Forecast Accuracy, Coefficient Bias and Bayesian Vector Autoregressions

Ronald Bewley
The University of New South Wales

A Bayesian Vector Autoregression (BVAR) can be thought of either as a method of alleviating the burden of the over-parameterisation usually associated with unrestricted VARs, or as a method of correcting coefficient bias when the time series are nonstationary. Monte Carlo evidence is provided to show that the latter appears to be a more important characteristic of BVARs in experiments using a four-equation cointegrated system, and with that system embedded in a ten-equation model containing six extraneous random walks. It is found that the BVAR model generally performs much better than a VAR in levels and is a viable alternative to a vector error correction model. It is also found that estimating constant terms when there is no drift in the data causes a major deterioration in forecasting performance.

1. INTRODUCTION

For more than two decades, vector autoregressive models (VARs) have been used for macroeconomic modelling and forecasting. The original Litterman (1986) model contained six variables, and six lags of each variable in each equation, making the demands on data extreme. This led Litterman to suggest that more accurate forecasting models might be produced by combining the evidence contained in the data with the Bayesian priors specifying each time series is a random walk. The resultant Bayesian VAR (BVAR), based on Theil and Goldberger's (1961) mixed estimation procedure, has become a standard and successful method of forecasting macroeconomic time series.

Although the 20 years that have elapsed since the pioneering days of VAR modelling produced significantly longer time series, the need for methods such as BVARs to reduce the demands of modelling on data have not diminished. Researchers have begun to develop larger models (Dungey and Pagan, 1999) and structural changes often prevent older data being used.

Recently, Abadir, Hadri and Tzvalis (1999) demonstrated that the bias in parameter estimates of VARs with I(1) data increases with the inclusion of extraneous I(1) variables. As a result, two major questions are posed for researchers. Should one block-segment models to reduce this parameter bias? Should the estimated number of unit roots be imposed upon such models before estimation?

Given that equations from classical econometric models often contain only three or four variables in an equation, it is reasonable to ask whether it is better to build VARs with, say, 10 equations, or a number of inter-linked VARs, each of only three or four variables. Since any additional variables, over and above those needed to capture the macroeconomic transmission mechanism increase the bias in the estimated coefficients, this is a significant, and as yet, largely unexplored question (Metin, 1995).

Clements and Hendry (1995) conducted a Monte Carlo study and reported that if the true number of unit roots is unknown, it is preferable to err on the side of too many cointegrating equations rather than too few, whilst Brandner and Kunst concluded to the contrary.

The second purpose of this paper is to illuminate this Clements-Hendry and Kunst-Brandner contradiction. The forecasting performance of Vector Error Correction (VEC) models based on an estimated number of unit roots is compared to the extremes of a VAR in levels and a VAR in differences (DVAR) and all three are compared to a BVAR specification.

Forecasts from VAR models with biased estimates produce forecasts that are unconditionally unbiased (Dufour, 1984) but this bias significantly contributes to forecast mean squared error (mse) and bias in a conditional sense. By directly comparing the forecast performance of BVARs with VECs in a Monte Carlo experiment, some light can be shed on whether BVARs are as successful as they are because of bias-correction or parameter reduction.

2. MONTE CARLO EXPERIMENT

The core of the data generating process (dgp) for this experiment is a four-equation cointegrated system with one common trend (when $|\phi| < 1$)

$$y_{it} = y_{4t} + \varepsilon_{it}, \qquad i = 1,2,3$$

$$\Delta y_{4t} = (1 - \theta L)\xi_{4t}$$

$$(1 - \phi L)\varepsilon_{it} = (1 - \theta L)\xi_{it} \qquad i = 1,2,3$$

where each ξ_t is a normally and independently distributed, mutually uncorrelated, zero mean, unit variance disturbance term.

In a second dgp, the core model is augmented by six extraneous and mutually uncorrelated random walks. From Abadir et al., it is known that the coefficient bias in unrestricted VARs increases when such extraneous random walks are included so that a comparison of the results from these two models goes part way to addressing the block-segmentation question.

In order to focus on the small sample properties of the estimators, the number of observations (T) is set to 75, although the results of additional experiments not reported here suggests that the same conclusions are reached with 150 observations, albeit in a more muted sense.

Two values of ϕ are chosen to reflect no cointegration ($\phi = 1$), and extreme cointegration ($\phi = 0$). Similarly, two values of θ are used. When $\theta = 0$, the true lag is one and, when $\theta = 0.75$, the true lag length is infinite, but can reasonably be approximated by a VAR(2) or VAR(3) with T = 75. In each of the eight experiments, 5,000 replications were conducted.

Forecast accuracy is measured by the trace of the forecast mse matrix for the four-equation model, or that four-equation partition for the ten-equation system. Thus, any deterioration in mse in the presence of extraneous random walks is directly attributable to their inclusion and not as a result of the model's inability to predict the additional variables.

Four basic variants of the model are estimated:
(a) the VEC model using both the Chao and Phillips (1999) joint procedure, VEC(J), and the more standard sequential procedure based on using the BIC to select the lag length, and Johansen and Juselius's (1990) maximum root test to select the number of common trends using a sequence of 5% tests, VEC(S).
(b) a VAR using BIC to select the lag length.
(c) a DVAR using BIC to select the lag length.
(d) Two BVARs, each using AIC to select the lag length, but with either the standard Minnesota priors (with hyperparameters of 0.2 for the diagonal and 0.2 for the off-diagonal terms), BVAR(M), or with looser priors on the off-diagonal (off-diagonal hyperparameter = 0.8), BVAR(L).

Each estimated model contained a constant term although none was present in the dgp. All of the models are compared to a no-change model which happens to be the correct model when $\phi = 1$ and $\theta = 0$. The mse's were computed for one-step ahead and ten-steps ahead.

3. FORECAST ACCURACY

One result, which is not reported, was particularly clear. Initially, each of the models was also estimated with an arbitrary lag length of four (Muscatelli and Hurn, 1992). There was extremely strong support for using an information criterion to select the lag length over arbitrarily setting it to four in both VAR and DVAR variants. In some cases the mse more than halved when the lag length was estimated. Since the BVAR method is designed to be less susceptible to inefficiency due to over-specifying the lag length, AIC was preferred to BIC for choosing the lag length for that variant. There was also no support for arbitrarily setting the lag length to four in this case. Thus, no further account will be taken of the popular, but arbitrary, procedure of not estimating, but imposing, the lag length.

The results of the mse forecast comparison are presented in Table 1. There is no clear ranking between the joint and sequential procedures for determining the number of common trends. It is known from Bewley and Yang (1995, 1998) that the case of mutually uncorrelated residuals (ε_{it}) used here is the worst case for the power of the Johansen test and it is dominated by the Bewley-Yang test in that part of the parameter space. Furthermore, the choice of $\phi = 0$ produces a particularly strong degree of cointegration and intermediate values of ϕ would make cointegration harder to detect. Thus, further work on choosing between the joint and sequential procedures is warranted.

Table 1: MSE Forecast Comparison

Model	4-equation model				10-equation model			
	$\phi = 1$		$\phi = 0$		$\phi = 1$		$\phi = 0$	
lead time	1	10	1	10	1	10	1	10
$\theta = 0.00$								
VEC(J)	10.0	115.1	10.7	54.3	10.1	113.5	16.2	59.7
VEC(S)	10.0	116.0	10.6	52.8	12.0	113.5	12.6	55.9
BVAR(M)	10.3	119.8	14.2	58.9	10.3	118.3	13.9	57.6
BVAR(L)	10.6	132.0	11.8	54.2	11.4	157.1	13.4	68.4
VAR	11.2	143.7	10.7	54.5	12.8	165.3	12.9	70.4
DVAR	10.0	115.1	16.2	60.3	10.1	113.5	16.2	59.7
No-Change	9.9	100.7	16.0	52.3	10.0	99.1	16.0	52.6
$\theta = 0.75$								
VEC(J)	13.5	339.6	14.9	151.4	15.5	327.9	16.2	155.1
VEC(S)	13.4	345.0	14.2	149.1	18.3	420.3	17.1	161.4
BVAR(M)	13.8	365.0	14.4	153.2	13.5	361.0	14.6	157.2
BVAR(L)	14.1	434.3	12.8	154.1	14.8	499.3	14.6	202.5
VAR	14.8	475.4	14.1	160.3	19.9	557.8	17.3	205.9
DVAR	13.2	340.4	15.9	155.2	15.5	327.8	16.2	155.1
No-Change	15.6	294.3	15.9	135.5	15.2	290.9	15.9	136.8

The BVAR methods do reasonably well with the stronger priors performing better in the larger model, whether or not there is cointegration in the dgp. This result may well be due to the increased bias of the unit roots in the larger model so that it is more important to approximate the unit roots than cointegrating relationships. However, in the smaller model with $\theta = 0$, the looser prior worked better when there is no cointegration, but the reverse is true when $\phi = 1$.

The VAR performs poorly, particularly at longer lead times, and in the larger model. Indeed, it performs worst in 11 of the 16 cases. The only times that the VAR does relatively well is with the one-step ahead forecasts when there is cointegration.

The DVAR, on the other hand, works reasonably well when there is no cointegration although it does perform much worse than the no-change model with the longer lead time. This behaviour is undoubtedly due to having to estimate a constant term in the DVAR which makes the angle of the forecast trajectory a significant issue.

A direct comparison of the BVAR and VEC models reveals that there is no systematic ordering between them, particularly if the better from each pair are compared. On the other hand, the VEC is much preferred to the VAR and on a par with, or better than, the DVAR.

When only longer-run forecasts are considered, some clearer rankings emerge. In particular, all models do relatively poorly compared to the no-change model, even when cointegration is present. Of course, each model could have been estimated with the constant term suppressed and these results suggest that option being worthy of further investigation either as a set of exact restrictions or in a Bayesian framework. Of course, when there is drift in the model, the no-change model would perform poorly and results not presented show that the same-change model, that is appropriate in the presence of drift, performs particularly poorly in the reported no-drift case. Indeed, the ten-step-ahead mse for the same-change model was of the order of ten times greater than for the no-change model. Since many models would contain variables both with and without drift, the naive models are not viable alternatives in empirical work.

By directly comparing the mse's for the four- and ten-equation models, it can be noted that there is some deterioration in the one-step ahead forecasting performance but, at ten steps ahead, the mse is much worse for the VAR and BVAR(L).

It follows from this that block-segmentation is not a major issue in building VEC models but could well be crucial in the construction of large unrestricted VARs. However, the superiority of the no-change model over the VEC model at long lead times, and the deterioration in mse in 1-step forecasts with model size, is indicative of possible gains that might be had, either from incorporating some Bayesian procedure in combination with the VEC model, or from block-segmentation for second-order gains in mse.

4. COEFFICIENT BIAS

These experiments also shed some light on the degree of coefficient bias present in the eight cases considered.

For simplicity and clarity, only the cases of the unrestricted VAR and the BVAR(M) will be considered with no moving average terms ($\theta = 0$) in the dgp. In this case, the dgp can also be written as

$$Y = Y_{-1}A + U$$

using obvious notation where Y is T x n and n is the number of equations. When there is no cointegration $A = I_n$, and when cointegration is present ($\phi = 0$)

$$A = \begin{bmatrix} 0 & 0 \\ \iota' & 1 \end{bmatrix}$$

for the four-equation model, where ι is a 3-element column vector of unit elements. That is, each variable depends only upon the fourth variable lagged one period. For the ten-equation model

$$A = \begin{bmatrix} 0 & 0 & 0 \\ \iota' & 1 & 0 \\ 0 & 0 & I \end{bmatrix}$$

The parameter estimates for the leading 4 x 4 estimates of A from VARs and BVAR(M)s with one lag are presented in Table 2.

In the left panels of Table 2, the underlying parameter matrix is an identity. Since the average own-coefficient in the four-equation OLS-estimated VAR model is 0.8621, the average bias is 0.1379 and this bias increases to 0.2650 with ten equations.

While the forecasts from these models are unbiased, and the Monte Carlo evidence supports this, the 'conditional bias' is of the order of 25% of the forecast pivot one-step ahead in the larger model but only half of this in the smaller model. Clearly, conditional bias increases rapidly with lead time. Therefore, there is a large cost in conditional bias and mse due to including irrelevant variables.

The BVAR priors do much to correct this OLS bias. The average OLS biases of 0.1379 and 0.2650 reduce to 0.0764 and 0.1109 in the BVAR smaller and larger models, respectively, so that the size of the model is of less of a consequence when BVARs are used.

When the VAR is estimated by OLS, the bias of the nonzero elements is much less that in the non-cointegrated case. What is surprising is the impact of the BVAR on the estimated A matrix. The priors pull the OLS estimates towards an identity matrix (the prior) but with a noticeably different impact on each equation.

While the estimates on the diagonal and the fourth row must be jointly considered, it is not entirely clear why the BVAR does so well compared to the VEC models in reducing the forecast mse over that from the VAR. Each series is nonstationary and it appears that there is little to gain by estimating only one common trend once each unit root has been bias-corrected.

Table 2: Estimated Coefficients for the VAR(1)

	$\phi = 1.0$				$\phi = 0.0$			
4 equations (m=0)								
	0.8571	-0.0029	-0.0011	-0.0003	-0.0246	0.0010	0.0010	0.0031
	0.0064	0.8654	0.0010	0.0011	-0.0026	0.0281	0.0007	-0.0023
OLS	-0.0016	0.0001	0.8626	-0.0013	-0.0006	0.0006	0.0281	-0.0014
	0.0028	0.0031	0.0028	0.8634	0.9591	0.9587	0.9581	0.9314
	0.9226	-0.0023	-0.0011	-0.0003	0.6263	-0.0086	-0.0033	-0.0002
	-0.0020	0.9259	-0.0014	-0.0004	-0.0029	0.6962	-0.0047	-0.0003
BVAR	-0.0018	-0.0022	0.9244	-0.0012	0.0227	0.0109	0.7915	-0.0006
	-0.0016	-0.0017	-0.0024	0.9215	0.1449	0.1180	0.0742	0.9363
10 equations (m=6)								
	0.7291	-0.0041	-0.0021	0.0002	-0.0947	0.0010	0.0014	0.0007
	0.0050	0.7385	-0.0016	-0.0009	-0.0056	-0.1008	-0.0048	-0.0035
OLS	0.0001	-0.0018	0.7363	-0.0013	0.0059	0.0064	-0.0935	0.0043
	-0.0055	-0.0051	-0.0032	0.7360	0.8886	0.8872	0.8914	0.7947
	0.8881	-0.0033	-0.0018	-0.0006	0.5714	-0.0104	-0.0045	-0.0005
	-0.0035	0.8913	-0.0024	-0.0008	-0.0072	0.6354	-0.0064	-0.0007
BVAR	-0.0032	-0.0039	0.8902	-0.0016	0.0138	0.0047	0.7296	-0.0011
	-0.0030	-0.0038	-0.0042	0.8868	0.1117	0.0923	0.0596	0.8955

5. CONCLUSIONS

It is common for researchers to estimate VAR forecasting models with up to ten equations using nonstationary data and relatively short samples of say 75-100 observations. A set of experiments has been devised in an attempt to mimic some of the problems encountered in this literature by embedding a four-equation VAR with three cointegrating equations within a ten-equation VAR that contains six extraneous random walks.

For example, the four cointegrated variables could be two domestic and two foreign interest rates that are mutually cointegrated and the remaining six variables could be other financial and real variables that happen to have little or no impact on the interest rate determination process.

Certain reasonably strong conclusions flow from the Monte Carlo experiments. There is no evidence to favour the sometimes-practiced approach of arbitrarily setting the lag length to, say, four for quarterly data and there is no evidence in favour of estimating an unrestricted VAR. In that sense, imposing too many unit roots is to be preferred to imposing too few. However, estimating the lag length and the number of common trends either jointly or sequentially is beneficial but the resultant VEC models do not typically dominate the BVAR models in these experiments with the Johansen procedure.

It is also shown that the coefficient estimates of VAR models are very biased in samples of this size, even though the forecasts themselves are unbiased. This establishes the need to consider the concept of a conditional, rather than an unconditional, forecast bias in a more general assessment of forecasting performance.

The experiments also suggest further work might be warranted. The constant terms in models with no underlying drift place a major role in producing poor long term forecasts suggesting that exact restrictions or strong priors on constants might prove to be beneficial.

Secondly, the joint success of the VEC and BVAR models is suggestive of some hybrid model outperforming both.

Finally, more experimentation is necessary to establish a ranking between the joint and sequential testing procedures for lag length and the number of common trends. It is known from Bewley and Yang (1995, 1998) that the power of the latter test critically depends upon a parameter not varied in this experiment and that other tests are more powerful than Johansen's in certain circumstances.

6. ACKNOWLEDGMENT

I gratefully acknowledge funding from the *Australian Research Council* (A79700597) and thank Stuart Nolan and Tommy Ng for research assistance.

7. REFERENCES

Abadir, K.M., K. Hadri and E. Tzavalis, The influence of VAR dimensions on estimator biases, *Econometrica*, 67(1), 63-181, 1999.

Bewley, R., and M. Yang, Tests for cointegration based on canonical correlation analysis, *Journal of the American Statistical Association*, 90(431), 990-996, 1995.

Bewley, R., and M. Yang, On the size and power of system tests for cointegration, *Review of Economics and Statistics*, 80(4), 675-679, 1998.

Brandner, P., and R.M. Kunst, Forecasting with vector autoregressions - the influence of cointegration: A Monte Carlo study, Research Memorandum No. 265, 1990. Institute for Advanced Studies, Vienna.

Chao, J.C., and P.C.B. Phillips, Model selection in partially nonstationary vector autoregressive processes with reduced rank structure, *Journal of Econometrics*, 91(2), 227-271, 1999.

Clements, M.P., and D.F. Hendry, Forecasting in cointegrated systems, *Journal of Applied Econometrics* 10(2), 127-146, 1995.

Dufour, J.-M., Unbiasedness of predictions from estimated vector autoregressions, *Econometric Theory* 1, 387-402, 1985.

Dungey, M., and A. Pagan, Towards a structural VAR model of the Australian economy, mimeo, LaTrobe University, 1999

Johansen, S., and K. Juselius (1990), Maximum likelihood estimation and inference on cointegration - with applications to the demand for money, *Oxford Bulletin of Economics and Statistics*, 52(2), 169-210, 1990.

Litterman, R.B., Forecasting with Bayesian vector autoregressions - five years of experience, *Journal of Business and Economic Statistics* 4, 25-37, 1986.

Metin, K., An integrated analysis of Turkish inflation, *Oxford Bulletin of Economics and Statistics*, 57(4), 513-531, 1995.

Muscatelli, V.A., and S. Hurn, Cointegration and dynamic time series models, *Journal of Economic Surveys*, 6(1), 1-43, 1992.

Theil, H., and A.S. Goldberger, On pure and mixed statistical estimation in economics, *International Economic Review* 2(1), 65-78, 1961.