International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 2, 2003, pp 19–25

VAR Modeling with Mixed Series

Md. Atikur Rahman Khan
Dept. of Population Science and Human Resource Development,
University of Rajshahi, Bangladesh

Md. Ayub Ali
Department of Statistics,
University of Rajshahi, Bangladesh

[Received October 27, 2002; Revised June 21, 2003; Accepted August 26, 2003]

Abstract

The purpose of the present study was to confirm a method of VAR (Vector Autoregression) modeling when the variables are mixed in nature (i.e., a mixture of stationary and non-stationary variables). Using restricted cross validity predictive power (RCVPP) and root mean squared forecast error (RMSFE), it was found that working at level gives better results in this case.

Keywords and Phrases: Vector Autoregression (VAR), Vector Error Correction (VEC), Restricted Cross Validity Predictive Power (RCVPP), and Root Mean Squared Forecast Error (RMSFE).

1 Introduction

The VAR modeling is very important for the forecast of time series variables when they are interrelated, however it has some limitations. The VAR model may deal with both stationary and non-stationary variables. If all the variables are non-stationary then they should be transformed (usually by differencing) to make it stationary (Gujarati, 1995). Sims (1980) recommended against differencing even if the variables contain a unit root (i.e., non-stationary). According to Sims (1980), Harvey (1990) suggests to work at level if the variables are mixed in nature. Again, prediction with non-stationary data may provide unsatisfactory forecast (Cleary and Hay, 1980; Gujarati, 1995; and Pankratz, 1991). The VAR analysis is to determine the interrelationships among the variables, not the parameter estimates. But, we could not avoid the modeling behavior and the interrelationships of the variables in the model. Although differencing throws away the comovements in the data set we may build two types of models to examine their predictive performance using levels of variables and difference of non-stationary series along with the stationary series. If a set of variables are all I(1) and their linear combination is I(0) then that set of variables are cointegrated (Engel and Granger 1987). But, for mixed series integrated orders are not same. So there arise complexities in regard to usual error correction (Johansen, 1996). In this case the approach proposed by Blanchard and Quah (1989) can be applied to build a VEC model. So, what we can do when the variables are mixed in nature (i.e., mixture of stationary and non-stationary variables), is a problem. Therefore, the purpose of the present study is to investigate a concrete decision whether we should work at level or not when the variables are mixed in nature.

2 Methods

If we include an intercept term, and exogenous variables in the model, then the model can be written as:

$$Y_t = \mu + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \beta x_t + u_t \tag{1}$$

where, μ is a vector of intercept terms (constants), Y_t is a vector of endogenous variables, x_t is a vector of exogenous variables, $A_1, A_2, \cdots, A_p$ are matrices of coefficients to be estimated, and u_t is a vector of innovations that are correlated with each other but are uncorrelated with their own lagged values and also with Y_{t-1} through Y_{t-p} and x_t After fitting a model we need to examine its validity. A model validation technique to examine the validity of the fitted models is known as restricted cross validity predictive power (Khan and Ali, 2003). The restricted cross validity predictive power, ρ^2_{rcv} is:

$$\rho^2_{rcv} = 1 - \frac{(n-1)(n-2)(n+1)}{n(n-k-1)(n-k-2)}(1-R^2) \tag{2}$$

where n is the number of observations, k is the number of predictors, R^2 is the correlation coefficient between observed and fitted response values that satisfies the restriction

$$R^2 \geq \frac{n(n-k-1)(n-k-2)}{(n-1)(n-2)(n+1)}; \quad n > k+2$$

Let us assume that R_1^2 and R_2^2 are the coefficients of multiple determinations from VAR model of order p_1 at level and that from the VEC model (based on the logic of Blanchard and Quah (1989)) of order p_2, respectively. Then the restricted cross validity predictive power for VAR model at level and that for VEC model can be given respectively as:

$$\rho^2_{1rcv} = 1 - \frac{(n-1)(n-2)(n+1)}{n(n-k_1-1)(n-k_1-2)}(1-R_1^2); \quad n > k_1+2 \tag{3}$$

and

$$\rho^2_{2rcv} = 1 - \frac{(m-1)(m-2)(m+1)}{m(m-k_2-1)(m-k_2-2)}(1-R_2^2); \quad m > k_2+2 \tag{4}$$

where, k_1 is the number of predictors in the VAR model, k_2 is the number of predictors in the VEC model, and m is the number of observation in the VEC model, i.e., $m = n-1$ as one point is lost for error correction. Now, let us assume that the number of predictors in both the cases are equal i.e. $k_1 = k_2 = k$ (say). Putting $m = n-1$ and $k_1 = k_2 = k$ we get:

$$\rho^2_{1rcv} = 1 - \frac{(n-1)(n-2)(n+1)}{n(n-k-1)(n-k-2)}(1-R_1^2); \quad n > k+2 \tag{5}$$

and

$$\rho^2_{2rcv} = 1 - \frac{(n-2)(n-3)n}{(n-1)(n-k-2)(n-k-3)}(1-R_2^2); \quad n > k+3 \tag{6}$$

From Eq.(5) and Eq.(6) we can write

$$\begin{aligned} \frac{1-\rho^2_{1rcv}}{1-\rho^2_{2rcv}} &= \frac{(n-1)(n+1)}{n(n-k-1)} \times \frac{(n-1)(n-k-3)}{n(n-3)} \times \frac{1-R_1^2}{1-R_2^2} \\ &= \frac{n^2-1}{n^2} \times \frac{n^2-nk-4n+k+3}{n^2-nk-4n+3k+3} \times \frac{1-R_1^2}{1-R_2^2} \end{aligned} \tag{7}$$

where, R_1^2 is obtained using n observed values and R_2^2 is obtained from $m = n-1$ transformed values. Since the exact relationship between R_1^2 and R_2^2 is unknown so we may replace $R_i^2 (i = 1, 2)$ with their population mean.

We know that for k regressors:

$$F = 1 - \frac{R^2/k}{(1-R^2)/(n-k-1)} \tag{8}$$

$$\begin{aligned} \Rightarrow \frac{R^2}{1-R^2} &= \frac{k}{n-k-1} \\ \Rightarrow R^2 &= \frac{1}{1+\frac{n-k-1}{k}F_1}, \text{where} F_1 = \frac{1}{F} \sim F_{n-k-1,k} \\ \Rightarrow R^2 &= \frac{1}{1+\frac{n-k-1}{k}F_1} \sim \beta_1\left(\frac{k}{2}, \frac{n-k-1}{2}\right) \end{aligned} \tag{9}$$

Thus, the expected value of R^2 is $\frac{k}{n-1}$. After replacing $R_i^2 (i = 1, 2)$ with their population means we can write Eq.(7) as:

$$\frac{1-\rho^2_{1rcv}}{1-\rho^2_{2rcv}} = \frac{n^2-1}{n^2} \times \frac{n^2-nk-4n+k+3}{n^2-nk-4n+3k+3} \times \frac{1-\frac{k}{n-1}}{1-\frac{k}{n-2}}$$

$$
\begin{aligned}
&= \frac{n^2-1}{n^2} \times \frac{n^2-nk-4n+k+3}{n^2-nk-4n+3k+3} \times \frac{n^2-nk-3n+2k+2}{n^2-nk-3n+k+2} \\
&= \frac{n^2-1}{n^2} \times \frac{n^2-nk-4n+k+3}{n^2-nk-4n+3k+3} \times \left(1+\frac{k}{n^2-nk-3n+k+2}\right) \\
&= \frac{n^2-1}{n^2} \times \left(1+\frac{2k}{n^2-nk-4n+3k+3}\right) \times \left(1+\frac{k}{n^2-nk-3n+k+2}\right) \\
&= \frac{n^2-1}{n^2} \times \left(1-\frac{k}{(n-k-2)(n-3)}\right) < 1
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \frac{1-\rho^2_{1rcv}}{1-\rho^2_{2rcv}} &< 1 \\
\Rightarrow \rho^2_{1rcv} &> \rho^2_{2rcv} \qquad (10)
\end{aligned}
$$

that is, the restricted cross validity predictive power is found to be higher for VAR modeling at level when the variables are mixed in nature.

$$
\begin{aligned}
\text{Again,} \quad \rho^2_{rcv} &= 1-\frac{(n-1)(n-2)(n+1)}{n(n-k-1)(n-k-2)}(1-R^2) \\
&= 1-\frac{(n-1)(n-2)(n+1)}{n(n-k-1)(n-k-2)} \times \frac{SS(e)}{SS(t)},
\end{aligned}
$$

where $SS(e)$ and $SS(t)$ are sum of squared error and sum of squared total, respectively.

Let us assume that we are interested in forecasting for r periods. Now, we divide $SS(e)$ into two parts $SS(e_1)$ and $SS(e_2)$ so that $SS(e) = SS(e_1) + SS(e_2)$, where $SS(e_1)$ is the sum of squared residuals of first $n-r$ units and $SS(e_2)$ is the sum of squared residuals of last r units.

$$
\text{Hence,} \quad \rho^2_{rcv} = 1-\frac{(n-1)(n-2)(n+1)}{n(n-k-1)(n-k-2)} \times \frac{SS(e_1)+SS(e_2)}{SS(t)}
$$

$$
\begin{aligned}
\Rightarrow (1-\rho^2_{rcv}) &= \frac{(n-1)(n-2)(n+1)}{n(n-k-1)(n-k-2)} \times \frac{SS(e_1)+SS(e_2)}{SS(t)} \\
&= \frac{(n-1)(n-2)(n+1)}{n(n-k-1)(n-k-2)} \times \frac{r}{SS(t)}\left(\frac{SS(e_1)}{r}+\frac{SS(e_2)}{r}\right) \\
&= \frac{(n-1)(n-2)(n+1)}{n(n-k-1)(n-k-2)} \times \frac{r}{SS(t)}\left(\frac{SS(e_1)}{r}+MSPE\right)
\end{aligned}
$$

where, $MSPE = \frac{SS(e_2)}{r}$ is the Mean Squared Prediction Error of r periods based on pre-head computation. If we consider consistent trend, i.e., the equality of pre-head and post-head computation, we can use $MSPE$ as $MSFE$.

Then $MSPE = \frac{n(n-k-1)(n-k-2)}{r(n-1)(n-2)(n+1)} SS(t)(1-\rho^2_{rcv}) - \frac{SS(e_1)}{r}$

$$\Rightarrow RMSFE = \sqrt{\frac{n(n-k-1)(n-k-2)}{r(n-1)(n-2)(n+1)} SS(t)(1-\rho^2_{rcv}) - \frac{SS(e_1)}{r}} \quad (11)$$

In Eq.(11) we have computed the RMSFE using pre-head computation process, a process in which we built model using full sample and compute RMSFE using last r sample points. However, this pre-head procedure would not serve the forecasting performance of a fitted model, because, the forecast should be based on ahead process. Thus, we need to modify Eq.(11) so that it could serve better for forecast purpose. Replacing $SS(t) = SS_n(t)$ by $SS_{n+r}(t)$, $SS(e_1)$ by $SS(e)$ and n by $(n+r)$ we get Eq.(11) as a modified equation. Then the modified equation may be given by

$$RMSFE = \sqrt{\frac{(n+r)(n+r-k-1)(n+r-k-2)}{r(n+r-1)(n+r-2)(n+r+1)} SS_{n+r}(t)(1-\rho^2_{rcv}) - \frac{SS(e)}{r}} \quad (12)$$

where, $SS_{n+r}(t)$ is the sum of squares total of n observed values and r forecasted values. From Eq.(12), we can explain that the RMSFE will decrease with the increasing RCVPP. Also, we have found that the VAR model at the level for all the variables has more RCVPP (Eq.10). Thus, the VAR model at level has more RCVPP as well as less RMSFE when the variables are mixed in nature and the numbers of predictors are equal in both the cases.

3 Numerical Example

To provide a numerical solution of our theory we have considered the annual Gross Domestic Product (GDP) and Government Consumption Expenditure (GCONS) of Bangladesh from the year 1974 to 2000 (Bangladesh Bank, 1974-2000). We used the data from 1974 to 1997 for modeling and that from 1998 to 2000 for examining the performance of the fitted model. A unit root test is performed first. We have tried to include the effect of intercept term and trend but no significant effect is found. Also, the lag specification was done by SC and AIC criteria. Using the critical values of MacKinnon (1996), it is found that the variable GDP is stationary at the level and the other variable GCONS becomes stationary after first differencing (Table 1). Thus, two series under study are mixed in nature. The test of Granger causality leads to a bi-directional causality between the variables (results in Table 1). So, we may study the interrelationship among the variables using VEC and VAR models to conduct a comparative study. Cointegration test is carried out and a VEC model is constructed for unrestricted VAR with suitable lag length. The comparative results shown in Table 1 divulge that the VAR model at the level of both the variables have more RCVPP as well as less RMSFE than the VEC model.

Table 1: Comparative Study

Description of Test and Model Specification			GCONS	GDP
Unit Root Test	Specification		None	None
	Stationary at		First Difference	Level
	DF-statistic		-6.447186	3.519646
	1% MacKinnon critical value		-2.6756	-2.6700
Granger Causality	Lag		1	1
	Probability		2.0E-06	0.00909
	Direction		Bi-directional	Bi-directional
VAR	Order p=2 No intercept or Trend or other exogenous variables	No. of Parameters estimated	4	4
		No. of observation after adj.	22	22
		SC	19.26395	12.63794
		R^2	0.808714	0.992536
		RCVPP	0.691206	0.987951
		RMSFE	1845787488	29450041
VEC	Intercept in the cointegratedt equation but no trend. The test VAR with no intercept or trend. Only one CE at 5% level (Johansen, 1991, 1996) with lag length one. Order of VEC is p=2, no exogenous variables.	No. of Parameter. est.	4	4
		No. of obs. after adj.	21	21
		SC	19.83419	12.54988
		R^2	0.690580	0.744827
		RCVPP	0.486756	0.576737
		RMSFE	2.97E+10	245262355

4 Conclusion

The VAR model at level is applicable rather than VEC model when the variables are mixed in nature. The VAR in difference (non-stationary at level) will be a misspecification if the variables are cointegrated; VEC models are appropriate in that occasion.

References

Bangladesh Bank (1974-2000). *Economic Trend, Various issues.*

Blanchard, O. J. and Quah, D. (1989). The Dynamic Effects of Aggregate Demand and Supply Disturbances, *The American Economic Review*, Vol.79, No.4, pp 655-673.

Engel, R.F. and Granger, C. W. J. (1987). Cointegration and Error Correction: Representation, Estimation and Testing, *Econometrica*, Vol.55, pp 251-276.

Gujarati, DN (1995). *Basic Econometrics*, 3rd ed., Mc Graw-Hill International.

Harvey A. C. (1990). *The Econometric Analysis of Time Series*, 2nd ed., Philip Alan.

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, *Econometrica*, 1991, Vol.59, pp.1551-1580.

Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press Inc., New York.

Khan, M. A. R. and Ali, M. A. (2003). Restricted Cross Validity Predictive Power, *Pakistan Journal of Statistics*, Vol. 19, No.2.

MacKinnon, J. G. (1996). Numerical Distribution Function for Unit Root and Cointegration Tests, *Journal of Applied Econometrics*, Vol.11, pp.601-618.

Mc Cleary, R. and Hay, R. A. (1980). *Applied Time Series Analysis for the Social Science*, Sega Publications, California.

Pankratz, A. (1991). *Forecasting with Dynamic Regression Models*, John Wiley & Sons Inc., New York.

Sims C. A. (1980). Macroeconomics and Reality, *Econometrica*, Vol. 48, No.1.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences*, Third Edition. Lawrence Erlbaum Associates, Inc., Publishers, Mahwah, New Jersey.