Dimension Reduction Methods for Multivariate Time Series

BigVAR

Will Nicholson
PhD Candidate
wbnicholson.com

github.com/wbnicholson/BigVAR

Department of Statistical Science
Cornell University

May 28, 2015

Acknowledgements

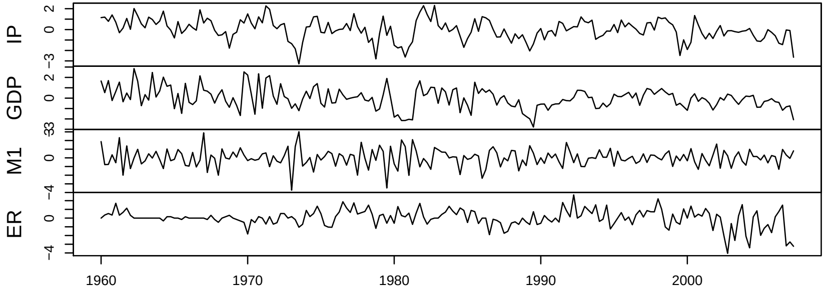
- Joint work with David S. Matteson (Cornell Statistics) and Jacob Bien (Cornell BSCB).
- Initial development of `BigVAR` package was supported by Google's Summer of Code in 2014.
- Amazon ec2 usage supported by an AWS in Education Research Grant.

Motivation

- The vector autoregression (VAR) has served as one of the core tools in modeling multivariate time series.
- Unfortunately, it is also heavily overparameterized.
- In certain scenarios, we may only desire forecasts from a subset of the included series (VARX).
- By applying structured penalties in both the VAR and VARX framework, we can substantially improve forecasting performance in a computationally efficient manner.

VARX Motivation

As an example, suppose that we are interested in forecasting 4 Canadian macroeconomic series: Industrial Production, M1, CA/US Exchange Rate



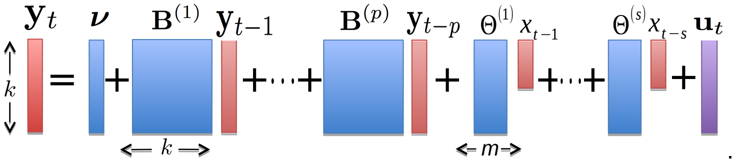
VARX

- Canada has a relatively small, open economy that is very interdependent with the US
- Empirical evidence suggests that US macroeconomic indicators can aid in predicting their Canadian counterparts.
- How can we effectively leverage the information from US macroeconomic series to improve Canadian forecasts?

Outline

- **Structured Regularization for Vector Autoregression with Exogenous Variables**
- Hierarchical Vector Autoregression
- BigVAR

VARX



- $\mathbf{y}_t \in \mathbb{R}^k$: modeled *endogenous* series
- $\mathbf{x}_t \in \mathbb{R}^m$: unmodeled *exogenous* series
- p, s : maximum lag orders
- $\mathbf{B}^{(\ell)} \in \mathbb{R}^{k \times k}$: endogenous coefficient matrix at lag ℓ
- $\boldsymbol{\theta}^{(j)} \in \mathbb{R}^{k \times m}$: exogenous coefficient matrix at lag j

The VAR can be viewed as a VARX in which $m = s = 0$

VARX Estimation

Provided k, s, m, p are small relative to T, we can forecast $\hat{\mathbf{y}}_{t+1}$ via least squares, by solving

$$\min_{\boldsymbol{\nu},\mathbf{B},\boldsymbol{\theta}} \|\mathbf{y}_t - \boldsymbol{\nu} - \sum_{\ell=1}^{p} \mathbf{B}^{(\ell)}\mathbf{y}_{t-\ell} - \sum_{j=1}^{s} \boldsymbol{\theta}^{(j)}\mathbf{x}_{t-j}\|_2^2, \tag{1}$$

where

$$\mathbf{B} = [\mathbf{B}^{(1)}, \dots, \mathbf{B}^{(p)}], \quad \boldsymbol{\theta} = [\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(s)}].$$

Equation (1) requires the estimation of $k(kp + ms + 1)$ least squares coefficients!

How can we reduce the parameter space of the VARX?

Existing Dimension Reduction Techniques

- Information Criterion based lag order selection (AIC/BIC) breaks down at high dimensions (Gonzalo and Pitarakis (2002)).
- Classical Bayesian Methods (Bayesian VAR) do not provide sparse solutions.
- Modern Bayesian approaches that impose sparsity (Stochastic Search Variable Selection) scale very poorly.
- Lasso-oriented regularization procedures have generally not been adapted to time-dependent problems.

Structured Lasso Penalties

We append structured convex penalties to (1) of the form

$$\lambda\left(\mathcal{P}_y(\mathbf{B}) + \mathcal{P}_x(\boldsymbol{\theta})\right), \tag{2}$$

in which

- $\lambda \geq 0$ is a penalty parameter
- $\mathcal{P}_y(\mathbf{B})$ endogenous group penalty structure
- $\mathcal{P}_x(\boldsymbol{\theta})$ exogenous group penalty structure

Penalties Considered:

- Group Lasso
- Sparse Group Lasso
- Lasso
- Nested Group Lasso

Why Penalized Regression?

- Fast, scalable solution algorithms
- Imposition of Sparsity
- Variable selection and estimation in one step
- **Reproducible**
- Not reliant on complex or subjective hyperparameters

Group Lasso

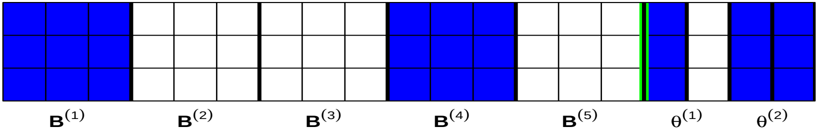
- $\mathbf{B}, \boldsymbol{\theta}$ are partitioned into natural "groupings"
- Within a group, coefficients will either all be nonzero or identically zero.
- We consider two endogenous groupings
  - By coefficient matrix, $\mathbf{B}^{(\ell)}$ (*Lag Group Lasso*)
  - Between "own" (diagonal of $\mathbf{B}^{(\ell)}$) and "other lags" (off diagonal elements of $\mathbf{B}^{(\ell)}$) (*Own/Other Group Lasso*)
- Each exogenous series is assigned to its own group

Example Group Lasso Sparsity Patterns (Shaded)

$(k = 3, p = 5; m = 2; s = 2)$

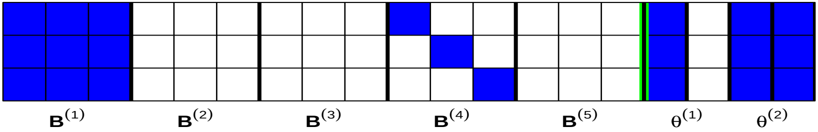
$$\mathcal{P}_y(\mathbf{B}) = \sqrt{k^2}\sum_{\ell=1}^{p}\|\mathbf{B}^{(\ell)}\|_2, \mathcal{P}_x(\boldsymbol{\theta}) = \sqrt{k}\sum_{j=1}^{s}\sum_{i=1}^{m}\|\boldsymbol{\theta}_{\cdot,i}^{(j)}\|_2.$$

Lag Group Lasso VARX



$$\mathcal{P}_y(\mathbf{B}) = \sqrt{k}\sum_{\ell=1}^{p}||\mathbf{B}_{\text{on}}^{(\ell)}||_2 + \sqrt{k(k-1)}\sum_{\ell=1}^{p}||\mathbf{B}_{\text{off}}^{(\ell)}||_2$$

Own/Other Group Lasso VARX



Sparse Group Lasso

- Limitations of the Group Lasso
  - If a group is active, all of its coefficients must be nonzero
  - Computational burden of including many small groups

- The Sparse Group Lasso (Simon et al. (2013)) addresses these shortcomings by adding an additional regularization parameter, α, to allow for within-group sparsity.

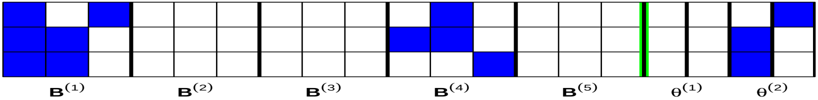
- The additional parameter is set based on a heuristic to control the degree of sparsity.

Example Sparse Group Lasso Sparsity Patterns (Shaded)

$$\mathcal{P}_y(\mathbf{B}) = (1-\alpha)\Big(\sqrt{k^2}\sum_{\ell=1}^{p}\|\mathbf{B}^{(\ell)}\|_2\Big) + \alpha\|\mathbf{B}\|_1,$$

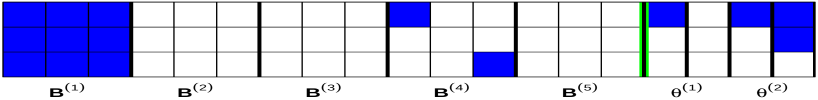
$$\mathcal{P}_x(\boldsymbol{\theta}) = (1-\alpha)\Big(\sqrt{k}\sum_{i=1}^{s}\sum_{j=1}^{m}\|\boldsymbol{\theta}^{(i)}_{\cdot,j}\|_2\Big) + \alpha\|\boldsymbol{\theta}\|_1.$$

Sparse Lag Group Lasso VARX



$$\mathcal{P}_y(\mathbf{B}) = (1-\alpha)\Big(\sqrt{k}\sum_{\ell=1}^{p}||\mathbf{B}^{(\ell)}_{\text{on}})||_2 + \sqrt{k(k-1)}\sum_{\ell=1}^{p}||\mathbf{B}^{(\ell)}_{\text{off}}||_2\Big) + \alpha\|\mathbf{B}\|_1$$

Sparse Own/Other Group Lasso VARX



Lasso

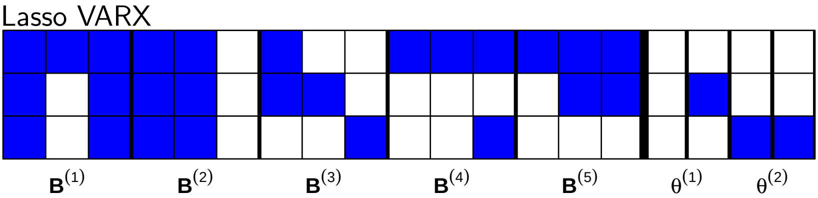
- The Lasso is the simplest grouping.

- Every parameter can be viewed as having its own group.

- Does not incorporate the VARX structure, but results in a comparably simpler optimization problem.

Example Lasso Sparsity Pattern (shaded)

$$\mathcal{P}_y(\mathbf{B}) = \|\mathbf{B}\|_1, \quad \mathcal{P}_x(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|_1$$



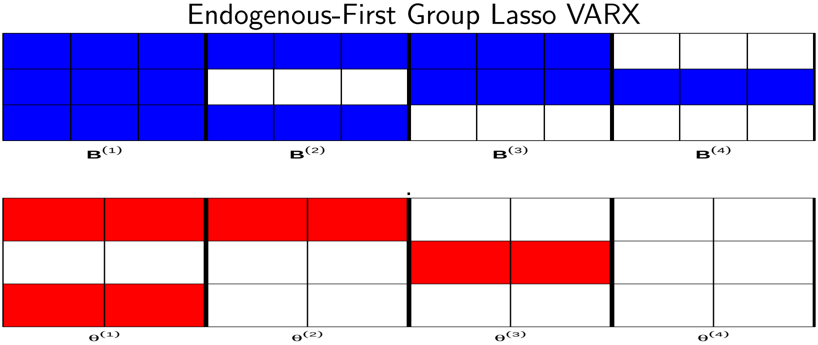
Nested Group Structures

- We have previously considered disjoint groupings that form a partition of $\mathbf{B}, \boldsymbol{\theta}$.
- In certain scenarios, one might wish to assign a relative important to endogenous versus exogenous variables.
- Our *Endogenous-First* Group Lasso penalty prioritizes endogenous series.
- At a given lag, an exogenous series can enter the model only if their endogenous counterpart is nonzero.

Example Endogenous-First Sparsity Pattern (shaded)

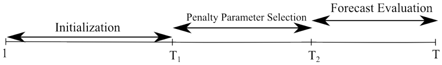
(k=3,p=4,m=2,s=4)

$$P_{x,y}(\mathbf{B},\boldsymbol{\theta}) = \sum_{\ell=1}^{p}\sum_{j=1}^{k}\left(\|[\mathbf{B}_j^{(\ell)},\boldsymbol{\theta}_{j,\cdot}^{(\ell)}]\|_2 + \|\boldsymbol{\theta}_{j,\cdot}^{(\ell)}\|_2\right)$$



Penalty Parameter Selection: "Rolling" Cross-Validation

- Following Banbura et al. (2009), we utilize a selection procedure that respects time dependence.
- Select $\hat{\lambda}$ from a grid of values $\lambda_1, \ldots, \lambda_n$.
- At T_1, we forecast $\hat{\mathbf{y}}_{T_1+1}^{\lambda_i}$ for $i = 1, \ldots, n$, and sequentially add observations until time T_2.
- We choose $\hat{\lambda}$ as the minimizer of MSFE.
- T_2 through T is used to evaluate the forecasting accuracy of $\hat{\lambda}$.



Application

- We can use our VARX framework to forecast the previously described 4 Canadian macroeconomic series ($k = 4$).
- Quarterly, ranging from Q1 1960 to Q4 2007 ($T \approx 200$).
- 20 US macroeconomic indicators procured from Koop (2011) used as exogenous predictors ($m = 20$).
- Quarter 1 of 1977 to Quarter 1 of 1992 is used for penalty parameter selection while Quarter 2 of 1992 to Quarter 4 of 2007 is used for forecast evaluation.

Results

Table : Out of sample MSFE of one-step ahead VARX forecasts of 4 Canadian macroeconomic indicators with 20 exogenous predictors $p = 4, s = 4$

Model/ VARX Penalty Structure	MSFE
Lasso	2.996
Lag Group Lasso	2.988
Own/Other Group Lasso	2.995
Sparse Lag Group Lasso	2.959
Sparse Own/Other Group Lasso	2.984
Endogenous-First VARX	3.033
VAR with lag selected by AIC	3.341
VAR with lag selected by BIC	3.201
Sample Mean	3.052
Random Walk	4.545

VAR Results

Table : Out of sample MSFE of one-step ahead VAR forecasts of 4 Canadian Macroeconomic Indicators $p = 4$

Model/VAR Penalty Structure	MSFE
Lasso	3.027
Lag Group Lasso	3.075
Own/Other Group Lasso	3.303
Sparse Lag Group Lasso	3.042
Sparse Own/Other Group Lasso	3.037

Outline

- Structured Regularization for Vector Autoregression with Exogenous Variables
- **Hierarchical Vector Autoregression**
- BigVAR

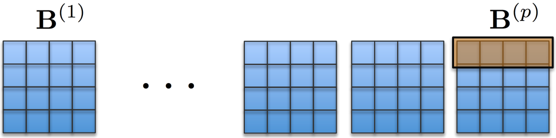
Hierarchical VAR (HVAR)

- The previous penalties remain agnostic with regard to lag order selection.
- We propose a *hierarchical* group lasso penalty that takes into account lag order in the VAR context.
- Distant lags are penalized before recent lags
- Allows for varying lag order across marginal models.
- We present three HVAR Penalties:
  - Componentwise
  - Own/Other
  - Elementwise

Componentwise HVAR

Maximum lag order can vary across marginal models, but within a series, all components have the same maximum lag.

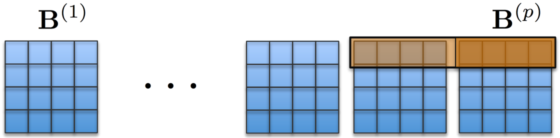
$$\mathcal{P}_y(\mathbf{B}) = \sum_{i=1}^{k} \sum_{\ell=1}^{p} \|\mathbf{B}_i^{(\ell:p)}\|_2$$



Componentwise HVAR

Maximum lag order can vary across marginal models, but within a series, all components have the same maximum lag.

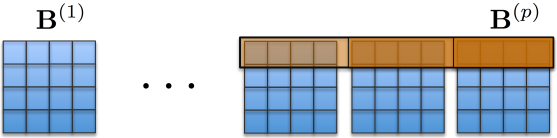
$$\mathcal{P}_y(\mathbf{B}) = \sum_{i=1}^{k} \sum_{\ell=1}^{p} \|\mathbf{B}_i^{(\ell:p)}\|_2$$



Componentwise HVAR

Maximum lag order can vary across marginal models, but within a series, all components have the same maximum lag.

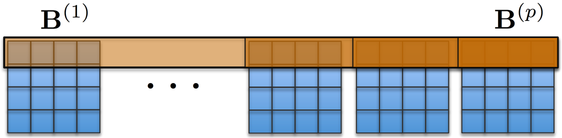
$$\mathcal{P}_y(\mathbf{B}) = \sum_{i=1}^{k}\sum_{\ell=1}^{p}\|\mathbf{B}_i^{(\ell:p)}\|_2$$



Componentwise HVAR

Maximum lag order can vary across marginal models, but within a series, all components have the same maximum lag.

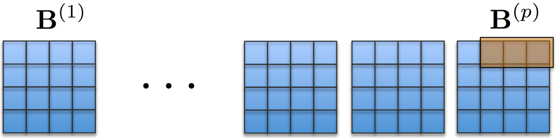
$$\mathcal{P}_y(\mathbf{B}) = \sum_{i=1}^{k} \sum_{\ell=1}^{p} \|\mathbf{B}_i^{(\ell:p)}\|_2$$



Own/Other HVAR

Similar to Componentwise, but within a lag prioritizes "own" lags over "other" lags.

$$\mathcal{P}_y(\mathbf{B}) = \sum_{i=1}^{k} \sum_{\ell=1}^{p} \left[\|\mathbf{B}_i^{(\ell:p)}\|_2 + \|(\mathbf{B}_{i,-i}^{(\ell)}, \mathbf{B}_i^{([\ell+1]:p)})\|_2 \right]$$

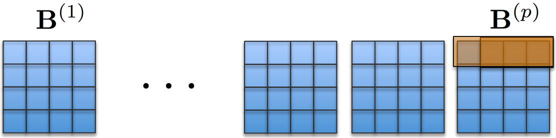


.

Own/Other HVAR

Similar to Componentwise, but within a lag prioritizes "own" lags over "other" lags.

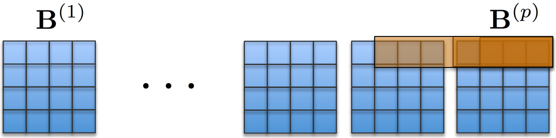
$$\mathcal{P}_y(\mathbf{B}) = \sum_{i=1}^{k} \sum_{\ell=1}^{p} \left[\|\mathbf{B}_i^{(\ell:p)}\|_2 + \|(\mathbf{B}_{i,-i}^{(\ell)}, \mathbf{B}_i^{([\ell+1]:p)})\|_2 \right]$$



Own/Other HVAR

Similar to Componentwise, but within a lag prioritizes "own" lags over "other" lags.

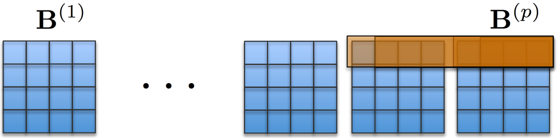
$$\mathcal{P}_y(\mathbf{B}) = \sum_{i=1}^{k} \sum_{\ell=1}^{p} \left[\|\mathbf{B}_i^{(\ell:p)}\|_2 + \|(\mathbf{B}_{i,-i}^{(\ell)}, \mathbf{B}_i^{([\ell+1]:p)})\|_2 \right]$$



Own/Other HVAR

Similar to Componentwise, but within a lag prioritizes "own" lags over "other" lags.

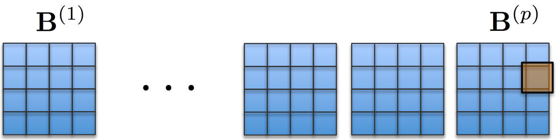
$$\mathcal{P}_y(\mathbf{B}) = \sum_{i=1}^{k} \sum_{\ell=1}^{p} \left[\|\mathbf{B}_i^{(\ell:p)}\|_2 + \|(\mathbf{B}_{i,-i}^{(\ell)}, \mathbf{B}_i^{([\ell+1]:p)})\|_2 \right]$$



Elementwise HVAR

The most general structure: in each marginal model, each series may have its own maximum lag.

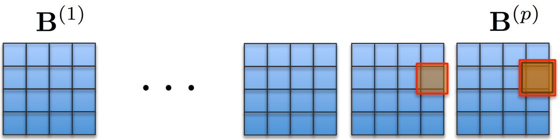
$$\mathcal{P}_y(\mathbf{B}) = \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{\ell=1}^{p} \|\mathbf{B}_{ij}^{(\ell:p)}\|_2$$



Elementwise HVAR

The most general structure: in each marginal model, each series may have its own maximum lag.

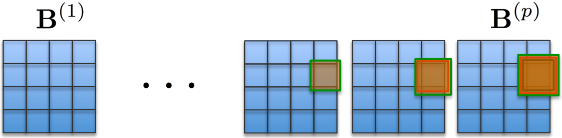
$$\mathcal{P}_y(\mathbf{B}) = \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{\ell=1}^{p}\|\mathbf{B}_{ij}^{(\ell:p)}\|_2$$



Elementwise HVAR

The most general structure: in each marginal model, each series may have its own maximum lag.

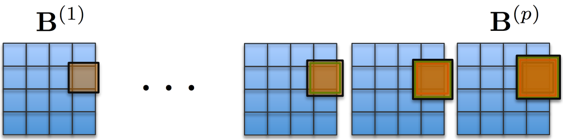
$$\mathcal{P}_y(\mathbf{B}) = \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{\ell=1}^{p} \|\mathbf{B}_{ij}^{(\ell:p)}\|_2$$



Elementwise HVAR

The most general structure: in each marginal model, each series may have its own maximum lag.

$$\mathcal{P}_y(\mathbf{B}) = \sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{\ell=1}^{p} \|\mathbf{B}_{ij}^{(\ell:p)}\|_2$$



.

Data Example

We consider forecasting 168 macroeconomic indicators (the "Large" model of Koop (2011), $k = 168$, $T \approx 200$, $p = 13$)

Class	**Method**	MSFE
HVAR	Componentwise	104.143
	Own-other	93.004
	Elementwise	93.172
VAR	Lasso	103.555
	Lag-weighted lasso	104.244
Other	Sample mean	120.120
	Random walk	212.368

Outline

- Structured Regularization for Vector Autoregression with Exogenous Variables
- Hierarchical Vector Autoregression
- **BigVAR**

BigVAR

- R package designed for penalized regression in a multivariate time series setting.
- All solution algorithms are optimized for use in time-dependent problems, written in C++ and linked via Rcpp.
- Utilizes s4 object classes.
- User-friendly interface.

Implementation Example (HVAR)

We consider forecasting 20 US Macroeconomic Indicators: the "Medium" model of Koop (2011)

BigVAR

`constructModel` creates an s4 object of class `BigVAR`

```r
library(BigVAR)
mod1 = constructModel(Y, p = 4, struct = "None", gran = c(5, 10), verbose = FALSE)
```

Arguments:

- Y: $T \times k$ time series or $T \times (k + m)$ endogenous and exogenous series
- p: maximum lag order for endogenous coefficients
- struct: Structured Penalty
- gran: Penalty Grid Options (depth and number of penalty parameters)
- verbose: option to display a progress bar
- VARX (optional): VARX specifications (k and s)
- For other (non-required) options see the package manual

Struct Options

Struct Argument	Penalty	VAR	VARX
"Lag"	Lag Group Lasso	X	X
"Diag"	Own/Other Group Lasso	X	X
"SparseLag"	Lag Sparse Group Lasso	X	X
"SparseDiag"	O/O Sparse Group Lasso	X	X
"None"	Lasso	X	X
"EF"	Endogenous-First VARX	.	X
"HVARC"	Componentwise Hierarchical	X	.
"HVAROO"	Own/Other Hierarchical	X	.
"HVARELEM"	Elementwise Hierarchical	X	.
"Tapered"	Lag Weighted Lasso	X	.

Estimation

- Fit models with the `BigVAR` method: `cv.BigVAR`
- Performs rolling cross validation, forecast evaluation, and compares against AIC, BIC, sample mean, and random walk benchmarks.
- Returns an object of class `BigVAR.results`, which inherits class `BigVAR` .

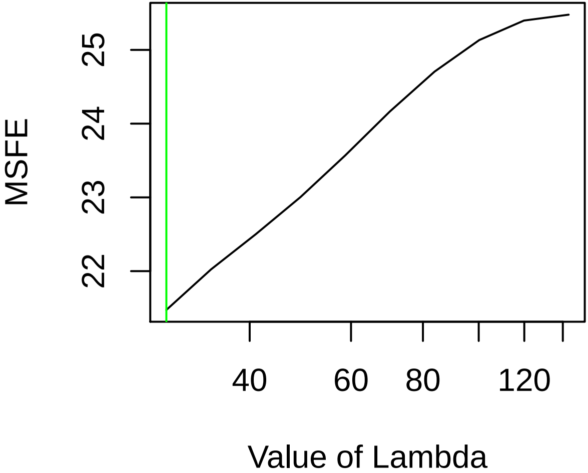
```
res = cv.BigVAR(mod1)
res
## *** BIGVAR MODEL Results ***
## Structure
## [1] "None"
## Maximum Lag Order
## [1] 4
## Optimal Lambda
## [1] 28.663
## Grid Depth
## [1] 5
## Index of Optimal Lambda
## [1] 10
## In-Sample MSFE
## [1] 21.474
## BigVAR Out of Sample MSFE
## [1] 12.221
## *** Benchmark Results ***
## Conditional Mean Out of Sample MSFE
## [1] 14.876
## AIC Out of Sample MSFE
## [1] 22.188
## BIC Out of Sample MSFE
## [1] 12.995
## RW Out of Sample MSFE
## [1] 29.14
```

Diagnostics

- The end-user has some flexibility with regard to the granularity of the penalty grid.
- Ideally, $\hat{\lambda}$ should be near the middle of the grid, to ensure that it is deep enough.
- If it is at the boundary, increasing the first parameter of gran may improve forecast performance.
- However, too large of a value will unncessarily increase computation time.
- `plot.BigVAR.Results` can be used to visualize this relationship

Diagnostics

```r
plot(res)
```

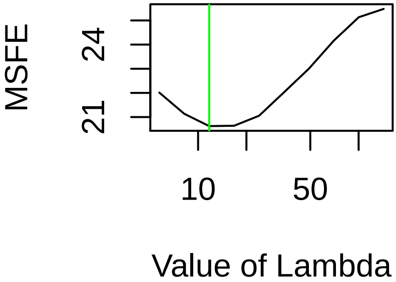


Diagnostics

```
mod1@Granularity = c(25, 10)
res2 <- cv.BigVAR(mod1)
mean(res2@OOSMSFE)

## [1] 12.19338

plot(res2)
```



Other Capabilities

- h-step ahead forecasts can be obtained by the `predict` method

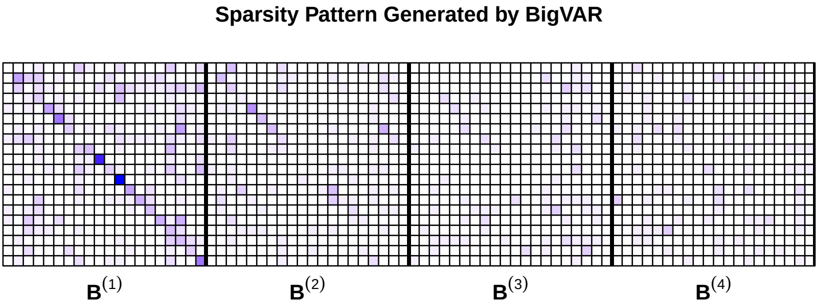
```
# Forecasts for GDP, CPI, Federal Funds Rate
predict(res2, 1)[1:3, ]

## [1] -0.1916194  0.6542203 -0.3422589
```

Other Capabilities

- Sparsity Plots depicting nonzero coefficients with `SparsityPlot.BigVAR.results`

```
SparsityPlot.BigVAR.results(res2)
```



Future Extensions

- Alternative cross validation procedures; incorporation of online learning.
- Extending `BigVAR` to incorporate structural econometric modeling.
- Penalized Maximum Likelihood as an alternative to least squares.

Our Papers

William B. Nicholson, Jacob Bien, and David S. Matteson. Hierarchical vector autoregression. arXiv preprint arXiv:1412.5250, 2014.

William B. Nicholson, David S. Matteson, and Jacob Bien. Structured Regularization for Large Vector Autoregression with Exogenous Variables. `http://www.wbnicholson.com/Nicholsonetal2015` , 2015.