Estimation of Vector Autoregressive Models 13

13.1 Introduction

In this chapter we derive the Least-Squares (LS) estimator for vectorautoregressive (VAR) models and its asymptotic distribution. For this end, we have to make several assumption which we maintain throughout this chapter.

Assumption 13.1. *The VAR process $\{X_t\}$ is generated by*

$$\Phi(\mathrm{L})X_t = Z_t$$
$$X_t - \Phi_1 X_{t-1} - \cdots - \Phi_p X_{t-p} = Z_t \qquad \text{with } Z_t \sim \mathrm{WN}(0, \Sigma),$$

Σ nonsingular, and admits a stationary and causal representation with respect to $\{Z_t\}$:

$$X_t = Z_t + \Psi_1 Z_{t-1} + \Psi_2 Z_{t-2} + \ldots = \sum_{j=0}^{\infty} \Psi_j Z_t = \Psi(\mathrm{L})Z_t$$

with $\sum_{j=0}^{\infty} \|\Psi_j\| < \infty$.

Assumption 13.2. *The residual process $\{Z_t\}$ is not only white noise, but also independently and identically distributed:*

$$Z_t \sim \mathrm{IID}(0, \Sigma).$$

Assumption 13.3. *All fourth moments of Z_t exist. In particular, there exists a finite constant $c > 0$ such that*

$$\mathbb{E}\left(Z_{it}Z_{jt}Z_{kt}Z_{lt}\right) \le c \qquad \text{for all } i, j, k, l = 1, 2, \ldots, n, \text{ and for all } t.$$

© Springer International Publishing Switzerland 2016

K. Neusser, *Time Series Econometrics*, Springer Texts in Business and Economics,
DOI 10.1007/978-3-319-32862-1_13

Note that the moment condition is automatically fulfilled by Gaussian processes. For the ease of exposition, we omit a constant in the VAR. Thus, we consider the demeaned process.

13.2 The Least-Squares Estimator

Let us denote by $\phi_{ij}^{(k)}$ the (i,j)-th element of the matrix Φ_k, $k = 1, 2, \ldots, p$, then the i-th equation, $i = 1, \ldots, n$, can be written as

$$X_{it} = \phi_{i1}^{(1)} X_{1,t-1} + \ldots + \phi_{in}^{(1)} X_{n,t-1} + \ldots + \phi_{i1}^{(p)} X_{1,t-p} + \ldots + \phi_{in}^{(p)} X_{n,t-p} + Z_{it}.$$

We can view this equation as a regression equation of X_{it} on all lagged variables $X_{1,t-1}, \ldots, X_{n,t-1}, \ldots, X_{1,t-p}, \ldots, X_{n,t-p}$ with error term Z_{it}. Note that the regressors are the same for each equation. The np regressors have coefficient vector $\left(\phi_{i1}^{(1)}, \ldots, \phi_{in}^{(1)}, \ldots, \phi_{i1}^{(p)}, \ldots, \phi_{in}^{(p)}\right)'$. Thus, the complete VAR(p) model has n^2p coefficients in total to be estimated. In addition, there are $n(n+1)/2$ independent elements of the covariance matrix Σ that have to be estimated too.

It is clear that the n different equations are linked through the regressors and the errors terms which in general have non-zero covariances $\sigma_{ij} = \mathbb{E} Z_{it} Z_{jt}$. Hence, it seems warranted to take a systems approach and to estimate all equations of the VAR jointly. Below, we will see that an equation-by-equation approach is, however, still appropriate.

Suppose that we have $T + p$ observations with $t = -p+1, \ldots, 0, 1, \ldots, T$, then we can write the regressor matrix for each equation compactly as a $T \times np$ matrix $\mathbf{X}$:

$$\mathbf{X} = \begin{pmatrix} X_{1,0} & \ldots & X_{n,0} & \ldots & X_{1,-p+1} & \ldots & X_{n,-p+1} \\ X_{1,1} & \ldots & X_{n,1} & \ldots & X_{1,-p+2} & \ldots & X_{n,-p+2} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{1,T-1} & \ldots & X_{n,T-1} & \ldots & X_{1,T-p} & \ldots & X_{n,T-p} \end{pmatrix}.$$

Using this notation, we can write the VAR for observations $t = 1, 2, \ldots, T$ as

$$\underbrace{(X_1, X_2, \ldots, X_T)}_{=Y} = \underbrace{(\Phi_1, \Phi_2, \ldots, \Phi_p)}_{=\Phi} \underbrace{\begin{pmatrix} X_0 & X_1 & \ldots & X_{T-1} \\ X_{-1} & X_0 & \ldots & X_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{-p+1} & X_{-p+2} & \ldots & X_{T-p} \end{pmatrix}}_{=\mathbf{X}'} + \underbrace{(Z_1, Z_2, \ldots, Z_T)}_{=Z}$$

or more compactly

$$Y = \Phi \mathbf{X}' + Z.$$

There are two ways to bring this equation system in the usual multivariate regression framework. One can either arrange the data according to observations or according to equations. Ordered in terms of observations yields:

$$\operatorname{vec} Y = \operatorname{vec}(\Phi \mathbf{X}') + \operatorname{vec} Z = (\mathbf{X} \otimes I_n) \operatorname{vec} \Phi + \operatorname{vec} Z \tag{13.1}$$

with $\operatorname{vec} Y = (X_{11}, X_{21}, \ldots, X_{n1}, X_{12}, X_{22}, \ldots, X_{n2}, \ldots, X_{1T}, X_{2T}, \ldots, X_{nT})'$. If the data are arranged equation by equation, the dependent variable is $\operatorname{vec} Y' = (X_{11}, X_{12}, \ldots, X_{1T}, X_{21}, X_{22}, \ldots, X_{2T}, \ldots, X_{n1}, X_{n2}, \ldots, X_{nT})'$. As both representations, obviously, contain the same information, there exists a $nT \times nT$ permutation or commutation matrix K_{nT} such that $\operatorname{vec} Y' = K_{nt} \operatorname{vec} Y$. Using the computation rules for the Kronecker product, the vec operator, and the permutation matrix (see Magnus and Neudecker 1988), we get for the ordering in terms of equations

$$\begin{aligned}
\operatorname{vec} Y' &= K_{nT} \operatorname{vec} Y = K_{nT} \big(\operatorname{vec}(\Phi \mathbf{X}') + \operatorname{vec} Z\big) \\
&= K_{nT}(\mathbf{X} \otimes I_n) \operatorname{vec} \Phi + K_{nT} \operatorname{vec} Z \\
&= (I_n \otimes \mathbf{X}) K_{n^2 p} \operatorname{vec} \Phi + K_{nT} \operatorname{vec} Z \\
&= (I_n \otimes \mathbf{X}) \operatorname{vec} \Phi' + \operatorname{vec} Z'
\end{aligned} \tag{13.2}$$

where $K_{n^2 p}$ is the corresponding $n^2 \times p$ permutation matrix relating $\operatorname{vec} \Phi$ and $\operatorname{vec} \Phi'$.

The error terms of the different equations are correlated because, in general, the covariances $\sigma_{ij} = \mathbb{E} Z_{it} Z_{jt}$ are nonzero. In the case of an arrangement by observation the covariance matrix of the error term $\operatorname{vec} Z$ is

$$\begin{aligned}
\mathbb{V} \operatorname{vec} Z &= \mathbb{E}(\operatorname{vec} Z)(\operatorname{vec} Z)' \\
&= \begin{pmatrix}
\sigma_1^2 \ldots \sigma_{1n} & 0 & \ldots & 0 & \ldots & 0 & \ldots & 0 \\
\vdots \quad \ddots \quad \vdots & \vdots & \ddots & \vdots & \ldots & \vdots & \ddots & \vdots \\
\sigma_{n1} \ldots \sigma_n^2 & 0 & \ldots & 0 & \ldots & 0 & \ldots & 0 \\
0 \quad 0 \quad 0 & \sigma_1^2 & \ldots & \sigma_{1n} & \ldots & 0 & \ldots & 0 \\
\vdots \quad \ddots \quad \vdots & \vdots & \ddots & \vdots & \ldots & \vdots & \ddots & \vdots \\
0 \ldots 0 & \sigma_{n1} & \ldots & \sigma_n^2 & \ldots & 0 & \ldots & 0 \\
\vdots \quad \vdots \quad \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 \ldots 0 & 0 & \ldots & 0 & \ldots & \sigma_1^2 & \ldots & \sigma_{1n} \\
\vdots \quad \ddots \quad \vdots & \vdots & \ddots & \vdots & \ldots & \vdots & \ddots & \vdots \\
0 \ldots 0 & 0 & \ldots & 0 & \ldots & \sigma_{n1} & \ldots & \sigma_n^2
\end{pmatrix} = I_T \otimes \Sigma.
\end{aligned}$$

In the second case, the arrangement by equation, the covariance matrix of the error term $\operatorname{vec} Z'$ is

$$\mathbb{V}\operatorname{vec} Z' = \mathbb{E}(\operatorname{vec} Z')(\operatorname{vec} Z')'$$

$$= \begin{pmatrix} \sigma_1^2 & \ldots & 0 & \sigma_{12} & \ldots & 0 & \ldots & \sigma_{1n} & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ldots & \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_1^2 & 0 & \ldots & \sigma_{12} & \ldots & 0 & \ldots & \sigma_{1n} \\ \sigma_{21} & \ldots & 0 & \sigma_2^2 & \ldots & 0 & \ldots & \sigma_{2n} & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ldots & \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_{21} & 0 & \ldots & \sigma_2^2 & \ldots & 0 & \ldots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \sigma_{n1} & \ldots & 0 & \sigma_{n2} & \ldots & 0 & \ldots & \sigma_n^2 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ldots & \vdots & \ddots & \vdots \\ 0 & \ldots & \sigma_{n1} & 0 & \ldots & \sigma_{n2} & \ldots & 0 & \ldots & \sigma_n^2 \end{pmatrix} = \Sigma \otimes I_T.$$

Given that the covariance matrix is not a multiple of the identity matrix, efficient estimation requires the use of generalized least squares (GLS). The GLS estimator minimizes the weighted sum of squared errors

$$S(\operatorname{vec}\Phi) = (\operatorname{vec} Z)'(I_T \otimes \Sigma)^{-1}(\operatorname{vec} Z) \longrightarrow \min_{\Phi}.$$

The solution of this minimization problem can be found in standard econometric textbooks like (Dhrymes 1978; Greene 2008; Hamilton 1994b) and is given by

$$\begin{aligned}
(\operatorname{vec}\widehat{\Phi})_{\mathrm{GLS}} &= \left((\mathbf{X}\otimes I_n)'(I_T\otimes\Sigma)^{-1}(\mathbf{X}\otimes I_n)\right)^{-1}(\mathbf{X}\otimes I_n)'(I_T\otimes\Sigma)^{-1}\operatorname{vec} Y \\
&= \left((\mathbf{X}'\otimes I_n)(I_T\otimes\Sigma^{-1})(\mathbf{X}\otimes I_n)\right)^{-1}(\mathbf{X}'\otimes I_n)(I_T\otimes\Sigma^{-1})\operatorname{vec} Y \\
&= \left((\mathbf{X}'\otimes\Sigma^{-1})(\mathbf{X}\otimes I_n)\right)^{-1}(\mathbf{X}'\otimes\Sigma^{-1})\operatorname{vec} Y \\
&= \left((\mathbf{X}'\mathbf{X})\otimes\Sigma^{-1}\right)^{-1}(\mathbf{X}'\otimes\Sigma^{-1})\operatorname{vec} Y \\
&= \left((\mathbf{X}'\mathbf{X})^{-1}\otimes\Sigma\right)(\mathbf{X}'\otimes\Sigma^{-1})\operatorname{vec} Y = \left(((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\otimes I_n\right)\operatorname{vec} Y \\
&= (\operatorname{vec}\widehat{\Phi})_{\mathrm{OLS}}
\end{aligned}$$

As the covariance matrix Σ cancels, the GLS and the OLS-estimator deliver numerically exactly the same solution. The reason for this result is that the regressors are the same in each equation. If this does not hold, for example when some coefficients are set a priori to zero, efficient estimation would require the use of GLS.

Further insights can be gained by rewriting the estimation problem in terms of the arrangement by equation (see Eq. (13.2)). For this purpose, multiply the above estimator from the left by the commutation matrix K_{n^2p}[1]:

$$\begin{aligned}(\operatorname{vec}\widehat{\Phi}')_{\mathrm{OLS}} &= K_{n^2p}(\operatorname{vec}\widehat{\Phi})_{\mathrm{OLS}} = K_{n^2p}\left(((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\otimes I_n\right)\operatorname{vec}Y\\ &= \left(I_n\otimes((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\right)K_{nT}\operatorname{vec}Y = \left(I_n\otimes((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\right)\operatorname{vec}Y'.\end{aligned}$$

This can be written in a more explicit form as

$$\begin{aligned}(\operatorname{vec}\widehat{\Phi}')_{\mathrm{OLS}} &= \begin{pmatrix}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' & 0 & \ldots & 0\\ 0 & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\end{pmatrix}\operatorname{vec}Y'\\ &= \begin{pmatrix}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y_1 & 0 & \ldots & 0\\ 0 & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y_2 & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y_n\end{pmatrix}\end{aligned}$$

where Y_i, $i = 1,\ldots,n$, stacks the observations of the i-th variable such that $Y_i = (X_{i1}, X_{i2}, \ldots, X_{iT})'$. Thus, the estimation of VAR as a system can be broken down into the estimation of n regression equations with dependent variable X_{it}. Each of these equations can then be estimated by OLS.

Thus, we have proven that

$$\operatorname{vec}\widehat{\Phi} = (\operatorname{vec}\widehat{\Phi})_{\mathrm{GLS}} = (\operatorname{vec}\widehat{\Phi})_{\mathrm{OLS}} = \left(((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\otimes I_n\right)\operatorname{vec}Y, \tag{13.3}$$

$$\operatorname{vec}\widehat{\Phi}' = (\operatorname{vec}\widehat{\Phi}')_{\mathrm{GLS}} = (\operatorname{vec}\widehat{\Phi}')_{\mathrm{OLS}} = \left(I_n\otimes((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\right)\operatorname{vec}Y'. \tag{13.4}$$

The least squares estimator can also be rewritten without the use of the *vec*-operator:

$$\widehat{\Phi} = Y\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

Under the assumptions stated in the Introduction Sect. 13.1, these estimators are consistent and asymptotically normal.

Theorem 13.1 (Asymptotic Distribution of OLS Estimator)**.** *Under the assumption stated in the Introduction Sect. 13.1, it holds that*

$$\operatorname{plim}\widehat{\Phi} = \Phi$$

[1] Alternatively, one could start from scratch and investigate the minimization problem $S(\operatorname{vec}\Phi') = (\operatorname{vec}Z')'(\Sigma^{-1}\otimes I_T)(\operatorname{vec}Z') \to \min_{\Phi}$.

and that

$$\text{by observation:} \qquad \sqrt{T}\left(\operatorname{vec}\widehat{\Phi} - \operatorname{vec}\Phi\right) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \mathbf{\Gamma}_p^{-1} \otimes \Sigma\right),$$

respectively,

$$\text{by equation:} \qquad \sqrt{T}\left(\operatorname{vec}\widehat{\Phi}' - \operatorname{vec}\Phi'\right) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \Sigma \otimes \mathbf{\Gamma}_p^{-1}\right)$$

where $\mathbf{\Gamma}_p = \operatorname{plim}\frac{1}{T}(\mathbf{X}'\mathbf{X})$.

Proof. See Sect. 13.3. □

In order to make use of this result in practice, we have to replace the matrices Σ and $\mathbf{\Gamma}_p$ by some estimate. A natural consistent estimate of $\mathbf{\Gamma}_p$ is given according to Proposition 13.1 by

$$\widehat{\mathbf{\Gamma}}_p = \frac{\mathbf{X}'\mathbf{X}}{T}.$$

In analogy to the multivariate regression model, a natural estimator for Σ can be obtained from the Least-Squares residuals $\widehat{Z}$:

$$\widehat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}\widehat{Z}_t\widehat{Z}_t' = \frac{\widehat{Z}\widehat{Z}'}{T} = \frac{(Y - \widehat{\Phi}\mathbf{X}')(Y - \widehat{\Phi}\mathbf{X}')'}{T}.$$

The property of this estimator is summarized in the proposition below.

Theorem 13.2. *Under the condition of Theorem 13.1*

$$\operatorname{plim}\sqrt{T}\left(\widehat{\Sigma} - \frac{ZZ'}{T}\right) = 0$$

Proof. See Sect. 13.3. □

An alternative, but asymptotically equivalent estimator $\widetilde{\Sigma}$ is obtained by adjusting $\widehat{\Sigma}$ for the degrees of freedom:

$$\widetilde{\Sigma} = \frac{T}{T - np}\widehat{\Sigma}. \tag{13.5}$$

If the VAR contains a constant, as is normally the case in practice, the degrees of freedom correction should be $T - np - 1$.

Small sample inference with respect to the parameters Φ can therefore be carried out using the approximate distribution

$$\operatorname{vec}\widehat{\Phi} \sim \mathrm{N}\left(\operatorname{vec}\Phi, \widehat{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}\right). \tag{13.6}$$

This implies that hypothesis testing can be carried out using the conventional t- and F-statistics. From a system perspective, the appropriate degree of freedom for the t-ratio would be $nT - n^2p - n$, taking a constant in each equation into account. However, as that the system can be estimated on an equation by equation basis, it seems reasonable to use $T - np - 1$ instead. This corresponds to a multivariate regression setting with T observation and $np + 1$ regressors, including a constant.

However, as in the univariate case the Gauss Markov theorem does not apply because the lagged regressors are correlated with past error terms. This results in biased estimates in small samples. The amount of the bias can be assessed and corrected either by analytical or bootstrap methods. For an overview, a comparison of the different corrections proposed in the literature, and further references see Engsteg and Pedersen (2014).

13.3 Proofs of the Asymptotic Properties of the Least-Squares Estimator

Lemma 13.1. *Given the assumptions made in Sect. 13.1, the process* $\{\operatorname{vec} Z_{t-j}Z'_{t-i}\}$, $i, j \in \mathbb{Z}$ *and* $i \neq j$, *is white noise.*

Proof. Using the independence assumption of $\{Z_t\}$, we immediately get

$$\begin{aligned}
\mathbb{E}\operatorname{vec} Z_{t-j}Z'_{t-i} &= \mathbb{E}(Z_{t-i} \otimes Z_{t-j}) = 0,\\
\mathbb{V}(\operatorname{vec} Z_{t-j}Z'_{t-i}) &= \mathbb{E}\left((Z_{t-i} \otimes Z_{t-j})(Z_{t-i} \otimes Z_{t-j})'\right)\\
&= \mathbb{E}\left((Z_{t-i}Z'_{t-i}) \otimes (Z_{t-j}Z'_{t-j})\right) = \Sigma \otimes \Sigma,\\
\Gamma_{\operatorname{vec} Z_{t-j}Z'_{t-i}}(h) &= \mathbb{E}\left((Z_{t-i} \otimes Z_{t-j})(Z_{t-i-h} \otimes Z_{t-j-h})'\right)\\
&= \mathbb{E}\left((Z_{t-i}Z'_{t-i-h}) \otimes (Z_{t-j}Z'_{t-j-h})\right) = 0, \qquad h \neq 0.
\end{aligned}$$

□

Under the assumption put forward in the Introduction, $\frac{1}{T}(\mathbf{X}'\mathbf{X})$ converges in probability for $T \to \infty$ to a $np \times np$ matrix $\boldsymbol{\Gamma}_p$. This matrix consists of p^2 blocks where each (i, j)-th block corresponds to the covariance matrix $\Gamma(i - j)$. Thus we have the following proposition:

Proposition 13.1. *Under the assumption stated in the Introduction Sect. 13.1*

$$\frac{\mathbf{X}'\mathbf{X}}{T} \xrightarrow{p} \mathbf{\Gamma}_p = \begin{pmatrix} \Gamma(0) & \Gamma(1) & \dots & \Gamma(p-1) \\ \Gamma'(1) & \Gamma(0) & \dots & \Gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma'(p-1) & \Gamma'(p-2) & \dots & \Gamma(0) \end{pmatrix}.$$

with $\mathbf{\Gamma}_p$ *being nonsingular.*

Proof. Write $\frac{1}{T}(\mathbf{X}'\mathbf{X})$ as

$$\frac{\mathbf{X}'\mathbf{X}}{T} = \begin{pmatrix} \widehat{\Gamma}(0) & \widehat{\Gamma}(1) & \dots & \widehat{\Gamma}(p-1) \\ \widehat{\Gamma}'(1) & \widehat{\Gamma}(0) & \dots & \widehat{\Gamma}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\Gamma}'(p-1) & \widehat{\Gamma}'(p-2) & \dots & \widehat{\Gamma}(0) \end{pmatrix}$$

where

$$\widehat{\Gamma}(h) = \frac{1}{T}\sum_{t=0}^{T-1} X_t X'_{t-h}, \qquad h = 0, 1, \dots, p-1.$$

We will show that each component $\widehat{\Gamma}(h)$ of $\frac{1}{T}\mathbf{X}'\mathbf{X}$ converges in probability to $\Gamma(h)$. Taking the causal representation of $\{X_t\}$ into account

$$\begin{aligned} \widehat{\Gamma}(h) &= \frac{1}{T}\sum_{t=0}^{T-1} X_t X'_{t-h} = \frac{1}{T}\sum_{t=0}^{T-1}\sum_{j=0}^{\infty}\sum_{i=0}^{\infty} \Psi_j Z_{t-j} Z'_{t-h-i} \Psi'_i \\ &= \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} \Psi_j \left(\frac{1}{T}\sum_{t=0}^{T-1} Z_{t-j} Z'_{t-h-i}\right) \Psi'_i \\ &= \sum_{j=0}^{\infty}\sum_{i=h}^{\infty} \Psi_j \left(\frac{1}{T}\sum_{t=0}^{T-1} Z_{t-j} Z'_{t-i}\right) \Psi'_{i-h}. \end{aligned}$$

According to Lemma 13.1 above $\{Z_{t-j}Z'_{t-i}\}$, $i \neq j$, is white noise. Thus,

$$\frac{1}{T}\sum_{t=0}^{T-1} Z_{t-j} Z'_{t-i} \xrightarrow{\text{p}} 0, \qquad i \neq j,$$

according to Theorem 11.1. Hence, for m fixed,

$$G_m(h) = \sum_{j=0}^{m} \sum_{\substack{i=h \\ i \neq j}}^{m+h} \Psi_j \left(\frac{1}{T} \sum_{t=0}^{T-1} Z_{t-j} Z'_{t-i} \right) \Psi'_{i-h} \xrightarrow{\ \mathrm{p}\ } 0.$$

Taking absolute values and expectations element-wise,

$$\begin{aligned}
\mathbb{E}\, |G_\infty(h) - G_m(h)| &= \mathbb{E} \left| \sum_{\substack{j>m \text{ or } i>m+h \\ i \neq j}} \Psi_j \left(\frac{1}{T} \sum_{t=0}^{T-1} Z_{t-j} Z'_{t-i} \right) \Psi'_{i-h} \right| \\
&\leq \sum_{\substack{j>m \text{ or } i>m+h \\ i \neq j}} |\Psi_j| \left(\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}|Z_{t-j} Z'_{t-i}| \right) |\Psi'_{i-h}| \\
&\leq \sum_{\substack{j>m \text{ or } i>m+h \\ i \neq j}} |\Psi_j| \left(\mathbb{E}|Z_1 Z'_2| \right) |\Psi'_{i-h}| \\
&\leq \sum_{\substack{j>m \text{ or } i>m \\ i \neq j}} |\Psi_j| \left(\mathbb{E}|Z_1 Z'_2| \right) |\Psi'_i| \\
&\leq \sum_{j>m} |\Psi_j| \left(\mathbb{E}|Z_1 Z'_2| \sum_i |\Psi'_i| \right) \\
&\qquad + \left(\sum_j |\Psi_j|\ \mathbb{E}|Z_1 Z'_2| \right) \sum_{i>m} |\Psi'_i|
\end{aligned}$$

As the bound is independent of T and converges to 0 as $m \to \infty$, we have

$$\lim_{m \to \infty} \limsup_{T \to \infty} \mathbb{E}\, |G_\infty(h) - G_m(h)| = 0.$$

The Basic Approximation Theorem C.14 then establishes that

$$G_\infty(h) \xrightarrow{\ \mathrm{p}\ } 0.$$

Henceforth

$$\begin{aligned}
\widehat{\Gamma}(h) &= G_\infty(h) + \sum_{j=h}^{\infty} \Psi_j \left(\frac{1}{T} \sum_{t=0}^{T-1} Z_{t-j} Z'_{t-j} \right) \Psi'_{j-h} \\
&= G_\infty(h) + \sum_{j=h}^{\infty} \Psi_j \left(\frac{1}{T} \sum_{t=0}^{T-1} Z_t Z'_t \right) \Psi'_{j-h} + \text{remainder}
\end{aligned}$$

where the remainder only depends on initial conditions[2] and is therefore negligible as $T \to \infty$. As

$$\frac{1}{T}\sum_{t=0}^{T-1} Z_t Z_t' \xrightarrow{\ \mathrm{p}\ } \Sigma,$$

we finally get

$$\widehat{\Gamma}(h) \xrightarrow{\ \mathrm{p}\ } \sum_{j=h}^{\infty} \Psi_j \Sigma \Psi_{j-h}' = \Gamma(h).$$

The last equality follows from Theorem 10.2. □

Proposition 13.2. *Under the assumption stated in the Introduction Sect. 13.1*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \operatorname{vec}(Z_t X_{t-1}', Z_t X_{t-2}', \ldots, Z_t X_{t-p}')$$
$$= \frac{1}{\sqrt{T}} \operatorname{vec}(Z\mathbf{X}) = \frac{1}{\sqrt{T}}(\mathbf{X}' \otimes I_n) \operatorname{vec} Z$$
$$\xrightarrow{\ d\ } \mathrm{N}(0, \mathbf{\Gamma}_p \otimes \Sigma)$$

Proof. The idea of the proof is to approximate $\{X_t\}$ by some simpler process $\{X_t^{(m)}\}$ which allows the application of the CLT for dependent processes (Theorem C.13). This leads to an asymptotic distribution which by the virtue of the Basic Approximation Theorem C.14 converges to the asymptotic distribution of the original process. Define $X_t^{(m)}$ as the truncated process from the causal presentation of X_t:

$$X_t^{(m)} = Z_t + \Psi_1 Z_{t-1} + \ldots + \Psi_m Z_{t-m}, \qquad m = p, p+1, p+2, \ldots$$

Using this approximation, we can then define the process $\{Y_t^{(m)}\}$ as

$$Y_t^{(m)} = \operatorname{vec}\left(Z_t X_{t-1}^{(m)\prime}, Z_t X_{t-2}^{(m)\prime}, \ldots, Z_t X_{t-p}^{(m)\prime}\right) = \begin{pmatrix} X_{t-1}^{(m)} \\ X_{t-2}^{(m)} \\ \vdots \\ X_{t-p}^{(m)} \end{pmatrix} \otimes Z_t.$$

[2]See the proof of Theorem 11.2.2 in Brockwell and Davis (1991) for details.

Due to the independence of $\{Z_t\}$ this process is a mean zero white noise process, but is clearly not independent. It is easy to see that the process is actually $(m+p)$-dependent with variance $\mathbf{V}_m$ given by

$$\begin{aligned}
\mathbf{V}_m = \mathbb{E}Y_t^{(m)}Y_t^{(m)\prime} &= \mathbb{E}\left(\left(\begin{pmatrix} X_{t-1}^{(m)} \\ X_{t-2}^{(m)} \\ \vdots \\ X_{t-p}^{(m)} \end{pmatrix} \otimes Z_t\right)\left(\begin{pmatrix} X_{t-1}^{(m)} \\ X_{t-2}^{(m)} \\ \vdots \\ X_{t-p}^{(m)} \end{pmatrix} \otimes Z_t\right)'\right) \\
&= \left(\mathbb{E}\begin{pmatrix} X_{t-1}^{(m)} \\ X_{t-2}^{(m)} \\ \dots \\ X_{t-p}^{(m)} \end{pmatrix}\begin{pmatrix} X_{t-1}^{(m)} \\ X_{t-2}^{(m)} \\ \dots \\ X_{t-p}^{(m)} \end{pmatrix}'\right) \otimes \mathbb{E}Z_tZ_t' \\
&= \begin{pmatrix} \Gamma^{(m)}(0) & \Gamma^{(m)}(1) & \dots & \Gamma^{(m)}(p-1) \\ \Gamma^{(m)}(1)' & \Gamma^{(m)}(0) & \dots & \Gamma^{(m)}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma^{(m)}(p-1)' & \Gamma^{(m)}(p-2)' & \dots & \Gamma^{(m)}(0) \end{pmatrix} \otimes \Sigma \\
&= \mathbf{\Gamma}_p^{(m)} \otimes \Sigma
\end{aligned}$$

where $\Gamma_p^{(m)}$ is composed of

$$\begin{aligned}
\Gamma^{(m)}(h) &= \mathbb{E}X_{t-1}^{(m)}X_{t-1-h}^{(m)\prime} \\
&= \mathbb{E}\left(Z_{t-1} + \Psi_1 Z_{t-2} + \ldots + \Psi_m Z_{t-1-m}\right) \\
&\qquad \left(Z_{t-1-h} + \Psi_1 Z_{t-2-h} + \ldots + \Psi_m Z_{t-1-m-h}\right)' \\
&= \sum_{j=h}^{m} \Psi_j \Sigma \Psi_{j-h}', \qquad h = 0, 1, \ldots, p-1.
\end{aligned}$$

Thus, we can invoke the CLT for $(m+p)$-dependent process (see Theorem C.13) to establish that

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} Y_t^{(m)}\right) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \mathbf{V}_m).$$

For $m \to \infty$, $\Gamma^{(m)}(h)$ converges to $\Gamma(h)$ and thus $\mathbf{\Gamma}_p^{(m)}$ to $\mathbf{\Gamma}_p$. Therefore, $\mathbf{V}_m \longrightarrow \mathbf{\Gamma}_p \otimes \Sigma$.

The variance of the approximation error is equal to

$$
\begin{aligned}
&\mathbb{V}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\operatorname{vec}(Z_tX'_{t-1},\ldots,Z_tX'_{t-p}) - Y_t^{(m)}\right)\right)\\
&= \frac{1}{T}\mathbb{V}\left(\sum_{t=1}^{T}\operatorname{vec}\left(Z_t(X_{t-1}-X_{t-1}^{(m)})',\ldots,Z_t(X_{t-p}-X_{t-p}^{(m)})'\right)\right)\\
&= \frac{1}{T}\mathbb{V}\left(\sum_{t=1}^{T}\operatorname{vec}\left(Z_t\left(\sum_{j=m+1}^{\infty}\Psi_jZ_{t-1-j}\right)',\ldots,Z_t\left(\sum_{j=m+1}^{\infty}\Psi_jZ_{t-p-j}\right)'\right)\right)\\
&= \mathbb{V}\left(\operatorname{vec}\left(Z_t\left(\sum_{j=m+1}^{\infty}\Psi_jZ_{t-1-j}\right)',\ldots,Z_t\left(\sum_{j=m+1}^{\infty}\Psi_jZ_{t-p-j}\right)'\right)\right)\\
&= \mathbb{E}\left(\left(\begin{pmatrix}\sum_{j=m+1}^{\infty}\Psi_jZ_{t-1-j}\\ \vdots\\ \sum_{j=m+1}^{\infty}\Psi_jZ_{t-p-j}\end{pmatrix}\otimes Z_t\right)\left(\begin{pmatrix}\sum_{j=m+1}^{\infty}\Psi_jZ_{t-1-j}\\ \vdots\\ \sum_{j=m+1}^{\infty}\Psi_jZ_{t-p-j}\end{pmatrix}\otimes Z_t\right)'\right)\\
&= \begin{pmatrix}\sum_{j=m+1}^{\infty}\Psi_j\Sigma\Psi_j' \;\ldots & \ldots\\ \vdots & \ddots & \vdots\\ \ldots & \ldots\; \sum_{j=m+1}^{\infty}\Psi_j\Sigma\Psi_j'\end{pmatrix}\otimes\Sigma.
\end{aligned}
$$

The absolute summability of Ψ_j then implies that the infinite sums converge to zero as $m \to \infty$. As $X_t^{(m)} \xrightarrow{\text{m.s.}} X_t$ for $m \to \infty$, we can apply the Basic Approximation Theorem C.14 to reach the required conclusion

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\operatorname{vec}(Z_tX'_{t-1}, Z_tX'_{t-2},\ldots,Z_tX'_{t-p}) \xrightarrow{\text{d}} \mathrm{N}(0,\boldsymbol{\Gamma}_p\otimes\Sigma).$$

□

Proof of Theorem 13.1

Proof. We prove the Theorem for the arrangement by observation. The prove for the arrangement by equation can be proven in a completely analogous way. Inserting the regression formula (13.1) into the least-squares formula (13.3) leads to:

$$
\begin{aligned}
\operatorname{vec}\widehat{\Phi} &= (((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\otimes I_n)(\mathbf{X}\otimes I_n)\operatorname{vec}\Phi + (((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\otimes I_n)\operatorname{vec}Z\\
&= \operatorname{vec}\Phi + (((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\otimes I_n)\operatorname{vec}Z. \qquad (13.7)
\end{aligned}
$$

Bringing vec Φ to the left hand side and taking the probability limit, we get using Slutzky's Lemma C.10 for the product of probability limits

$$\begin{aligned}\operatorname{plim}(\operatorname{vec}\widehat{\Phi} - \operatorname{vec}\Phi) &= \operatorname{plim}\operatorname{vec}\left(Z\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right)\\ &= \operatorname{vec}\left(\operatorname{plim}\frac{Z\mathbf{X}}{T}\operatorname{plim}\left(\frac{\mathbf{X}'\mathbf{X}}{T}\right)^{-1}\right) = 0.\end{aligned}$$

The last equality follows from the observation that Proposition 13.1 implies $\operatorname{plim}\frac{\mathbf{X}'\mathbf{X}}{T} = \mathbf{\Gamma}_p$ nonsingular and that Proposition 13.2 implies $\operatorname{plim}\frac{Z\mathbf{X}}{T} = 0$. Thus, we have established that the Least-Squares estimator is consistent.

Equation (13.7) further implies:

$$\begin{aligned}\sqrt{T}(\operatorname{vec}\widehat{\Phi} - \operatorname{vec}\Phi) &= \sqrt{T}\left(((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \otimes I_n\right)\operatorname{vec}Z\\ &= \left(\left(\frac{\mathbf{X}'\mathbf{X}}{T}\right)^{-1} \otimes I_n\right)\frac{1}{\sqrt{T}}(\mathbf{X}' \otimes I_n)\operatorname{vec}Z\end{aligned}$$

As $\operatorname{plim}\frac{\mathbf{X}'\mathbf{X}}{T} = \mathbf{\Gamma}_p$ nonsingular, the above expression converges in distribution according to Theorem C.10 and Proposition 13.2 to a normally distributed random variable with mean zero and covariance matrix

$$(\mathbf{\Gamma}_p^{-1} \otimes I_n)(\mathbf{\Gamma}_p \otimes \Sigma)(\mathbf{\Gamma}_p^{-1} \otimes I_n) = \mathbf{\Gamma}_p^{-1} \otimes \Sigma$$

□

Proof of Theorem 13.2

Proof.

$$\begin{aligned}\widehat{\Sigma} &= \frac{(Y - \widehat{\Phi}\mathbf{X}')(Y - \widehat{\Phi}\mathbf{X}')'}{T}\\ &= \frac{(Y - \Phi\mathbf{X}' + (\Phi - \widehat{\Phi})\mathbf{X}')(Y - \Phi\mathbf{X}' + (\Phi - \widehat{\Phi})\mathbf{X}')'}{T}\\ &= \frac{1}{T}(Z + (\Phi - \widehat{\Phi})\mathbf{X}')(Z + (\Phi - \widehat{\Phi})\mathbf{X}')'\\ &= \frac{ZZ'}{T} + \frac{Z\mathbf{X}}{T}(\Phi - \widehat{\Phi})' + (\Phi - \widehat{\Phi})\frac{\mathbf{X}'Z'}{T} + (\Phi - \widehat{\Phi})\frac{\mathbf{X}'\mathbf{X}}{T}(\Phi - \widehat{\Phi})'\end{aligned}$$

Applying Theorem C.7 and the results of Propositions 13.1 and 13.2 shows that

$$\frac{Z\mathbf{X}(\Phi - \widehat{\Phi})'}{\sqrt{T}} \xrightarrow{\mathrm{p}} 0$$

and

$$(\Phi - \widehat{\Phi})\frac{\mathbf{X}'\mathbf{X}}{T}\sqrt{T}(\Phi - \widehat{\Phi})' \xrightarrow{\text{p}} 0.$$

Hence,

$$\sqrt{T}\left(\frac{(Y - \widehat{\Phi}\mathbf{X}')(Y - \widehat{\Phi}\mathbf{X}')'}{T} - \frac{ZZ'}{T}\right) = \sqrt{T}\left(\widehat{\Sigma} - \frac{ZZ'}{T}\right) \xrightarrow{\text{p}} 0$$

$\square$

13.4 The Yule-Walker Estimator

An alternative estimation method can be derived from the Yule-Walker equations. Consider first a VAR(1) model. The Yule-Walker equation in this case simply is:

$$\begin{aligned}\Gamma(0) &= \Phi\Gamma(-1) + \Sigma \\ \Gamma(1) &= \Phi\Gamma(0)\end{aligned}$$

or

$$\begin{aligned}\Gamma(0) &= \Phi\Gamma(0)\Phi' + \Sigma \\ \Gamma(1) &= \Phi\Gamma(0).\end{aligned}$$

The solution of this system of equations is:

$$\begin{aligned}\Phi &= \Gamma(1)\Gamma(0)^{-1} \\ \Sigma &= \Gamma(0) - \Phi\Gamma(0)\Phi' = \Gamma(0) - \Gamma(1)\Gamma(0)^{-1}\Gamma(0)\Gamma(0)^{-1}\Gamma(1)' \\ &= \Gamma(0) - \Gamma(1)\Gamma(0)^{-1}\Gamma(1)'.\end{aligned}$$

Replacing the theoretical moments by their empirical counterparts, we get the *Yule-Walker estimator* for Φ and Σ:

$$\begin{aligned}\widehat{\Phi} &= \widehat{\Gamma}(1)\widehat{\Gamma}(0)^{-1}, \\ \widehat{\Sigma} &= \widehat{\Gamma}(0) - \widehat{\Phi}\widehat{\Gamma}(0)\widehat{\Phi}'.\end{aligned}$$

In the general case of a VAR(p) model the Yule-Walker estimator is given as the solution of the equation system

$$\widehat{\Gamma}(h) = \sum_{j=1}^{p} \widehat{\Phi}_j \widehat{\Gamma}(h-j), \qquad k = 1, \ldots, p,$$

$$\widehat{\Sigma} = \widehat{\Gamma}(0) - \widehat{\Phi}_1 \widehat{\Gamma}(-1) - \ldots - \widehat{\Phi}_p \widehat{\Gamma}(-p)$$

As the least-squares and the Yule-Walker estimator differ only in the treatment of the starting values, they are asymptotically equivalent. In fact, they yield very similar estimates even for finite samples (see e.g. Reinsel (1993)). However, as in the univariate case, the Yule-Walker estimator always delivers, in contrast to the least-square estimator, coefficient estimates with the property $\det(I_n - \hat{\Phi}_1 z - \ldots - \hat{\Phi}_p z^p) \neq 0$ for all $z \in \mathbb{C}$ with $|z| \leq 1$. Thus, the Yule-Walker estimator guarantees that the estimated VAR possesses a causal representation. This, however, comes at the price that the Yule-Walker estimator has a larger small-sample bias than the least-squares estimator, especially when the roots of $\Phi(z)$ get close to the unit circle (Tjøstheim and Paulsen 1983; Shaman and Stine 1988; Reinsel 1993). Thus, it is generally preferable to use the least-squares estimator in practice.