Identification and Goodness of Fit Tests for SVAR Models with Application to the Effects of the Quantitative Easing Policy by the Bank of Japan *

BY NAOYA KATAYAMA
Institute of International Business, National Cheng Kung University
and Faculty of Economics, Kansai University,
3-3-35 Yamate-cho, Suita, Osaka 564-8680, Japan
katayama@kansai-u.ac.jp

Abstract. Although there are some methods for constructing structural vector autoregressive (SVAR) models, there are fewer statistical methods for identifying and diagnosing the structural form of the model. This means that there is no way to confirm contemporaneous causal order even if the model seems satisfactorily estimated. We therefore propose some tests for identifying and make diagnoses of the structural form of an SVAR model. The proposed tests use the (inverse) covariance matrix of the error of the reduced form of the VAR models, where the errors are possibly uncorrelated but non-independent. Monte Carlo experiments illustrate the finite sample performance for the proposed test. The proposed tests apply to the effects of the quantitative easing policy by the Bank of Japan for the period January 2009 to December 2014.

Keywords. VAR models; Structural VAR models; AB model; Goodness-of-fit test.

1 Introduction

Suppose that the $K$-dimensional time series $(X_t)$ is generated by a vector autoregressive model, namely the VAR($p$) model

$$X_t = \mu + \sum_{i=1}^{p} \Phi_i X_{t-i} + U_t, \quad t = 0, \pm 1, \pm 2, \ldots,$$

(1)

where $\Phi_i$ is a $K \times K$ matrix whose elements are three-times continuously differentiable functions of an unknown $p_0$ parameter vector $\theta_0 \in \Theta \subset \mathbb{R}^{p_0}$ with $p_0 \leq pK^2$, and $E(U_t) = 0$, $\text{Var}(U_t) = \Sigma_U$ and $\Sigma_U$ are positive definite.

The VAR($p$) model is considered to be a “reduced form” model and is therefore merely a vehicle for summarizing the dynamic properties of data. It is thus a conventional approach to finding a model with instantaneously uncorrelated residuals that models the observable variables directly.

1.1 Why we have to use SVAR?

The following example is from Breitung et. al. (2004) and clarifies difference of VAR and SVAR models.

*Preliminary edition. Please do not circulate.
Example 1 (Keynsian Model) For simplicity, economic models assume that $X_t = U_t$. Let $q_t$, $i_t$, $m_t$ denote output, an interest rate, and real money, respectively. The errors of the corresponding reduced form VAR are denoted by $U_t = (u_t^q, u_t^i, u_t^m)$. The IS-LM model reflecting a traditional Keynesian view is

\[
u^q_t = -a_{12} u^i_t + b_{11} \varepsilon_t^{IS} , \quad (\text{IS curve}),
\]

\[
u^i_t = -a_{21} u^q_t - a_{23} u^m_t + b_{22} \varepsilon_t^{LM} , \quad (\text{inverse LM curve}),
\]

\[
u^m_t = b_{33} \varepsilon_t^m , \quad (\text{money supply curve}).
\]

This corresponds to the AB-model, which can be written

\[
\begin{bmatrix}
  1 & a_{12} & 0 \\
  a_{21} & 1 & a_{23} \\
  0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
  u^q_t \\
  u^i_t \\
  u^m_t
\end{bmatrix}
= \begin{bmatrix}
  b_{11} & 0 & 0 \\
  0 & b_{22} & 0 \\
  0 & 0 & b_{33}
\end{bmatrix}
\begin{bmatrix}
  \varepsilon_t^{IS} \\
  \varepsilon_t^{LM} \\
  \varepsilon_t^m
\end{bmatrix}.
\]

However, the reduced form of a VAR model corresponds to

\[
X_t = A^{-1} B \varepsilon_t,
\]

which makes interpretation of the matrix $A^{-1} B$ difficult.

Henceforth, we consider a model based on $AX_t$. Because realistic representation of the econometric model includes the correlation structures of past data, we now examine the model

\[
AX_t = \mu^* + \sum_{i=1}^p \Phi_i^* X_{t-i} + AU_t , \quad (2)
\]

\[
AU_t = B \varepsilon_t , \quad (3)
\]

where $\mu^* = A \mu$, $\Phi_i^* = A \Phi_i$ and $(\varepsilon_t)$ is a white noise sequence with mean zero and variance $\Sigma_\varepsilon = I_K$. We further assume that both $A$ and $B$ are non-singular. This model is called a structural VAR (SVAR) model. Especially, (3) is called an AB model. SVAR models are widely used by applied statisticians and has been used since the 1980s. From these equations, we have:

\[
X_t = \mu^* + \Phi_0^* X_t + \sum_{i=1}^p \Phi_i^* X_{t-i} + B \varepsilon_t , \quad (4)
\]

where $\Phi_0^* = I - A$ denoting the contemporaneous coefficients and $\Phi_i^*, i = 1, \ldots, p$ denoting lag-$i$ coefficients.

Rewriting the AB model as $U_t = A^{-1} B \varepsilon_t$, we have that the (inverse) covariance of $(U_t)$ denote contemporaneous effects for respective elements. Covariance of the error is usually decomposed into the product of the matrix using Cholesky decomposition. This decomposition yields the contemporaneous effects among the data. However, an important point to note is that the Cholesky decomposition is only one type of identification restriction for representing contemporaneous effects.

1.2 Identification AB models by Cholesky decomposition is difficult

Cholesky decomposition proposes triangular matrices, which implies that the first variable has an instantaneous cause for all other variables. The second variable may have an immediate impact on the last $K - 2$ components of $X_t$, and so on. To establish such an ordering may be quite a difficult exercise in practice. We now present following example.
Example 2 (An Artificial Model) We simulated the 3-dimensional VAR(2) model \( X_t = \begin{bmatrix} 0.6 I_3 X_{t-1} + 0.2 I_3 X_{t-2} + U_t, \end{bmatrix} \) where \( U_t = (u_t^A, u_t^B, u_t^C)' \),

\[
A U_t = B \varepsilon_t \iff \begin{bmatrix} 1 & 0.2 & 0.4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_t^A \\ u_t^B \\ u_t^C \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_t^A \\ \varepsilon_t^B \\ \varepsilon_t^C \end{bmatrix},
\]

and \((\varepsilon_t) \sim NID(0, I_3)\) with sample size 500. Because \( U_t = A^{-1} \varepsilon_t \),

\[
\Sigma_U = E(U_t U_t') = A^{-1} A^{-1}' = \begin{bmatrix} 1.2 & -0.2 & -0.4 \\ -0.2 & 1.0 & 0.0 \\ -0.4 & 0.0 & 1.0 \end{bmatrix}, \quad \Sigma_U^{-1} = A' A = \begin{bmatrix} 1.00 & 0.20 & 0.40 \\ 0.20 & 1.04 & 0.08 \\ 0.40 & 0.08 & 1.16 \end{bmatrix},
\]

so the sample covariance matrix \( \hat{S}_n = n^{-1} \sum \hat{U}_t \hat{U}_t' \) is a consistent estimator of \( \Sigma_U \), and \( \hat{S}_n^{-1} \) is a consistent estimator of \( \Sigma_U^{-1} \).

The Cholesky decomposition of \( \hat{S}_n^{-1} \) is approximately \( A \), which implies: ”\( B \) or \( C \) ⇒ \( A \)”. If we change the ordering according to

\[
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.4 & 0.2 & 1 \end{bmatrix} \begin{bmatrix} u_t^C \\ u_t^B \\ u_t^A \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_t^C \\ \varepsilon_t^B \\ \varepsilon_t^A \end{bmatrix},
\]

then the Cholesky decomposition for this \( A \) is

\[
\begin{bmatrix} 1 & 0.07 & 0.32 \\ 0 & 1 & 0.20 \\ 0 & 0 & 1 \end{bmatrix}
\]

which implies: ”\( A \) ⇒ \( B \) ⇒ \( C \)”. Therefore, incorrect ordering may provide an incorrect conclusion.

1.3 Other identification methods

Some methods other than Cholesky decomposition are mentioned below.

1. Theoretical models approach: Enders (2014, Section 5.11) reviewed several approaches from theoretical models, in particular coefficient restriction of the AB model and variance restriction.

   - Coefficient restriction gives short-term restrictions on the dynamics of the model, for example, restrictions from macro-economic theory or symmetry restrictions from open-economy models.
   - Variance restriction, for example, the diagonal \( B \) is equal to one.
   - Some elements of the long-run impact matrix equal zero, where the long-run impact matrix is given by \((I_K - \Phi_1 - \cdots - \Phi_p)^{-1} A^{-1} B\). Examples in the literature include Blancard and Quah (1989) and Galí (1999).

2. Hypothesis testing approach. Tests based on the (inverse) variance matrix have been investigated less. There are only three existing tests:

   - Parametric test for AB model-based LR test statistics. See Lütkepohl (2007, (9.3.6)): \( n \log \{ \det(\hat{S}_n) \} - \log \{ \det(\hat{A}^{-1} \hat{B} \hat{B}' (\hat{A}^{-1})') \} \).
• Pairwise partial correlation test from Swanson and Granger (1997).

However, these tests make an IID (normality) assumption, and nobody has fully discussed the (partial) covariance based test from the view point of identification and diagnostics of the AB model.

3. Data driven approach:

• Independent component analysis. See Moneta et al. (2013).
• Information criteria, such as AIC (Akaike information criterion) and BIC (Bayesian information criterion). Lütkepohl (2007, chapter 4) discussed various strategies for selecting VAR models with IID errors. Mainassara & Kokonendji (2015) examined the performance of order selection for SVAR models under weak white noise (WWN). Modified information criteria are also presented.

1.4 The aim of the paper

These alternative methods described above do not sufficiently discuss identification and goodness-of-fit tests for the selected models. Therefore, we propose a test for the (partial) covariance matrix based test. The highlights are as follows.

• Our test is applicable to SVAR models with uncorrelated errors but not necessary independent, which is useful in applications to financial data. Hereafter, we call IID noise strong white noise and we call uncorrelated noise WWN.
• The proposed tests are useful for identifying for recursive over-identified AB models.
• The proposed goodness-of-fit tests are based on the Browne (1984)’s covariance matrix tests, which detects incorrect ordering of recursive AB models.
• As an application, we analyse the effects of the quantitative easy policy of the Bank of Japan for the period 2009 to 2014.
• Our methods may be applicable to instantaneous Granger causality tests.
• The approach can be extended to testing for the block triangular A or B for identification.
• The approach can be extended to structural vector error correction models.

Mathematical Lemmas and proofs are given in the appendix.

2 Three examples of identification methods

Let \( U_t = (u_{1t}, \ldots, u_{Kt})' \). To discuss identification, we first parametrize the (inverse) covariance matrices from (3). Rewriting the AB model as \( U_t = A^{-1}B \varepsilon_t \), we have

\[
\Sigma_U = A^{-1}B B' A^{-1}'.
\]

Furthermore, the inverse matrix of \( \Sigma_U \) is

\[
\Sigma_U^{-1} = A'B^{-1'}B^{-1}A.
\]

From these equations, directly identifying the AB model based on the sample (partial) correlation matrix is not an easy task, even if we know some of the elements of \( \Sigma_U \) or \( \Sigma_U^{-1} \) are restricted to zero or one. For example, Cholesky decomposition rewrites a positive-definite symmetric matrix \( \Sigma \) as the product of a lower triangular matrix \( Q \) and its transpose. However,
the zero elements of $\Sigma$ do not always correspond to zero elements of $Q$ other than in the first columns of $\Sigma$ and $Q$. See, for example, Hamilton (1994, Section 4.4). Therefore, we propose another tests for the restriction of the matrices $A$ and $B$ in the following examples. These tests are based on an idea from Lütkepohl (2007, Section 3.6), which discussed testing for Granger’s instantaneous causality.

Natural estimators of $\Sigma_U$ and $\Sigma_U^{-1}$ are given by sample analogues of the mean of the residual process, $(\hat{U}_t \hat{U}_t')$, where $(\hat{U}_t)$ denotes a $K$-dimensional vector of residuals from the reduced form of the VAR($p$) model (1), and $\hat{U}_t = X_t - A_1 X_{t-1} - \cdots - A_p X_{t-p}$. We put

$$S_n = \frac{1}{n} \sum_{t=1}^{n} U_t U_t' \quad \text{and} \quad \hat{S}_n = \frac{1}{n} \sum_{t=1}^{n} \hat{U}_t \hat{U}_t'.$$

**Assumption 1** Denote the vech operator by $v(\cdot)$.

1. $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$.

2. $\sqrt{n} v(S_n - \Sigma_U) \xrightarrow{d} N(0, \Omega)$, where $K^* = K(K+1)/2$ and $\Omega$ is a $K^* \times K^*$ symmetric, non-singular matrix.

Assumption 1 is based on Lütkepohl (2007, Chapter 3) under IID and Boubacar Manassara and Franço (2011) under WWN.

Lemma 4.1 shows that the distribution of $\hat{S}_n$ is asymptotically equivalent to that of $S_n$. Therefore, Assumption 1.2 guarantees asymptotic normality of $\hat{S}_n$. When $(U_t)$ is IID, $v(U_t U_t')$ is also IID and $\Omega$ can be expressed as $2D_K^T(\Sigma_U \otimes \Sigma_U) D_K^{\prime T}$, where $D_K = (D_K' D_K)^{-1} D_K'$ is the Moore–Penrose generalized inverse of the duplication matrix $D_K$. See, for example, Lütkepohl (2007, Proposition 3.4). However, under the WWN assumption, $v(U_t U_t')$ is no longer IID and we need an additional assumption. Davidson (1994, Chapter 24) reviewed several central limit theorems for dependent processes, including martingale difference processes, near epoch dependence functions of strong mixing processes, and stationary ergodic sequences. The asymptotic variance $\Omega$ is described as $2\pi$ times the spectral density matrix at the zero frequency of the vector $v(U_t U_t')$.

From (5) and (6), The matrices $\Sigma_U$ and $\Sigma_U^{-1}$ play major rule to identify $A$ and $B$. For identification, it would be important to examine the correlation structure because each elements of these depend on the magnitude of the variables. Let $\Sigma_U = (\sigma_{ij})$ and $\Sigma_U^{-1} = (\sigma^{ij})$. The correlation between $u_{it}$ and $u_{jt}$ is given by

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}}, \quad R = (\rho_{ij}) = \text{dg}(\Sigma_U)^{-1/2} \Sigma_U \text{dg}(\Sigma_U)^{-1/2}, \quad (7)$$

where $\text{dg}(\Sigma_U)^{-1/2} = \text{dg}(\sigma_{11}^{-1/2}, \ldots, \sigma_{KK}^{-1/2})$. The partial correlation between $u_{it}$ and $u_{jt}$ given the other $K - 2$ elements is given by

$$\rho^{ij} = \frac{-\sigma^{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}}, \quad (8)$$

and the partial correlation matrix $R_{\text{inv}} = (\rho^{ij})$ has 1s along the principal diagonal and non-diagonal $(i, j)$ element is given by $\rho^{ij}$. See, for example, Fujikoshi et al. (2010, Chapter 4). Therefore, $\sigma_{ij} = 0$ indicates the correlation between $u_{it}$ and $u_{jt}$ equal to zero, and $\sigma^{ij} = 0$ indicates the correlation between $u_{it}$ and $u_{jt}$ given the other $K - 2$ elements equal to zero. These results imply that the (inverse) correlation matrices contain information about the structural effect.
where $R$ null hypothesis of (10), we have

Example 4 (AB model) Suppose that $A = I_K$, that $B$ is a lower triangular matrix and that $U_t = B z_t$. It follows that $\Sigma_U = BB'$, which is obtained from a Cholesky decomposition. Therefore, we now consider the test

$$H_0 : R_0 v(B) = 0 \text{ vs } H_A : R_0 v(B) \neq 0,$$

where $R_0$ is a $(q \times K^*)$ matrix choosing the elements from strict lower triangular part of $B$. This matrix is full row rank consists of known zero or one elements.

Let $\hat{B}$ be the lower triangular matrix obtained from a Cholesky decomposition of $\hat{S}_n$ such that $\hat{S}_n = \hat{B}\hat{B}'$. Under Assumption 1 and the null hypothesis in (9), we have

$$Q_n^B = n v(\hat{B})'R_0'\{ R_0H_n\Omega_nH_n' R_0 \}^{-1} R_0 v(\hat{B}) \overset{d}{\rightarrow} \chi^2(q),$$

where $H_n = H_Q(\hat{S}_n)$ and $H_Q(\cdot)$ is defined in Lemma 3.1.

Example 4 (AB model) Suppose that $A$ an upper triangular matrix where has 1s along the principal diagonal and that $B$ is a diagonal matrix. It follows from (6) that $A$ and $B$ in $\Sigma_U^{-1} = A'(BB')^{-1}A$ are obtained from a Cholesky decomposition. See, for example, Hamilton (1994, Section 4.4). We now consider a test for

$$H_0 : R_0 v(A') = 0 \text{ vs } H_A : R_0 v(A') \neq 0,$$

where $R_0$ is defined similarly to Example 3. We obtain the following result, which follows from Lemma 3.1 (b). Let $\hat{A}$ be the upper triangular matrix and $\hat{B}$ the diagonal matrix obtained from a Cholesky decomposition of $\hat{S}_n^{-1}$ such that $\hat{S}_n^{-1} = \hat{A}'\hat{B}^{-1}\hat{B}^{-1}\hat{A}$. Then, under Assumption 1 and the null hypothesis of (10), we have

$$Q_n^A = n v(\hat{A})'R_0'\{ R_0H_n\Omega_nH_n' R_0 \}^{-1} R_0 v(\hat{A}) \overset{d}{\rightarrow} \chi^2(q),$$

where $H_n = L_K(\hat{B} \otimes I_K)L_K' H_Q(\hat{S}_n^{-1})H_{\text{inv}}(\hat{S}_n)$, $H_Q(\cdot)$ and $H_{\text{inv}}(\cdot)$ are defined in Lemmas 3.1 and 3.3.

Example 5 (Partial correlation) Swanson and Granger (1997) discussed testing for pairwise correlation for modelling an acyclic model (or an over-identified A model). We now consider tests for joint test for partial correlation matrix: some of non-diagonal elements of $\Sigma_U^{-1}$ (or equivalently, partial correlation matrix) are zeros:

$$H_0 : R_0 v(\Sigma_U^{-1}) = 0 \text{ vs } H_A : R_0 v(\Sigma_U^{-1}) \neq 0,$$

where $R_0$ is defined similarly to Example 3. We have from Lemma 3.3 that:

$$\sqrt{n} \left\{ v(\hat{S}_n^{-1}) - v(\Sigma_U^{-1}) \right\} \overset{d}{\rightarrow} N(0, H_{\text{inv}}(\Sigma_U)\Omega H_{\text{inv}}(\Sigma_U)),$$

where $H_{\text{inv}}(\cdot)$ is given by Lemma 3.3. Then, under Assumption 1 and the null hypothesis of (11), we have

$$Q_n^{\text{inv}} = n v(\hat{S}_n^{-1})'R_0'\{ R_0H_{\text{inv}}(\hat{S}_n)\Omega_nH_{\text{inv}}(\hat{S}_n)' R_0 \}^{-1} R_0 v(\hat{S}_n^{-1}) \overset{d}{\rightarrow} \chi^2(q).$$

Put $\hat{S}_n^{-1} = (\hat{\rho}^{ij})$. Then the partial correlation matrix $\hat{R}_{\text{inv}} = (\hat{\rho}^{ij})$ is given by:

$$\hat{\rho}^{ij} = \begin{cases} -\frac{\hat{\rho}^{ij}}{\hat{s}^{ii}\hat{s}^{jj}} & i \neq j \\ 1 & i = j \end{cases}$$
We have from (12) and Lemma 3.2 that:

\[ Q_{n}^{\text{inv2}} = n \, \nu(\hat{R}_{1}^{\text{inv}})^{'} R_{0}^{'} (R_{0} H_{n} \Omega_{n} H_{n}^{'} R_{0}^{'} )^{-1} R_{0} \, \nu(\hat{R}_{1}^{\text{inv}}) \xrightarrow{d} \chi^{2}(q), \quad (13) \]

where \( H_{n} = L_{K} H_{P}(\hat{S}_{n}^{-1}) H_{P}^{'} (\hat{S}_{n}^{-1}). \)

**Remark 1** The identification examples discuss the restrictions of \( A, B \) or these products directly. The alternate way to examine the restrictions of the impulse responses to have values that are motivated by the economic theory. The long-run restriction of the impulse response is first investigated by Blanchard and Quah (1989). The sign restrictions of the impulse response are examined by several papers. See, e.g., Baumeister and Hamilton (2015), Martin et al (2013, Section 14.2.4) and references therein. The asymptotic distribution of the (accumulated or total accumulated) impulse responses is given by Lütkepohl (1990, Proposition 1).

### 3 Goodness-of-fit test for the (inverse) covariance matrix

The proposed tests described in the previous subsection are useful for identifying the structure of the AB-model. This subsection discusses goodness-of-fit tests for the AB-model.

We consider the following restrictions on \( A \) and \( B \):

\[
\text{vec}(A) = R_{A} \gamma_{A} + r_{A} \quad \text{and} \quad \text{vec}(B) = R_{B} \gamma_{B} + r_{B},
\]

where \( R_{A} \) and \( R_{B} \) are suitable fixed matrices of zeros and ones, \( \gamma_{A} \) and \( \gamma_{B} \) are vectors of free parameters and \( r_{A} \) and \( r_{B} \) are vectors of fixed parameters. We also define \( \gamma_{0} = (\gamma_{A}^{'} \gamma_{B}^{'} )^{'} \) to be a \( l \)-dimensional parameter vector and \( \hat{\gamma}_{n} \) to be the estimator of \( \gamma_{0} \). Putting \( \Sigma(\gamma_{0}) = A^{-1} B B^{'} A^{-1} \) and \( \Sigma(\gamma_{0})^{-1} = A^{'} B^{-1} B^{-1} A \), we consider the following tests:

\[
H_{0} : \Sigma_{U} = \Sigma(\gamma_{0}) \quad \text{vs} \quad H_{A} : \Sigma_{U} \neq \Sigma(\gamma_{0}),
\]

\[
H_{0} : \Sigma_{U}^{-1} = \Sigma(\gamma_{0})^{-1} \quad \text{vs} \quad H_{A} : \Sigma_{U}^{-1} \neq \Sigma(\gamma_{0})^{-1}.
\]

The estimator of \( \gamma_{0} \) has been discussed in Amisano & Giannini (1997), Breitung et al. (2004), Lütkepohl (2007) and Boubacar Maïnassara (2011). We assume that:

**Assumption 2**

1. \( \sqrt{n}(\hat{\gamma}_{n} - \gamma_{0}) = O_{P}(1). \)

2. \( \text{rank}(\Delta) = l, \) where \( \Delta = \Delta(\gamma_{0}) \) and

\[
\Delta \equiv \frac{\partial \nu(\Sigma(\gamma_{0}))}{\partial \gamma_{0}^{'}} = [-2D_{K}^{\Delta}(\Sigma_{U} \otimes A^{-1}) R_{A}^{'} : 2D_{K}^{\Delta}(A^{-1} B \otimes A^{-1}) R_{B}].
\]

The matrix \( \Delta \) is used for local identification of the AB model. The assumption above guarantees that the Fisher information matrix is non-singular. See, for example, Lütkepohl (2007, Chapter 9). It also guarantees that both the partial derivatives of both \( \nu(\Sigma(\gamma_{0})) \) and \( \nu(\Sigma(\gamma_{0})^{-1}) \) with respect to \( \gamma_{0} \) are full column rank. See Lemma 5.1 below.

Following the tests from Browne (1984, Proposition 4), we first construct the test statistic for (15). The test statistics \( e_{n} = e_{n}(\hat{\gamma}_{n}) = \nu(\hat{S}_{n} - \Sigma(\hat{\gamma}_{n})) \) may be written as

\[
e_{n} = \nu(\hat{S}_{n} - \Sigma_{U}) + \nu \{ \Sigma_{U} - \Sigma(\gamma_{0}) \} - \nu \{ \Sigma(\hat{\gamma}_{n}) - \Sigma(\gamma_{0}) \}
\]

(17)

The second term on the right-hand side is zero if the null hypothesis of (15) is true. Using the mean value theorem, the third term on the right-hand side is approximately \(-\Delta(\hat{\gamma}_{n} - \gamma_{0}). \) Because \( \Delta \) is full column rank, there exists an orthogonal complement matrix for \( \Delta, \) denoted
\( \Delta_\perp \), which is a \( K^* \times (K^* - l) \) matrix with full column rank such that \( \Delta_\perp' \Delta = 0 \). Therefore, multiplying on the left by \( \Delta_\perp' \) under the null hypothesis, we have

\[
\Delta_\perp' \left\{ \hat{S}_n - \Sigma(\hat{\gamma}_n) \right\} = \Delta_\perp' v \left( \hat{S}_n - \Sigma_U \right) + o_p(n^{-1/2}). \tag{18}
\]

From this analysis, we consider the following test statistics:

\[
Q_n = n e_n' \Delta_\perp(\Delta_\perp'^{-1}\Omega_n \Delta_\perp) - 1 \Delta_\perp' e_n = n e_n' \Omega_n^{-1} e_n - n e_n' \Omega_n^{-1} \Delta_n(\Delta_n'^{-1} \Omega_n^{-1} \Delta_n)' - 1 \Delta_n'^{-1} \Omega_n^{-1} e_n,
\]

where \( \Delta_n = \Delta(\hat{\gamma}_n) \) and \( \Delta_\perp \) is its orthogonal complement. The second term on the right-hand side may be regarded as a correction term for a non-efficient estimator. If we use an estimator satisfying \( \Delta(\hat{\gamma}_n)' \Omega_n^{-1} e_n(\hat{\gamma}_n) = 0 \), for example, \( \hat{\gamma}_n = \arg min, e_n(\gamma)' \Omega_n^{-1} e_n(\gamma) \), then the second term is equal to zero. One of way computing \( \Delta_\perp \) from \( \Delta_n \) is by QR decomposition, which is built into standard statistical software such as R and MATLAB.

We similarly have the test statistic for (16). From Lemma 5.2, we define \( \tilde{\Delta} = \Delta(\gamma_0) \) as \( \tilde{\Delta} = \partial \bar{v}(\Sigma(\gamma_0)^{-1})/\partial \gamma_0' = H_\gamma \Delta \), where \( H_\gamma = D_K' \{ \Sigma(\gamma_0)^{-1} \otimes \Sigma(\gamma_0)^{-1} \} D_K \), and define \( \Delta_n = \tilde{\Delta}(\hat{\gamma}_n) \) and \( \Delta_\perp \) as its orthogonal complement. We now consider the following test statistic:

\[
\tilde{Q}_n = n \tilde{e}_n' \tilde{\Delta}_\perp(\tilde{\Delta}_\perp'^{-1} \tilde{\Omega}_n \tilde{\Delta}_\perp) - 1 \tilde{\Delta}_\perp' \tilde{e}_n = n \tilde{e}_n' \tilde{\Omega}_n^{-1} \tilde{e}_n - n \tilde{e}_n' \tilde{\Omega}_n \tilde{\Delta}_n(\tilde{\Delta}_n'^{-1} \tilde{\Omega}_n^{-1} \tilde{\Delta}_n)' - 1 \tilde{\Delta}_n'^{-1} \tilde{\Omega}_n^{-1} \tilde{e}_n,
\]

where \( \tilde{e}_n = v \{ \hat{S}_n^{-1} - \Sigma(\hat{\gamma}_n)^{-1} \} \).

The asymptotic distributions of \( Q_n \) and \( \tilde{Q}_n \) are given by the following theorem.

**Theorem 1** Under Assumptions 1 and 2, we have the following as \( n \to \infty \).

1. **Testing for (15) under the null hypothesis**, \( Q_n \xrightarrow{d} \chi^2(K^* - l) \).
2. **Testing for (16) under the null hypothesis**, \( \tilde{Q}_n \xrightarrow{d} \chi^2(K^* - l) \).

**Example 6 (An Artificial Model (cont. from Example 2))** We compute the test statistics \( Q_n \) and \( \tilde{Q}_n \) in the correctly specified case, \( U_t = (u_t^A, u_t^B, u_t^C)' \), and in the incorrectly specified case, \( U_t^{Rev} = (u_t^C, u_t^B, u_t^A)' \).

<table>
<thead>
<tr>
<th>Ordering</th>
<th>DF</th>
<th>( Q_n ) value</th>
<th>( Q_n ) p-value</th>
<th>( \tilde{Q}_n ) value</th>
<th>( \tilde{Q}_n ) p-value</th>
<th>log likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>( U_t )</td>
<td>1</td>
<td>0.02</td>
<td>0.88</td>
<td>0.02</td>
<td>0.88</td>
<td>-737.0</td>
</tr>
<tr>
<td>( U_t^{Rev} )</td>
<td>1</td>
<td>2.88</td>
<td>0.09</td>
<td>2.90</td>
<td>0.09</td>
<td>-738.6</td>
</tr>
</tbody>
</table>

The log likelihood is computed as in Lütkepohl (2007, Section 9.3).

3.1 **Empirical size and power for (in)correct ordering**

Using models in Example 2, we conducted simulation studies where we computed \( Q_n \) and \( \tilde{Q}_n \). We supposed that the \( p = 2 \) is known and LRV estimator is Andrew’s estimator, and \( B \) is also supposed to be known. Both of the tests are 3 DF. We consider sample size \( n = 200, 400, 600, 800 \) and use 10,000 replications of each experiment. Table 1 represent the relative rejection frequencies (in %). The rows of \( U_t \) correspond to the correctly specified model. The rows of \( U_t^{Rev} \) correspond to the incorrectly specified model.

The tests of \( Q_n \) have noticeable over-rejection probabilities for \( n = 200 \). The means and variances are larger than those for \( \chi_3^2 \). This behaviour diminishes as sample size increases. \( \tilde{Q}_n \) is comparatively close to nominal significance level. When the causal ordering is incorrect, rows of \( U_t^{Rev} \), the both test statistics seems reject the models satisfactorily.
mean 4.74 3.64 3.42 3.17 3.35 3.15
variance 21.55 8.31 14.60 8.39 9.87 6.92

\( Q_n \) | \( \hat{Q}_n \) | \( Q_n \) | \( \hat{Q}_n \) | \( Q_n \) | \( \hat{Q}_n \) | \( Q_n \) | \( \hat{Q}_n \)
---|---|---|---|---|---|---|---
\( n = 200 \) | \( n = 400 \) | \( n = 600 \) | \( n = 800 \)

\( U_t \)
1% | 8.50 | 2.10 | 4.50 | 2.20 | 3.10 | 1.70 | 1.50 | 0.90
5% | 18.60 | 8.60 | 11.80 | 7.20 | 8.60 | 5.80 | 6.70 | 4.60
10% | 25.00 | 15.40 | 18.00 | 14.20 | 14.80 | 11.00 | 13.60 | 10.30

\( U_t^{Rev} \)
1% | 42.90 | 33.20 | 65.70 | 63.30 | 84.50 | 84.10 | 94.60 | 94.90
5% | 64.70 | 57.10 | 83.80 | 82.30 | 94.30 | 94.60 | 99.60 | 99.30
10% | 71.90 | 68.40 | 90.50 | 89.60 | 97.50 | 97.90 | 99.70 | 99.70

Table 1: Empirical size and power of Example 2 (DF= 3)

4 Real data examples

4.1 Japanese macroeconomic data

We now analyse the effects of the quantitative easing policy by the Bank of Japan for the period January 2009 to December 2014 (monthly data, \( n = 72 \)) using the following standard VAR model:

\[
X_t = \begin{bmatrix} CPI \\ CI \\ RAJ \\ ARINF \\ MB \end{bmatrix}_t = \mu + \Phi_1 \begin{bmatrix} CPI \\ CI \\ RAJ \\ ARINF \\ MB \end{bmatrix}_t + \Phi_2 \begin{bmatrix} CPI \\ CI \\ RAJ \\ ARINF \\ MB \end{bmatrix}_{t-1} + U_t,
\]

where \( CPI \) is the consumer price index, \( CI \) is the index of total domestic expenditure, \( RAJ \) is the rate of job availability, \( ARINF \) is the expected rate of inflation, and \( MB \) is the monetary base. This model follows from Maekawa et al. (2015). They discussed the causal relation of the model from reduced form of a VAR(2) model, where \( p = 2 \) is selected by Akaike’s information criterion. They infer the following causal order:

\[
MB \rightarrow ARINF \rightarrow RAJ \rightarrow CI \rightarrow CPI.
\]

This section provides contemporaneous and lagged coefficients of (4).

\[
X_t = \mu^* + \Phi_0 X_t + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + B \varepsilon_t,
\]

where \( \Phi_0 = I - A \) is a strict upper triangular matrix. We tentatively assume that a recursive AB model in Example 4: \( B \) is a diagonal matrix and \( A \) is an upper triangular matrix with ones as the diagonal elements.

We first estimate reduced form of the VAR(2) coefficients imposing zero restriction of \( \mu, \Phi_1 \) and \( \Phi_2 \) by R function restrict in vars package. Estimated parameters are all significant. The goodness-of-fit test (test for lag-1,2,3 correlated errors) for the Breusch-Godfrey LM statistic is not rejected with \( p \)-value 0.43. Therefore we next identify the AB model using the residual sequences.
4.2 Identification of the AB model

The sample correlation matrix is:

<table>
<thead>
<tr>
<th></th>
<th>CPI</th>
<th>CI</th>
<th>RAJ</th>
<th>ARINF</th>
<th>MB</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI</td>
<td>1.00</td>
<td>-0.62</td>
<td>-0.17</td>
<td>-0.19</td>
<td>0.02</td>
</tr>
<tr>
<td>CI</td>
<td>-0.62</td>
<td>1.00</td>
<td>0.30</td>
<td>-0.11</td>
<td>-0.24</td>
</tr>
<tr>
<td>RAJ</td>
<td>-0.17</td>
<td>0.30</td>
<td>1.00</td>
<td>0.09</td>
<td>-0.10</td>
</tr>
<tr>
<td>ARINF</td>
<td>0.19</td>
<td>-0.11</td>
<td>0.09</td>
<td>1.00</td>
<td>0.30</td>
</tr>
<tr>
<td>MB</td>
<td>0.02</td>
<td>-0.24</td>
<td>-0.10</td>
<td>0.30</td>
<td>1.00</td>
</tr>
</tbody>
</table>

and the partial correlation matrix is:

<table>
<thead>
<tr>
<th></th>
<th>CPI</th>
<th>CI</th>
<th>RAJ</th>
<th>ARINF</th>
<th>MB</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI</td>
<td>1.00</td>
<td>-0.65</td>
<td>0.06</td>
<td>-0.30</td>
<td>-0.08</td>
</tr>
<tr>
<td>CI</td>
<td>-0.65</td>
<td>1.00</td>
<td>0.27</td>
<td>-0.25</td>
<td>-0.20</td>
</tr>
<tr>
<td>RAJ</td>
<td>0.06</td>
<td>0.27</td>
<td>1.00</td>
<td>0.16</td>
<td>-0.06</td>
</tr>
<tr>
<td>ARINF</td>
<td>-0.30</td>
<td>-0.25</td>
<td>0.16</td>
<td>1.00</td>
<td>0.25</td>
</tr>
<tr>
<td>MB</td>
<td>-0.08</td>
<td>-0.20</td>
<td>-0.06</td>
<td>0.25</td>
<td>1.00</td>
</tr>
</tbody>
</table>

We conduct joint tests using Lemma 3 for zero-restriction of the underlined elements in the matrices above and concluded that these elements are not significant. However, the zero restriction of the (partial) correlation matrices cannot identify $A$ as discussed in Section 2. Therefore, following from Example 4, we estimate $A$ by Cholesky decomposition of $S^{-1}$:

<table>
<thead>
<tr>
<th></th>
<th>CPI</th>
<th>CI</th>
<th>RAJ</th>
<th>ARINF</th>
<th>MB</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI</td>
<td>1.00</td>
<td>0.09**</td>
<td>-0.01</td>
<td>0.39**</td>
<td>0.93</td>
</tr>
<tr>
<td>CI</td>
<td>0</td>
<td>1.00</td>
<td>-0.58*</td>
<td>0.76</td>
<td>20.83</td>
</tr>
<tr>
<td>RAJ</td>
<td>:</td>
<td>:.</td>
<td>1.00</td>
<td>-0.84</td>
<td>11.6</td>
</tr>
<tr>
<td>ARINF</td>
<td>:</td>
<td>:.</td>
<td>1.00</td>
<td>-4.18*</td>
<td></td>
</tr>
<tr>
<td>MB</td>
<td>0</td>
<td>...</td>
<td>...</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

In this matrix, * and ** denotes the results of marginal significance test discussed in Example 4, where * denotes significant 5% and ** denotes significant 1%. We next conduct joint test of other strictly upper triangular elements with zero-restriction by the chi-squared test in Example 4. For the elements with asterisk, the null hypothesis of joint zero-restrictions is rejected with p-value 0.00. Reversely, for the elements without asterisk, the null hypothesis of joint zero-restrictions is not rejected with p-value 0.70. Therefore, we tentatively identified $A$ by zero restrictions toward all elements without asterisk.

4.3 Estimation and goodness-of-fit test of the AB model

The estimation of the AB model is computed from $\text{SVAR}$ function in $\text{vars}$ package. The estimated matrix $A$ is:

$$
\hat{A} = 
\begin{array}{ccccc}
\text{CPI} & \text{CI} & \text{RAJ} & \text{ARINF} & \text{MB} \\
1 & 0.09(0.01) & 0.00 & 0.40(0.13) & 0.00 \\
0 & 1 & -0.59(0.21) & 0.00 & 0.00 \\
\vdots & \vdots & 1 & 0.00 & 0.00 \\
\vdots & \vdots & \vdots & 1 & -4.18(1.78) \\
0 & \cdots & \cdots & 0 & 1 \\
\end{array}
$$
and the estimated matrix $B$ is:

$$\hat{B} = dg [0.16(0.01) \quad 1.31(0.11) \quad 0.74(0.06) \quad 0.15(0.01) \quad 0.01(0.00)]' .$$

where $(\cdot)$ denotes standard error. We have confirmed that all parameters are significant. Finally, from Theorem 1, we conduct goodness-of-fit tests for these models. The results are shown in the table below.

<table>
<thead>
<tr>
<th>statistic</th>
<th>DF</th>
<th>value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_n$</td>
<td>6</td>
<td>3.09</td>
<td>0.80</td>
</tr>
<tr>
<td>$\hat{Q}_n$</td>
<td>6</td>
<td>4.87</td>
<td>0.56</td>
</tr>
<tr>
<td>LR</td>
<td>6</td>
<td>5.80</td>
<td>0.45</td>
</tr>
</tbody>
</table>

where the LR test statistic is computed according to Lütkepohl (2007, Section 9.3). All tests accept the over-identified model. We also note that the tentative causal order (19) is confirmed by the goodness-of-fit tests.

The estimation results of (20) are as follows:

$$\hat{\Phi}_0^* =
\begin{bmatrix}
\text{CPI} & \text{CI} & \text{RAJ} & \text{ARINF} & \text{MB} \\
0 & -0.09 & 0.00 & -0.40 & 0.00 \\
0 & 0 & 0.59 & 0.00 & 0.00 \\
\vdots & \vdots & 0 & 0.00 & 0.00 \\
\vdots & \vdots & \vdots & 0 & 4.18 \\
0 & \cdots & \cdots & 0 & 0 \\
\end{bmatrix}$$

$$\hat{\Phi}_1^* =
\begin{bmatrix}
\text{CPI} & \text{CI} & \text{RAJ} & \text{ARINF} & \text{MB} \\
1.20 & 0.05 & -0.07 & -0.52 & -3.11 \\
-1.58 & 0.00 & 0.80 & 0.00 & 16.05 \\
-0.50 & 0.00 & 1.36 & 0.00 & 0.00 \\
-0.08 & 0.00 & 0.00 & 1.30 & 4.17 \\
0.00 & 0.00 & 0.00 & 0.02 & 1.00 \\
\end{bmatrix}$$

$$\hat{\Phi}_2^* =
\begin{bmatrix}
\text{CPI} & \text{CI} & \text{RAJ} & \text{ARINF} & \text{MB} \\
0.00 & -0.03 & 0.02 & 0.19 & -1.26 \\
0.00 & 0.36 & -0.26 & 0.00 & 8.05 \\
0.00 & 0.15 & -0.44 & 0.00 & 13.67 \\
0.00 & 0.00 & 0.00 & -0.48 & 1.34 \\
0.00 & 0.00 & 0.00 & -0.02 & 0.00 \\
\end{bmatrix}$$

Figures 1 draws the contemporaneous causal effects from $\hat{\Phi}_0^*$, where the solid arrows correspond to positive effects and dashed arrows correspond to negative effects. Similarly, Figure 2 draws the contemporaneous and lagged effects from MB based on the results of $\Phi_i^*$, $i = 0, 1, 2$. Similarly, Figures 3–6 draw the effects from ARINF, RAJ, CI, CPI, respectively. We find as follows:

- Some negative arrows are shown toward CPI, which contradict to the economic theory.
- MB (monetary base) show significant positive effect to other variables except for CPI.
- ARINF (expected inflation rate) does not show significant positive effect to RAJ, CI, CPI, which is consistent with impulse response analysis of Maekawa et al. (2015).
Figure 1: Plot of result from $\hat{\Phi}_0$. The solid arrows indicate positive effects, dashed arrows negative ones.

Figure 2: Plot of result from $\hat{\Phi}_i$, $i = 0, 1, 2$ for MB. The solid arrows indicate positive effects, dashed arrows negative ones.

Figure 3: Plot of result from $\hat{\Phi}_i$, $i = 0, 1, 2$ for RAJ. The solid arrows indicate positive effects, dashed arrows negative ones.
Figure 4: Plot of result from $\hat{\Phi}_1^i$, $i = 0, 1, 2$ for RAJ. The solid arrows indicate positive effects, dashed arrows negative ones.

Figure 5: Plot of result from $\hat{\Phi}_1^i$, $i = 0, 1, 2$ for CI. The solid arrows indicate positive effects, dashed arrows negative ones.

Figure 6: Plot of result from $\hat{\Phi}_1^i$, $i = 0, 1, 2$ for CPI. The solid arrows indicate positive effects, dashed arrows negative ones.
Acknowledgments

We would like to thank the editors, and the referees for their valuable comments and suggestions. We would also thank Dr. Kouichi Maekawa, Dr. Shuichi Komura, Dr. Shuichi Nagata for their allowance use of the data in Section 4, valuable comments and suggestions.

This work is supported by JSPS Kakenhi (Grant Numbers 26380278, 26380279), Kansai University’s overseas research program for the academic year of 2015, and Ministry of Science and Technology, TAIWAN for the project from MOST 104-2811-H-006-005.

APPENDIX

Lemma 1 Suppose \( \hat{s}_n = v(\hat{S}_n) \), \( \sigma_U = v(\Sigma_U) \) and \( f : \mathbb{R}^{K^*} \rightarrow \mathbb{R}^M \) is a continuously differentiable function which maps \( \sigma_U \) on an \( M \)-dimensional vector with \( \partial f(\sigma_U)/\partial \sigma_U' \neq 0 \). Then using delta method (see e.g., Serfling (1980, pp.122-124)) and Lemma 4.1, we have:

\[
\sqrt{n}(\hat{S}_n - \sigma_U) \xrightarrow{d} N(0, \Omega) \tag{A.1}
\]

\[
\sqrt{n}\{f(\hat{s}_n) - f(\sigma_U)\} \xrightarrow{d} N(0, F\Omega F') \tag{A.2}
\]

where \( F = F(\sigma_U) = \partial F(\sigma_U)/\partial \sigma_U' \). In addition, if \( \text{rank}(F) = M \),

\[
T_n = n\{f(\hat{s}_n) - f(\sigma_U)\}'(F_n\Omega_nF_n')^{-1}\{f(\hat{s}_n) - f(\sigma_U)\} \xrightarrow{d} \chi^2(M), \tag{A.3}
\]

where \( F_n \) and \( \Omega_n \) are consistent estimator of \( F \) and \( \Omega \), respectively.

Lemma 2 Let \( A, B, C, D \) be matrices with appropriate dimensions.

1. \( (A \otimes B)' = (A' \otimes B') \).
2. \( (A \otimes B)(C \otimes D) = (AC \otimes BD) \).
3. \( \text{vec}(ABC) = (C' \otimes A)\text{vec}(B) \).
4. Suppose that \( \sqrt{n}\text{vec}[^{\hat{A}}-A:]^{\hat{B}}-B:]^{\hat{C}}-C] \xrightarrow{d} N(0, \Sigma) \). Then \( \sqrt{n}\text{vec}(^{\hat{A}}B\hat{C}-ABC) \xrightarrow{d} N(0, \Sigma_{ABC}) \), where \( \Sigma_{ABC} = H\Sigma H' \), \( H = [C' B' \otimes I_A : C' \otimes A : I_C \otimes AB], I_A \) and \( I_B \) are identity matrices with \( I_A A = A \) and \( C \otimes C = C \). Especially, if \( \Sigma \) is block diagonal with diagonal elements matrices are \( \Sigma_A, \Sigma_B, \Sigma_C \) respectively, \( \Sigma_{ABC} = (C' B' \otimes I_A)\Sigma_A (BC \otimes I) + (C' \otimes A)\Sigma_B (C \otimes A') + (I_C \otimes AB)\Sigma_C (I \otimes B'A') \).
5. Suppose that, for square \( (K \times K) \) matrices \( ^{\hat{A}}A \) and \( A \), \( \sqrt{n}\text{vec}(^{\hat{A}}-A) \xrightarrow{d} N(0, \Sigma_A) \). Then

\[
\sqrt{n}\text{vec}(^{\hat{A}}-A) \xrightarrow{d} \begin{cases} N(0, D_K^+\Sigma_A D_K^+) & \text{if } ^{\hat{A}}A \text{ and } A \text{ are both symmetric,} \\
N(0, L_K \Sigma_A L_K') & \text{otherwise} \end{cases}
\]

In the proof the rules for matrix manipulations stated in Schott (2005), Magnus and Neudecker (1988) and Magnus (1988) will be used. The three references will be abbreviated by S, MN and M, respectively.
Proof of Lemma 2. We have 1–3 from (M, Section 1.10). From 3,
\[
\sqrt{n} \text{vec}(\hat{A}BC - ABC) = \sqrt{n} \text{vec} \left\{ \mathbb{I}_A (\hat{A} - A) \hat{B}C + A(\hat{B} - B)C - AB(\hat{C} - C) \mathbb{I}_C \right\} \\
= [\hat{C}'\hat{B}' \otimes I_A ; \hat{C}' \otimes A ; \mathbb{I}_C \otimes AB] \text{vec} [\hat{A} - A; \hat{B} - B; \hat{C} - C].
\]
This proves 4. Finally, 5 follows from (M, Theorems 4.1(iv), 5.3(i), Theorem 6.7(i)). □

Define $K_{mn}$ is the commutation matrix defined so that $\text{vec}(G) = K_{mn} \text{vec}(G')$ for any $(m \times n)$ matrix $G$, and $L_K$ is the $(K^* \times K^2)$ elimination matrix defined so that $\text{vec}(F) = L_K \text{vec}(F)$ for any $K \times K$ matrix $F$.

The following results are easily obtained from Lemma 1 of Lütkepohl (1981), (M, Theorems 10.8 and 10.10), and (S, Corollary 9.4.1):

**Lemma 3** Suppose that $\Xi$ is symmetric, positive definite $(K \times K)$ matrix and $\sqrt{n} v(\Xi_n - \Xi) \overset{d}{\to} N(0, \Sigma_{\Xi})$ with positive definite symmetric matrix $\Sigma_{\Xi}$. Then it holds that:

1. **[Convergence of factorized matrices of $\Xi_n$]** Suppose that $Q$ is $(K \times K)$ non singular lower triangular matrix such that $\Xi = QQ'$, and $Q_n$ is $(K \times K)$ non singular lower triangular matrix such that $\Xi_n = Q_nQ_n'$.

   (a) Then $v(Q)$ is a totally differentiable function of $v(\Xi)$ and
   \[
   \sqrt{n} v(Q_n - Q) \overset{d}{\to} N(0, \Sigma_Q),
   \]
   where $\Sigma_Q = H_{Q(\Xi)} = H_Q(\Xi)' \Sigma_{\Xi} H_Q(\Xi)$, $H_Q(\Xi) = [L_K(1_{K^2} + K_{KK})(Q \otimes I_K)D_K']^{-1}$. Especially, if $R_0 v(\Sigma_Q) = 0$, $\sqrt{n} R_0 v(Q_n) \overset{d}{\to} N(0, R_0 \Sigma_Q R_0')$.

   (b) In addition, suppose that $S$ and $T$ are $(K \times K)$ matrices such that $Q = ST$, where $S$ is a lower triangular with $\text{dg}(S) = (1, \ldots, 1)'$ and $T = \text{dg}(Q)$. Similarly, corresponding to $S$ and $T$, suppose that $S_n$ and $T_n$ are $(K \times K)$ matrices such that $Q_n = S_nT_n$.

   Then
   \[
   \sqrt{n} R_0 v(S_n - S) \overset{d}{\to} N(0, \Sigma_S),
   \]
   where $\Sigma_S = L_K C_S H_S(\Xi) \Sigma_{\Xi} H_S(\Xi)' C_S' L_K$, $C_S = I_K - (I_K \otimes S) \Psi_K \Psi_K$, $H_S(\Xi) = \{\text{dg}(Q)^{-1} \otimes I_K\} L_K H_Q(\Xi)$, Especially, if $R_0 v(S) = 0$,
   \[
   \sqrt{n} R_0 v(S_n) \overset{d}{\to} N(0, R_0 L_K H_S(\Xi) \Sigma_{\Xi} H_S(\Xi)' L_K R_0')
   \]
   and the asymptotic variance is non singular.

2. **[Convergence of normalized matrices of $\Xi_n$]** Put $P = P(\Xi) = \frac{\text{dg}(\Xi)^{-1/2} \Xi \text{dg}(\Xi)^{-1/2}}{\text{dg}(\Xi)}$ and $P_n = P(\Xi_n)$. Then
   \[
   \sqrt{n} v(P_n - P) \overset{d}{\to} N(0, \Sigma_P),
   \]
   where $\Sigma_P = L_K C_P H_P(\Xi) \Sigma_{\Xi} H_P(\Xi)' C_P' L_K$, $C_P = I - N_K (1 \otimes P) N_K \Psi_K \Psi_K$, $H_P(\Xi) = \{\text{dg}(\Xi)^{-1/2} \otimes \text{dg}(\Xi)^{-1/2}\} D_K$. Especially, if $R_0 v(P) = 0$,
   \[
   \sqrt{n} R_0 v(P_n) \overset{d}{\to} N(0, R_0 L_K H_P(\Xi) \Sigma_{\Xi} H_P(\Xi)' L_K R_0')
   \]
   and the asymptotic variance is non singular.

3. **[Convergence of inverse matrix of $\Xi_n$]** $\sqrt{n} v(\Xi_n^{-1} - \Xi^{-1}) \overset{d}{\to} N(0, H_{\text{Inv}}(\Xi) \Sigma_{\Xi} H_{\text{Inv}}(\Xi)'$, where $H_{\text{Inv}}(\Xi)$ is a $(K \times K)$ non singular matrix defined by $H_{\text{Inv}}(\Xi) = D_K(\Xi^{-1} \otimes \Xi^{-1}) D_K$.}
Proof of Lemma 3. The proof of 1 (a) is obtained from Lemma 1 of Lütkepohl (1989) and 3 is obtained from (M, Theorems 10.8 and 4.8(iii)) and (S, Corollary 9.4.1). The proof of 2 is similar manner of the proof of (M, Theorem 10.10). We present outlines only. Putting:

\[ W_n = \text{dg}(\Xi)^{-1/2} (\Xi - \Xi) \text{dg}(\Xi)^{-1/2}, \]

we have: \( \sqrt{n} \text{vec}(W_n) \xrightarrow{d} N(0, \mathbf{H}_P(\Xi) \Sigma_\Xi \mathbf{H}_P(\Xi)'), \)

which yields (A.5). For the case of \( R_0 v(P) = 0, \) it is enough to show that the second term of right hand side of (A.6) is negligible. It is follows that \( R_0 v(P) = 0 \) is equivalent to \( R_0 v(P \text{dg}(A)) = 0 \) and \( R_0 v(\text{dg}(A)) = 0, \) where all elements of \( \text{dg}(A) \) are non-zero. Since \( L_K \mathbf{H}_P(\Xi) = L_K \{ \text{dg}(\Xi)^{-1/2} \otimes \text{dg}(\Xi)^{-1/2} \} L_K' \) using the proof of (M, Theorem 5.12), \( L_K \mathbf{H}_P(\Xi) \) is non singular using (M, Theorem 5.7 (ii)). Therefore the asymptotic variance is non singular as \( R_0 \) is full raw rank.

Therefore, left of the proof is 1 (b). This is obtained from the similar manner of the proof of 2. It follows from Hamilton (1994, Section 4.4) that \( (T, T_n) \) is non singular using (M, Theorem 5.7 (ii)). Therefore the asymptotic variance is non singular.

Under Assumptions 1, the following hold as \( n \to \infty:\)

1. \( \sqrt{n} \{ v(S_n) - v(S_n) \} \xrightarrow{p} 0, \)
2. \( \sqrt{n} \{ v(S_n) - v(S_n) \} \xrightarrow{d} N(0, \Omega), \)
3. \( \sqrt{n} \{ v(S_n^{-1}) - v(S_n^{-1}) \} \xrightarrow{d} N(0, \mathbf{H}_{\text{inv}}(S_n) \Omega \mathbf{H}_{\text{inv}}(S_n)'), \)

Lemma 4 Under Assumptions 1, the following hold as \( n \to \infty:\)

1. \( \sqrt{n} \{ v(S_n) - v(S_n) \} \xrightarrow{p} 0, \)
2. \( \sqrt{n} \{ v(S_n) - v(S_n) \} \xrightarrow{d} N(0, \Omega), \)
3. \( \sqrt{n} \{ v(S_n^{-1}) - v(S_n^{-1}) \} \xrightarrow{d} N(0, \mathbf{H}_{\text{inv}}(S_n) \Omega \mathbf{H}_{\text{inv}}(S_n)'), \)

16
Proof. For 1, it is enough to show that
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \text{vec}(\hat{U}_t \hat{U}_t' - U_t U_t') \xrightarrow{p} 0. \tag{A.12}
\]

Using a Taylor series expansion around \( \hat{b}_n = \theta_0 \), we have:
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \text{vec}(\hat{U}_t \hat{U}_t' - U_t U_t') = \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{\partial U_t}{\partial \theta_0} \otimes U_t + U_t \otimes \frac{\partial U_t}{\partial \theta_0} \right\} \sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1),
\]
which we use the facts that \( \text{vec}(yx') = x \otimes y \) and \( d(x \otimes y) = (dx) \otimes y + x \otimes (dy) \). See, for example, Schott (2005, Chap. 9) Because \( \sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1) \), (A.12) follows from the law of large numbers of martingale sequences:
\[
\frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{\partial U_t}{\partial \theta_0} \otimes U_t + U_t \otimes \frac{\partial U_t}{\partial \theta_0} \right\} \xrightarrow{p} 0.
\]

The proof of 2 is readily obtained from 1 and Assumption 1.2. For 3, a Taylor series expansion around \( v(\hat{b}_S) = v(U_t) \) gives
\[
v(\hat{S}_n^{-1}) = v(\Sigma_U^{-1}) - \hat{H} \left\{ v(\hat{S}_n) - v(\Sigma_U) \right\} + o_p(n^{-1/2}),
\]
where we use a matrix derivative formula from Schott (2005, Corollary 9.4.1). The rest of the proof of 3 is obvious from 2. \( \square \)

Lemma 5
1. \( \Delta \equiv \frac{\partial v(\Sigma(\theta_0))}{\partial \gamma_0'} = [-2D_K^+(\Sigma_U \otimes A^{-1}) R_A : 2D_K^+(A^{-1} B \otimes A^{-1}) R_B], \)
2. \( \tilde{\Delta} \equiv \frac{\partial v(\Sigma(\theta_0)^{-1})}{\partial \gamma_0'} = -\hat{H} \gamma \Delta, \) where \( \hat{H}_\gamma = D_K^+ \{ \Sigma(\theta_0)^{-1} \otimes \Sigma(\theta_0)^{-1} \} D_K. \)

Proof. The proof of 5 follows from the local identification of the AB model. See, for example, Lütkepohl (2007, Proposition 9.3). Using the chain rule for matrix derivatives, we obtain
\[
\frac{\partial v(\Sigma(\theta_0)^{-1})}{\partial \gamma_0'} = \frac{\partial v(\Sigma(\theta_0)^{-1})}{\partial v(\Sigma(\theta_0)^{-1})'} \frac{\partial v(\Sigma(\theta_0)^{-1})}{\partial \gamma_0'},
\]
where we have again used a matrix derivative formula from Schott (2005, Corollary 9.4.1). \( \square \)

Proof of Theorem 1 The proof of 1 is readily obtained from (18) and Lemma 4. For the proof of 2, note that
\[
\bar{e}_n = v \left( \hat{S}_n^{-1} - \Sigma_U^{-1} \right) + v \left\{ \Sigma_U^{-1} - \Sigma(\theta_0)^{-1} \right\} - v \left\{ \Sigma(\hat{\gamma}_n)^{-1} - \Sigma(\theta_0)^{-1} \right\}, \tag{A.13}
\]
We have from Lemma 5.2 and Lemma 4.3 that, under the null hypothesis,
\[
\sqrt{n}\tilde{\Delta}_n \bar{e}_n = \sqrt{n}\tilde{\Delta}_n' v \left( \hat{S}_n^{-1} - \Sigma_U^{-1} \right) + o_p(1)
\rightarrow N(0, \tilde{\Delta}_n' \hat{H} \Omega \hat{H}' \tilde{\Delta}_n).
\]
Because the rank of the asymptotic variance is \( \text{rank}(\tilde{\Delta}_n) = K^* - l \), we obtain the desired result. \( \square \)
REFERENCES


