VARshrink: Shrinkage Estimation Methods for Vector Autoregressive Models

Namgil Lee
Kangwon National University

Heon-Young Yang
Kangwon National University

Sung-Ho Kim
KAIST

Abstract

We introduce an R software package, **VARshrink**, for providing shrinkage estimation methods for vector autoregressive (VAR) models. Contrary to the standard ordinary least squares method, shrinkage estimation methods can be applied to high-dimensional VAR models with dimensionality greater than the number of observations. While existing R packages for VAR shrinkage are mostly based on parametric, Bayesian estimation methods using informative priors, the **VARshrink** aims to be an integrative R package delivering nonparametric, parametric, and semiparametric methods in a unified and consistent manner. The current version of **VARshrink** contains four shrinkage estimation methods, which are the multivarate ridge regression, a nonparametric shrinkage method, a full Bayesian approach using noninformative priors, and a semiparametric Bayesian approach using informative priors. VAR estimation problems are formulated as a multivariate regression form, so that all the shrinkage estimation methods can be run by one interface function named `VARshrink()`. We clearly present mathematical expressions of shrinkage estimators of the shrinkage estimation methods. The effective number of parameters can be calculated based on a shrinkage intensity parameter value, which plays an important role for further statistical analyses. We provide a sample R code for demonstration of the package and comparison of performances of the shrinkage estimation methods.

Keywords: Bayes estimation, vector autoregression (VAR), high-dimensionality, shrinkage, multivariate time series.

1. Introduction

Let $\mathbf{y}_t = (y_{t1}, y_{t2}, \ldots, y_{tK})^\top$ denote a $K \times 1$ vector of endogenous variables. A vector autoregressive (VAR) model of order p can be expressed as

$$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{C}\mathbf{d}_t + \boldsymbol{\epsilon}_t, \tag{1}$$

where $\mathbf{d}_t$ is an $L \times 1$ vector of deterministic regressors, $\boldsymbol{\epsilon}_t$ is a $K \times 1$ noise vector, and $\mathbf{A}_1, \ldots, \mathbf{A}_p$

and **C** are coefficient matrices (Hamilton 1994; Tsay 2005).

1.1. Backgrounds

Recently, several R packages have been developed for parameter estimation and forecasting using stochastic time series models. The **forecast** package provides various methods and tools for univariate time series models such as the ARIMA and ETS models (Hyndman, Athanasopoulos, Bergmeir, Caceres, Chhay, O'Hara-Wild, Petropoulos, Razbash, Wang, Yasmeen, R Core Team, Ihaka, Reid, Shaub, Tang, and Zhou 2018). The **MTS** package has been developed for a wide variety of multivariate linear time series models and multivariate volatility models such as the VARMA, multivariate EWMA, and low-dimensional BEKK models (Tsay and Wood 2018). The **vars** package provides methods for multivariate time series analysis using the VAR, SVAR, and SVEC models (Pfaff and Stigler 2018).

In this study, we focus on shrinkage estimation of VAR model parameters. Shrinkage estimation methods have been playing crucial roles in high-dimensional statistical modeling; see, e.g., Beltrachini, von Ellenrieder, and Muravchik (2013), Böhm and von Sachs (2009), Fiecas and Ombao (2011) and Ledoit and Wolf (2004). For VAR models, several shrinkage estimation methods have been suggested such as a Stein-type nonparametric shrinkage estimation method (Opgen-Rhein and Strimmer 2007b), Bayesian VARs using informative priors (Bańbura, Giannone, and Reichlin 2010; Doan, Litterman, and Sims 1984; Koop and Korobilis 2010; Litterman 1986), Bayesian VARs using noninformative priors (Ni and Sun 2005; Sun and Ni 2004), and a semiparametric Bayesian approach adopting a modified K-fold cross validation (Lee, Choi, and Kim 2016).

Due to its popularity in macroeconomic time series analysis, several Bayesian VAR methods have been implemented in R packages. For example, the function `BVAR()` in **MTS** implements a Bayesian VAR method using an informative prior (Tsay and Wood 2018); the package **bvarsv** implements Bayesian VAR models with stochastic volatility and time-varying parameters (Krueger 2015; Koop and Korobilis 2010; Primiceri 2005).

On the other hand, we note that other types of shrinkage methods which have been implemented for other purposes than multivariate time series analysis can be applied to shrinkage estimation of VAR parameters. For instance, the function `cov.shrink()` in the package **corpcor** has been implemented to compute shrinkage estimates of covariances, but it has been applied to estimate VAR coefficients in Opgen-Rhein and Strimmer (2007b; Schäfer, Opgen-Rhein, Zuber, Ahdesmäki, Silva, and Strimmer 2017). Moreover, we note that VAR models can be reformulated into multivariate regression problems, so that penalized least squares methods can be used for shrinkage estimation of VAR parameters, e.g., the functions `lm.gls()` for generalized least squares and `lm.ridge()` for ridge regression in the package **MASS** (Ripley, Venables, Bates, Hornik, Gebhardt, and Firth 2018); the function `glmnet()` for Lasso and Elastic-Net regularized generalized linear models in the package `glmnet()` (Friedman, Hastie, Tibshirani, Simon, Narasimhan, and Qian 2018); the function `linearRidge()` for ridge regression in the package **ridge** (Moritz and Cule 2018).

1.2. Main Purpose

While Bayesian approaches have been widely used in the literature, we note that nonparametric and semiparametric approaches have advantages in the case of high-dimensional VARs with more than several hundreds of time series variables due to their relatively low computational

costs (Opgen-Rhein and Strimmer 2007b). Despite of their relatively high computational costs, Bayesian approaches can impose proper assumptions on the multivariate time series data flexibly, such as VAR roots near unity and correlations between noise processes (Lee *et al.* 2016). In this sense, a semiparametric approach can be a trade-off between nonparametric and parametric approaches (Lee *et al.* 2016).

In this study, we developed an integrative R package, **VARshrink**, for implementing nonparametric, parametric, and semiparametric approaches for shrinkage estimation of VAR model parameters. By providing a simple interface function, `VARshrink()`, for running various types of shrinkage estimation methods, the performance of the methods can be easily compared. We note that the package **vars** implemented an ordinary least squares method for VAR models by the function `VAR()`. The function `VARshrink()` was built to extend `VAR()` to shrinkage estimation methods, so that the **VARshrink** package takes advantage of the tools and methods available in the **vars** package. For example, we demonstrate the use of model selection criteria such as AIC and BIC in this paper, where the methods `AIC()`, `BIC()`, and `logLik()` can handle multivariate t-distribution and the effective number of parameters in **VARshrink**.

This paper is organized as follows. In Section 2, we explain the general distribution assumption for the noise term in a Bayesian framework. We also present the formulation of VAR models in a multivariate regression problem, which simplifies implementation of the package. In Section 3, we describe the common interface function and the four shrinkage estimation methods included in the package. We clearly present closed form expressions for the shrinkage estimators inferred by the shrinkage methods, so that we can indicate the role of the shrinkage intensity parameters in each method. In addition, we explain how the effective number of parameters can be computed for shrinkage estimators. In Section 4, we present numerical experiments using benchmark data and simulated data for comparing performances of the shrinkage estimation methods. Discussion and conclusions are provided in Section 5.

---

2. Models

In the VAR model in (1), we assume that the noise vectors are independent and identically distributed from a scale mixture of multivariate normal distributions, including multivariate normal distributions, multivariate t-distributions (Ni and Sun 2005), and multivariate Laplace distributions (MLD) (Eltoft, Kim, and Lee 2006). By incorporating scale mixture distributions in the model assumption, we can handle outliers appropriately and produce robust estimates (West 1984). In specific, we consider that a noise vector can be expressed as a product

$$\boldsymbol{\epsilon}_t = \mathbf{z}_t q_t^{-1/2}, \tag{2}$$

where $\mathbf{z}_t \sim \mathrm{N}_K(\mathbf{0}, \boldsymbol{\Sigma})$ is a $K \times 1$ vector having a multivariate normal distribution with the mean vector of zeros and the covariance matrix of $\boldsymbol{\Sigma}$, and $q_t \sim \mathrm{Gamma}(\nu/2, \nu/2)$ is a random variable having a gamma distribution with shape $\alpha = \nu/2$ and rate $\beta = \nu/2$. The following statements describe characteristics of the distributions in more detail.

- If $1 < \nu < \infty$, then $\boldsymbol{\epsilon}_t$ has a multivariate t-distribution with the degree of freedom ν.

The probability density function (p.d.f.) for $\boldsymbol{\epsilon}_t$ can be expressed as

$$f\left(\boldsymbol{\epsilon}_t\right) = g\left(\left\|\mathbf{V}^{-1/2}\boldsymbol{\epsilon}_t\right\|^2\right)|\mathbf{V}|^{-1/2}, \tag{3}$$

with

$$g(x) \propto \left(1 + \frac{x}{\nu}\right)^{-(\nu+K)/2}. \tag{4}$$

Define

$$h(x) = -2\frac{\mathrm{d}}{\mathrm{d}x}\log g(x) = \frac{\nu+K}{\nu+x}. \tag{5}$$

- If $\nu \to \infty$, the distribution for $\boldsymbol{\epsilon}_t$ converges to the multivariate normal distribution, $\mathrm{N}_K(\mathbf{0},\boldsymbol{\Sigma})$. The p.d.f. for $\boldsymbol{\epsilon}_t$ can be expressed as (3) with $g(x) \propto \exp(-x/2)$ and $h(x) = 1$.

- If $\nu = 1$, the distribution for $\boldsymbol{\epsilon}_t$ converges to the multivariate Cauchy distribution. The p.d.f. for $\boldsymbol{\epsilon}_t$ can be expressed as (3) with $g(x) \propto (1+x)^{-(1+K)/2}$ and $h(x) = (1+K)/(1+x)$.

- If $\nu \approx 0$, the distribution cannot be obtained in a simple closed form, but the p.d.f. for $\boldsymbol{\epsilon}_t$ can still be written with $g(x) \propto x^{-K/2}$ and $h(x) = K/x$.

  In this case, the expression for q_t during the iteration process in semiparametric Bayesian approach (Lee *et al.* 2016) is equivalent to the IRLS procedure for L1 norm minimization (Chartrand and Yin 2008); see Section 3.4 for more detail.

  In addition, we note that the p.d.f. $f^{MLD}(\boldsymbol{\epsilon}_t)$ of the multivariate Laplace distribution (MLD) can be expressed as in (3) with

  $$g^{MLD}(x) \propto \frac{\tilde{K}_{(K-2)/2}(\sqrt{2x/\beta})}{(\sqrt{\beta x/2})^{(K-2)/2}}$$

  for some $\beta > 0$ (Eltoft *et al.* 2006). Since $\tilde{K}_\alpha(z)$, the modified Bessel function of the second kind, can be approximated by $\tilde{K}_\alpha(z) \approx \Gamma(\alpha)2^{\alpha-1}z^{-\alpha}$ for small values of $0 < z \leq \sqrt{\alpha+1}$, the p.d.f. of MLD can be approximated by $g^{MLD}(x) \propto x^{-K/2+1}$ for small values of $0 < x \leq \sqrt{K\beta/4}$. Therefore, we consider that the multivariate t-distribution is approximately close to the MLD when $\nu \to 0$.

In general, we can rewrite the model equation (1) in the form of a multivariate regression as

$$\mathbf{y}_t = \boldsymbol{\Psi}^\top \mathbf{x}_t + \boldsymbol{\epsilon}_t, \tag{6}$$

where $\boldsymbol{\Psi} = (\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p, \mathbf{C})^\top$ is a $(Kp+L) \times K$ matrix of coefficients, $\mathbf{x}_t = (\mathbf{y}_{t-1}^\top, \dots, \mathbf{y}_{t-p}^\top, \mathbf{d}_t^\top)^\top$ is a $(Kp+L) \times 1$ vector of regressors. For estimation of VAR parameters from the observed time series data $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T$, we define data matrices as

$$\begin{aligned} \mathbf{Y} &= (\mathbf{y}_{p+1}, \mathbf{y}_{p+2}, \dots, \mathbf{y}_T)^\top \in \mathbb{R}^{N \times K}, \\ \mathbf{X} &= (\mathbf{x}_{p+1}, \mathbf{x}_{p+2}, \dots, \mathbf{x}_T)^\top \in \mathbb{R}^{N \times (Kp+L)}, \end{aligned} \tag{7}$$

with $N = T - p$. Then, we can rewrite (6) in a matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\Psi} + \mathbf{E} \in \mathbb{R}^{N \times K}, \tag{8}$$

with $\mathbf{E} = (\boldsymbol{\epsilon}_{p+1}, \dots, \boldsymbol{\epsilon}_T)^\top$ and $N = T - p$.

3. Shrinkage Estimation Methods

In this section, we will describe the shrinkage estimation methods implemented in this package. The methods are used for estimating the VAR coefficient matrix $\boldsymbol{\Psi}$ alone or both of the $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ in (8).

We provide a common R function interface `VARshrink()` for running the estimation methods, which is defined by

```
VARshrink(y, p = 1, type = c("const", "trend", "both", "none"),
  season = NULL, exogen = NULL, method = c("ridge", "ns", "fbayes",
  "sbayes", "kcv"), lambda = NULL, lambda_var = NULL, dof = Inf, ...)
```

The input arguments are described as follows.

- `y`: A T-by-K matrix of endogenous variables
- `p`: Integer for the lag order
- `type`: Type of deterministic regressors to include.

    1) "const" - the constant: $\mathbf{d}_t = 1$.
    2) "trend" - the trend: $\mathbf{d}_t = t$.
    3) "both" - both the constant and the trend: $\mathbf{d}_t = (t, 1)^\top$.
    4) "none" - no deterministic regressors.

- `season`: An integer value of frequency for inclusion of centered seasonal dummy variables.
- `exogen`: A T-by-L matrix of exogenous variables.
- `method`:

    1) "ridge" - multivariate ridge regression.
    2) "ns" - a Stein-type nonparametric shrinkage method.
    3) "fbayes" - a full Bayesian shrinkage method using noninformative priors.
    4) "sbayes" - a semiparametric Bayesian shrinkage method using parameterized cross validation.
    5) "kcv" - a semiparametric Bayesian shrinkage method using K-fold cross validation

- `lambda, lambda_var`: Shrinkage parameter value(s). Use of this parameter is slightly different for each method: the same value does not imply the same shrinkage estimates. See description in the following subsections for the use of shrinkage parameters in each method.
- `dof`: Degree of freedom of multivariate t-distribution for noise. Valid only for `method = "fbayes"` and `method = "sbayes"`. `dof = Inf` means multivariate normal distribution.

The output value is an object of class "varshrinkest", which inherits the class "varest" in the package **vars**, and an object of class "varest" can be obtained by the function `VAR()` in the package **vars**. The input arguments and the output value indicate that `VARshrink()` is an

Table 1: Structure of the package VARshrink.

Function or method	Class	Methods for class	Functions for class
VAR	varshrinkest, varest	coef, fevd, fitted, irf, logLik, Phi, plot, predict, print, Psi, resid, summary	Acoef_sh, arch.test_sh, Bcoef_sh, BQ_sh, causality_sh, normality.test_sh, restrict_sh, roots_sh, serial.test_sh, stability_sh
fevd	varfevd	plot, print	
irf	varshirf, varirf	plot, print	
predict	varprd	plot, print	fanchart
summary	varshsum, varsum	print	
arch.test_sh	varcheck	plot, print	
normality.test_sh	varcheck	plot, print	
serial.test_sh	varcheck	plot, print	
stability_sh	varstabil	plot, print	

extension of `VAR()` in the **vars** package. As a result, almost all the methods and functions included in the package **vars** are available for the package **VARshrink**, such as `fevd()`, `Phi()`, `plot()`. The methods and functions available in the **VARshrink** package are summarized in Table 1. The names of the functions for class "varshrinkest" has suffix "_sh" in order to distinguish them from the functions for class "varest". We remark that the classes "varshrinkest", "varshirf", and "varshsum" inherit the classes "varest", "varirf", and "varsum" of the **vars** package, respectively.

3.1. Multivariate Ridge Regression

The ridge regression method is a kind of penalized least squares (PLS) method, which produces a biased estimate of the VAR coefficient (Hoerl and Kennard 1970). Formally speaking, the ridge regression estimator of $\mathbf{\Psi}$ can be obtained by minimizing the penalized sum of squared prediction errors (SSPE) as

$$\widehat{\mathbf{\Psi}}^{\mathrm{R}}(\lambda) = \arg\min_{\mathbf{\Psi}} \ \frac{1}{N} \left\| \mathbf{Y} - \mathbf{X}\mathbf{\Psi} \right\|_F^2 + \lambda \left\| \mathbf{\Psi} \right\|_F^2, \tag{9}$$

where $\|\mathbf{A}\|_F = \sqrt{\sum_i \sum_j a_{ij}^2}$ is the Frobenius norm of a matrix $\mathbf{A}$, $N = T - p$, and $\lambda \geq 0$ is called the regularization parameter or the shrinkage parameter. The ridge regression estimator $\widehat{\mathbf{\Psi}}^{\mathrm{R}}(\lambda)$ can be expressed in the closed form

$$\widehat{\mathbf{\Psi}}^{\mathrm{R}}(\lambda) = \left(\mathbf{X}^\top \mathbf{X} + N\lambda \mathbf{I} \right)^{-1} \mathbf{X}^\top \mathbf{Y}, \qquad \lambda \geq 0. \tag{10}$$

The shrinkage parameter λ for the ridge regression can be automatically determined by using the generalized cross-validation (GCV) (Golub, Heath, and Wahba 1979). The GCV selects the value of λ where the GCV score given below is minimized:

$$GCV(\lambda) = \frac{1}{N} \left\| (\mathbf{I} - \mathbf{H}(\lambda)\mathbf{Y}) \right\|_{\mathrm{F}}^2 \Bigg/ \left[\frac{1}{N} Trace(\mathbf{I} - \mathbf{H}(\lambda)) \right]^2 , \tag{11}$$

where $\mathbf{H}(\lambda) = \mathbf{X}^\top \left(\mathbf{X}^\top \mathbf{X} + N\lambda \mathbf{I} \right)^{-1} \mathbf{X}^\top$.

In this package, the interface to the shrinkage estimation methods is provided by the function `VARshrink(method = "ridge", ...)`. If the input argument `lambda` is set at a value or a vector of values, then corresponding GCV score is computed automatically for each λ value, and the VAR coefficients with the smallest GCV score is selected. If `lambda = NULL`, then the default value of `c(0.0001, 0.0005, 0.001, 0.005, ..., 10, 50)` is used.

For example, simulated time series data of length $T = 100$ were generated based on a multivariate normal distribution for noise and a VAR model with $p = 1$, $K = 2$, $\mathbf{A}_1 = 0.5\mathbf{I}_2$, $\mathbf{C} = (0.2, 0.7)^\top$, and $\mathbf{\Sigma} = 0.1^2\mathbf{I}_2$ as follows:

```
R> set.seed(1000)
R> myCoef <- list(A = list(matrix(c(0.5, 0, 0, 0.5), 2, 2)), c = c(0.2, 0.7))
R> myModel <- list(Coef = myCoef, Sigma = diag(0.1^2, 2), dof = Inf)
R> Y <- simVARmodel(numT = 100, model = myModel, burnin = 10)
R> resu_estim <- list()
```

Then, the multivariate ridge regression can be carried out for VAR models as follows. The result printed on the screen shows all the `lambda` values considered and the corresponding GCV values. The VAR parameters are estimated using the lambda value with the minimum GCV value.

```
R> resu_estim$`Ridge regression` <-
R+   VARshrink(Y, p = 1, type = "const", method = "ridge", lambda = NULL)
R> resu_estim$`Ridge regression`


VAR Shrinkage Estimation Results:
=================================

Estimated coefficients for equation y1:
=======================================
Call:
y1 = y1.l1 + y2.l1 + const

     y1.l1      y2.l1       const
0.61424699 0.07318362 0.05264740


Estimated coefficients for equation y2:
```

```
=======================================
Call:
y2 = y1.l1 + y2.l1 + const


    y1.l1      y2.l1       const
0.1195550 0.3254618 0.9055450



lambda: 1e-04 5e-04 0.001 0.005 0.01 0.05 0.1 0.5 1 5 10 50 (estimated: TRUE)
GCV:   0.0187 0.0186 0.0186 0.0192 0.02 0.0227 0.025 0.0647 0.1506 0.8406 1.2754 1.9329
```

The method `summary()` is available for objects of class "varshrinkest" as follows:

```
R> summary(resu_estim$`Ridge regression`)
```

```
VAR Shrinkage Estimation Results:
=================================
Endogenous variables: y1, y2
Deterministic variables: const
Sample size: 99
Log Likelihood: 188.793
Roots of the characteristic polynomial:
0.6419 0.2978
Call:
VARshrink(y = Y, p = 1, type = "const", method = "ridge", lambda = NULL)



Estimation results for equation y1:
===================================
y1 = y1.l1 + y2.l1 + const

      Estimate Std. Error t value Pr(>|t|)
y1.l1  0.61425    0.07636   8.045 2.27e-12 ***
y2.l1  0.07318    0.07549   0.969    0.335
const  0.05265    0.10888   0.484    0.630
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.08913 on 96.16244 degrees of freedom
Multiple R-Squared: 0.3864, Adjusted R-squared: 0.3747
F-statistic: 32.95 on 1.837559 and 96.16244 DF,  p-value: 4.725e-11


Estimation results for equation y2:
===================================
```

```
y2 = y1.l1 + y2.l1 + const

      Estimate Std. Error t value Pr(>|t|)
y1.l1  0.11955    0.08632   1.385 0.169249
y2.l1  0.32546    0.08534   3.814 0.000242 ***
const  0.90555    0.12308   7.357  6.3e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.1008 on 96.16244 degrees of freedom
Multiple R-Squared: 0.1219, Adjusted R-squared: 0.1051
F-statistic: 7.263 on 1.837559 and 96.16244 DF,  p-value: 0.00157



Scale matrix, Sigma, of multivariate t distribution for noise:
          y1          y2
y1  0.0077160 -0.0006829
y2 -0.0006829  0.0098611

Degrees of freedom of multivariate t distribution for noise:
[1] Inf

Correlation matrix of Sigma:
         y1         y2
y1  1.00000 -0.07829
y2 -0.07829  1.00000
```

3.2. Nonparametric Shrinkage (NS)

The nonparametric shrinkage (NS) estimation method for VAR models, proposed by Opgen-Rhein and Strimmer (2007b), produces an estimate of $\boldsymbol{\Psi}$ based on a James-Stein type shrinkage of sample covariance matrices (Opgen-Rhein and Strimmer 2007a; Schäfer and Strimmer 2005).

We will briefly describe the NS method. In the NS method, the data matrices $\mathbf{X}$ and $\mathbf{Y}$ are mean-corrected so that each column has the mean of zero. Let $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}] \in \mathbb{R}^{N \times (Kp+K+L)}$ be a combined data matrix. The sample covariance matrix of $\mathbf{Z}$ is partitioned as

$$\mathbf{S}_{ZZ} = \frac{1}{N-1}\mathbf{Z}^{\top}\mathbf{Z} = \begin{bmatrix} \mathbf{S}_{XX} & \mathbf{S}_{XY} \\ \mathbf{S}_{XY}^{\top} & \mathbf{S}_{YY} \end{bmatrix} \in \mathbb{R}^{(Kp+K+L) \times (Kp+K+L)}, \tag{12}$$

where $\mathbf{S}_{XX} = (N-1)^{-1}\mathbf{X}^{\top}\mathbf{X}$, $\mathbf{S}_{XY} = (N-1)^{-1}\mathbf{X}^{\top}\mathbf{Y}$, and $\mathbf{S}_{YY} = (N-1)^{-1}\mathbf{Y}^{\top}\mathbf{Y}$. The matrix $\mathbf{S}_{ZZ}$ can be decomposed as

$$\mathbf{S}_{ZZ} = \mathbf{D}_{Z}^{1/2}\mathbf{R}_{ZZ}\mathbf{D}_{Z}^{1/2}, \tag{13}$$

where $\mathbf{R}_{\mathrm{ZZ}}$ is the sample correlation matrix and $\mathbf{D}_{\mathrm{Z}} = \mathrm{diag}(s_{11}, s_{22}, \ldots, s_{Kp+K+L,Kp+K+L})$ is a diagonal matrix with diagonal elements of sample variances. Shrinkage estimates of the correlation matrix and the variances can be written as

$$\widehat{\mathbf{R}}_{\mathrm{ZZ}} = (1 - \lambda)\mathbf{R}_{\mathrm{ZZ}} + \lambda\mathbf{I} \quad \text{and} \quad \widehat{\mathbf{D}}_{\mathrm{Z}} = \mathrm{diag}(\hat{s}_{11}, \hat{s}_{22}, \ldots, \hat{s}_{Kp+K+L,Kp+K+L}) \qquad (14)$$

with

$$\hat{s}_{ii} = (1 - \lambda_v)s_{ii} + \lambda_v s_{\mathrm{med}}, \quad i = 1, 2, \ldots, Kp + K + L, \qquad (15)$$

where s_{med} is a median of all the sample variances s_{ii}, and $0 \leq \lambda, \lambda_v \leq 1$ are shrinkage parameters. The estimated covariance matrix can be computed by

$$\widehat{\mathbf{S}}_{\mathrm{ZZ}}(\lambda, \lambda_v) = \widehat{\mathbf{D}}_{\mathrm{Z}}^{1/2}\widehat{\mathbf{R}}_{\mathrm{ZZ}}\widehat{\mathbf{D}}_{\mathrm{Z}}^{1/2}. \qquad (16)$$

The values of the shrinkage parameters λ and λ_v are determined by the James-Stein type shrinkage method, which we call the NS method described in (Opgen-Rhein and Strimmer 2007a; Schäfer and Strimmer 2005).

The ordinary least squares estimate $\widehat{\mathbf{\Psi}}^{\mathrm{OLS}}$ of $\mathbf{\Psi}$ is given by $\widehat{\mathbf{\Psi}}^{\mathrm{OLS}} = \mathbf{S}_{\mathrm{XX}}^{-1}\mathbf{S}_{\mathrm{XY}}$. We define the NS estimate of $\mathbf{\Psi}$ as

$$\widehat{\mathbf{\Psi}}^{\mathrm{N}}(\lambda, \lambda_v) = \widehat{\mathbf{S}}_{\mathrm{XX}}^{-1}\widehat{\mathbf{S}}_{\mathrm{XY}}, \qquad 0 \leq \lambda, \lambda_v \leq 1. \qquad (17)$$

where $\widehat{\mathbf{S}}_{\mathrm{XX}}$ and $\widehat{\mathbf{S}}_{\mathrm{XY}}$ are parts of the estimated covariance matrix,

$$\widehat{\mathbf{S}}_{\mathrm{ZZ}}(\lambda, \lambda_v) = \begin{bmatrix} \widehat{\mathbf{S}}_{\mathrm{XX}} & \widehat{\mathbf{S}}_{\mathrm{XY}} \\ \widehat{\mathbf{S}}_{\mathrm{XY}}^{\top} & \widehat{\mathbf{S}}_{\mathrm{YY}} \end{bmatrix}. \qquad (18)$$

In the package **VARshrink**, the function `VARshrink(method = "ns", ...)` provides an interface with the NS method. In specific, the package **corpcor** (Schäfer *et al.* 2017) includes the R function `cov.shrink()`, which can determine λ and λ_v and estimate the covariance matrix $\widehat{\mathbf{S}}_{\mathrm{ZZ}}(\lambda, \lambda_v)$. The function `VARshrink()` in the **VARshrink** package infers the NS estimates of VAR coefficients, $\widehat{\mathbf{\Psi}}^{\mathrm{N}}(\lambda, \lambda_v)$, by using the covariance matrix $\widehat{\mathbf{S}}_{\mathrm{ZZ}}(\lambda, \lambda_v)$. If the input arguments `lambda` and `lambda_var` are set as `lambda = NULL` and `lambda_var = NULL`, then λ and λ_v are determined automatically. We can find the estimated λ and λ_v values on the printed screen. For example,

```
R> resu_estim$`Nonparametric shrinkage` <-
R+   VARshrink(Y, p = 1, type = "const", method = "ns",
R+                     lambda = NULL, lambda_var = NULL)
R> resu_estim$`Nonparametric shrinkage`


VAR Shrinkage Estimation Results:
=================================

Estimated coefficients for equation y1:
=======================================
Call:
y1 = y1.l1 + y2.l1 + const
```

```
    y1.l1      y2.l1       const
0.56430137 0.06139375 0.08916350


Estimated coefficients for equation y2:
=======================================
Call:
y2 = y1.l1 + y2.l1 + const

    y1.l1      y2.l1      const
0.1080570 0.2537706 1.0116241


lambda: 0.104881997093098 (estimated: TRUE)
lambda_var: 1 (estimated: TRUE)
```

The `summary()` shows statistical inference on the estimated VAR coefficients together with the estimated scale matrix Σ of multivariate t-distribution for noise.

```
R> summary(resu_estim$`Nonparametric shrinkage`)


VAR Shrinkage Estimation Results:
=================================
Endogenous variables: y1, y2
Deterministic variables: const
Sample size: 99
Log Likelihood: 188.443
Roots of the characteristic polynomial:
0.5844 0.2337
Call:
VARshrink(y = Y, p = 1, type = "const", method = "ns", lambda = NULL,
    lambda_var = NULL)


Estimation results for equation y1:
===================================
y1 = y1.l1 + y2.l1 + const

      Estimate Std. Error t value Pr(>|t|)
y1.l1 0.564301   0.007794  72.406  < 2e-16 ***
y2.l1 0.061394   0.007352   8.350 4.57e-13 ***
const 0.089164   0.088732   1.005    0.317
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.08873 on 97.81547 degrees of freedom
Multiple R-Squared: 0.3446, Adjusted R-squared: 0.3434
F-statistic: 278.7 on 0.1845331 and 97.81547 DF,  p-value: 7.636e-12
```

```
Estimation results for equation y2:
====================================
y2 = y1.l1 + y2.l1 + const

      Estimate Std. Error t value Pr(>|t|)
y1.l1 0.108057   0.008769   12.32   <2e-16 ***
y2.l1 0.253771   0.008273   30.68   <2e-16 ***
const 1.011624   0.099839   10.13   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.09984 on 97.81547 degrees of freedom
Multiple R-Squared: 0.08139,    Adjusted R-squared: 0.07965
F-statistic: 46.96 on 0.1845331 and 97.81547 DF,  p-value: 0.0003688
```

```
Scale matrix, Sigma, of multivariate t distribution for noise:
          y1         y2
y1  0.0077791 -0.0006733
y2 -0.0006733  0.0098486

Degrees of freedom of multivariate t distribution for noise:
[1] Inf

Correlation matrix of Sigma:
        y1        y2
y1  1.00000 -0.07693
y2 -0.07693  1.00000
```

3.3. Full Bayesian Shrinkage

Ni and Sun (2005) and Sun and Ni (2004) studied Bayesian estimation methods using noninformative priors for VAR models, where they used Markov Chain Monte Carlo (MCMC) methods for estimating coefficient matrices, noise covariance matrices, and other hyperparameters in Bayesian VAR models. In Ni and Sun (2005), various Bayesian estimators were compared using several types of loss functions, noninformative priors, and multivariate normal and t-distributions. Among them, Bayesian estimators using a certain type of noninformative priors showed higher accuracies than the other Bayesian estimators in simulated experiments.

The noninformative prior for coefficient matrices was called the shrinkage prior, and the prior for the noise covariance matrix was called the reference prior.

In the package **VARshrink**, we can obtain Bayesian estimators using the shrinkage prior and the reference prior, which are the noninformative priors that yielded the highest accuracies in the simulation results (Ni and Sun 2005). As a Bayesian estimator of VAR parameters, the minimizer of the posterior expectation of quadratic loss is computed, which is the mean of the posterior distribution. Additionally, the minimizer of the posterior expectation of LINEX loss is also computed (Ni and Sun 2005; Zellner 1986). In this section, we will explain the model assumptions in more detail.

First, the noise vectors are independent and identically distributed from multivariate t-distributions with the degree of freedom ν, i.e., $\boldsymbol{\epsilon}_t \sim t_\nu(\mathbf{0}, \boldsymbol{\Sigma})$. It can be expressed as a hierarchical model as

$$(\boldsymbol{\epsilon}_t | q_t) \sim \mathrm{N}_K(\mathbf{0}, q_t^{-1}\boldsymbol{\Sigma}).$$
$$q_t \sim \mathrm{Gamma}(\nu/2, \nu/2). \tag{19}$$

Second, we denote the vectorized version of $\boldsymbol{\Psi} \in \mathbb{R}^{(J/K) \times K}$ by $\boldsymbol{\psi} = \mathrm{vec}(\boldsymbol{\Psi}) \in \mathbb{R}^J$. Here $J = K(Kp + L)$. The shrinkage prior $\pi_{\mathrm{S}}(\boldsymbol{\psi})$ for $\boldsymbol{\psi} \in \mathbb{R}^J$ is taken as $\pi_{\mathrm{S}}(\boldsymbol{\psi}) \propto \|\boldsymbol{\psi}\|^{-(J-2)}$. By introducing a latent variable $\lambda > 0$, the shrinkage prior can also be expressed as a scale mixture of multivariate normals as[1]

$$(\boldsymbol{\psi} | \lambda) \sim \mathrm{N}_J(\mathbf{0}, \lambda^{-1}\mathbf{I}_J),$$
$$\pi(\lambda) \propto 1. \tag{20}$$

Third, the reference prior $\pi_{\mathrm{R}}(\boldsymbol{\Sigma})$ for $\boldsymbol{\Sigma}$ is taken as

$$\pi_{\mathrm{R}}(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1} \prod_{1 \leq i \leq j \leq K} (\lambda_i - \lambda_j)^{-1}, \tag{21}$$

where $\lambda_1 > \lambda_2 > \cdots > \lambda_K$ are eigenvalues of $\boldsymbol{\Sigma}$.

Note that no hyperparameters are involved in the shrinkage prior and the reference prior, since they are noninformative priors. The Gibbs MCMC method makes use of the hierarchical expression of $\boldsymbol{\epsilon}_t$. The Gibbs MCMC samples the parameters $(\boldsymbol{\psi}, \lambda, \boldsymbol{\Sigma}, \mathbf{Q}, \nu)$ with $\mathbf{Q} = \mathrm{diag}(q_{p+1}, \ldots, q_T)$ from conditional posterior distributions (Ni and Sun 2005). We remark that the mean of the conditional posterior distribution, $\pi(\boldsymbol{\psi} | \lambda, \boldsymbol{\Sigma}, \mathbf{Q}, \nu; \mathbf{Y})$ of $\boldsymbol{\psi}$ is given by

$$\widehat{\boldsymbol{\psi}}^F(\lambda) = \left[\left(\boldsymbol{\Sigma}^{-1} \otimes \left(\mathbf{X}^\top \mathbf{Q} \mathbf{X} \right) \right) + \lambda \mathbf{I}_J \right]^{-1} \mathrm{vec}\left(\mathbf{X}^\top \mathbf{Q} \mathbf{Y} \boldsymbol{\Sigma}^{-1} \right), \qquad \lambda > 0. \tag{22}$$

Note that if $\boldsymbol{\Sigma} = \mathbf{I}_K$ and $q_{p+1} = \cdots = q_T = 1$, then the estimator becomes the ridge regression estimator. That is, Bayesian shrinkage estimators have more flexible and abundant expressions, even though more computational effort is required to estimate more parameters.

In the package **VARshrink**, the function VARshrink(method = "fbayes", ...) plays the role as an interface with the full Bayesian shrinkage method with the shrinkage prior and the reference prior. The shrinkage parameter λ is not set at a fixed value, i.e., the arguments lambda and lambda_var are of no use here. Instead, we can estimate the mean of the posterior distribution, $\hat{\lambda}^{-1} = \mathbb{E}\left[\lambda^{-1} | \mathbf{Y}\right]$ during the MCMC process (Ni and Sun 2005).

There are several arguments to be specified as follows.

---

[1] Ni and Sun (2005) used the letter δ to denote $\delta = \lambda^{-1}$.

- `dof`: If `dof = Inf`, then a multivariate normal distribution is applied and and the weight **Q** is not estimated. If `dof` is a finite value, then we apply the multivariate t-distribution with a fixed degree of freedom $\nu = $ `dof` and estimate **Q**. If `dof = NULL`, we estimate both the degree of freedom ν and the weight **Q**. In this case, the package **ars** is required for sampling the parameter ν from its conditional posterior distribution.

- `burnincycle, mcmccycle`: The Gibbs MCMC method samples a series of parameter values of the length, `burnincycle + mcmccycle`. `burnincycle` is the number of sampled parameter values to be discarded in the beginning, and `mcmccycle` is the number to be attained and used for computing the parameter estimates. By default, we set `burnincycle = 1000` and `mcmccycle = 2000`.

For example, we run the full Bayesian shrinkage method with a fixed $\nu = 6$ as follows.

```
R> resu_estim$`Full Bayes (fixed dof)` <-
R+   VARshrink(Y, p = 1, type = "const", method = "fbayes", dof = 6,
R+                       burnincycle = 1000, mcmccycle = 2000)
R> resu_estim$`Full Bayes (fixed dof)`
```

```
VAR Shrinkage Estimation Results:
=================================


Estimated coefficients for equation y1:
=======================================
Call:
y1 = y1.l1 + y2.l1 + const


     y1.l1       y2.l1        const
0.63856333 0.05398115 0.07042395



Estimated coefficients for equation y2:
=======================================
Call:
y2 = y1.l1 + y2.l1 + const


    y1.l1       y2.l1       const
0.1028193 0.3133995 0.9326989



Sigma for noise:
              [,1]            [,2]
[1,]   0.0073772943 -0.0002864258
[2,]  -0.0002864258  0.0079703260
dof for noise: 6 (estimated: FALSE)
lambda: 1.46381861993198 (estimated: TRUE)
```

By using `summary()`, we can find the estimated scale matrix Σ of multivariate t-distribution for noise.

```r
R> summary(resu_estim$`Full Bayes (fixed dof)`)
```

```
VAR Shrinkage Estimation Results:
=================================
Endogenous variables: y1, y2
Deterministic variables: const
Sample size: 99
Log Likelihood: 184.085
Roots of the characteristic polynomial:
0.6548 0.2971
Call:
VARshrink(y = Y, p = 1, type = "const", method = "fbayes", dof = 6,
    burnincycle = 1000, mcmccycle = 2000)


Estimation results for equation y1:
===================================
y1 = y1.l1 + y2.l1 + const

      Estimate Std. Error t value Pr(>|t|)
y1.l1  0.63856    0.07898   8.085 1.87e-12 ***
y2.l1  0.05398    0.08314   0.649    0.518
const  0.07042    0.12042   0.585    0.560
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.08917 on 96.04011 degrees of freedom
Multiple R-Squared: 0.4031, Adjusted R-squared: 0.391
F-statistic:  33.1 on 1.959891 and 96.04011 DF,  p-value: 1.608e-11


Estimation results for equation y2:
===================================
y2 = y1.l1 + y2.l1 + const

      Estimate Std. Error t value Pr(>|t|)
y1.l1   0.1028     0.0893   1.151  0.25242
y2.l1   0.3134     0.0940   3.334  0.00122 **
const   0.9327     0.1362   6.850 7.02e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.1008 on 96.04011 degrees of freedom
Multiple R-Squared: 0.1111,  Adjusted R-squared: 0.09297
F-statistic: 6.125 on 1.959891 and 96.04011 DF,   p-value: 0.003331
```

```
Scale matrix, Sigma, of multivariate t distribution for noise:
           [,1]        [,2]
[1,]   0.0073773 -0.0002864
[2,]  -0.0002864  0.0079703
```

```
Degrees of freedom of multivariate t distribution for noise:
[1] 6
```

```
Correlation matrix of Sigma:
          [,1]      [,2]
[1,]   1.00000 -0.03735
[2,]  -0.03735  1.00000
```

Instead, we can estimate the degree of freedom, ν, of multivariate t-distribution by the argument `dof = NULL` as follows:

```
R> resu_estim$`Full Bayes (estim dof)` <-
R+   VARshrink(Y, p = 1, type = "const", method = "fbayes", dof = NULL,
R+             burnincycle = 1000, mcmccycle = 2000)
R> resu_estim$`Full Bayes (estim dof)`
```

```
VAR Shrinkage Estimation Results:
=================================
```

```
Estimated coefficients for equation y1:
=======================================
Call:
y1 = y1.l1 + y2.l1 + const
```

```
     y1.l1      y2.l1       const
0.63942247 0.05771602 0.06449797
```

```
Estimated coefficients for equation y2:
=======================================
Call:
y2 = y1.l1 + y2.l1 + const
```

```
    y1.l1      y2.l1       const
```

```
0.1024030 0.3152221 0.9299175


Sigma for noise:
              [,1]           [,2]
[1,]   0.0069592834  -0.0001524324
[2,]  -0.0001524324   0.0079022459
dof for noise: 8.21302751071667 (estimated: TRUE)
lambda: 1.41018436005034 (estimated: TRUE)
```

3.4. Semiparametric Bayesian Shrinkage

Whereas full Bayesian shrinkage methods estimate all the hyperparameters including latent variables via MCMC methods, semiparametric Bayesian methods estimate some of the hyperparameters by a certain nonparametric method and estimate the rest in a Bayesian framework. The semiparametric approach is advantageous especially when the dimensionality of the model is so high that standard MCMC methods are not computationally tractable.

Lee *et al.* (2016) assumed scale mixtures of multivariate normal distributions for noise vectors as in (2) or (19). The prior distribution for the model coefficients, $\boldsymbol{\psi} = \mathrm{vec}(\boldsymbol{\Psi})$, was set as multivariate normal distributions, similarly to (20). Here, we can choose either the conjugate prior (CJ) and non-conjugate (NCJ) prior as follows:

(i) The conjugate prior for $\boldsymbol{\psi} \in \mathbb{R}^J$ is expressed by

$$(\boldsymbol{\psi}|\lambda, \boldsymbol{\Sigma}) \sim \mathrm{N}_J\left(\mathbf{0}, \frac{1-\lambda}{(N-1)\lambda}\boldsymbol{\Sigma} \otimes \mathbf{I}\right), \qquad 0 < \lambda < 1. \tag{23}$$

(ii) The non-conjugate prior for $\boldsymbol{\psi} \in \mathbb{R}^J$ is expressed by

$$(\boldsymbol{\psi}|\lambda) \sim \mathrm{N}_J\left(\mathbf{0}, \frac{1-\lambda}{(N-1)\lambda}\mathbf{I}_J\right), \qquad 0 < \lambda < 1. \tag{24}$$

The scale matrix $\boldsymbol{\Sigma}$ for noise is included in the equation for conjugate prior distribution but not for the non-conjugate prior distribution. The non-conjugate prior is quite similar to the full Bayesian formulation in (20). However, the main difference is that, in the semiparametric Bayesian approach, the shrinkage parameter λ should be estimated explicitly via a nonparametric method, but in the full Bayes approach, it is a latent variable which should be sampled and estimated implicitly via an MCMC method.

The prior distribution for $\boldsymbol{\Sigma}$ was set as an inverse Wishart distribution as

$$(\boldsymbol{\Sigma}|\mathbf{L}_0, m_0) \sim \mathrm{InvWishart}(\mathbf{L}_0, m_0), \qquad \mathbf{L}_0 \succ \mathbf{0}, m_0 > K - 1. \tag{25}$$

where $\mathbf{L}_0 \succ \mathbf{0}$ means that $\mathbf{L}_0$ is positive definite.

Once the shrinkage parameter λ is set at a fixed value, the other parameters, $\boldsymbol{\psi}, \boldsymbol{\Sigma}$, and $\mathbf{Q}$ can be estimated iteratively in a Bayesian framework efficiently. Briefly speaking, we consider estimating the parameters $\boldsymbol{\psi}$ and $\boldsymbol{\Sigma}$ at the maximum point (i.e., the mode) of the marginal

posterior density function $\pi(\boldsymbol{\psi}, \boldsymbol{\Sigma}|\lambda; \mathbf{Y})$. In the case of the non-conjugate prior, the mode, $(\widehat{\boldsymbol{\psi}}^{S}(\lambda), \widehat{\boldsymbol{\Sigma}}^{S}(\lambda))$, is expressed as

$$\widehat{\boldsymbol{\psi}}^{S}(\lambda) = \left[\left((\widehat{\boldsymbol{\Sigma}}^{S})^{-1} \otimes \left(\mathbf{X}^{\top} \widehat{\mathbf{Q}}^{S} \mathbf{X} \right) \right) + \frac{(N-1)\lambda}{1-\lambda} \mathbf{I}_{J} \right]^{-1} \mathrm{vec} \left(\mathbf{X}^{\top} \widehat{\mathbf{Q}}^{S} \mathbf{Y} (\widehat{\boldsymbol{\Sigma}}^{S})^{-1} \right), \qquad (26)$$

and

$$\widehat{\boldsymbol{\Sigma}}^{S}(\lambda) = \frac{1}{m_0 + T + K + 1} \left(\mathbf{L}_0 + \mathbf{Y}^{\top} \widehat{\mathbf{Q}} \mathbf{Y} - \mathbf{Y}^{\top} \widehat{\mathbf{Q}} \mathbf{X} \widehat{\boldsymbol{\Psi}}^{S}(\lambda) \right), \qquad (27)$$

for $0 < \lambda < 1$, where $\widehat{\mathbf{Q}}^{S} = \mathrm{diag}(\hat{q}_{p+1}, \ldots, \hat{q}_T)$ is obtained in an iterative manner.[2] The values $\widehat{\mathbf{Q}}^{S}$ is also expressed as

$$\hat{q}_t = h \left(\left\| (\widehat{\boldsymbol{\Sigma}}^{S})^{-1/2} \hat{\boldsymbol{\epsilon}}_t \right\|^2 \right), \qquad t = p + 1, \ldots, T, \qquad (28)$$

where $h(x)$ is defined in (5) depending on the noise distribution and $\hat{\boldsymbol{\epsilon}}_t$ is the residual given by $\hat{\boldsymbol{\epsilon}}_t = \mathbf{y}_t - \widehat{\boldsymbol{\Psi}}^{S\top} \mathbf{x}_t$ (Lee *et al.* 2016).

The shrinkage parameter λ is determined via an internal nonparametric method, which is called the parameterized cross validation (PCV). This algorithm can be considered as a modified K-fold cross validation, especially for estimating the shrinkage parameter of VAR models; see Lee *et al.* (2016) for a detailed explanation.

In addition, the semiparametric Bayesian shrinkage method adopts the idea of shrinking both the correlations and variances from the NS method. As a result, there are two shrinkage parameters λ and λ_v, where $0 \leq \lambda \leq 1$ is used for the shrinkage estimation of the VAR coefficient matrix Ψ and noise covariance matrix $\boldsymbol{\Sigma}$ while $0 \leq \lambda_v \leq 1$ is used for the shrinkage estimation of the variances of the variables $y_{tj}, j = 1, \ldots, K$.

In the package **VARshrink**, the function `VARshrink(method = "sbayes", ...)` is for the semiparametric Bayesian shrinkage method. There are several input arguments to these functions as follows:

- `dof`: If `dof = Inf`, we use a multivariate normal distribution for noise vectors. Otherwise, `dof` can be set as a finite number ν between 0 and ∞ for the multivariate t-distribution with the degree of freedom ν. If `dof = NULL`, then ν is automatically selected by the K-fold cross validation, with $K =$ `num_fold`. `dof = Inf` by default.
- `lambda, lambda_var`: `lambda = NULL` and `lambda_var = NULL` implies that the shrinkage parameters λ and λ_v are estimated automatically with $0 \leq \lambda, \lambda_v \leq 1$.
- `prior_type`: Either `prior_type = "NCJ"` or `prior_type = "CJ"`, which implies non-conjugate prior or conjugate prior. The default value is "NCJ".
- `num_folds`: The number of folds for the parameterized cross validation method for determining λ value. `num_folds = 5` by default. It works only when `lambda = NULL` or `dof = NULL`.
- `m0`: The value of the hyperparameter m_0 of the inverse Wishart prior of $\boldsymbol{\Sigma}$. `m0 = ` K by default. On the other hand, the other hyperparameter $\mathbf{L}_0$ of the inverse Wishart prior is set by $\mathbf{L}_0 = (m_0 + K + 1)\mathbf{I}_K$.

---

[2]We note that the expression in (26) is equivalent to that in (22) in a full Bayesian framework.

For example, we can run the semiparametric shrinkage method as follows.

```
R> resu_estim$`Semi Bayes (fixed dof)` <-
R+    VARshrink(Y, p = 1, type = "const", method = "sbayes", dof = 6,
R+             lambda = NULL, lambda_var = NULL, prior_type = "NCJ",
R+             num_folds = 5, m0 = ncol(Y))
R> resu_estim$`Semi Bayes (fixed dof)`


VAR Shrinkage Estimation Results:
=================================

Estimated coefficients for equation y1:
=======================================
Call:
y1 = y1.l1 + y2.l1 + const

     y1.l1      y2.l1      const
0.52361471 0.05712005 0.07157308


Estimated coefficients for equation y2:
=======================================
Call:
y2 = y1.l1 + y2.l1 + const

     y1.l1      y2.l1      const
0.09327854 0.36697792 0.97005425



Sigma for noise:
              y1            y2
y1  0.0065088952 -0.0003156327
y2 -0.0003156327  0.0106188219
dof for noise: 6 (estimated: FALSE)
lambda: 0.00318699518988467 (estimated: TRUE)
lambda_var: 0.940206098485031 (estimated: TRUE)

R> summary(resu_estim$`Semi Bayes (fixed dof)`)


VAR Shrinkage Estimation Results:
=================================
Endogenous variables: y1, y2
Deterministic variables: const
Sample size: 99
Log Likelihood: 127.887
```

```
Roots of the characteristic polynomial:
0.5524 0.3382
Call:
VARshrink(y = Y, p = 1, type = "const", method = "sbayes", lambda = NULL,
    lambda_var = NULL, dof = 6, prior_type = "NCJ", num_folds = 5,
    m0 = ncol(Y))


Estimation results for equation y1:
===================================
y1 = y1.l1 + y2.l1 + const


      Estimate Std. Error t value Pr(>|t|)
y1.l1  0.52361    0.08491   6.167 1.65e-08 ***
y2.l1  0.05712    0.09226   0.619    0.537
const  0.07157    0.13375   0.535    0.594
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.09877 on 96.0093 degrees of freedom
Multiple R-Squared: 0.2713, Adjusted R-squared: 0.2561
F-statistic: 17.95 on 1.990699 and 96.0093 DF,  p-value: 2.492e-07


Estimation results for equation y2:
===================================
y2 = y1.l1 + y2.l1 + const


      Estimate Std. Error t value Pr(>|t|)
y1.l1  0.09328    0.13092   0.713   0.4779
y2.l1  0.36698    0.14224   2.580   0.0114 *
const  0.97005    0.20622   4.704 8.55e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.1523 on 96.0093 degrees of freedom
Multiple R-Squared: 0.06678,    Adjusted R-squared: 0.04743
F-statistic: 3.451 on 1.990699 and 96.0093 DF,  p-value: 0.03589


Scale matrix, Sigma, of multivariate t distribution for noise:
          y1            y2
y1  0.0065089 -0.0003156
y2 -0.0003156  0.0106188
```

```
Degrees of freedom of multivariate t distribution for noise:
[1] 6

Correlation matrix of Sigma:
          y1       y2
y1  1.00000 -0.03797
y2 -0.03797  1.00000
```

We can also let the software package to choose the degree of freedom parameter ν automatically by setting dof = NULL. In this case, the package uses simply a K-fold cross validation to find an optimal value of ν.

```
R> resu_estim$`Semi Bayes (estim dof)` <-
R+   VARshrink(Y, p = 1, type = "const", method = "sbayes", dof = NULL,
R+             lambda = NULL, lambda_var = NULL, prior_type = "NCJ",
R+             num_folds = 5, m0 = ncol(Y))
R> resu_estim$`Semi Bayes (estim dof)`


VAR Shrinkage Estimation Results:
=================================

Estimated coefficients for equation y1:
=======================================
Call:
y1 = y1.l1 + y2.l1 + const

      y1.l1        y2.l1        const
 0.50699465 -0.05884566  0.20997630


Estimated coefficients for equation y2:
=======================================
Call:
y2 = y1.l1 + y2.l1 + const

      y1.l1        y2.l1        const
0.07138759 0.37448706 0.97355673


Sigma for noise:
              y1             y2
y1  5.001873e-03 -9.978788e-06
y2 -9.978788e-06  6.991886e-03
dof for noise: 0.5 (estimated: TRUE)
```

```
lambda: 0.00145763877257461 (estimated: TRUE)
lambda_var: 0.940206098485031 (estimated: TRUE)
```

3.5. K-fold Cross Validation for Semiparametric Shrinkage

The **VARshrink** includes an implementation of the K-fold cross validation (CV) method for selecting shrinkage parameters. In specific, the K-fold CV method can select the λ and λ_v values for the semiparametric Bayesian shrinkage estimator described in Section 3.4. We recall that the semiparametric method selects a λ value by using the PCV method and selects a λ_v value by a Stein-type method.

The K-fold CV method can be run as follows. The arguments to VARshrink() are same to those for the semiparametric Bayesian shrinkage method except for the argument method = "kcv".

```
R> resu_estim$`K-fold CV (fixed dof)` <-
R+   VARshrink(Y, p = 1, type = "const", method = "kcv", dof = 6,
R+             lambda = NULL, lambda_var = NULL, prior_type = "NCJ",
R+             num_folds = 5, m0 = ncol(Y))
R> resu_estim$`K-fold CV (fixed dof)`


VAR Shrinkage Estimation Results:
=================================

Estimated coefficients for equation y1:
=======================================
Call:
y1 = y1.l1 + y2.l1 + const

     y1.l1      y2.l1      const
0.62604182 0.05632303 0.07184976


Estimated coefficients for equation y2:
=======================================
Call:
y2 = y1.l1 + y2.l1 + const

     y1.l1      y2.l1      const
0.09004368 0.29111218 0.96959392


Sigma for noise:
             y1            y2
y1  0.0075080166 -0.0003445399
y2 -0.0003445399  0.0081370205
```

Table 2: Sum of squared errors of VAR coefficients estimated by the shrinkage methods.

	SSE
Ridge regression	0.063
Nonparametric shrinkage	0.080
Full Bayes (fixed dof)	0.068
Full Bayes (estim dof)	0.067
Semi Bayes (fixed dof)	0.030
Semi Bayes (estim dof)	0.024
K-fold CV (fixed dof)	0.071
K-fold CV (estim dof)	0.062

```
dof for noise: 6 (estimated: FALSE)
lambda: 0.001 (estimated: TRUE)
lambda_var: 0.001 (estimated: TRUE)
```

Degree of freedom of multivariate t-distribution for noise can be selected automatically by setting the argument `dof = Inf` as follows.

```
R> resu_estim$`K-fold CV (estim dof)` <-
R+   VARshrink(Y, p = 1, type = "const", method = "kcv", dof = NULL,
R+            lambda = NULL, lambda_var = NULL, prior_type = "NCJ",
R+            num_folds = 5, m0 = ncol(Y))
```

After all, the function `calcSSE_Acoef()` computes the sum of squared errors (SSEs) between two VAR model parameters, $\{\mathbf{A}_k^{(1)}\}$ and $\{\mathbf{A}_k^{(2)}\}$, as $SSE = \sum_{k=1}^{p} \sum_{i,j} (\mathbf{A}_{kij}^{(1)} - \mathbf{A}_{kij}^{(2)})^2$.

```
R> resu_sse <- data.frame(SSE = sapply(resu_estim,
R+   function(x) calcSSE_Acoef(Acoef_sh(x), myCoef$A)))
```

---

4. Numerical Experiments

In this section, we apply the shrinkage estimation methods in the package **VARshrink** to a benchmark data set and simulated data sets. Using the benchmark data set, we demonstrate the use of information criteria such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for comparison of VAR models. In addition, by simulated experiments, we compute an expected value of the SSE of each shrinkage estimation method for evaluating the accuracy.

4.1. Benchmark Data

The Canada data set is a benchmark macroeconomic data included in the package **vars**. It contains four time series variables representing employment(`e`), labor productivity(`prod`), real
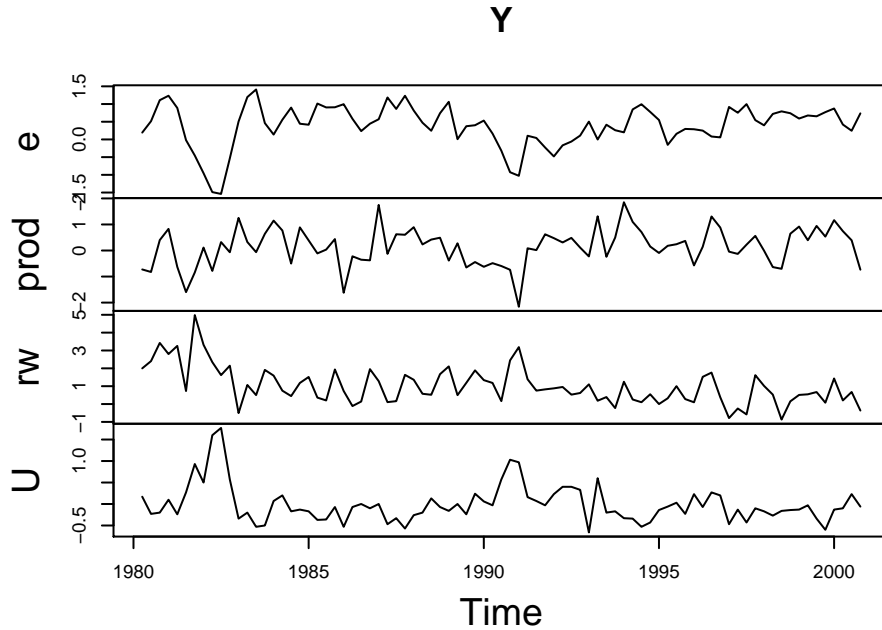
Figure 1: The benchmark data set obtained by differencing the Canada data.

wage(`rw`), and unemployment rate(`U`), with 84 observations. We took difference on the data to remove trend, yielding $T = 83$. Figure 1 shows the differencing result.

```
R> data(Canada, package = "vars")
R> Y = diff(Canada)
R> plot(Y, cex.lab = 1.3)
```

The shrinkage methods are run with the option `const = "const"` since the mean of the data has not been corrected. We set the lag order p at `p = 1, 2, 3` to compare several VAR models. For the semiparametric Bayesian method (`method = "sbayes"`), the degree of freedom ν was automatically selected. In addition, we ran the K-fold CV method (`method = "kcv"`) for choosing λ and λ_v values of the semiparametric Bayesian shrinkage estimator.

```
R> set.seed(1000)
R> resu_model <- array(NA, dim = c(5, 2, 3),
R+    dimnames = list(c("Ridge regression", "Nonparametric shrinkage",
R+                      "Full Bayes", "Semi Bayes", "K-fold CV"),
R+                c("AIC", "BIC"), c("p=1", "p=2", "p=3")))
R> for (p in 1:3) {
R+    EstimRidge <- VARshrink(Y, p = p, type = "const", method = "ridge")
R+    resu_model["Ridge regression", , p] <- c(AIC(EstimRidge), BIC(EstimRidge))
R+
R+    EstimNS <- VARshrink(Y, p = p, type = "const", method = "ns")
R+    resu_model["Nonparametric shrinkage", , p] <-
```

Table 3: Information criteria (AIC, BIC) for model comparison.

	AIC.p=1	BIC.p=1	AIC.p=2	BIC.p=2	AIC.p=3	BIC.p=3
Ridge regression	465.8	504.6	442.9	509.3	445.3	525.9
Nonparametric shrinkage	462.5	496.6	434.6	489.0	430.5	502.9
Full Bayes	458.3	500.4	447.1	514.7	450.6	532.2
Semi Bayes	502.6	549.6	483.8	567.3	496.2	607.8
K-fold CV	526.9	574.3	477.2	526.2	469.6	540.0

```
R+     c(AIC(EstimNS), BIC(EstimNS))
R+
R+   EstimFB <- VARshrink(Y, p = p, type = "const", method = "fbayes", dof = NULL)
R+   resu_model["Full Bayes", , p] <- c(AIC(EstimFB), BIC(EstimFB))
R+
R+   EstimSB <- VARshrink(Y, p = p, type = "const", method = "sbayes",
R+                   dof = NULL, prior_type = "NCJ")
R+   resu_model["Semi Bayes", , p] <- c(AIC(EstimSB), BIC(EstimSB))
R+
R+   EstimKCV <- VARshrink(Y, p = p, type = "const", method = "kcv",
R+                   dof = NULL, prior_type = "NCJ")
R+   resu_model["K-fold CV", , p] <- c(AIC(EstimKCV), BIC(EstimKCV))
R+ }
```

We can compare several models by computing their AIC and BIC. The following results in Table 3 indicate that the NS method produced better VAR coefficients than those of the other methods, and that the AIC took the minimum at $p = 3$ while the BIC took the minimum at $p = 2$.

The estimated parameters by the NS method with $p = 2$ can be analyzed further by using the methods and functions in Table 1. For example, we can perform time series forecasting as in Figure 2.

```
R> plot(predict(VARshrink(Y, p = 2, type = "const", method = "ns")), names = "U")
```

4.2. Comparative Simulation

For comparison of estimation accuracies by the shrinkage estimation methods, we generated multivariate time series data of length $T = 20, 40, 80, 160$ from $K = 20$ dimensional VAR models of order $p = 1$. We note that when $T = 20$, the sample size $N = T - p = 19$ becomes smaller than the dimensionality $K = 20$. We set the VAR coefficient matrix, $\mathbf{A}_1$, to have the diagonal entries of 0.6 and off-diagonal entries of zeros except for K entries randomly selected on the lower triangular part of the matrix. The nonzero off-diagonal entries were uniformly randomly drawn from the interval $[-1, -0.2] \cup [0.2, 1]$. Since the coefficient matrix is lower triangular, the diagonal entries determine its eigenvalues, which lead to weak stationarity (Hamilton 1994; Tsay 2005). The noise ϵ_t was randomly sampled from the multivariate normal distribution whose covariance matrix has diagonal entries of 1 and off-diagonal entries
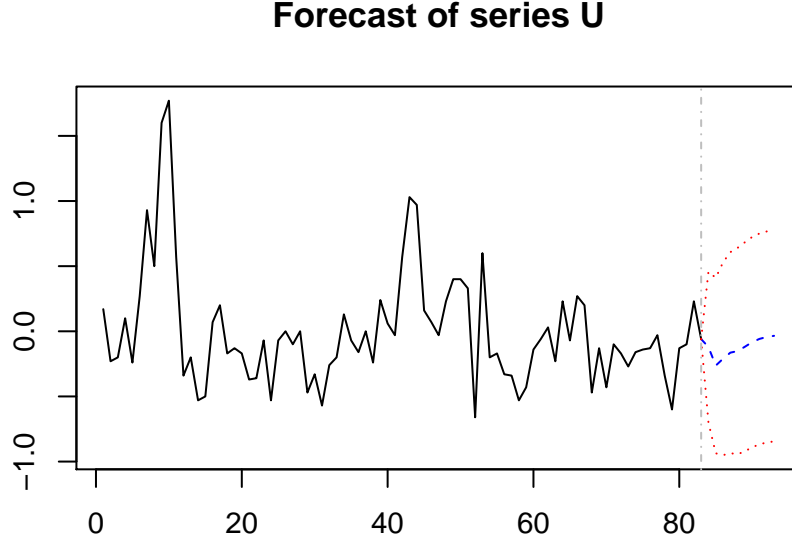
Forecast of series U

Figure 2: A 10-step ahead Forecasting of time series by the VAR model estimated by the nonparametric shrinkage method. The differenced Canada data were modeled by a VAR(2) model selected at the minimum BIC.

of 0.5. The mean of sum of squared errors (MSSE) of estimated VAR coefficients, $\widehat{\mathbf{A}}_1$, was computed by

$$\sum_{i=1}^{K}\sum_{j=1}^{K}\left(\left(\widehat{\mathbf{A}}_1\right)_{ij} - (\mathbf{A}_1)_{ij}\right)^2. \tag{29}$$

We repeated the experiments 50 times and drew boxplots to compare the SSEs of the shrinkage methods.

```
R> set.seed(1000)
R> p <- 1; K <- 20; NumTimePTS <- c(20, 40, 80, 160)
R> const_vector <- c(rep(0.2, 5), rep(0.7, 15))
R> Sig <- diag(0.5, K) + matrix(0.5, K, K)
R> resu_SSE <- array(0, dim = c(50, 4, length(NumTimePTS)), dimnames =
R+   list(1:50, c("Ridge", "NS", "FBayes", "SBayes"), NumTimePTS))
R> for (idT in 1:length(NumTimePTS)) {
R+   numT = NumTimePTS[idT]
R+   for (r in 1:50) {
R+     Ad <- createVARCoefs_ltriangular(p = p, K = K, diag_val = 0.6,
R+            num_nonzero = K, const_vector = const_vector, range_max = 1)
R+     Md <- list(Coef = Ad, Sigma = Sig, dof = Inf)
R+     Y <- simVARmodel(numT = numT, model = Md, burnin = 20)
R+     EstimRG <- VARshrink(Y, p, "const", method = "ridge")
```
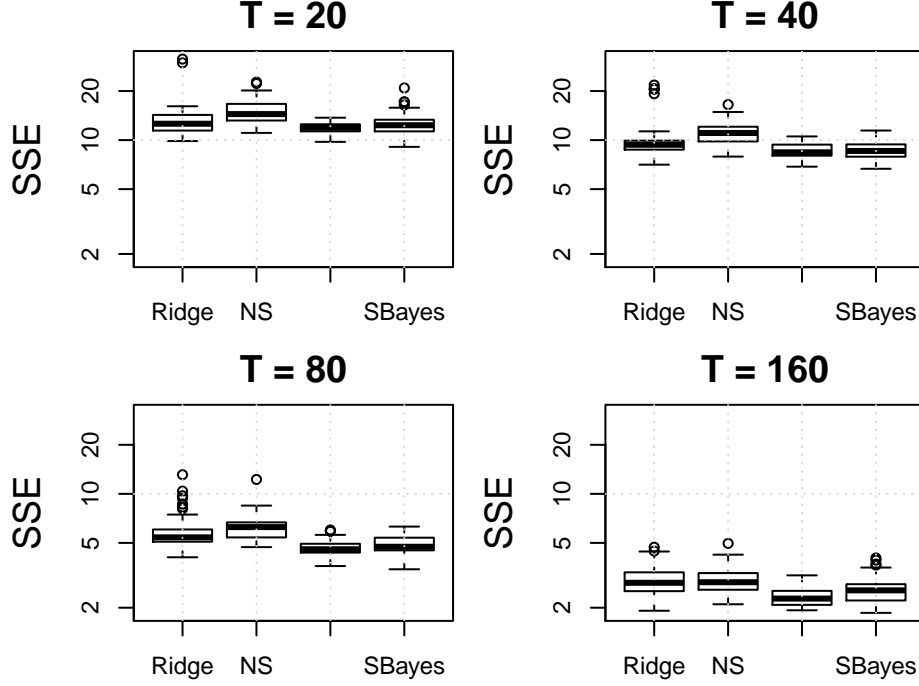
Figure 3: Comparison of the mean of sum of squared errors (MSSE) of the VAR coefficients estimated by the four shrinkage methods. The multivariate time series data were randomly generated from VAR models of order $p = 1$ and dimensionality $K = 20$, and the noise term was randomly drawn from a multivariate normal distribution.

```
R+      EstimNS <- VARshrink(Y, p, "const", method = "ns")
R+      EstimFB <- VARshrink(Y, p, "const", method = "fbayes", dof = NULL)
R+      EstimSB <- VARshrink(Y, p, "const", method = "sbayes", prior_type = "NCJ")
R+      resu_SSE[r, 1, idT] <- calcSSE_Acoef(Acoef_sh(EstimRG), Ad$A)
R+      resu_SSE[r, 2, idT] <- calcSSE_Acoef(Acoef_sh(EstimNS), Ad$A)
R+      resu_SSE[r, 3, idT] <- calcSSE_Acoef(Acoef_sh(EstimFB), Ad$A)
R+      resu_SSE[r, 4, idT] <- calcSSE_Acoef(Acoef_sh(EstimSB), Ad$A)
R+   }
R+ }

pdf
  2
```

Figure 3 shows the boxplots of the MSSEs of the VAR coefficients estimated by the four shrinkage methods. The full Bayesian approach (FBayes) and the semiparametric Bayesian approach (SBayes) yielded the best accuracies among the four methods.

---

5. Conclusions

We wrote an R software package **VARshrink** for shrinkage estimation of VAR model parameters. The shrinkage methods included in this package are the multivariate ridge regression (Hoerl and Kennard 1970; Golub *et al.* 1979), a nonparametric shrinkage method (Opgen-Rhein and Strimmer 2007b), a full Bayesian shrinkage method (Ni and Sun 2005), and a semiparametric Bayesian shrinkage method (Lee *et al.* 2016). An advantage of this package is the integration of the nonparametric, parametric, and semiparametric methods in one frame via a common interface function `VARshrink()`, which makes it easy and convenient to use various types of shrinkage estimation methods. Moreover, we note that the shrinkage estimation methods implemented in the current version have not been widely implemented in R packages in the context of VAR models.

We demonstrated the use of model selection criteria as AIC and BIC by using benchmark time series data. We note that computation of the log-likelihood of an estimated model must consider the actual distribution assumption of each method. We explained that the multivariate normal distribution is not the only choice for a distribution of noise, but another distributions such as the multivariate t-distribution can be chosen. Moreover, an effective number of parameters must be calculated for computing the AIC and BIC values. In this case, a large number of shrinkage parameter value leads to a small value of the effective number of parameters. In the package **VARshrink**, effective number of parameters is computed automatically based on the selected shrinkage parameter value.

Shrinkage methods are quite crucial especially for high dimensional VAR models. Bayesian approaches have been developed widely for VAR models in the literature for various purposes. However, the computational time for carrying out MCMC processes may be intractably high for high dimensional VAR models. For this reason, it is important to use nonparametric and semiparametric shrinkage estimation methods together to produce computationally feasible estimates of model parameters. The R package **VARshrink** is the first step to an integrative and general types of software packages for VAR models. Moreover, this package can be extended to include other shrinkage methods such as Bayesian shrinkage methods using several types of different prior distributions.

References

Bańbura M, Giannone D, Reichlin L (2010). "Large Bayesian vector auto regressions." *Journal of Applied Econometrics*, **25**(1), 71–92. `doi:10.1002/jae.1137`.

Beltrachini L, von Ellenrieder N, Muravchik CH (2013). "Shrinkage Approach for Spatiotemporal EEG Covariance Matrix Estimation." *IEEE Transactions on Signal Processing*, **61**(7), 1797–1808. `doi:10.1109/TSP.2013.2238532`.

Böhm H, von Sachs R (2009). "Shrinkage estimation in the frequency domain of multivariate time series." *Journal of Multivariate Analysis*, **100**(5), 913 – 935. `doi:10.1016/j.jmva.2008.09.009`.

Chartrand R, Yin W (2008). "Iteratively reweighted algorithms for compressive sensing." In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3869–3872. `doi:10.1109/ICASSP.2008.4518498`.

Doan T, Litterman R, Sims C (1984). "Forecasting and conditional projection using realistic prior distributions." *Econometric Reviews*, **3**(1), 1–100. `doi:10.1080/07474938408800053`.

Eltoft T, Kim T, Lee TW (2006). "On the multivariate Laplace distribution." *IEEE Signal Processing Letters*, **13**(5), 300–303. `doi:10.1109/LSP.2006.870353`.

Fiecas M, Ombao H (2011). "The generalized shrinkage estimator for the analysis of functional connectivity of brain signals." *The Annals of Applied Statistics*, **5**(2A), 1102–1125. `doi:10.1214/10-AOAS396`.

Friedman J, Hastie T, Tibshirani R, Simon N, Narasimhan B, Qian J (2018). "glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models." R Package Version 2.0-16. Available at `https://cran.r-project.org/web/packages/glmnet`.

Golub GH, Heath M, Wahba G (1979). "Generalized cross-validation as a method for choosing a good ridge parameter." *Technometrics*, **21**(2), 215–223. `doi:10.1080/00401706.1979.10489751`.

Hamilton JD (1994). *Time Series Analysis*. Princeton University Press, Princeton.

Hoerl AE, Kennard RW (1970). "Ridge regression: Biased estimation for nonorthogonal problems." *Technometrics*, **12**(1), 55–67. `doi:10.1080/00401706.1970.10488634`.

Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F, R Core Team, Ihaka R, Reid D, Shaub D, Tang Y, Zhou Z (2018). "forecast: Forecasting Functions for Time Series and Linear Models." R Package Version 8.4. Available at `https://cran.r-project.org/web/packages/forecast`.

Koop G, Korobilis D (2010). "Bayesian multivariate time series methods for empirical macroeconomics." *Foundations and Trends in Econometrics*, **3**(4), 267–358. `doi:10.1561/0800000013`.

Krueger F (2015). "bvarsv: Bayesian Analysis of a Vector Autoregressive Model with Stochastic Volatility and Time-Varying Parameters." R Package Version 1.1. Available at `https://cran.r-project.org/web/packages/bvarsv`.

Ledoit O, Wolf M (2004). "A well-conditioned estimator for large-dimensional covariance matrices." *Journal of Multivariate Analysis*, **88**(2), 365 – 411. `doi:10.1016/S0047-259X(03)00096-4`.

Lee N, Choi H, Kim SH (2016). "Bayes shrinkage estimation for high-dimensional VAR models with scale mixture of normal distributions for noise." *Computationl Statistics & Data Analysis*, **101**, 250–276. `doi:10.1016/j.csda.2016.03.007`.

Litterman RB (1986). "Forecasting with Bayesian Vector Autoregressions: Five Years of Experience." *Journal of Business & Economic Statistics*, **4**(1), 25–38. `doi:10.2307/1391384`.

Moritz S, Cule E (2018). "ridge: Ridge Regression with Automatic Selection of the Penalty Parameter." R Package Version 2.3. Available at `https://cran.r-project.org/web/packages/ridge`.

Ni S, Sun D (2005). "Bayesian Estimates for Vector Autoregressive Models." *Journal of Business & Economic Statistics*, **23**(1), 105–117. `doi:10.1198/073500104000000622`.

Opgen-Rhein R, Strimmer K (2007a). "Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach." *Statistical Applications in Genetics and Molecular Biology*, **6**(1), 9. `doi:10.2202/1544-6115.1252`.

Opgen-Rhein R, Strimmer K (2007b). "Learning causal networks from systems biology time course data: an effective model selection procedure for the vector autoregressive process." *BMC Bioinformatics*, **8**(2), S3. `doi:10.1186/1471-2105-8-S2-S3`.

Pfaff B, Stigler M (2018). "vars: VAR Modelling." R Package Version 1.5-3. Available at `https://cran.r-project.org/web/packages/vars`.

Primiceri GE (2005). "Time Varying Structural Vector Autoregressions and Monetary Policy." *The Review of Economic Studies*, **72**(3), 821–852. `doi:10.1111/j.1467-937X.2005.00353.x`.

Ripley B, Venables B, Bates DM, Hornik K, Gebhardt A, Firth D (2018). "MASS: Support Functions and Datasets for Venables and Ripley's MASS." R Package Version 7.3-51.1. Available at `https://cran.r-project.org/web/packages/MASS`.

Schäfer J, Opgen-Rhein R, Zuber V, Ahdesmäki M, Silva APD, Strimmer K (2017). "corpcor: Efficient estimation of covariance and (partial) correlation." R Package Version 1.6.9. Available at `http://CRAN.R-project.org/package=corpcor`.

Schäfer J, Strimmer K (2005). "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics." *Statistical Applications in Genetics and Molecular Biology*, **4**(1), 32. `doi:10.2202/1544-6115.1175`.

Sun D, Ni S (2004). "Bayesian analysis of vector-autoregressive models with noninformative priors." *Journal of Statistical Planning and Inference*, **121**(2), 291 – 309. `doi:10.1016/S0378-3758(03)00116-2`.

Tsay RS (2005). *Analysis of Financial Time Series*. Wiley series in probability and statistics, 2nd edition. John Wiley & Sons, Hoboken, NJ. ISBN 978-0-471-69074-0.

Tsay RS, Wood D (2018). "MTS: All-Purpose Toolkit for Analyzing Multivariate Time Series (MTS) and Estimating Multivariate Volatility Models." R Package Version 1.0. Available at `https://cran.r-project.org/web/packages/MTS`.

West M (1984). "Outlier models and prior distributions in Bayesian linear regression." *Journal of the Royal Statistical Society. Series B (Methodological)*, **46**(3), 431–439.

Zellner A (1986). "Bayesian estimation and prediction using asymmetric loss functions." *Journal of the American Statistical Association*, **81**(394), 446–451. `doi:10.1080/01621459.1986.10478289`.

Affiliation:

Namgil Lee
Kangwon National University
Department of Information Statistics, Kangwon National University, Chuncheon, Gangwon
24341, South Korea
E-mail: namgil.lee@kangwon.ac.kr
URL: https://sites.google.com/site/namgil

Sung-Ho Kim
KAIST
Department of Mathematical Sciences, KAIST, Daejeon 305-701, South Korea
E-mail: sung-ho.kim@kaist.edu