Stat

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) https://doi.org/10.1002/sta4.203

Vector autoregression and envelope model

Lei Wang[a] and Shanshan Ding[a]*

Received 15 July 2018; Accepted 16 August 2018

Vector autoregression is an important technique for modelling multivariate time series and has been widely used in a variety of applications. Owing to its fast growth of parameters with the dimension of the time series vector, dimension reduction is often desirable in multivariate time series analysis. The envelope model is a new approach to achieve dimension reduction and allows efficient estimation in multivariate analysis. In this paper, we provide the first work to explore the application and extension of envelope models to multivariate time series data. We present the envelope and partial envelope formulations for vector autoregression and elaborate model selection, parameter estimation and asymptotic results. Simulations and real data analysis demonstrate the efficiency gains of the envelope vector autoregression models compared with the standard models in terms of estimation. Meanwhile, the envelope models can excel in prediction improvement. © 2018 John Wiley & Sons, Ltd.

Keywords: dimension reduction; envelope model; partial envelope model; reducing subspace; vector autoregression

1 Introduction

Multivariate time series (MTS) data are widely available in a variety of fields including medicine, finance, science, and engineering. For example, in finance, price change in one market can easily and instantly work on another market. With economic globalization, financial markets are more dependent on each other than ever before, and one must consider them jointly to better understand the dynamic structure of the global finance. An MTS consists of multiple time series that are modelled simultaneously. The most successful, flexible, and easy-to-use model for the analysis of MTS data is the vector autoregression (VAR) model, put forward by Sims (1980). It is a natural extension of the univariate autoregressive model to dynamic MTS. Usually, a univariate autoregression is a single-equation, single-variable linear model in which the current value of a variable is explained by its own lagged values. VAR generalizes the univariate autoregressive model by allowing for more than one evolving variable. Each variable is in turn explained by its own lagged values, plus current and past values of the other variables.

For MTS, dimension reduction is desirable because the number of parameters in a model grows very fast with the dimension of the vector of time series. Therefore, finding simplifying structures or factors is beneficial for modelling MTS in order to reduce the number of parameters and gain efficiency. Some traditional dimension reduction methods for multivariate linear regression, such as principal component, canonical analysis, and reduced-rank models, have been studied and applied in MTS. Anderson (1963) first outlined the possible functions of factor analysis in time series analysis, discussed some of the problems and difficulties that arise, and pointed out limitations on its usefulness. Recently, Pan & Yao (2008) proposed a new method for estimating common factors of multiple time series that is applicable to some non-stationary time series. Box & Tiao (1977) proposed a canonical transformation of an r-dimensional stationary autoregressive process. Velu (1986) investigated reduced-rank coefficient models for multiple time series. Krawczak & Szkatula (2014) introduced a new approach referred to as symbolic essential attributes

[a]**Department of Applied Economics and Statistics, University of Delaware, Newark, DE 19716, USA.**

*__Email: sding@udel.edu__

Stat

(wileyonlinelibrary.com) https://doi.org/10.1002/sta4.203

The ISI's Journal for the Rapid Dissemination of Statistics Research

approximation to reduce the dimensionality of multidimensional time series. Matilainen et al. (2017) combined ideas from sliced inverse regression and blind source separation methods to obtain linear combinations of the explaining time series, which are ordered according to their relevance with respect to the response.

Cook et al. (2010) proposed a nascent technique called the envelope model that achieves dimension reduction and simultaneous modelling of multivariate linear regression and has the potential to produce substantial gains in efficiency in multivariate analysis. This approach is based on the construction of a link between the mean function and the covariance matrix, using the minimal reducing subspace of the latter that accommodates the former. Its motivation came from the observation that some characteristics of the response vector might be unaffected by changes in the predictors. For multivariate linear regression,

$$Y = B_0 + B_1 X + \varepsilon, \tag{1}$$

where $Y \in R^r$, the predictor vector $X \in R^p$, the coefficient matrix $B_1 \in R^{r\times p}$, and the error vector ε follows a multivariate normal distribution $N(0, \Sigma)$; the envelope model makes use of the stochastic relationships among the elements of Y and identifies a part of the response that is immaterial to changes in X. Excluding this immaterial part in the estimation of B_1 can lead to substantial gains in efficiency. It has been shown that the new envelope estimator is asymptotically more efficient than the standard estimator without enveloping.

Since the seminal work by Cook et al. (2010), researchers have extended envelope models to more general and broad contexts. For example, Su & Cook (2011) introduced the partial envelope model that leads to a parsimonious method for multivariate linear regression when some of the predictors are of special interest. Cook et al. (2013) established connections between envelopes and partial least squares (PLS) and demonstrated advantages of envelope methods over PLS. Cook et al. (2015) incorporated the idea of enveloping into reduced-rank regression and demonstrated efficiency gains. Su et al. (2016) developed a sparse envelope model that performs response variable selection under the envelope model. Khare et al. (2017) proposed a comprehensive Bayesian framework for estimation and model selection in envelope models. In addition, Li & Zhang (2017) and Ding & Cook (2018) studied envelope models in matrix-variate and tensor settings.

With the popularity of the envelope method and the demand of parsimonious modelling of MTS, we provide the first work to explore the application and extension of envelope models to MTS with a target on VAR models. The new methods reduce the number of parameters in a VAR estimation and increase efficiency in both estimation and prediction. The efficiency gains sometimes are massive and equivalent to taking hundreds of additional observations. Meanwhile, as more recent time series lags might play an important role to influence the current time series, we also exert the advantage of partial envelope models to MTS that target on a portion of lag effects and demonstrate improvement.

The rest paper is organized as follows. We first review the envelope model proposed by Cook et al. (2010) in Section 2. We then formulate envelopes and partial envelopes for the VAR model in Section 3. Sections 4–6 present model selection, parameter estimation, and asymptotic results, respectively. Simulation studies for evaluating the numerical performance of the envelope VAR models are provided in Section 7. We apply the envelope VAR and partial envelope VAR models to real time series data in Section 8 and conclude the paper in Section 9.

2 Review of envelope model

The envelope model was originally proposed to achieve efficiency gains in multivariate linear regression (1). It is based on the construction of a link between the mean function and the covariance matrix, using the minimal reducing subspace of the latter that accommodates the former. This can lead to a parsimonious multivariate regression model,

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) https://doi.org/10.1002/sta4.203

or the envelope model, where the number of parameters can be reduced. In addition, the envelope estimator of the unknown coefficient matrix $B_1 \in R^{r\times p}$ in (1) has the potential to achieve massive gains in efficiency than has the standard estimator of B_1, and these gains will be passed on to other tasks like prediction.

By considering that some characteristics of the response vector could be unaffected by changes in the predictors, enveloping is to make use of the stochastic relationships among the elements of Y and identifies a part of the response that is immaterial to changes in X. More specifically, let $(\Gamma, \Gamma_0) \in R^{r\times r}$ be an orthogonal matrix. Then Y can be decomposed into two parts, $\Gamma^T Y$ and $\Gamma_0^T Y$. Assume that (i) $\Gamma_0^T Y \,|\, X \sim \Gamma_0^T Y$ and (ii) $\Gamma_0^T Y \perp \Gamma^T Y \,|\, X$, where "$\sim$" means identically distributed and "$\perp$" indicates independence. Condition (i) implies that the distribution of $\Gamma_0^T Y$ does not depend on X. Thus, $\Gamma_0^T Y$ does not carry information about B_1. Condition (ii) implies that $\Gamma_0^T Y$ does not carry any information about B_1 through its conditional correlation with $\Gamma^T Y$. These two conditions together imply that $\Gamma_0^T Y$ does not carry any information about B_1 directly or indirectly, and therefore, $\Gamma_0^T Y$ is immaterial to the regression, and only $\Gamma^T Y$ is material to the regression. We call $\Gamma^T Y$ and $\Gamma_0^T Y$ the material part and immaterial part, respectively.

Cook et al. (2010) showed that (i) and (ii) are equivalent to the following algebraic conditions: (a) $\mathcal{B} \subseteq \text{span}(\Gamma)$, where $\mathcal{B} = \text{span}(B_1)$, the column span of B_1, and (b) $\Sigma = \Sigma_1 + \Sigma_2 = P_\Gamma \Sigma P_\Gamma + Q_\Gamma \Sigma Q_\Gamma$. Here, P_Γ indicates the projection matrix onto Γ or $\text{span}(\Gamma)$, and $Q_\Gamma = I - P_\Gamma$. When (b) holds, $\text{span}(\Gamma)$ is a reducing subspace of Σ. Thus, (a) and (b) together indicate that $\text{span}(\Gamma)$ is a reducing subspace of Σ that contains $\mathcal{B}$. To achieve maximum dimension reduction, the Σ-envelope of $\mathcal{B}$, denoted by $\mathcal{E}_\Sigma(\mathcal{B})$ or $\mathcal{E}$, is defined as the smallest reducing subspace of Σ that contains $\mathcal{B}$. Let $u = \dim(\mathcal{E}_\Sigma(B))$ be the envelope dimension ($u \le p$). For simplicity, we use $\Gamma \in R^{r\times u}$ to be a semi-orthogonal basis of $\mathcal{E}_\Sigma(B)$. Consequently, $\mathcal{E}_\Sigma(\mathcal{B})$ decomposes the total variation into variations related to the material and immaterial parts of Y : $\Sigma_1 = \text{var}(P_\Gamma Y \,|\, X)$ and $\Sigma_2 = \text{var}(Q_\Gamma Y \,|\, X)$. We call (1) an envelope model when the structure of (a) and (b) is considered. Because B_1 is related only to the material part, the decomposition of Σ suggests that excluding the immaterial information makes estimation of B_1 more efficient. In particular, massive efficiency gains can be obtained when $||\Sigma_2|| \gg ||\Sigma_1||$ (Cook et al., 2010).

To formulate the envelope model, because $\mathcal{B}$ is constrained in $\text{span}(\Gamma)$, we can write $B_1 = \Gamma\eta$, where $\eta \in R^{u\times p}$ is the coordinate of B_1 relative to Γ. As Σ can be decomposed into Σ_1 and Σ_2, we have $\Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T$, where Γ_0 is a semi-orthogonal basis of the orthogonal subspace $\mathcal{E}^\perp$ and is a completion of Γ, $\Omega \in R^{u\times u}(>0)$ is the coordinate of Σ_1 relative to Γ, and $\Omega_0 \in R^{(r-u)\times(r-u)}(>0)$ is the coordinate of Σ_2 relative to Γ_0. Thus, the coordinate form of the envelope model is

$$Y = B_0 + \Gamma\eta X + \varepsilon, \quad \Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T. \tag{2}$$

If $u < r$, some immaterial information for the model estimation exists. Thus, by connecting B_1 to only the material subspace, one can gain estimation efficiency and reduce number of parameters. If $u = r$, then $\mathcal{E}_\Sigma(\mathcal{B}) = R^r$, which implies that there is no immaterial information and the envelope model reduces to the standard model. For more intuition on the envelope model, we refer to the review section in Ding & Cook (2018).

3 Envelope VAR model

3.1 Envelope model formulation for VAR

The VAR is a stochastic process model that is used to capture the linear interdependencies among multiple time series. Let $Y_t = (y_{1t}, y_{2t}, \ldots, y_{qt})^T$ denote a random vector of q time series. The p-lag vector autoregressive model (VAR(p)) has the form

$$Y_t = \mu + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \varepsilon_t, \quad t = 1, \ldots, n, \tag{3}$$

where β_i are $q \times q$ coefficient matrices and ε_t is a $q \times 1$ unobservable white noise vector process (serially uncorrelated or independent) and is usually assumed to be normal with mean zero and time invariant covariance matrix Σ. Let $\beta(B) = I_q - \beta_1 B - \ldots - \beta_p B^p$ be the backshift operator for (3). Then the VAR(p) series is weak stationary if all solutions of the determinant equation $|\beta(B)| = 0$ are greater than one in the modulus.

We define the envelope VAR model on the basis of the following reformulation:

$$Y_t = \mu + \beta Y_{p,t} + \varepsilon_t, \tag{4}$$

where $Y_{p,t} = (Y_{t-1}^T, Y_{t-2}^T, \ldots, Y_{t-p}^T)^T \in R^{qp}$ is the combined p-lag predictor vector with each $Y_{t-i}, i = 1, 2, \ldots, p$, having q time series variables, $\beta = (\beta_1, \beta_2, \ldots, \beta_p) \in R^{q \times pq}$ is the matrix of unknown coefficients, and ε_t follows $N(0, \Sigma)$.

Suppose that there is a subspace $S \subseteq R^q$ such that (i) the marginal distribution of $Q_S Y_t$ does not depend on the lag effect $Y_{p,t}$ and (ii) given the predictor vector $Y_{p,t}$, $P_S Y_t$ and $Q_S Y_t$ are uncorrelated. Here, P_S is the projection matrix onto S and Q_S is the orthogonal projection. Then a change in $Y_{p,t}$ can affect the distribution of Y_t only via $P_S Y_t$. Informally, we think of $P_S Y_t$ as the part of Y_t that is material to the regression, while $Q_S Y_t$ is $Y_{p,t}$-invariant and thus is immaterial. Likewise, we have (a) $\mathcal{B} \subseteq \text{span(S)}$, where $\mathcal{B} = \text{span}(\beta)$, and (b) $\Sigma = \Sigma_1 + \Sigma_2 = P_S \Sigma P_S + Q_S \Sigma Q_S$; thus, span($S$) is a reducing subspace of Σ that contains $\mathcal{B}$. The intersection of all reducing subspaces of Σ that contain $\mathcal{B}$ is called the Σ-envelope of $\mathcal{B}$, denoted as $\mathcal{E}_\Sigma(\mathcal{B})$, or $\mathcal{E}$ for brevity. The envelope $\mathcal{E}_\Sigma(\mathcal{B})$ serves to distinguish $P_S Y_t$ and the maximal $Y_{p,t}$-invariant $Q_S Y_t$ in the estimation of the VAR parameters. By removing immaterial variation, the envelope VAR model can acquire substantial increases in efficiency, sometimes equivalent to taking hundreds of additional observations.

Suppose that the dimension of the envelope $\mathcal{E}_\Sigma(\mathcal{B})$ is d. For now, we assume it is known. The selection of d is discussed in Section 4. Let $\Phi \in R^{q \times d}$ be a semi-orthogonal basis of $\mathcal{E}_\Sigma(\mathcal{B})$ and Φ_0 be a completion of Φ. Then reparameterizing (4) in terms of the basis matrix of $\mathcal{E}_\Sigma(\mathcal{B})$, we have its coordinate form:

$$Y_t = \mu + \Phi U Y_{p,t} + \varepsilon_t, \quad \Sigma = \Sigma_1 + \Sigma_2 = \Phi \Omega \Phi^T + \Phi_0 \Omega_0 \Phi_0^T, \tag{5}$$

where $\beta = \Phi U$, $U \in R^{d \times qp}$ carries the coordinates of β with respect to Φ, $\Omega \in R^{d \times d}$, and $\Omega_0 \in R^{(q-d) \times (q-d)}$ are positive definite and carry the coordinates of Σ with respect to Φ and Φ_0, respectively. When $d < q$, some immaterial information exists in the VAR model; thus, the envelope VAR approach can be more efficient. If $d = q$, then $\mathcal{E}_\Sigma(\mathcal{B}) = R^q$, indicating that there is no immaterial information and the envelope VAR model reduces to the standard VAR model.

3.2 Partial envelope model formulation for VAR

Partial envelope model was originally developed from considering that a subset of the predictors is of special interest in multivariate linear regression (Su & Cook, 2011). By targeting on a portion of predictors, the partial envelope is often a smaller subspace than the envelope introduced in Section 2. Thus, it can potentially eliminate more immaterial information for the estimation of the target parameters. Similar rationale can be used to build the partial envelope VAR model, which can offer gains that may not possible with the envelope VAR model.

In MTS, we might be more interested in some lag effects, and the rest orders might be less of interest or used to account for error dependency. For example, the effect from immediate time lags might be of main interest. Without loss of generality, suppose that Y_{t-1} is the main interest lag effect. In this case, we might partition the columns of β into $\beta_1 \in R^{q \times q}$ and $\beta_2 \in R^{q \times q(p-1)}$. Then we have

$$Y_t = \mu + \beta_1 Y_{t-1} + \beta_2 Y_{p-1,t-1} + \varepsilon_t, \tag{6}$$

where $Y_{t-1} = (y_{1,t-1}, y_{2,t-1}, \ldots, y_{q,t-1})^T \in R^q$ is of specific interest, $Y_{p-1,t-1} = (Y_{t-2}^T, \ldots, Y_{t-p}^T)^T \in R^{q(p-1)}$, and the error vector $\varepsilon_t \sim N(0, \Sigma)$. We then adopt the idea of partial envelope model and only consider the Σ-envelope for

$\mathcal{B}_1 = \text{span}(\beta_1)$, denoted by $\mathcal{E}_\Sigma(\mathcal{B}_1)$ or $\mathcal{E}_1$, and leave β_2 as unconstrained parameters. Correspondingly, $\mathcal{B}_1 \subseteq \mathcal{E}_\Sigma(\mathcal{B}_1)$ and $\Sigma = P_{\mathcal{E}_1}\Sigma P_{\mathcal{E}_1} + Q_{\mathcal{E}_1}\Sigma Q_{\mathcal{E}_1}$. This is the same as the envelope VAR structure, except that the enveloping is relative to $\mathcal{B}_1$ instead of the larger space $\mathcal{B}$. We refer to $\mathcal{E}_\Sigma(\mathcal{B})$ as the full envelope. Because $\mathcal{B}_1 \subseteq \mathcal{B}$, the partial envelope is contained in the full envelope, that is, $\mathcal{E}_\Sigma(\mathcal{B}_1) \subseteq \mathcal{E}_\Sigma(\mathcal{B})$. Thus, the partial envelope is likely to identify more immaterial variants and gain more efficiency in terms of estimating β_1. Suppose the dimension of $\mathcal{E}_\Sigma(\mathcal{B}_1)$ is d_1. Then $d_1 \le d \le q$. Let $\Phi_1 \in R^{q\times d_1}$ be a semi-orthogonal matrix, whose columns form a basis for $\mathcal{E}_\Sigma(\mathcal{B}_1)$. Let $\Phi_{10} \in R^{q\times(q-d_1)}$ be a completion of Φ_1 and a basis of $\mathcal{E}_1^\perp$. In addition, let $U_1 \in R^{d_1\times q}$ be the coordinates of β_1 in terms of the basis matrix Φ_1. Then model (6) can be reparameterized to the partial envelope VAR model as

$$Y_t = \mu + \Phi_1 U_1 Y_{t-1} + \beta_2 Y_{p-1,t-1} + \varepsilon_t, \quad \Sigma = \Sigma_{\mathcal{E}_1} + \Sigma_{\mathcal{E}_1^\perp} = \Phi_1\Omega_1\Phi_1^T + \Phi_{10}\Omega_{10}\Phi_{10}^T, \tag{7}$$

where Ω_1 and Ω_{10} are the coordinates of $\Sigma_{\mathcal{E}_1}$ and $\Sigma_{\mathcal{E}_1^\perp}$ relative to Φ_1 and Φ_{10}, respectively.

The stationarity condition of the envelope VAR or partial envelope VAR models is the same as that of the standard VAR model except now β_S or part of β_S have some low-rank structure.

4 Model selection

To fit a VAR model, we first need to determine the time lag p. The lag length for the VAR model can be determined using model selection criteria. The general approach is to fit the VAR models with $p = 0, \ldots, p_{\max}$ and choose the value of p which minimizes some information criteria. Typical information criteria for the VAR models have the form

$$\text{IC}(p) = \ln|\bar{\Sigma}(p)| + c_n \cdot \varphi(p), \tag{8}$$

where $\bar{\Sigma}(p) = \frac{1}{n}\sum_{t=1}^n \hat{\varepsilon}_t\hat{\varepsilon}_t'$ is the estimated residual covariance matrix, c_n is a sequence indexed by the sample size n, and $\varphi(p)$ is a penalty function that penalizes large VAR models. The standard VAR(p) models have $\varphi(p) = pq^2 + q + q(q+1)/2$ free parameters. The dimension q is kept constant, without imposing any restrictions on the error covariance matrix, acting as if the VAR(p) model has $\varphi(p) = pq^2$ parameters. The three most common information criteria are the Akaike information criterion (AIC), Schwarz–Bayesian [Bayesian information criterion (BIC)], and Hannan–Quinn (HQ):

$$\begin{aligned}
\text{AIC}(p) &= \ln|\bar{\Sigma}(p)| + \frac{2}{n}pq^2 \\
\text{BIC}(p) &= \ln|\bar{\Sigma}(p)| + \frac{\ln(n)}{n}pq^2 \\
\text{HQ}(p) &= \ln|\bar{\Sigma}(p)| + \frac{2ln(\ln(n))}{n}pq^2.
\end{aligned} \tag{9}$$

The AIC asymptotically overestimates the order with positive probability, whereas the BIC and HQ criteria estimate the order consistently under fairly general conditions if the true order p is less than or equal to $p_{\max}$. Hence, in our numerical studies, we select p by BIC for order selection. The order p can also be selected by cross validation when normality is not assumed.

Once p is selected, the VAR order is determined. We next need to select the dimension d of the envelope, or d_1 of the partial envelope. The dimension of the envelope or the partial envelope can also be determined by using an information criterion:

$$\begin{aligned}
\text{IC}(d) &= \ln|\bar{\Sigma}(d)| + c_n \cdot \varphi(d), \\
\text{IC}(d_1) &= \ln|\bar{\Sigma}(d_1)| + c_n \cdot \varphi(d_1),
\end{aligned} \tag{10}$$

where the basic structure of the information criterion is similar as previously described for p. We replace p with d or d_1 and keep c_n the same; meanwhile, we use $\varphi(d)$ or $\varphi(d_1)$ to represent the number of parameters under the envelope VAR or partial envelope VAR model. For example, for the envelope VAR model, the number of parameters to be estimated is $q + qpd + q(q-d) + \frac{d(d+1)}{2} + \frac{(q-d)(q-d+1)}{2}$. After simplifying, we have $\varphi(d) = q + dqp + q(q+1)/2$. For the partial envelope model (7), $\varphi(d_1) = q + d_1 q + q^2(p-1) + q(q+1)/2$.

Recall that the standard VAR model has $q \times qp + q + q(q+1)/2$ parameters. Compared with the standard model, the envelope VAR model reduces the number of parameters by $(q-d) \times qp$, while the partial envelope model reduces the number of parameters by $(q-d_1)q$.

We again choose BIC for envelope dimension determination. When d or d_1 is equal to q, the envelope or partial envelope model reduces to the standard VAR. There is no efficiency gain. When d or d_1 is less than q, we expect that the envelope or partial envelope VAR model would gain potential efficiency. We might also determine p and d simultaneously with the information criteria. The results are similar in this case.

5 Parameter estimation

We next present the estimation for the envelope VAR model. For a given lag p and an envelope dimension d, there are unknown parameters μ, $\mathcal{E}_\Sigma(\mathcal{B})$, U, Ω, and Ω_0 to be estimated. Under normality, we use the maximum likelihood estimation (MLE) for the parameter estimation.

The conditional log likelihood function $L(\mu, \mathcal{E}_\Sigma(\mathcal{B}), U, \Omega, \Omega_0)$ for $Y_t \mid Y_{p,t}, t = p+1, \ldots, n$, can be expressed as

$$\begin{aligned} L = &-(nq/2)\log(2\pi) - (n/2)\log|\Phi\Omega\Phi^T + \Phi_0\Omega_0\Phi_0^T| \\ &- (1/2)\sum_{t=p+1}^{n}(Y_t - \mu - \Phi U Y_{p,t})^T(\Phi\Omega\Phi^T + \Phi_0\Omega_0\Phi_0^T)^{-1}(Y_t - \mu - \Phi U Y_{p,t}). \end{aligned} \tag{11}$$

Without loss of generality, suppose that the sample predictors are centred, and then the maximum likelihood estimator of μ is simply $\hat{\mu} = \bar{Y}_t = \frac{1}{n-p}\sum_{t=p+1}^{n} Y_t$. Substitute this back to the log likelihood function and then decompose $Y_t - \bar{Y}_t = P_\Phi(Y_t - \bar{Y}_t) + Q_\Phi(Y_t - \bar{Y}_t)$. After a series of derivations (see the Appendix), given a basis Φ, the MLEs of U, Ω, and Ω_0 can all be represented as functions of Φ. Thus, the conditional log likelihood can finally be simplified as

$$L_1 = -(nq/2)\left(\log(2\pi) + 1\right) - (n/2)\log(\hat{\Sigma}_{Y_t}) - (n/2)\log|\Phi^T\hat{\Sigma}_{\text{res}}\Phi| - (n/2)\log|\Phi^T\hat{\Sigma}_{Y_t}^{-1}\Phi|, \tag{12}$$

where $\hat{\Sigma}_{Y_t}$ is the sample covariance matrix of Y_t and $\hat{\Sigma}_{\text{res}}$ denotes the sample covariance matrix of residuals from the regression of Y_t on $Y_{p,t}$. The estimates of the remaining parameters require the estimator of $\mathcal{E}_\Sigma(\mathcal{B})$, or of its basis. To estimate the $\mathcal{E}_\Sigma(\mathcal{B})$, Cook et al. (2010) solved the manifold optimization problem:

$$\hat{\mathcal{E}}_\Sigma(\mathcal{B}) = \underset{\text{span}(\Phi)\in\mathcal{G}(q,d)}{\text{argmin}} \log|\Phi^T\hat{\Sigma}_{\text{res}}\Phi| + \log|\Phi^T\hat{\Sigma}_{Y_t}^{-1}\Phi|, \tag{13}$$

where $\mathcal{G}(q,d)$ denotes a $q \times d$ Grassmann manifold, which is the set of all d-dimensional subspaces in a q-dimensional space, meaning that the minimum in (13) is taken over all semi-orthogonal matrices $\Phi \in R^{q\times d}$. However, a Grassmannian optimization can be relatively slow. Therefore, we consider a non-Grassmannian optimization method proposed in Cook et al. (2016).

To illustrate the idea, without loss of generality, assume that the first d rows of Φ, denoted by Φ_a, is full rank. Then Φ can be formulated as

(wileyonlinelibrary.com) https://doi.org/10.1002/sta4.203

$$\Phi = \begin{bmatrix} \Phi_a \\ \Phi_b \end{bmatrix} = \begin{bmatrix} I_u \\ A \end{bmatrix} \times \Phi_a = \Phi_A \Phi_a, \tag{14}$$

where $A = \Phi_b \Phi_a^{-1} \in R^{(q-d)\times d}$ is an unconstrained matrix and $\Phi_A = (I_d, A^T)^T$. As Φ_a is full rank, then Φ_A is basis of the envelope $\mathcal{E}_\Sigma(\mathcal{B})$. With the use of the relationship, the objective function in (13) can be reparameterized as a function of only A as

$$-2\log|\Phi_A^T \Phi_A| + \log|\Phi_A^T \hat{\Sigma}_{\text{res}} \Phi_A| + \log|\Phi_A^T \hat{\Sigma}_{Y_t}^{-1} \Phi_A|. \tag{15}$$

In (15), the minimization over A is unconstrained. Hence, it can be easily solved by standard optimization methods (Cook et al., 2016). Once A is estimated, we can accordingly obtain a basis estimator, denoted by $\hat{\Phi}$ for $\mathcal{E}_\Sigma(\mathcal{B})$. Correspondingly, the MLEs of the remaining parameters in the envelope VAR model (5) can be obtained, denoted by $\hat{U}$, $\hat{\Omega}$, and $\hat{\Omega}_0$ (see the Appendix). Hence, the envelope MLEs of β and Σ are $\hat{\beta} = \hat{\Phi}\hat{U}$ and $\hat{\Sigma} = \hat{\Phi}\hat{\Omega}\hat{\Phi}^T + \hat{\Phi}_0\hat{\Omega}_0\hat{\Phi}_0^T$. It can be shown that $\hat{\beta} = P_{\hat{\Phi}}\tilde{\beta}$, where $\tilde{\beta}$ is the standard MLE of β without enveloping. Therefore, the envelope MLE of β is the projection of the standard MLE onto the (estimated) envelope.

For the partial envelope model, the maximum likelihood estimator $\hat{\beta}_1$ of β_1 is the projection of the estimator of β_1 from the standard model onto $\hat{\mathcal{E}}_\Sigma(\mathcal{B}_1)$. The maximum likelihood estimator $\hat{\beta}_2$ of β_2 is the estimated coefficient matrix from the ordinary least squares fit of the residuals $Y_t - \bar{Y}_t - \hat{\beta}_1 Y_{t-1}$ on $Y_{p-1,t-1}$. The derivation is similar to that of Su & Cook (2011).

6 Asymptotic results

In this section, we provide the asymptotic distribution for the envelope VAR estimators and compare it with that of the standard VAR estimator. We show that the envelope VAR estimators are asymptotically more efficient than the standard estimator is.

Let $\tilde{\beta}$ and $\tilde{\Sigma}$ be the MLEs of β and Σ under the standard VAR model (4) without enveloping. Then $(\text{vec}(\tilde{\beta})^T, \text{vech}(\tilde{\Sigma})^T)^T$ is asymptotically jointly normal with mean $(\text{vec}(\beta)^T, \text{vech}(\Sigma)^T)^T$ and covariance matrix

$$\Delta = \begin{bmatrix} G^{-1} \otimes \Sigma & 0 \\ 0 & 2C_q(\Sigma \otimes \Sigma)C_q \end{bmatrix},$$

where "vec" and "vech" represent the vector operator and vector half operator (Henderson & Searle, 1979), "$\otimes$" stands for Kronecker product, $G = E(Y_{p,t}Y_{p,t}^T)$, and C_q is the "contraction" matrix such that $\text{vech}(\Sigma) = C_q \text{vec}(\Sigma)$.

Let $\psi = (\text{vec}(U)^T, \text{vec}(\Phi)^T, \text{vech}(\Omega)^T, \text{vech}(\Omega_0)^T)^T$. Denote

$$h(\psi) = \begin{bmatrix} \text{vec}(\beta) \\ \text{vech}(\Sigma) \end{bmatrix} = \begin{bmatrix} \text{vec}(\Phi U) \\ \text{vech}(\Phi\Omega\Phi^T + \Phi_0\Omega_0\Phi_0^T) \end{bmatrix}. \tag{16}$$

The following proposition gives the asymptotic distribution for the MLE of the envelope VAR model.

Proposition 1

Suppose that the envelope VAR(p) model (5) *is stationary, and* ϵ_t *follows multivariate normal distribution with mean zero and positive definite covariance matrix* Σ*. Then (i) the envelope MLE* $(\text{vec}(\hat{\beta})^T, \text{vech}(\hat{\Sigma})^T)^T$ *is asymptotically normal with mean* $(\text{vec}(\beta)^T, \text{vech}(\Sigma)^T)^T$ *and covariance matrix* $\Lambda = H(H^T\Delta^{-1}H)^\dagger H^T$*, where* $H = \frac{\partial h(\psi)}{\partial \psi^T}$*; (ii)* $\Delta - \Lambda \geq 0$*. Thus, the envelope VAR estimator is asymptotically more efficient than the standard VAR estimator.*

Because in MTS one is particularly interested in making inference on β, in the following, we give the marginal asymptotic distribution of $\text{vec}(\hat{\beta})$.

Proposition 2

Suppose that the envelope VAR (p) model is stationary, and the error ϵ_t follows multivariate normal distribution with mean zero and positive definite covariance matrix Σ. Then the envelope MLE vec($\hat{\beta}$) is asymptotically normal with mean vec(β) and covariance matrix

$$avar[\sqrt{n}vec(\hat{\beta})] = G^{-1} \otimes \Phi\Omega\Phi^T + (U^T \otimes \Phi_0)V^{-1}(U \otimes \Phi_0^T), \tag{17}$$

where $V = UGU^T \otimes \Omega_0^{-1} + \Omega \otimes \Omega_0^{-1} + \Omega^{-1} \otimes \Omega_0 - 2I_d \otimes I_{q-d}$.

The proof of the propositions is similar to that in Cook et al. (2010). The results of Proposition 1 can be obtained by employing the asymptotic properties of overparameterized structural models by Shapiro (1986), and the results of Proposition 2 can be achieved directly from Proposition 1 by simplifying and partitioning the joint asymptotic covariance matrix Λ. The outline of the proof is given in the Appendix.

If $d = q$, then $\Phi\Omega\Phi^T = \Sigma$, and the second term on the right-hand side of (17) vanishes. The envelope estimator $\hat{\beta}$ reduces to the standard estimator $\tilde{\beta}$. When $d < q$, the first term on the right-hand side of (17) is the asymptotic variance of $\hat{\beta}$ when Φ is known, and the second term can be interpreted as the "cost" of estimating $\mathcal{E}_\Sigma(\mathcal{B})$. The total on the right does not exceed $G^{-1} \otimes \Sigma$, which is the asymptotic variance of $\tilde{\beta}$ from the standard model. Recall that

$$\text{avar}[\sqrt{n}\text{vec}(\tilde{\beta})] = G^{-1} \otimes \Sigma = G^{-1} \otimes \Phi\Omega\Phi^T + G^{-1} \otimes \Phi_0\Omega_0\Phi_0^T. \tag{18}$$

Subtracting the asymptotic variance of vec($\hat{\beta}$) from (18), we have

$$\text{avar}[\sqrt{n}\text{vec}(\tilde{\beta})] - \text{avar}[\sqrt{n}\text{vec}(\hat{\beta})] = G^{-1} \otimes \Phi_0\Omega_0\Phi_0^T - (U^T \otimes \Phi_0)V^{-1}(U \otimes \Phi_0^T) \geq 0. \tag{19}$$

Asymptotic results for the partial envelope model can be similarly derived as for the envelope model. The only difference is that now we have

$$h(\psi) = \begin{bmatrix} \text{vec}(\beta_1) \\ \text{vec}(\beta_2) \\ \text{vech}(\Sigma) \end{bmatrix} = \begin{bmatrix} \text{vec}(\Phi_1 U_1) \\ \text{vec}(\beta_2) \\ \text{vech}(\Phi_1\Omega_1\Phi_1^T + \Phi_{10}\Omega_{10}\Phi_{10}^T) \end{bmatrix},$$

where $\psi = (\text{vec}(U_1)^T, \text{vec}(\Phi_1)^T, \text{vec}(\beta_2)^T, \text{vech}(\Omega_1)^T, \text{vech}(\Omega_{10})^T)^T$. The results in Proposition 1 similarly hold.

In later applications, instead of using asymptotic variances of the estimators of β, we use their asymptotic standard errors (ase) to compare the envelope or partial envelope VAR model with the standard VAR model. Specifically, we demonstrate the advantages or efficiency of the envelope methods by assessing if the element-wise ratios of ase[$\sqrt{n}$vec($\tilde{\beta}$)] to ase[$\sqrt{n}$vec($\hat{\beta}$)] are larger than 1. We also computed bootstrap standard errors (SEs), which that showed similar results.

7 Simulation studies

In this section, we evaluate the performance of the envelope VAR model numerically and compare it with that of the VAR model. The envelope VAR model can be implemented using the R software package *Renvlp*. We first generated data on the basis of envelope models with four parameter settings, by varying the size of the envelope dimension d and the size of the time series vector q: (i) $q = 10$, $d = 2$, $\Omega = \sigma^2 I_d$, $\Omega_0 = \sigma_0^2 I_{q-d}$; (ii) $q = 8$, $d = 2$, $\Omega = \sigma^2 I_d$, $\Omega_0 = \sigma_0^2 I_{q-d}$; (iii) $q = 5$, $d = 1$, $\Omega = \sigma^2 I_d$, $\Omega_0 = \sigma_0^2 I_{q-d}$; and (iv) $q = 3$, $d = 1$, $\Omega = \sigma^2 I_d$, $\Omega_0 = \sigma_0^2 I_{q-d}$; $\sigma^2 = 0.1$ and

(wileyonlinelibrary.com) https://doi.org/10.1002/sta4.203

$\sigma_0^2 = 1$ for all cases. The semi-orthogonal matrix Φ was generated by orthogonalized matrix of independent uniform $(0, 1)$ random variables. The elements of μ and U and the first observation Y_1 were generated from standard normal variables. Then $Y_t, t = 2, \ldots, n$, were sequentially obtained by VAR (1) model. All the four series are stationary. We then fitted the VAR and envelope VAR models to the data and evaluated their estimation accuracy of the coefficient parameters by comparing the estimated β with the true β, according to the following criterion: $||\hat{\beta} - \beta||_F$, where "$||\cdot||_F$" represents the Frobenius norm. We used seven different sample sizes, $n = 100, 200, 300, 500, 800, 1000$, and 1500. Within each sample size, 100 replicates were simulated. The average estimation and prediction errors were computed over the 100 replicates for each sample size and each model. The envelope dimensions were selected by BIC. We also divided simulated data into training and testing sets by selecting the first 80% of data as training set and the rest 20% of data as testing set. Cross validation is not meaningful as the target of MTS is to predict future values. We used different models to obtain coefficient estimations with the training set and then obtained prediction errors for the testing set.

Table I shows the envelop dimension selection results for d over different sample sizes for the four settings. For example, when $q = 10$ and $d = 2$, there are 99% of times that true envelope dimension was selected. We see that over all the sample sizes and settings, the BIC can correctly select the envelope dimension with very high percentages.

Table II shows the mean, maximum, and minimum SE ratios of $\tilde{\beta}$ from the VAR model relative to $\hat{\beta}$ from the envelope VAR model over 100 runs. The mean ratios range from 7.03 to 19.83. The maximum ratios range from 13.75 to 96.67, and the minimum ratios range from 1 to 4.94. The results approve that the envelope VAR model is more efficient than the standard VAR model, and efficiency gains are massive in some cases.

Table I. Results of envelope dimension selection by Bayesian information criterion.

Setting	100 (%)	200 (%)	300 (%)	500 (%)	800 (%)	1000 (%)	1500 (%)
$q = 10, d = 2$	99	100	100	100	100	100	100
$q = 8, d = 2$	100	100	100	100	100	100	100
$q = 5, d = 1$	99	99	100	100	100	100	100
$q = 3, d = 1$	97	99	100	100	100	100	100

Table II. The average, maximum, and minimum ratios of standard errors of $\tilde{\beta}$ to standard errors of $\hat{\beta}$ over 100 runs.

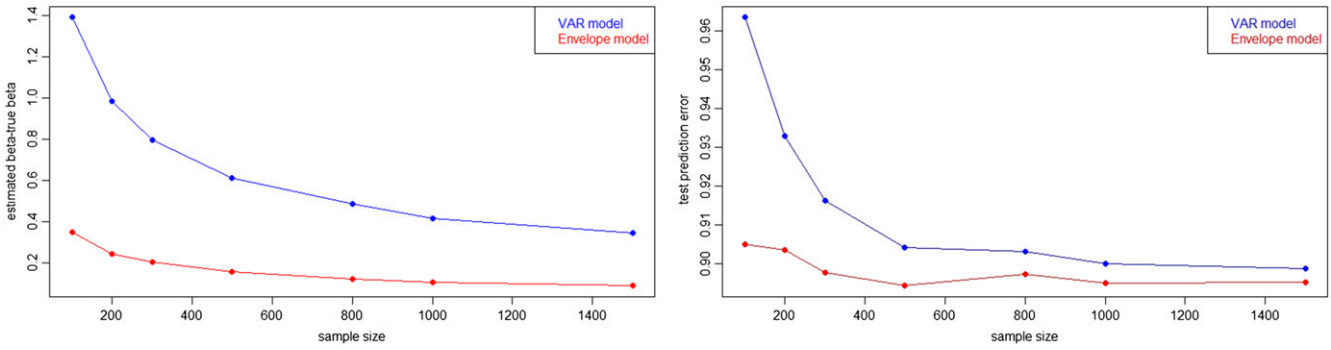
Setting	Level	100	200	300	500	800	1000	1500
$q = 10, d = 2$	Mean	9.57	9.31	9.10	9.15	9.09	9.12	9.09
	Max	30.82	26.50	25.54	23.83	23.15	23.34	22.18
	Min	1.57	3.72	4.14	4.27	4.72	4.79	4.94
$q = 8, d = 2$	Mean	7.57	7.23	7.18	7.11	7.05	7.04	7.03
	Max	23.48	17.57	16.50	15.43	14.45	14.37	13.75
	Min	1.58	1.69	1.80	1.81	1.81	1.85	1.87
$q = 5, d = 1$	Mean	10.89	10.57	10.56	10.55	10.51	10.48	10.50
	Max	28.14	21.81	22.59	19.77	19.03	18.56	17.85
	Min	1.25	1.14	3.64	4.06	4.34	4.32	4.36
$q = 3, d = 1$	Mean	19.46	19.63	19.74	19.57	19.67	19.77	19.83
	Max	96.67	87.92	81.85	81.31	77.72	75.71	74.53
	Min	1.00	1.04	3.88	4.08	4.19	4.21	4.37

(wileyonlinelibrary.com) https://doi.org/10.1002/sta4.203

Stat
The ISI's Journal for the Rapid Dissemination of Statistics Research

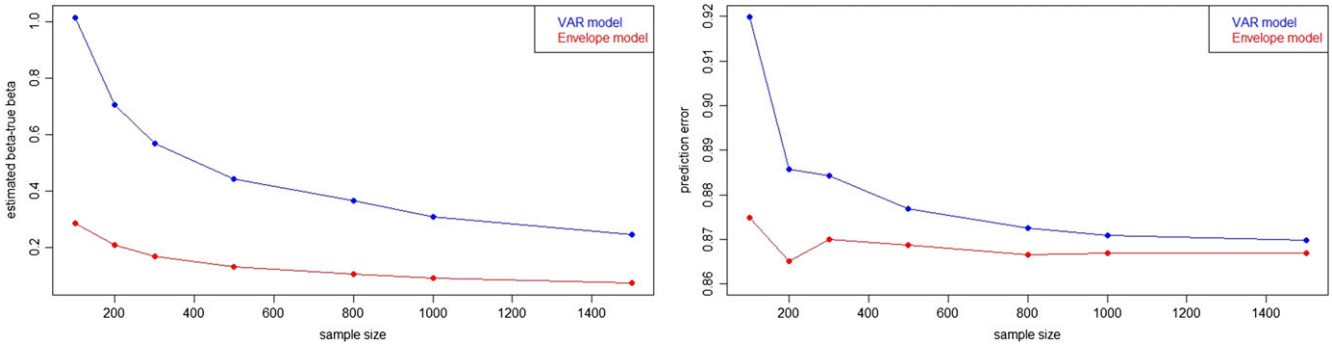
Figures 1–4 demonstrate the estimation and prediction accuracy of the envelope VAR model compared with the VAR model. The left panels show the results of estimation errors $||\hat{\beta} - \beta||_F$. We can see that for all settings, the estimation errors decrease as sample size increases. The estimation errors of the envelope model are smaller than those of the standard VAR model, especially when sample size is small. For example, when sample size is 100, the estimation errors of the envelope VAR model are 3.5 times smaller than those of the VAR model. The envelope model estimates more accurately.

The right panels in Figures 1–4 are the testing set prediction errors. It can be seen that over all sample sizes and different settings, the prediction errors of the envelope VAR model are smaller than those of the standard VAR model. At small sample size levels, the difference of the prediction errors between the two models is even larger. When the time series dimension q is small, the prediction errors do not monotonously decrease with increasing sample sizes and show some fluctuations. However, for large q, the prediction errors decrease with increasing sample sizes. The results in Figures 1–4 together indicate that the envelope VAR model can effectively improve parameter estimates and enhance prediction performance over the standard VAR model. The partial envelope VAR models show similar results. Owing to page limitation, we will demonstrate the performance of the partial envelope VAR model in the real application.



(a) Estimation error comparison (b) Testing set prediction error comparison

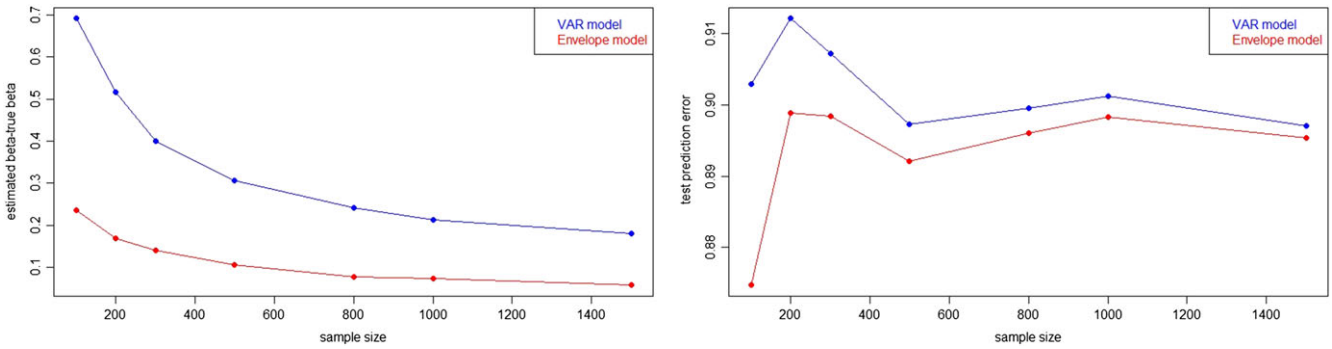
Figure 1. $q = 10, d = 2$.



(a) Estimation error comparison (b) Testing set prediction error comparison

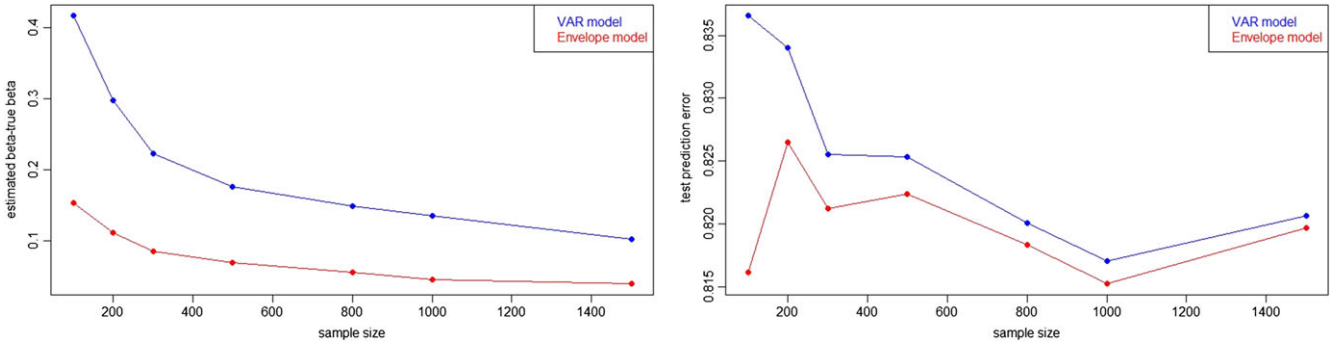
Figure 2. $q = 8, d = 2$.

(wileyonlinelibrary.com) https://doi.org/10.1002/sta4.203



(a) Estimation error comparison (b) Testing set prediction error comparison

Figure 3. $q = 5, d = 1$.



(a) Estimation error comparison (b) Testing set prediction error comparison

Figure 4. $q = 3, d = 1$.

8 Data applications

In this section, we demonstrate the efficiency gains of the envelope/partial envelope VAR models based on three real data examples.

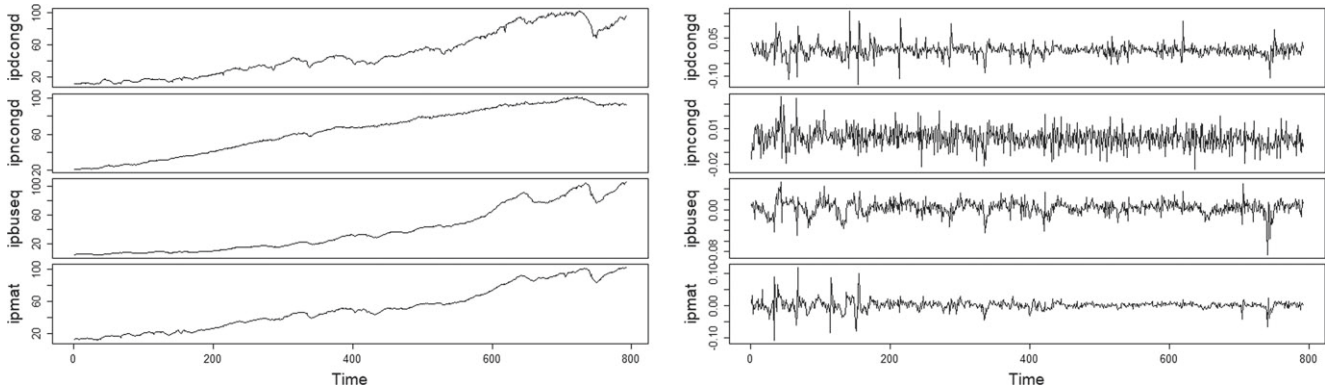
8.1 Envelope model

The first data set consists of: US monthly industrial production index from January 1947 to December 2012 for 792 data points. The original data are from the Federal Reserve Bank of St. Louis and are seasonally adjusted. The data are also used as an example in Tsay (2013). The four components are durable consumer goods (IPDCONGD), non-durable consumer goods (IPNCONGD), business equivalent (IPBUSEQ), and materials (IPMAT). The time series graph is shown in the left panel of Figure 5. Data were pretreated with the log transformation and differencing to be stabilized and to reduce trend. After this, we obtain stationary time series, which are shown in the right panel of Figure 5. We then applied the VAR and envelope VAR to the stationary series. The VAR order was determined to be 2 by BIC, and the error term behaved closely as a white noise process (Durbin & Watson, 1951). Now, the standard VAR model is structured with $Y_t \in R^4$, $Y_{p,t} \in R^8$, $\beta = (\beta_1, \beta_2) \in R^{4\times 8}$, $\beta_1 \in R^{4\times 4}$ for the first order and $\beta_2 \in R^{4\times 4}$ for the second order. To evaluate model prediction performance, we divided data into the training and testing sets by taking first 80%

Stat

(wileyonlinelibrary.com) https://doi.org/10.1002/sta4.203

The ISI's Journal for the Rapid Dissemination of Statistics Research



(a) Time series of original data

(b) Time series after transformation

Figure 5. Consumer goods time series.

Table III. Asymptotic standard errors of VAR and envelope VAR models.

	VAR SE_v	Envelope VAR SE_e	Ratio SE_v/SE_e
β_1	1.14, 3.30, 2.35, 1.82	1.10, 3.16, 2.52, 1.75	1.04, 1.04, 1.04, 1.04
	0.36, 1.03, 0.74, 0.57	0.10, 0.28, 0.27, 0.24	3.61, 3.71, 2.70, 2.42
	0.54, 1.57, 1.12, 0.87	0.53, 1.54, 1.10, 0.85	1.02, 1.02, 1.02, 1.02
	0.69, 1.99, 1.42, 1.10	0.44, 1.29, 0.93, 0.79	1.55, 1.55, 1.53, 1.39
β_2	1.13, 3.29, 2.29, 1.92	1.08, 3.15, 2.20, 1.84	1.04, 1.04, 1.04, 1.04
	0.35, 1.03, 0.72, 0.60	0.11, 0.28, 0.35, 0.17	3.35, 3.68, 2.04, 3.48
	0.54, 1.56, 1.09, 0.91	0.53, 1.54, 1.07, 0.90	1.02, 1.02, 1.02, 1.02
	0.68, 1.98, 1.38, 1.16	0.44, 1.29, 0.92, 0.75	1.55, 1.54, 1.50, 1.54

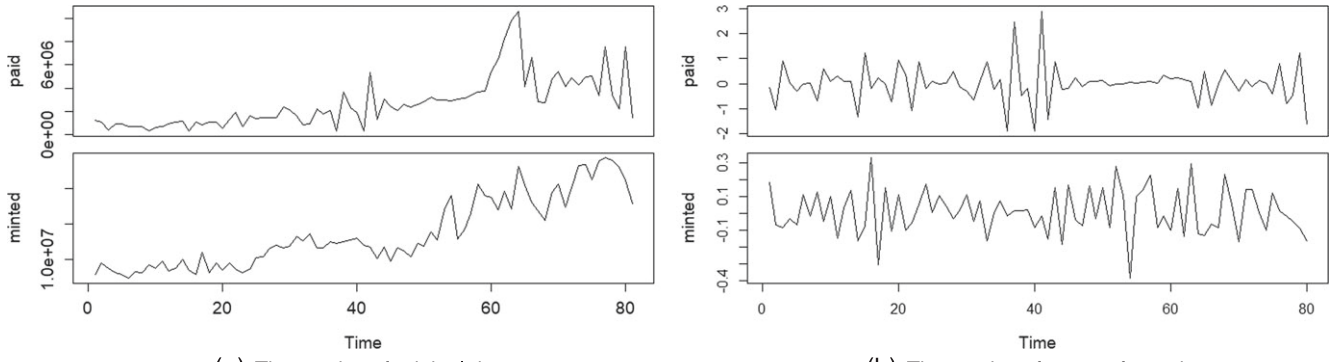
Note: SE, standard error; VAR, vector autoregression.

of data as training data and the rest as the testing data. For the envelope VAR model, the dimension d was selected to be 2 by BIC. We fitted both the VAR and envelope VAR models and compared the results. The asymptotic SEs for each element in the estimated β from two models are summarized in Table III, with subscripts "e," "pe," and "v" indicating the results from the envelope VAR model, partial envelope model, and standard VAR model, respectively, where the subscripts for the partial envelope were used in a later example.

It can be seen from Table III that the envelope VAR model provides smaller asymptotic SEs for the estimation of all the elements of β. The ratios of SE_v to SE_e range from 1.02 to 3.68. We further evaluated the sum of the squared prediction errors (SSPEs) on the basis of testing set for both models. It shows that the envelope VAR model and the standard VAR model have comparable prediction performance, both with SSPEs around 0.0136. Although the envelope VAR model does not lead to much improvement in prediction, it indeed improves estimation efficiency in this case.

The second data are about silver minted and payments made to New Spain between 1720 and 1800 and are available at http://users.stat.ufl.edu/ winner/data/treas1700.dat. There are two series, *situados* paid to New Spain (paid) and silver minted (minted). Same as the first data, the second data were also transformed to stationary by differencing and the log transformation. The series before and after transformation are shown in Figure 6. The VAR order was selected to be 3 based on BIC, and the error term is close to a white noise process. Then the setting of VAR model is $Y_t \in R^2$, $Y_{p,t} \in R^6$, $\beta = (\beta_1, \beta_2, \beta_3) \in R^{2\times 6}$. The dimension d of the envelope VAR model was selected to be 1

(wileyonlinelibrary.com) https://doi.org/10.1002/sta4.203



(a) Time series of original data (b) Time series after transformation

Figure 6. Silver minted and payments made to New Spain time series.

Table IV. Asymptotic standard errors of VAR and envelope VAR models.

	VAR SE_v	Envelope VAR SE_v	Ratio SE_v/SE_e
β_1	0.92, 4.32	0.92, 4.32	1.00, 1.00
	0.21, 0.99	0.17, 0.10	1.26, 9.95
β_2	1.11, 4.73	1.11, 4.73	1.00, 1.00
	0.25, 1.08	0.12, 0.08	1.00, 11.00
β_3	0.91, 4.31	0.91, 4.31	1.00, 1.00
	0.21, 0.98	0.08, 0.08	2.63, 11.99

Note: SE, standard error; VAR, vector autoregression.

by BIC. The asymptotic SEs for each element in the estimated β from the two models are summarized in Table IV. The ratios of asymptotic SEs between the VAR and envelope VAR estimates are between 1 and 11.99, showing that envelope model produces efficient gains by largely reducing the SEs of the estimates. Correspondingly, the prediction error (SSPE) of the envelope model for the testing set is 0.467, which is smaller than 0.469 of the standard VAR model. The envelope model improves both estimation and prediction in this example.

8.2 Partial envelope model

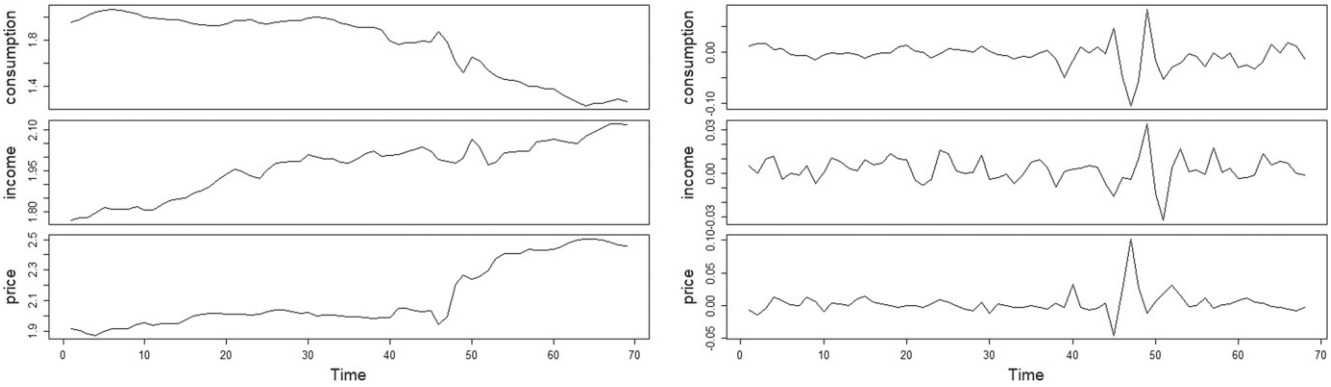
The third data are annual consumption of spirits in the UK from 1870 to 1938. The data are originally from Shapiro (1986) and are available at http://users.stat.ufl.edu/~winner/data/spirits.dat. The series in the data set include indexed consumption, income, and price of spirits. Similarly, we differenced and log transformed the data and acquired stationary time series, shown in Figure 7. The order of the VAR model was selected to be 2 by BIC. Hence, the VAR model has $Y_t \in R^3$, $Y_{p,t} \in R^6$, and $\beta = (\beta_1, \beta_2) \in R^{3\times 6}$. We applied the VAR, envelope VAR, and partial envelope VAR models to these data. The partial envelope VAR model can also be implemented with the R software package *Renvlp*. The dimension d of the envelope model and d_1 of the partial envelope model were determined to be 2 and 1, respectively, by BIC. The asymptotic SEs for the estimated β from the VAR, envelope VAR, and partial envelope VAR models are shown in Table V.

The third column in Table V shows that the ratios of asymptotic SEs of β estimation for the VAR and envelope VAR models are only a little larger than 1, indicating that the envelope model does not gain much efficiency in this case. However, according to the last column in Table V, the partial envelope model largely reduces the coefficient estimation variation with SE_v/SE_{pe} ratios ranging from 1.29 to 5.42 for the estimated β_1, meaning that by partially enveloping part of the parameter space, one can potentially reduce more immaterial information and gain more efficiency.

Stat

(wileyonlinelibrary.com) https://doi.org/10.1002/sta4.203

The ISI's Journal for the Rapid Dissemination of Statistics Research



(a) Time series of original data (b) Time series after transformation

Figure 7. UK consumption, income, and price of spirits time series.

9 Discussion

In this paper, we present the application and extension of the envelope and partial envelope models to MTS, especially to the VAR settings. The new methods can identify material and immaterial information in VAR, and by eliminating immaterial variation, they can achieve substantial efficiency gains in model estimation, which has been shown in our theoretical results and numerical studies including both simulation and real data analysis. The envelope VAR model can also help improve prediction performance although testing set error reduction might not always be guaranteed. When there is no reduction, the envelope VAR model performs very similarly to the standard VAR model in prediction but can achieve potential efficiency gains in estimation. Thus, it can still be beneficial. The idea of enveloping is mainly applied to response reduction in the VAR models. Future works on envelopes and VAR models can be extended to the setting of predictor reduction by exploiting similar rationales in Cook et al. (2013). This direction can also help build connections between envelopes and PLS regression in VAR settings and demonstrate potential advantages of the envelope methods over PLS in time series modelling. In addition, we mainly target on studying the envelop/partial envelope VAR models. In reality, more complex MTS processes such as VAR and moving average (VARMA) and seasoned VAR or VARMA models are worth considering. Furthermore, dimension reduction with simultaneous variable selection (Su et al., 2016; Qian et al., 2018) for MTS modelling is also a meaningful direction to explore. The proposed method can be further extended to matrix or tensor settings (Hoff, 2015; Ding & Cook, 2014; 2015a,b). We leave these works for future investigation.

Appendix. Maximum likelihood estimation

The conditional log likelihood function $L(\mu, \mathcal{E}_\Sigma(\mathcal{B}), U, \Omega, \Omega_0)$ for $Y_t \,|\, Y_{p,t}, t = p+1, \ldots, n$, is equivalent to

$$\begin{aligned} L = &-(nq/2)\log(2\pi) - (n/2)\log|\Omega| - (n/2)\log|\Omega_0| \\ &- (1/2)\sum_{t=p+1}^{n} (Y_t - \mu - \Phi U Y_{p,t})^T (\Phi\Omega^{-1}\Phi^T + \Phi_0\Omega_0^{-1}\Phi_0^T)^{-1}(Y_t - \mu - \Phi U Y_{p,t}). \end{aligned} \tag{A1}$$

Suppose the sample predictors are centred without loss of generality, and then the maximum likelihood estimator of μ is $\hat{\mu} = \bar{Y}_t$. Substituting this into the log likelihood function and then decomposing $Y_t - \bar{Y}_t = P_\Phi(Y_t - \bar{Y}_t) + Q_\Phi(Y_t - \bar{Y}_t)$ and simplifying, we arrive at the first partially maximized log likelihood:

$$L_1(U, \mathcal{E}_\Sigma(\mathcal{B}), \Omega, \Omega_0) = -(nr/2)\log(2\pi) + L_{11}(U, \mathcal{E}_\Sigma(\mathcal{B}), \Omega) + L_{12}(\mathcal{E}_\Sigma(\mathcal{B}), \Omega_0), \tag{A2}$$

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) https://doi.org/10.1002/sta4.203

where

$$L_{11}(U, \mathcal{E}_\Sigma(\mathcal{B}), \Omega) = -(n/2)\log|\Omega| - 1/2 \sum_{t=p+1}^{n} \{\Phi^T(Y_t - \bar{Y}_t) - UY_{p,t}\}^T \Omega^{-1} \{\Phi^T(Y_t - \bar{Y}_t) - UY_{p,t}\},$$

$$L_{12}(\mathcal{E}_\Sigma(\mathcal{B}), \Omega_0) = -(1/2)\log|\Omega_0| - 1/2 \sum_{t=p+1}^{n} (Y_t - \bar{Y}_t)^T \Phi_0 \Omega_0^{-1} \Phi_0^T (Y_t - \bar{Y}_t)\}.$$

Holding Φ fixed, L_{11} can be seen as the log likelihood for the multivariate regression of $\Phi^T(Y_t - \bar{Y}_t)$ on $Y_{p,t}$, and thus, L_{11} is maximized over U at the value $U = \Phi^T\tilde{\beta}$, where $\tilde{\beta}$ is the MLE of β from the standard VAR fit. Substituting this into L_{11} and simplifying, we obtain a partially maximized version of L_{11} as

$$L_{21}(\mathcal{E}_\Sigma(\mathcal{B}), \Omega) = -(n/2)\log|\Omega| - (1/2)\sum_{i=1}^{n}(\Phi^T r_i)^T \Omega^{-1} \Phi^T r_i,$$

where r_i is the ith residual vector from the fit of the standard VAR model. From this, it follows immediately that, with fixed Φ, L_{21} is maximized over Ω at $\hat{\Omega} = \Phi^T \hat{\Sigma}_{\text{res}} \Phi$. Substituting $\hat{\Omega}$ back into L_{21} leads to $L_{31}(\mathcal{E}_\Sigma(\mathcal{B})) = -(n/2)\log|\Phi^T \hat{\Sigma}_{\text{res}} \Phi| - nd/2$. By similar reasoning, the value of Ω_0 that maximizes $L_{12}(\mathcal{E}_\Sigma(\mathcal{B}), \Omega_0)$ is $\hat{\Omega}_0 = \Phi_0^T \hat{\Sigma}_{Y_t} \Phi_0$. This leads to the partial maximization of L_{21} to be $L_{22}(\mathcal{E}_\Sigma(\mathcal{B})) = -(n/2)\log|\Phi_0^T \hat{\Sigma}_{Y_t} \Phi_0| - n(q-d)/2$.

Combining L_{31} and L_{22}, we arrive at the partially maximized form:

$$L_2(\mathcal{E}_\Sigma(\mathcal{B})) = -(nq/2)\log(2\pi) - nq/2 - (n/2)\log|\Phi^T \hat{\Sigma}_{\text{res}} \Phi| - (n/2)\log|\Phi_0^T \hat{\Sigma}_{Y_t} \Phi_0|.$$

The result enables us to conclude that $\hat{\mathcal{E}}_\Sigma(\mathcal{B})$ is equal to $\arg\max L_2(\mathcal{E}_\Sigma(\mathcal{B}))$. In addition, by Lemma 2.4 of Cook et al. (2010), suppose that $A \in R^{t\times t}$ is non-singular and that the column partitioned matrix $(\alpha, \alpha_0) \in R^{t\times t}$ is orthogonal, then $|\alpha_0^T A \alpha_0| = |A| + |\alpha^T A^{-1} \alpha|$. With the use of this result, given Φ, the conditional log likelihood can be written as

$$L_1 = -(nq/2)\log(2\pi) - nq/2 - (n/2)\log(\hat{\Sigma}_{Y_t}) - (n/2)\log|\Phi^T \hat{\Sigma}_{\text{res}} \Phi| - (n/2)\log|\Phi^T \hat{\Sigma}_{Y_t}^{-1} \Phi|.$$

Proof of Proposition 1

Let $\theta = (vec(\beta)^T, vech(\Sigma)^T)^T$ and $\tilde{\theta}$ be the standard MLE of θ. Under the envelope setting, $\theta = h(\psi)$. Let $F(\tilde{\theta}, \theta) = -L(h(\psi)) + L(\tilde{\theta})$, where L denotes the log likelihood function of the VAR model. Then it is easy to verify that $F(\tilde{\theta}, \theta)$ satisfies the assumptions of Proposition 3.1 in Shapiro (1986). As the envelope VAR model (5) *is overparameterized and $\sqrt{n}\left(\tilde{\theta} - \theta\right)$ converges to $N(0, \Delta)$, by applying Proposition 4.1 in Shapiro (1986), we have $\sqrt{n}\left(h(\hat{\psi}) - h(\psi)\right)$ converges in distribution to a multivariate normal random vector with mean zero and covariance matrix $H(H^T \Delta^{-1} H)^\dagger H^T$, where $H = \frac{\partial h(\psi)}{\partial \psi^T}$.*

To prove (ii), note that

$$\Delta - \Lambda = \Delta - H(H^T\Delta^{-1}H)^\dagger H^T = \Delta^{\frac{1}{2}}\left(I - \Delta^{-\frac{1}{2}} H(H^T\Delta^{-1}H)^\dagger H^T \Delta^{-\frac{1}{2}}\right)\Delta^{\frac{1}{2}} = \Delta^{\frac{1}{2}}\left(I - P_{\Delta^{-\frac{1}{2}}H}\right)\Delta^{\frac{1}{2}}$$
$$= \Delta^{\frac{1}{2}} Q_{\Delta^{-\frac{1}{2}}H} \Delta^{\frac{1}{2}}.$$

As $Q_{\Delta^{-\frac{1}{2}}H}$ is the projection matrix onto the orthogonal subspace of span($\Delta^{-\frac{1}{2}}H$), it is positive semi-definite. This completes the proof. □

Proof of Proposition 2

According to (16)*, it can be shown that*

$$H = \begin{pmatrix} I_{qp} \otimes \Phi & U^T \otimes I_q & 0 & 0 \\ 0 & 2C_q(\Phi\Omega \otimes I_q - \Phi \otimes \Phi_0\Omega_0\Phi_0^T) & C_q(\Phi \otimes \Phi)E_d & C_q(\Phi_0 \otimes \Phi_0)E_{q-d} \end{pmatrix},$$

where $E_q \in R^{q^2 \times q(q+1)/2}$ *is the expansion matrix such that* $\mathrm{vec}(A) = E_q \mathrm{vech}(A)$ *for any symmetric matrix* $A \in R^{q \times q}$*. Then by simplifying* Λ *and partitioning the results into two blocks, the asymptotic covariance matrix of* $\mathrm{vec}(\hat{\beta})$ *can be obtained.* □

Acknowledgments

The authors sincerely thank the editor, associate editor, and anonymous referees for their valuable comments that helped improve this manuscript. Ding's research is partially supported by DE-CTR ACCEL/NIH U54 GM104941 SHoRe award and the University of Delaware GUR award.

References

Anderson, T W (1963), 'The use of factor analysis in the statistical analysis of multiple time series', *Psychometrika*, **28**, 1–25.

Box, GEP & Tiao, GC (1977), 'A canonical analysis of multiple time series', *Biometrika*, **64**(2), 356–365.

Cook, RD, Forzani, L & Su, Z (2016), 'A note on fast envelope estimation', *Journal of Multivariate Analysis*, **150**, 42–54.

Cook, RD, Forzani, L & Zhang, X (2015), 'Envelopes and reduced-rank regression', *Biometrika*, **102**(2), 439–456.

Cook, RD, Helland, IS & Su, Z (2013), 'Envelopes and partial least squares regression', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **75**(5), 851–877.

Cook, RD, Li, B & Chiaromonte, F (2010), 'Envelope models for parsimonious and efficient multivariate linear regression', *Statistica Sinica*, **20**(3), 927–1010.

Ding, S & Cook, RD (2014), 'Dimension folding PCA and PFC', *Statistica Sinica*, **24**, 463–492.

Ding, S & Cook, RD (2015a), 'Tensor sliced inverse regression', *Journal of Multivariate Analysis*, **133**, 216–231.

Ding, S & Cook, RD (2015b), 'Higher-order sliced inverse regression', *Wiley Interdisciplinary Reviews: Computational Statistics*, **7**, 249–257.

Ding, S & Cook, RD (2018), 'Matrix variate regressions and envelope models', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **80**(2), 387–408.

Durbin, J & Watson, GS (1951), 'Testing for serial correlation in least square regression II', *Biometrika*, **38**, 159–177.

Henderson, VH & Searle, SR (1979), 'Vec and vech operators for matrices, with some users in Jacobians and multivariate statistics', *The Canadian Journal of Statistics*, **7**(1), 65–81.

Hoff, PD (2015), 'Multilinear tensor regression for longitudinal relational data', *The Annals of Applied Statistics*, **9**(3), 1169–1193.

(wileyonlinelibrary.com) https://doi.org/10.1002/sta4.203

Khare, K, Pal, S & Su, Z (2017), 'A Bayesian approach for envelope models', *The Annuals of Statistics*, **45**(1), 196–222.

Krawczak, M & Szkatula, G (2014), 'An approach to dimensionality reduction in time series', *Information Sciences*, **260**(1), 15–36.

Li, L & Zhang, X (2017), 'Parsimonious tensor response regression', *Journal of the American Statistical Association*, **112**, 1131–1146.

Matilainen, M, Croux, C, Nordhausen, K & Oja, H (2017), 'Supervised dimension reduction for multivariate time series', *Econometrics and Statistics*, **4**, 57–69.

Pan, J & Yao, Q (2008), 'Modelling multiple time series via common factors', *Biometrika*, **95**(2), 365–379.

Qian, W, Ding, S & Cook, RD (2018), 'Sparse minimum discrepancy approach to sufficient dimension reduction with simultaneous variable selection in ultrahigh dimension', *Journal of the American Statistical Association*. to appear https://doi.org/10.1080/01621459.2018.1497498.

Shapiro, A (1986), 'Asymptotic theory of overparameterized structural models', *Journal of the American Statistical Association*, **81**(393), 142–149.

Sims, CA (1980), 'Macroeconomics and reality', *Econometrica*, **48**(1), 1–48.

Su, Z & Cook, RD (2011), 'Partial envelopes for efficient estimation in multivariate linear regression', *Biometrika*, **98**(1), 133–146.

Su, Z, Zhu, G, Chen, X & Yang, Y (2016), 'Sparse envelope model: Efficient estimation and response variable selection in multivariate linear regression', *Biometrika*, **103**(3), 579–593.

Tsay, RS (2013), *Multivariate Time Series Analysis with R and Financial Applications*, *Wiley*, New York.

Velu, RP (1986), 'Reduced rank models for multiple time series', *Biometrika*, **73**(1), 105–118.