

Combining Forecasts from Linear and Nonlinear Models Using Sophisticated Approaches

Mamdouh A. M. Abdelsalam^{1,2} & Doaa Akl Ahmed³

¹ Department of Economics, University of Leicester, UK

² Department of Economics, University of Mnofia, Egypt

³ Department of Economics, University of Benha, Egypt

Correspondence: Astley Clarke Building, University of Leicester, Leicester, LE1 7RH, UK. Tel: 44-7474-5400-2015. E-mail: maa64@le.ac.uk

Received: August 21, 2015

Accepted: September 15, 2015

Online Published: October 25, 2015

doi:10.5539/ijef.v7n11p190

URL: <http://dx.doi.org/10.5539/ijef.v7n11p190>

Abstract

This paper aims at improving the prediction accuracy through using combining forecasts approaches. In forecast combination, the crucial issue is the selection of the weights to be assigned to each model. In addition to traditional methods, we propose, also, two sophisticated approaches. These suggested methods are modified Bayesian Moving Average (BMA) and Extended Time-varying coefficient (ETVC). The first technique is based on merging the traditional BMA with other frequentist combination schemes to avoid the subjective prior inside the traditional Bayesian technique. The suggested ETVC approach provides consistent time-varying parameters even if there are some measurement errors, omitted variables bias and if the true functional form is unknown. Concerning the included models, we consider both linear and nonlinear models in order to calculate the forecasts of quarterly Egyptian CPI inflation. We find that our proposed scheme ETVC is superior to the best model and all other static combination schemes including the time-varying scheme based on the random walk coefficients updated (TVR) approach. Additionally, the suggested modified Bayesian approach improves the traditional BMA and overcomes the problem of depending on the arbitrary choice for the initial priors.

Keywords: forecast combination, linear models, nonlinear models, forecasting, time-varying coefficients, bayesian forecast combination

1. Introduction

Economic forecasting is an essential tool for economic policy-making. It is believed that the best forecast can be obtained by estimating a parametric model based on a particular dataset and, then, generating predictions from the fitted model. Since different model specifications have the heterogeneous information, they yield different forecasts. Usually, we use some criteria to select the best model and to eliminate the other ones. However, these rejected projections may have some marginal information which is not contained in the best predictor. Therefore, as confirmed empirically, the inclusion of these predictions to form one combined forecast can improve the accuracy of predictions (Clemen, 1989; Armstrong, 1989). Furthermore, if there is structural instability in the data, it is recommended to obtain the average forecast of different models to deal with this variability (Ravazzolo, Van Dijk, & Verbeek, 2007).

In combining forecasts, the primary concern is to calculate the optimal weights that correspond to each model in order to minimize a particular loss function. Methods of obtaining the optimal weights are divided into two groups: namely, Bayesian Model Averaging (BMA); and Frequentist Model Averaging (FMA). Granger and Ramanathan (1984) introduced the FMA methodology by developing Bates and Granger's (1969) method. This technique is based on averaging different predictions in order to minimise a defined loss function. They proposed employing the coefficients of restricted Ordinary Least Square (OLS) as weights for the competing models. Terui and Van Dijk (2002) extended the OLS combination approach by representing the dynamic forecast combination where the weights were estimated as a random walk process. Additionally, Hansen (2007) suggested using the Mallows Model Averaging (MMA) technique based on Mallows' (1973) criterion. Furthermore, Hansen (2008) applied MMA to estimate the weights associated with different forecasts. However, there is widespread criticism of the latter approach because it assumes that all considered models are nested so that the results are sensitive to

changing the order of regressors inside the model.

On the other hand, the BMA combines different forecasts by computing the conditional posterior probability for each model and, then, generating weights by calculating the mean of these probabilities. Initially, the BMA uses the Bayesian Information Criterion (BIC), developed by Schwarz (1978), for model selection where the former could be considered to be a simple form of the BMA approach. There is very extensive literature on BMA where the core of this approach is to account for the future uncertainty in terms of probabilities (Lahiri & Martin, 2010) (Note 1). Recent studies of forecast combination in economics showed significant interest in the application of the BMA method. However, there is extensive criticism of the BMA because its results are dependent on using some subjective priors. Those priors should be predetermined for each model and parameters. In other words, BMA priors are based on the researchers' arbitrary choices.

This paper employs different techniques to choose the optimum weights. Therefore, it utilises not only the aforementioned traditional approaches but proposes, also, two more methodologies: namely, modified BMA and Extended Time-varying coefficient (ETVC). Our aim is to demonstrate that, in comparison with traditional methods, these sophisticated techniques improve the accuracy of combining forecasts. The suggested modified BMA avoids the subjective priors inside the traditional Bayesian approach by incorporating traditional BMA with others frequentist combinations schemes. Precisely, we propose the use of the weights of both Granger and Ramanathan (1984) and Inverse Mean Square Forecasting Errors (IMSFE) as priors for the modified BMA.

On the other hand, the proposed ETVC technique is based on Swamy, Tavlas, Hall, and Hondroyannis (2010) and Hall, Swamy, and George (2014). The first authors developed the Time-varying coefficient (TVC) as a method for estimating consistent parameters even though there is uncertainty about the exact functional form. Additionally, the estimated coefficients are consistent either in cases of omitting some relevant variables, or if there are measurement errors in the included variables. Therefore, due to its above mentioned advantages, we believe that the ETVC method is quite important in generating the optimum weights. Furthermore, it can improve the combination of forecasts since it allows us to imitate the unknown functional form for the included variable. Additionally, it can add more information inside the combination scheme since it is based on some transformations of the predictors in the state variables and not only in the linear form. Finally, ETVC provides more flexibility in the prediction process; this implies that many different types of available forecasts can be considered in order to obtain the best identification.

Concerning the employed models, which we use to calculate the individual predictions, we apply both linear and nonlinear specifications. That is because, sometimes, it is difficult to argue whether the underlying time series is linear or nonlinear. Also, nonlinear models are very useful when the relationships are subject to regime changes. Furthermore, the specific data generating process may change its features from linear to nonlinear (Terui & Kariya, 1997). Therefore, by incorporating those different models, we can compare the forecasting accuracy of each type. The included nonlinear models are Generalized Autoregressive Conditional Heteroscedasticity (GARCH) and its threshold extension (TARCH). Also, we consider the Autoregressive conditional variance, skewness and kurtosis (GARCHSK-M), the Neural Network (NN), and Markov-switching (MS) models. The latter model is a piece-wise linear model since the data generating process is linear within each regime. Then, we utilise two structural linear models, namely, Bayesian Vector Autoregressive (BVAR) and the modern semi-structural "DSGE-VAR" models. Finally, we estimate Time-Varying Coefficients Autoregressive model. We apply those models to predict Egypt's quarterly CPI inflation. Then, by using both traditional and developed techniques proposed in this paper, we use the predictions, resulting from these different specifications, to compute the combined forecast.

Consequently, this research contributes to the literature by suggesting two more sophisticated combination schemes and compares them with the traditional methods. Additionally, we employ a range of both linear and nonlinear models to forecast future inflation which we use as inputs in our combination methods. Finally, we apply our methodology to quarterly Egyptian CPI inflation; this was not investigated in previous studies. Our motivation for choosing Egypt as a case study is the Central Bank of Egypt's announcement of its plan to move to a full-fledged inflation targeting framework when its prerequisites are satisfied. Consequently, it must have accurate models to predict future inflation since it is the core of this regime.

The results indicate that the Semi-Structural model is the best model according to all employed prediction criteria. Furthermore, both the time-varying Autoregressive and TARCH models provide good forecasts whereas the linear BVAR model has the lowest prediction accuracy. Regarding the combination techniques, our proposed ETVC technique dominates the best model and all other static combination approaches. Also, we compared the ETVC's forecasting ability with the time-varying scheme with random walk coefficients updated (TVR) and

found that it was inferior to the ETCV approach. Additionally, the suggested modified Bayesian methodology improves the traditional BMA and overcomes the problem of subjective choice of initial priors.

The paper is arranged as follows. Section 2 discusses the employed combination methods. Section 3 presents the applied models. Section 4 displays the empirical results of the eight different models. Section 5 is assigned to assess the performance of the alternative forecast combination approaches. Finally, Section 6 concludes and suggests some policy recommendations.

2. Combination Methodology

This paper applies different procedures for choosing the optimal weights in combining forecasts. In addition to the two proposed methods, namely modified BMA and ETVC averaging with time-varying weights, these measures include simple schemes, Frequentist Combination schemes, and Bayesian Combination approaches.

Suppose that we have k available projections $\hat{y}_{T+h,1}, \hat{y}_{T+h,2}, \dots, \hat{y}_{T+h,k}$ resulting from k different models and they are used to compute a forecast of y_{T+h} . Assume, also, that the combined point forecast is a function of those individual k predictions plus the vector of weights associated with each of those models, $w_{i,T+h}$. Thus, this combined forecast could be written as: $\hat{y}_{T+h} = g(\hat{y}_{T+h,1}, \hat{y}_{T+h,2}, \dots, \hat{y}_{T+h,k}, w_{i,T+h})$. The forecast error is given by $e_{T+h} = y_{T+h} - g(\hat{y}_{T+h,1}, \dots, \hat{y}_{T+h,k}, \hat{w}_{T+h})$. We can calculate the optimum values of the weights by minimising the loss function (L) as follows:

$$\min_{w_{T+h}} E[L(e_{T+h}(w_{T+h}) | \hat{y}_{T+h,1}, \dots, \hat{y}_{T+h,k})] \quad (1)$$

This loss function is supposed to be a function only of the forecast error. Additionally, the vector \hat{w}_{T+h} contains the optimal weights which satisfy equation (1). According to Ravazzolo et al. (2007), equation (1) allows for the inclusion of both nonlinear and time-varying approaches of combinations. In order to achieve a closed form solution of equation (1), the loss function is assumed to be the Mean Squared Forecast Error (MSFE) as expressed in equation (2).

$$L(e_{T+1}(w_{T+1})) = \theta(\hat{y}_{T+1} - y_{T+1})^2 \quad \theta > 0 \quad (2)$$

Thus, in this study, we assume that the loss function is the MSE which is given a fixed value of $\theta = 1$. Now, our discussion focuses on the different approaches applied to compute the weights assigned to each model. These approaches are divided into four categories: Simple; frequentist; Bayesian; and the proposed approaches.

2.1 Simple Forecasts Combination Schemes

As indicated by Timmermann (2006), there is no need to estimate any parameters or to compute the variance-covariance matrix in simple forecast combination schemes. In this research, we use two simple approaches which are equal weights (EQ) and Inverse Mean Square Forecast error (IMSFE); these are explained briefly below:

2.1.1 Equal Weights (EQ)

This approach is considered to be the simplest method for calculating the combination of weights. Despite its simplicity, many studies found that it worked better than many of the complicated techniques. It computes those weights as the mathematical average of all available individual forecasts:

$$w_i = \frac{1}{k} \quad (3)$$

2.1.2 Inverse Mean Square Forecast Error (IMSFE)

This combination scheme assumes that the models, which have less forecasting error, should be associated with higher weights. Hence, calculating the combination weights depends on the inverse forecasting error for the available forecasting models; and it can be obtained as follows:

$$w_i = \frac{\left(\frac{1}{MSFE_i}\right)}{\sum_{i=1}^k \left(\frac{1}{MSFE_i}\right)} \quad (4)$$

2.2 Frequentist Combination Schemes (FR)

This research applies two frequentists combination approaches, namely, Granger-Ramanathan (GR) and MMA.

These are covered below:

2.2.1 Granger-Ramanathan Combination (GR)

This technique depends on estimating OLS regression to compute the weights of each single model by imposing some constraints. These restrictions include no intercept, nonnegative coefficients and, finally, the summation of these estimates should equal unity. Therefore, the constrained regression model can be represented by the following:

$$y_t = \beta^1 y_t^1 + \beta^2 y_t^2 + \dots + \beta^k y_t^k + u_t \quad (5)$$

given that $\sum_{i=1}^k w_i = 1, w_i \geq 0$

The y_t^1 , y_t^2 and y_t^k are the forecasts resulting from the first, second and k^{th} model. (Note 2).

2.2.2 Mallows Model Averaging (MMA)

Hansen (2007; 2008) introduced (MMA) as a method for models combination; this could be expressed as:

$$MMA_T(w) = (y - \mu(w))^T (y - \mu(w)) + 2\sigma^2 m(w) \quad (6)$$

Where:

$$\mu(w) = \sum_{i=1}^K w^i \mu^i = \sum_{i=1}^K w^i y^i \hat{c}^i = P(w)y \quad (7)$$

$$P(w) = \sum_{i=1}^K w^i y^i (y^{iT} y^i)^{-1} y^{iT}$$

Where $P(w)$ is the averaged projection matrix and, given that the sample variance is an alternative of the unknown variance of population $m(w)$ refers to the number of effective coefficients as follows:

$$m(w) = \sum_{i=1}^K w^i m^i \quad (8)$$

For some fixed integer N , the weights are derived by minimizing equation (6) with constraints $w^i \in H_k^*(N)$. Hansen (2008) adopted the following discrete set H_k^* for weights:

$$H_k^*(N) \equiv \{w^m \in [0, \frac{1}{N}, \frac{2}{N}, \dots, 1] : \sum_{i=1}^K w^i m^i\} \quad (9)$$

2.3 Bayesian Combination

The logic behind the BMA is that, if there are k potential models, only one of them is the true one. Then, we estimate the posterior distribution as the weighted average of the conditional predictive densities for the included models. The predictive density of y_{t+h} providing the available observed data till the time t , Q_T , is estimated using the posterior probabilities as follows: (Note 3).

$$p(y_{T+h}|Q_T) = \sum_{i=1}^K p(M_i|Q_T) p(y_{T+h}|Q_T, M_i) \quad (10)$$

Where K is the number of considered models, $p(M_i|Q_T)$ is the posterior probability of model M_i . Additionally, $p(y_{T+h}|Q_T, M_i)$ is the conditional predictive density conditional on the M_i and the function Q_T . Based on Q_T and other information set in model M_i , we can calculate the conditional predictive density as:

$$p(y_{T+h}|F_T, M_i) = \int p(y_{T+h}|\theta_i, F_T, M_i) p(\theta_i|F_i, M_i) d\theta_i \quad (11)$$

Where $\int p(y_{T+h}|\theta_i, Q_T, M_i)$ can be defined as the conditional predictive density of y_{T+h} given θ_i , Q_T and M_i . Then, model M ' posterior probability can be estimated by: (Note 4).

$$p(M_i|D) = w_i = \frac{P(D|M_i) P(M_i)}{\sum_{i=1}^K P(D|M_i) P(M_i)} \quad (12)$$

Where D is specified dataset and $P(M_i)$ is the prior probability associated for the model M_i .

2.4 The Proposed Combination Approaches

This subsection presents the two proposed combination schemes: namely, the modified BMA; and ETVC averaging methods.

2.4.1 The modified Bayesian Approach

We propose a combination of both frequentist and Bayesian schemes by using the results of the former combination methods as priors inside the latter technique. This suggestion is advantageous since it should improve forecasting accuracy and, also, it overcomes the subjective choice in choosing priors inside the

Bayesian approach. Therefore, using the weights w_{if} resulting from previous methods and substituting the results of both OLS and IMSFE as priors for the same Bayesian identification as in the previous procedure, gives the following as follows:

$$P(M_i) = f(w_{if}) \quad (13)$$

2.4.2 ETVC Averaging Approach

As indicated in the introduction, Swamy et al. (2010) used the basis of the theorem of Swamy and Mehta (1975) to derive the TVC technique as a method for estimating consistent parameters. According to this theory, any nonlinear functional form can be specified precisely as a linear model; however, it allows the parameters to vary over time. This indicates that, by using the TVC model, we can always estimate any relationship even though we do not have enough knowledge about the specification of the true function. Therefore, we can obtain consistent coefficients taking into consideration any nonlinearity that may exist in the actual dataset or that some relevant variables may be missing from the estimated models.

Forecast combination can face the misspecification problem exactly like individual models. Specifically, combined predictions can have the same omission bias like other least squares regression models. Therefore, the estimation of optimum weights may be exposed to the omission problem since it is not expected to include all possible specifications. Additionally, the relationship between the forecasted variable and the regressors (i.e. alternative predictors) inside the combination scheme may be nonlinear (Landram, Shah, & Landram, 2011; Dong, 2002). Consequently, traditional estimation methods, such as OLS are invalid while ETVC is appropriate under these circumstances.

Based on the TVC approach, we can express any nonlinear relations to average different forecasts as follows:

$$y_t = \beta_{1t}y_t^1 + \dots + \beta_{kt}y_t^k \quad (t = 1, \dots, T) \quad (14)$$

Thus, y_t^k represents the predictions resulting from the different models and β_{it} are the time-varying coefficients (or weights) for each model. As indicated in the introduction, the TVC approach depends on the selection of some relevant variables to feed the coefficients drivers. Initially, the selection of those drivers depends on a bench of "arbitrary" assumptions which, usually, are challenging in conducting this approach. Hall et al. (2014) classified the complete set of biased TVCs into two subgroups: namely, biased, and unbiased components. The biased part contains the coefficients associated with the misspecification of the model. Therefore, this component should be eliminated in order to obtain the consistent estimates of the correct functional form. We follow Hall et al.'s (2014) technique to obtain consistent estimates of the weights assigned to individual models in our forecast combination exercise. The following are the assumptions that should be imposed to implement operationally the above mentioned technique (Note 5).

Assumption 1: each coefficient can be expressed as a linear function of a group of variables (coefficient drivers) plus a random error. On the assumption that we have π_j fixed parameters, and z_{dt} coefficient drivers, the time-varying coefficients is written as:

$$\beta_{it} = \pi_{j0} + \sum_{d=0}^{p-1} \pi_{jd} z_{dt} + \epsilon_{jt} \quad (j = 0, 1, \dots, k-1) \quad (15)$$

Assumption 2: we can categorise the group of coefficient drivers and the constant in equation (15) into three different subgroups, namely, A_{1j}, A_{2j}, A_{3j} . Therefore, A_{1j} is related to the variability in the true parameter because of the nonlinearity of the relationship. Also, both A_{2j}, A_{3j} are associated with the omitted variable bias and the measurement error bias respectively. Moreover, the selected driver set should satisfy some prerequisites. As a rule, the total drivers should produce a well-specified relationship and it should have high predictive power.

In this paper, we base our estimate of the ETVC on the full coefficients method in TVC approach with some driver sets inside each time-varying coefficient. In addition to some nonlinear transformations from each alternative forecast, we include some polynomial components to be able to account for any bias and, also, for any possible nonlinearity inside the actual data of the considered forecasted variable (Note 6).

2.4.3 Time-Varying with Random Walk Updating Coefficients (TVR)

Finally, it is useful to compare our proposed ETVC combination scheme with other time-varying coefficients methods. One of the most common alternatives is the Time Varying with Random Walk Coefficients (TVR) method. According to this methodology, the combination of different point forecasts can be written in the form: (Note 7).

$$y_t = \beta_{1t}y_t^1 + \dots + \beta_{kt}y_t^k + u_t \quad (16)$$

$$\beta_t = \beta_{t-1} + v_t \quad (17)$$

Where, y_t and y_t^k are the actual series and the predictions of the various models. Additionally, the time-varying coefficients β_{it} are modelled as random walk processes and error terms; u_t and v_t are independent and distributed normally with zero means and constant variances.

Both non-constant combinations schemes can be identified in state space form in which the first equation is assigned to the variable of interest with other state equations for each time-varying coefficient. The state space can be estimated by utilizing the predictive Kalman Filter algorithm as recursive relationships with respect to the predictions (Harvey, 1989). However, compared to TVR, the ETVC is a more sophisticated approach for fitting the non-constant coefficients since the former incorporates extensive information to overcome many misspecifications in the underlying relationship as mentioned earlier. This means that ETVC is expected to give better imitation for the real data generation process; this is the ultimate aim of all econometricians.

3. Models

This section presents the employed linear and nonlinear specifications. The nonlinear models include time-varying conditional volatility models, GARCH-M and TARCH-M models. Then, we present the GARCHSK-M model which assumes that the conditional distribution is time-varying in the first four moments. Additionally, we employ the Neural Network (NN) and Markov Switching (MS) models. This is followed by introducing two structural linear models, namely, BVAR and the modern Dynamic Structure General Equilibrium (DSGE) augmented with the Vector Autoregressive model which yields the (DSGE-VAR) model. Finally, we present briefly the Time-Varying Coefficients Autoregressive model.

3.1 GARCH Model with Generalised Error Distribution

We employ the GARCH model, developed by Bollerslev (1986), as an extension of the ARCH model introduced by Engle in 1982. The basic GARCH model estimates the conditional volatility as a function of its past lags as well as lagged squared errors. This model is estimated using the Maximum Likelihood (ML) approach. Therefore, in the GARCH(1,1) model, the conditional variance equation can be written as:

$$h_t = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 h_{t-1} \quad (19)$$

Where ε_{t-1}^2 are the past squared errors. Given that the variance must be strictly positive, the parameters of equation (19) must be greater than zero (i.e., $\beta_0 \geq 0$, $\beta_1 \geq 0$, $\beta_2 \geq 0$). Furthermore, to guarantee that h_t is stationary, the sum of ARCH and GARCH parameters must be less than unity (i.e., $\beta_1 + \beta_2 < 1$).

3.2 TARCH-M Model with T-Distribution

There is criticism of the GARCH (p,q) models because they assume that the variance responds similarly to positive and negative shocks. Glosten, Jagannathan, and Runkle (1993) developed the TARCH specification to capture the asymmetric response of financial time series to different signs of the shock. According to this specification, the conditional volatility can be written as:

$$h_t = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 h_{t-1} + \beta_3 \varepsilon_{t-1}^2 (\varepsilon_{t-1} < 0) \quad (20)$$

The following conditions are required to ensure that conditional volatility is strictly positive; $\beta_0 > 0$, $\beta_1 > 0$, $\beta_1 + \beta_3 > 0$, $\beta_2 \geq 0$. The asymmetry parameter β_3 can be positive or negative depending on the shock. Also, this coefficient is used to measure the contributions of shocks to persistence over both the short-run and the long-run; $(\beta_1 + \beta_3/2)$ and $(\beta_1 + \beta_2 + \beta_3/2)$ respectively.

3.3 Modelling Conditional Variance, Skewness and Kurtosis

Based on the Gram-Charlier series expansion of the normal density, Leon, Rubio, and Serna (2005) proposed a new methodology to estimate jointly time-varying conditional second, third and fourth moments. Specifically, they assumed that each of these moments is generated by a GARCH-type process. Let GARCHSK-M indicates the time-varying higher order moments model when the conditional variance, skewness and kurtosis follow a GARCH (1,1) specification. Therefore, the GARCHSK-M model is estimated in steps starting with the GARCH (1,1) model of inflation. Then, the estimated parameters are used as starting values for the equations of mean and variance in the GARCHSK-M model. Thus, the variance equation takes the form described in equation (19) while mean, skewness and kurtosis equations are represented by the following set of equations:

$$\text{Mean equation: } \pi_t = \sum_{i=1}^n \alpha_i \pi_{t-i} + \varepsilon_t \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2) \quad (21)$$

$$\varepsilon_t = \eta_t \sqrt{h_t} \quad ; \quad \eta_t \sim (0,1) \quad E(\varepsilon_t | I_{t-1}) \approx (0, h_t)$$

$$\text{Skewness equation: } s_t = \gamma_0 + \gamma_1 \eta_{t-1}^3 + \gamma_2 s_{t-1} \quad (22)$$

$$\text{Kurtosis equation: } k_t = \delta_0 + \delta_1 \eta_{t-1}^4 + \delta_2 k_{t-1} \quad (23)$$

Where ε_t is the error term, η_t is the standardized residuals, h_t, s_t , and k_t are conditional volatility, skewness and kurtosis corresponding to η_t respectively. They determined that $E_{t-1}(\eta_t) = 0, E_{t-1}(\eta_t^2) = 1, E_{t-1}(\eta_t^3) = s_t$ and $(\eta_t^4) = k_t$.

3.4 Neural Network (NN) (Note 8)

Artificial NN models comprise complex nonlinear relationships to generate forecasts based on the brain's simple mathematical approaches. They can be seen as a network of "neurons" categorized with particular layers. As inputs come in the bottom layers, outputs or forecasts come in the top layers and "hidden neurons" come as intermediate layers. Thus, by using lagged values as inputs inside the neural network system, the NN model can be utilized to estimate nonlinear autoregressive models for a particular variable. In this case, the relationship can be represented in the form $NNT(p, k)$ where p is the number of lags, and the k is the number of the hidden nodes inside the layers. We employ the Feed-Forward networks with one hidden layer and the three nodes approach, which depends on the training sample to fit the data. Then, we obtain the out-of-sample forecasts depending on a learning algorithm to minimize a particular loss function (Hyndman & Athanasopoulos, 2013).

3.5 Markov Switching Regression

This model is based on decomposing a series in a finite sequence of distinct stochastic processes or regimes. Therefore, the current process in each regime is linear but the combination of the processes generates a nonlinear regime. The autoregressive model, which is subject to changes in the autoregressive parameter, can be expressed as in the following system. Equations (24) and (25) assume that we have two regimes, (Brooks, 2002; Laurini & Portugal, 2002):

$$\pi_t = \alpha_1 + \beta_{i1} \sum_{i=1}^p \pi_{t-i} + \varepsilon_{1t} \quad m = 1 \quad (24)$$

$$\pi_t = \alpha_2 + \beta_{i2} \sum_{i=1}^p \pi_{t-i} + \varepsilon_{1t} \quad m = 2 \quad (25)$$

The parameters α_1 and β_{i1} capture the behaviour of the series when the current system is one while α_2 and β_{i2} describe the behaviour of the series in the second regime. In this paper, we employ the Markov Chain method and assume that the probability of a variable s_t conditional on some particular j value depends only on its previous value s_{t-1} . This is represented by the following equation:

$$P\{s_t = j | s_{t-1} = k, \dots\} = P\{s_t = j | s_{t-1} = i, \dots\} = P_{ij} \quad (26)$$

Where i, j , give the probability that state j follows the state i . The key feature of this Markov transition matrix of the first order is that the probability of transition to the next regime depends only on the current state (Laurini & Portugal, 2002).

3.6 Bayesian Vector Autoregressive (BVAR)

Vector Autoregressive (VAR) is a common workhorse in forecasting purposes. In this study, our VAR model includes four variables which are incorporated usually in the case of fitting inflation inside the small open economy. Moreover, we employ the Bayesian approach that combines between the initialised priors and data fit summed up in the final posteriors. This type of model is advantageous in comparison with the classical VAR, especially in case of small samples, since it allows for more degrees of freedom by including priors in the initial fitting values. The included variables are CPI inflation, GDP growth rate, changes in real exchange rate and oil prices.

3.7 Semi Structure Model

Dynamic Stochastic General Equilibrium (DSGE) models are employed regularly as they provide analytical tools to understand better the equilibrium relationships inside the economy. However, DSGE models have received many criticisms in terms of forecasting accuracy. Therefore, many efforts have been made to improve their predictions (Ingram & Whiteman, 1994; Schorfheide, 2000; Del Negro & Schorfheide, 2004; and Del Negro, Schorfheide, Smets, & Wouters, 2007; Gupta & Steinbach, 2013). These studies recommended that both the structural DSGE and the VAR models be merged. According to Del Negro and Schorfheide (2004), the DSGE estimated parameters might provide useful information for VAR parameters. Therefore, subject to the relative weights (λ) assigned to each type of data, the VAR model should be estimated based on both the actual and the DSGE priors. This study uses that methodology to compute the forecasts from the DSGE-VAR model based on the optimization assumptions for micro-agents (Woodford, 2003). The corresponding model encompasses a representative householder, a sequence of monopolistic competitive firms and the central bank (For more details, see Appendix A) BVAR posterior estimate is conditional on the value of the relative weights

λ associated with both models DSGE and the unrestricted VAR. Where the optimal value of λ can be estimated such that it maximizes $P(Y|\lambda)$ as follows:

$$\hat{\lambda} = \arg \max_{\lambda \in \Lambda} P(Y|\lambda) \quad (29)$$

3.8 Time-Varying Coefficients Autoregressive

We analyse a linear autoregressive model with time-varying coefficients that can be presented in the following form:

$$Y_t = \delta_{0t} + \delta_{1t}Y_{t-1} + \dots + \delta_{pt}Y_{t-p} + \varepsilon_t \quad (30)$$

The underlying estimates follow a random walk process. The model can be characterised in a state space form which can be solved by a predictive Kalman Filter algorithm (Note 9).

$$\hat{\beta}_{t|t-1} = E(\beta_t | Y_1, \dots, Y_{t-1}) \quad (31)$$

$$\hat{\Sigma}_{t|t-1} = E[(\beta_t - \hat{\beta}_{t|t-1})(\beta_t - \hat{\beta}_{t|t-1})'] \quad (32)$$

Consequently, Y_t is CPI inflation and $\hat{\Sigma}_t$ is the estimated covariance matrix and the optimal lag length is determined based on AIC criteria.

4. Empirical Results

4.1 Data and Preliminary Check

We employ quarterly data sourced from International Financial Statistics (IFS) for the period 1957:1 to 2015:1. We include variables of CPI, nominal exchange rate, Gross domestic product (GDP), nominal interest rate (r) for Egypt as well as world oil prices (oil) and CPI of the USA. Inflation data is computed as quarterly changes in the logarithm of the CPI. We chose the sample to include the largest number of available observations in order to provide more accurate results. Table 1 displays the basic descriptive statistics for the data which, according to Jarque-Bera (JB) test statistic, is unlikely to be drawn from normal distribution

Table 1. Descriptive statistics of the CPI inflation (1957:1 to 2015:1)

Statistic	value
Mean	0.020815
Median	0.013453
Maximum	0.115602
Minimum	-0.032790
Std. Dev.	0.022941
Skewness	1.097721
Kurtosis	4.351560
JB	64.25123
JB- p-value	0.000000

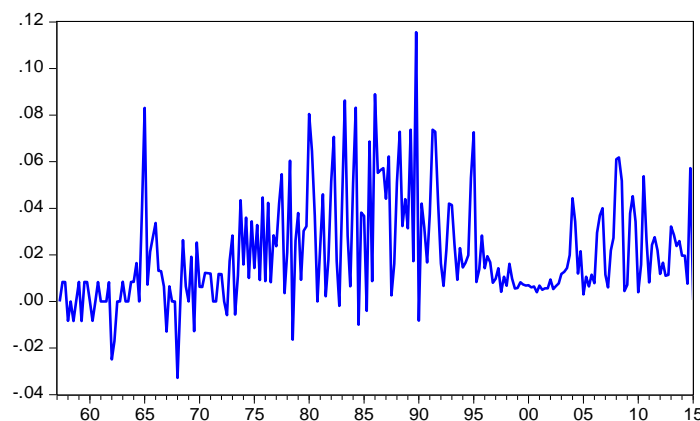


Figure 1. Quarterly inflating rate for the period (1957:1 to 2015:1)

We follow Box-Jenkins approach in selecting the best specification of the mean equation for the models which allows for volatility modelling. For both the GARCH and GARCHSK-M model, the selected specification

includes first and fourth lags of inflation while the TARCH specification is represented as an ARMA(1,2) process. These specifications are selected according to both AIC and SIC criteria and they are free of the serial correlation between the errors. However, the serial autocorrelation amongst the residuals of these different models exists in the sequences of ε_t^2 , ε_t^3 and ε_t^4 . Furthermore, from the ARCH LM test, there are indications of the ARCH effects in the residuals. Therefore, the models, which assume time-varying conditional variance and higher order moments are more suitable in modelling inflation. Also, we add two dummies in the volatility equations to account for the shift to the open door policy in 1974 and the start of Economics Reform and Structural Adjustment Programme (ERSAP) in May 1991.

4.2 Results

We estimated the models over the period (1957:1-2000:2). Therefore, we used the rest of the observations to predict out-of-sample inflation in order to allow a suitable number of observations to be employed in the combination of forecasts. Table 2 presents the results of the first three models, GARCH-M, TARCH-M, and the GARCHSK-M. As shown in the Table, the volatility persistence parameter β_2 is positive and significant in all models with the lowest magnitude being in the GARCHSK-M model. Concerning the volatility effect in the mean equation for the GARCHSK-M model, the estimated parameter is both positive and significant. In addition, the effects of shocks to variance are significant in all models with the lowest magnitude being in the GARCHSK-M model. With reference to the conditional skewness, the shocks parameter is significant with a negative sign while the persistence parameter is both positive and significant. In the same way, the shocks to conditional kurtosis and the persistence parameters are both positive and significant. Furthermore, the lagged kurtosis coefficient is greater than that of the lagged volatility whereas the shock effects on kurtosis are the smallest when compared to those effects of shocks to volatility and skewness.

Table 2. GARCH-M (GED), TARCH-M (t-dist) and GARCHSK-M models

GARCH (GED): Mean equation: $\pi_t = \alpha_1 \pi_{t-1} + \alpha_2 \pi_{t-4} + \varepsilon_t$ Variance equation: $h_t = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 h_{t-1} + \kappa_1 d74 + \kappa_2 d91$							
TARCH (t-dist): Mean equation: $\pi_t = \alpha_1 AR(1) + \alpha_1 MA(1) + \alpha_1 MA(2) + \varepsilon_t$ Variance equation: $h_t = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 h_{t-1} + \beta_3 \varepsilon_{t-1}^2 (\varepsilon_{t-1} < 0) + \kappa_1 d74 + \kappa_2 d91$							
GARCHSK-M: Mean equation: $\pi_t = \mu h_t + \alpha_1 \pi_{t-1} + \alpha_2 \pi_{t-4} + \varepsilon_t$ Variance equation: $h_t = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 h_{t-1} + \kappa_1 d74 + \kappa_2 d91$ Skewness Equation: $s_t = \gamma_0 + \gamma_1 \eta_{t-1}^3 + \gamma_2 s_{t-1}$ Kurtosis Equation: $k_t = \delta_0 + \delta_1 \eta_{t-1}^4 + \delta_2 k_{t-1}$							
Model		GARCH-M (GED)		GARCHSK-M		TARCH-M (t-dist)	
		estimate	p-value	estimate	p-value	Estimate	p-value
Mean Equation	α_1	0.421581	0.0000	16.04779	0.0000	0.946401	0.0000
	α_2	0.359132	0.0000	0.396783	0.0000	-0.719429	0.0000
	α_3			0.230260	0.0000	-0.082415	0.3384
Variance equation	β_0	5.75×10^{-5}	0.0185	0.000115	0.0000	8.25×10^{-5}	0.0004
	β_1	0.564472	0.0004	0.331738	0.0000	0.408893	0.0195
	β_2	0.406056	0.0007	0.342286	0.0000	0.435064	0.0000
	β_3					-0.414592	0.0349
	k_1	0.000257	0.0000	0.000402	0.0000	0.000351	0.0000
	k_2	-0.000314	0.0000	0.000355	0.0000	-0.000424	0.0000
	GED	1.5					
Skewness Equation	γ_0			-0.148656	0.0000		
	γ_1			-0.028112	0.0000		
	γ_2			0.014555	0.0035		
Kurtosis Equation	δ_0			1.309328	0.0000		
	δ_1			0.003846	0.0000		
	δ_2			0.442644	0.0000		
Log-likelihood		420.4933		837.5127		426.0869	
L-Jung Q-Stat							
ε_t (lag 10)		13.046	0.221	10.286	0.416	10.510	0.161
ε_t^2 (lag 10)		3.4926	0.967	7.5362	0.674	2.5021	0.991

All models are estimated using ML estimation using Marquardt algorithm. Significant p-values are indicated by bold.

Regarding the Artificial feed-forward Network model, its fitment is based on a single hidden layer with three nodes and four lags of quarterly inflation. This allows us to obtain the filter in-sample forecasts as training period and, then, the out-of-sample predictions. Figure 2 shows the in-sample filter of the model while Figure 3 displays the out-of-sample forecasts where the red lines represent the estimated series and the blue lines represent the actual data.

In the MS model, the suitable number of regimes is selected according to the AIC. Figure 4 displays the results of the prediction, filtered, and smoothed probabilities of the model. The filtered probabilities are computed using the information up to period $t-1$ to infer the probabilities at moment t .

With respect to the estimation of the structural models BVAR and DSGE-VAR, they are estimated over the period 1982:1 2000:2 using the Bayesian approach. The BVAR model includes an exogenous dummy variable to account for the ERSAP structural break. The selected lag length is chosen according to three criteria which are the Likelihood ratio (LR), the Final prediction error (FBE), and the AIC. Diagnostic tests show that the model is well-specified.

On the other hand, the DSGE-VAR model consists of some linear equations that describe the endogenous variables dynamics which can be solved by using the standard Blanchard-Kahn condition and the Metropolis-Hastings algorithm to derive the posterior distribution. Table 3 presents the estimation results of the DSGE-VAR model; it shows that $\hat{\lambda}$ is 0.6749. (Note 10).

In order to estimate the Time-Varying Autoregressive model, we utilise the predictive Kalman Filter approach with random walk coefficients as explained previously. Also, AIC lag criterion indicates that the best lags are the first and the fourth including a constant term in each equation. Figure 6 shows the time-varying for the coefficients.

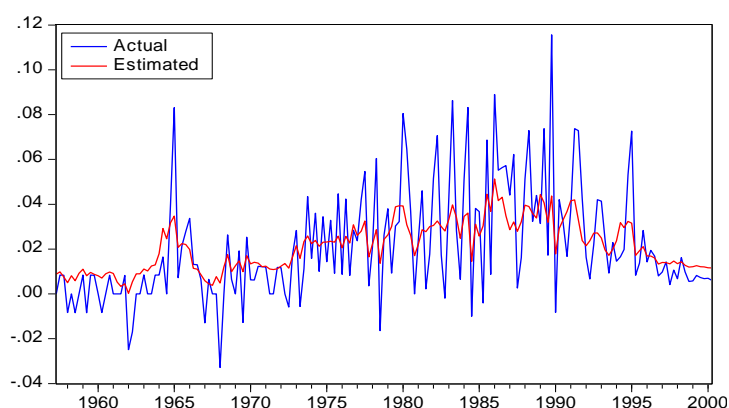


Figure 2. In-sample filter for one step ahead in neural network

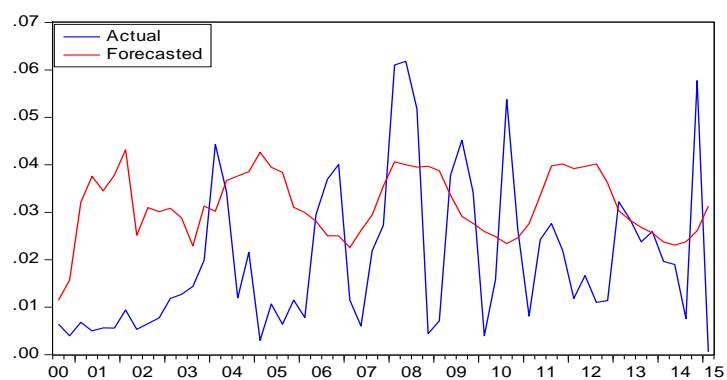


Figure 3. Out-sample forecasts for one step ahead in neural network

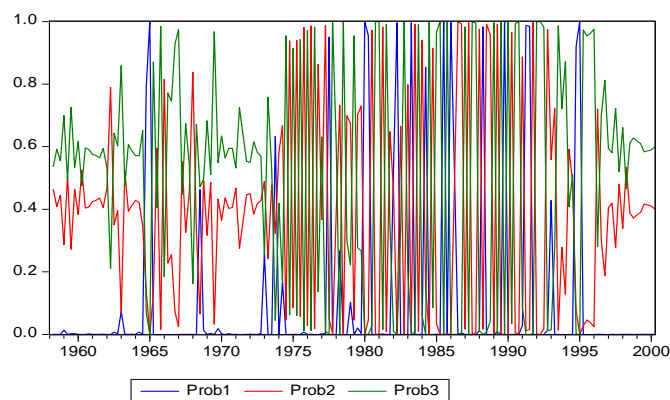


Figure 4. Different probabilities regimes

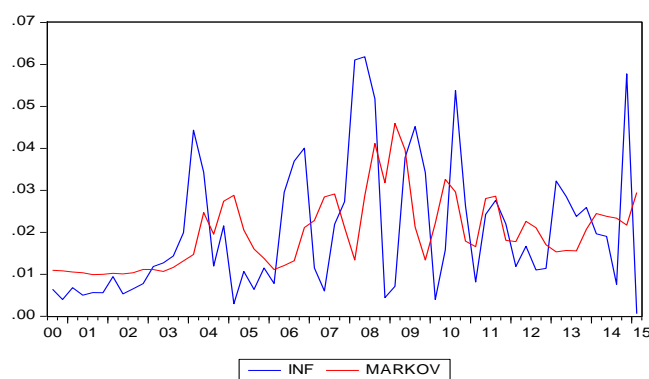


Figure 5. Out-sample forecasts for one-step-ahead in MS model

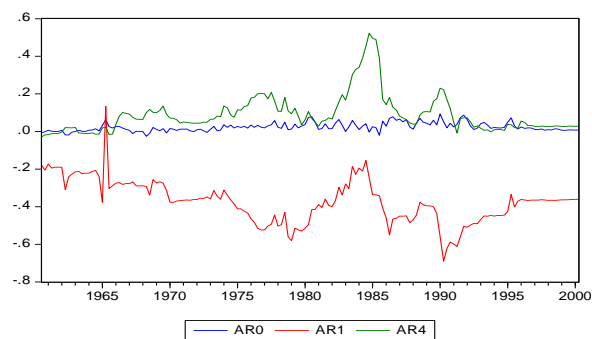


Figure 6. Time-varying autoregressive coefficients

Table 3. DSGE-VAR estimation

	Prior Mean	Pos. mean
κ	0.300	0.040
τ	2.000	3.4869
ψ_1	1.500	1.3204
ψ_2	0.125	1.5057
ρ_R	0.500	0.1073
ρ_g	0.800	0.9761
ρ_z	0.300	0.2064
λ	1.000	0.6749
$\epsilon_{g,t}$	0.875	0.3489
$\epsilon_{z,t}$	0.630	0.4281
$\epsilon_{R,t}$	0.251	0.1065
Calibrated Parameters		
r^*	0.5	
γ	0.5	

5. Forecasting Evaluation

5.1 Forecasting Performance of Estimated Models

Table 4 shows the different measures used to assess the accuracy of predictions. The first two criteria, namely, RMSE and MSE depend on the scale of the dependent variable which implies that they are relative measures in order to compare forecasts across different models. According to this criterion, the smaller the error, the better is the related model's forecasting ability. With respect to Theil inequality coefficient, it must lie between zero and one, where zero is a sign of perfect fit.

Table 5 presents the results of different measures for the one-step-ahead out-of-sample forecasts. According to the three employed criteria, the DSGE-VAR model is the best model whereas, the BVAR model is the worst one. Additionally, based on all the conducted measures, the TVAR model provides good forecasts. Also, although the TARCH model is good in terms of both RMSE and MSE measures, its performance is poor in terms of the TIC measure.

Table 4. Different criterions of predictive power

Criterion	Formula
1. Root Mean square error	$RMSE = \sqrt{\frac{1}{N} \sum_{t=T+1}^{T+N} (\hat{\pi}_t - \pi_t)^2}$
2. Mean square error	$MSE = \frac{1}{N} \sum_{t=T+1}^{T+N} (\hat{\pi}_t - \pi_t)^2$
3. Theil inequality coefficient	$TIC = \frac{\sqrt{\frac{1}{N} \sum_{t=T+1}^{T+N} (\hat{\pi}_t - \pi_t)^2}}{\sqrt{\frac{1}{N} \sum_{t=T+1}^{T+N} \hat{\pi}_t^2} + \sqrt{\frac{1}{N} \sum_{t=T+1}^{T+N} \pi_t^2}}$

Table 5. Out-of sample forecasts power of different models for one step ahead

	BVAR	DSGE-VAR	MR	TV	NN	GARCH (GED)	TARCH (t-dist)	GARCHSK-M
1. RMSE	0.0209 [8]	0.01448 [1]	0.0161 [5]	0.01505 [2]	0.02045 [7]	0.0159 [4]	0.0151 [3]	0.0199 [6]
2. MSE	0.00044 [8]	0.00021 [1]	0.00026 [5]	0.00022 [2]	0.000418 [7]	0.00025 [4]	0.00023 [3]	0.00039 [6]
3. TIC	0.4812 [8]	0.30119 [1]	0.3566 [4]	0.32960 [2]	0.367229 [5]	0.39265 [7]	0.368979 [6]	0.35154 [3]

The numbers in the square brackets indicate rankings of the models where [1] indicates the best models according to the corresponding measure.

5.2 Combination Results

The main aim of any forecasts combination procedure is to improve the prediction accuracy of the individual forecasts. Therefore, the good combination scheme should be superior to all individual forecasts and it should perform well compared to the other competing combination methods. In our analysis, we compare the forecasting performance of the different forecasting combination schemes and the best model in terms of MSE and RMSE. Table 6 reports the results of comparing the prediction power of all combination methods whereas Table (B) presents the weights associated with the individual models according to the different static combinations schemes. Additionally, the time-varying weights, based on both random walk updated coefficients and ETVC, are shown in graphs (B1) and (B2) respectively. In general, we can observe that the dynamic combination technique using ETVC dominates the best model and all other static combination schemes and, also, the traditional time-varying with random walk updated coefficients in our exercises.

Specifically, the dynamic combination scheme using ETVC is the most accurate combination procedure since it outperforms the best individual model and all other combination methods in terms of all forecasting measures. Additionally, as shown in Table 6, the best forecast is superior to both simple and frequentist schemes since the

GR, MMA, MFE and EQ combination methods have higher values of MSE, RMSE and TIC in comparison with the best individual model. Concerning the Bayesian combination schemes, the suggested improvements in substituting initial priors based on the other combination schemes proved to be very successful since the traditional Bayesian approach had values of 0.000185 and 0.0136 for MSE and RMSE respectively. Additionally, the Bayesian combination augmented with IRMSFE scheme achieved 0.000180 and 0.0134 respectively, and, for the above-mentioned criteria, the Bayesian combination augmented with OLS scheme achieved 0.00017964 and 0.0134 respectively. It is worth noting that we did not face the puzzle of forecasts combination literature in which the EQ approach outperforms sophisticated combinations methods. There are two reasons for us investigating this matter. Firstly, we have initial heterogeneous models which imply that each model has its specific information and some specific features. Secondly, we used some more sophisticated combinations methods such as ETVC and the modified Bayesian combination technique; these proved that it works very well with averaging different forecasts.

Table 6. Out-of-sample forecasts power of different combination methods for one step ahead

	RMSE	MSE	TIC
The best model	0.014485	0.0002098	0.3011
	[6]	[6]	[7]
EQ	0.014486	0.0002099	0.31129
	[7]	[7]	[8]
MFE	0.014489	0.0002099	0.31288
	[8]	[8]	[9]
MMA	0.015667	0.000239	0.35284
	[10]	[10]	[10]
GR	0.014523	0.000211	0.30
	[9]	[9]	[6]
TV-R	0.013851	0.00019185	0.291
	[5]	[5]	[4]
ETVC	0.011486	0.00013194	0.243
	[1]	[1]	[1]
BMA	0.01360	0.0001851	0.296
	[4]	[4]	[5]
BMA-OLS	0.01340	0.00017964	0.2885
	[2]	[2]	[2]
BMA-MFE	0.01342	0.000180	0.2895
	[3]	[3]	[3]

6. Conclusion and Policy Implications

By using Egyptian quarterly inflation data, the paper aimed at improving the inflation prediction using forecast combination of predictions of different linear and nonlinear models. In choosing the optimal weights associated with alternative models, we used not only the traditional approaches but, also, we proposed two advanced approaches: namely, modified BMA; and ETVC. In order to avoid the arbitrary choice, we based the proposed modified BMA on using the weights of some frequentist combination methods as priors inside the traditional Bayesian technique. The ETVC technique allowed us to compute consistent optimal weights even if there were measurement errors, misspecification or if the correct functional form was not identified. Consequently, it added more information inside the combination scheme since it was dependent on some transformations of the predictors in the state variables and not only in the linear form.

The results indicate that, according to MSE and RMSE criteria, the Semi-Structural model is the best model while the BVAR gives the worst forecasts. Concerning the combination techniques, the proposed ETVC approach dominates the best model, the time-varying scheme with random walk coefficients (TVR) and all other static combination schemes. Furthermore, the suggested modified Bayesian approach improves the traditional BMA and overcomes the problem of subjective choice for the initial priors.

Based upon the above conclusions, when generating the inflation predictions, the Central Bank of Egypt should consider a range of different models including linear and nonlinear models. Also, we recommend that the Central Bank of Egypt does not depend on a single model for prediction purposes. Instead, we recommend that the

sophisticated combinations schemes, namely ETVc and the modified BMA, should be used to improve the published forecasts for inflation.

This research could be extended in different ways. Firstly, we based our forecasts on some linear and nonlinear models and recommend that more models be incorporated such as the bilinear or the Dynamic Factor models. Secondly, we recommend that the ETVc combination scheme be extended by incorporating the Bayesian probability techniques since we consider that this would be a worthwhile starting point for future studies.

References

- Armstrong, J. (1989). Combining Forecasts: The End of the Beginning or the Beginning of the End. *International Journal of Forecasting*, 5. [http://dx.doi.org/10.1016/0169-2070\(89\)90013-7](http://dx.doi.org/10.1016/0169-2070(89)90013-7)
- Balke, N. S., & Wynne, M. A. (1996). *An Equilibrium Analysis of Relative Price Changes and Aggregate Inflation*. Federal Reserve Bank of Dallas, Working Papers 96-09. [http://dx.doi.org/10.1016/S0304-3932\(99\)00051-3](http://dx.doi.org/10.1016/S0304-3932(99)00051-3)
- Ball, L., & Mankiw, N. G. (1995). Relative-Price Changes as Aggregate Supply Shocks. *The Quarterly Journal of Economics*, 110(1). <http://dx.doi.org/10.3386/w4168>
- Bates, J. M., & Granger, C. W. J. (1969). The Combination of Forecasts. *Operational Research Quarterly*, 20. <http://dx.doi.org/10.2307/3008764>
- Blanchard, O. J., & Kahn, C. M. (1980). The Solution of Linear Difference Models under Rational Expectations. *Econometrica*, *Econometric Society*, 48(5).
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics*, 31. [http://dx.doi.org/10.1016/0304-4076\(86\)90063-1](http://dx.doi.org/10.1016/0304-4076(86)90063-1)
- Brooks, C. (2002). *Introductory Econometrics for Finance*. Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511841644>
- Christiano, L. J. (2007). Comment on Del Negro, Schorfheide, Smets and Wouters. *Journal of Business and Economic Statistics*, 25(2).
- Clemen, R. (1989). Combining Forecasts: A Review and Annotated Bibliography. *International Journal of Forecasting*, 5. [http://dx.doi.org/10.1016/0169-2070\(89\)90012-5](http://dx.doi.org/10.1016/0169-2070(89)90012-5)
- Del Negro, M., & Schorfheide, F. (2003). *Take Your Model Bowling: Forecasting with General Equilibrium Models*. Federal Reserve Bank of Atlanta, Economic Review.
- Del Negro, M., & Schorfheide, F. (2004). Priors from General Equilibrium Models for VARs. *International Economic Review*, 45(2). <http://dx.doi.org/10.1111/j.1468-2354.2004.00139.x>
- Del Negro, M., Schorfheide, F., Smets, F., & Wouters, R. (2007). On the Fit of the New Keynesian Models. *Journal of Business and Economic Statistics*, 25(2). <http://dx.doi.org/10.1198/073500107000000016>
- Dong, J. (2002). *A Nonlinear Combining Forecast Method Based On Fuzzy Neural Network*. First International Conference on Machine Learning and Cybernetics, China, 4-5 November. <http://dx.doi.org/10.1109/ICMLC.2002.1175421>
- Engle, R. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation. *Econometrica*, 52, 267-287.
- Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance*, 48(5). <http://dx.doi.org/10.1111/j.1540-6261.1993.tb05128.x>
- Granger, C. W. J., & Ramanathan, R. (1984). Improved Methods of Combining Forecasts. *Journal of Forecasting*, 3. <http://dx.doi.org/10.1002/for.3980030207>
- Gupta, R., & Steinbach, R. (2013). A DSGE-VAR Model for Forecasting Key South African Macroeconomic Variables. *Economic Modelling*, 33. <http://dx.doi.org/10.1016/j.econmod.2013.03.012>
- Hall, S. G., Swamy, P. A., & George, S. T. (2014). *Time Varying Coefficient Models: A Proposal for Selecting the Coefficient Driver Sets*. University of Leicester Working Paper No. 14/18.
- Hansen, B. E. (2007). Least Squares Model Averaging. *Econometrica*, 75(4). <http://dx.doi.org/10.1111/j.1468-0262.2007.00785.x>
- Hansen, B. E. (2008). Least-Squares Forecast Averaging. *Journal of Econometrics*, 146(2).

- <http://dx.doi.org/10.1016/j.jeconom.2008.08.022>
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781107049994>
- Hoeting, J. A., Raftery, A. E., & Madigan, D. (2002). Bayesian Variable and Transformation Selection in Linear Regression. *Journal of Computational and Graphical Statistics*, 11. <http://dx.doi.org/10.1198/106186002501>
- Hyndman, R., & Athanasopoulos, J. (2013). *Forecasting: principles and practice*. OTexts.
- Ingram, B., & Whiteman, C. (1994). Supplanting the Minnesota Priors Forecasting Macroeconomic Time Series Using Real Business Cycle Model Priors. *Journal of Monetary Economics*, 34(4).
- Lahiri, K., & Martin, G. (2010). Bayesian Forecasting in Economics. *International Journal of Forecasting*, 26(2). [http://dx.doi.org/10.1016/0304-3932\(94\)90030-2](http://dx.doi.org/10.1016/0304-3932(94)90030-2)
- Landram, F. G., Mendez, F., Shah, V., & Landram, S. V. (2011). A Fresh Approach to Combining Forecasts. *Southwestern Economic Review*, 36.
- Laurini, M. P., & Portugal, M. S. (2002). *Markov Switching Based Nonlinear Tests for Market Efficiency Using the R /US Exchange Rate*. Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Economia.
- Leon, A., Rubio, G., & Serna, G. (2005). Autoregressive Conditional Volatility, Skewness and Kurtosis. *The Quarterly Review of Economics and Finance*, 45.
- Madigan, D. M., & Raftery, A. E. (1994). Model Selection and Accounting for Model Uncertainty In Graphical Models Using Occam's Window. *Journal of the American Statistical Association*, 89. <http://dx.doi.org/10.1080/01621459.1994.10476894>
- MalloWS, C. L. (1973). Some comments on C p. *Technometrics*, 15. <http://dx.doi.org/10.1080/00401706.1973.10489103>
- Ravazzolo, F., Van Dijk, H. K., & Verbeek, M. (2007). *Predictive Gains from Forecast Combination Using Time-varying Model Weight*. Econometric Institute Report 2007-26. <http://dx.doi.org/10.2139/ssrn.1012574>
- Schorfheide, F. (2000). Loss Function-Based Evaluation of DSGE Models. *Journal of Applied Econometrics*, 15(6). <http://dx.doi.org/10.1002/jae.582>
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6. <http://dx.doi.org/10.2307/2958889>
- Song, H., Li, G., Witt, S. F., & Athanasopoulos, G. (2011). Forecasting Tourist Arrivals Using Time-Varying Parameter Structural Time Series Models. *International Journal of Forecasting*, 27(3). <http://dx.doi.org/10.1016/j.ijforecast.2010.06.001>
- Swamy, P. A. V. B., & Mehta, J. S. (1975). Bayesian and Non-Bayesian Analysis of Switching Regressions and a Random Coefficient Regression Model. *Journal of the American Statistical Association*, 70. <http://dx.doi.org/10.1080/01621459.1975.10482478>
- Swamy, P. A., Tavlas, G. S., Hall, S. G., & Hondroyannis, G. (2010). Estimation of Parameters in the Presence of Model Misspecification and Measurement Error. *Studies in Nonlinear Dynamics and Econometrics*, 14(1). <http://dx.doi.org/10.2202/1558-3708.1743>
- Terui, N., & Kariya, T. (1997). Testing Gaussianity and linearity of Japanese Stock returns. *Financial Engineering and the Japanese Markets*, 4. <http://dx.doi.org/10.1023/A:1009692319131>
- Terui, N., & Van Dijk, H. K. (2002). Combined Forecasts from Linear and Nonlinear Time Series Models. *International Journal of Forecasting*, 18(3). [http://dx.doi.org/10.1016/S0169-2070\(01\)00120-0](http://dx.doi.org/10.1016/S0169-2070(01)00120-0)
- Timmermann, A. (2006). *Forecast Combinations*. In G. Elliott, W. J. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. 1).
- Woodford, M. (2003). *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton: Princeton University Press.

Notes

Note 1. For more details on the available literature on BMA, see Hoeting, Raftery, and Madigan (2002).

Note 2. The OLS estimation with these assumptions is coded as a quadratic program inside R software.

Note 3. Our analysis is based on Madigan and Raftery (1994).

Note 4. All Bayesian weights and calculations are estimated by using BMS package inside R software.

Note 5. Coefficient drivers are just a combination of some variables that jointly explain changes in β_{it} to a reasonable extent.

Note 6. The full coefficients method means that incorporating both the unbiased and biased components.

Note 7. For more details, see Herman and van Dijk (2002) and Song et al. (2011).

Note 8. The model is estimated using forecasting package inside R software.

Note 9. For more details, see Harvey (1989).

Note 10. For more details, see for example Schorfheide (2000). Del Negro and Schorfheide (2004) reports that $\hat{\lambda}$ is around 0.6.

Appendix A:

DSGE-VAR Estimation Method

The key equations include:

$$x_t = E_t x_{t+1} - \tau^{-1}(R_t - E_t[\pi_{t+1}]) + (1 - \rho_g)g_t + \rho_z \frac{1}{\tau} z_t \quad A.1$$

$$\pi_t = \frac{\gamma}{r^*} E_t[\pi_{t+1}] + \kappa(x_t - g_t) \quad A.2$$

$$R_t = \rho_R R_{t-1} + (1 - \rho_R)(\psi_1(\pi_t) + \psi_2 x_t) + \epsilon_{R,t} \quad A.3$$

Where the shocks processes are expressed as:

$$z_t = \rho_z z_{t-1} + \epsilon_{z,t} \quad A.4$$

$$g_t = \rho_g g_{t-1} + \epsilon_{g,t} \quad A.5$$

Equation (A.1) can be defined as an intertemporal consumption Euler equation while equation (A.2) controls inflation behaviour. In addition, equation (A.3) is the monetary policy rule. Moreover, the shock process $\epsilon_{z,t}$ and $\epsilon_{g,t}$ can be interpreted as technology shock and the shock of government spending respectively.

The DSGE-VAR system can be estimated by following, Christiano (2007). Also, let y_t refers to the column vector of n observed variables, which is assumed to be also the same variables in the VAR, then:

$$Y_t = X\Phi + U \quad \text{and} \quad u_t \sim N(0, \Sigma_u) \quad A.6$$

Where Y_t is a matrix of y'_t , U is a matrix of u'_t , X is a matrix of rows $y'_{t-1}, y'_{t-2}, \dots, y'_{t-p}$ and finally, $\Phi = \phi'_0, \phi'_1, \dots, \phi'_p$. The BVAR posterior is estimated conditional on the value of the relative weights λ associated with both models, DSGE and the unrestricted VAR. Suppose that the set of all possible values for λ is Λ as $\{\Lambda \equiv \lambda_1, \lambda_2, \dots, \lambda_i, \dots, \lambda_q\}$ for all i and $\lambda_i > 0$.

The parameters of the VAR could be computed analytically such that maximizes $P(Y|\lambda)$ as follows:

$$\hat{\lambda} = \arg \max_{\lambda \in \Lambda} P(Y|\lambda) \quad A.7$$

Where, $P(Y|\lambda)$ is the marginal data density of the model. That is:

$$P(Y|\lambda) = \int_{\Sigma_u^*, \Phi^*, \theta^*} p(Y, \theta, \Sigma_u, \Phi|\lambda) d(\Sigma_u, \Phi, \theta) \quad A.8$$

Where: θ is DSGE parameters be denoted by the vector; $\Sigma_u^*, \Phi^*, \theta^*$ are potential values for the parameters Σ_u, Φ, θ respectively.

Appendix B

Results of forecast combinations

Table B1. Different weight of static models

	BVAR	DSGE-VAR	MR	TV	NN	GARCH (GED)	TARCH (t-dist)	GARCHSK-M
EQ	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
MFE	0.100876	0.145764	0.130754	0.1403	0.103248	0.133094	0.139841	0.133094
GR	0.1	0.1	0.007	0.08	0.117	0	0.274	0
MMA	0.433	0.232	0.04	0.1	0.04	0.15	0.005	0.15
BMA	0.194373	0.389252	0.078021	0.066659	0.055536	0.062658	0.075417	0.062658
BMA-OLS	0.104538	0.328523	0.096681	0.09267	0.088625	0.092289	0.09781	0.092289
BMA-MFE	0.117024	0.332757	0.096221	0.087866	0.079628	0.087559	0.097799	0.087559

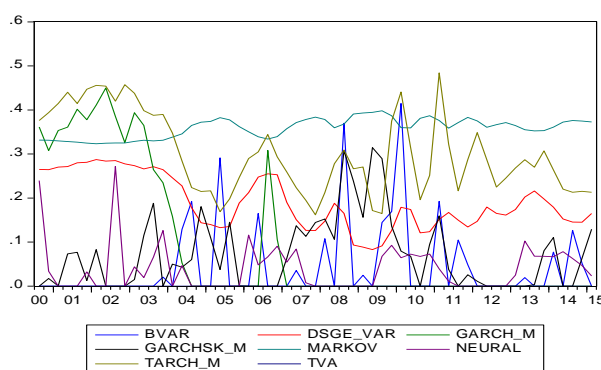


Figure B1. Weights assigned to different models according to Random Walk

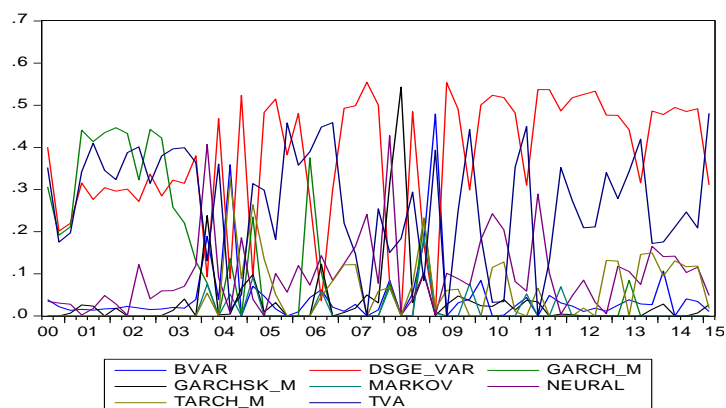


Figure B2. Weights assigned to different models according to ETVC

Note. The estimation of ETVC is based on the full coefficients method in TVC approach, with two driver sets inside each time varying-coefficient which include the first lag for the associated point forecast variable and its quadratic form to account for the nonlinearity inside the combination scheme. As we tried to augment more nonlinear terms, but the quadratic forms were the most significant ones.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).