korobilis, 2013)). We use a tight Minnesota prior for β_0 specified in Koop and Korobilis (2013), which tends to be similar to the Normal prior. Further, we also allow for the estimation of the shrinkage hyperparameter in a time-varying manner which is computationally less demanding than re-estimating the shrinkage priors and the model at each point in time. We therefore specify the prior mean to be equal to $E(\beta_0) = 0$, after transforming the data to stationarity. The Minnesota prior covariance matrix for β_0 is a diagonal such that $var(\beta_0) = \underline{V}$ and $\underline{V}_i$ denotes the diagonal element. The prior covariance matrix is then defined through:

$$\underline{V}_i = \begin{cases} \frac{\gamma}{r^2} & \text{for coefficients on r for r=1,...,p} \\ \underline{a} & \text{for the intercept} \end{cases} \tag{11}$$

Where p is the lag length. $\underline{V}$ and γ are the key hyperparameter controlling the degree of shrinkage on the VAR coefficients. As in Koop and Korobilis (2013) we use one shrinkage parameter to simplify computation. This approach differs slightly to the Minnesota prior in that it contains two shrinkage parameters which are set to fixed values. $\underline{a}$ is set to equal 10^2. To produce reasonable forecast performance in large VARs and TVP-VARs, a large degree of shrinkage is necessary. We therefore use a wide grid for $\gamma \in [10^5, 0.001, 0.005, 0.01, 0.05, 0.1]$.

We also augment the model space with models of different dimensions. In particular we do dynamic model selection for a small (including only three variables), medium (including seven variables) and large (including 21 variables) TVP-VAR. As discussed in Koop and Korobilis (2013) and also used by Ding and Karlsson (2014) working with TVP-VAR of different dimensions, y_t will be of different dimension and therefore predictive densities $p_j(y_{t-1}|y^{t-i-1})$ will not be comparable. This can be resolved by using the predictive densities for the small VAR (these are variables that are included in all models). In this analysis it means that the dynamic model selection is determined by the joint predictive likelihood for economic growth, core inflation and the 3-month Treasury bill rate.

2.2 FAVAR model

Although work by Koop and Korobilis (2013) and Bańbura et al. (2010) provide techniques to shrink the parameter space in order to make Large VAR estimation and analysis feasible, it may be that using other methods such as data shrinkage from factor augmented VARs provide better forecasts of core inflation. We use all data included in the large VAR (excluding core inflation itself) to estimate factors for a FAVAR model. The transformed data is standardised.

In order to determine the number of factors to use we implement a modified Bai and Ng (2002) information criterion as implemented by Alessi et al. (2010). This method chooses the number of factors by minimising the variance of the idiosyncratic component of the approximate factor model. This is subject to a penalisation in order to avoid over-parameterisation. The information criterion is:

$$\hat{r}^T_{c,N} = \underset{0 \le k \le r_{max}}{\operatorname{argmin}} IC^{T*}_{\alpha,N}(k) \tag{12}$$

where

$$IC^{T*}_{\alpha,N}(k) = log[\frac{1}{NT}\sum_{i=1}^{N}\sum_{i=1}^{T}(x_{it} - \hat{\gamma}_i^{(k)}\hat{F}_t^{(k)})^2] + ckp_a(N,T) \text{for a=1,2} \tag{13}$$

For k common factors, N is the number of variables, T the number of observations, $x_{it} - \hat{\gamma}_i^{(k)}\hat{F}_t^{(k)}$ the idiosyncratic error, c an arbitrary positive real number and $p_a(N,T)$ the penalty function. Alessi et al. (2010) propose multiplying the penalty function by c since Hallin and Liška (2007) show that a penalty function, $p(N,T)$ leads to consistent estimation of r, the number of factors, if and only if $cp(N,T)$ does as well.

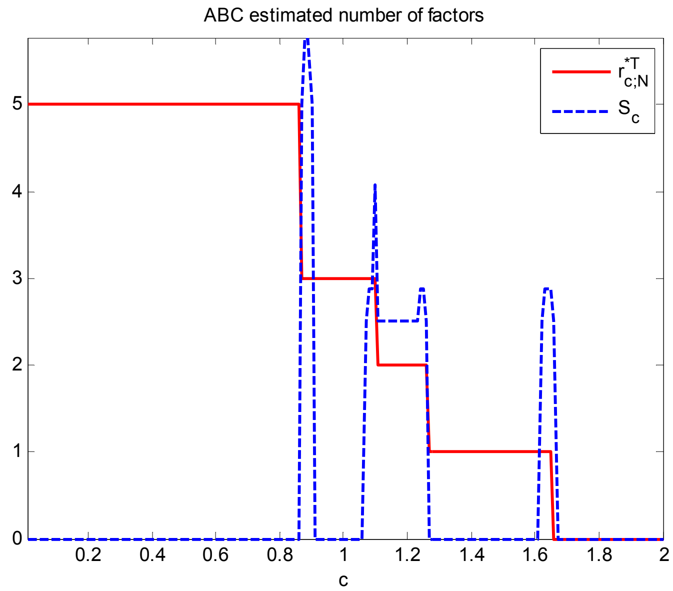
The only information available regarding the behaviour of $\hat{r}^T_{c,N}$ can be gleaned from analysing subsamples of sizes (n_j, t_j). For any j, we can compute $\hat{r}^{t_j}_{c,n_j}$ which is a monotonic non-increasing function in c. Therefore, there exist moderate values of c such that $\hat{r}^T_{c,N}$ converges from above to r. This result, however, needs to be independent of j for the criterion to be stable. This is measured by the variance of

$\hat{r}^{t_j}_{c,n_j}$ as a function of j:

$$S_c = \frac{1}{J}\sum_{j=1}^{J}[\hat{r}^{t_j}_{c,n_j} - \frac{1}{J}\sum_{j=1}^{J}\hat{r}^{t_j}_{c,n_j}]^2 \tag{14}$$

Figure 1 shows the estimated number of factors for our model. The vertical axis represents the number of factors while the horizontal axis represents an arbitrary positive real number c. We run the results over a number subsample sizes in order to get a robust result. To determine the number of factors we have to find the first value of $\hat{r}^{T}_{c,N}$ where S_c is zero. The results suggest that the number of factors should be three.

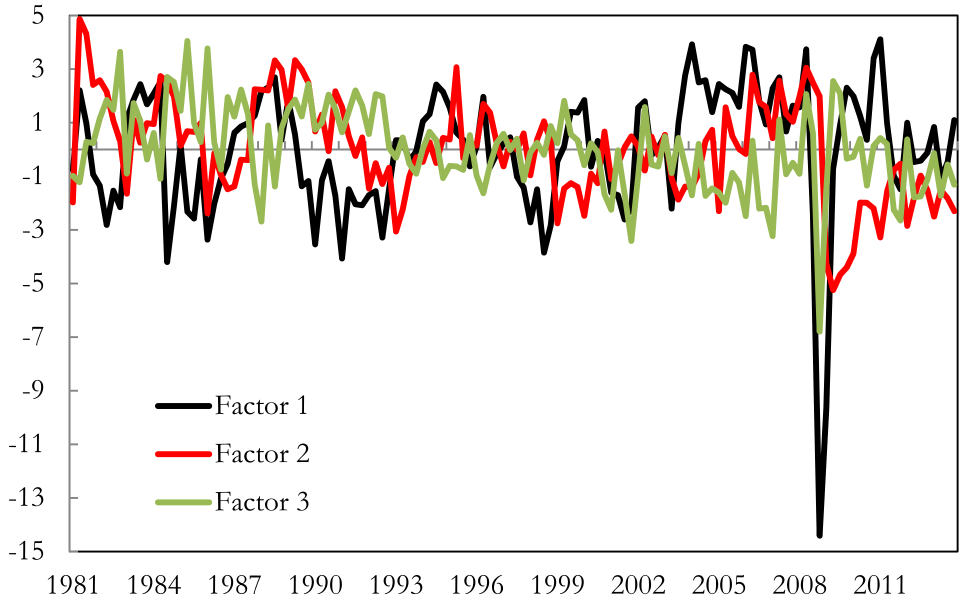
Figure 1: Estimating the number of factors



Other methods were also used including the original Bai and Ng (2002) information criterion and a method proposed by Onatski (2010). The Bai and Ng (2002) method did not converge, a common problem with smaller datasets. According to Onatski (2010), three factors were also chosen.

The estimated factors from this analysis are plotted in figure 2.

Figure 2: Estimated factors



2.3 Structural break models

Since structural breaks tend to be common in macroeconomic data and are one of the major reasons for poor forecasting (Stock and Watson, 1996; Ang and Bekaert, 2002; Clements and Hendry, 1998; and Bauwens et al., 2011), we also consider a structural break model to forecast core inflation in South Africa. We consider a combination of the PPT and the KP priors. The model uses the PPT prior for the break process and the KP prior in conditional mean and variance. We use the same framework as in Bauwens et al. (2011) and a detailed discussion is presented there. We specify the linear regression model framework for the structural break models as:

$$y_t = Z_t \beta_{s_t} + \sigma_{s_t} \varepsilon_t \tag{15}$$

Where y_t is the dependent variable, Z_t contains the lagged dependent variables or lagged exogenous variables available for forecasting y_t, ε_t is i.i.d. $N(0,1)$. β_{s_t} determines the conditional mean coefficients and σ_{s_t} represents volatilities. This regression allows for β_{s_t} and σ_{s_t} to vary over time with $s_t \in 1,...,K$ a random variable indicating which regime applies at time t.

We use the KP prior in conditional mean and variance which adopts a hierarchical prior motivated by the state space literature on time-varying parameter models (discussed in detail in Bauwens et al., 2011). The random walk evolution of coefficients is specified as:

$$\beta_j = \beta_{j-1} + \mu_j \tag{16}$$

Where μ_j is i.i.d. $N_m(0, B_0)$ which is equivalent to $\beta_j|\beta_{j-1} \sim N_m(\beta_{j-1}, B_0)$. The parameters β_{j-1} and B_0 are unknown and can de estimated from the data. This means that when a structural break occurs, the conditional mean coefficients are drawn from a distribution centred at β_{j-1}. It is the most recent regime which has the most influence on conditional mean coefficients in a new regime.

To model the break process, we consider an approach in Chib (1998) and used in PPT. Assume that the restricted Markov process for s_t is given by:

$$Pr(s_t = 1|s_{t-1} = i) = p_i \tag{17}$$

and

$$Pr(s_t = i+1|s_{t-1} = i) = 1 - p_i \tag{18}$$

This equation is interpreted as a hierarchical prior and implies a geometric prior distribution fo $d_i = \tau_t - \tau_{t-1}$ - which measure the durations of regimes. Therefore if regime i holds at time $t-1$, then at time t the process can either remain in regime i with probability p_i or a break occurs and the process moves to regime $i+1$ with probability $1-p_i$. To select the number of breaks, we rely on the specification in Bauwens et al. (2011) and set the maximum breaks allowed to five such that: $K = 1, ..., K^{max}$.

3 Data

The data used is motivated by a generalised New Keynesian Phillips curve as in Koop and Korobilis (2012) and Stock and Watson (1999). Table 1 provides details of the 21 variables included in the dataset, the VAR these variables are used in as well as the transformation imposed. The data is quarterly and ranges from 1981Q1 to 2013Q4. All data is transformed to be stationary (see transformation in table 1). These include activity variables such as real GDP and capacity utilisation; labour market variables such as unit labour cost, wages and employment; financial variables such as stock returns and money stock; and other prices such as producer price inflation, oil prices and non-energy commodity prices. Note that the start and end date of our sample is purely driven by data availability on the various variables used, at the time of writing this paper.

Table 1: Data series used in the small, medium and large VARs

Variable	Transformation*	Description	VAR*
RGDP	Log first diff.	Gross domestic product at market prices (GDP)	S,M,L
CORE	Log first diff.	Headline CPI less interest on mortgages, food, petrol and electricity	S,M,L
TB3	Levels	Treasury bills: 91 days tender rate	S,M,L
NEER	Log first diff.	Nominal effective exchange rate of the rand: Average for the period - 15 trading partners	M,L
OIL	Log first diff.	Brent crude oil spot price (USD)	M,L
FORPRD	Log first diff.	Foreign wholesale price index (trade weighted) (own calculation)	M,L
ULC	Log first diff.	Manufacturing: Unit labour costs	M,L
PCE	Log first diff.	Final consumption expenditure by households: Total	L
GFCF	Log first diff.	Gross fixed capital formation (Investment)	L
JSE	Log first diff.	Johannesburg Stock Exchange (JSE) All Share index	L
M3	Log first diff.	Money supply: M3	L
CREDIT	Log first diff.	All monetary institutions: Total domestic credit extension	L
LEAD_FOR	Log first diff.	Leading indicator of all the main trading partner countries	L
RETAIL	Log first diff.	Retail sales	L
WAGES	Log first diff.	Total salaries and wages in the manufacturing sector	L
EMPL_PVT	Log first diff.	Employment in private sector (own calculation)	L
INCOME	Log first diff.	Disposable income of households	L
IP	Log first diff.	Industrial production (own calculation)	L
UTIL	Levels	Manufacturing: Utilisation of production capacity - Total	L
PPI	Log first diff.	Manufacturing Producer Price Index	L
COM_NENG	Log first diff.	World bank commodity price index: non-energy (USD)	L

*Log first diff=logged and the first difference was used, S=Small VAR, M=Medium VAR, L=Large VAR

4 Empirical results

The main results of this paper are presented in Table 2 and Table 3. These show the iterated forecasts for horizons 1 to 8 quarters ($h = 1, \ldots, 8$) with a forecast evaluation period of 2000Q1 to 2013Q4, i.e., the starting point of the out-of-sample period corresponds to the starting quarter of the inflation targeting era in south African monetary policy decision. The VAR models are estimated with $p = 1$ based on the Bayesian Information Criteria (BIC). The following models are presented:

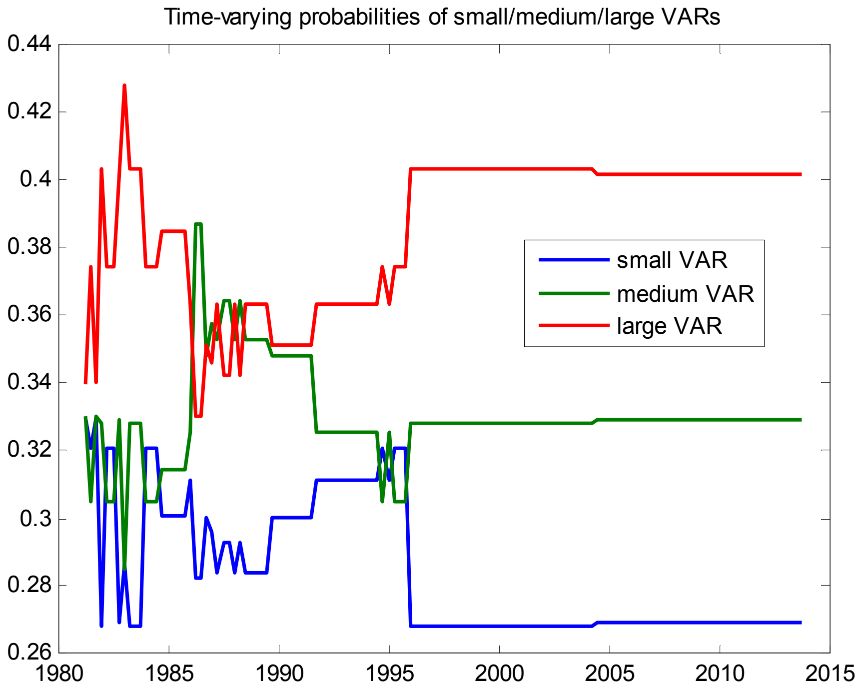
- A full approach which uses all three VAR model sizes using DMS; referred to as dynamic dimension selection (DDS). This is labelled TVP-VAR-DDS in the tables;
- TVP-VAR model using the three different size VARs including a small (S) VAR using three variables; a medium (M) VAR using seven variables; and a large (L) VAR using 21 variables;
- Heteroscedastic VARs using the three dimensions setting $\lambda = 1$ and $\kappa = 0.96$;
- Homoscedastic VARs using the three dimenstions setting $\lambda = 1$ and $\kappa = 1$;
- A structural breaks model using PPT and KP priors;
- A random walk model;
- TVP-AR models;
- A small VAR estimated using Ordinary Least Squares (OLS);
- FAVAR models;
- and an AR(1) model using OLS.

Since the use of iterated forecast increases the computational burden we follow Koop and Korobilis (2013) and do the predictive simulation in two ways. First, we use the VAR coefficients which hold at time T to forecast variables at time $T+h$. This method assumes that the VAR coefficients remain unchanged between T and $T+h$. Second, we assume that these coefficients change out-of-sample and simulate from the random walk state equation 1 to produce draws of β_{T+h} and is labelled as $\beta_{T+h} \sim RW$ in the tables. Both methods provide β_{T+h}, which we use to simulate draws of y_{T+h} conditional on β_{T+h} to approximate the predictive density.

To evaluate the forecast performance we use the MSFE and the predictive likelihood. The MSFE and the predictive likelihood in Table 2 and 3 are presented as relative to the random walk model. This means that the numbers in Table 2 are the ratios of a particular model specification divided by the random walk model. For Table 3, the results presented are the sum of log predictive likelihood of different models minus the sum of lag predictive likelihood obtained for the random walk model.

DDS forecasts use the TVP-VAR of dimension with the highest probability. We therefore plot the time-varying probabilities associated with the TVP-VAR of each dimension in Figure 3. Between 1981 and 1985 and from 1988 to the end of our sample period, the DMS uses large TVP-VARs to produce forecasts. The medium TVP-VAR performs better than the large TVP-VAR only during 1986 and 1987, while the probability of the small TVP-VARs is constantly lower than the other two TVP-VARs dimensions.

Figure 3: Estimated Dynamic Dimension Selection probabilities of the small, medium and large TVP-VARs



In general, most models (full model, TVP-AR, small and medium TVP-VARs and FAVAR) and different specification (excluding the VAR with homoscedastic errors) perform better than the random walk model. The models perform particularly much better for $h = 3,4,5,6$. The full model and the TVP-AR are preferred for core inflation across all horizons and model specifications. For the small and medium TVP-VARs as well as the FAVAR only the VAR with homoscedastic errors is outperformed by the random walk model at some horizons. The poor performance of the homoscedastic VAR model highlights the importance of allowing for heteroscedastic errors in getting the shape of the predictive density. In general, these results show that the models employed in this paper provide an effective way of estimating even large VAR with heteroscedastic errors and choosing prior shrinkage.

The large TVP-VAR and the benchmark models tend to compete with the random walk model since they perform better in some horizons and worse in others. The large TVP-VAR outperforms the random walk model only when $h = 3,4.5,6,7$. While for both VARs with homoscedastic errors and VARs with heteroscedastic errors perform worse than the random walk, the heteroscedastic VARs outperforms homoscedastic VARs. The $AR(1)$ structural break model does well relative to the random walk only from $h = 3$.

From Table 2, there are no significant gains when simulating β_{T+h} from the random walk model compared to just assuming that the VAR coefficients remain unchanged over the forecast horizon. The noticeable comparison can be made between models where $\lambda = 0.99$ and models with $\lambda = \lambda_t$. Models where the forgetting factor is pre specified outperform models where the forgetting factors are allowed to change over time.

In summary, the full model, TVP-AR, the small and medium TVP-VAR models, as well as the FAVAR (excluding the VAR with homoscedastic errors) perform better on average relative to the other models and the random walk model. Specifically, the small TVP-VAR has the smallest MSFE relative to all models employed.

With regards to the predictive likelihood results presented in Table 3, all VAR specifications perform significantly better than the random walk model, confirming somewhat the results presented in Table 2. Even in this case, models with $\lambda = 0.99$ perform better than models where $\lambda = \lambda_t$. Also, the VARs with heteroscedastic errors outperform the VARs with homoscedastic errors. Even with the predictive likelihood, the benchmark models tend to perform poorly relative to the random walk model. Only the $AR(1)$ structural break model performs better than the random walk model for $h = 3$ onwards. When taking the average of the models, the full model, TVP-AR, the small and medium TVP-VAR models, as well as the FAVAR (excluding the VAR with homoscedastic errors) perform better on average relative to the other models - as in the case in Table 2. Similar to the results using the MSFE, the small TVP-VAR has the largest predictive likelihoods relative to all other models for all specifications.

5 Conclusion

In this paper, we use a suite of econometric models to forecast quarterly core inflation in South Africa using 21 variables for the period covering 1981Q1 to 2013Q4. The forecasts are evaluated using the MSFE and the predictive likelihood relative to the random walk model for $h = 1, ..., 8$. We find that most VAR models (specifically the small TVP-VARs and excluding the large TVP-VARs) perform better than the random walk model and other benchmark models for both forecast evaluation methods and for all horizons. Allowing for structural break does not improve the forecast performance for core inflation. The structural model only performs better than the random walk model for $h = 3$ onwards, but is outperformed by other models. Further, the forecasts where we allow for heteroscedastic errors in getting the shape of the predictive density outperform VARs with homoscedastic errors. We also find that models with $\lambda = 0.99$ perform better than models where the forgetting factors are allowed to change over time. Overall, our results imply that additional information on the GDP growth rate and interest rate is sufficient to forecast core inflation accurately, but the relationship between these three variables needs to be modelled in a time-varying (nonlinear) fashion.

Camba-Mendez and Kapetanios (2005) used disaggregated price indices to forecast core inflation by employing factor models. In light of this, future research could be aimed at forecasting South African core inflation using disaggregated price indices based on time-varying models, to see if such disaggregated information on price can produce more accurate forecasts than those obtained from GDP growth rate and interest rates.

Table 2: Relative MSFE relative to the random walk model for core inflation

	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
Full Model								
TVP-VAR-DDS, λ=0.99,$\beta_{T+h} = \beta_T$	0.91	0.66	0.53	0.53	0.48	0.55	0.64	0.62
TVP-VAR-DDS, λ= λ_t,B_{T+h}~RW	0.91	0.66	0.53	0.53	0.48	0.55	0.65	0.62
TVP-AR								
TVP-AR, λ=0.99,$\beta_{T+h} = \beta_T$	0.80	0.57	0.45	0.41	0.40	0.44	0.50	0.50
TVP-AR, λ= λ_t, $\beta_{T+h} = \beta_T$	0.85	0.67	0.54	0.48	0.47	0.52	0.58	0.60
TVP-AR, λ=0.99, B_{T+h}~RW	0.79	0.55	0.45	0.42	0.40	0.43	0.49	0.50
TVP-AR, λ= λ_t, B_{T+h}~RW	0.86	0.67	0.54	0.48	0.48	0.52	0.58	0.60
Small VAR								
TVP-VAR, λ=0.99, $\beta_{T+h} = \beta_T$	0.73	0.47	0.38	0.35	0.35	0.39	0.45	0.45
TVP-VAR, λ= λ_t, $\beta_{T+h} = \beta_T$	0.75	0.52	0.43	0.39	0.39	0.44	0.50	0.51
TVP-VAR, λ=0.99, B_{T+h}~RW	0.73	0.47	0.38	0.35	0.35	0.39	0.44	0.45
TVP-VAR, λ= λ_t, B_{T+h}~RW	0.75	0.53	0.43	0.39	0.39	0.44	0.49	0.51
VAR, Heteroscedastic	0.77	0.58	0.47	0.43	0.43	0.48	0.53	0.57
VAR, Homoscedastic	1.02	1.02	0.84	0.74	0.74	0.83	0.86	0.96
Medium VAR								
TVP-VAR, λ=0.99, $\beta_{T+h} = \beta_T$	0.85	0.52	0.42	0.42	0.40	0.44	0.49	0.53
TVP-VAR, λ= λ_t, $\beta_{T+h} = \beta_T$	0.91	0.64	0.52	0.50	0.48	0.54	0.58	0.64
TVP-VAR, λ=0.99, B_{T+h}~RW	0.86	0.52	0.42	0.42	0.40	0.44	0.49	0.53
TVP-VAR, λ= λ_t, B_{T+h}~RW	0.92	0.64	0.52	0.50	0.49	0.54	0.59	0.64
VAR, Heteroscedastic	0.97	0.74	0.60	0.56	0.55	0.61	0.65	0.72
VAR, Homoscedastic	2.41	2.23	1.79	1.53	1.51	1.69	1.70	1.86
Large VAR								
TVP-VAR, λ=0.99, $\beta_{T+h} = \beta_T$	1.12	1.03	0.86	0.80	0.77	0.86	0.91	1.01
TVP-VAR, λ= λ_t, $\beta_{T+h} = \beta_T$	1.20	1.18	0.98	0.88	0.86	0.94	0.98	1.08
TVP-VAR, λ=0.99, B_{T+h}~RW	1.13	1.02	0.85	0.80	0.77	0.86	0.91	1.00
TVP-VAR, λ= λ_t, B_{T+h}~RW	1.20	1.18	0.97	0.87	0.85	0.94	0.98	1.07
VAR, Heteroscedastic	1.34	1.37	1.13	1.00	0.97	1.08	1.11	1.23
VAR, Homoscedastic	2.35	2.36	1.91	1.65	1.61	1.78	1.80	1.93
FAVAR								
TVP-VAR, λ=0.99, $\beta_{T+h} = \beta_T$	0.94	0.66	0.51	0.48	0.48	0.53	0.59	0.57
TVP-VAR, λ= λ_t, $\beta_{T+h} = \beta_T$	0.97	0.75	0.59	0.54	0.54	0.60	0.66	0.67
TVP-VAR, λ=0.99, B_{T+h}~RW	0.94	0.65	0.50	0.48	0.48	0.52	0.59	0.57
TVP-VAR, λ= λ_t, B_{T+h}~RW	0.95	0.75	0.59	0.54	0.54	0.60	0.66	0.66
VAR, Heteroscedastic	0.99	0.83	0.65	0.59	0.59	0.66	0.72	0.73
VAR, Homoscedastic	2.52	2.12	1.76	1.52	1.49	1.63	1.68	1.92
Benchmark Models								
Random Walk	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Small VAR OLS	1.00	1.01	0.99	0.96	0.99	1.14	1.18	1.26
AR(1) OLS	1.03	1.06	1.07	1.03	1.08	1.25	1.28	1.37
AR(1) Structural Breaks	1.68	1.18	0.96	0.87	0.85	0.91	0.90	0.94
Average performance								
Excluding benchmark models	1.08	0.89	0.72	0.65	0.64	0.71	0.76	0.81
Including benchmark models	1.09	0.91	0.75	0.69	0.68	0.75	0.80	0.85

Table 3: Sum of log predictive likelihoods relative to the random walk model for core inflation

	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8
Full Model								
TVP-VAR-DDS, λ=0.99,$\beta_{T+h} = \beta_T$	83.13	62.05	56.36	52.90	49.98	47.22	45.62	45.03
TVP-VAR-DDS, $\lambda = \lambda_t$,$B_{T+h} \sim RW$	83.56	62.06	56.70	52.64	49.99	46.70	45.46	45.16
TVP-AR								
TVP-AR, λ=0.99,$\beta_{T+h} = \beta_T$	82.29	67.25	61.50	55.62	53.43	53.15	48.57	49.04
TVP-AR, $\lambda = \lambda_t$, $\beta_{T+h} = \beta_T$	80.59	63.50	57.66	51.86	49.79	49.09	45.19	45.16
TVP-AR, λ=0.99, $B_{T+h} \sim RW$	81.85	67.49	60.96	54.88	53.70	52.85	48.91	49.17
TVP-AR, $\lambda = \lambda_t$, $B_{T+h} \sim RW$	80.53	63.62	57.34	51.76	49.84	49.09	45.51	45.30
Small VAR								
TVP-VAR, λ=0.99, $\beta_{T+h} = \beta_T$	85.87	71.46	65.29	59.35	56.43	55.49	51.31	51.34
TVP-VAR, $\lambda = \lambda_t$, $\beta_{T+h} = \beta_T$	85.02	69.35	62.59	56.94	54.21	53.29	49.24	48.81
TVP-VAR, λ=0.99, $B_{T+h} \sim RW$	85.97	71.39	64.71	59.07	56.01	55.29	51.26	51.07
TVP-VAR, $\lambda = \lambda_t$, $B_{T+h} \sim RW$	84.89	69.22	62.46	56.67	54.09	53.11	49.25	48.94
VAR, Heteroscedastic	84.05	67.32	60.79	54.86	52.40	51.37	47.19	46.33
VAR, Homoscedastic	76.02	46.95	43.88	35.76	34.23	35.85	29.74	25.23
Medium VAR								
TVP-VAR, λ=0.99, $\beta_{T+h} = \beta_T$	85.53	68.39	62.30	55.04	53.14	51.87	48.94	48.00
TVP-VAR, $\lambda = \lambda_t$, $\beta_{T+h} = \beta_T$	82.46	64.32	58.24	51.19	49.14	47.83	45.14	43.61
TVP-VAR, λ=0.99, $B_{T+h} \sim RW$	85.56	68.29	62.10	54.79	52.91	51.54	48.91	47.57
TVP-VAR, $\lambda = \lambda_t$, $B_{T+h} \sim RW$	82.25	64.38	58.18	50.79	49.25	47.89	44.84	43.83
VAR, Heteroscedastic	80.82	61.11	55.14	48.33	46.18	44.67	42.51	40.68
VAR, Homoscedastic	54.22	30.17	23.57	19.97	20.10	18.58	16.18	15.56
Large VAR								
TVP-VAR, λ=0.99, $\beta_{T+h} = \beta_T$	77.56	52.92	45.93	40.06	38.34	36.39	34.57	33.24
TVP-VAR, $\lambda = \lambda_t$, $\beta_{T+h} = \beta_T$	75.32	49.61	42.49	37.35	35.60	34.56	32.43	31.71
TVP-VAR, λ=0.99, $B_{T+h} \sim RW$	77.51	53.22	46.01	40.22	38.13	36.72	34.41	33.50
TVP-VAR, $\lambda = \lambda_t$, $B_{T+h} \sim RW$	75.26	49.61	42.98	37.72	35.92	34.56	32.69	32.21
VAR, Heteroscedastic	72.10	44.81	38.02	33.02	31.31	30.13	28.37	27.68
VAR, Homoscedastic	52.86	25.14	18.68	10.94	11.07	11.05	9.95	14.81
FAVAR								
TVP-VAR, λ=0.99, $\beta_{T+h} = \beta_T$	83.76	63.10	58.07	51.40	49.44	47.99	44.79	45.89
TVP-VAR, $\lambda = \lambda_t$, $\beta_{T+h} = \beta_T$	82.24	60.00	54.81	48.44	46.38	44.81	41.73	42.69
TVP-VAR, λ=0.99, $B_{T+h} \sim RW$	83.95	63.14	58.03	51.53	49.33	48.10	45.05	46.08
TVP-VAR, $\lambda = \lambda_t$, $B_{T+h} \sim RW$	82.52	59.85	54.82	48.47	46.71	44.91	42.00	42.77
VAR, Heteroscedastic	80.77	57.74	52.31	46.39	44.24	42.58	39.36	40.06
VAR, Homoscedastic	48.09	17.37	4.82	2.46	2.95	3.61	6.21	8.86
Benchmark Models								
Random Walk	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Small VAR OLS	-1.47	-3.62	-5.44	-7.37	-9.29	-10.57	-13.00	-15.81
AR(1)	-0.73	-2.18	-3.49	-5.11	-6.75	-7.95	-10.54	-13.39
AR(1) Structural Breaks	-34.11	-7.82	6.09	11.99	17.09	20.30	24.05	26.96
Average performance								
Excluding benchmark models	75.80	55.68	49.62	43.97	42.17	41.08	38.34	38.19
Including benchmark models	69.93	51.78	46.43	41.21	39.57	38.57	35.96	37.32

References

Alessi, L., M. Barigozzi, and M. Capasso (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters 80*(23), 1806–1813.

Ang, A. and G. Bekaert (2002). Regime switches in interest rates. *Journal of Business & Economic Statistics 20*(2), 163–182.

Bahramian, P., M. Balcilar, R. Gupta, and P. T. Kanda (2014). Forecasting south african inflation using non-linear models: A weighted loss-based evaluation. Technical report.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bańbura, M., D. Giannone, and L. Reichlin (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics 25*(1), 71–92.

Bauwens, L., G. Koop, D. Korobilis, and J. V. Rombouts (2011). A comparison of forecasting procedures for macroeconomic series: the contribution of structural break models. *CIRANO-Scientific Publications* (2011s-13).

Blignaut, Z., G. Farrell, V. Munyama, and L. Rangasamy (2009). A note on the trimmed mean measure of core inflation in south africa. *South African Journal of Economics 77*(4), 538–552.

Bryan, M. F. and S. Cecchetti (1993). The consumer price index as a measure of inflation. *Economic Review* (Q IV), 15–24.

Bryan, M. F. and S. G. Cecchetti (1994). Measuring core inflation. In *Monetary Policy*, pp. 195–219. The University of Chicago Press.

Camba-Mendez, G. and G. Kapetanios (2005). Forecasting euro area inflation using dynamic factor measures of underlying inflation. *Journal of Forecasting 24*(7), 491–503.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of econometrics 86*(2), 221–241.

Clements, M. and D. Hendry (1998). *Forecasting economic time series*. Cambridge University Press.

Ding, S. and S. Karlsson (2014). Model averaging and variable selection in var models.

Du Plessis, S., G. Du Rand, and K. Kotzé (2015). Measuring core inflation in south africa. *South African Journal of Economics*.

Fagin, S. L. (1964). Recursive linear regression theory, optimal filter theory, and error analysis of optimal systems. In *IEEE International Convention Record*, Volume 12, pp. 216–245.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Science & Business Media.

Gupta, R., P. T. Kanda, M. P. Modise, and A. Paccagnini (2015). Dsge model-based forecasting of modelled and nonmodelled inflation variables in south africa. *Applied Economics 47*(3), 207–221.

Hallin, M. and R. Liška (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association 102*(478), 603–617.

Jazwinski, A. H. (2007). *Stochastic processes and filtering theory*. Courier Corporation.

Kapetanios, G. (2004). A note on modelling core inflation for the uk using a new dynamic factor estimation method and a large disaggregated price index dataset. *Economics Letters 85*(1), 63–69.

Koop, G. and D. Korobilis (2012). Forecasting inflation using dynamic model averaging*. *International Economic Review 53*(3), 867–886.

Koop, G. and D. Korobilis (2013). Large time-varying parameter vars. *Journal of Econometrics 177*(2), 185–198.

Koop, G. and S. M. Potter (2007). Estimation and forecasting in models with multiple breaks. *The Review of Economic Studies 74*(3), 763–789.

Morana, C. (2007). A structural common factor approach to core inflation estimation and forecasting. *Applied Economics Letters 14*(3), 163–169.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics 92*(4), 1004–1016.

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann (2006). Forecasting time series subject to multiple structural breaks. *The Review of Economic Studies 73*(4), 1057–1084.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski (2005). Using bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review 133*(5), 1155–1174.

Raftery, A. E., M. Kárnỳ, and P. Ettler (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics 52*(1), 52–66.

Ranchhod, S. et al. (2013). Measures of new zealand core inflation. *Reserve Bank of New Zealand Bulletin 76*(1).

Rangasamy, L. (2009). Inflation persistence and core inflation: The case of south africa. *South African Journal of Economics 77*(3), 430–444.

RiskMetrics (1996). Technical document. *New York-JP Morgan*.

Ruch, F. and D. Bester (2013). Towards a measure of core inflation using singular spectrum analysis. *South African Journal of Economics 81*(3), 307–329.

Stock, J. H. and M. W. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics 14*(1), 11–30.

Stock, J. H. and M. W. Watson (1999). Forecasting inflation. *Journal of Monetary Economics 44*(2), 293–335.

Sun, T. (2004). *Forecasting Thailand's core inflation*. Number 4-90. International Monetary Fund.

Woglom, G. (2005). Forecasting south african inflation. *South African Journal of Economics 73*(2), 302–320.