Lecture 3: Vector Autoregressive (VAR) Models

Reference: Chapter 2 of the textbook.

1 Introduction

Vector autoregressive models are perhaps the most widely used multivariate time series models. They are the dynamic version of multivariate multiple linear regressions commonly used in multivariate statistical analysis. We start with stationary VAR models. As before, in this course, \( \{a_t\} \) is a sequence of serially uncorrelated random vectors with mean zero and positive-definite covariance matrix \( \Sigma_a \), which is denoted by \( \Sigma_a > 0 \).

A \( k \)-dimensional time series follows a VAR(\( p \)) model if

\[
\mathbf{z}_t = \phi_0 + \sum_{i=1}^{p} \phi_i \mathbf{z}_{t-i} + \mathbf{a}_t,
\]

where \( \phi_0 \) is a \( k \)-dimensional constant vector and \( \phi_i \) are \( k \times k \) real-valued matrices. For a VAR(\( p \)) model, we shall discuss the following topics:

1. Model structure and Granger causality
2. Relation to transfer function models (distributed lag model)
3. Stationarity condition
4. Invertibility condition
5. Moment equations: multivariate Yule-Walker equations
6. Implied component models
7. MA representation

2 VAR(1) Models

3 VAR(2) Models

4 VAR(\( p \)) Models

5 Estimation

[This handout follows Chapter 2 of Tsay (2014).]
Given the data, we have, for a VAR($p$) model,

$$z_t = \phi_0 + \phi_1 z_{t-1} + \cdots + \phi_p z_{t-p} + a_t, \quad t = p+1, \ldots, T.$$  

1. Least squares estimates (generalized LSE)

A VAR($p$) model can be written as

$$z_t' = x_t' \beta + a_t',$$

where $x_t = (1, z_{t-1}', \ldots, z_{t-p}')'$ is a $(kp+1)$-dimensional vector and $\beta' = [\phi_0, \phi_1, \ldots, \phi_p]$ is a $k \times (kp+1)$ matrix. Putting in matrix form, the data for estimation is

$$Z = X \beta + A,$$  

where $Z$ is a $(T-p) \times k$ matrix with $i$-row being $z_{p+i}'$, $X$ is a $(T-p) \times (kp+1)$ design matrix with $i$th row being $x_{p+i}'$, and $A$ is a $(T-p) \times k$ matrix with $i$th row being $a_{p+i}'$. The matrix representation in Eq. (2) is particularly convenient for VAR($p$) model. For example, column $j$ of $\beta$ contains parameters associated with $z_{jt}$. From Eq. (2), we obtain

$$\text{vec}(Z) = (I_k \otimes X)\text{vec}(\beta) + \text{vec}(A).$$  

Note that the covariance matrix of vec($A$) is $\Sigma_a \otimes I_{T-p}$.

The GLS estimate of $\beta$ is obtained by minimizing

$$S(\beta) = [\text{vec}(A)]'(\Sigma_a \otimes I_{T-p})^{-1}\text{vec}(A)$$

$$\quad = [\text{vec}(Z - X\beta)]'(\Sigma_a^{-1} \otimes I_{T-p})\text{vec}(Z - X\beta)$$

$$\quad = tr[(Z - X\beta)\Sigma_a^{-1}(Z - X\beta)'].$$

The last equality holds because $\Sigma_a$ is a symmetric matrix and we use $tr(DBC) = \text{vec}(C')'(B' \otimes I)\text{vec}(D)$. From Eq. (4), we have

$$S(\beta) = [\text{vec}(Z) - (I_k \otimes X)\text{vec}(\beta)]'(\Sigma_a^{-1} \otimes I_{T-p})[\text{vec}(Z) - (I_k \otimes X)\text{vec}(\beta)]$$

$$\quad = [\text{vec}(Z)' - \text{vec}(\beta)'(I_k \otimes X')][\Sigma_a^{-1} \otimes I_{T-p}][\text{vec}(Z) - (I_k \otimes X)\text{vec}(\beta)]$$

$$\quad = \text{vec}(Z)'(\Sigma_a^{-1} \otimes I_{T-p})\text{vec}(Z) - 2\text{vec}(\beta)'(\Sigma_a^{-1} \otimes X')\text{vec}(Z)$$

$$\quad + \text{vec}(\beta)'(\Sigma_a^{-1} \otimes X'X)\text{vec}(\beta).$$

Taking partial derivatives of $S(\beta)$ with respect to vec($\beta$), we obtain

$$\frac{\partial S(\beta)}{\partial \text{vec}(\beta)} = -2(\Sigma_a^{-1} \otimes X')\text{vec}(Z) + 2(\Sigma_a^{-1} \otimes X'X)\text{vec}(\beta).$$

Equating to zero gives the normal equations

$$(\Sigma_a^{-1} \otimes X'X)\text{vec}(\beta) = (\Sigma_a^{-1} \otimes X')\text{vec}(Z).$$

Consequently, the GLS estimate of an VAR($p$) model is

$$\text{vec}(\hat{\beta}) = (\Sigma_a^{-1} \otimes X'X)^{-1}(\Sigma_a^{-1} \otimes X')\text{vec}(Z)$$

$$\quad = [I_k \otimes (X'X)^{-1}](\Sigma_a^{-1} \otimes X')\text{vec}(Z)$$

$$\quad = [I_k \otimes (X'X)^{-1}X]\text{vec}(Z).$$

$$\quad = \text{vec}((X'X)^{-1}(X'Z)).$$
where the last equality holds because \( \text{vec}(DB) = (I \otimes D)\text{vec}(B) \). In other words, we obtain

\[
\hat{\beta} = (X'X)^{-1}(X'Z) = \left[ \sum_{t=p+1}^{T} x_t x_t' \right]^{-1} \sum_{t=p+1}^{T} x_t z_t',
\]

(9)

which interestingly does not depend on \( \Sigma_a \).

**Remark.** The result in Eq. (9) shows that one can obtain the GLS estimate of a VAR\((p)\) model equation-by-equation. That is, one can consider the \( k \) multiple linear regressions of \( z_{it} \) on \( x_t \) separately, where \( i = 1, \ldots, k \). This estimation method is convenient when one considers parameter constraints in a VAR\((p)\) model.

**Ordinary Least Squares (OLS) Estimate**

Readers may notice that the GLS estimate of VAR\((p)\) model in Eq. (9) is identical to that of the ordinary least squares estimate of the multivariate multiple linear regression in Eq. (2). Replacing \( \Sigma_a \) in Eq. (5) by \( I_k \), we have the objective function of the ordinary least squares estimation

\[
S_o(\beta) = \text{tr}[(Z - X\beta)(Z - X\beta)'].
\]

(10)

The derivations discussed above continue to hold step-by-step with \( \Sigma_a \) replaced by \( I_k \). One thus obtains the same estimate given in Eq. (9) for \( \beta \). The fact that the GLS estimate is the same as the OLS estimate for a VAR\((p)\) model was first shown in Zellner (1962). In what follows, we refer to the estimate in Eq. (9) simply as the least squares (LS) estimate.

The residual of the LS estimate is

\[
\hat{a}_t = z_t - \hat{\phi}_0 - \sum_{i=1}^{p} \hat{\phi}_i z_{t-i}, \quad t = p + 1, \ldots, T
\]

and let \( \hat{A} \) be the residual matrix, i.e. \( \hat{A} = Z - X\hat{\beta} = [I_{T-p} - X(X'X)^{-1}X']Y \). The LS estimate of the innovational covariance matrix \( \Sigma_a \) is

\[
\hat{\Sigma}_a = \frac{1}{T - (k + 1)p} \sum_{t=p+1}^{T} \hat{a}_t \hat{a}_t' = \frac{1}{T - (k + 1)p - 1} \hat{A}' \hat{A},
\]

where the denominator is determined by \( [T - p - (kp + 1)] \), which is the effective sample size less the number of parameters in the equation for each component \( z_{it} \). By Eq. (2), we see that

\[
\hat{\beta} - \beta = (X'X)^{-1}X' \hat{A}.
\]

(11)

Since \( E(\hat{A}) = 0 \), we see that the LS estimate is an unbiased estimator. The LS estimate of a VAR\((p)\) model has the following properties.

**Theorem 2.1.** For the stationary VAR\((p)\) model, assume that \( a_t \) are independent and identically distributed with mean zero and positive definite covariance matrix \( \Sigma_a \). Then, (i)
$E(\hat{\beta}) = \beta$, where $\beta$ is defined in Eq. (2), (ii) $E(\hat{\Sigma}_a) = \Sigma_a$, (iii) the residual \( \hat{A} \) and the LS estimate \( \beta \) are uncorrelated, and (iv) the covariance of the parameter estimates is

$$\text{Cov}[\text{vec}(\hat{\beta})] = \hat{\Sigma}_a \otimes (X'X)^{-1} = \hat{\Sigma}_a \otimes \left( \sum_{t=p+1}^{T} x_t x'_t \right)^{-1}.$$  

2. Maximum likelihood estimates under normality

Assume further that $a_t$ of the VAR($p$) model follows a multivariate normal distribution. Let $z_{h:q}$ denote the observations from $t = h$ to $t = q$ (inclusive). The conditional likelihood function of the data can be written as

$$L(z_{(p+1):T}|z_{1:p}, \beta, \Sigma_a) = \prod_{t=p+1}^{T} p(z_t|z_{1:(t-1)}, \beta, \Sigma_a)$$

$$= \prod_{t=p+1}^{T} p(a_t|z_{1:(t-1)}, \beta, \Sigma_a)$$

$$= \prod_{t=p+1}^{T} p(a_t|\beta, \Sigma_a)$$

$$= \prod_{t=p+1}^{T} \frac{1}{(2\pi)^{k/2}|\Sigma_a|^{1/2}} \exp\left[ -\frac{1}{2} a'_t \Sigma_a^{-1} a_t \right]$$

$$\propto |\Sigma_a|^{-(T-p)/2} \exp \left[ -\frac{1}{2} \sum_{t=p+1}^{T} tr(a'_t \Sigma_a^{-1} a_t) \right].$$

The log-likelihood function then becomes

$$\ell(\beta, \Sigma_a) = c - \frac{T-p}{2} \log(|\Sigma_a|) - \frac{1}{2} \sum_{t=p+1}^{T} tr(a'_t \Sigma_a^{-1} a_t)$$

$$= c - \frac{T-p}{2} \log(|\Sigma_a|) - \frac{1}{2} tr \left( \Sigma_a^{-1} \sum_{t=p+1}^{T} a_t a'_t \right),$$

where $c$ is a constant, and we use the properties that $tr(CD) = tr(DC)$ and $tr(C + D) = tr(C + D)$. Noting that $\sum_{t=p+1}^{T} a_t a'_t = A'A$, where $A = Z - X\beta$ is the error matrix in Eq. (2), we can rewrite the log-likelihood function as

$$\ell(\beta, \Sigma_a) = c - \frac{T-p}{2} \log(|\Sigma_a|) - \frac{1}{2} S(\beta),$$

where $S(\beta)$ is given in Eq. (5).

Since the parameter matrix $\beta$ only appears in the last term of $\ell(\beta, \Sigma_a)$, maximizing the log-likelihood function over $\beta$ is equivalent to minimizing $S(\beta)$. Consequently, the maximum likelihood (ML) estimate of $\beta$ is the same as its LS estimate. Next, taking the partial derivative of the log-likelihood function with respective to $\Sigma_a$, we obtain

$$\frac{\partial \ell(\hat{\beta}, \Sigma_a)}{\partial \Sigma_a} = -\frac{T-p}{2} \Sigma_a^{-1} + \frac{1}{2} \Sigma_a^{-1} \hat{A}'\hat{A} \Sigma_a^{-1}.$$  

(13)
Equating the prior normal equation to zero, we obtain the maximum likelihood estimate of $\Sigma_a$ as

$$\hat{\Sigma}_a = \frac{1}{T-p} \hat{A}' \hat{A} = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{a}_t \hat{a}_t'.$$

(14)

This result is the same as that for the multiple linear regression. The ML estimate of $\Sigma_a$ is only asymptotically unbiased. Finally, the Hessian matrix of $\beta$ can be obtained by taking the partial derivative of Eq. (7), namely

$$-\frac{\partial^2 \ell(\beta, \Sigma_a)}{\partial \text{vec}(\beta) \partial \text{vec}(\beta)} = \frac{1}{2} \frac{\partial^2 S(\beta)}{\partial \text{vec}(\beta) \partial \text{vec}(\beta)'} = \Sigma_a^{-1} \otimes X'X.$$

Inverse of the Hessian matrix provides the asymptotic covariance matrix of the ML estimate of vec($\beta$). Next, taking derivative of Eq. (13), we obtain

$$\frac{\partial^2 \ell(\hat{\beta}, \Sigma_a)}{\partial \text{vec}(\Sigma_a) \partial \text{vec}(\Sigma_a)'} = \frac{T-p}{2} (\Sigma_a^{-1} \otimes \Sigma_a^{-1}) - \frac{1}{2} (|\Sigma_a^{-1} \otimes \Sigma_a^{-1}| \hat{A}' \hat{A} \Sigma_a^{-1})$$

$$- \frac{1}{2} \Sigma_a^{-1} \hat{A}' \hat{A} (\Sigma_a^{-1} \otimes \Sigma_a^{-1}).$$

Consequently, we have

$$-E \left( \frac{\partial^2 \ell(\hat{\beta}, \Sigma_a)}{\partial \text{vec}(\Sigma_a) \partial \text{vec}(\Sigma_a)'} \right) = \frac{T-p}{2} (\Sigma_a^{-1} \otimes \Sigma_a^{-1}).$$

This result provides asymptotic covariance matrix for the ML estimates of elements of $\Sigma_a$.

**Theorem 2.2.** Suppose that the innovation $a_t$ of a stationary VAR($p$) model follows a multivariate normal distribution with mean zero and positive-definite covariance matrix $\Sigma_a$. Then, the maximum likelihood estimates are vec($\hat{\beta}$) = $(X'X)^{-1} X'Z$ and $\hat{\Sigma}_a = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{a}_t \hat{a}_t'$. Also, (i) $(T-p)\hat{\Sigma}_a$ is distributed as $W_{k,T-(k+1)p-1}(\Sigma_a)$, a Wishart distribution, and (ii) vec($\hat{\beta}$) is normally distributed with mean vec($\beta$) and covariance matrix $\Sigma_a \otimes (X'X)^{-1}$, and (iii) vec($\hat{\beta}$) is independent of $\hat{\Sigma}_a$, where $Z$ and $X$ are defined in Eq. (2). Furthermore, $\sqrt{T}[\text{vec}(\hat{\beta}) - \beta]$ and $\sqrt{T}[\text{vec}(\hat{\Sigma}_a) - \text{vec}(\Sigma_a)]$ are asymptotically normally distributed with mean zero and covariance matrices $\Sigma_a \otimes G^{-1}$ and $2\Sigma_a \otimes \Sigma_a$, respectively, where $G = E(x_t x_t')$ with $x_t$ defined in Eq. (2).

Finally, given the data set $\{z_1, \ldots, z_T\}$, the maximized likelihood of a VAR($p$) model is

$$L(\hat{\beta}, \Sigma_a|z_{1:p}) = (2\pi)^{-k(T-p)/2} |\hat{\Sigma}_a|^{-(T-p)/2} \exp[-\frac{k(T-p)}{2}].$$

(15)

This value is useful in likelihood ratio tests to be discussed later.

3. Properties of the estimates

Assume that the fourth moments of $a_t$ are finite. Specifically, $a_t = (a_{1t}, \ldots, a_{kt})'$ is continuous and satisfies

$$E|a_{it}a_{jt}a_{ut}a_{vt}| < \infty, \quad \text{for } i, j, u, v = 1, \ldots, k \text{ and all } t.$$
Lemma 2.3. If the VAR($p$) process $z_t$ of Eq. (1) is stationary and satisfies the condition in Eq. (16), then, as $T \to \infty$, we have

(i) $\frac{X'X}{(T-p)} \to_p G$,

(ii) $\frac{1}{\sqrt{T-p}} \text{vec}(X'A) = \frac{1}{\sqrt{T-p}}(I_k \otimes X') \text{vec}(A) \to_d N(0, \Sigma_a \otimes G)$,

where $\to_p$ and $\to_d$ denote convergence in probability and distribution, respectively, $X$ and $A$ are defined in Eq. (2), and $G$ is a nonsingular matrix given by

$$G = \begin{bmatrix} 1 & 0' \\ 0 & \Gamma_0^* \end{bmatrix} + \begin{bmatrix} 0 \\ u \end{bmatrix} [0, u']$$

where $0$ is a $kp$-dimensional vector of zeros, $\Gamma_0^*$ is defined below

\begin{equation}
\Gamma_0^* = \begin{bmatrix} \Gamma_0 & \Gamma_1 & \cdots & \Gamma_{p-1} \\ \Gamma_1' & \Gamma_0 & \cdots & \Gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{p-1}' & \Gamma_{p-2}' & \cdots & \Gamma_0 \end{bmatrix}_{kp \times kp},
\end{equation}

and $u = 1_p \otimes \mu$ with $1_p$ being a $p$-dimensional vector of 1.

Theorem 2.3. Suppose that the VAR($p$) time series $z_t$ in Eq. (1) is stationary and its innovation $a_t$ satisfies the assumption in Eq. (16). Then, as $T \to \infty$,

(i) $\hat{\beta} \to_p \beta$,

(ii) $\sqrt{T-p}[\text{vec}(\hat{\beta}) - \text{vec}(\beta)] = \sqrt{T-p}[\text{vec}(\hat{\beta} - \beta)] \to_d N(0, \Sigma_a \otimes G^{-1})$,

where $G$ is defined in Lemma 2.3.

Proof. By Eq. (11), we have

$$\hat{\beta} - \beta = \left( \frac{X'X}{T-p} \right)^{-1} \left( \frac{X'A}{T-p} \right) \to_p 0,$$

because the last term approaches $0$. This establishes the consistency of $\hat{\beta}$. For result (ii), we can use Eq. (8) to obtain

$$\sqrt{T-p}[\text{vec}(\hat{\beta}) - \text{vec}(\beta)] = \sqrt{T-p}[I_k \otimes (X'X)^{-1}X'] \text{vec}(A)\right) = \sqrt{T-p}[I_k \otimes (X'X)^{-1}][I_k \otimes X'] \text{vec}(A)$$

$$= \left[I_k \otimes \left( \frac{X'X}{T-p} \right)^{-1} \right] \frac{1}{\sqrt{T-p}}[I_k \otimes X'] \text{vec}(A).$$

Therefore, the limiting distribution of $\sqrt{T-p}[\text{vec}(\hat{\beta}) - \text{vec}(\beta)]$ is the same as that of

$$\left(I_k \otimes G^{-1}\right) \frac{1}{\sqrt{T-p}}[I_k \otimes X'] \text{vec}(A).$$
Hence, by Lemma 2.3, the limiting distribution of \( \sqrt{T-p} [\text{vec}(\hat{\beta}) - \text{vec}(\beta)] \) is normal and the covariance matrix is

\[
(I_k \otimes G^{-1})(\Sigma_a \otimes G)(I_k \otimes G^{-1}) = \Sigma_a \otimes G^{-1}.
\]

The proof is complete.

4. Bayesian estimation, including \textit{Minnesota prior} of Litterman (1986)

6 \hspace{1em} \textbf{Order selection}

1. Sequential likelihood ratio tests

2. Information criteria: AIC, BIC, and HQ

7 \hspace{1em} \textbf{Model checking}

1. Residual cross-correlation matrices and their properties

Let \( \hat{A} = Z - X\hat{\beta} \) be the residual matrix of a fitted VAR(p) model, using the notation in Eq. (2). The ith row of \( \hat{A} \) is \( \hat{a}_{p+i} = z_{p+i} - \hat{\phi}_0 - \sum_{i=1}^{p} \hat{\phi}_i z_{t-i} \). The lag-\( \ell \) cross covariance matrix of the residual series is defined as

\[
\hat{C}_\ell = \frac{1}{T-p} \sum_{t=p+\ell+1}^{T} \hat{a}_t \hat{a}_{t-\ell}'.
\]

Note that \( \hat{C}_0 = \hat{\Sigma}_a \) is the residual covariance matrix. In matrix notation, we can rewrite the lag-\( \ell \) residual cross-covariance matrix \( \hat{C}_\ell \) as

\[
\hat{C}_\ell = \frac{1}{T-p} \hat{A}' B^\ell \hat{A}, \quad \ell \geq 0
\]

where \( B \) is a \( (T-p) \times (T-p) \) back-shift matrix defined as

\[
B = \begin{bmatrix}
0 & 0_{T-p-1}' \\
I_{T-p-1} & 0_{T-p-1}'
\end{bmatrix},
\]

where \( 0_h \) is the \( h \)-dimensional vector of zero. The lag-\( \ell \) residual cross-correlation matrix is define as

\[
\hat{R}_\ell = \hat{D}^{-1} \hat{C}_\ell \hat{D}^{-1},
\]

where \( \hat{D} \) is the diagonal matrix of the standard errors of the residual series, that is, \( \hat{D} = \sqrt{\text{diag}(\hat{C}_0)} \). In particular, \( \hat{R}_0 \) is the residual correlation matrix.

For model checking, we consider the asymptotic joint distribution of the residual cross-covariance matrices \( \hat{\Xi}_m = [\hat{C}_1, \ldots, \hat{C}_m] \). Using the notation in Eq. (18), we have

\[
\hat{\Xi}_m = \frac{1}{T-p} \hat{A}' [B\hat{A}, B^2\hat{A}, \ldots, B^m\hat{A}] = \frac{1}{T-p} \hat{A}' B_m (I_m \otimes \hat{A}),
\]
where $B_m = [B, B^2, \ldots, B^m]$ is a $(T - p) \times m(T - p)$ matrix.

Since $\hat{A} = A - X(\hat{\beta} - \beta)$ and letting $T_p = T - p$, we have

$$T_p \hat{\Xi}_m = A' B_m (I_m \otimes A) - A' B_m [I_m \otimes X(\hat{\beta} - \beta)] - (\hat{\beta} - \beta)' X' B_m (I_m \otimes A) + (\hat{\beta} - \beta)' X' B_m [I_m \otimes X(\hat{\beta} - \beta)].$$

**Lemma 2.4.** Suppose that $z_t$ follows a stationary VAR($p$) model of Eq. (1) with $a_t$ being a white noise process with mean zero and positive covariance matrix $\Sigma_a$. Also, assume that the assumption in Eq. (16) holds and the parameter matrix $\beta$ of the model in Eq. (2) is consistently estimated via methods discussed before and the residual cross covariance matrix is defined in Eq. (18). Then, \( \sqrt{T_p} \text{vec}(\hat{\Xi}_m) \) has the same limiting distribution as

$$\sqrt{T_p} \text{vec}(\hat{\Xi}_m) - \sqrt{T_p} H \text{vec}[(\hat{\beta} - \beta)'],$$

where $\hat{\beta}$ is the theoretical counterpart of $\hat{\beta}_m$ obtained by dividing the first term of Eq. (21) by $T_p$, and $H = H' \otimes I_k$ with $H = H' \otimes I_k$ with

$$H_\ast = \begin{bmatrix} 0' & 0' & \cdots & 0' \\ \Sigma_a & \psi_1 \Sigma_a & \cdots & \psi_{m-1} \Sigma_a \\ 0_k & \Sigma_a & \cdots & \psi_{m-2} \Sigma_a \\ \vdots & \vdots & \cdots & \vdots \\ 0_k & 0_k & \cdots & \psi_m \Sigma_a \end{bmatrix}_{(kp+1) \times km},$$

where $0$ is a $k$-dimensional vector of zero, $0_k$ is a $k \times k$ matrix of zero and $\psi_i$ are the coefficient matrices of the MA representation of the VAR($p$) model.

**Lemma 2.5.** Assume that $z_t$ is a stationary VAR($p$) series satisfying the conditions of Lemma 2.4, then

$$\begin{bmatrix} \frac{1}{\sqrt{T_p}} \text{vec}(A' X) \\ \sqrt{T_p} \text{vec}(\Xi_m) \end{bmatrix} \rightarrow_d N \left( 0, \begin{bmatrix} G & H_\ast \\ H_\ast' & I_m \otimes \Sigma_a \end{bmatrix} \otimes \Sigma_a \right),$$

where $G$ is defined in Lemma 2.3 and $H_\ast$ is defined in Lemma 2.4.

**Theorem 2.4.** Suppose that $z_t$ follows a stationary VAR($p$) model of Eq. (1) with $a_t$ being a white noise process with mean zero and positive covariance matrix $\Sigma_a$. Also, assume that the assumption in Eq. (16) holds and the parameter matrix $\beta$ of the model in Eq. (2) is consistently estimated by a method discussed before and the residual cross covariance matrix is defined in Eq. (18). Then,

$$\sqrt{T_p} \text{vec}(\hat{\Xi}_m) \rightarrow_d N(0, \Sigma_{c,m}),$$

where

$$\Sigma_{c,m} = (I_m \otimes \Sigma_a - H_\ast' G^{-1} H_\ast) \otimes \Sigma_a$$

$$= I_m \otimes \Sigma_a \otimes \Sigma_a - \hat{H} (\Gamma_0^\ast)^{-1} \otimes \Sigma_a) \hat{H}'.$$
where $H_*$ and $G$ are defined in Lemma 2.5, $\Gamma_0$ is the expanded covariance matrix defined in Eq. (17), and $\bar{H} = \bar{H}_* \otimes I_k$ with $\bar{H}_*$ being a submatrix of $H_*$ with the first row of zeros removed.

Let $D$ be the diagonal matrix of the standard errors of the components of $a_t$, that is, $D = \text{diag}\{\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{kk}}\}$, where $\Sigma_a = [\sigma_{ij,a}]$.

**Theorem 2.5.** Assume that the conditions of Theorem 2.4 hold. Then,

$$\sqrt{T} p \text{vec}(\hat{\xi}_m) \rightarrow_d N(0, \Sigma_{r,m}),$$

where $\Sigma_{r,m} = [(I_m \otimes R_0) - H_0' G^{-1} H_0] \otimes R_0$, where $R_0$ is the lag-0 cross-correlation matrix of $a_t$, $H_0 = H_*(I_m \otimes D^{-1})$, and $G$, as before, is defined in Lemma 2.3.

2. Multivariate portmanteau statistics

Let $R_\ell$ be the theoretical lag-$\ell$ cross-correlation matrix of residuals $a_t$. The hypothesis of interest in model checking is

$$H_0 : R_1 = \cdots = R_m = 0 \quad \text{versus} \quad H_a : R_j \neq 0 \quad \text{for some } 1 \leq j \leq m,$$

where $m$ is a pre-specified positive integer. The Portmanteau statistic is often used to perform the test. For residual series, the statistic becomes

$$Q_k(m) = T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} tr(\hat{R}_\ell' \hat{R}_0^{-1} \hat{R}_\ell \hat{R}_0^{-1})$$

$$= T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} tr(\hat{R}_\ell' \hat{R}_0^{-1} \hat{R}_\ell \hat{R}_0^{-1} D^{-1} D)$$

$$= T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} tr(\hat{C}_\ell' \hat{C}_0^{-1} \hat{C}_\ell \hat{C}_0^{-1})$$

$$= T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} tr(\hat{C}_\ell' \hat{C}_0^{-1} \hat{C}_\ell \hat{C}_0^{-1}). \quad (23)$$

**Theorem 2.6.** Suppose that $z_t$ follows a stationary VAR($p$) model of Eq. (1) with $a_t$ being a white noise process with mean zero and positive covariance matrix $\Sigma_a$. Also, assume that the assumption in Eq. (16) holds and the parameter matrix $\beta$ of the model in Eq. (2) is consistently estimated by a method discussed earlier and the residual cross covariance matrix is defined in Eq. (18). Then, the test statistic $Q_k(m)$ is asymptotically distributed as a chi-square distribution with $(m - p) k^2$ degrees of freedom.

3. Model simplification and refinements

**Testing Zero Parameters**

Let $\hat{\omega}$ is a $v$-dimensional vector consisting of the target parameters. In other words, $v$ is the
number of parameters to be fixed to zero. Let \( \omega \) be the counterpart of \( \hat{\omega} \) in the parameter matrix \( \beta \) in Eq. (2). The hypothesis of interest is

\[
H_0 : \omega = 0 \quad \text{versus} \quad H_a : \omega \neq 0.
\]

Clearly, there exists a \( v \times k(p + 1) \) locating matrix \( K \) such that

\[
K \text{vec}(\beta) = \omega, \quad \text{and} \quad K \text{vec}(\hat{\beta}) = \hat{\omega}.
\]

By Theorem 2.3 and properties of multivariate normal distribution, we have

\[
\sqrt{T_p}(\hat{\omega} - \omega) \rightarrow d N[0, K(\Sigma_a \otimes G^{-1})K'],
\]

where \( T_p = T - p \) is the effective sample size. Consequently, under \( H_0 \), we have

\[
T_p\hat{\omega}'[K(\Sigma_a \otimes G^{-1})K']^{-1}\hat{\omega} \rightarrow d \chi^2_v,
\]

where \( v = \text{dim}(\omega) \).

8 Linear constraints and Granger causality test

The prior test can be generalized to linear constraints, namely

\[
\text{vec}(\beta) = J\gamma + r,
\]

where \( J \) is \( k(p + 1) \times P \) constant matrix of rank \( P \), \( r \) is a \( k(p + 1) \)-dimensional constant vector, and \( \gamma \) denotes a \( P \)-dimensional vector of unknown parameters.

9 Forecasting

1. Mean squares errors

We can use VAR(1) model to demonstrate forecasting and mean-reverting of stationary VAR processes.

2. Forecasts with parameter uncertainty

With estimated parameters, we have \( \ell \)-step ahead forecast error

\[
\hat{e}_h(\ell) = z_{h+\ell} - \hat{z}_h(\ell) = z_{h+\ell} - z_h(\ell) + z_h(\ell) - \hat{z}_h(\ell) = e_h(\ell) + [z_h(\ell) - \hat{z}_h(\ell)].
\]

The two terms of the forecast errors \( \hat{e}_h(\ell) \) are therefore uncorrelated and we have

\[
\text{Cov}[\hat{e}_h(\ell)] = \text{Cov}[e_h(\ell)] + E[[z_h(\ell) - \hat{z}_h(\ell)][z_h(\ell) - \hat{z}_h(\ell)]]
\equiv \text{Cov}[e_h(\ell)] + \text{MSE}[z_h(\ell) - \hat{z}_h(\ell)],
\]

where the notation \( \equiv \) is used to denote equivalence. Letting \( T_p = T - p \) denote the effective sample size in estimation, we assume that the parameter estimates satisfy

\[
\sqrt{T_p\text{vec}(\hat{\beta}' - \beta')} \rightarrow d N(0, \Sigma_{\beta'}).\]
Since $z_h(\ell)$ is a differentiable function of vec$(\beta')$, one can show that

$$
\sqrt{T_p}[\hat{z}_h(\ell) - z_h(\ell)|F_h] \rightarrow_d N\left(0, \frac{\partial z_h(\ell)}{\partial \text{vec}(\beta')'} \Sigma_\beta \frac{\partial z_h(\ell)'}{\partial \text{vec}(\beta')'}\right).
$$

This result suggests that we can approximate the mean square error (MSE) in Eq. (29) by

$$
\Omega_\ell = E\left[\frac{\partial z_h(\ell)}{\partial \text{vec}(\beta')'} \Sigma_\beta \frac{\partial z_h(\ell)'}{\partial \text{vec}(\beta')'}\right].
$$

If we further assume that $a_t$ is multivariate normal, then we have

$$
\sqrt{T_p}[\hat{z}_h(\ell) - z_h(\ell)] \rightarrow_d N(0, \Omega_\ell).
$$

Consequently, we have

$$
\text{Cov}[\hat{e}_h(\ell)] = \text{Cov}[e_h(\ell)] + \frac{1}{T_p} \Omega_\ell. \tag{30}
$$

It remains to derive the quantity $\Omega_\ell$. To this end, we need to obtain the derivatives $\frac{\partial z_h(\ell)}{\partial \text{vec}(\beta')'}$.

Using VAR($p$) model, we have

$$
z_h(\ell) = JP^\ell x_h, \quad \ell \geq 1, \tag{31}
$$

where

$$
P = \begin{bmatrix} 1 & 0_{kp}^t \\ \nu & \Phi \end{bmatrix}_{(kp+1) \times (kp+1)}, \quad J = [0_k, I_k, 0_{k \times (p-1)}]_{k \times (kp+1)};
$$

and $\nu = [\phi_0', 0_{k(p-1)}]'$, where $\Phi$ is the companion matrix of $\phi(B)$ defined before, $0_m$ is an $m$-dimensional vector of zero, and $0_{m \times n}$ is an $m \times n$ matrix of zero. This is a generalized version of the $\ell$-step ahead forecast of a VAR(1) model to include the constant vector $\phi_0$ in the recursion and it can be shown by mathematical induction. Using Eq. (31), we have

$$
\frac{\partial z_h(\ell)}{\partial \text{vec}(\beta')'} = \frac{\partial \text{vec}(JP^\ell x_h)}{\partial \text{vec}(\beta')'} = (x_h' \otimes J) \frac{\partial \text{vec}(P^\ell)}{\partial \text{vec}(\beta')'}
$$

$$
= (x_h' \otimes J) \left[ \sum_{i=0}^{\ell-1} (P^i)' \otimes P_{\ell-i} \right] \frac{\partial \text{vec}(P^\ell)}{\partial \text{vec}(\beta')'}
$$

$$
= (x_h' \otimes J) \left[ \sum_{i=0}^{\ell-1} (P^i)' \otimes P_{\ell-i} \right] (I_{kp+1} \otimes J')
$$

$$
= \sum_{i=0}^{\ell-1} x_h(P^i)' \otimes J P^i J'
$$

$$
= \sum_{i=0}^{\ell-1} x_h(P^i)' \otimes \psi_i,
$$

where
where we have used the fact that $J \mathbf{P}^i \mathbf{J} = \psi_i$. Using the least squares estimate $\hat{\beta}$, we have, via Eq. (??), $\Sigma_{\beta'} = G^{-1} \otimes \Sigma_a$. Therefore,

$$\Omega_\ell = E \left[ \frac{\partial z_h(\ell)}{\partial \text{vec}(\beta')} (G^{-1} \otimes \Sigma_a) \frac{\partial z_h(\ell)'}{\partial \text{vec}(\beta')} \right]$$

$$= \sum_{i=0}^{\ell-1} \sum_{j=0}^{\ell-1} E[x_h'(P')^{\ell-1-i}G^{-1}P^{\ell-1-j}x_h] \otimes \psi_i \Sigma_a \psi_j'$$

$$= \sum_{i=0}^{\ell-1} \sum_{j=0}^{\ell-1} E[tr(x_h'(P')^{\ell-1-i}G^{-1}P^{\ell-1-j}x_h)] \psi_i \Sigma_a \psi_j'$$

$$= \sum_{i=0}^{\ell-1} \sum_{j=0}^{\ell-1} tr[(P')^{\ell-1-i}G^{-1}P^{\ell-1-j}E(x_hx_h')] \psi_i \Sigma_a \psi_j'$$

$$= \sum_{i=0}^{\ell-1} \sum_{j=0}^{\ell-1} tr[(P')^{\ell-1-i}G^{-1}P^{\ell-1-j}G] \psi_i \Sigma_a \psi_j'. \quad (32)$$

In particular, if $\ell = 1$, then

$$\Omega_1 = tr(I_{kp+1})\Sigma_a = (kp + 1)\Sigma_a,$$

and

$$\text{Cov}[\hat{z}_h(1)] = \Sigma_a + \frac{kp + 1}{T_p} \Sigma_a = \frac{T_p + kp + 1}{T_p} \Sigma_a.$$ 

What is the main implication of the result?

### 10 Impulse response function

It is also known as the multiplier analysis.

1. Definition, cumulative impulse response function

Consider the MA representation of a vector time series (assuming zero mean),

$$z_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots.$$

Suppose we like to study the impact of $a_0 = (1, 0, \ldots, 0)'$ on the system. That is, we like to know what would happen to the system when the shock $a_{10} = 1$ and all the other shocks are zero. From the model, it is easily seen that

$$z_0 = a_0, \quad z_1 = \psi_1 a_0 = \psi_1[1], \quad z_2 = \psi_2 a_0 = \psi_2[1], \cdots$$

where $\psi_i[1]$ denotes the 1-th column of $\psi_i$. The cumulative impact of $a_0$ on the system to time $t = n$ is then

$$\psi_n[1] = \sum_{i=0}^{n} \psi_i[1].$$
The same argument applies to shock occurring at the $i$th component of $a_0$. Consequently, the cumulative impact on the system to time $t = n$ can be written as

$$\psi_n = \sum_{i=0}^{n} \psi_i,$$

and the total multipliers or long-run effects is

$$\psi_\infty = \sum_{i=0}^{\infty} \psi_i.$$

2. Orthogonal innovations The statement of the prior subsection implicitly assumes there is no correlation in the components of $a_0$. In real applications, components of $a_t$ are typically correlated so that one cannot change one component without affecting the others. A more formal definition of impact of $a_{1t}$ on the future observation $z_{t+j}$ for $j > 0$, is

$$\text{effect} = \frac{\partial z_{t+j}}{\partial a_{1t}} = \psi_j \frac{\partial a_t}{\partial a_{1t}} = \psi_j \Sigma_a [1, 1] \sigma_{a,11}^{-1},$$

under the linear model framework. In the above, I use the simple linear regression relation $a_{it} = (\sigma_{a,i1}/\sigma_{a,11})a_{1t} + \epsilon_{it}$.

To simplify the concept, one can perform orthogonalization on $a_t$ such as the Cholesky decomposition of $\Sigma_a$, namely

$$\Sigma_a = U'U,$$

where $U$ is an upper triangular matrix with positive diagonal elements. Let $\eta_t = (U')^{-1}a_t$. Then, $\text{cov}(\eta_t) = I_k$ and we have

$$z_t = \psi(B)a_t = \psi(B)U'(U')^{-1}a_t = [\psi(B)U']\eta_t.$$

We can then apply the concept of the prior section using modified $\psi$-weight matrices $\psi_j U'$. An weakness of this approach to impulse response functions is that the Cholesky decomposition depends on the ordering of elements of $a_t$. Thus, care must be exercised in multiplier analysis.

11 Forecast error covariance decomposition

12 R demonstration and examples

13 MTS commands used

1. VARorder: order selection (including VARorderI)
2. VAR: estimation
3. refVAR: refined VAR estimation
4. VARpsi: compute $\psi$-weight matrices
5. VARpred: prediction
6. VARirf: impulse response function
7. VARchi: Testing zero parameter constraints
8. BVAR: Bayesian VAR estimation
9. MTSdiag: model checking