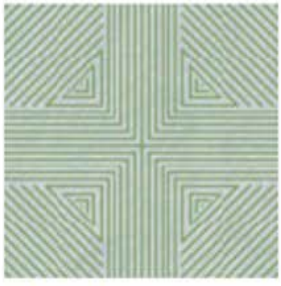
CENTRAL BANK OF
THE REPUBLIC OF AZERBAIJAN

WORKING PAPER SERIES
№01/2016

FORECASTING INFLATION IN POST-OIL BOOM YEARS: A CASE FOR NON-LINEAR MODELS?

Vugar Ahmadov, Shaig Adigozalov, Salman Huseynov,
Fuad Mammadov, Vugar Rahimov

06 April 2016



Forecasting inflation in post-oil boom years: A case for non-linear models?[1]

Vugar Ahmadov[2], Shaig Adigozalov[3], Salman Huseynov[4],

Fuad Mammadov[5], Vugar Rahimov[6]

6 April, 2016

ABSTRACT

In this study, we investigate relative performance of various non-linear models against that of an autoregressive model in forecasting future inflation. We find that non-linear models have trivial forecast superiority over the univariate autoregressive model in terms of central forecast accuracy. They also perform poorly when their forecasts are measured against those of the 3 variables VAR model. In addition, we also show that non-linear models cannot beat the random walk in terms of central forecast accuracy which is in line with the previous literature on Azerbaijan during the post-oil boom years. However, we also demonstrate that non-linear models still have clear forecast advantage over both linear and random walk models in predicting forecast density.

JEL classification: C11, C13, C32, C53
Keywords: Inflation; Forecasting; Bayesian methods; Non-linear models

[1] Authors would like to thank to Mehdi Mehdiyev, Ramiz Rahmanov and their colleagues at the CRD for their valuable comments and suggestions. Finally, the authors are also grateful to the BCC program, the SECO and the Geneva Graduate Institute for their support. The views expressed are those of the authors and do not necessarily reflect the views of the Central Bank of the Republic of Azerbaijan.
[2] Vugar Ahmadov – Center for Research and Development, Central Bank of Azerbaijan, email: vugar_ahmadov@cbar.az
[3] Shaig Adigozalov – Center for Research and Development, Central Bank of Azerbaijan, email: shaig_adigozalov@cbar.az
[4] Salman Huseynov – Centre for Research and Development, Central Bank of Azerbaijan and Institute of Control Systems, National Academy of Sciences, email: salman_huseynov@cbar.az
[5] Fuad Mammadov – Center for Research and Development, Central Bank of Azerbaijan, email: fuad_mammadov@cbar.az
[6] Vugar Rahimov – Center for Research and Development, Central Bank of Azerbaijan, email: vugar_rahimov@cbar.az

"Prediction is very difficult, especially about the future"
Niels Bohr

I. Introduction

Forecasting is in the heart of any decision making process, especially when it deals with far future and when its consequences cannot be undone later. Central banks are one of the places where forecasting is undertaken as a routine task and exercised on daily basis. The reason for this is obvious – any effects of central bank decisions on the economy can only be observed with longer time lags. However, forecasting is not an easy exercise – it requires considerable amount of time and resources. Failure to correctly forecast economic developments also imposes reputation costs on central banks and wreaks havoc on their credibility. But success at forecasting also depends on the intrinsic predictability of future economic developments. That is, the question is whether future economic developments are predictable given information set available at present?

In a very interesting paper, Hendry and Mizon (2013) discuss unpredictability in economic modeling and forecasting. They describe three important sources of unpredictability in economic analysis: intrinsic unpredictability, instance unpredictability and extrinsic unpredictability. They show that our workhorse forecast models (for instance, DSGE models) become unreliable when they are most needed. Edge and Gurkaynak (2010) demonstrate that DSGE models do not fare well in forecasting inflation during Great Moderation years. They show that no model is good at forecasting inflation – in fact, inflation become unforecastable during that period. This finding is in line with Stock and Watson (2007) results. They also show that despite significant decline in volatility during Great Moderation years, forecasting inflation has not become easier – its persistence has declined over time. Similarly, Atkenson and Ohanion (2001) find that since 1984, Phillips curve forecasts have lost their superiority against naïve inflation forecasting models. According to D'Agostino, Giannone and Surico (2006), the ability of relatively sophisticated models in predicting inflation and real activity in US has deteriorated remarkably during the Great Moderation period.

In an inflation forecasting exercise Huseynov, Ahmadov and Adigozalov (2014) show that almost similar results hold for an oil exporting country, namely Azerbaijan. They demonstrate that despite considerable decline in economic volatility during post-oil boom years, naïve models have gained significant advantage over more complex ones. In fact, according to them, it is impossible to beat a random walk forecast for inflation regardless of how sophisticated models one employs. However, in their forecasting study, they mostly utilize linear models and assume no changes in economic dynamics during that period. In contrary to their study, we investigate inflation forecasting in post-oil boom years using non-linear methods. We assess the relative performance of non-linear models and examine whether they can better forecast inflation during the same period.

For forecasting exercise we estimate various regime change models encompassing threshold models, smooth transition models, Markov switching models and time varying parameter models. We utilize both univariate and multivariate version of aforementioned methods and compare their performance with our baseline autoregressive model.

We use quarterly data and estimate our models for the period Q1 2003- Q4 2014. We find that for the post-oil boom period of Q1 2012-Q4 2014, non-linear models barely beat autoregressive model forecasts when they are judged with central forecast accuracy. The quality of their forecasts even fall inferior to mean forecasts generated by 3 variables VAR models. Most importantly, we also show that non-linear model cannot outperform naïve models forecasts, such as random walk forecasts, a finding that is in line with Huseynov, Ahmadov and Adigozalov (2014). They also find that it is very hard to beat random walk forecasts regardless of the degree of sophistication in models. That is, our results coupled with the evidences from the aforementioned study demonstrate that it is hard to explain inflation dynamics with different models at hand and in fact, inflation is unforecastable for this time period. However, we also find evidence that though non-linear models are not good at producing central forecast tendency, they can still be confidently used to give probability distributions to future outcomes of inflation.

The structure of the paper is designed as follows: the Section II discusses data and forecasting methods employed in this study, Section III presents results from forecast comparison experiment, Section IV discusses probable implications and causes of the change in inflation process, and finally, Section V concludes.

II. Data and Methodology

In this paper, we draw on quarterly data on domestic CPI, real non-oil GDP and reserve money covering the period Q1 2003 – Q4 2014. Domestic CPI and GDP figures are collected from the State Statistical Committee's Statistical Bulletins, while reserve money (in manat) is taken from the database of Central Bank of Azerbaijan. We exclude the value added of the mining sector from the overall GDP to obtain non-oil GDP figures. Note that quarterly real GDP figures are calculated at 2005 constant prices. All variables are seasonally adjusted applying TRAMO-SEATS package and quarter-on-quarter changes of these variables are obtained using seasonally adjusted figures.

In this paper, we employ univariate and multivariate (3 variables) non-linear time series methods and compare their forecast performance with that of our baseline autoregressive model, namely, an AR(2) process fitted to inflation. We estimate our models for the period Q1 2013 – Q4 2011 and keep the last 12 observations (Q1 2012 – Q4 2014) for the conduct of out-of-sample forecast. We use the iterated forecasting approach and recursive scheme to carry out forecasting exercise for horizons $h = \{1, 2, 4, 6\}$. We measure central forecast accuracy of each model using Root Mean Squared Forecast Error (RMSFE) and calculate relative RMSFE against

the baseline model. We also employ log predictive score to judge forecast accuracy of their predictive density against the baseline model. We choose the best lag order of each model based on Schwarz Information Criterion (SIC).

Markov-switching models

The first set of non-linear models that we employ is a Markov-switching (MS) model with 2 separate regimes. We label the first regime as *"low inflation regime" (regime 0)* and the second one as *"high inflation regime" (regime 1)*. Because we can determine which model parameter(s) to shift, we obtain various combinations of switching parameters, thus different specifications.

As a univariate MS model, we will employ an autoregressive *p* order (MS AR(p)) process for inflation y_t which can be specified as follows:

$$y_t = \mu_{S_t} + \sum_{i=1}^{p} \rho_{i,S_{t-i}} y_{t-i} + \sigma_{S_t} \varepsilon_t \tag{1}$$

where we define parameters in the above regression as:

$$\mu_{S_t} = \mu_0 (1 - S_t) + \mu_1 S_t = \mu_0 + (\mu_1 - \mu_0) S_t \tag{2}$$

$$\rho_{i,S_{t-i}} = \rho_{i,0} (1 - S_{t-i}) + \rho_{i,1} S_{t-i} = \rho_{i,0} + (\rho_{i,1} - \rho_{i,0}) S_{t-i} \tag{3}$$

$$\sigma_{S_t} = \sigma_0 (1 - S_t) + \sigma_1 S_t = \sigma_0 (1 + \frac{(\sigma_1 - \sigma_0)}{\sigma_0} S_t) = \sigma_0 (1 + h S_t) \tag{4}$$

where μ_{S_t} is the regression constant (μ_0 and μ_1 are the regression means for the first and the second regimes respectively), $\rho_{i,S_{t-p}}$ is the slope coefficient for the i^{th} order regression term, σ_{S_t} is the standard deviation, ε_t is a zero mean and unit variance shock, S_t, $t = 1,2,...,T$ is Markov-switching random variable taking value 1 at the *regime 1* and value 0 at the *regime 0*.

We will assume that the latent variable S_t follows a two-state, first order Markov-switching process with the following transition matrix:

$$P = \begin{bmatrix} p & 1-q \\ 1-p & q \end{bmatrix} \tag{5}$$

where the probability of switching from the *regime 0* to the *regime 0* is $\Pr(S_t = 0 \mid S_{t-1} = 0) = p$ and the probability of switching from the *regime 1* to the *regime 1* is $\Pr(S_t = 1 \mid S_{t-1} = 1) = q$. Note that in our specification transition probabilities are assumed to be time-invariant and constant over time.

By determining which regression parameters to switch, we can obtain different specifications for the above regression equation (1). In the first specification we only allow intercept of the

regression to shift but other parameters to be the same across two regimes. That is, only the equation (2) is valid for this specification, and other parameters of the regression remains the same across two regimes, i.e., $\rho_{i,0} = \rho_{i,1} = \rho_i$ for all autoregressive terms and $\sigma_0 = \sigma_1 = \sigma$ for both regimes. Note that this specification is different from Hamilton (1989) where he defines a mean adjusted autoregressive process as follows:

$$(y_t - \mu_{S_t}) = \sum_{i=1}^{p} \rho_i (y_{t-i} - \mu_{S_{t-i}}) + \sigma\varepsilon_t \tag{6}$$

Although the equation (5) is defined for 2 regimes, due to its autoregressive nature one needs to take into account 2^{p+1} parameter combinations. Hamilton (1994) shows (chapter 22) how to re-write this specification and re-define transition matrix P in terms of 2^{p+1} regimes. However, our first specification with 2 regimes allows only 2 different parameter combinations in contrary to Hamilton (1989), so we do not need to re-define the transition matrix.

In the second specification for $AR(p)$ process, we assume that only intercept and volatility are changing across regimes. In other words, we assume that equations (2) and (4) applies to the second specification, but autoregressive coefficients remain constant, i.e., $\rho_{i,0} = \rho_{i,1} = \rho_i$ for all lagged terms. Note that as in the previous specification with two-state Markov-switching regimes, we need to take into account 2 different parameter combinations despite the autoregressive nature of the process.

In the third specification, we allow all parameters to shift across 2 regimes. Because of the autoregressive nature of the process and the definition of autoregressive coefficients, we need to take into account 2^{p+1} parameter combinations. Therefore, we transform the two-state transition matrix P in equation (5) into a new transition matrix with 2^{p+1} different regimes. Note that one can sidestep this complication by defining the autoregressive terms as follows:

$$\rho_{i,S_{t-i}} = \rho_{i,0}(1 - S_t) + \rho_{i,1} S_t = \rho_{i,0} + (\rho_{i,1} - \rho_{i,0}) S_t \tag{7}$$

In addition to our univariate AR specifications, we also model non-oil output, reserve money and inflation by employing *pth* order VAR specification with a two-state Markov-switching process as follows:

$$Y_t = \mu_{S_t} + \sum_{i=1}^{p} A_i Y_{t-i} + \Omega_{S_t}^{1/2} \varepsilon_t \tag{8}$$

where Y_t is a *3x1* vector, μ_{S_t} is a regime dependent mean vector, A_i is a corresponding autoregressive coefficient matrix for the *i* period lagged vector term Y_{t-i} and Ω_{S_t} is a regime dependent variance-covariance matrix.

In the VAR framework, we prepare inflation forecasts using two different specifications: (i) in the first specification, we only allow structural breaks (or regime changes) in the mean of the

regression and (ii) in the second specification, we allow joint shifts in the mean as well as in the variance-covariance matrix of the regression. Note that in both specifications autoregressive coefficient matrix do not change across regimes.

In all of our AR and VAR specifications, we apply Bayesian methods and Gibbs sampling algorithm to obtain posterior distribution of model parameters and forecast density. We use Minnesota type priors for coefficient vectors. For $AR(p)$ specifications, we set the prior mean for the coefficient of the first lag to 0.5 and for other lags prior values are set to 0. The hyperparameters of priors are set to values mostly employed in the literature (Canova (2007), p.380) $\lambda_1 = 0.2, \lambda_2 = 0.5\left(for\ VAR\right), \lambda_3 = 1$ except $\lambda_4 = 10$. For VAR specifications, the same priors are employed (see Appendix for more details).

Threshold and smooth transition models

The second set of non-linear models that we employ in our forecasting exercise is threshold and smooth transition models. As in MS specifications, we employ these models to describe 2 different regimes in the economy. Parameter switches across different regimes are governed by a threshold variable. In these models, though threshold variable is known, a difficulty emerges due to an unobserved threshold level. In our exercises, the threshold variable is a lagged term of the variable in AR or VAR specifications – models that are also labeled as self-exciting threshold models.

A two-regime threshold autoregressive process ($AR(p)$) can be described as follows:

$$y_t = [\mu_0 + \sum_{i=1}^{p} \rho_{i,0} y_{t-i} + \sigma_0 \varepsilon_t] S_t + [\mu_1 + \sum_{i=1}^{p} \rho_{i,1} y_{t-i} + \sigma_1 \varepsilon_t](1 - S_t) \tag{9}$$

where

$$S_t = \begin{cases} 1 if\ y_{t-d} > c \\ 0 if\ y_{t-d} \le c \end{cases} \tag{10}$$

where d is a positive integer and c is a threshold value. As in the case of the *MS AR(p)* process, we define three specifications (i) only regression intercept shifts (ii) intercept and volatility shift (iii) all regression parameters shift. Note that equation (9) depicts shifts in all parameters of the regression (third specification) and other specifications can be obtained by proper adjustments in the specification above.

A self-exciting threshold VAR model can be expressed as follows:

$$Y_t = [\mu_0 + \sum_{i=1}^{p} A_i Y_{t-i} + \Omega_0^{1/2} \varepsilon_t] S_t + [\mu_1 + \sum_{i=1}^{p} A_i Y_{t-i} + \Omega_1^{1/2} \varepsilon_t](1 - S_t) \tag{11}$$

where

$$S_t = \begin{cases} 1 \, if \; Y_{i,t-d} > c \\ 0 \, if \; Y_{i,t-d} \leq c \end{cases} \tag{12}$$

where $Y_{i,t-d}$ is a *d* period lagged *ith* variable of the vector Y_t.

In the smooth transition VAR model, the specification (11) also applies, but now the state vector S_t is a continuous variable given by:

$$S_t = \begin{cases} 1 - \dfrac{1}{1 + \exp(-\gamma(Y_{i,t-d} - c))} \\ \dfrac{1}{1 + \exp(-\gamma(Y_{i,t-d} - c))} \end{cases} \tag{13}$$

These models can also be estimated using classical methods (see, for instance, Tsay (1989, 1998), Terasvirta (1994)), for example, non-linear OLS or Maximum Likelihood methods. However, we prefer to apply Bayesian methods (see Chen and Lee (1995), Chen (1998)) using both Gibbs and Metropolis-Hasting algorithms (for more details, see the Appendix).

We use the same priors for coefficient vectors in our AR and VAR specifications as in the MS case. Besides, we draw a candidate threshold value from $c_{new} = c_{old} + \Psi^{1/2}\xi$, $\xi \sim N(0,1)$ in the Metropolis-Hastings step. We choose the scaling factor Ψ to ensure that the acceptance rate remains between 15% and 55%. The priors for mean and variance of the threshold value are obtained from a training sample undertaken with the first twelve observations. In our exercises, we assume a flat prior for delay parameter *d*, but restrict its maximum value to 4. In addition, for smooth transition VAR, we assume a gamma distribution for γ, its hyperparameters being set to 1.25 and 1 respectively.

Time varying parameter models

The third set of non-linear models that we utilize in our forecasting exercise is a time varying parameter (TVP) AR model with stochastic volatility. Huseynov, Ahmadov and Adigozalov (2014) employ TVP-VAR models in their forecasting exercise. But due to large number of variables in their TVP-VAR (30 variables), they opt to use forgetting factor algorithm proposed by Koop and Korobilis (2013) in their exercise. Here, as in we will use Bayesian methods (see, for instance, Primiceri (2004), Barnett, *et al* (2012), Baumeister, *et al* (2010)) and MCMC algorithm in our TVP-AR model.

A TVP-AR process with stochastic volatility can be expressed as follows:

$$y_t = \mu_t + \sum_{i=1}^{p} \rho_{i,t} y_{t-i} + \sigma_t \varepsilon_t \tag{14}$$

Letting $\theta_t = \{\mu_t, \rho_{i,t}\}$ AR parameters evolve as random walks:

$$\theta_t = \theta_{t-1} + \zeta_t \tag{15}$$

where $\zeta_t \sim N(0, Q)$. The variance of the error term evolves as:

$$\ln \sigma_t^2 = \ln \sigma_{t-1}^2 + \upsilon_t \tag{16}$$

$$\upsilon_t \sim N(0, g) \tag{17}$$

We can estimate AR model using Bayesian methods and applying Carter-Kohn and Metropolis-Hastings algorithm (see Blake and Mumtaz (2012)).

III. Forecast Comparisions

We measure forecast accuracy of each model using Root Mean Squared Forecast Error (RMSFE) and log predictive score. By construction, RMSFE shows the central forecast accuracy of the model and defined as follows:

$$RMSFE_{m,h} = \sqrt{\frac{1}{N_h} \sum_{t=T}^{T+N_h-1} (\pi_{t+1} - \pi_{t+1}^{m,h})^2} \tag{18}$$

where N_h is the number of time periods *h-period ahead* forecast is evaluated , π_{t+1} denotes the actual value of quarterly inflation at the forecast evaluation period *t+1*, and $\pi_{t+1}^{m,h}$ *h*-period ahead forecast (made *h* periods in the past) for model *m*. The relative forecasting strength of each model for *h* period ahead forecast is calculated based on the relative RMSFE using the AR (2) specification for inflation as our baseline models:

$$RMSFE_h^{REL} = \frac{RMSFE_{m,h}}{RMSFE_{b,h}} \tag{19}$$

where *b* denotes baseline model. The model with superior forecasting power should possess a relative RMSFE value less than unity.[7]

The second criterion for forecast accuracy that we appeal to is log predictive score. We apply it to compare the quality of probabilistic forecasts by giving a numerical value employing the whole predictive distribution and the event that realizes. Based on the joint predictive density function of $y_{t+1}, y_{t+2}, ..., y_{T+h}$, it is expressed as:

[7] There are more formal methods to test for equal forecast accuracy of different models or forecast encompassing, either nested or non-nested models under different forecasting schemes (see for example, Deibold and Mariano (1995), Giacomini and White (2006), Clark and McCracken (2001), McCracken (2004), etc.).

$$S(h,m) = \sum_{t=T}^{T+N_h-1} \log p(y_{t+1},...,y_{t+h} \mid Y_t, m) \tag{20}$$

We report the difference between the log score of the model *m* and our baseline model. If the difference is positive, then this implies that the model *m* outperforms the baseline model in terms of predictive density accuracy.

Table 1. Relative RMSFE and difference of log score for different models (baseline AR(2))

		RMSFE				Log score			
		1Q	**2Q**	**4Q**	**6Q**	**1Q**	**2Q**	**4Q**	**6Q**
TVP-AR	**all**	0.875	0.874	0.873	0.956	0.194	0.148	0.050	0.009
MS-AR	**const**	1.109	1.132	1.147	1.130	0.875	0.820	0.789	0.714
	const+vol	0.975	0.970	0.970	1.024	0.990	0.969	1.021	0.914
	all	0.955	0.955	1.029	1.084	0.980	0.980	0.925	0.859
MS-VAR	**const**	1.343	1.086	1.097	1.225	-0.762	-0.870	-1.261	-1.591
	const+vol	0.975	0.799	0.845	0.886	-1.245	1.337	1.803	-2.175
TAR	**const**	0.951	0.992	1.117	1.066	0.076	0.056	-0.009	-0.002
	const+vol	0.950	0.990	1.117	1.109	0.317	0.279	0.110	0.124
	all	1.001	0.992	1.098	1.089	0.324	0.286	0.098	-0.023
TVAR	**all**	0.969	0.929	0.904	0.931	0.441	0.524	0.328	0.125
STVAR	**all**	1.541	1.388	1.383	1.317	-1.134	-1.038	-1.051	-1.039

Note: MS-(V)AR denotes Markov-switching (Vector)Autoregressive models. *"const"* specification allows switches only in intercept of regression, *"const + vol"* switches both in intercept and volatility, whereas *"all"* specifies switches in all parameters of the regression. Similarly, "TAR" is an acronym for Threshold Autoregressive (AR) models, "TVP-AR" is time varying parameter AR process, "STVAR" denotes smooth transition VAR models.

Table 1 reports the relative RMSFE and the difference of the log scores of the models with respect to our baseline model AR(2). According to the relative RMSFE criteria, univariate Markov-switching and threshold models (*"all"* and *"const+vol"* specifications) demonstrate some advantage over the baseline model in predicting inflation for 1 and 2 quarters ahead forecast horizon. However, the gain in forecast accuracy for those models is not that substantial and hovers around 3%-5%. For longer horizons those models loss their superiority in terms of forecast accuracy against the baseline model. In contrary, it seems that our univariate TVP-AR specification with stochastic volatility which assumes time varying parameter for all regression parameters do perform well in comparison to the baseline model. It exhibits relative gain in forecast accuracy around 13% for all forecast horizons except for 6 quarters ahead where its superiority is around 5%.

In addition, it seems that the first specification (*"const"*) of both MS-AR and TAR models where we only allow intercepts of regression to switch, but keep other parameters of the

regression the same across regimes is inferior to the baseline model. Less accuracy of the first specification (*"const"*) is also applicable for multivariate MS-VAR specification and almost true regardless of the forecast horizon.

Regarding multivariate models, MS-VAR (*"const+vol"* specification) and TVAR (*"all"* specification) model forecasts can beat the baseline model forecasts for all forecast horizons. In the MS-VAR model the maximum gain in forecast accuracy is nearly 21% (for 2 quarters ahead) while in TVAR model it is around 10% (for 4 quarters ahead). However, this cannot be applied to the STVAR model forecasts which fall inferior to those of the baseline model for all forecast horizons.

Relative RMSFE results helps to reveal central tendency in forecast accuracy of each model in comparison to the baseline model. Beside that we also appeal to difference of log scores to measure predictive density accuracy of each model. Table 1 shows that all specifications of univariate MS-AR models possess clear advantage over the baseline model in terms of this forecast accuracy criterion. All specifications of the univariate TAR as well as TVP-AR models also demonstrate some advantage over the baseline model.

Except the TVAR model, all multivariate models fare worse in terms of density forecast accuracy relative to the baseline model. Among them, multivariate Markov-switching models are the worst performers (the worst is *"const"* specification) according to this criterion.

Overall, though univariate regime switching models do not exhibit substantial gains in terms of central forecast accuracy, their superiority in terms of predictive density accuracy is significantly larger. It seems that univariate TVP-AR model performs well in terms of both central forecast and predictive density accuracy. In general, multivariate models are not good at predictive density accuracy (except the TVAR model) and they also do not possess significant advantage in terms of central forecast accuracy except the second specification of MS-VAR. The finding that non-linear models generate noisier central forecasts but demonstrate clear advantage over linear models is in line with Alessandri and Mumtaz (2013). They also show that non-linear models clearly outperform linear ones in predicting distributions, but generate inferior central forecasts.

IV. Discussion

In the previous section, we demonstrate that non-linear models do possess clear advantage in predicting distributions for the forecast exercise period Q1 2011 - Q4 2014. We also collect evidence that some univariate and multivariate models outperform the baseline model in terms of central forecast accuracy.

It is also interesting to assess forecast performance of non-linear models against a multivariate linear model. For this comparison, we estimate a VAR specification using classical regression methods with the same 3 variables also being employed in the non-linear multivariate models. Table 2 in the Appendix A summarizes the results for this exercise. Clearly, only one non-linear model, namely TVP-AR model can barely outperform VAR central forecasts. No other univariate or multivariate models can even enjoy a minor advantage in forecast accuracy against the VAR model forecasts. However, as in the case of our univariate baseline model, non-linear models demonstrate clear superiority in terms of predictive density against the baseline VAR model.

For almost the same period, Huseynov, Ahmadov and Adigozalov (2014) show that it is difficult to beat a random walk or naïve forecast employing relatively sophisticated models. Table 3 in the Appendix A displays the results where we compare forecast performances of non-linear models against the random walk forecast. Obviously, no non-linear model can outperform random walk forecast for any forecast horizon when central tendency is used as an accuracy measure. This finding is not surprising given their minor advantage over the linear models. However, non-linear models do in fact deliver superior performance compared to the random walk forecast in predicting distributions.

Figure 1. Regimes from TAR model

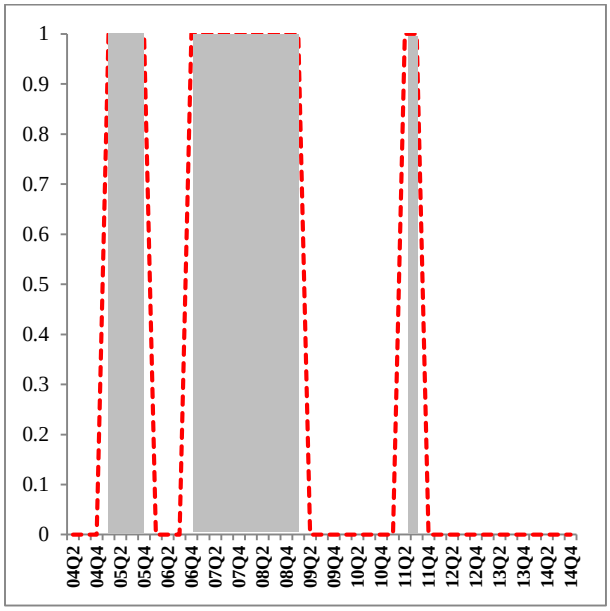
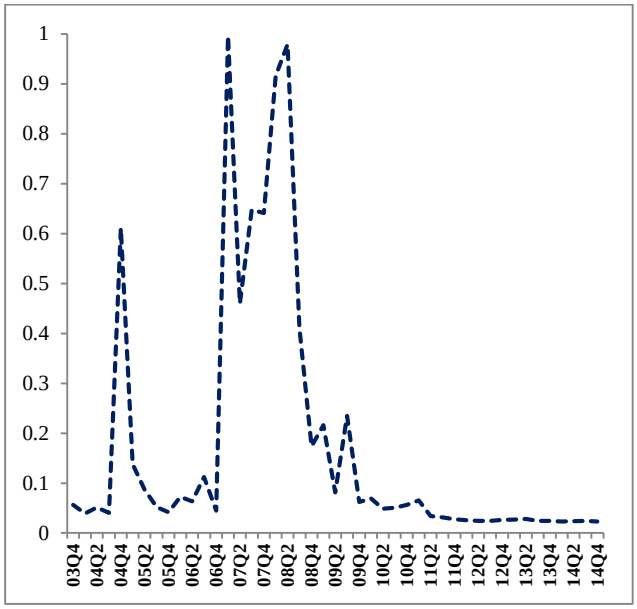


Figure 2. Regimes from MS-AR model



Note: Figure 1 depicts 2 different regimes obtained from TAR model specification with constant and volatility switching, but other paramers remaining the same across regimes. TAR model regimes takes 2 different discrete values. *Regime 0 (low inflation regime)* expresses the first regime and takes the value 0 whereas *regime 1 (high inflation regime)* defines the second regime and takes the value 1. Similarly, Figure 2 depicts filtered probabilities obtained from Markov-switching model with 2 regimes allowing only constant and volatility switching. Filtered probabilities display the probability of the *regime 1 (high inflation regime)*. Because there are only 2 regimes, the probabilities of the *regime 0* can be obtained by deducting filtered probabilities from one.

Therefore, as Huseynov, Ahmadov and Adigozalov (2014), we also find that random walk forecasts are very difficult to beat for that period. Huseynov, Ahmadov and Adigozalov (2014) show that there might occur some structural changes in the inflation or economic dynamics in the country during the years of 2003-2014. We also demonstrate that inflation dynamics show some traces of changes over the time period that we investigate. This pattern is clearly captured by our MS-AR or TAR specifications (Figures 1-2). Similarly, the aforementioned study also finds strong evidence of the probable regime changes during the same period. However, as our exercise with non-linear models also put forward capturing this change is not sufficient to gain forecast advantage over the random walk forecasts. When we combine our finding with those of Huseynov, Ahmadov and Adigozalov (2014) it seems that inflation become unforecastable for that period for linear as well as non-linear models when central tendency is utilized as an accuracy measure. But the good news is that non-linear models can exhibit clear advantage over random walk as well as linear model forecasts in predicting distributions. That is, even it is difficult to forecast central tendency in inflation process for quarters ahead, one can still confidently attach a non-trivial probability to future realizations of inflation using non-linear models that we develop here.

V. Conclusion

In this study, we estimate various non-linear models accounting for two different regimes and compare their forecasting performance with that of linear models and random walk forecasts. We find that non-linear models produce noisier mean forecasts and experience hard times in beating linear model forecasts when central forecast tendency is used as an accuracy measure. They even lose to the forecasts generated by naïve models such as random walk or no-change forecasts. This finding is in line with Huseynov, Ahmadov and Adigozalov (2014) who also show that random walk forecasts are hard to beat for the similar time frame. Though non-linear models generate noisier central forecasts they outperform linear model and random walk forecasts in predicting distributions. That is, though non-linear models are not good at predicting central forecast tendency, they can still be used confidently to attach a non-trivial probability to future inflation outcomes.

References

Alessandri, Piergiorgio, and Mumtaz, Haroon, 2013, "Financial conditions and density forecasts for US output and inflation", Bank of England, Center for Central Bank Studies, Joint Research Paper No. 4

Atkeson, Andrew, and Lee E. Ohanian, 2001, "Are Phillips curves useful for forecasting inflation?", *Federal Reserve Bank of Minneapolis Quarterly Review,* Vol. 25(1), pp. 2-11

Barnett, Alina, Mumtaz, Haroon, and Theodoridis, Konstantinos, 2012, "Forecasting UK GDP growh, inflation and interest rates under structural change: a comparison of models with time-varying parameters", Bank of England WP no. 450

Baumeister, Christiane, Liu, Philip, and Mumtaz, Haroon, 2010, "Changes in the transmission of monetary policy: evidence from a time-varying factor-augmented VAR", Bank of England WP no. 401

Blake, Andrew, and Mumtaz, Haroon, 2012, *Applied Bayesian Econometrics for Central Bankers,* Bank of England, Center for Central Banking Studies, Technical Handbook no. 4

Canova, Fabio, 2007, *Methods for Applied Macroeconomic Research,* Princeton University Press

Chen, Cathy W.S., 1998, "A Bayesian analysis of generalized threshold autoregressive models", *Statistics and Probability Letters,* Vol. 40, pp. 15-22

Chen, Cathy, W.S., and Lee, Jack E., 1995, "Bayesian inference of threshold autoregressive models", *Journal of Time Series Analysis,* Vol. 16(5) pp. 483-492

D'Agostino, Antonello, Giannone, Domenico, and Surico, Paolo, 2006, "(Un)predictability and macroeconomic stability", ECB Working Paper no. 605

Edge, Rochelle M., and Gürkaynak, Refet S., 2010, "How useful are estimated DSGE model forecasts for central bankers?", *Brookings Papers on Economic Activity,* Vol. 41(2), pp.209-259

Hamilton, James, 2014, "Macroeconomic regimes and regime switches", *Handbook of Macroeconomics,* Vol. 2

Hamilton, James, 1994, *Time Series Analysis,* Princeton University Press

Hamilton, James, 1989, "A new approach to the economic analysis of nonstationary times series and business cycle", *Econometrica,* Vol. 57(2), pp. 357-384

Hendry, David F., and Mizon, Grayham E., 2014, "Unpredictability in economic analysis, econometric modeling and forecasting", *Journal of Econometrics,* Vol. 182, pp. 186-195

Huseynov, Salman, and Ahmadov, Vugar, Adigozalov, Shaig, 2014, "Beating a random walk: "Hard Times" for forecasting inflation in post-oil boom years?", MPRA paper 63515

Groen, Jan, and Mumtaz, Haroon, 2008, "Investigating the structural stability of the Phillips curve relationship", Bank of England WP no. 350

Kim, Chang-Jin, and Nelson, Charles R., 1999, *State Space Models with Regime Switching,* MIT Press

Koop, Gary and Korobilis, Dimitris, 2013, "Large time varying parameter VARs", *Journal of Econometrics,* Vol. 177(2), pp. 185-198

Krolzig, Hans-Martin, 1997, *Markov Switching Vector Autoregressions,* Springer

Primiceri, Giorgio E., 2004, "Time varying structural vector autoregression and monetary policy", *Review of Economic Studies,* Vol. 72(3), pp. 821-852

Simon, John, 1996, "A Markov-switching model of inflation in Australia", Reserve Bank of Australia, Discussion Paper no. 9611

Sims, Christopher and Zha, Tao, 2005, "Were there regime switches in US monetary policy?", CEPS WP no. 110

Stock, James H., and Watson, W. Mark, 2007, "Why has US inflation become harder to forecast?", *Journal of Money, Credit and Banking,* Vol. 39(1), pp. 3-33

Terasvirta, Timo, 1994, "Specification, estimation and evaluation of smooth transition autoregressive models", *Journal of the American Statistical Association,* Vol. 89(425), pp. 208-218

Tsay, Ruey S., 1998, "Testing and modeling multivariate threshold models", *Journal of the American Statistical Association,* Vol. 93(443), pp. 1188-1202

Tsay, Ruey S., 1989, "Testing and modeling threshold autoregressive processes", *Journal of the American Statistical Association,* Vol. 84(405), pp. 231-240

APPENDIX A: Tables

Table 2. Relative RMSFE and difference of log scores for different models (baseline VAR(2))

		RMSFE				Log score			
		1Q	**2Q**	**4Q**	**6Q**	**1Q**	**2Q**	**4Q**	**6Q**
TVP-AR	**all**	0.993	1.025	0.989	0.966	0.172	0.052	0.147	0.115
MS-AR	**const**	1.258	1.328	1.299	1.142	0.853	1.328	1.299	1.142
	const+vol	1.106	1.137	1.099	1.035	0.968	0.873	1.126	1.038
	all	1.083	1.120	1.165	1.095	0.958	0.880	1.030	0.983
MS-VAR	**const**	1.523	1.274	1.243	1.238	-1.267	-1.433	-1.699	-2.051
	const+vol	1.106	0.937	0.958	0.895	-0.783	-0.966	-1.157	-1.470
TAR	**const**	1.136	1.163	1.244	1.101	0.055	-0.041	0.095	0.122
	const+vol	1.078	1.161	1.266	1.120	0.295	0.183	0.214	0.137
	all	1.079	1.163	1.266	1.078	0.302	0.190	0.202	0.101
TVAR	**all**	1.099	1.090	1.024	0.941	0.420	0.428	0.432	0.250
STVAR	**all**	1.748	1.628	1.566	1.331	-1.155	-1.134	-0.947	-0.915

Note: MS-(V)AR denotes Markov-switching (Vector)Autoregressive models. *"const"* specification allows switches only in intercept of regression, *"const + vol"* switches both in intercept and volatility, whereas *"all"* specifies switches in all parameters of the regression. Similarly, "TAR" is an acronym for Threshold Autoregressive (AR) models, "TVP-AR" is time varying parameter AR process, "STVAR" denotes smooth transition VAR models.

Table 3. Relative RMSFE and difference of log scores for different models (baseline RW)

		RMSFE				Log score			
		1Q	**2Q**	**4Q**	**6Q**	**1Q**	**2Q**	**4Q**	**6Q**
TVP-AR	**all**	1.688	1.677	1.548	2.342	0.263	-0.045	-0.258	-0.311
MS-AR	**const**	2.139	2.171	2.033	2.768	0.944	0.627	0.488	0.412
	const+vol	1.881	1.859	1.720	2.508	1.059	0.777	0.720	0.612
	all	1.842	1.832	1.824	2.654	1.049	0.783	0.624	0.556
MS-VAR	**const**	2.589	2.083	1.945	2.999	-1.176	-1.529	-2.105	-2.477
	const+vol	1.880	1.533	1.498	2.169	-0.693	-1.063	-1.563	-1.893
TAR	**const**	1.931	1.901	1.946	2.667	0.146	-0.137	-0.311	-0.304
	const+vol	1.833	1.899	1.981	2.715	0.386	0.086	-0.191	-0.290
	all	1.834	1.902	1.981	2.612	0.393	0.093	-0.204	-0.325
TVAR	**all**	1.869	1.782	1.602	2.280	0.510	0.331	0.026	-0.177
STVAR	**all**	2.972	2.662	2.451	3.226	-1.065	-1.230	-1.353	-1.342

Note: MS-(V)AR denotes Markov-switching (Vector)Autoregressive models. *"const"* specification allows switches only in intercept of regression, *"const + vol"* switches both in intercept and volatility, whereas *"all"* specifies switches in all parameters of the regression. Similarly, "TAR" is an acronym for Threshold Autoregressive (AR) models, "TVP-AR" is time varying parameter AR process, "STVAR" denotes smooth transition VAR models.

APPENDIX B: Markov-switching models

Consider an MS AR model:

$$y_t = \mu_{S_t} + \sum_{i=1}^{p} \rho_{i,S_{t-i}} y_{t-i} + \sigma_{S_t} \varepsilon_t \tag{B-1}$$

where S_t follows a two-state Markov chain. Collecting $\theta = \{\mu_{S_t}, \rho_{1,S_{t-1}}, \rho_{2,S_{t-2}}, ..., \rho_{p,S_{t-p}}\}$ as the Gibbs-algorithm is cycled through the following steps:

1. Given starting values for AR parameters and variance as well as transition probabilities, the latent state vector S_t is obtained using multi-move Gibbs sampling drawn from $f(\breve{S}_T \mid \tilde{Y}_T, \mu_{S_t}, \rho_{1,S_t}, \rho_{2,S_t}, ..., \rho_{p,S_t}, \sigma_{S_t}, p, q)$ where $\tilde{S}_T = [S_1, S_2, ..., S_T]$ and $\tilde{Y}_T = [Y_1, Y_2, ..., Y_T]$. Kim and Nelson (1999, Chapter 9) show that the Markov property of the state variable implies that

$$f(\breve{S}_T \mid \tilde{Y}_T) = f(S_T \mid Y_T) \prod_{t=1}^{T-1} f(S_t \mid S_{t+1}, \tilde{Y}_T) \tag{B-2}$$

This density can be simulated in two steps:

(a) Run Hamilton (1989) filter to get $f(S_t \mid \tilde{Y}_T), t = 1,2,3..., T$ and save them. The last iteration of the filter provides us with $f(S_T \mid \tilde{Y}_T)$ from which S_T is generated.

(b) To generate S_t conditional on $\tilde{Y}_T$ and $S_{t+1}, t = T-1, T-2, ..., 1$, we use the following result:

$$f(S_t \mid \tilde{Y}_T) \propto f(S_{t+1} \mid S_s) f(S_t \mid \tilde{Y}_t) \tag{B-3}$$

where $f(S_{t+1} \mid S_s)$ is a transition probability, and $f(S_t \mid \tilde{Y}_t)$ has been saved from step (a). Kim and Nelson(1999) show that one can sample S_t from (B-3) as follows:

(i) first, calculate $\Pr(S_t = 1 \mid S_{t+1}, \tilde{Y}_T)$ in the following way:

$$\Pr(S_t = 1 \mid S_{t+1}, \tilde{Y}_T) = \frac{f(S_{t+1} \mid S_t = 1) f(S_t = 1 \mid \tilde{Y}_T)}{\sum_{j=0}^{1} f(S_{t+1} \mid S_t = j) f(S_t = j \mid \tilde{Y}_T)} \tag{B-4}$$

(ii) generate a random number from a uniform distribution. If the generated number is less than or equal to the calculated value of $\Pr(S_t = 1 \mid S_{t+1}, \tilde{Y}_T)$, set $S_t = 1$. Otherwise, set it to 0.

2. Conditional on $\tilde{S}_T$, transition probabilities *p* and *q* are independent of the data set. Using beta distributions as conjugate priors for transition probabilities, Kim and Nelson (1999, Chapter 9) show how to draw transition probabilities from posterior beta distribution. Given $\tilde{S}_T$

vector from the previous step, draw transition probabilities from two independent beta distributions:

$$p \mid \tilde{S}_T \sim beta(u_{00} + n_{00}, u_{01} + n_{01}) \quad \text{(B-5)}$$

$$q \mid \tilde{S}_T \sim beta(u_{11} + n_{11}, u_{10} + n_{10}) \quad \text{(B-6)}$$

where u_{ij} are hyperparameters of the priors and n_{ij} refers to transition from state *i* to *j*, which can be easily counted given $\tilde{S}_T = [S_1, S_2, ..., S_T]$.

3. Conditional on θ, $\tilde{S}_T$ and $\tilde{Y}_T$, sample σ_0^2 and σ_1^2 from inverted Gamma distribution. Recall that $\sigma_{S_t}^2 = \sigma_0^2(1 + hS_t)$ and $\sigma_1^2 = \sigma_0^2(1 + h)$. First, one can generate σ_0^2 conditional on h and then generate $\tilde{h} = 1 + h$ conditional on σ_0^2.

   One can draw variances from posterior distribution as follows:

   (i) To generate σ_0^2 conditional on h, divide both sides of equation (A-1) by $\sqrt{(1 + hS_t)}$. Given an inverted Gamma distribution as a conjugate prior for σ_0^2, the posterior distribution for σ_0^2 takes the form:

   $$\sigma_0^2 \mid h, \theta, \tilde{S}_T, \tilde{Y}_T \sim IG(\frac{\nu_1}{2}, \frac{\delta_1}{2}) \quad \text{(B-7)}$$

   where $\nu_1 = \nu_0 + T$, $\delta_1 = \delta_0 + e^{*\prime} e^{*}$, ν_0 and δ_0 are respective priors from Inverted Gamma distribution, T is the sample size and e^{*} is regression residuals calculated from adjusted equation (B-1).

   (ii) To generate $\tilde{h} = 1 + h$ conditional on σ_0^2, divide both sides of the equation (B-1) by σ_0. Given a conjugate prior for $\tilde{h}$, the posterior distribution is as follows:

   $$\tilde{h} \mid \sigma_0^2, \theta, \tilde{S}_T, \tilde{Y}_T \sim IG(\frac{\nu_3}{2}, \frac{\delta_3}{2}) \quad \text{(B-8)}$$

   where $\nu_3 = \nu_2 + T_1$, $\delta_3 = \delta_2 + e^{**\prime} e^{**}$, ν_2 and δ_2 are respective priors from Inverted Gamma distribution, T_1 is the cardinality of the set $N_1 = \{t : S_t = 1\}$. and e^{**} is regression residuals calculated from adjusted equation (B-1) for the values of y_t for which $S_t = 1$.

   Once $\tilde{h} = 1 + h$ is generated from the above posterior distribution σ_1^2 is calculated by $\sigma_1^2 = \sigma_0^2(1 + h)$.

4. Conditional on $\tilde{S}_T$, $\tilde{Y}_T$, σ_0^2 and σ_1^2, sample θ from posterior distribution as follows:

   (i) First, divide both sides of the equation (B-1) by σ_{S_t} .

   (ii) Assuming Minnesota priors for coefficient vector, draw coefficient vector from normal distribution:

$$\theta \mid \sigma_0^2, \sigma_1^2, \tilde{S}_T, \tilde{Y}_T \sim N(b_1, B_1) \tag{B-9}$$

   where

$$b_1 = (B_0^{-1} + X'X)^{-1}(B_0^{-1}b_0 + X'Y) \tag{B-10}$$

$$B_1 = (B_0^{-1} + X'X)^{-1} \tag{B-11}$$

   where Y is $T \times 1$ vector including left hand side (LHS) variable y_t and X is a $T \times 2(p+1)$ matrix including p lagged terms of LHS variable, their elementwise multiplication with $\tilde{S}_T$, a vector of ones and $\tilde{S}_T$, b_0 is the mean and B_0 is the variance-covariance of the prior distribution respectively. Recall that because of our specification for autoregressive terms, one needs to take lags of $\tilde{S}_T$ vector and then implement elementwise multiplication.

Note that the above algorithm is designed for our third specification allowing all parameters of the regression to switch across regimes. For the first and second specification, one needs to make minor adjustments to this algorithm accordingly.

Recall that in our third specification, we need to take into account 2^{p+1} parameter combinations. Thus, we transform transition matrix P in equation (5) into $2^{p+1} \times 2^{p+1}$ matrix and run Hamilton (1989) filter in the step 1 (a) and obtain $f(S_t = 1 \mid \tilde{Y}_T)$ by summing respective rows (in our case, even rows) of the output from Hamilton (1989) filter.

The Gibbs algorithm for MS VAR models is also similar to that of MS AR algorithm, except here one works with vector of variables rather than a single variable. One can refer to Krolzig (1997, Chapters 2, 8, 9) and Canova (2007, Chapter 11) for more details.

APPENDIX C: Threshold models

Consider a threshold AR model:

$$y_t = I(y_{t-d} <= c)(\mu_0 + \sum_{i=1}^{p} \rho_{i,0} y_{t-i} + \sigma_0 \varepsilon_t) + I(y_{t-d} > c)(\mu_1 + \sum_{i=1}^{p} \rho_{i,1} y_{t-i} + \sigma_1 \varepsilon_t)$$

where $I(\cdot)$ is an indicator function. Note that in our STAR model $I(\cdot) = I(\frac{1}{1+\exp(-\gamma(Y_{i,t-d} - c))})$

Let's collect regression parameters into a vector $\theta = \{\mu_0, \rho_{1,0}, \rho_{2,0}, ..., \rho_{p,0}, \mu_1, \rho_{1,1}, \rho_{2,1}, ..., \rho_{p,1}\}$.

Applying Bayesian methods, one can apply Gibbs and Metropolis-Hastings algorithms jointly to sample regression parameters from posterior distributions:

1. Conditional on c^{j-1}, construct S_t using equations (9) for TAR model.
2. Given from the previous step, derive OLS version of parameter vector $\hat{\theta}_{OLS}$
3. Conditional on c, σ_0 and σ_1, draw θ from normal distribution.
4. Conditional on θ and c, draw σ_0 and σ_1 from inverted Gamma distribution.
5. Given θ^j, σ_0^j and σ_1^j, c^j is generated by
$$c^j = c^{j-1} + \sigma_c u_c$$
6. If the ratio $\frac{L(y_t; \theta^j, c^j, \sigma_0^j, \sigma_1^j) p(\theta^j) p(c^j) p(\sigma_0^j) p(\sigma_1^j)}{L(y_t; \theta^{j-1}, c^{j-1}, \sigma_0^{j-1}, \sigma_1^{j-1}) p(\theta^{j-1}) p(c^{j-1}) p(\sigma_0^{j-1}) p(\sigma_1^{j-1})}$ is greater than a random variable generated from the uniform distribution over the unit interval then the draw is accepted – set $c^{j-1} = c^j$ and proceed to step 1. Otherwise, the draw is discarded – set $c^{j-1} = c^{j-1}$ and proceed to step 1.