ROMANIAN REVIEW OF SOCIAL SCIENCES

Romanian Review of Social Sciences (2013), No.5

rrss.univnt.ro

A STRATEGY TO IMPROVE THE FORECASTS BASED ON VAR MODELS USING HODRICK-PRESCOTT FILTER

Mihaela Simionescu
Institute for Economic Forecasting of the Romanian Academy, Bucharest
E-mail: mihaela_mb1@yahoo.com

Abstract.

The main goal of this research is to find a suitable strategy to improve the accuracy of predictions based on a VAR model. The forecasts of annual inflation and unemployment rate made for Romania on the horizon 2010-2012 using a VAR(1) model were improved by proposing new predictions based on a VAR(2) model. But, in this last case the VAR(2) models used differentiated data series for inflation and unemployment rate that were smoothed using Hodrick-Prescott filter. However, the naïve forecasts based on random walk model provided more accurate predictions for both variables.

Key-words: forecasts, accuracy, VAR models, Hodrick-Prescott filter

1. Introduction

The objective of this research is to propose a strategy of improving the forecasts based on VAR models. Many econometric models were used to put in evidence the relationship between inflation and unemployment rate. In Romania the Phillips curve is not checked, another method being necessary. A VAR model is a good choice, a reciprocal causality being identified between the two variables.

Transformed variables are used in this study because VAR models work with stationary data series. Two VAR models were proposed, for the second one the stationary data series being filtered using Hodrick-Prescott filter. The predictions based on the last model are more accurate than the ones based on the first model. However, naïve forecasts outperformed the predictions based on the proposed VAR models.

2. Literature

In literature some of the researchers identify two directions regarding the macroeconomic modelling: keynesian models and VAR models, introduced by Sims (1980).

Some authors, as Eigner (2009), analyse the forecasting process using two dimensions: univariate forecasting (that without model, like smoothing techniques and forecasting with model like ARIMA procedure) and multivariate forecasting (open loop system (multiple regression and transfer functions) and close loop system (VAR and SVAR for stationary form and VEC and SVEC for non-stationary form of data)).

Sims introduced VAR models as a reply to simultaneous equations models, because the last ones, according to Andrei and Bourbonnais (2008), did not take into account essential information from data set like: low accuracy of predictions variables causality and restrictions elimination related to exogenous character of the variables. First of all, we have to check if the data has one of the following properties: stationarity or co-integration. Therefore, there are structural vectorial autoregressive models and vector errors correction models (VECM). Dynamic model of type VECM is used in making predictions and in evaluating the effects of random shocks over variables. For VAR models if we choose different lags, the estimations will be different. VAR models are used in making predictions by Litterman in 1986, this author introducing a Bayesian change. MCNees made predictions using VAR models, noticing that the root mean squared error for some of the studied variables decrease when the forecasting horizon is reduced from one to eight quarters. Predictions based on VAR models were built by many authors, some of them being Fildes (1988), Edlund and Karlsson (1993), Webb (1995), Simkins (1995), Ramos (2003), Thomakos and Gómez (2004), Guerrero (2006), Kano (2008), Moench (2008), Sinclair and Stekler (2009).

Clark and McCracken (2006) explained the ability of different methods of improving the forecasts accuracy in real time if an unstable VAR model is used.

African Institute of Applied Economics shows that nowadays the VAR models are used to check the forecasts based on NKAPC models (New Keynesian – Augmented Philips Curve).

VAR models used to analyse the policies should generate unconditioned forecasts, but also conditioned ones in simulation policies. The predictions are based on structural VAR models. However, this type of models was rather criticised because it generates hard to interpret and low accuracy forecasts. Pecican, Tănăsoiu and Iacob (2001) enumerated the limits of VAR models: non-theoretical character of the model, questionable interpretation of estimation results and superficial identification of lag length.

Gupta and Kabundi (2009) proved for data taken from south-african economy that a dynamic factor model generated better forecasts than a Bayesian VAR model for variables like GDP rate, inflation rate and nominal interest rate.

VAR models used to analyse the policies should be unconditional, but also conditional. The predictions are based on the structural form of a VAR model, even if this type of models is quite criticized because of the low degree of accuracy.

The accuracy of forecasts based on VAR models can be measured using the trace of the mean-squared forecasts error matrix, according to or generalized forecasts error second moment, according to Clements and Hendry (2003).

Robinson W. (1998) got a better accuracy for predictions based on VAR model for some macroeconomic variables with respect to other models like transfer functions.

Lack C. (2006) found out that combined forecasts based on VAR models are a good strategy of improving the predictions accuracy.

3. Forecasts based on VAR models. The accuracy assessment and improvement

The variables used in building VAR models are the inflation rate and the unemployment rate of Romania. Annual data are used, the time series horizon being 1991-2012. The index of prices used to determine the inflation rate is expressed in comparable prices (1990=100). The initial data series are not stationary, being integrated of order 1. Therefore, the data were stationarized by making a differentiation of first order. The new variables are denoted by delta_ir and delta_ur, being used to construct the VAR models. Most of the lag length criteria recommend a VAR with lag 1.

We will try to construct VAR models using the filtered data series based on Hodrick-Prescott filter (lambda=100). The Hodrick–Prescott (HP) filter is very used in macroeconomics to extract the trend of the data series and separate the cyclical component of the time series. The smoothed data gotten are more sensitive to long term changes.

The initial data series is composed of trend and cyclical component:

$inf_t = tr_t + c_t$.

Hodrick and Prescott (1997) suggest the solve of the minimization problem:

$$\min_{\{tr_t\}t=1,T} \sum_{t=1}^{T}(inf_t - tr_t)^2 + \gamma \sum_{t=2}^{T-1}(\nabla^2 tr_{t+1})^2$$

γ- penalty parameter

The solution to the above equation can be written, according to Bratu (Simionescu) (2013) as:

$inf_t = (\gamma F + I_T) \cdot tr_T$

inf_t- vector of the initial data series of the inflation rate

$$F=\begin{bmatrix} 1 & -2 & 1 & 0 & \dots & & & & & & & \\ -2 & 5 & -4 & 1 & 0 & \dots & \cdots & & & & & 0 \\ 1 & -4 & 6 & -4 & 1 & 0 & & & & & & \\ & & \vdots & & & & \ddots & & \vdots & & & \\ & & & & & & & & & & & \cdots \\ & & & & & & \cdots & 0\ 1 & -4 & 6 & -4 & 1 & 0 \\ & & & & & & & 0\ 0 & 1 & -4 & 6 & -4 & 1 \\ 0 & & 0 & & & & \dots & 0 & 0 & 1 & -4\ 5 & -2 \end{bmatrix}$$

The trend is calculated as: $tr_T = [(\gamma \cdot F + I_T \quad)]^{-1} \cdot inf_T$.

Figure. 1: Filtered inflation rate data series using Hodrick-Prescott filter

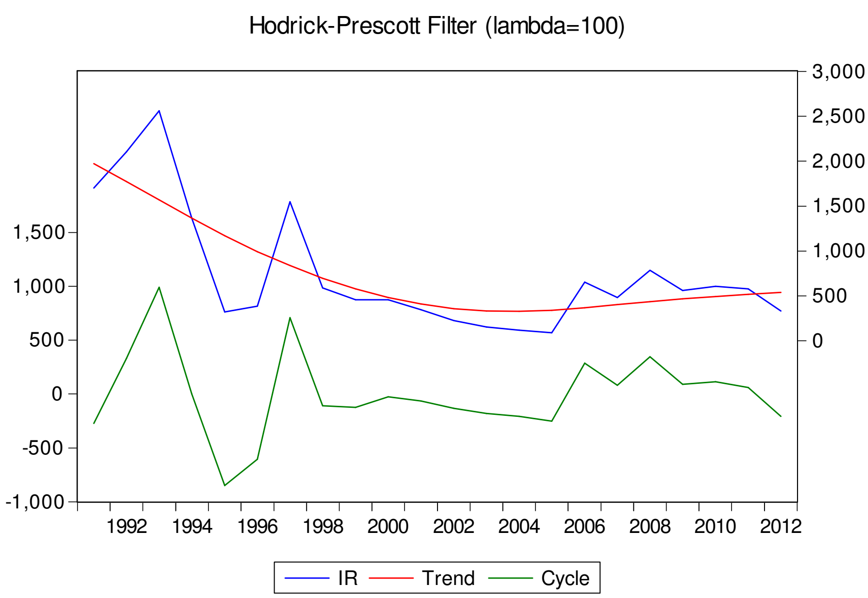
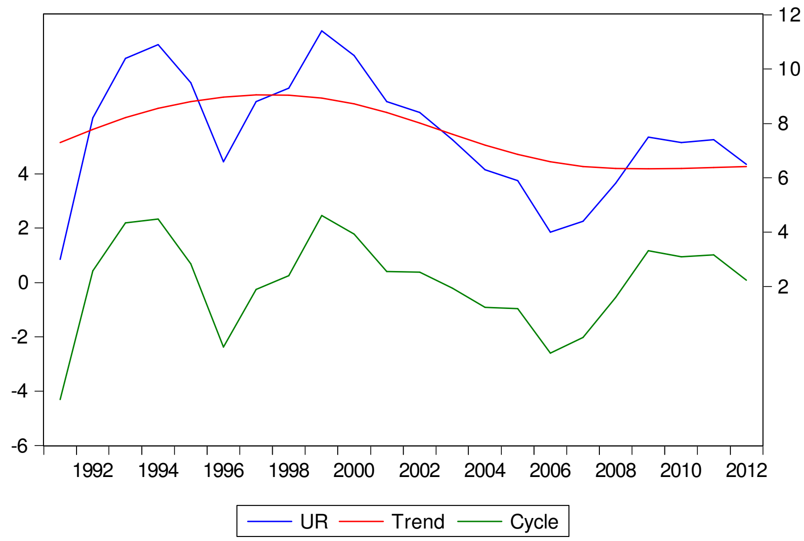


Figure 2: Filtered unemployment rate data series using Hodrick-Prescott filter

Hodrick-Prescott Filter (lambda=100)



The new data series are not stationary and a differentiation of different orders did not solve the problem.

Another possibility is to filter the stationary data obtained by making the differentiation of order 1. New variables are denoted by: f_d_ir and f_d_ur, the data sets being stationary.

Table 1: The Phillips-Perron test for checking the stationarity of the data series

Variable	Model with trend and constant	Model without trend and constant	Model with constant
d_ir	PP= -5.059481 1%, 5%, respectively 10% critical values: -4.5348 -3.6746 -3.2762	PP= -2.545634 1%, 5%, respectively 10% critical values: -2.6968 -1.9602 -1.6251	PP= -5.065676 1%, 5%, respectively 10% critical values: -3.8304 -3.0294 -2.6552

d_ur	PP= -6.951421 1%, 5%, respectively 10% critical values: -4.571559 -3.690814 -3.286909	PP= -2.69946 1%, 5%, respectively 10% critical values: -1.961409 -1.606610 -2.046578	PP= -3.751905 1%, 5%, respectively 10% critical values: -3.857386 -3.040391 -2.660551
f_d_ir	PP= -4.571559 1%, 5%, respectively 10% critical values: -1.758757 -3.690814 -3.286909	PP= -2.699769 1%, 5%, respectively 10% critical values: -0.451949 -1.961409 -1.606610	PP= -3.857386 1%, 5%, respectively 10% critical values: -3.040391 -2.660551 -1.059920
f_d_ur	PP= -3.885282 1%, 5%, respectively 10% critical values: -4.616209 -3.710482 -3.297799	PP= -3.992245 1%, 5%, respectively 10% critical values: -2.708094 -1.962813 -1.606129	PP= -4.067445 1%, 5%, respectively 10% critical values: -3.886751 -3.052169 -2.666593

Four of the lag length criteria (AIC, SC, HQ and FPE) recommend the selection of a lag equal with two. Gutierrez, Souza and de Carvalho Guillen O.T. (2009) showed that AIC is the most suitable criterion, taking into account that for small size samples SC criterion tends to select an underparametrized model.

Table 2: Lag length criteria for VAR models

VAR(1) model						
Lag	LogL	LR	FPE	AIC	SC	HQ
0	-162.8573	NA*	908447.9	19.39497	19.49300	19.40472
1	-159.4452	5.619892	980537.1*	19.46414*	19.75822*	19.49337*
VAR(2) model						
Lag	LogL	LR	FPE	AIC	SC	HQ

0	-106.6389	NA	2709.263	13.57986	13.67643	13.58481
1	-5.480953	164.3816	0.014528	1.435119	1.724840	1.449955
2	14.70491	27.75557*	0.001988*	-0.588114*	-0.105246*	-0.563387*

According to AIC, HQ, FPE and SC values, the best model for the differentiated variables is a VAR(1).

All the lag length criteria indicated that a VAR(2) would be the most suitable choice for the filtered values of the differentiated variables . The VAR model used to build the one-step-ahead forecasts are presented in the ***Appendix 1***.

The Granger causality test is applied for stationary data series in order to establish if a variable is cause for the other one. In Granger acceptance, a variable X is cause for Y if better predictions result when the information provided by X is taken into account.

Table 3: VAR Granger causality tests

Hypothesis	Prob.
d_ir does not Granger cause d_ur	0.2233
d_ur does not Granger cause d_ir	0.2460
f_d_ir does not Granger cause f_d_ur	0.3856
f_d_ur does not Granger cause f_d_ir	0.4375

The results of Granger causality test show that d_ur is the cause of d_ir, but also d_ir is the cause of d_ur. A reciprocal causality is met also for the other variables: f_d_ir and f_d_ur.

VAR Residual Portmanteau Tests are used to test the errors' autocorrelation for both identified models (VAR(1) and VAR(2)). The assumptions of the test are formulated as:

H0: the errors are not autocorrelated

H1: the errors are autocorrelated

Table 4: Residual Portmanteau test for checking errors' autocorrelation

Model	Lags	Q-Stat	Prob.	Adj Q-Stat	Prob.
VAR(1)	1	1.861990	NA*	1.978365	NA*

	2	6.775239	0.1483	7.546713	0.1097
	3	8.571188	0.3797	9.727508	0.2847
	4	10.81789	0.5446	12.66550	0.3938
	5	12.40660	0.7155	14.91618	0.5308
	6	14.58583	0.7996	18.28407	0.5687
	7	16.93701	0.8513	22.28108	0.5625
	8	17.37672	0.9408	23.11165	0.7274
	9	18.77391	0.9694	26.08068	0.7600
	10	20.72929	0.9804	30.82945	0.7128
	11	21.29929	0.9933	32.44445	0.7964
	12	22.33602	0.9973	35.96935	0.8002
VAR(2)	1	1.861990	NA*	1.978365	NA*
	2	6.775239	0.1483	7.546713	0.1097
	3	8.571188	0.3797	9.727508	0.2847
	4	10.81789	0.5446	12.66550	0.3938
	5	12.40660	0.7155	14.91618	0.5308
	6	14.58583	0.7996	18.28407	0.5687
	7	16.93701	0.8513	22.28108	0.5625
	8	17.37672	0.9408	23.11165	0.7274
	9	18.77391	0.9694	26.08068	0.7600
	10	20.72929	0.9804	30.82945	0.7128
	11	21.29929	0.9933	32.44445	0.7964
	12	22.33602	0.9973	35.96935	0.8002

For the lag 1 up to 12, the probabilities (Prob.) of the tests are greater than 0.05, fact that implies that there is not enough evidence to reject the null hypothesis (H0). So, we do not have enough reasons to say that the errors are auto-correlated. So, after the application of Residual Portmanteau Test, the conclusion is that there are not autocorrelations between errors for VAR(1) and VAR(2) models.

The homoscedasticity is checked using a VAR Residual LM test. If the value of LM statistic is greater than the critical value, the errors series is heteroscedastic.

Table 5: Residual LM test for checking errors' homoscedasticity

Model	Lags	LM-Stat	Prob
VAR(1)	1	7.325802	0.1196

	2	5.287560	0.2590
	3	1.779197	0.7763
	4	2.197819	0.6994
	5	1.558136	0.8163
	6	2.015393	0.7329
	7	2.413127	0.6603
	8	0.444662	0.9787
	9	1.476469	0.8308
	10	3.068824	0.5464
	11	7.325802	0.1196
	12	5.287560	0.2590
VAR(2)	1	7.325802	0.1196
	2	5.287560	0.2590
	3	1.779197	0.7763
	4	2.197819	0.6994
	5	1.558136	0.8163
	6	2.015393	0.7329
	7	2.413127	0.6603
	8	0.444662	0.9787
	9	1.476469	0.8308
	10	3.068824	0.5464
	11	1.111814	0.8924
	12	1.194993	0.8789

LM test shows that there is a constant variance of the errors for both models, because of the values greater than 0.05 for the probability.

The normality tests are applied under the Cholesky (Lutkepohl) orthogonalization. If the Jarque-Bera statistic is lower than the critical value there is not enough evidence to reject the normal distribution of the errors.

Table 6: Jarque-Bera test for checking normal distribution

Model	Component	Jarque-Bera	df	Prob.
VAR(1)	1	0.130640	2	0.9368
	2	0.278859	2	0.8699
	Joint	0.409499	4	0.9817

VAR(2)	1	0.637404	2	0.7271
	2	0.361869	2	0.8345
	Joint	0.999273	4	0.9099

The Residual normality test provided probabilities greater than 0.05, fact that implies that the errors series has a normal distribution when Cholesky (Lutkepohl) Orthogonalization is applied.

These VAR models are used to make forecasts on the horizon 2010-2012. Then, the accuracy of forecasts is checked to establish the better model.

Table 7: Predictions of inflation and unemployment rate based on VAR(1) model

Type of forecasts/Years	Inflation rate compared to the previous period (%)	Unemployment rate (%)
One-step-ahead forecasts		
2010	5.731	7.635
2011	9.106	7.265
2012	8.544	7.364

For both variables in 2011 the predicted values increased compared to the forecast in 2010, but a slightly decrease is observed in 2012. However, this evolution is in contrast with the real situation that shows a decrease tendency.

Table 8: Predictions of inflation and unemployment rate based on VAR(2) model

Type of forecasts/Years	Inflation rate compared to the previous period (%)	Unemployment rate (%)
One-step-ahead forecasts		
2010	5.947	7.2
2011	6.785	6.9
2012	4.17	7.2

For the inflation rate in 2011 an increase is observed compared to the previous year, but a rather high decrease is anticipated for 2012. The unemployment rate decreased in 2011 and then it went to the predicted level for 2012.

The prediction error is computed as the difference between the effective value and the forecasted one of a variable X and it is denoted by e_X. For the number of forecasts on the horizon it is used the notation "n". The most frequently used statistical measures for assessing the forecasts accuracy, according to Bratu (2012), are : Root Mean Squared Error (RMSE) : $RMSE=\sqrt{\frac{1}{n}\sum_{j=1}^{n}e_X^2}$, Mean error (ME) : $ME=\frac{1}{n}\sum_{j=1}^{n}e_X$ and Mean absolute error (MAE) : $MAE=\frac{1}{n}\sum_{j=1}^{n}|\ e_X\ |$. RMSE is influenced by outliers. These absolute measures depend on the unit of measurement, this disadvantage being eliminated unless if the indicators are expressed as percentage.

U Theil's statistic, used in making comparisons between predictions, can be used in two variants, presented also by the Australian Treasury.

The next notations are used:

a- actual/registered value of the analyzed variable

p- value for the predicted variable

t- time

e- error (difference between actual value and the forecasted one)

n- number of periods

U1 takes value between 0 and 1, a closer value to zero indicating a better accuracy for that prediction. If there are alternative forecasts for the same variable, the one with the lowest value of U1 is the most accurate.

$$U_1=\frac{\sqrt{\sum_{t=1}^{n}[a_t-p_t]^2}}{\sqrt{\sum_{t=1}^{n}a_t^2}+\sqrt{\sum_{t=1}^{n}p_t^2}}$$

Instead of U1, the mean absolute scaled error can be computed (MASE= mean | es_t |.), the result being the same: $es_t=\frac{e_t}{\frac{1}{n-1}\sum_{i=2}^{n}|X_i-X_{i-1}|}$

For making comparisons with the naive forecasts U2 Theil's coefficient is used.

$$U_2 = \sqrt{\frac{\sum_{i=1}^{n-1}\left[\frac{p_{t+1}-a_{t+1}}{a_t}\right]^2}{\sum_{i=1}^{n-1}\left[\frac{a_{t+1}-a_t}{a_t}\right]^2}}$$

If U_2 =1=> there are not differences in terms of accuracy between the two forecasts to compare
If U_2 <1=> the forecast to compare has a higher degree of accuracy than the naive one
If U_2 >1=> the forecast to compare has a lower degree of accuracy than the naive one

Table 9: The accuracy of forecasts based on VAR(1) models

Accuracy measure	Inflation rate forecasts	Unemployment rate forecasts
ME	-2.7237	-2.3519
MAE	2.9630	2.3519
MAPE	0.7324	0.5734
RMSE	3.5735	2.6358
MASE	1.6819	0.9519
U2	2.4127	1.6394

Source: own calculations using Excel

The forecasts based on VAR(1) models on the horizon 2010-2012 have a low degree of accuracy. However, the unemployment rate predictions are more accurate than the inflation ones, according to MASE values. The naïve forecasts based on random walk are superior to the VAR ones in this case, having a value greater than 1 for U2. The errors for unemployment rate forecasts represent in average 57,34% of the registered value. The percentage is rather higher for inflation predictions on the horizon 2010-2012, being 73,24%. Both types of predictions are overestimated, this overestimation being persistent for unemployment rate. All the accuracy indicators have lower values for unemployment rate forecasts.

Table 10: The accuracy of forecasts based on VAR(2) models

Accuracy measure	Inflation rate forecasts	Unemployment rate forecasts
ME	-0.5640	-0.8667
MAE	0.6593	0.8667
MAPE	0.1492	0.1510
RMSE	1.3100	1.7720
MASE	0.0690	0.0765
U2	1.9863	1.2371

Source: own calculations using Excel

All the accuracy indicators have lower values for the forecasts based on VAR(2) models. However, the values of U2 are greater than 1, fact that shows the superiority of random walk models. The predictions are still overestimated, but the errors for the inflation rate are only 14,92% of the registered value during 2010-2012. MASE indicators, but also the other ones put in evidence that inflation forecasts are more accurate than those made for the unemployment on the horizon 2010-2012.

4. Conclusions

Two VAR models were proposed for transformed variables based on the annual inflation rate and unemployment rate. The main assumptions were checked to test the validity of the proposed models. The residuals are not serial correlated and homoscedastic. There is not enough evidence to reject the hypothesis of normal distributions. A VAR(1) model was proposed for the differentiated inflation and unemployment rates and a VAR(2) model for these filtered variables. The predictions based on these models were assessed, resulting a higher accuracy for the VAR(2) forecasts that use filtered variables based on Hodrick-Prescott filter.

However, the naïve forecasts are better than those based on mentioned VAR models, the recent assumption in literature regarding the superiority of simple predictions is checked also in this case.

Appendix 1

Model used to make predictions for 2010-2012

Forecasts horizon	VAR(1) model	VAR(2) model
2010-2012	DELTA_IR =	F_D_IR =

	0.0215574796162*DELTA_IR(-1) - 93.8357834332*DELTA_UR(-1) - 74.2642116223 DELTA_UR = 0.000775168016877*DELTA_IR(-1) + 0.202122675386*DELTA_UR(-1) - 0.0326537893885	0.964032415782*F_D_IR(-1) - 0.0143112874561*F_D_IR(-2) - 3.08703563918*F_D_UR(-1) - 15.7133463761*F_D_UR(-2) + 11.4832475032 F_D_UR = 0.00214188660948*F_D_IR(-1) - 0.00177897936239*F_D_IR(-2) + 1.79846893207*F_D_UR(-1) - 0.793276660413*F_D_UR(-2) + 0.0367573329808

References

Andrei T., Bourbonnais R. (2008), *Econometrie*, Editura Economică, Bucureşti

Athanasopoulos G., Teixeira de Cavalho Guillen O., Isster J.V., Vahid F. (2010), Model selection, estimation and forecasting in VAR models with short-run and long-run restrictions, *Working papers series*, No. 253, Banco Central do Brasil, p. 1-49

Bratu M. (2012), *Strategies to Improve the Accuracy of Macroeconomic Forecasts in USA*, LAP LAMBERT Academic Publishing, ISBN-10: 3848403196, ISBN-13**:** 978-3848403196, http://www.amazon.com/Strategies-Improve-Accuracy-Macroeconomic-Forecasts/dp/3848403196

Bratu (Simionescu) M. (2013), *Filters or Holt Winters technique to improve the forecasts for USA inflation rate ?*, Acta Universitatis Danubius. Œconomica Vol 9, issue no. 1/2013

Clark, T. E., McCraken, M. W. (2006), ”Forecast with Small Macroeconomic VARS in the Presence of Instabilities”, *The Federal Reserve Bank of Kansas City Economic Research Department*

Clements M.P., Hendry D.F. (2003), “On the limitations of comparing mean squared forecast errors (with discussion)”, *Journal of Forecasting*, Volume 12, p. 617-639

Gupta R. & Kabundi A. (2009), "Forecasting Macroeconomic Variables Using Large Datasets: Dynamic Factor Model versus Large-Scale BVARs," Working Papers 143, Economic Research Southern Africa.

Gutierrez C.E.C., Souza R.C., de Carvalho Guillen O.T. (2009), Selection of optimal lag length in cointegrated VAR models with weak form of common cyclical features, article available at
http://mpra.ub.uni-muenchen.de/22550/1/Gutierrez-Souza-Guillen_revis_o_final.pdf

Hodrick R. & Prescott E.C. (1997). Postwar U.S. Business Cycles: An empirical investigation, *Journal of Money, Credit and Banking,* 1-16, p. 84-90.

Lack C. (2006), Forecasting Swiss inflation using VAR models, *Swiss National Bank Economic Studies*, No. 2/2006

Pecican E., Tănăsoiu O., Iacob A.I. (2001), *Modele macroeconomice*, Editura ASE

Robinson W. (1998), Forecasting inflation using VAR Analysis, *Bank of Jamaica*, article available at:
http://www.boj.org.jm/uploads/pdf/papers_pamphlets/papers_pamphlets_forecasting_inflation_using_var_analysis.pdf

Sims C.A. (1980), ”Macroeconomics and reality”, *Econometrica,* volume 48, Issuse 1, p. 1-48

Teräsvirta T., van Dijk D., Medeiros M.C. (2005), ”Linear models, smooth transition autoregressions, and neural networks for forecasting, macroeconomic time series: A re-examination”, *International Journal of Forecasting*, Volume 21, Issue 4, Octombrie-Decembrie 2005, p. 755-774

Eigner F. (2009), ”Forecasting and VAR models”, *UK Econometric Forecasting*, Prof. Kunst