Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 72(2) 2018 113–119

An Empirical Evaluation of Parameter Shrinkage Techniques for Vector Autoregression Models

B.S. Yashavanth[1], K.N. Singh[2], A.K. Paul[2], Amrender Kumar[3], D.R. Singh[3] and Pal Singh[2]
[1]*ICAR-National Academy of Agricultural Research Management, Hyderabad*
[2]*ICAR-Indian Agricultural Statistics Research Institute, New Delhi*
[3]*ICAR-Indian Agricultural Research Institute, New Delhi*

Received 10 January 2017; Revised 13 July 2017; Accepted 13 April 2018

SUMMARY

The vector autoregressive (VAR) models are being widely used for modeling and forecasting multiple time series. But the problem of over parameterization has restricted their use only for small number of time series. The Constrained VAR, Bayesian VAR and Least Absolute Shrinkage and Selection Operator (LASSO) applied for VAR to shrink the parameter estimates have gain importance in this regard. The present study is directed at empirically comparing the performance of Constrained VAR, Bayesian VAR and LASSO VAR using the data on annual estimated landings of six different marine fish species in India. The results from the forecast evaluation techniques indicated that all three techniques are equally good in forecasting the estimated fish landings whereas LASSO VAR outperformed Bayesian VAR in shrinking the parameters of the model to zero.

Keywords: Bayesian VAR, LASSO, Over parameterization, Shrinkage.

1. INTRODUCTION

Time series analysis techniques are being used since decades for modelling and forecasting of production, yield, prices of crop produce and forewarning of incidence of pests and diseases. In recent times, univariate (generally Autoregressive Integrated Moving Average (ARIMA)) models have been used to model and forecast several agricultural commodity variables (Padhan, 2012; Paul *et al*., 2013; Prabhakaran *et al*. 2013; Kim *et al*. 2013). When more than one time series variables have to be modeled together simultaneously utilizing the interrelations between time series variables, vector autoregressive (VAR) model is generally preferred because of its simple model building procedure (Gujarathi *et al.*, 2009). Use of VAR models for multivariate analysis is prominent in literature. VAR models have been used extensively for modelling agricultural and economic indicators (Primicer, 2005; Gutierrez *et al*., 2014) globally. One of the major setback that researchers often encounter while using the VAR models is the problem of over parameterization i.e., as the number of study variables increases along with the number of lags used in the model, the number of parameters that are to be estimated increases geometrically. As it is known that, if there are $k$time series to be modelled with p lags, then the number of parameters to be estimated in the model will be k^2p+k, which may go very large.

One solution to address this problem is to go for Constrained VAR (CVAR) (Lutkepohl, 1993) where the non-significant parameters obtained in VAR model are forced to take the value zero. Another solution is the Bayesian VAR model denoted by BVAR (Koop and Korobollis, 2009; Banbura *et al*., 2010) where the model parameters are treated as random variables and prior probabilities are assigned to them. The general idea is to use informative priors to shrink the unrestricted model towards a parsimonious benchmark. Many prior distributions are available in the literature and the performance of a VAR model depends on the choice of prior distribution (Sevinc and Ergun, 2009) and hyper parameters (Carrierro *et al*., 2009) making BVAR a subjective procedure.

Corresponding author: B.S. Yashavanth
E-mail address: yashavanthbs@gmail.com

Researchers have used Least Absolute Selection and Shrinkage Algorithm (Tibshirani, 1996), denoted by LASSO, where the parameters are penalized with an L_1 constraint along with its several variants to address the over parameterized VAR models (Hsu *et al*., 2008; Gefand, 2014; Nicholson, 2014). The scope of this article is directed to study BVAR and LASSO VAR (LVAR) models and empirically evaluate their performance with VAR and CVAR using a real data on annual landings of six different fish species in India.

2. MATERIALS AND METHODS

2.1 Vector Autoregression

In a univariate autoregression of order p, the study variable is regressed upon on p lags of itself, whereas in a k variable VAR of order p, k different equations are built. In each equation, a variable is regressed on p lags of itself, and p lags of every other variable. The key point is that, in contrast to the univariate case, vector autoregressions allow for cross-variable dynamics. Each variable is related not only to its own past, but also to the past of all the other variables under consideration.

Suppose there are k time series components $\{Y_{1t}\},\{Y_{2t}\},\ldots,\{Y_{kt}\}$ for t=1,2,3,…,n at equally spaced time intervals. We can represent these components by a vector $\mathbf{Y}_t = (\mathbf{Y}_{1t}, \mathbf{Y}_{2t}, \ldots, \mathbf{Y}_{kt})^T$ which is called as a vector of time series which can be modeled by a vector autoregressive model. The time series $\mathbf{Y}_t$ follows a Vector Autoregressive model of order p, VAR (p), if it satisfies

$$\mathbf{Y}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}_1 \mathbf{Y}_{t-1} + \ldots + \boldsymbol{\Phi}_p \mathbf{Y}_{t-p} + \mathbf{e}_t, \quad p>0, \qquad (1)$$

where $\boldsymbol{\mu}$ is a k-dimensional parameter vector, $\boldsymbol{\Phi}_j$ for j=1,…,p are $k \times k$ parameter matrices and $\mathbf{e}_t$ is a sequence of serially uncorrelated random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}$. Using the back-shift operator B, the VAR (p) model can be written as

$$\left(\mathbf{I} - \boldsymbol{\Phi}_1 B - \ldots - \boldsymbol{\Phi}_p B^p\right)\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{e}_t, \qquad (2)$$

where I is the $k \times k$ identity matrix. This representation can be written in a compact form as

$$\boldsymbol{\Phi}(B)\mathbf{Y}_t = \boldsymbol{\mu} + \mathbf{e}_t,$$

where $\boldsymbol{\Phi}(B) = \mathbf{I} - \boldsymbol{\Phi}_1 B - \ldots - \boldsymbol{\Phi}_p B^p$ is a matrix polynomial. A VAR (p) model assumes that Cov $(\mathbf{Y}_t, \mathbf{e}_t) = \boldsymbol{\Sigma}$, the covariance matrix of $\mathbf{e}_t$; and Cov $(\mathbf{Y}_{t-l}, \mathbf{e}_t) = \mathbf{0}$ for $l > 0$;

2.2 Bayesian Vector Autoregression

A simple BVAR model is similar to a simple unrestricted VAR model. In both the cases, each variable is expressed as a function of the past values of all the other variables of interest. But the difference is that in BVAR, the parameters of the model are treated as random variables. For the VAR(p) model in equation 1, a prior distribution, $\pi(\theta)$ which is not conditioned upon any realized observation is given to all the parameters, $\boldsymbol{\theta} = (\boldsymbol{\Phi}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ that are to be estimated. This is used to form the posterior distribution $p(\theta|Y_T)$ which is the distribution of the parameters conditional on the observed data. Another important element in BVAR is the distribution of the observed data conditional on parameters, i.e., the likelihood function given as:

$$L(\mathbf{Y}_T | \boldsymbol{\theta}) = \prod_{t=1}^{T} f\left(y_t | \mathbf{Y}_{t-1}, \boldsymbol{\theta}\right) \qquad (3)$$

According to Baye's rule, the posterior distribution is proportional to the product of the likelihood function and the prior distribution as below

$$p(\boldsymbol{\theta} | \mathbf{Y}_T) = \frac{L(\mathbf{Y}_T | \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int L(\mathbf{Y}_T | \boldsymbol{\theta})\pi(\boldsymbol{\theta}) d\boldsymbol{\theta}} \propto L(\mathbf{Y}_T | \boldsymbol{\theta})\pi(\boldsymbol{\theta}) \qquad (4)$$

where $\int L(\mathbf{Y}_T | \boldsymbol{\theta})\pi(\boldsymbol{\theta}) d\boldsymbol{\theta} = p(\mathbf{Y}_T)$ is the marginal likelihood.

The point estimates and their measures of precision can be found on integrating the marginal distributions of the parameters. The posterior distribution depends upon the prior beliefs as well as the data which makes selection of prior distribution very crucial for BVAR modeling. Many prior distributions are available in the literature: the Minnesota prior, the diffuse and Normal-Wishart prior, the Normal-Diffuse prior, the extended natural conjugate prior to name few. However, in this study, the widely considered Minnesota prior is used.

The Minnesota priors are proposed by Litterman (1986). These priors restrict the lag structure of the over parameterized VAR model by imposing tightness on the parameters that are to be estimated. The first mean-lag $(diag(\boldsymbol{\theta}_{\pi_1}))$ is chosen to be 0.9 for variables modelled in levels, and 0 for variables modelled in first difference. The rest of the lag-coefficients in $\boldsymbol{\theta\pi}$ are presumed to be 0. The shrinkage hyperparameters are used to specify the variance-covariance matrix, Ω_Π, of the coefficients. These form the tightness structure which is specified as below:

$$diag(\Omega_\Pi) = \begin{cases} (\frac{\lambda_1}{m^{\lambda_3}})^2 & \text{for own lags} \\ (\frac{\lambda_1 \lambda_2}{p^{\lambda_3}})^2 \frac{\sigma_q^2}{\sigma_r^2} & \text{for lags of variable q in equation r} \end{cases} \quad (5)$$

where p is the number of lags and λ_i are the hyper parameters. σ_q^2 is the variance of residuals from a univariate AR(p) parameter estimations for variable q. The differences in scales and units of measurement are controlled by σ_q^2 / σ_r^2. The hyperparameters impose tightness according to the table 1.

Table 1. Range of values for the hyper parameters of Bayesian VAR

Hyperparameter	Description	Allowed range
λ_1	Overall shrinkage	$\lambda_1 > 0$
λ_2	Cross-variable shrinkage	$0 < \lambda_2 \le 1$
λ_3	Lag decay	$\lambda_3 > 0$

2.3 LASSO - Vector Autoregression

LASSO, a least square method where the parameters are penalized with an L_1 constraint is originally proposed for linear regression set up. This has been extended to the VAR model in equation (1), by applying L_1 penalty to the convex least squares objective function

$$\frac{1}{2}\|\mathbf{Y} - \boldsymbol{\mu}\mathbf{1}' - \boldsymbol{\Phi}\mathbf{Z}\|_F^2 + \lambda\|\boldsymbol{\Phi}\|_1 \quad (6)$$

In which $\|\mathbf{X}\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}|x_{ij}|^2$ is the square of the Frobenius norm of X, $\|\mathbf{X}\|_1 = \sum_{jk}|X_{jk}|$ is the L_1 norm and $\lambda \geq 0$ is a penalty parameter. An L_1 penalty will induce sparsity in the coefficient matrix **Φ** by zeroing individual entries. The rolling cross-validation technique proposed by Nicholson *et al.* (2014) can be used for estimating the optimum value for the tuning parameter λ. This is done by dividing the data into three periods of length T/3; one each for initialization, training and forecast evaluation. The procedure is illustrated in the Fig. 1.

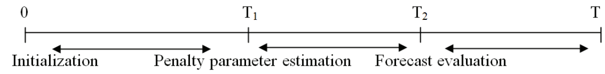


Fig. 1. Illustration of rolling cross-validation

The time indices are given by T_1 and T_2. The period T_1+1 through T_2 is used for training and T_2+1 through T is used for evaluation of forecast accuracy in a rolling manner. Then the one-step ahead mean-square forecast error (MSFE) is minimized. If $\hat{y}_{t+1}^{\lambda}$ is the one-step ahead forecast based on all observations from 1,…,t, then the MSFE to be minimized is given as

$$MSFE(\lambda) = \frac{1}{(T_2 - T_1 - 1)}\sum_{t=T_1}^{T_2-1}\|\hat{y}_{t+1}^{\lambda} - y_{t+1}\|_F^2 \quad (7)$$

2.4 Forecast evaluation technique

The forecasting ability of different models is assessed with respect to two common performance measures, viz. the root mean squared error (RMSE) and the mean absolute percentage error (MAPE) of each series in the VAR model.

The RMSE measures the overall performance of a model and for a series it is given by

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}$$

where, y_t is the actual value for time t, $\hat{y}_t$ is the predicted value for time t, and n is the number of predictions. The second criterion, the MAPE is a measure of average error for each point forecast which is given as below for a series

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{y}_t - y_t}{y_t}\right| \times 100$$

where the symbols have the same meaning as above. The model with least RMSE and MAPE values is considered as the best model suitable for the data.

2.5 Data

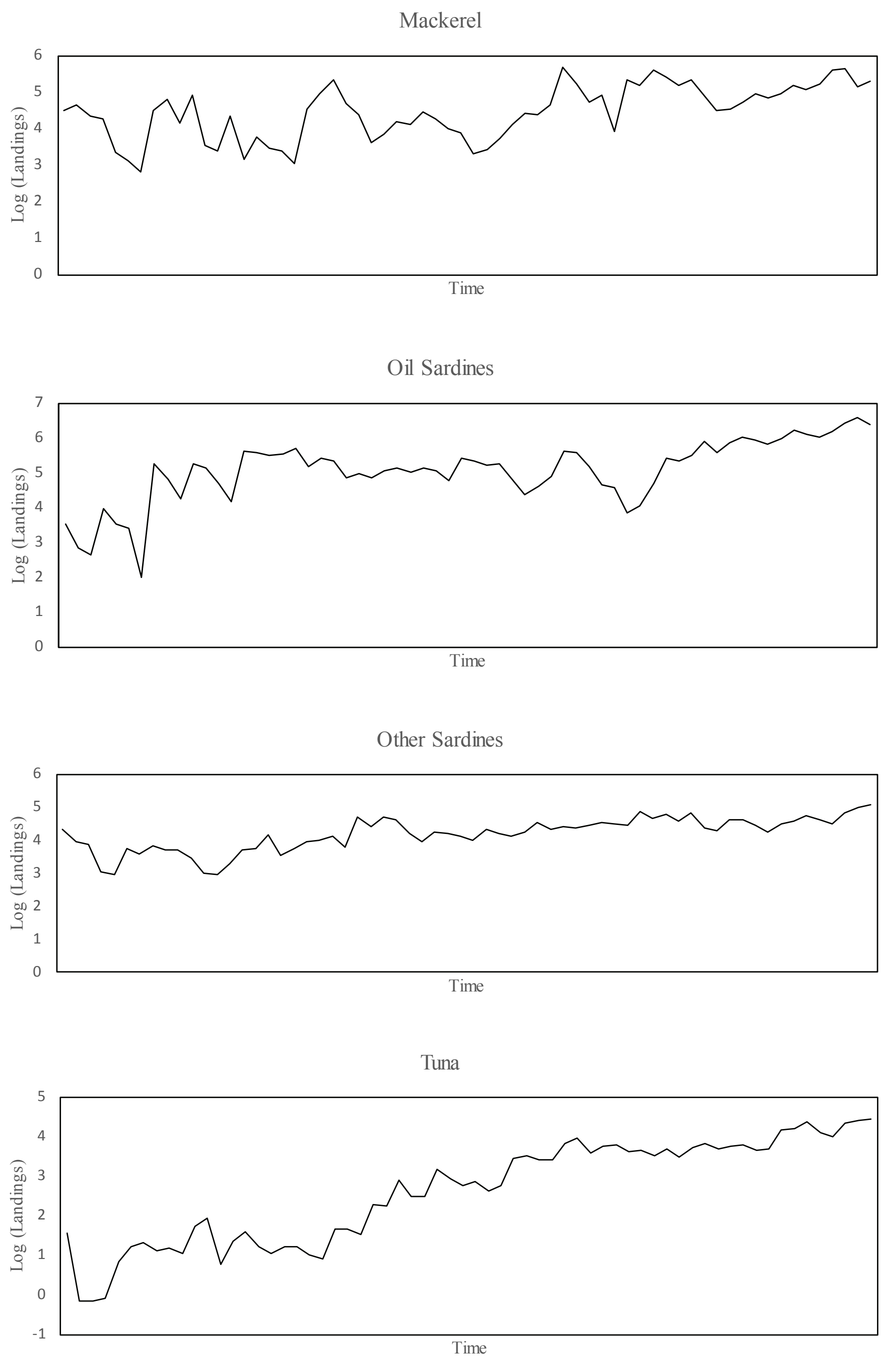
The above discussed methodology is applied to a real time series data of estimated annual landings of six different commercially important marine fish species (Oil Sardines, Mackerel, Other Sardines, Ribbon fish, Tuna and Seerfish) in India. The data consists of 64 observations for each variable (1950-2013). The first

58 observations are used for model fitting and last 6 observations are used for model evaluation. The data is available at *www.cmfri.org.in*. The analysis is carried out using SAS 9.3 and 'BigVAR' package in R.

3. RESULTS AND DISCUSSION

The BVAR and LVAR are applied to the real data on annual landings of six different fish species. For computational advantage, the natural logarithms of the values are considered for the analysis. The transformed data are plotted against time in Fig. 2. A perusal of the figure indicates that the series have an upward trend hinting at possible nonstationarity of the series.

To check the stationarity of the data, the Phillips-Perron unit root test is performed at 5% level of significance, the results of which are given in table 2.



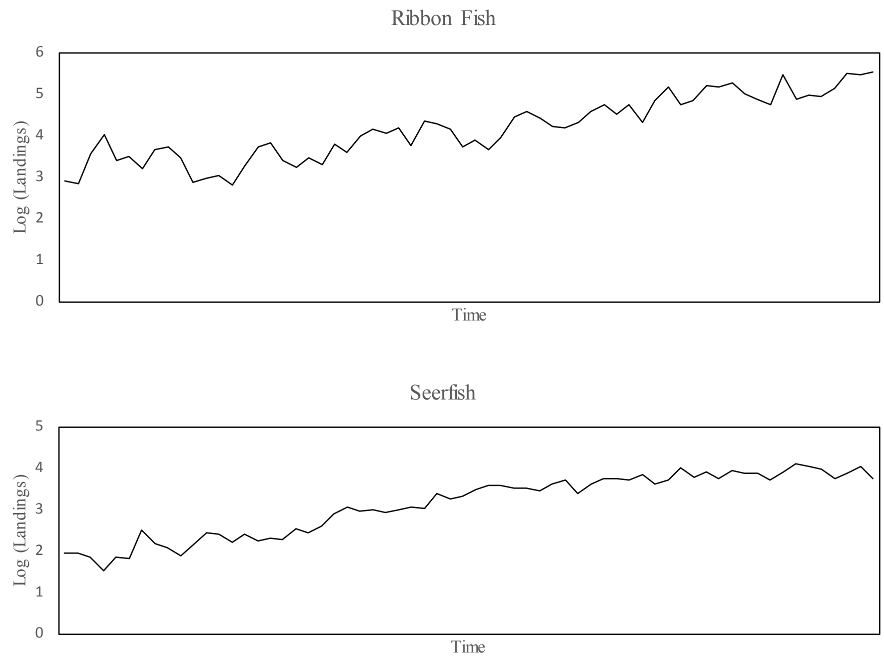


Fig. 2. Time plot of annual landings of different fish species in logarithmic scale

The results indicated that the series are stationary with respect to the trend.

Table 2. Results of the Phillips-Perron unit root test for the data on fish landings

Variable	Type	Rho	Pr < Rho	Tau	Pr < Tau
Oil Sardines	Single mean	-14.682	0.032	-2.938	0.047
	Trend	-22.81	0.022	-3.746	0.026
Mackerel	Single mean	-18.754	0.009	-3.241	0.022
	Trend	-28.051	0.005	-4.292	0.006
Other Sardines	Single mean	-10.024	0.117	-2.278	0.181
	Trend	-28.259	0.005	-4.581	0.002
Ribbon Fish	Single mean	-6.524	0.290	-1.950	0.307
	Trend	-33.167	0.001	-4.646	0.002
Tuna	Single mean	-2.241	0.743	-0.901	0.781
	Trend	-30.464	0.002	-4.609	0.002
Seerfish	Single mean	-1.527	0.828	-0.882	0.787
	Trend	-25.141	0.012	-3.941	0.016

Once the stationarity of the data series are established, the VAR, BVAR and LVAR of orders up to 6 are fitted to the data. The Akaike Information Criterion, AIC,(table 3) is used for choosing the appropriate order of the models. According to this criterion, the model which provides the least value for AIC among all the candidate models for the given data is selected as the preferred model.

Table 3. Akaike Information Coefficients of VAR, BVAR & LVAR models of different order

Order	Akaike Information Coefficients		
	VAR	Bayesian VAR	Lasso VAR
1	**-10.98**	**-10.98**	**0.764**
2	-9.259	-9.178	0.764
3	-8.255	-7.799	0.764
4	-6.895	-6.2	0.764
5	-6.192	-4.441	2.768
6	-6.968	-2.938	3.945

Based on the AIC, the VAR(1) and BVAR(1) were selected as the best models suitable for the data. For the LVAR model, all the four models from VAR(1) to VAR(4) provided the same AIC resulting in choosing the LVAR(1) model parsimoniously. Also, for the LVAR(1) model, different values are tried for the tuning parameter λ and the optimum value is chosen by minimizing the Mean Squared Forecasting Error (MSFE). The Least value of MSFE (=1.086) was obtained for λ=0.828. The values of MSFE for different values of λ are plotted in Fig. 3. Once the VAR(1) model is fitted, the parameter estimates were tested for their significance using the regular *t*-test. The parameters that are found to be non-significant at α=0.10 are constrained to zero and hence the parameters of the CVAR(1) are estimated.

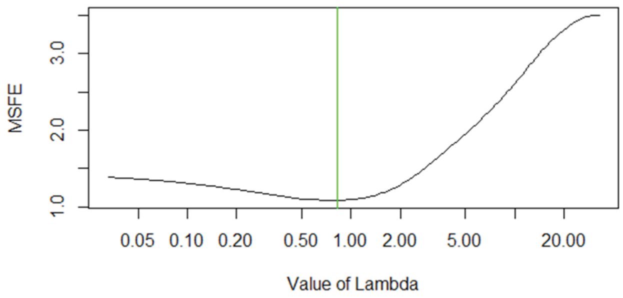


Fig 3. Mean Squared Forecasting Errors for different values of λ

The parameter estimates obtained from VAR, CVAR, BVAR and LVAR are given in table 4, table 5, table 6 and table 7, respectively. Table 4 also gives the *p*-values of the *t*-test for the significance of parameters. If we look at the estimates obtained from VAR and BVAR models, there is no much difference with respect to the number of parameters shrunk. Whereas, the LVAR is successful in shrinking 12 parameter estimates exactly to zero corroborating its better performance over VAR and BVAR.

Table 4. Parameter estimates of the VAR (1) model

μ	Φ_1					
1.187 *(0.269)*	0.599 *(<0.001)*	-0.010 *(0.945)*	-0.125 *(0.659)*	-0.025 *(0.911)*	-0.197 *(0.325)*	0.649 *(0.076)*
0.740 *(0.416)*	-0.122 *(0.198)*	0.508 *(<0.001)*	0.233 *(0.336)*	-0.076 *(0.689)*	-0.163 *(0.339)*	0.580 *(0.063)*
1.391 *(0.003)*	-0.040 *(0.397)*	-0.028 *(0.644)*	0.440 *(<0.001)*	0.044 *(0.646)*	-0.043 *(0.616)*	0.385 *(0.015)*
0.205 *(0.690)*	-0.061 *(0.257)*	0.073 *(0.279)*	0.050 *(0.716)*	0.567 *(<0.001)*	-0.143 *(0.139)*	0.563 *(0.002)*
-1.304 *(0.057)*	0.022 *(0.754)*	0.050 *(0.568)*	-0.244 *(0.178)*	0.363 *(0.013)*	0.521 *(<0.001)*	0.568 *(0.015)*
0.103 *(0.750)*	0.051 *(0.134)*	0.009 *(0.830)*	0.023 *(0.785)*	0.070 *(0.302)*	0.132 *(0.032)*	0.652 *(<0.001)*

Table 5. Parameter estimates of the Constrained VAR (1) model

μ	Φ_1					
1.120	0.645	0	0	0	0	0.220
1.134	0	0.537	0	0	0	0.294
1.461	0	0	0.399	0	0	0.336
0.679	0	0	0	0.582	0	0.344
-1.435	0	0	0	0.264	0.569	0.480
0.466	0	0	0	0	0.131	0.753

Table 6. Parameter estimates of the Bayesian VAR (1) model

μ	Φ1					
1.225	0.601	-0.012	-0.107	-0.025	-0.167	0.588
0.792	-0.114	0.507	0.227	-0.070	-0.138	0.531
1.399	-0.036	-0.026	0.446	0.042	-0.030	0.358
0.25	-0.056	0.067	0.058	0.569	-0.119	0.514
-1.29	0.024	0.046	-0.212	0.344	0.541	0.533
0.094	0.048	0.010	0.021	0.070	0.123	0.670

Table 7. Parameter estimates of the LVAR (1) model

μ	Φ_1					
1.383	0.590	0	0	0	0	0.224
1.289	-0.089	0.472	0.158	0	0	0.267
1.649	-0.002	0	0.367	0.006	0.011	0.303
0.763	-0.022	0.057	0.003	0.519	0	0.352
-1.44	-0.024	0	0	0.300	0.589	0.380
0.501	0.038	0	0	0.044	0.190	0.575

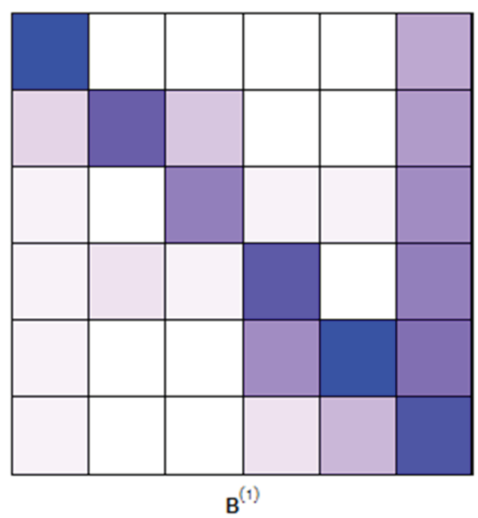


Fig 4. Sparsity pattern of the LVAR(1) model

The Fig. 4 gives the sparsity pattern of the parameter estimates obtained by the LVAR(1) model. The grey scale of the grid indicates the value of the weight of the parameter estimates; blank grids indicating that the estimates are exactly shrunk to zero.

Table 8. Forecasting performances of VAR, BVAR and LVAR models for training data

Variable	MAPE (%)				RMSE			
	VAR	CVAR	BVAR	LVAR	VAR	CVAR	BVAR	LVAR
Oil Sardines	9.770	9.595	9.747	9.806	0.570	0.581	0.570	0.579
Mackerel	8.955	9.680	8.958	9.233	0.483	0.507	0.484	0.493
Other Sardines	4.859	4.862	4.852	4.862	0.244	0.248	0.244	0.250
Ribbon Fish	5.616	6.089	5.674	6.040	0.274	0.287	0.274	0.283
Tuna	3.545	3.932	3.574	3.925	0.362	0.371	0.362	0.371
Seerfish	5.222	5.448	5.220	5.407	0.173	0.179	0.173	0.176

The measures of the forecast accuracy, given in table 8 and table 9 for training and testing data sets, respectively, indicate that both BVAR and LVAR have performed almost equally in forecasting the annual fish landings for different species. But when we look at the number of non-zero parameters in the model, the LVAR has performed better.

Table 9. Forecasting performances of VAR, BVAR and LVAR models for testing data

Variable	MAPE (%)				RMSE			
	VAR	CVAR	BVAR	LVAR	VAR	CVAR	BVAR	LVAR
Oil Sardines	7.277	6.556	7.351	8.198	0.528	0.490	0.533	0.589
Mackerel	6.360	4.611	6.309	6.858	0.405	0.312	0.402	0.427
Other Sardines	3.003	3.142	3.022	3.523	0.176	0.185	0.178	0.217
Ribbon Fish	4.980	5.258	5.061	5.924	0.323	0.337	0.328	0.390
Tuna	3.143	2.927	3.020	3.830	0.179	0.149	0.173	0.174
Seerfish	7.045	5.869	6.925	4.847	0.312	0.259	0.306	0.219

4. CONCLUSIONS

In the present study, the Bayesian VAR and LASSO VAR methods which are useful inaddressing the over parameterized VAR models are briefly studied.These methodologies are also empirically evaluated with VAR and Constrained VAR using the data on estimated annual landings of six different fish species. The models obtained are compared with each other using forecast evaluation techniques. Both the methods are found to perform evenly in forecasting as evident by RMSE and MAPE whereas LASSO VAR is found to be successful in shrinking more number of parameters to zero.

REFERENCES

Banbura, M., Giannone, D. and Reichlin, L. (2010). Large bayesian vector auto regressions. *J. App. Economet.,* **25**, 71-92.

Carriero, A., Kapetanios, G. and Marcellino M. (2009). Forecasting exchange rates with a large bayesian VAR, *Int. J. Forecasting,* **25(2)**, 400-417.

Gefand, D. (2014). Bayesian doubly adaptive elastic-net Lasso for VAR shrinkage. *Int. J. Forecasting*, **30**, 1-11.

Gutierrez, L., Piras, F. and Roggero, P. P. (2014). A global vector autoregression model for the analysis of wheat export prices. *Amer. J. Agril. Economics Advance Access*, 1-18.

Hsu, N.J., Hung, H.L. and Chang, Y.M. (2008). Subset selection for vector autoregressive processes using lasso. *Comp. Statist. Data Analysis,* **52**, 3645–3657.

Kim, B., Shin, S., Kim, Y.Y., Yum, S. and Kim, J. (2013). Forecasting demand of agricultural tractor, riding type rice transplanter and combine harvester by using an ARIMA model. *J. Biosyst. Engg.*, **38(1)**, 9-17.

Koop, G. and Korobilis, D. (2009). Bayesian multivariate time series methods for empirical macroeconomics. *Munich Personal RePEc Archive*, Paper No. 20125.

Litterman, R.B. (1986). Forecasting with Bayesian vector autoregressions - Five years of experience. *J. Busi. Eco. Statist.,* **4**, 25-38.

Lutkepohl, H. (1993). *Introduction to multiple time series analysis*. Springer-Verlag, Berlin.

Nicholson, W. B., Matteson, D. S. and Bien, J. (2014). Structured regularization for large vector autoregressions. *Technical report*, University of Cornell, 2014

Padhan, P.C., 2012. Application of ARIMA model for forecasting agricultural productivity in India*, J. Agril. Soc. Sci.,* **8**: 50-56.

Paul, R.K., Panwar, S., Sarkar, S. K., Kumar, A., Singh, K. N., Farooqi, S and Choudhary, V.K. (2013). Modelling and forecasting of meat exports from India. *Agril. Eco. Res. Review,* **26(2)**, 249-255.

Prabakaran, K., Sivapragasam, C., Jeevapriya, C and Narmatha, A. (2013). Forecasting cultivated areas and production of wheat in India using ARIMA model, *Golden Res. Thoughts*. **3(3)**.

Primicer, G.E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *The Review of Economic Studies*, **72(3)**, 821-852.

Sevinc, V. and Ergun, G. (2009). Usage of different prior distributions in Bayesian vector autoregressive models. *Hacettepe J. Math. Statist.*, **38(1)**, 85-93.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc.* **58(B)**, 267-288.