MODELLING COLOMBO CONSUMER PRICE INDEX: A VECTOR AUTOREGRESSIVE APPROACH

M.C. Alibuhtto

Department of Mathematical Sciences, Faculty of Applied Sciences,
South Eastern University of Sri Lanka.
E-mail: mcabuhtto@seu.ac.lk

ABSTRACT

The consumer price index measures changes in the price level of consumer goods and services purchased by households, which reflect inflation. Various studies have been conducted to modelling and forecasting Consumer Price Index (CPI) by developed countries .However, such studies have not been reported in Sri Lanka. This paper is an attempt to modelling the Colombo Consumer Price Index (CCPI) by using monthly CCPI data from January 2003 to May 2011. For this purpose, Stepwise Regression, Principal Component Analysis and Vector Autoregressive (VAR) approach were used. The VAR model with the first principal component of selected CCPI components was identified the best fitted model for the CCPI series. The model was also tested to an independent data set using CCPI from February 2010 to May 2011.

Keywords: *CCPI, Inflation, Modelling, VAR*

Introduction

The Consumer Price Index (CPI) is an indicator to measure the average change in the prices paid by consumers for a specific basket of goods and services over time in a country. This “shopping basket” represents a different items consist of common consumer goods and services which are purchased by an average household. The weights for each item in the shopping basket are determined based by the amount spent on these items by households in a given country.

Some international standards for economic statistics have evolved primarily in order to enable internationally comparable statistics to be compiled. However, individual countries also stand to benefit from international standards. The CPI standards described in this manual draw upon the collective experience and expertise accumulated in many countries. All countries can benefit by having easy access to this experience and expertise. In many countries, the CPIs were first compiled mainly in order to adjust wages to compensate for the loss of purchasing power which is caused by inflation. (Sylvester Young, 2008).

The CCPI was introduced by the Department of Census and Statistics (DCS), in 2007, which is based on the Household Income and Expenditure Survey (HIES) conducted by the DCS in 2002.The HIES data represents more up to date consumer patterns for a much larger sample size, as well as an increased coverage area within Colombo for price collection.

The CCPI is based on prices of food and non-alcoholic beverages, cloth and footwear, housing, water, electricity, gas and other fuels, furnishing, household equipment and routine maintenance of the house, health, transport, communication, recreation and culture, education, miscellaneous and goods and services that people buy for their daily living.

The modelling and forecasting is usually carried out in order to provide an aid to decision making and planning the future. Forecasting CCPI are importantinputs for government, businesses sector, policy makers, investors, workers and various individuals for various applications.

The objective of this study is to modelling CCPI by applying a vector auto regressive (VAR) model.The VAR model was introduced by Sims

(1980) and this model is the most successful,flexible, and easy to use models for the analysis of multivariate time series. It isa natural extension of the univariate autoregressive model to dynamic multivariate time series. The VAR model has proven to be especially useful fordescribing the dynamic behavior of economic andfinancial time series andfor forecasting.

Economic literature on issue of CCPI modellingand forecasting is concerned with the positive relation of inflation level. Kenny et al. (1998) conducted a study to develop a multiple time series models for forecasting Irish Inflation. For this purpose, the Bayesian approaches to the estimation of VAR models were employed. A large selection of inflation indicators is assessed as potential candidates for inclusion in a VAR model. The results of this study confirm that the significant improvement in forecasting performance, which can be obtained by the use of Bayesian techniques.

Genberg and Chang (2007) conducted a study to develop a multivariate time series model to forecast output growth and inflation in the Hong Kong economy. For this purpose, the three types of VAR (unrestricted VAR, Bayesian VAR and conditional VAR) models were used. Based on their study, the results suggest that the Bayesian VAR framework incorporating external influences provide a useful tool to produce more accurate forecasts relative to the unrestricted VARs and univariate time series models, and conditional forecasts have the potential to further improve upon the Bayesian models.

Enders (2004), Fritzer et al. (2002), Lutkepol (2001) and many other authors suggest that for the calculation of forecasts of economic indicators VAR models should be applied because all variables in these models are endogenous, and,therefore, not a single variable may be removed when explanations for the behaviour of other variables areoffered. For the forecasting of economic indicators two types of VAR models may be applied: simple, or unrestricted, VAR models and models with certain restrictions on exogenous indicators present in them, or restricted VAR models.

Materials and Methods

Data: The secondary data on monthly CCPI data from January 2003 to May 2011 were considered for the analysis and it was collected from Department of Census and Statistics. The collected monthly CCPI data from January 2003 to January 2010 ('training set') is used for model fitting and data from February 2010 to May 2011 ('validation set') is used for validation of the model.

Vector Autoregressive Models (VAR)

VARis an econometric model has been used primarily in macroeconomics to capture the relationship and independencies between important economic variables. They do not rely heavily on economic theory except for selecting variables to be included in the VARs. The VAR can be considered as a means of conducting causality tests, or more specifically Granger causality tests.

VAR can be used to test the Causality as Granger-Causality requires that lagged values of variable 'X' are related to subsequent values in variable 'Y', keeping constant the lagged values of variable 'Y' and any other explanatory variables. In connection with Granger causality, VAR model provides a natural framework to test the Granger causality between each set of variables. VAR model estimates and describe the relationships and dynamics of a set of endogenous variables. For a set of 'n' time series variables $y_t = (y_{1t}, y_{2t}, \ldots\ldots\ldots, y_{nt})$, a VAR model with exogenous variables of order p (VARX (p)) can be written as:

$$y_t = A_0 + \sum_{i=1}^{p} A_i y_{t-i} + \sum_{j=1}^{m} B_j X_j + \varepsilon_t \longrightarrow (1)$$

Where ε_t - error term

Model Selection Criteria: The following statistical measures were used to find an appropriate model for CCPI.

$$AIC = \log\left(\frac{rss}{n}\right) + \left(\log(n) * \frac{k}{n}\right) \longrightarrow \quad (2)$$

$$BIC = \log\left(\frac{rss}{n}\right) + \left(2 * \frac{k}{n}\right) \longrightarrow \quad (3)$$

Where;k = number of coefficient estimated, rss = residual sum of

square, n = number of observations

Results and Discussions

The objective of this study is modelling CCPI series with VAR models.For this purpose two different approach were used.

1. VAR model with some Components of CCPI were selected by using stepwise regression method.
2. VAR model with common value for all CCPI componets by using principal component techniques.

VAR Model between CCPI versus Its Selected Components

The stepwise regression method was used to find suitable variables among ten components of CCPI series taking one lag behind. Both the probabillity of entry a varaible and to remove a variable was set as 0.05. The results of stepwise regression are tabulated in table 1.

Table 1: Output ofthe stepwiseregression analysis on CCPIversus FD (-1), CL (-1),…, MS (-1)Alpha-to-Enter: 0.05 Alpha-to-Removes: 0.05

Step	1	2	3	4	VIF
Constant β_0	23.064	17.406	-2.324	-7.647	-
FD(-1)β_1 t-Value p-Value	0.847 100.17 0.000	0.641 48.95 0.000	0.541 32.76 0.000	0.551 34.12 0.000	29.667
HO (-1) β_2 t-Value p-Value	-	0.214 16.60 0.000	0.194 18.88 0.000	0.186 18.37 0.000	11.978
MS (-1) β_3 t-Value p-Value	-	-	0.284 7.63 0.000	0.554 5.64 0.000	203.717
FU (-1)β_4 t-value p-value	-	-	-	-0.242 -2.95 0.004	186.991
S	3.44	1.65	1.26	1.21	
R-sq(adj)	99.18	99.81	99.89	99.90	
Mallows Cp	629.2	83.1	17.2	10.0	

Based on the above results, the selected regression equation can be written as:

$$CCPI_t = -7.647 + 0.551FD_{t-1} + 0.186HO_{t-1} - 0.242FU_{t-1} + 0.554MS_{t-1} \quad (A) \longrightarrow$$

Results in table1indicates that four indices can be selected from ten indices and thesefour indices are significant at 5% significance level. The selected indices are Food (FD), Housing (HO), Furnishing (FU) and Miscellaneous (MS). It should be noted that VIF of each varaible is very high confirming the excisatnce of highly significant multicollinearity among the four explantory varaibles and consequently the above model is not recommended to forecast. It is also found that the constant term in the model is significantly different from zero. The time series plot of the four selected indices and CCPI series are shown in figure 1. It indicates that the CCPI and selected four indices show upward trends and these all series are non-stationary.

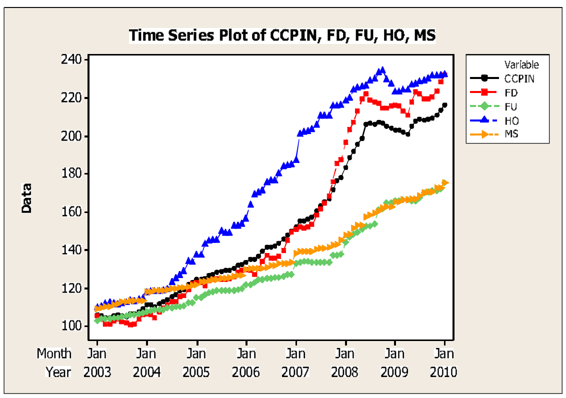

Figure 1: The time series plot of selected four indices and CCPIseries.

Identification of Lag Order of VAR Model Selection Criteria

Results in table 2 indicates that the minimum values of Schwarz Information Criteria (SIC) and Hannan-Quinn (HQ) statistic were obtained at lag 1. Therefore, it can be concluded that the optimal lag length of this model is one. Thus, Granger Causality test was carried out for CCPI, FD, HO, FU and MS and the results are shown in table 3.

Table 2: Values of SIC and HQ statistics at various lag orders

Lag	SIC	HQ
0	30.84	30.75
1	17.84*	17.30*
2	18.47	17.48
3	18.90	17.49
4	19.54	17.46
5	20.37	18.01

Table 3:
Results of Granger Causality test between CCPI and its components

Null Hypothesis	F-Statistics	P-Value
FD does not Granger Cause CCPI	3.90	0.041
CCPI does not Granger Cause FD	2.70	0.103
HO does not Granger Cause CCPI	14.08	0.000
CCPI does not Granger Cause HO	6.25	0.014
FU does not Granger Cause CCPI	12.40	0.000
CCPI does not Granger Cause FU	28.02	0.000
MS does not Granger Cause CCPI	4.31	0.032
CCPI does not Granger Cause MS	8.16	0.001

Table 3 indicates that thesevennull hypotheses are rejected at 5% significance level (p-value < 0.05). The F-Statistic values are significant and provide strong evidence for the argument that there is bi-directional linear granger causality between CCPI and selected indices of CCPI (HO, FU and MS) but FD and CCPI has only unidirectional granger causality relationship.

VAR Model with FD, HO, FU and MS

The parameter estimates of VAR model for CCPI versus FD, HO, FU and MS are shown in table 4.

Table 4:
Results of parameter estimation of identified VAR model

Variables	Coefficients	Stand. Error	t-value	P-value
CCPI (-1)	0.329	0.115	2.859	0.005
FD (-1)	0.391	0.065	6.062	0.000
HO (-1)	0.140	0.021	6.692	0.000
FU (-1)	-0.183	0.069	-2.649	0.008
MS (-1)	0.313	0.056	5.606	0.000

Table 4 indicates that the estimates of all parameters are significant at 5% significance level (p-value < 0.05). As the constant term was not significant in this model and model without constant term was considered. The final VAR model with components of CCPI can be wrriten as:

$$CCPI_t = 0.329CCPI_{t-1} + 0.391FD_{t-1} + 0.140HO_{t-1} - 0.183FU_{t-1} + 0.313MS_{t-1} \longrightarrow (B)$$

The suitable Principal Component (PC) for all components of CCPI

It is obvious that the correlations between all components of CCPI are interrelated and fairly large. Also, there is significant multicollinearity exists among the components of CCPI, the data set can be used to reach the possibility of dimensional reduction among indices. The results of communalities of CCPI components are tabulated in table5.

Table 5: Results of communalities of CCPI components

Variable	Initial	Extraction
Food	1.000	0.974
Clothing	1.000	0.979
Housing	1.000	0.885
Furnishing	1.000	0.995
Health	1.000	0.887
Transport	1.000	0.942
Communication	1.000	0.269
Recreation	1.000	0.859
Education	1.000	0.960
Miscellaneous	1.000	0.992

Table 5 indicates that the maximum of 99.5% the variance in furnishing and the minimum of 26.9% of the variance in communication index are accounted by the extracted factors. Hence, the communication index can be dropped to find the principal component because it has less communality value. The results of communalities of nine components of CCPI are tabulated in table 6.

Table 6: Results of communalities of nine components of CCPI.

Variable	Initial	Extraction
Food	1.000	0.968
Clothing	1.000	0.984
Housing	1.000	0.895
Furnishing	1.000	0.996
Health	1.000	0.874
Transport	1.000	0.940
Recreation	1.000	0.867
Education	1.000	0.974
Miscellaneous	1.000	0.998

Table 6 indicates that the maximum of 99.8% of the variance in miscellaneous and the minimum of 86.7% of the variance in recreation index are accounted by the extracted factors.

Table 7: Eigen values of the correlation matrix of CCPI components

consumer goods and services purchased by households, which reflect inflation. Various studies have been conducted to modelling and forecasting Consumer Price Index (CPI) by developed countries .However, such studies have not been reported in Sri Lanka. This paper is an attempt to modelling the Colombo Consumer Price Index (CCPI) by using monthly CCPI data from January 2003 to May 2011. For this purpose, Stepwise Regression, Principal Component Analysis and Vector Autoregressive (VAR) approach were used. The VAR model with the first principal component of selected CCPI components was identified the best fitted model for the CCPI series. The model was also tested to an independent data set using CCPI from February 2010 to May 2011.

Keywords: CCPI, Inflation, Modelling, VAR

Table 7 indicates that the first component accounts for 94.399 % of the variance. All remaining components are not significant. Hence, the first component has been chosen.

Table 8: Eigen scores of the first Principal Component (PC)

Variable	Eigen scores
Food(FD)	0.338
Clothing(CL)	0.340
Housing(HO)	0.325
Furnishing(FU)	0.342
Health(HL)	0.321
Transport(TR)	0.333
Recreation(RE)	0.319
Education(ED)	0.339
Misc(MS)	0.343

According to the table 8, the nine components can be reduced to single Principal Component (PC) and a new variable is denoted by PCand it can be written as:

$$PC = 0.338FD + 0.340CL + 0.325HO + 0.342FU + 0.321HL + 0.333TR + 0.319RE + 0.339ED + 0.343MS \longrightarrow (C)$$

VAR model between CCPI versus first PC

VAR Lag Order Selection Criteria

Table 9: Results of lag order selection

Lag	SIC	HQ
0	17.38	17.34
1	8.45	8.34
2	8.44*	8.27*
3	8.54	8.28
4	8.68	8.35
5	8.73	8.34

Results in table 9indicates that the minimum values of SIC and HQ statistic were obtained at lag 2. Therefore, it can be concluded that the optimal lag length of this model is two. Thus, Granger Causality test was carried out for CCPI and PC and the results are shown in table 10.

Table 10: Results of Granger Causality test between CCPI and PC

Null Hypothesis	F_Value	P_Value
CCPI (N) does not Granger Cause PC	15.268	0.000
PC does not Granger Cause CCPI (N)	9.141	0.000

Table 10 indicates that both the null hypotheses are rejected at 5% significance level (p-value < 0.05). The F-Statistic values are significant and provide strong evidence for the argument that there is a bi-directional linear Granger causality between CCPI and PC.

VAR Model with PC

The parameter estimates of VAR model for CCPI versus PC are shown in table 11.

Table 11: Results of parameter estimation

Variables	Coefficients	Stand. Error	t-value	P-value
CCPI (-1)	1.762	0.132	13.346	0.000
CCPI (-2)	-0.623	0.146	-4.279	0.000
PC (-1)	-0.258	0.063	-4.080	0.000
PC (-2)	0.201	0.059	3.426	0.001
C	4.377	1.712	2.556	0.012

Table 11 indicates that the estimates of all parameters are significant at 5% significance level (p-value < 0.05). The fitted VAR model can be wrriten as:

$$CCPI_t = 1.762CCPI_{t-1} - 0.623CCPI_{t-2} - 0.258PC_{t-1} + 0.201PC_{t-2} + 4.377 \longrightarrow (D)$$

Comparisons between Fitted two VAR Models

In the model estimation, the AIC and SIC values from each estimated models are computed. AIC and SIC values will be used in order to estimate which model is a better model for CCPI. For this purpose, the model with the lowest AIC and SIC values are concluded to be a better model. The results are reported in table 12.

Table 12: Comparison of the fitted two VAR models

Model	Log likelihood	AIC	SIC
B	-144.67	3.51	3.60
D	-136.44	3.41	3.55

The results indicate that the both AIC and SIC values from modelD is the lowest compared with modelB. Also, log likelihood value is high for modelD. Therefore, it shows that the modelD is the best model for forecasting monthly CCPI series.

3.5.4 Forecasting Performance of the Selected VAR Models

Table 13: Results of forecast performance statistics

Model	Data set	MAPE
B	Training set	0.61
	Validation set	0.81
D	Training set	0.59
	Validation set	0.62

Table 13 indicates that the mean of percentage error (MAPE) for validation set of the modelDis lower than the modelB. Therefore, it can be concluded that the modelDis a better forecast model for monthly CCPI series.

Conclusions and Recommendations

This study aimed to modelling CCPI series using VAR approach. Two types of VAR models were estimated and model was also tested to an independent data set using CCPI from February 2010 to May 2011. The comparative performance of these VAR models have checked and verified by using the model selection procedure (AIC and SIC). The comparison indicates that the VAR model with the first principal component of selected CCPI components was identified the best fitted model to forecast the CCPI series. The error series of the fitted model was found to be a white noise process.

References

Enders, W., (2004), Applied Econometric Time Series, New York: John Wiley and Sons.

Fritzer, F., Moser, G., and Scharler, J., (2002), Forecasting Austrian HICP and Its Components Using VAR and ARIMA Models, Oesterreichische National bank Working Paper No. 73, pp. 1-45.

Genberg, H., and Chang, J., (2007), A VAR Framework for Forecasting Hong Kong's Output and Inflation, Research Department, Hong Kong Monetary Authority, Working Paper 02.

Kenny, G., Meyler, A., and Quinn, T., (1998), Bayesian VAR Models for Forecasting Irish Inflation, Central Bank of Ireland, Technical Paper, 4/RT/98.

Lutkepol, H. (2001). Vector Autoregressive and VectorError Correction Models. Institute of Statistics and Econometrics.

Sylvester Young, D., (2008), Global Wage Report 2008/09, International Labour Office (ILO), Geneva, ISBN 978-92-2-121499-1

Sims, (1980), Macroeconomics and Reality, Econometrica, 48, pp. 1-48.