3

Estimation of Vector Autoregressive Processes

3.1 Introduction

In this chapter, it is assumed that a K-dimensional multiple time series $y_1, \dots, y_T$ with $y_t = (y_{1t}, \dots, y_{Kt})'$ is available that is known to be generated by a stationary, stable VAR(p) process

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t. \tag{3.1.1}$$

All symbols have their usual meanings, that is, $\nu = (\nu_1, \dots, \nu_K)'$ is a $(K \times 1)$ vector of intercept terms, the A_i are $(K \times K)$ coefficient matrices and u_t is white noise with nonsingular covariance matrix Σ_u. In contrast to the assumptions of the previous chapter, the coefficients $\nu, A_1, \dots, A_p$, and Σ_u are assumed to be unknown in the following. The time series data will be used to estimate the coefficients. Note that notationwise we do not distinguish between the stochastic process and a time series as a realization of a stochastic process. The particular meaning of a symbol should be obvious from the context.

In the next three sections, different possibilities for estimating a VAR(p) process are discussed. In Section 3.5, the consequences of forecasting with estimated processes will be considered and, in Section 3.6, tests for causality are described. The distribution of impulse responses obtained from estimated processes is considered in Section 3.7.

3.2 Multivariate Least Squares Estimation

In this section, multivariate least squares (LS) estimation is discussed. The estimator obtained for the standard form (3.1.1) of a VAR(p) process is considered in Section 3.2.1. Some properties of the estimator are derived in Sections 3.2.2 and 3.2.4 and an example is given in Section 3.2.3.

3.2.1 The Estimator

It is assumed that a time series $y_1, \dots, y_T$ of the y variables is available, that is, we have a sample of size T for each of the K variables for the same sample period. In addition, p presample values for each variable, $y_{-p+1}, \dots, y_0$, are assumed to be available. Partitioning a multiple time series in sample and presample values is convenient in order to simplify the notation. We define

$$\begin{aligned}
&Y := (y_1, \dots, y_T) && (K \times T),\\
&B := (\nu, A_1, \dots, A_p) && (K \times (Kp+1)),\\
&Z_t := \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix} && ((Kp+1) \times 1),\\
&Z := (Z_0, \dots, Z_{T-1}) && ((Kp+1) \times T),\\
&U := (u_1, \dots, u_T) && (K \times T),\\
&\mathbf{y} := \operatorname{vec}(Y) && (KT \times 1),\\
&\boldsymbol{\beta} := \operatorname{vec}(B) && ((K^2p+K) \times 1),\\
&\mathbf{b} := \operatorname{vec}(B') && ((K^2p+K) \times 1),\\
&\mathbf{u} := \operatorname{vec}(U) && ((KT \times 1).
\end{aligned} \tag{3.2.1}$$

Here vec is the column stacking operator as defined in Appendix A.12.

Using this notation, for $t = 1, \dots, T$, the VAR(p) model (3.1.1) can be written compactly as

$$Y = BZ + U \tag{3.2.2}$$

or

$$\begin{aligned}
\operatorname{vec}(Y) &= \operatorname{vec}(BZ) + \operatorname{vec}(U)\\
&= (Z' \otimes I_K) \operatorname{vec}(B) + \operatorname{vec}(U)
\end{aligned}$$

or

$$\mathbf{y} = (Z' \otimes I_K)\boldsymbol{\beta} + \mathbf{u}. \tag{3.2.3}$$

Note that the covariance matrix of $\mathbf{u}$ is

$$\Sigma_{\mathbf{u}} = I_T \otimes \Sigma_u. \tag{3.2.4}$$

Thus, multivariate LS estimation (or GLS estimation) of $\boldsymbol{\beta}$ means to choose the estimator that minimizes

$$\begin{aligned}
S(\boldsymbol{\beta}) &= \mathbf{u}'(I_T \otimes \Sigma_u)^{-1}\mathbf{u} = \mathbf{u}'(I_T \otimes \Sigma_u^{-1})\mathbf{u} \\
&= [\mathbf{y} - (Z' \otimes I_K)\boldsymbol{\beta}]'(I_T \otimes \Sigma_u^{-1})[\mathbf{y} - (Z' \otimes I_K)\boldsymbol{\beta}] \\
&= \operatorname{vec}(Y - BZ)'(I_T \otimes \Sigma_u^{-1}) \operatorname{vec}(Y - BZ) \\
&= \operatorname{tr}\left[(Y - BZ)'\Sigma_u^{-1}(Y - BZ)\right]. \qquad (3.2.5)
\end{aligned}$$

In order to find the minimum of this function we note that

$$\begin{aligned}
S(\boldsymbol{\beta}) &= \mathbf{y}'(I_T \otimes \Sigma_u^{-1})\mathbf{y} + \boldsymbol{\beta}'(Z \otimes I_K)(I_T \otimes \Sigma_u^{-1})(Z' \otimes I_K)\boldsymbol{\beta} \\
&\quad - 2\boldsymbol{\beta}'(Z \otimes I_K)(I_T \otimes \Sigma_u^{-1})\mathbf{y} \\
&= \mathbf{y}'(I_T \otimes \Sigma_u^{-1})\mathbf{y} + \boldsymbol{\beta}'(ZZ' \otimes \Sigma_u^{-1})\boldsymbol{\beta} - 2\boldsymbol{\beta}'(Z \otimes \Sigma_u^{-1})\mathbf{y}.
\end{aligned}$$

Hence,

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 2(ZZ' \otimes \Sigma_u^{-1})\boldsymbol{\beta} - 2(Z \otimes \Sigma_u^{-1})\mathbf{y}.$$

Equating to zero gives the *normal equations*

$$(ZZ' \otimes \Sigma_u^{-1})\widehat{\boldsymbol{\beta}} = (Z \otimes \Sigma_u^{-1})\mathbf{y} \qquad (3.2.6)$$

and, consequently, the LS estimator is

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= ((ZZ')^{-1} \otimes \Sigma_u)(Z \otimes \Sigma_u^{-1})\mathbf{y} \\
&= ((ZZ')^{-1}Z \otimes I_K)\mathbf{y}. \qquad (3.2.7)
\end{aligned}$$

The Hessian of $S(\boldsymbol{\beta})$,

$$\frac{\partial^2 S}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = 2(ZZ' \otimes \Sigma_u^{-1}),$$

is positive definite which confirms that $\widehat{\boldsymbol{\beta}}$ is indeed a minimizing vector. Strictly speaking, for these results to hold, it has to be assumed that ZZ' is nonsingular. This result will hold with probability 1 if y_t has a continuous distribution which will always be assumed in the following.

It may be worth noting that the multivariate LS estimator $\widehat{\boldsymbol{\beta}}$ is identical to the ordinary LS (OLS) estimator obtained by minimizing

$$\bar{S}(\boldsymbol{\beta}) = \mathbf{u}'\mathbf{u} = [\mathbf{y} - (Z' \otimes I_K)\boldsymbol{\beta}]'[\mathbf{y} - (Z' \otimes I_K)\boldsymbol{\beta}] \qquad (3.2.8)$$

(see Problem 3.1). This result is due to Zellner (1962) who showed that GLS and LS estimation in a multiple equation model are identical if the regressors in all equations are the same.

The LS estimator can be written in different ways that will be useful later on:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= ((ZZ')^{-1}Z \otimes I_K)\left[(Z' \otimes I_K)\boldsymbol{\beta} + \mathbf{u}\right] \\
&= \boldsymbol{\beta} + ((ZZ')^{-1}Z \otimes I_K)\mathbf{u} \qquad (3.2.9)
\end{aligned}$$

or

$$\begin{aligned}
\operatorname{vec}(\widehat{B}) &= \widehat{\boldsymbol{\beta}} = ((ZZ')^{-1}Z \otimes I_K) \operatorname{vec}(Y) \\
&= \operatorname{vec}(YZ'(ZZ')^{-1}).
\end{aligned}$$

Thus,

$$\begin{aligned}
\widehat{B} &= YZ'(ZZ')^{-1} \\
&= (BZ+U)Z'(ZZ')^{-1} \\
&= B + UZ'(ZZ')^{-1}.
\end{aligned} \tag{3.2.10}$$

Another possibility for deriving this estimator results from postmultiplying

$$y_t = BZ_{t-1} + u_t$$

by Z'_{t-1} and taking expectations:

$$E(y_t Z'_{t-1}) = BE(Z_{t-1}Z'_{t-1}). \tag{3.2.11}$$

Estimating $E(y_t Z'_{t-1})$ by

$$\frac{1}{T}\sum_{t=1}^{T} y_t Z'_{t-1} = \frac{1}{T} YZ'$$

and $E(Z_{t-1}Z'_{t-1})$ by

$$\frac{1}{T}\sum_{t=1}^{T} Z_{t-1} Z'_{t-1} = \frac{1}{T} ZZ',$$

we obtain the normal equations

$$\frac{1}{T}YZ' = \widehat{B}\frac{1}{T}ZZ'$$

and, hence, $\widehat{B} = YZ'(ZZ')^{-1}$. Note that (3.2.11) is similar but not identical to the system of Yule-Walker equations in (2.1.37). While central moments about the expectation $\mu = E(y_t)$ are considered in (2.1.37), moments about zero are used in (3.2.11).

Yet another possibility to write the LS estimator is

$$\widehat{\mathbf{b}} = \operatorname{vec}(\widehat{B}') = (I_K \otimes (ZZ')^{-1}Z) \operatorname{vec}(Y'). \tag{3.2.12}$$

In this form, it is particularly easy to see that multivariate LS estimation is equivalent to OLS estimation of each of the K equations in (3.1.1) separately. Let b'_k be the k-th row of B, that is, b_k contains all the parameters of the k-th equation. Obviously $\mathbf{b}' = (b'_1, \dots, b'_k)$. Furthermore, let $y_{(k)} = (y_{k1}, \dots, y_{kT})'$ be the time series available for the k-th variable, so that

$$\operatorname{vec}(Y') = \begin{bmatrix} y_{(1)} \\ \vdots \\ y_{(K)} \end{bmatrix}.$$

With this notation $\widehat{b}_k = (ZZ')^{-1}Zy_{(k)}$ is the OLS estimator of the model $y_{(k)} = Z'b_k + u_{(k)}$, where $u_{(k)} = (u_{k1}, \dots, u_{kT})'$ and $\widehat{\mathbf{b}}' = (\widehat{b}'_1, \dots, \widehat{b}'_K)$.

3.2.2 Asymptotic Properties of the Least Squares Estimator

Because small sample properties of the LS estimator are difficult to derive analytically, we focus on asymptotic properties. Consistency and asymptotic normality of the LS estimator are easily established if the following results hold:

$$\Gamma := \text{plim } ZZ'/T \text{ exists and is nonsingular} \tag{3.2.13}$$

and

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\text{vec}(u_t Z'_{t-1}) = \frac{1}{\sqrt{T}}\text{ vec}(UZ') &= \frac{1}{\sqrt{T}}(Z\otimes I_K)\mathbf{u} \\ &\xrightarrow[T\to\infty]{d} \mathcal{N}(0, \Gamma\otimes\Sigma_u),\end{aligned} \tag{3.2.14}$$

where, as usual, $\xrightarrow{d}$ denotes convergence in distribution. It follows from a theorem due to Mann & Wald (1943) that these results are true under suitable conditions for u_t, if y_t is a stationary, stable VAR(p). For instance, the conditions stated in the following definition are sufficient.

Definition 3.1 (*Standard White Noise*)
A white noise process $u_t = (u_{1t}, \ldots, u_{Kt})'$ is called *standard white noise* if the u_t are continuous random vectors satisfying $E(u_t) = 0, \Sigma_u = E(u_t u_t')$ is nonsingular, u_t and u_s are independent for $s \neq t$, and, for some finite constant c,

$$E|u_{it}u_{jt}u_{kt}u_{mt}| \leq c \qquad \text{for } i,j,k,m = 1,\ldots,K, \text{ and all } t.$$

■

The last condition means that all fourth moments exist and are bounded. Obviously, if the u_t are normally distributed (Gaussian) they satisfy the moment requirements. With this definition it is easy to state conditions for consistency and asymptotic normality of the LS estimator. The following lemma will be essential in proving these large sample results.

Lemma 3.1
If y_t is a stable, K-dimensional VAR(p) process as in (3.1.1) with standard white noise residuals u_t, then (3.2.13) and (3.2.14) hold. ■

Proof: See Theorem 8.2.3 of Fuller (1976, p. 340). ■

The lemma holds also for other definitions of standard white noise. For example, the convergence result in (3.2.14) follows from a central limit theorem for martingale differences or martingale difference arrays (see Proposition C.13) by noting that $w_t = \text{vec}(u_t Z'_{t-1})$ is a martingale difference sequence under quite general conditions. The convergence result in (3.2.13) may then be obtained from a suitable weak law of large numbers (see Proposition C.12). In the next proposition the resulting asymptotic properties of the LS estimator are stated formally.

Proposition 3.1 (*Asymptotic Properties of the LS Estimator*)
Let y_t be a stable, K-dimensional VAR(p) process as in (3.1.1) with standard white noise residuals, $\widehat{B} = YZ'(ZZ')^{-1}$ is the LS estimator of the VAR coefficients B and all symbols are as defined in (3.2.1). Then,

$$\text{plim } \widehat{B} = B$$

and

$$\sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sqrt{T} \text{ vec}(\widehat{B} - B) \xrightarrow{d} \mathcal{N}(0, \Gamma^{-1} \otimes \Sigma_u) \tag{3.2.15}$$

or, equivalently,

$$\sqrt{T}(\widehat{\mathbf{b}} - \mathbf{b}) = \sqrt{T} \text{ vec}(\widehat{B}' - B') \xrightarrow{d} \mathcal{N}(0, \Sigma_u \otimes \Gamma^{-1}), \tag{3.2.16}$$

where $\Gamma = \text{plim } ZZ'/T$. ■

Proof: Using (3.2.10),

$$\text{plim}(\widehat{B} - B) = \text{plim}\left(\frac{UZ'}{T}\right) \text{plim}\left(\frac{ZZ'}{T}\right)^{-1} = 0$$

by Lemma 3.1, because (3.2.14) implies plim $UZ'/T = 0$. Thus, the consistency of $\widehat{B}$ is established.

Using (3.2.9),

$$\begin{aligned}
\sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \sqrt{T}((ZZ')^{-1}Z \otimes I_K)\mathbf{u} \\
&= \left(\left(\frac{1}{T}ZZ'\right)^{-1} \otimes I_K\right) \frac{1}{\sqrt{T}}(Z \otimes I_K)\mathbf{u}.
\end{aligned}$$

Thus, by Proposition C.2(4) of Appendix C, $\sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ has the same asymptotic distribution as

$$\left[\text{plim}\left(\frac{1}{T}ZZ'\right)^{-1} \otimes I_K\right] \frac{1}{\sqrt{T}}(Z \otimes I_K)\mathbf{u} = (\Gamma^{-1} \otimes I_K)\frac{1}{\sqrt{T}}(Z \otimes I_K)\mathbf{u}.$$

Hence, the asymptotic distribution of $\sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is normal by Lemma 3.1 and the covariance matrix is

$$(\Gamma^{-1} \otimes I_K)(\Gamma \otimes \Sigma_u)(\Gamma^{-1} \otimes I_K) = \Gamma^{-1} \otimes \Sigma_u.$$

The result (3.2.16) can be established with similar arguments (see Problem 3.2). ■

As mentioned previously, if u_t is *Gaussian* (normally distributed) white noise, it satisfies the conditions of Proposition 3.1 so that consistency and asymptotic normality of the LS estimator are ensured for stable Gaussian (normally distributed) VAR(p) processes y_t. Note that normality of u_t implies normality of the y_t for stable processes.

In order to assess the asymptotic dispersion of the LS estimator, we need to know the matrices Γ and Σ_u. From (3.2.13) an obvious consistent estimator of Γ is

$$\widehat{\Gamma} = ZZ'/T. \tag{3.2.17}$$

Because $\Sigma_u = E(u_t u_t')$, a plausible estimator for this matrix is

$$\begin{aligned} \widetilde{\Sigma}_u &= \frac{1}{T}\sum_{t=1}^{T} \widehat{u}_t\widehat{u}_t' = \frac{1}{T}\widehat{U}\widehat{U}' = \frac{1}{T}(Y - \widehat{B}Z)(Y - \widehat{B}Z)' \\ &= \frac{1}{T}[Y - YZ'(ZZ')^{-1}Z][Y - YZ'(ZZ')^{-1}Z]' \\ &= \frac{1}{T}Y[I_T - Z'(ZZ')^{-1}Z][I_T - Z'(ZZ')^{-1}Z]'Y' \\ &= \frac{1}{T}Y(I_T - Z'(ZZ')^{-1}Z)Y'. \end{aligned} \tag{3.2.18}$$

Often an adjustment for degrees of freedom is desired because in a regression with fixed, nonstochastic regressors this leads to an unbiased estimator of the covariance matrix. Thus, an estimator

$$\widehat{\Sigma}_u = \frac{T}{T - Kp - 1}\widetilde{\Sigma}_u \tag{3.2.19}$$

may be considered. Note that there are $Kp + 1$ parameters in each of the K equations of (3.1.1) and, hence, there are $Kp+1$ parameters in each equation of the system (3.2.2). Of course, $\widehat{\Sigma}_u$ and $\widetilde{\Sigma}_u$ are asymptotically equivalent. They are consistent estimators of Σ_u if the conditions of Proposition 3.1 hold. In fact, a bit more can be shown.

Proposition 3.2 (*Asymptotic Properties of the White Noise Covariance Matrix Estimators*)
Let y_t be a stable, K-dimensional VAR(p) process as in (3.1.1) with standard white noise innovations and let $\bar{B}$ be an estimator of the VAR coefficients B so that $\sqrt{T}\,\mathrm{vec}(\bar{B} - B)$ converges in distribution. Furthermore, using the symbols from (3.2.1), suppose that

$$\bar{\Sigma}_u = (Y - \bar{B}Z)(Y - \bar{B}Z)'/(T - c),$$

where c is a fixed constant. Then

$$\operatorname{plim}\sqrt{T}(\bar{\Sigma}_u - UU'/T) = 0. \tag{3.2.20}$$

■

Proof:

$$\begin{aligned}\frac{1}{T}(Y-\bar{B}Z)(Y-\bar{B}Z)' &= (B-\bar{B})\left(\frac{ZZ'}{T}\right)(B-\bar{B})'+(B-\bar{B})\frac{ZU'}{T}\\ &\quad +\frac{UZ'}{T}(B-\bar{B})'+\frac{UU'}{T}.\end{aligned}$$

Under the conditions of the proposition, $\text{plim}(B-\bar{B})=0$. Hence, by Lemma 3.1,

$$\text{plim }(B-\bar{B})ZU'/\sqrt{T}=0$$

and

$$\text{plim }\left[(B-\bar{B})\frac{ZZ'}{T}\sqrt{T}(B-\bar{B})'\right]=0$$

(see Appendix C.1). Thus,

$$\text{plim }\sqrt{T}\left[(Y-\bar{B}Z)(Y-\bar{B}Z)'/T-UU'/T\right]=0.$$

Therefore, the proposition follows by noting that $T/(T-c)\to 1$ as $T\to\infty$.■

The proposition covers both estimators $\widehat{\Sigma}_u$ and $\widetilde{\Sigma}_u$. It implies that the feasible estimators $\widetilde{\Sigma}_u$ and $\widehat{\Sigma}_u$ have the same asymptotic properties as the estimator

$$\frac{UU'}{T}=\frac{1}{T}\sum_{t=1}^{T}u_t u_t'$$

which is based on the unknown true residuals and is therefore not feasible in practice. In particular, if $\sqrt{T}\text{ vec}(UU'/T-\Sigma_u)$ converges in distribution, $\sqrt{T}\text{ vec}(\widehat{\Sigma}_u-\Sigma_u)$ and $\sqrt{T}\text{ vec}(\widetilde{\Sigma}_u-\Sigma_u)$ will have the same limiting distribution (see Proposition C.2 of Appendix C.1). Moreover, it can be shown that the asymptotic distributions are independent of the limiting distribution of the LS estimator $\widehat{B}$. Another immediate implication of Proposition 3.2 is that $\widetilde{\Sigma}_u$ and $\widehat{\Sigma}_u$ are consistent estimators of Σ_u. This result is established next.

Corollary 3.2.1
Under the conditions of Proposition 3.2,

$$\text{plim }\widetilde{\Sigma}_u=\text{plim }\widehat{\Sigma}_u=\text{plim }UU'/T=\Sigma_u.$$

■

Proof: By Proposition 3.2, it suffices to show that plim $UU'/T=\Sigma_u$ which follows from Proposition C.12(4) because

$$E\left(\frac{1}{T}UU'\right) = \frac{1}{T}\sum_{t=1}^{T} E(u_t u_t') = \Sigma_u$$

and

$$\mathrm{Var}\left(\frac{1}{T}\mathrm{vec}(UU')\right) = \frac{1}{T^2}\sum_{t=1}^{T} \mathrm{Var}[\mathrm{vec}(u_t u_t')] \leq \frac{T}{T^2} g \underset{T\to\infty}{\longrightarrow} 0,$$

where g is a constant upper bound for $\mathrm{Var}[\mathrm{vec}(u_t u_t')]$. This bound exists because the fourth moments of u_t are bounded by Definition 3.1. ■

If y_t is stable with standard white noise, Proposition 3.1 and Corollary 3.2.1 imply that $(\widehat{\beta}_i - \beta_i)/\widehat{s}_i$ has an asymptotic standard normal distribution. Here β_i $(\widehat{\beta}_i)$ is the i-th component of $\boldsymbol{\beta}$ $(\widehat{\boldsymbol{\beta}})$ and $\widehat{s}_i$ is the square root of the i-th diagonal element of

$$(ZZ')^{-1} \otimes \widehat{\Sigma}_u. \tag{3.2.21}$$

This result means that we can use the "t-ratios" provided by common regression programs in setting up confidence intervals and tests for individual coefficients. The critical values and percentiles may be based on the asymptotic standard normal distribution. Because it was found in simulation studies that the small sample distributions of the "t-ratios" have fatter tails than the standard normal distribution, one may want to approximate the small sample distribution by some t-distribution. The question is then what number of degrees of freedom (d.f.) should be used. The overall model (3.2.3) may suggest a choice of d.f. $= KT - K^2p - K$ because in a standard regression model with nonstochastic regressors the d.f. of the "t-ratios" are equal to the sample size minus the number of estimated parameters. In the present case, it seems also reasonable to use d.f. $= T - Kp - 1$ because the multivariate LS estimator is identical to the LS estimator obtained for each of the K equations in (3.2.2) separately. In a separate regression for each individual equation, we would have T observations and $Kp + 1$ parameters. If the sample size T is large and, thus, the number of degrees of freedom is large, the corresponding t-distribution will be very close to the standard normal so that the choice between the two becomes irrelevant for large samples. Before we look a little further into the problem of choosing appropriate critical values, let us illustrate the foregoing results by an example.

3.2.3 An Example

As an example, we consider a three-dimensional system consisting of first differences of the logarithms of quarterly, seasonally adjusted West German fixed investment (y_1), disposable income (y_2), and consumption expenditures (y_3) from File E1 of the data sets associated with this book. We use only data from 1960–1978 and reserve the data for 1979–1982 for a subsequent analysis. The original data and first differences of logarithms are plotted in Figures 3.1 and 3.2, respectively. The original data have a trend and are thus considered to be nonstationary. The trend is removed by taking first differences of logarithms. We will discuss this issue in some more detail in Part II. Note that the value for 1960.1 is lost in the differenced series.

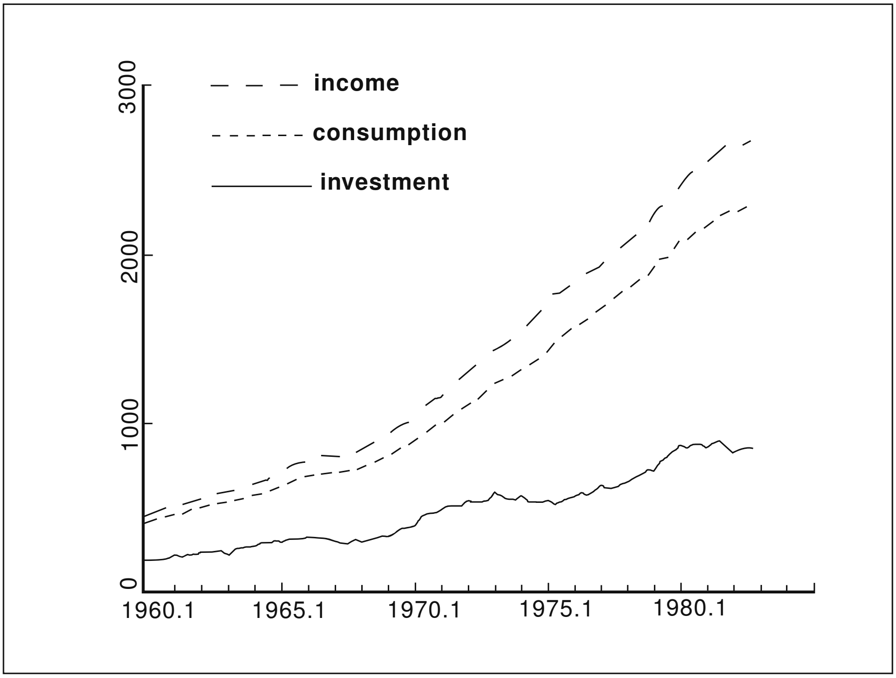

Fig. 3.1. West German investment, income, and consumption data.

Let us assume that the data have been generated by a VAR(2) process. The choice of the VAR order $p = 2$ is arbitrary at this point. In the next chapter, criteria for choosing the VAR order will be considered. Because the VAR order is two, we keep the first two observations of the differenced series as presample values and use a sample size of $T = 73$. Thus, we have a (3×73) matrix Y, $B = (\nu, A_1, A_2)$ is (3×7), Z is (7×73) and $\boldsymbol{\beta}$ and $\mathbf{b}$ are both (21×1) vectors.

The LS estimates are

$$\begin{aligned}\widehat{B} &= (\widehat{\nu}, \widehat{A}_1, \widehat{A}_2) = YZ'(ZZ')^{-1} \\ &= \begin{bmatrix} -.017 & -.320 & .146 & .961 & -.161 & .115 & .934 \\ .016 & .044 & -.153 & .289 & .050 & .019 & -.010 \\ .013 & -.002 & .225 & -.264 & .034 & .355 & -.022 \end{bmatrix}. \quad (3.2.22)\end{aligned}$$

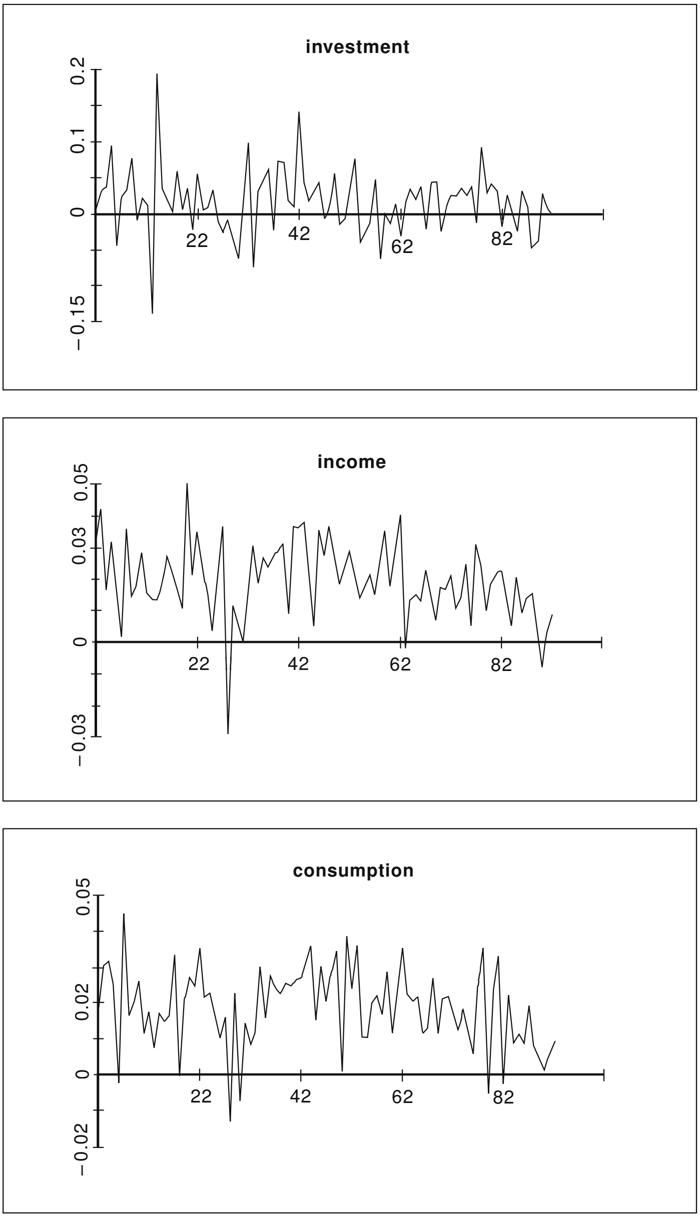

Fig. 3.2. First differences of logarithms of West German investment, income, and consumption.

To check the stability of the estimated process, we determine the roots of the polynomial $\det(I_3 - \widehat{A}_1 z - \widehat{A}_2 z^2)$ which is easily seen to have degree 6. Its roots are

$$z_1 = 1.753,\ z_2 = -2.694,\ z_{3/4} = -0.320 \pm 2.008i,\ z_{5/6} = -1.285 \pm 1.280i.$$

Note that these roots have been computed using higher precision than the three digits in (3.2.22). They all have modulus greater than 1 and, hence, the stability condition is satisfied.

We get

$$\begin{aligned}\widehat{\Sigma}_u &= \frac{1}{T-Kp-1}(YY' - YZ'(ZZ')^{-1}ZY') \\ &= \begin{bmatrix} 21.30 & .72 & 1.23 \\ .72 & 1.37 & .61 \\ 1.23 & .61 & .89 \end{bmatrix} \times 10^{-4}\end{aligned} \tag{3.2.23}$$

as estimate of the residual covariance matrix Σ_u. Furthermore,

$$\begin{aligned}&\widehat{\Gamma}^{-1} = (ZZ'/T)^{-1} \\ &= T\begin{bmatrix} .14 & .17 & -.69 & -2.51 & .10 & -.67 & -2.57 \\ \bullet & 7.39 & 1.24 & -10.56 & 1.80 & 1.08 & -8.70 \\ \bullet & \bullet & 139.81 & -87.40 & -4.58 & 30.21 & -50.88 \\ \bullet & \bullet & \bullet & 207.22 & .84 & -55.35 & 73.82 \\ \bullet & \bullet & \bullet & \bullet & 7.33 & -.03 & -9.31 \\ \bullet & \bullet & \bullet & \bullet & \bullet & 134.19 & -82.64 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & 207.71 \end{bmatrix}.\end{aligned}$$

Dividing the elements of $\widehat{B}$ by square roots of the corresponding diagonal elements of $(ZZ')^{-1} \otimes \widehat{\Sigma}_u$ we get the matrix of t-ratios:

$$\begin{bmatrix} -0.97 & -2.55 & 0.27 & 1.45 & -1.29 & 0.21 & 1.41 \\ 3.60 & 1.38 & -1.10 & 1.71 & 1.58 & 0.14 & -0.06 \\ 3.67 & -0.09 & 2.01 & -1.94 & 1.33 & 3.24 & -0.16 \end{bmatrix}. \tag{3.2.24}$$

We may compare these quantities with critical values from a t-distribution with d.f. $= KT - K^2p - K = 198$ or d.f. $= T - Kp - 1 = 66$. In both cases, we get critical values of approximately ± 2 for a two-tailed test with significance level 5%. Thus, the critical values are approximately the same as those from a standard normal distribution.

Apparently quite a few coefficients are not significant under this criterion. This observation suggests that the model contains unnecessarily many free parameters. In subsequent chapters, we will discuss the problem of choosing the VAR order and possible restrictions for the coefficients. Also, before an estimated model is used for forecasting and analysis purposes, the assumptions underlying the analysis should be checked carefully. Checking the model adequacy will be treated in greater detail in Chapter 4.

3.2.4 Small Sample Properties of the LS Estimator

As mentioned earlier, it is difficult to derive small sample properties of the LS estimator analytically. In such a case it is sometimes helpful to use *Monte*

Carlo methods to get some idea about the small sample properties. In a Monte Carlo analysis, specific processes are used to artificially generate a large number of time series. Then a set of estimates is computed for each multiple time series generated and the properties of the resulting empirical distributions of these estimates are studied (see Appendix D). Such an approach usually permits rather limited conclusions only because the findings may depend on the particular processes used for generating the time series. Nevertheless, such exercises give some insight into the small sample properties of estimators.

In the following, we use the bivariate VAR(2) example process (2.1.15),

$$y_t = \begin{bmatrix} .02 \\ .03 \end{bmatrix} + \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0 & 0 \\ .25 & 0 \end{bmatrix} y_{t-2} + u_t \tag{3.2.25}$$

with error covariance matrix

$$\Sigma_u = \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix} \times 10^{-4} \tag{3.2.26}$$

to investigate the small sample properties of the multivariate LS estimator. With this process we have generated 1000 bivariate time series of length $T = 30$ plus 2 presample values using independent standard normal errors, that is, $u_t \sim \mathcal{N}(0, \Sigma_u)$. Thus the 1000 bivariate time series are generated by a stable Gaussian process so that Propositions 3.1 and 3.2 provide the asymptotic properties of the LS estimators.

In Table 3.1, some empirical results are given. In particular, the empirical mean, variance, and mean squared error (MSE) of each parameter estimator are given. Obviously, the empirical means differ from the actual values of the coefficients. However, measuring the estimation precision by the empirical variance (average squared deviation from the mean in 1000 samples) or MSE (average squared deviation from the true parameter value), the coefficients are seen to be estimated quite precisely even with a sample size as small as $T = 30$. This is partly a consequence of the special properties of the process.

In Table 3.1, empirical percentiles of the t-ratios are also given together with the corresponding percentiles from the t- and standard normal distributions (d.f. $= \infty$). Even with the presently considered relatively small sample size the percentiles of the three distributions that might be used for inference do not differ much. Consequently, it does not matter much which of the theoretical percentiles are used, in particular, because the empirical percentiles, in many cases, differ quite a bit from the corresponding theoretical quantities. This example shows that the asymptotic results have to be used cautiously in setting up small sample tests and confidence intervals. On the other hand, this example also demonstrates that the asymptotic theory does provide some guidance for inference. For example, the empirical 95th percentiles of all coefficients lie between the 90th and the 99th percentile of the standard normal distribution given in the last row of the table. Of course, this is just one example and not a general finding.

Table 3.1. Empirical percentiles of t-ratios of parameter estimates for the example process and actual percentiles of t-distributions for sample size $T = 30$

parameter	empirical mean	empirical variance	MSE	1.	5.	10.	50.	90.	95.	99.
				empirical percentiles of t-ratios						
$\nu_1 = .02$	.041	.0011	.0015	−1.91	−1.04	−0.64	0.62	1.92	2.29	3.12
$\nu_2 = .03$	.038	.0005	.0006	−2.30	−1.40	−1.02	0.25	1.65	2.11	2.83
$\alpha_{11,1} = .5$	.41	.041	.049	−2.78	−2.18	−1.74	−0.43	0.92	1.28	2.01
$\alpha_{21,1} = .4$	.40	.018	.018	−2.61	−1.74	−1.28	0.04	1.28	1.71	2.65
$\alpha_{12,1} = .1$	.10	.078	.078	−2.27	−1.67	−1.35	−0.03	1.29	1.67	2.38
$\alpha_{22,1} = .5$	.44	.030	.034	−2.69	−1.97	−1.59	−0.35	0.89	1.30	2.06
$\alpha_{11,2} = 0$	−.05	.056	.058	−2.75	−1.93	−1.50	−0.24	1.02	1.38	2.09
$\alpha_{21,2} = .25$	.29	.023	.024	−1.99	−1.32	−0.99	0.20	1.45	1.81	2.48
$\alpha_{12,2} = 0$	−.07	.053	.058	−2.48	−1.91	−1.61	−0.28	0.97	1.39	2.03
$\alpha_{22,2} = 0$	−.01	.023	.024	−2.71	−1.72	−1.36	−0.03	1.18	1.53	2.18
	degrees of freedom(d.f.)			percentiles of t-distributions: 1.	5.	10.	50.	90.	95.	99.
	$T - Kp - 1 = 25$			−2.49	−1.71	−1.32	0	1.32	1.71	2.49
	$K(T - Kp - 1) = 50$			−2.41	−1.68	−1.30	0	1.30	1.68	2.41
	∞ (normal distribution)			−2.33	−1.65	−1.28	0	1.28	1.65	2.33

In an extensive study, Nankervis & Savin (1988) investigated the small sample distribution of the "t-statistic" for the parameter of a *univariate* AR(1) process. They found that it differs quite substantially from the corresponding t-distribution, especially if the sample size is small ($T < 100$) and the parameter lies close to the instability region. Analytical results on the bias in estimating VAR models were derived by Nicholls & Pope (1988) and Tjøstheim & Paulsen (1983). What should be learned from our Monte Carlo investigation and these remarks is that asymptotic distributions in the present context can only be used as rough guidelines for small sample inference. That, however, is much better than having no guidance at all.

3.3 Least Squares Estimation with Mean-Adjusted Data and Yule-Walker Estimation

3.3.1 Estimation when the Process Mean Is Known

Occasionally a VAR(p) model is given in *mean-adjusted form*,

$$(y_t - \mu) = A_1(y_{t-1} - \mu) + \cdots + A_p(y_{t-p} - \mu) + u_t. \tag{3.3.1}$$

Multivariate LS estimation of this model form is straightforward if the mean vector μ is known. Defining

$$
\begin{aligned}
&Y^0 := (y_1 - \mu, \dots, y_T - \mu) && (K \times T), \\
&A := (A_1, \dots, A_p) && (K \times Kp), \\
&Y_t^0 := \begin{bmatrix} y_t - \mu \\ \vdots \\ y_{t-p+1} - \mu \end{bmatrix} && (Kp \times 1), \\
&X := (Y_0^0, \dots, Y_{T-1}^0) && (Kp \times T), \\
&\mathbf{y}^0 := \operatorname{vec}(Y^0) && (KT \times 1), \\
&\boldsymbol{\alpha} := \operatorname{vec}(A) && (K^2 p \times 1),
\end{aligned}
\tag{3.3.2}
$$

we can write (3.3.1), for $t = 1, \dots, T$, compactly as

$$Y^0 = AX + U \tag{3.3.3}$$

or

$$\mathbf{y}^0 = (X' \otimes I_K)\boldsymbol{\alpha} + \mathbf{u}, \tag{3.3.4}$$

where U and $\mathbf{u}$ are defined as in (3.2.1). The LS estimator is easily seen to be

$$\widehat{\boldsymbol{\alpha}} = ((XX')^{-1}X \otimes I_K)\mathbf{y}^0 \tag{3.3.5}$$

or

$$\widehat{A} = Y^0 X'(XX')^{-1}. \tag{3.3.6}$$

If y_t is stable and u_t is standard white noise, it can be shown that

$$\sqrt{T}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{\boldsymbol{\alpha}}}), \tag{3.3.7}$$

where

$$\Sigma_{\widehat{\boldsymbol{\alpha}}} = \Gamma_Y(0)^{-1} \otimes \Sigma_u \tag{3.3.8}$$

and $\Gamma_Y(0) := E(Y_t^0 Y_t^{0\prime})$.

3.3.2 Estimation of the Process Mean

Usually μ will not be known in advance. In that case, it may be estimated by the vector of sample means,

$$\overline{y} = \frac{1}{T}\sum_{t=1}^{T} y_t. \tag{3.3.9}$$

Using (3.3.1), $\overline{y}$ can be written as

$$
\begin{aligned}
\overline{y} &= \mu + A_1 \left[\overline{y} + \frac{1}{T}(y_0 - y_T) - \mu \right] + \cdots \\
&\quad + A_p \left[\overline{y} + \frac{1}{T}(y_{-p+1} + \cdots + y_0 - y_{T-p+1} - \cdots - y_T) - \mu \right] \\
&\quad + \frac{1}{T} \sum_{t=1}^{T} u_t .
\end{aligned}
$$

Hence,

$$
(I_K - A_1 - \cdots - A_p)(\overline{y} - \mu) = \frac{1}{T} z_T + \frac{1}{T} \sum_t u_t, \tag{3.3.10}
$$

where

$$
z_T = \sum_{i=1}^{p} A_i \left[\sum_{j=0}^{i-1} (y_{0-j} - y_{T-j}) \right].
$$

Evidently,

$$
E(z_T/\sqrt{T}) = \frac{1}{\sqrt{T}} E(z_T) = 0
$$

and

$$
\mathrm{Var}(z_T/\sqrt{T}) = \frac{1}{T} \mathrm{Var}(z_T) \underset{T\to\infty}{\longrightarrow} 0
$$

because y_t is stable. In other words, $z_T/\sqrt{T}$ converges to zero in mean square. It follows that $\sqrt{T}(I_K - A_1 - \cdots - A_p)(\overline{y} - \mu)$ has the same asymptotic distribution as $\sum u_t/\sqrt{T}$ (see Appendix C, Proposition C.2). Hence, noting that, by a central limit theorem (e.g., Fuller (1976) or Proposition C.13),

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} u_t \xrightarrow{d} \mathcal{N}(0, \Sigma_u), \tag{3.3.11}
$$

if u_t is standard white noise, we get the following result:

Proposition 3.3 (*Asymptotic Properties of the Sample Mean*)
If the VAR(p) process y_t given in (3.3.1) is stable and u_t is standard white noise, then

$$
\sqrt{T}(\overline{y} - \mu) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\overline{y}}), \tag{3.3.12}
$$

where

$$
\Sigma_{\overline{y}} = (I_K - A_1 - \cdots - A_p)^{-1} \Sigma_u (I_K - A_1 - \cdots - A_p)'^{-1}.
$$

In particular, $\operatorname{plim} \overline{y} = \mu$. ■

The proposition follows from (3.3.10), (3.3.11), and Proposition C.15 of Appendix C. The limiting distribution in (3.3.11) holds even in small samples for Gaussian white noise u_t.

Because $\mu = (I_K - A_1 - \cdots - A_p)^{-1}\nu$ (see Chapter 2, Section 2.1), an alternative estimator for the process mean is obtained from the LS estimator of the previous section:

$$\widehat{\mu} = (I_k - \widehat{A}_1 - \cdots - \widehat{A}_p)^{-1}\widehat{\nu}. \tag{3.3.13}$$

Using again Proposition C.15 of Appendix C, this estimator is also consistent and has an asymptotic normal distribution,

$$\sqrt{T}(\widehat{\mu} - \mu) \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial \mu}{\partial \boldsymbol{\beta}'}(\Gamma^{-1} \otimes \Sigma_u)\frac{\partial \mu'}{\partial \boldsymbol{\beta}}\right), \tag{3.3.14}$$

provided the conditions of Proposition 3.1 are satisfied. It can be shown that

$$\frac{\partial \mu}{\partial \boldsymbol{\beta}'}(\Gamma^{-1} \otimes \Sigma_u)\frac{\partial \mu'}{\partial \boldsymbol{\beta}} = \Sigma_{\overline{y}} \tag{3.3.15}$$

and, hence, the estimators $\widehat{\mu}$ and $\overline{y}$ for μ are asymptotically equivalent (see Section 3.4). This result suggests that it does not matter asymptotically whether the mean is estimated separately or jointly with the other VAR coefficients. While this holds asymptotically, it will usually matter in small samples which estimator is used. An example will be given shortly.

3.3.3 Estimation with Unknown Process Mean

If the mean vector μ is unknown, it may be replaced by $\overline{y}$ in the vectors and matrices in (3.3.2) giving $\widehat{X}, \widehat{Y}^0$ and so on. The resulting LS estimator,

$$\widehat{\widehat{\boldsymbol{\alpha}}} = ((\widehat{X}\widehat{X}')^{-1}\widehat{X} \otimes I_K)\widehat{\mathbf{y}}^0,$$

is asymptotically equivalent to $\widehat{\boldsymbol{\alpha}}$. More precisely, it can be shown that, under the conditions of Proposition 3.3,

$$\sqrt{T}(\widehat{\widehat{\boldsymbol{\alpha}}} - \boldsymbol{\alpha}) \xrightarrow{d} \mathcal{N}(0, \Gamma_Y(0)^{-1} \otimes \Sigma_u), \tag{3.3.16}$$

where $\Gamma_Y(0) := E(Y_t^0 Y_t^{0\prime})$. This result will be discussed further in the next section on maximum likelihood estimation for Gaussian processes.

3.3.4 The Yule-Walker Estimator

The LS estimator can also be derived from the Yule-Walker equations given in Chapter 2, (2.1.37). They imply

$$\Gamma_y(h) = [A_1, \ldots, A_p] \begin{bmatrix} \Gamma_y(h-1) \\ \vdots \\ \Gamma_y(h-p) \end{bmatrix}, \qquad h > 0,$$

or

$$\begin{aligned} [\Gamma_y(1), \ldots, \Gamma_y(p)] &= [A_1, \ldots, A_p] \begin{bmatrix} \Gamma_y(0) & \ldots & \Gamma_y(p-1) \\ \vdots & \ddots & \vdots \\ \Gamma_y(-p+1) & \ldots & \Gamma_y(0) \end{bmatrix} \\ &= A\Gamma_Y(0) \end{aligned} \tag{3.3.17}$$

and, hence,

$$A = [\Gamma_y(1), \ldots, \Gamma_y(p)]\, \Gamma_Y(0)^{-1}.$$

Estimating $\Gamma_Y(0)$ by $\widehat{X}\widehat{X}'/T$ and $[\Gamma_y(1), \ldots, \Gamma_y(p)]$ by $\widehat{Y}^0\widehat{X}'/T$, the resulting estimator is just the LS estimator,

$$\widehat{\widehat{A}} = \widehat{Y}^0\widehat{X}'(\widehat{X}\widehat{X}')^{-1}. \tag{3.3.18}$$

Alternatively, the moment matrices $\Gamma_y(h)$ may be estimated using as many data as are available, including the presample values. Thus, if a sample $y_1, \ldots, y_T$ and p presample observations $y_{-p+1}, \ldots, y_0$ are available, μ may be estimated as

$$\overline{y}^* = \frac{1}{T+p} \sum_{t=-p+1}^{T} y_t$$

and $\Gamma_y(h)$ may be estimated as

$$\widehat{\Gamma}_y(h) = \frac{1}{T+p-h} \sum_{t=-p+h+1}^{T} (y_t - \overline{y}^*)(y_{t-h} - \overline{y}^*)'. \tag{3.3.19}$$

Using these estimators in (3.3.17), the so-called *Yule-Walker estimator* for A is obtained. For stable processes, this estimator has the same asymptotic properties as the LS estimator. However, it may have less attractive small sample properties (e.g., Tjøstheim & Paulsen (1983)).

The Yule-Walker estimator always produces estimates in the stability region (see Brockwell & Davis (1987, §8.1) for a discussion of the univariate case). In other words, the estimated process is always stable. This property is sometimes regarded as an advantage of the Yule-Walker estimator. It is responsible for possibly considerable bias of the estimator, however. Also, in practice, it may not be known a priori whether the data generation process of a given multiple time series is stable. In the unstable case, LS and Yule-Walker estimation are not asymptotically equivalent anymore (see also the discussion in Reinsel (1993, Section 4.4)). Therefore, enforcing stability may not be a good strategy in practice. The LS estimator is usually used in the following.

3.3.5 An Example

To illustrate the results of this section, we use again the West German investment, income, and consumption data. The variables y_1, y_2, and y_3 are defined as in Section 3.2.3, the sample period ranges from 1960.4 to 1978.4, that is, $T = 73$ and the data for 1960.2 and 1960.3 are used as presample values. Using only the sample values we get

$$\overline{y} = \begin{bmatrix} .018 \\ .020 \\ .020 \end{bmatrix} \tag{3.3.20}$$

which is different, though not substantially so, from

$$\widehat{\mu} = (I_3 - \widehat{A}_1 - \widehat{A}_2)^{-1}\widehat{\nu} = \begin{bmatrix} .017 \\ .020 \\ .020 \end{bmatrix} \tag{3.3.21}$$

as obtained from the LS estimates in (3.2.22).

Subtracting the sample means from the data we get, based on (3.3.18),

$$\widehat{\widehat{A}} = (\widehat{\widehat{A}}_1, \widehat{\widehat{A}}_2) = \begin{bmatrix} -.319 & .143 & .960 & -.160 & .112 & .933 \\ .044 & -.153 & .288 & .050 & .019 & -.010 \\ -.002 & .224 & -.264 & .034 & .354 & -.023 \end{bmatrix}. \tag{3.3.22}$$

This estimate is clearly distinct from the corresponding part of (3.2.22), although the two estimates do not differ dramatically.

If the two presample values are used in estimating the process means and moment matrices we get

$$\widehat{A}_{YW} = \begin{bmatrix} -.319 & .147 & .959 & -.160 & .115 & .932 \\ .044 & -.152 & .286 & .050 & .020 & -.012 \\ -.002 & .225 & -.264 & .034 & .355 & -.022 \end{bmatrix} \tag{3.3.23}$$

which is the Yule-Walker estimate. Although the sample size is moderate, there is a slight difference between the estimates in (3.3.22) and (3.3.23).

3.4 Maximum Likelihood Estimation

3.4.1 The Likelihood Function

Assuming that the distribution of the process is known, maximum likelihood (ML) estimation is an alternative to LS estimation. We will consider ML estimation under the assumption that the VAR(p) process y_t is Gaussian. More precisely,

$$\mathbf{u} = \mathrm{vec}(U) = \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix} \sim \mathcal{N}(0, I_T \otimes \Sigma_u). \tag{3.4.1}$$

In other words, the probability density of $\mathbf{u}$ is

$$f_{\mathbf{u}}(\mathbf{u}) = \frac{1}{(2\pi)^{KT/2}} |I_T \otimes \Sigma_u|^{-1/2} \exp\left[-\frac{1}{2}\mathbf{u}'(I_T \otimes \Sigma_u^{-1})\mathbf{u}\right]. \tag{3.4.2}$$

Moreover,

$$\mathbf{u} = \begin{bmatrix} I_K & 0 & \dots & 0 & \dots & \dots & 0 \\ -A_1 & I_K & & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & & & \vdots \\ -A_p & -A_{p-1} & \dots & I_K & & & 0 \\ 0 & -A_p & & & \ddots & & \vdots \\ \vdots & & \ddots & & & \ddots & \vdots \\ 0 & 0 & \dots & -A_p & \dots & \dots & I_K \end{bmatrix} (\mathbf{y} - \boldsymbol{\mu}^*)$$

$$+ \begin{bmatrix} -A_1 & -A_2 & \dots & -A_p \\ -A_2 & -A_3 & \dots & 0 \\ \vdots & & & \vdots \\ -A_p & 0 & \dots & 0 \\ & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} (Y_0 - \boldsymbol{\mu}), \tag{3.4.3}$$

where $\mathbf{y} := \mathrm{vec}(Y)$ and $\boldsymbol{\mu}^* := (\mu', \dots, \mu')'$ are $(TK \times 1)$ vectors and $Y_0 := (y_0', \dots, y_{-p+1}')'$ and $\boldsymbol{\mu} := (\mu', \dots, \mu')'$ are $(Kp \times 1)$. Consequently, $\partial\mathbf{u}/\partial\mathbf{y}'$ is a lower triangular matrix with unit diagonal which has unit determinant. Hence, using that $\mathbf{u} = \mathbf{y} - \boldsymbol{\mu}^* - (X' \otimes I_K)\boldsymbol{\alpha}$,

$$\begin{aligned} f_{\mathbf{y}}(\mathbf{y}) &= \left|\frac{\partial\mathbf{u}}{\partial\mathbf{y}'}\right| f_{\mathbf{u}}(\mathbf{u}) \\ &= \frac{1}{(2\pi)^{KT/2}} |I_T \otimes \Sigma_u|^{-1/2} \\ &\quad \times \exp\Big[-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}^* - (X' \otimes I_K)\boldsymbol{\alpha})'(I_T \otimes \Sigma_u^{-1}) \\ &\qquad \times (\mathbf{y} - \boldsymbol{\mu}^* - (X' \otimes I_K)\boldsymbol{\alpha})\Big], \end{aligned} \tag{3.4.4}$$

where X and $\boldsymbol{\alpha}$ are as defined in (3.3.2). For simplicity, the initial values Y_0 are assumed to be given fixed numbers. Hence, we get a log-likelihood function

$$\ln l(\mu, \boldsymbol{\alpha}, \Sigma_u)$$

$$
\begin{aligned}
&= -\frac{KT}{2}\ln 2\pi - \frac{T}{2}\ln|\Sigma_u| \\
&\quad -\frac{1}{2}\left[\mathbf{y} - \boldsymbol{\mu}^* - (X'\otimes I_K)\boldsymbol{\alpha}\right]'\left(I_T\otimes\Sigma_u^{-1}\right)\left[\mathbf{y} - \boldsymbol{\mu}^* - (X'\otimes I_K)\boldsymbol{\alpha}\right] \\
&= -\frac{KT}{2}\ln 2\pi - \frac{T}{2}\ln|\Sigma_u| - \frac{1}{2}\sum_{t=1}^{T}\left[(y_t-\mu) - \sum_{i=1}^{p}A_i(y_{t-i}-\mu)\right]' \\
&\quad \times\Sigma_u^{-1}\left[(y_t-\mu) - \sum_{i=1}^{p}A_i(y_{t-i}-\mu)\right] \\
&= -\frac{KT}{2}\ln 2\pi - \frac{T}{2}\ln|\Sigma_u| \\
&\quad -\frac{1}{2}\sum_t\left(y_t - \sum_i A_i y_{t-i}\right)'\Sigma_u^{-1}\left(y_t - \sum_i A_i y_{t-i}\right) \\
&\quad + \mu'\left(I_K - \sum_i A_i\right)'\Sigma_u^{-1}\sum_t\left(y_t - \sum_i A_i y_{t-i}\right) \\
&\quad -\frac{T}{2}\mu'\left(I_K - \sum_i A_i\right)'\Sigma_u^{-1}\left(I_K - \sum_i A_i\right)\mu \\
&= -\frac{KT}{2}\ln 2\pi - \frac{T}{2}\ln|\Sigma_u| - \frac{1}{2}\mathrm{tr}[(Y^0 - AX)'\Sigma_u^{-1}(Y^0 - AX)], \qquad (3.4.5)
\end{aligned}
$$

where $Y^0 := (y_1 - \mu, \dots, y_T - \mu)$ and $A := (A_1, \dots, A_p)$ are as defined in (3.3.2). These different expressions of the log-likelihood function will be useful in the following.

3.4.2 The ML Estimators

In order to determine the ML estimators of $\mu, \boldsymbol{\alpha}$, and Σ_u, the system of first order partial derivatives is needed:

$$
\begin{aligned}
\frac{\partial \ln l}{\partial \mu} &= \left(I_K - \sum_i A_i\right)'\Sigma_u^{-1}\sum_t\left(y_t - \sum_i A_i y_{t-i}\right) \\
&\quad -T\left(I_K - \sum_i A_i\right)'\Sigma_u^{-1}\left(I_K - \sum_i A_i\right)\mu \\
&= [I_K - A(\mathbf{j}\otimes I_K)]'\,\Sigma_u^{-1}\left[\sum_t (y_t - \mu - AY_{t-1}^0)\right], \qquad (3.4.6)
\end{aligned}
$$

where Y_t^0 is as defined in (3.3.2) and $\mathbf{j} := (1, \dots, 1)'$ is a $(p \times 1)$ vector of ones,

$$\begin{aligned}\frac{\partial \ln l}{\partial \boldsymbol{\alpha}} &= (X \otimes I_K)(I_T \otimes \Sigma_u^{-1})\left[\mathbf{y} - \boldsymbol{\mu}^* - (X' \otimes I_K)\boldsymbol{\alpha}\right] \\ &= (X \otimes \Sigma_u^{-1})(\mathbf{y} - \boldsymbol{\mu}^*) - (XX' \otimes \Sigma_u^{-1})\boldsymbol{\alpha}, \end{aligned} \tag{3.4.7}$$

$$\frac{\partial \ln l}{\partial \Sigma_u} = -\frac{T}{2}\Sigma_u^{-1} + \frac{1}{2}\Sigma_u^{-1}(Y^0 - AX)(Y^0 - AX)'\Sigma_u^{-1}. \tag{3.4.8}$$

Equating to zero gives the system of normal equations which can be solved for the estimators:

$$\widetilde{\mu} = \frac{1}{T}\left(I_K - \sum_i \widetilde{A}_i\right)^{-1} \sum_t \left(y_t - \sum_i \widetilde{A}_i y_{t-i}\right), \tag{3.4.9}$$

$$\widetilde{\boldsymbol{\alpha}} = ((\widetilde{X}\widetilde{X}')^{-1}\widetilde{X} \otimes I_K)(\mathbf{y} - \widetilde{\boldsymbol{\mu}}^*), \tag{3.4.10}$$

$$\widetilde{\Sigma}_u = \frac{1}{T}(\widetilde{Y}^0 - \widetilde{A}\widetilde{X})(\widetilde{Y}^0 - \widetilde{A}\widetilde{X})', \tag{3.4.11}$$

where $\widetilde{X}$ and $\widetilde{Y}^0$ are obtained from X and Y^0, respectively, by replacing μ with $\widetilde{\mu}$.

3.4.3 Properties of the ML Estimators

Comparing these results with the LS estimators obtained in Section 3.3, it turns out that the ML estimators of μ and $\boldsymbol{\alpha}$ are identical to the LS estimators. Thus, $\widetilde{\mu}$ and $\widetilde{\boldsymbol{\alpha}}$ are consistent estimators if y_t is a stationary, stable Gaussian VAR(p) process and $\sqrt{T}(\widetilde{\mu} - \mu)$ and $\sqrt{T}(\widetilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ are asymptotically normally distributed. This result also follows from a more general maximum likelihood theory (see Appendix C.6). In fact, that theory implies that the covariance matrix of the asymptotic distribution of the ML estimators is the limit of T times the inverse information matrix. The information matrix is

$$\mathcal{I}(\boldsymbol{\delta}) = -E\left[\frac{\partial^2 \ln l}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'}\right] \tag{3.4.12}$$

where $\boldsymbol{\delta}' := (\mu', \boldsymbol{\alpha}', \boldsymbol{\sigma}')$ with $\boldsymbol{\sigma} := \mathrm{vech}(\Sigma_u)$. Note that vech is a column stacking operator that stacks only the elements on and below the main diagonal of Σ_u. It is related to the vec operator by the $(\frac{1}{2}K(K+1) \times K^2)$ elimination matrix $\mathbf{L}_K$, that is, $\mathrm{vech}(\Sigma_u) = \mathbf{L}_K \mathrm{vec}(\Sigma_u)$ or, defining $\boldsymbol{\omega} := \mathrm{vec}(\Sigma_u)$, $\boldsymbol{\sigma} = \mathbf{L}_K \boldsymbol{\omega}$ (see Appendix A.12). For instance, for $K = 3$,

$$\begin{aligned}\boldsymbol{\omega} &= \mathrm{vec}(\Sigma_u) = \mathrm{vec}\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} \\ &= (\sigma_{11}, \sigma_{12}, \sigma_{13}, \sigma_{12}, \sigma_{22}, \sigma_{23}, \sigma_{13}, \sigma_{23}, \sigma_{33})' \end{aligned}$$

and

$$\boldsymbol{\sigma} = \operatorname{vech}(\Sigma_u) = \mathbf{L}_3\, \boldsymbol{\omega} = \begin{bmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{13} \\ \sigma_{22} \\ \sigma_{23} \\ \sigma_{33} \end{bmatrix}. \tag{3.4.13}$$

Note that in $\boldsymbol{\delta}$ we collect only the potentially different elements of Σ_u.

The asymptotic covariance matrix of the ML estimator $\widetilde{\boldsymbol{\delta}}$ is known to be

$$\lim_{T\to\infty} [\mathcal{I}(\boldsymbol{\delta})/T]^{-1}. \tag{3.4.14}$$

In order to determine this matrix, we need the second order partial derivatives of the log-likelihood. From (3.4.6) to (3.4.8) we get

$$\frac{\partial^2 \ln l}{\partial \mu\, \partial \mu'} = -T \left(I_K - \sum_i A_i \right)' \Sigma_u^{-1} \left(I_K - \sum_i A_i \right), \tag{3.4.15}$$

$$\frac{\partial^2 \ln l}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}'} = -(XX' \otimes \Sigma_u^{-1}), \tag{3.4.16}$$

$$\begin{aligned} \frac{\partial^2 \ln l}{\partial \boldsymbol{\omega}\, \partial \boldsymbol{\omega}'} &= \frac{T}{2}(\Sigma_u^{-1} \otimes \Sigma_u^{-1}) - \frac{1}{2}(\Sigma_u^{-1} \otimes \Sigma_u^{-1}\, UU'\Sigma_u^{-1}) \\ &\quad -\frac{1}{2}(\Sigma_u^{-1} UU' \Sigma_u^{-1} \otimes \Sigma_u^{-1}), \end{aligned} \tag{3.4.17}$$

where $\boldsymbol{\omega} = \operatorname{vec}(\Sigma_u)$ (see Problem 3.3),

$$\begin{aligned} \frac{\partial^2 \ln l}{\partial \mu\, \partial \boldsymbol{\alpha}'} &= -\left[I_K - (\mathbf{j}' \otimes I_K) A' \right] \Sigma_u^{-1} \sum_t Y_{t-1}^{0\prime} \otimes I_K \\ &\quad - \left(\sum_t u_t' \Sigma_u^{-1} \otimes I_K \right) (I_K \otimes \mathbf{j}' \otimes I_K) \frac{\partial \operatorname{vec}(A')}{\partial \boldsymbol{\alpha}'} \end{aligned} \tag{3.4.18}$$

(see Problem 3.4),

$$\frac{\partial^2 \ln l}{\partial \boldsymbol{\omega}\, \partial \mu'} = \frac{1}{2}(\Sigma_u^{-1} \otimes \Sigma_u^{-1}) \left[(I_K \otimes U) \frac{\partial \operatorname{vec}(U')}{\partial \mu'} + (U \otimes I_K) \frac{\partial \operatorname{vec}(U)}{\partial \mu'} \right] \tag{3.4.19}$$

(see Problem 3.5), and

$$\frac{\partial^2 \ln l}{\partial \boldsymbol{\omega}\, \partial \boldsymbol{\alpha}'} = -\frac{1}{2}(\Sigma_u^{-1} \otimes \Sigma_u^{-1}) \left[(I_K \otimes UX') \frac{\partial \operatorname{vec}(A')}{\partial \boldsymbol{\alpha}'} + (UX' \otimes I_K) \right] \tag{3.4.20}$$

(see Problem 3.6).

It is obvious from (3.4.18) that

$$\lim\ T^{-1}E\left(\frac{\partial^2 \ln l}{\partial\mu\,\partial\boldsymbol{\alpha}'}\right) = 0 \tag{3.4.21}$$

because $E(\sum_t Y_{t-1}^0/T) \to 0$. Furthermore, from (3.4.19), it follows that

$$E\left(\frac{\partial^2 \ln l}{\partial\boldsymbol{\omega}\,\partial\mu'}\right) = 0 \tag{3.4.22}$$

because $E(U) = 0$ and $\partial\,\mathrm{vec}(U')/\partial\mu'$ is a matrix of constants. Moreover, from (3.4.20), we have

$$\lim\ T^{-1}E\left(\frac{\partial^2 \ln l}{\partial\boldsymbol{\omega}\,\partial\boldsymbol{\alpha}'}\right) = 0 \tag{3.4.23}$$

because $E(UX'/T) \to 0$. Thus, $\lim \mathcal{I}(\boldsymbol{\delta})/T$ is block diagonal and we get the asymptotic distributions of $\mu, \boldsymbol{\alpha}$, and $\boldsymbol{\sigma}$ as follows.

Multiplying minus the inverse of (3.4.15) by T gives the asymptotic covariance matrix of the ML estimator for the mean vector μ, that is,

$$\sqrt{T}(\widetilde{\mu} - \mu) \xrightarrow{d} \mathcal{N}\left(0, \left(I_K - \sum_{i=1}^{p} A_i\right)^{-1} \Sigma_u \left(I_K - \sum_{i=1}^{p} A_i'\right)^{-1}\right). \tag{3.4.24}$$

Hence, $\widetilde{\mu}$ has the same asymptotic distribution as $\overline{y}$ (see Proposition 3.3). In other words, the two estimators for μ are asymptotically equivalent and, under the present conditions, this fact implies that $\overline{y}$ is asymptotically efficient because the ML estimator is asymptotically efficient. The asymptotic equivalence of $\widetilde{\mu}$ and $\overline{y}$ can also be seen from (3.4.9) (see the argument prior to Proposition 3.3 and Problem 3.7).

Taking the limit of T^{-1} times the expectation of minus (3.4.16) gives $\Gamma_Y(0) \otimes \Sigma_u^{-1}$. Note that $E(XX'/T)$ is not strictly equal to $\Gamma_Y(0)$ because we have assumed fixed initial values $y_{-p+1}, \dots, y_0$. However, asymptotically, as T goes to infinity, the impact of the initial values vanishes. Thus, we get

$$\sqrt{T}(\widetilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{d} \mathcal{N}(0, \Gamma_Y(0)^{-1} \otimes \Sigma_u). \tag{3.4.25}$$

Of course, this result also follows from the equivalence of the ML and LS estimators.

Noting that $E(UU') = T\Sigma_u$, it follows from (3.4.17) that

$$E\left(\frac{\partial^2 \ln l}{\partial\boldsymbol{\omega}\,\partial\boldsymbol{\omega}'}\right) = -\frac{T}{2}(\Sigma_u^{-1} \otimes \Sigma_u^{-1}). \tag{3.4.26}$$

Denoting by $\mathbf{D}_K$ the $(K^2 \times \frac{1}{2}K(K+1))$ duplication matrix (see Appendix A.12) so that $\boldsymbol{\omega} = \mathbf{D}_K\boldsymbol{\sigma}$, we get

$$\frac{\partial^2 \ln l}{\partial \boldsymbol{\sigma}\, \partial \boldsymbol{\sigma}'} = \frac{\partial \boldsymbol{\omega}'}{\partial \boldsymbol{\sigma}} \frac{\partial^2 \ln l}{\partial \boldsymbol{\omega}\, \partial \boldsymbol{\omega}'} \frac{\partial \boldsymbol{\omega}}{\partial \boldsymbol{\sigma}'} = \mathbf{D}_K' \frac{\partial^2 \ln l}{\partial \boldsymbol{\omega}\, \partial \boldsymbol{\omega}'} \mathbf{D}_K$$

and, hence,

$$\sqrt{T}(\widetilde{\boldsymbol{\sigma}} - \boldsymbol{\sigma}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widetilde{\boldsymbol{\sigma}}}) \tag{3.4.27}$$

with

$$\begin{aligned} \Sigma_{\widetilde{\boldsymbol{\sigma}}} &= -TE\left(\frac{\partial^2 \ln\, l}{\partial \boldsymbol{\sigma}\, \partial \boldsymbol{\sigma}'}\right)^{-1} = 2\left[\mathbf{D}_K'(\Sigma_u^{-1} \otimes \Sigma_u^{-1})\mathbf{D}_K\right]^{-1} \\ &= 2\mathbf{D}_K^+(\Sigma_u \otimes \Sigma_u)\mathbf{D}_K^{+\prime}, \end{aligned} \tag{3.4.28}$$

where $\mathbf{D}_K^+ = (\mathbf{D}_K'\mathbf{D}_K)^{-1}\mathbf{D}_K'$ is the Moore-Penrose generalized inverse of the duplication matrix $\mathbf{D}_K$ and Rule (17) from Appendix A.12 has been used. In summary, we get the following proposition.

Proposition 3.4 (*Asymptotic Properties of ML Estimators*)
Let y_t be a stationary, stable Gaussian VAR(p) process as in (3.3.1). Then the ML estimators $\widetilde{\mu}$, $\widetilde{\boldsymbol{\alpha}}$, and $\widetilde{\boldsymbol{\sigma}} = \text{vech}(\widetilde{\Sigma}_u)$ given in (3.4.9)–(3.4.11) are consistent and

$$\sqrt{T}\begin{bmatrix} \widetilde{\mu} - \mu \\ \widetilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \\ \widetilde{\boldsymbol{\sigma}} - \boldsymbol{\sigma} \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} \Sigma_{\widetilde{\mu}} & 0 & 0 \\ 0 & \Sigma_{\widetilde{\boldsymbol{\alpha}}} & 0 \\ 0 & 0 & \Sigma_{\widetilde{\boldsymbol{\sigma}}} \end{bmatrix}\right), \tag{3.4.29}$$

so that $\widetilde{\mu}$ is asymptotically independent of $\widetilde{\boldsymbol{\alpha}}$ and $\widetilde{\Sigma}_u$ and $\widetilde{\boldsymbol{\alpha}}$ is asymptotically independent of $\widetilde{\mu}$ and $\widetilde{\Sigma}_u$. The covariance matrices are

$$\Sigma_{\widetilde{\mu}} = \left(I_K - \sum_i A_i\right)^{-1} \Sigma_u \left(I_K - \sum_i A_i'\right)^{-1},$$

$$\Sigma_{\widetilde{\boldsymbol{\alpha}}} = \Gamma_Y(0)^{-1} \otimes \Sigma_u,$$

$$\Sigma_{\widetilde{\boldsymbol{\sigma}}} = 2\mathbf{D}_K^+(\Sigma_u \otimes \Sigma_u)\mathbf{D}_K^{+\prime}.$$

They may be estimated consistently by replacing the unknown quantities by their ML estimators and estimating $\Gamma_Y(0)$ by $\widetilde{X}\widetilde{X}'/T$. ■

In this section, we have chosen to consider the mean-adjusted form of a VAR(p) process. Of course, it is possible to perform a similar derivation for the standard form given in (3.1.1). In that case the ML estimators of ν and $\boldsymbol{\alpha}$ are not asymptotically independent though. Their joint asymptotic distribution is identical to that of $\widehat{\boldsymbol{\beta}}$ given in Proposition 3.1. From Proposition 3.2 we know that the asymptotic distribution of $\widetilde{\boldsymbol{\sigma}}$ remains unaltered. In the next section, we will investigate the consequences of forecasting with estimated rather than known processes.

3.5 Forecasting with Estimated Processes

3.5.1 General Assumptions and Results

In Chapter 2, Section 2.2, we have seen that the optimal h-step forecast of the process (3.1.1) is

$$y_t(h) = \nu + A_1 y_t(h-1) + \cdots + A_p y_t(h-p), \tag{3.5.1}$$

where $y_t(j) = y_{t+j}$ for $j \leq 0$. If the true coefficients $B = (\nu, A_1, \ldots, A_p)$ are replaced by estimators $\widehat{B} = (\widehat{\nu}, \widehat{A}_1, \ldots, \widehat{A}_p)$, we get a forecast

$$\widehat{y}_t(h) = \widehat{\nu} + \widehat{A}_1 \widehat{y}_t(h-1) + \cdots + \widehat{A}_p \widehat{y}_t(h-p), \tag{3.5.2}$$

where $\widehat{y}_t(j) = y_{t+j}$ for $j \leq 0$. Thus, the forecast error is

$$\begin{aligned} y_{t+h} - \widehat{y}_t(h) &= [y_{t+h} - y_t(h)] + [y_t(h) - \widehat{y}_t(h)] \\ &= \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} + [y_t(h) - \widehat{y}_t(h)], \end{aligned} \tag{3.5.3}$$

where the Φ_i are the coefficient matrices of the canonical MA representation of y_t (see (2.2.9)). Under quite general conditions for the process y_t, the forecast errors can be shown to have zero mean, $E[y_{t+h} - \widehat{y}_t(h)] = 0$, so that the forecasts are unbiased even if the coefficients are estimated. Because we do not need this result in the following, we refer to Dufour (1985) for the details and a proof. All the u_s in the first term on the right-hand side of the last equality sign in (3.5.3) are attached to periods $s > t$, whereas all the y_s in the second term correspond to periods $s \leq t$, if estimation is done with observations from periods up to time t only. Therefore, the two terms are uncorrelated. Hence, the MSE matrix of the forecast $\widehat{y}_t(h)$ is of the form

$$\begin{aligned} \Sigma_{\widehat{y}}(h) &:= \mathrm{MSE}[\widehat{y}_t(h)] = E\{[y_{t+h} - \widehat{y}_t(h)][y_{t+h} - \widehat{y}_t(h)]'\} \\ &= \Sigma_y(h) + \mathrm{MSE}[y_t(h) - \widehat{y}_t(h)], \end{aligned} \tag{3.5.4}$$

where

$$\Sigma_y(h) = \sum_{i=0}^{h-1} \Phi_i \Sigma_u \Phi_i'$$

(see (2.2.11)). In order to evaluate the last term in (3.5.4), the distribution of the estimator $\widehat{B}$ is needed. Because we have not been able to derive the small sample distributions of the estimators considered in the previous sections but we have derived the asymptotic distributions instead, we cannot hope for more than an asymptotic approximation to the MSE of $y_t(h) - \widehat{y}_t(h)$. Such an approximation will be derived in the following.

There are two alternative assumptions that can be made in order to facilitate the derivation of the desired result:

(1) Only data up to the forecast origin are used for estimation.
(2) Estimation is done using a realization (time series) of a process that is independent of the process used for prediction and has the same stochastic structure (for instance, it is Gaussian and has the same first and second moments as the process used for prediction).

The first assumption is the more realistic one from a practical point of view because estimation and forecasting are usually based on the same data set. In that case, because the sample size is assumed to go to infinity in deriving asymptotic results, either the forecast origin has to go to infinity too or it has to be assumed that more and more data at the beginning of the sample become available. Because the forecast uses only p vectors y_s prior to the forecast period, these variables will be asymptotically independent of the estimator $\widehat{B}$ (they are asymptotically negligible in comparison with all the other observations going into the estimate). Thus, asymptotically the first assumption implies the same results as the second one. In the following, for simplicity, the second assumption will therefore be used. Furthermore, it will be assumed that for $\boldsymbol{\beta} = \operatorname{vec}(B)$ and $\widehat{\boldsymbol{\beta}} = \operatorname{vec}(\widehat{B})$ we have

$$\sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{\boldsymbol{\beta}}}). \tag{3.5.5}$$

Samaranayake & Hasza (1988) and Basu & Sen Roy (1986) give a formal proof of the result that the MSE approximation obtained in the following remains valid under assumption (1) above.

With the foregoing assumptions it follows that, conditional on a particular realization $Y_t = (y_t', \dots, y_{t-p+1}')'$ of the process used for prediction,

$$\sqrt{T}\,[\widehat{y}_t(h) - y_t(h)|Y_t] \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial y_t(h)}{\partial \boldsymbol{\beta}'} \Sigma_{\widehat{\boldsymbol{\beta}}} \frac{\partial y_t(h)'}{\partial \boldsymbol{\beta}}\right) \tag{3.5.6}$$

because $y_t(h)$ is a differentiable function of $\boldsymbol{\beta}$ (see Appendix C, Proposition C.15(3)). Here T is the sample size (time series length) used for estimation. This result suggests the approximation of $\mathrm{MSE}\,[\widehat{y}_t(h) - y_t(h)]$ by $\Omega(h)/T$, where

$$\Omega(h) = E\left[\frac{\partial y_t(h)}{\partial \boldsymbol{\beta}'} \Sigma_{\widehat{\boldsymbol{\beta}}} \frac{\partial y_t(h)'}{\partial \boldsymbol{\beta}}\right]. \tag{3.5.7}$$

In fact, for a Gaussian process y_t,

$$\sqrt{T}\,[\widehat{y}_t(h) - y_t(h)] \xrightarrow{d} \mathcal{N}(0, \Omega(h)). \tag{3.5.8}$$

Hence, we get an approximation

$$\Sigma_{\widehat{y}}(h) = \Sigma_y(h) + \frac{1}{T}\Omega(h) \tag{3.5.9}$$

for the MSE matrix of $\widehat{y}_t(h)$.

From (3.5.7) it is obvious that $\Omega(h)$ and, thus, the approximate MSE $\Sigma_{\widehat{y}}(h)$ can be reduced by using an estimator that is asymptotically more efficient than $\widehat{\boldsymbol{\beta}}$, if such an estimator exists. In other words, efficient estimation is of importance in order to reduce the forecast uncertainty.

3.5.2 The Approximate MSE Matrix

To derive an explicit expression for $\Omega(h)$, the derivatives $\partial y_t(h)/\partial\boldsymbol{\beta}'$ are needed. They can be obtained easily by noting that

$$y_t(h) = J_1 \mathbf{B}^h Z_t, \tag{3.5.10}$$

where $Z_t := (1, y_t', \dots, y_{t-p+1}')'$,

$$\mathbf{B} := \underset{[(Kp+1)\times(Kp+1)]}{\begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ \nu & A_1 & A_2 & \dots & A_{p-1} & A_p \\ 0 & I_K & 0 & \dots & 0 & 0 \\ 0 & 0 & I_K & & 0 & 0 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & \dots & I_K & 0 \end{bmatrix}} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ & B & & \\ 0 & I_{K(p-1)} & & 0 \end{bmatrix}$$

and

$$J_1 := [\underbrace{0}_{(K\times 1)} : I_K : \underbrace{0 : \cdots : 0}_{(K\times K(p-1))}] \quad [K \times (Kp+1)].$$

The relation (3.5.10) follows by induction (see Problem 3.8). Using (3.5.10), we get

$$\begin{aligned}
\frac{\partial y_t(h)}{\partial \boldsymbol{\beta}'} &= \frac{\partial \operatorname{vec}(J_1 \mathbf{B}^h Z_t)}{\partial \boldsymbol{\beta}'} = (Z_t' \otimes J_1) \frac{\partial \operatorname{vec}(\mathbf{B}^h)}{\partial \boldsymbol{\beta}'} \\
&= (Z_t' \otimes J_1) \left[\sum_{i=0}^{h-1} (\mathbf{B}')^{h-1-i} \otimes \mathbf{B}^i \right] \frac{\partial \operatorname{vec}(\mathbf{B})}{\partial \boldsymbol{\beta}'} \\
&\qquad \text{(Appendix A.13, Rule (8))} \\
&= (Z_t' \otimes J_1) \left[\sum_{i=0}^{h-1} (\mathbf{B}')^{h-1-i} \otimes \mathbf{B}^i \right] (I_{Kp+1} \otimes J_1') \\
&\qquad \text{(see the definition of } \mathbf{B}) \\
&= \sum_{i=0}^{h-1} Z_t' (\mathbf{B}')^{h-1-i} \otimes J_1 \mathbf{B}^i J_1' \\
&= \sum_{i=0}^{h-1} Z_t' (\mathbf{B}')^{h-1-i} \otimes \Phi_i,
\end{aligned} \tag{3.5.11}$$

where $\Phi_i = J_1 \mathbf{B}^i J_1'$ follows as in (2.1.17). Using the LS estimator $\widehat{\boldsymbol{\beta}}$ with asymptotic covariance matrix $\Sigma_{\widehat{\boldsymbol{\beta}}} = \Gamma^{-1} \otimes \Sigma_u$ (see Proposition 3.1), the matrix $\Omega(h)$ is seen to be

$$\begin{aligned}
\Omega(h) &= E\left[\frac{\partial y_t(h)}{\partial \boldsymbol{\beta}'} (\Gamma^{-1} \otimes \Sigma_u) \frac{\partial y_t(h)'}{\partial \boldsymbol{\beta}} \right] \\
&= \sum_{i=0}^{h-1} \sum_{j=0}^{h-1} E(Z_t' (\mathbf{B}')^{h-1-i} \Gamma^{-1} \mathbf{B}^{h-1-j} Z_t) \otimes \Phi_i \Sigma_u \Phi_j' \\
&= \sum_i \sum_j E[\operatorname{tr}(Z_t' (\mathbf{B}')^{h-1-i} \Gamma^{-1} \mathbf{B}^{h-1-j} Z_t)] \Phi_i \Sigma_u \Phi_j' \\
&= \sum_i \sum_j \operatorname{tr}[(\mathbf{B}')^{h-1-i} \Gamma^{-1} \mathbf{B}^{h-1-j} E(Z_t Z_t')] \Phi_i \Sigma_u \Phi_j' \\
&= \sum_{i=0}^{h-1} \sum_{j=0}^{h-1} \operatorname{tr}[(\mathbf{B}')^{h-1-i} \Gamma^{-1} \mathbf{B}^{h-1-j} \Gamma] \Phi_i \Sigma_u \Phi_j', \qquad (3.5.12)
\end{aligned}$$

provided y_t is stable so that

$$\Gamma := \operatorname{plim}(ZZ'/T) = E(Z_t Z_t').$$

Here $Z := (Z_0, \dots, Z_{T-1})$ is the $((Kp+1) \times T)$ matrix defined in (3.2.1).

For example, for $h = 1$,

$$\Omega(1) = (Kp+1)\Sigma_u.$$

Hence, the approximation

$$\Sigma_{\widehat{y}}(1) = \Sigma_u + \frac{Kp+1}{T} \Sigma_u = \frac{T+Kp+1}{T} \Sigma_u \qquad (3.5.13)$$

of the MSE matrix of the 1-step forecast with estimated coefficients is obtained. This expression shows that the contribution of the estimation variability to the forecast MSE matrix $\Sigma_{\widehat{y}}(1)$ depends on the dimension K of the process, the VAR order p, and the sample size T used for estimation. It can be quite substantial if the sample size is small or moderate. For instance, considering a three-dimensional process of order 8 which is estimated from 15 years of quarterly data (i.e., $T = 52$ plus 8 presample values needed for LS estimation), the 1-step forecast MSE matrix Σ_u for known processes is inflated by a factor $(T + Kp + 1)/T = 1.48$. Of course, this approximation is derived from asymptotic theory so that its small sample validity is not guaranteed. We will take a closer look at this problem shortly. Obviously, the inflation factor $(T+Kp+1)/T \to 1$ for $T \to \infty$. Thus the MSE contribution due to sampling variability vanishes if the sample size gets large. This result is a consequence of estimating the VAR coefficients consistently. An expression for $\Omega(h)$ can also be derived on the basis of the mean-adjusted form of the VAR process (see Problem 3.9).

In practice, for $h > 1$, it will not be possible to evaluate $\Omega(h)$ without knowing the AR coefficients summarized in the matrix B. A consistent estimator $\widehat{\Omega}(h)$ may be obtained by replacing all unknown parameters by their LS estimators, that is, $\mathbf{B}$ is replaced by $\widehat{\mathbf{B}}$ which is obtained by using $\widehat{B}$ for B, Σ_u is replaced by $\widehat{\Sigma}_u, \Phi_i$ is estimated by $\widehat{\Phi}_i = J_1\widehat{\mathbf{B}}^i J_1'$, and Γ is estimated by $\widehat{\Gamma} = ZZ'/T$. The resulting estimator of $\Sigma_{\widehat{y}}(h)$ will be denoted by $\widehat{\Sigma}_{\widehat{y}}(h)$ in the following.

The foregoing discussion is of importance in setting up interval forecasts. Assuming that y_t is Gaussian, an approximate $(1-\alpha)100\%$ interval forecast, h periods ahead, for the k-th component $y_{k,t}$ of y_t is

$$\widehat{y}_{k,t}(h) \pm z_{(\alpha/2)}\widehat{\widehat{\sigma}}_k(h) \tag{3.5.14}$$

or

$$\left[\widehat{y}_{k,t}(h) - z_{(\alpha/2)}\widehat{\widehat{\sigma}}_k(h),\ \widehat{y}_{k,t}(h) + z_{(\alpha/2)}\widehat{\widehat{\sigma}}_k(h)\right], \tag{3.5.15}$$

where $z_{(\alpha)}$ is the upper $\alpha 100$-th percentile of the standard normal distribution and $\widehat{\widehat{\sigma}}_k(h)$ is the square root of the k-th diagonal element of $\widehat{\Sigma}_{\widehat{y}}(h)$. Using Bonferroni's inequality, approximate joint confidence regions for a set of forecasts can be obtained just as described in Section 2.2.3 of Chapter 2.

3.5.3 An Example

To illustrate the previous results, we consider again the investment/income/consumption example of Section 3.2.3. Using the VAR(2) model with the coefficient estimates given in (3.2.22) and

$$y_{T-1} = y_{72} = \begin{bmatrix} .02551 \\ .02434 \\ .01319 \end{bmatrix} \quad \text{and} \quad y_T = y_{73} = \begin{bmatrix} .03637 \\ .00517 \\ .00599 \end{bmatrix}$$

results in forecasts

$$\begin{aligned} \widehat{y}_T(1) &= \widehat{\nu} + \widehat{A}_1 y_T + \widehat{A}_2 y_{T-1} = \begin{bmatrix} -.011 \\ .020 \\ .022 \end{bmatrix}, \\ \widehat{y}_T(2) &= \widehat{\nu} + \widehat{A}_1 \widehat{y}_T(1) + \widehat{A}_2 y_T = \begin{bmatrix} .011 \\ .020 \\ .015 \end{bmatrix}, \end{aligned} \tag{3.5.16}$$

and so on.

The estimated forecast MSE matrix for $h = 1$ is

$$\begin{aligned} \widehat{\Sigma}_{\widehat{y}}(1) &= \frac{T+Kp+1}{T}\widehat{\Sigma}_u = \frac{73+6+1}{73}\widehat{\Sigma}_u \\ &= \begin{bmatrix} 23.34 & .785 & 1.351 \\ .785 & 1.505 & .674 \\ 1.351 & .674 & .978 \end{bmatrix} \times 10^{-4}, \end{aligned} \tag{3.5.17}$$

where $\widehat{\Sigma}_u$ from (3.2.23) has been used. We need $\widehat{\Phi}_1$ for evaluating

$$\widehat{\Sigma}_{\widehat{y}}(2) = \widehat{\Sigma}_y(2) + \frac{1}{T}\widehat{\Omega}(2),$$

where

$$\widehat{\Sigma}_y(2) = \widehat{\Sigma}_u + \widehat{\Phi}_1\widehat{\Sigma}_u\widehat{\Phi}_1'$$

and

$$\begin{aligned}\widehat{\Omega}(2) &= \sum_{i=0}^{1}\sum_{j=0}^{1} \operatorname{tr}\left[(\widehat{\mathbf{B}}')^{1-i}(ZZ'/T)^{-1}\widehat{\mathbf{B}}^{1-j}(ZZ'/T)\right]\widehat{\Phi}_i\widehat{\Sigma}_u\widehat{\Phi}_j' \\ &= \operatorname{tr}[\widehat{\mathbf{B}}'(ZZ')^{-1}\widehat{\mathbf{B}}ZZ']\widehat{\Sigma}_u + \operatorname{tr}(\widehat{\mathbf{B}}')\widehat{\Sigma}_u\widehat{\Phi}_1' \\ &\quad + \operatorname{tr}(\widehat{\mathbf{B}})\widehat{\Phi}_1\widehat{\Sigma}_u + \operatorname{tr}(I_{Kp+1})\widehat{\Phi}_1\widehat{\Sigma}_u\widehat{\Phi}_1'.\end{aligned}$$

From (2.1.22) we know that $\Phi_1 = A_1$. Hence, we use $\widehat{\Phi}_1 = \widehat{A}_1$ from (3.2.22). Thus, we get

$$\widehat{\Sigma}_y(2) = \begin{bmatrix} 23.67 & .547 & 1.226 \\ .547 & 1.488 & .554 \\ 1.226 & .554 & .952 \end{bmatrix} \times 10^{-4}$$

and

$$\widehat{\Omega}(2) = \begin{bmatrix} 10.59 & .238 & .538 \\ .238 & .675 & .233 \\ .538 & .233 & .422 \end{bmatrix} \times 10^{-3}.$$

Consequently,

$$\widehat{\Sigma}_{\widehat{y}}(2) = \begin{bmatrix} 25.12 & .580 & 1.300 \\ .580 & 1.581 & .586 \\ 1.300 & .586 & 1.009 \end{bmatrix} \times 10^{-4}. \tag{3.5.18}$$

Assuming that the data are generated by a Gaussian process, we get the following approximate 95% interval forecasts:

$$\begin{aligned} &\widehat{y}_{1,T}(1) \pm 1.96\widehat{\widehat{\sigma}}_1(1) \text{ or } -.011 \pm .095, \\ &\widehat{y}_{2,T}(1) \pm 1.96\widehat{\widehat{\sigma}}_2(1) \text{ or } \quad .020 \pm .024, \\ &\widehat{y}_{3,T}(1) \pm 1.96\widehat{\widehat{\sigma}}_3(1) \text{ or } \quad .022 \pm .019, \\ &\widehat{y}_{1,T}(2) \pm 1.96\widehat{\widehat{\sigma}}_1(2) \text{ or } \quad .011 \pm .098, \\ &\widehat{y}_{2,T}(2) \pm 1.96\widehat{\widehat{\sigma}}_2(2) \text{ or } \quad .020 \pm .025, \\ &\widehat{y}_{3,T}(2) \pm 1.96\widehat{\widehat{\sigma}}_3(2) \text{ or } \quad .015 \pm .020. \end{aligned} \tag{3.5.19}$$

In Figure 3.3, some more forecasts of the three variables with two-standard error bounds to each side are depicted. The intervals indicated by the dashed bounds may be interpreted as approximate 95% forecast intervals for the individual forecasts. If the region enclosed by the dashed lines is viewed as a joint confidence region for all 4 forecasts, a lower bound for the (approximate) probability content is $(100-4\times5)\% = 80\%$. In the figure it can be seen that for investment and income the actually observed values for 1979 ($t = 77, \dots, 80$) are well inside the forecast regions, whereas two of the four consumption values are outside that region.

3.5.4 A Small Sample Investigation

It is not obvious that the MSE and interval forecast approximations derived in the foregoing are reasonable in small samples because the MSE modification has been based on asymptotic theory. To investigate the small sample behavior of the predictor with estimated coefficients, we have used again 1000 realizations of the bivariate VAR(2) process (3.2.25)/(3.2.26) of Section 3.2.4 and we have computed forecast intervals for the period following the last sample period. In Table 3.2, the proportions of actual values falling in these intervals are reported for sample sizes of $T = 30$ and 100.

Table 3.2. Accuracy of forecast intervals in small samples based on 1000 bivariate time series

MSE used in interval construction	% forecast interval	percent of actual values falling in the forecast interval: $T = 30$, y_1	$T = 30$, y_2	$T = 100$, y_1	$T = 100$, y_2
	90	86.5	85.7	89.7	89.4
$\Sigma_y(1)$	95	92.6	91.8	94.5	94.0
	99	98.1	98.0	99.0	98.5
	90	89.3	88.2	90.4	90.0
$\Sigma_{\hat{y}}(1)$	95	94.4	94.1	95.3	94.6
	99	99.0	98.4	99.3	98.8
	90	85.2	84.2	89.6	88.5
$\widehat{\Sigma}_y(1)$	95	90.5	90.4	94.7	93.9
	99	98.4	96.5	98.9	98.3
	90	88.1	86.9	90.3	89.1
$\widehat{\Sigma}_{\hat{y}}(1)$	95	93.4	92.7	95.2	94.0
	99	99.4	97.8	99.1	98.5

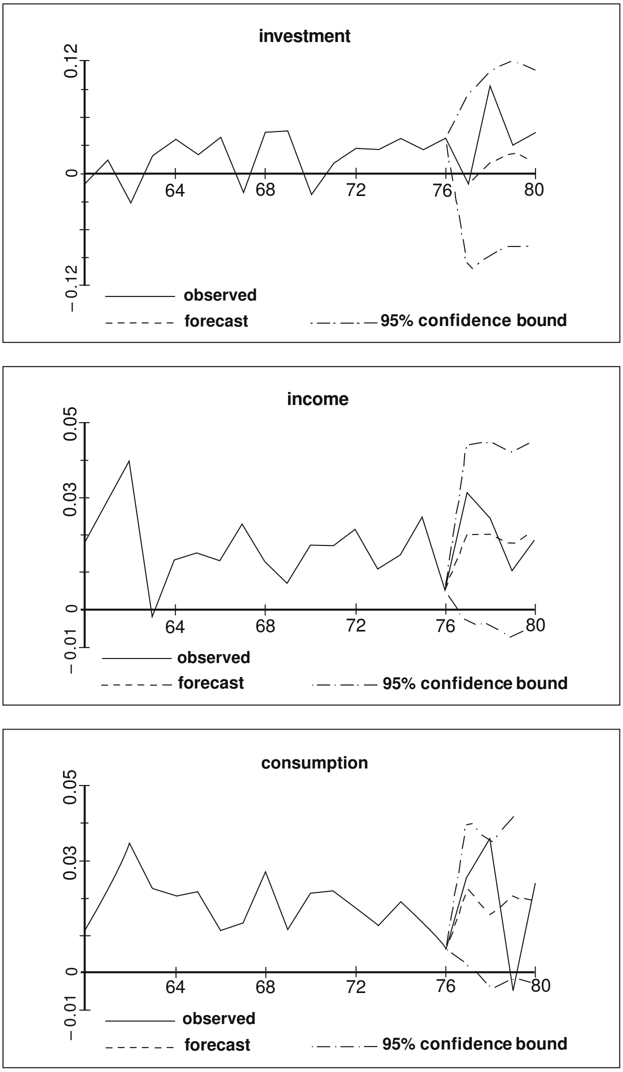

Fig. 3.3. Forecasts of the investment/income/consumption system.

Obviously, for $T = 30$, the theoretical and actual percentages are in best agreement if the approximate MSEs $\Sigma_{\hat{y}}(h)$ are used in setting up the forecast intervals. On the other hand, only forecast intervals based on $\widehat{\Sigma}_y(h) = \sum_{i=0}^{h-1} \widehat{\Phi}_i \widehat{\Sigma}_u \widehat{\Phi}_i'$ and $\widehat{\Sigma}_{\hat{y}}(h)$ are feasible in practice when the actual process coefficients are unknown and have to be estimated. Comparing only the results based on these two MSE matrices shows that it pays to use the asymptotic approximation $\widehat{\Sigma}_{\hat{y}}(h)$.

In Table 3.2, we also give the corresponding results for $T = 100$. Because the estimation uncertainty decreases with increasing sample size, one would expect that now the theoretical and actual percentages are in good agreement for all MSEs. This is precisely what can be observed in the table. Nevertheless, even now the use of the MSE adjustment in $\widehat{\Sigma}_{\widehat{y}}(1)$ gives slightly more accurate interval forecasts.

3.6 Testing for Causality

3.6.1 A Wald Test for Granger-Causality

In Chapter 2, Section 2.3.1, we have partitioned the VAR(p) process y_t in subprocesses z_t and x_t, that is, $y_t' = (z_t', x_t')$ and we have defined Granger-causality from x_t to z_t and vice versa. We have seen that this type of causality can be characterized by specific zero constraints on the VAR coefficients (see Corollary 2.2.1). Thus, in an estimated VAR(p) system, if we want to test for Granger-causality, we need to test zero constraints for the coefficients. Given the results of Sections 3.2, 3.3, and 3.4 it is straightforward to derive *asymptotic* tests of such constraints.

More generally we consider testing

$$H_0 : C\boldsymbol{\beta} = c \quad \text{against} \quad H_1 : C\boldsymbol{\beta} \neq c, \tag{3.6.1}$$

where C is an $(N \times (K^2p + K))$ matrix of rank N and c is an $(N \times 1)$ vector. Assuming that

$$\sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(0, \Gamma^{-1} \otimes \Sigma_u) \tag{3.6.2}$$

as in LS/ML estimation, we get

$$\sqrt{T}(C\widehat{\boldsymbol{\beta}} - C\boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}\left[0, C(\Gamma^{-1} \otimes \Sigma_u)C'\right] \tag{3.6.3}$$

(see Appendix C, Proposition C.15) and, hence,

$$T(C\widehat{\boldsymbol{\beta}} - c)' \left[C(\Gamma^{-1} \otimes \Sigma_u)C'\right]^{-1} (C\widehat{\boldsymbol{\beta}} - c) \xrightarrow{d} \chi^2(N). \tag{3.6.4}$$

This statistic is the *Wald statistic* (see Appendix C.7).

Replacing Γ and Σ_u by their usual estimators $\widehat{\Gamma} = ZZ'/T$ and $\widehat{\Sigma}_u$ as given in (3.2.19), the resulting statistic

$$\lambda_W = (C\widehat{\boldsymbol{\beta}} - c)' \left[C((ZZ')^{-1} \otimes \widehat{\Sigma}_u)C'\right]^{-1} (C\widehat{\boldsymbol{\beta}} - c) \tag{3.6.5}$$

still has an asymptotic χ^2-distribution with N degrees of freedom, provided y_t satisfies the conditions of Proposition 3.2, because under these conditions $[C((ZZ')^{-1} \otimes \widehat{\Sigma}_u)C']^{-1}/T$ is a consistent estimator of $\left[C(\Gamma^{-1} \otimes \Sigma_u)C'\right]^{-1}$. Hence, we have the following result.

Proposition 3.5 (*Asymptotic Distribution of the Wald Statistic*)
Suppose (3.6.2) holds. Furthermore, $\text{plim}(ZZ'/T) = \Gamma$, $\text{plim}\ \widehat{\Sigma}_u = \Sigma_u$ are both nonsingular and $H_0 : C\boldsymbol{\beta} = c$ is true, with C being an $(N \times (K^2p + K))$ matrix of rank N. Then

$$\lambda_W = (C\widehat{\boldsymbol{\beta}} - c)'[C((ZZ')^{-1} \otimes \widehat{\Sigma}_u)C']^{-1}(C\widehat{\boldsymbol{\beta}} - c) \xrightarrow{d} \chi^2(N).$$

■

In practice, it may be useful to make adjustments to the statistic or the critical values of the test to compensate for the fact that the matrix $\Gamma^{-1} \otimes \Sigma_u$ is unknown and has been replaced by an estimator. Working in that direction, we note that

$$NF(N,T) \underset{T \to \infty}{\xrightarrow{d}} \chi^2(N), \tag{3.6.6}$$

where $F(N,T)$ denotes an F random variable with N and T degrees of freedom (d.f.) (Appendix C, Proposition C.3). Because an $F(N,T)$-distribution has a fatter tail than the $\chi^2(N)$-distribution divided by N, it seems reasonable to consider the test statistic

$$\lambda_F = \lambda_W / N \tag{3.6.7}$$

in conjunction with critical values from some F-distribution. The question is then what numbers of degrees of freedom should be used? From the foregoing discussion it is plausible to use N as the numerator degrees of freedom. On the other hand, any sequence that goes to infinity with the sample size qualifies as a candidate for the denominator d.f. The usual F-statistic for a regression model with nonstochastic regressors has denominator d.f. equal to the sample size minus the number of estimated parameters. Therefore we may use this number here too. Note that, in the model (3.2.3), we have a vector $\mathbf{y}$ with KT observations and $\boldsymbol{\beta}$ contains $K(Kp+1)$ parameters. Alternatively, we will argue shortly that $T - Kp - 1$ is also a reasonable number for the denominator d.f. Hence, we have the approximate distributions

$$\lambda_F \approx F(N, KT - K^2p - K) \approx F(N, T - Kp - 1). \tag{3.6.8}$$

3.6.2 An Example

To see how this result can be used in a test for Granger-causality, let us consider again our example system from Section 3.2.3. The null hypothesis of no Granger-causality from income/consumption (y_2, y_3) to investment (y_1) may be expressed in terms of the coefficients of the VAR(2) process as

$$H_0 : \alpha_{12,1} = \alpha_{13,1} = \alpha_{12,2} = \alpha_{13,2} = 0. \tag{3.6.9}$$

This null hypothesis may be written as in (3.6.1) by defining the (4×1) vector $c = 0$ and the (4×21) matrix

$$C = \begin{bmatrix} 0\,0\,0 & 0\,0\,0 & 1\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 1\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 1\,0\,0 & 0\,0\,0 \\ 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 0\,0\,0 & 1\,0\,0 \end{bmatrix}.$$

With this notation, using the estimation results from Section 3.2.3,

$$\lambda_F = \widehat{\boldsymbol{\beta}}' C' \left[C((ZZ')^{-1} \otimes \widehat{\Sigma}_u) C' \right]^{-1} C \widehat{\boldsymbol{\beta}}/4 = 1.59. \tag{3.6.10}$$

In contrast, the 95th percentile of the $F(4, 3 \cdot 73 - 9 \cdot 2 - 3) = F(4, 198) \approx F(4, 73 - 3 \cdot 2 - 1) = F(4, 66)$-distribution is about 2.5. Thus, in a 5% level test, we cannot reject Granger-noncausality from income/consumption to investment.

In this example, the denominator d.f. are so large (namely 198 or 66) that we could just as well use λ_W in conjunction with a critical value from a $\chi^2(4)$-distribution. The 95th percentile of that distribution is 9.49 and, thus, it is about four times that of the F-test while $\lambda_W = 4\lambda_F$.

In an example of this type it is quite reasonable to use $T - Kp - 1$ denominator d.f. for the F-test because all the restrictions are imposed on coefficients from one equation. Therefore λ_F actually reduces to an F-statistic related to one equation with $Kp + 1$ parameters which are estimated from T observations. The use of $T - Kp - 1$ d.f. may also be justified by arguments that do not rely on the restrictions being imposed on the parameters of one equation only, namely by appealing to the similarity between the λ_F statistic and Hotelling's T^2 (e.g., Anderson (1984)).

Many other tests for Granger-causality have been proposed and investigated (see, e.g., Geweke, Meese & Dent (1983)). In the next chapter, we will return to the testing of hypotheses and then an alternative test will be considered.

3.6.3 Testing for Instantaneous Causality

Tests for instantaneous causality can be developed in the same way as tests for Granger-causality because instantaneous causality can be expressed in terms of zero restrictions for $\boldsymbol{\sigma} = \mathrm{vech}(\Sigma_u)$ (see Proposition 2.3). If y_t is a stable Gaussian VAR(p) process and we wish to test

$$H_0 : C\boldsymbol{\sigma} = 0 \quad \text{against} \quad H_1 : C\boldsymbol{\sigma} \neq 0, \tag{3.6.11}$$

we may use the asymptotic distribution of the ML estimator given in Proposition 3.4 to set up the Wald statistic

$$\lambda_W = T\widetilde{\boldsymbol{\sigma}}' C' [2C\mathbf{D}_K^+ (\widetilde{\Sigma}_u \otimes \widetilde{\Sigma}_u) \mathbf{D}_K^{+\prime} C']^{-1} C\widetilde{\boldsymbol{\sigma}}, \tag{3.6.12}$$

where $\mathbf{D}_K^+$ is the Moore-Penrose inverse of the duplication matrix $\mathbf{D}_K$ and C is an $(N \times K(K+1)/2)$ matrix of rank N. Under H_0, λ_W has an asymptotic χ^2-distribution with N degrees of freedom.

Alternatively, a Wald test of (3.6.11) could be based on the lower triangular matrix P which is obtained from a Choleski decomposition of Σ_u. Noting that instantaneous noncausality implies zero elements of Σ_u that correspond to zero elements of P, we can write H_0 from (3.6.11) equivalently as

$$H_0 : C\text{vech}(P) = 0. \tag{3.6.13}$$

Because vech(P) is a continuously differentiable function of $\boldsymbol{\sigma}$, the asymptotic distribution of the estimator P obtained from decomposing $\widetilde{\Sigma}_u$ follows from Proposition C.15(3) of Appendix C:

$$\sqrt{T}\,\text{vech}(\widetilde{P} - P) \xrightarrow{d} \mathcal{N}(0, \bar{H}\Sigma_{\widetilde{\boldsymbol{\sigma}}}\bar{H}'), \tag{3.6.14}$$

where

$$\bar{H} = \frac{\partial \text{vech}(P)}{\partial \boldsymbol{\sigma}'} = [\mathbf{L}_K(I_{K^2} + \mathbf{K}_{KK})(P \otimes I_K)\mathbf{L}_K']^{-1}$$

(see Appendix A.13, Rule (10)). Here $\mathbf{K}_{mn}$ is the commutation matrix defined such that $\text{vec}(G) = \mathbf{K}_{mn}\text{vec}(G')$ for any $(m \times n)$ matrix G and $\mathbf{L}_K$ is the $(\frac{1}{2}K(K+1) \times K^2)$ elimination matrix defined such that $\text{vech}(F) = \mathbf{L}_K\,\text{vec}(F)$ for any $(K \times K)$ matrix F (see Appendix A.12.2). A Wald test of (3.6.13) may therefore be based on

$$\lambda_W = T\text{vech}(\widetilde{P})'C'[C\widehat{\bar{H}}\widehat{\Sigma}_{\widetilde{\boldsymbol{\sigma}}}\widehat{\bar{H}}'C']^{-1}C\,\text{vech}(\widetilde{P}) \xrightarrow{d} \chi^2(N), \tag{3.6.15}$$

where hats denote the usual estimators. Although the two tests based on $\widetilde{\boldsymbol{\sigma}}$ and $\widetilde{P}$ are derived from the same asymptotic distribution, they may differ in small samples. Of course, in the previous discussion we may replace $\widetilde{\Sigma}_u$ by the asymptotically equivalent estimator $\widehat{\Sigma}_u$.

In our investment/income/consumption example, suppose we wish to test for instantaneous causality between (income, consumption) and investment. Following Proposition 2.3, the null hypothesis of no causality is

$$H_0 : \sigma_{21} = \sigma_{31} = 0 \quad \text{or} \quad C\boldsymbol{\sigma} = 0,$$

where σ_{ij} is a typical element of Σ_u and

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

For this hypothesis, the test statistic in (3.6.12) assumes the value $\lambda_W = 5.46$. Alternatively, we may test

$$H_0 : p_{21} = p_{31} = 0 \quad \text{or} \quad C\,\text{vech}(P) = 0,$$

where p_{ij} is a typical element of P. The corresponding value of the test statistic from (3.6.15) is $\lambda_W = 5.70$. Both tests are based on asymptotic $\chi^2(2)$-distributions and therefore do not reject the null hypothesis of no instantaneous causality at a 5% level. Note that the critical value for a 5% level test is 5.99.

3.6.4 Testing for Multi-Step Causality

In Section 2.3.1, we have also discussed the possibility of extending the information set and considering causality between two variables in a system that includes further variables. Using the same ideas as in the definition of Granger-causality resulted in the definition of h-step causality. This concept implies nonlinear restrictions for the VAR coefficients for which the usual application of the Wald principle does not result in a valid test. The following example from Lütkepohl & Burda (1997) illustrates the problem.

Consider a three-dimensional VAR(1) process:

$$\begin{bmatrix} z_t \\ y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \alpha_{zz} & \alpha_{zy} & \alpha_{zx} \\ \alpha_{yz} & \alpha_{yy} & \alpha_{yx} \\ \alpha_{xz} & \alpha_{xy} & \alpha_{xx} \end{bmatrix} \begin{bmatrix} z_{t-1} \\ y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} u_{z,t} \\ u_{y,t} \\ u_{x,t} \end{bmatrix}. \tag{3.6.16}$$

From (2.3.24) we know that a test of ∞-step noncausality from y_t to z_t ($y_t \not\rightarrow_{(\infty)} z_t$) needs to check $h = 2$ restrictions on the VAR coefficient vector. They are of the following nonlinear form:

$$r(\boldsymbol{\alpha}) = \begin{bmatrix} R\boldsymbol{\alpha} \\ R\boldsymbol{\alpha}^{(2)} \end{bmatrix} = (I_2 \otimes R) \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\alpha}^{(2)} \end{bmatrix},$$

where

$$R = [0\,0\,0 \;\; 1\,0\,0 \;\; 0\,0\,0],$$

$\boldsymbol{\alpha} = \operatorname{vec}(A_1)$ and $\boldsymbol{\alpha}^{(2)} = \operatorname{vec}(A_1^2)$, with A_1 being the coefficient matrix of the process in (3.6.16). Hence,

$$r(\boldsymbol{\alpha}) = \begin{bmatrix} \alpha_{zy} \\ \alpha_{zz}\alpha_{zy} + \alpha_{zy}\alpha_{yy} + \alpha_{zx}\alpha_{xy} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{3.6.17}$$

Denoting the covariance matrix of the asymptotic distribution of $\sqrt{T}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ as usual by $\Sigma_{\widehat{\boldsymbol{\alpha}}}$ and a consistent estimator by $\widehat{\Sigma}_{\widehat{\boldsymbol{\alpha}}}$, the Wald statistic for testing these restrictions has the form

$$\lambda_W = Tr(\widehat{\boldsymbol{\alpha}})' \left(\widehat{\frac{\partial r}{\partial \boldsymbol{\alpha}'}} \widehat{\Sigma}_{\widehat{\boldsymbol{\alpha}}} \widehat{\frac{\partial r'}{\partial \boldsymbol{\alpha}}} \right)^{-1} r(\widehat{\boldsymbol{\alpha}}),$$

where $\widehat{\partial r}/\partial \boldsymbol{\alpha}'$ is an estimator of $\partial r/\partial \boldsymbol{\alpha}'$ (see Appendix C.7). The statistic has an asymptotic $\chi^2(2)$-distribution under the null hypothesis, provided the matrix

$$\frac{\partial r}{\partial \boldsymbol{\alpha}'} \Sigma_{\widehat{\boldsymbol{\alpha}}} \frac{\partial r'}{\partial \boldsymbol{\alpha}}$$

is nonsingular. In the present case, the latter condition is unfortunately not satisfied for all relevant parameter values.

To see this, note that the matrix of first order partial derivatives of the function $r(\boldsymbol{\alpha})$ is

$$\frac{\partial r}{\partial \boldsymbol{\alpha}'} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{zy} & 0 & 0 & \alpha_{zz}+\alpha_{yy} & \alpha_{zy} & \alpha_{zx} & \alpha_{xy} & 0 & 0 \end{bmatrix}.$$

The restrictions (3.6.17) are satisfied if

$$\alpha_{zy} = \alpha_{zx} = 0, \alpha_{xy} \neq 0, \tag{3.6.18}$$

or

$$\alpha_{zy} = \alpha_{xy} = 0, \alpha_{zx} \neq 0, \tag{3.6.19}$$

or

$$\alpha_{zy} = \alpha_{zx} = \alpha_{xy} = 0. \tag{3.6.20}$$

Clearly, $\partial r/\partial \boldsymbol{\alpha}'$ has rank 1 only and, thus,

$$\mathrm{rk}\left(\frac{\partial r}{\partial \boldsymbol{\alpha}'} \Sigma_{\widehat{\boldsymbol{\alpha}}} \frac{\partial r'}{\partial \boldsymbol{\alpha}}\right) = 1,$$

if (3.6.20) holds. Hence, the standard Wald statistic will not have its asymptotic $\chi^2(2)$-distribution under the null hypothesis $r(\boldsymbol{\alpha}) = 0$ if (3.6.20) holds.

Lütkepohl & Burda (1997) discussed a possibility to circumvent the problem by simply drawing a random variable from a normal distribution and adding it to the second restriction. Thereby a nonsingular distribution of the modified restriction vector is obtained and a Wald type statistic can be constructed for this vector.

More generally, Lütkepohl & Burda (1997) proposed the following approach for testing the null hypothesis that the K_y-dimensional vector y_t is not h-step causal for the K_z-dimensional vector z_t ($y_t \not\rightarrow_{(h)} z_t$) if additional K_x variables x_t are present in the system of interest. Using the notation from Section 2.3.1, that is, $\mathbf{A}$ is defined as in the VAR(1) representation (2.1.8), $J := [I_K : 0 : \cdots : 0]$ is a $(K \times Kp)$ matrix, $A^{(j)} := J\mathbf{A}^j$, and $\boldsymbol{\alpha}^{(j)} := \mathrm{vec}(A^{(j)})$, the hypotheses of interest can be stated as

$$H_0 : (I_h \otimes R)\, \mathsf{a}^{(h)} = 0 \quad \text{against} \quad H_1 : (I_h \otimes R)\, \mathsf{a}^{(h)} \neq 0, \tag{3.6.21}$$

where R is a $(pK_zK_y \times pK^2)$ matrix, as defined in (2.3.23), and

$$\mathsf{a}^{(h)} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\alpha}^{(2)} \\ \vdots \\ \boldsymbol{\alpha}^{(h)} \end{bmatrix}.$$

Let $\widehat{\mathsf{a}}^{(h)}$ be the estimator corresponding to $\mathsf{a}^{(h)}$ based on the multivariate LS estimator $\widehat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$. Furthermore, we denote by $\operatorname{diag}(D)$ a diagonal matrix which has the diagonal elements of the square matrix D on its main diagonal and define the $(hpK_zK_y \times hpK_zK_y)$ matrix

$$\widehat{\Sigma}_w(h) = \begin{bmatrix} 0 & 0 \\ 0 & I_{h-1} \otimes \operatorname{diag}(R\widehat{\Sigma}_{\widehat{\boldsymbol{\alpha}}}R') \end{bmatrix}.$$

Moreover, we define a random vector $w_\lambda^{(h)} \sim \mathcal{N}(0, \lambda\widehat{\Sigma}_w(h))$ which is drawn independently of $\widehat{\boldsymbol{\alpha}}$. Here $\lambda > 0$ is some fixed real number. Lütkepohl & Burda (1997) defined the following modified Wald statistic for testing the pair of hypotheses in (3.6.21):

$$\begin{aligned} \lambda_W^{mod} &= T\left((I_h \otimes R)\,\widehat{\mathsf{a}}^{(h)} + \frac{w_\lambda^{(h)}}{\sqrt{T}}\right)' \\ &\quad \times \left[(I_h \otimes R)\,\widehat{\Sigma}_{\widehat{\mathsf{a}}}(h)\,(I_h \otimes R') + \lambda\widehat{\Sigma}_w(h)\right]^{-1} \\ &\quad \times \left((I_h \otimes R)\,\widehat{\mathsf{a}}^{(h)} + \frac{w_\lambda^{(h)}}{\sqrt{T}}\right). \end{aligned}$$

Here $\widehat{\Sigma}_{\widehat{\mathsf{a}}}(h)$ is a consistent estimator of the asymptotic covariance matrix of $\sqrt{T}(\widehat{\mathsf{a}}^{(h)} - \mathsf{a}^{(h)})$. It can be shown that

$$\lambda_W^{mod} \xrightarrow{d} \chi^2(hpK_zK_y)$$

under H_0. Notice that there is no need to add anything to the first pK_zK_y components of $(I_h \otimes R)\widehat{\mathsf{a}}^{(h)}$ because they are equal to $R\widehat{\boldsymbol{\alpha}}$ which has a nonsingular asymptotic distribution.

Clearly, adding some random term to $\widehat{\mathsf{a}}^{(h)}$ reduces the efficiency of the procedure and is likely to result in a loss in power of the test relative to a procedure which does not use this device. In particular, if the noise term is substantial in relation to the estimated variance, there may be some loss in power. Therefore, the amount of noise (the variance of the noise) is linked to the variance of the estimator through $\Sigma_w(h)$. Moreover, the quantity λ may be chosen close to zero. Thereby the loss in efficiency can be made arbitrarily small.

There are in fact also other possibilities to avoid the problems related to the Wald test. One way to get around it is to impose zero restrictions directly on the VAR coefficients prior to analyzing multi-step causality. The relevant subset models will be discussed in Chapter 5.

3.7 The Asymptotic Distributions of Impulse Responses and Forecast Error Variance Decompositions

3.7.1 The Main Results

In Chapter 2, Section 2.3.2, we have seen that the coefficients of the MA representations

$$y_t = \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i}, \quad \Phi_0 = I_K, \tag{3.7.1}$$

and

$$y_t = \mu + \sum_{i=0}^{\infty} \Theta_i w_{t-i} \tag{3.7.2}$$

are sometimes interpreted as impulse responses or dynamic multipliers of the system of variables y_t. Here $\mu = E(y_t)$, the $\Theta_i = \Phi_i P$, $w_t = P^{-1}u_t$, and P is the lower triangular Choleski decomposition of Σ_u such that $\Sigma_u = PP'$. Hence, $\Sigma_w = E(w_t w_t') = I_K$. In this section, we will assume that the Φ_i's and Θ_i's are unknown and they are computed from the estimated VAR coefficients and error covariance matrix. We will derive the asymptotic distributions of the resulting estimated Φ_i's and Θ_i's. In these derivations, we will not need the existence of MA representations (3.7.1) and (3.7.2). We will just assume that the Φ_i's are obtained from given coefficient matrices $A_1, \dots, A_p$ by recursions

$$\Phi_i = \sum_{j=1}^{i} \Phi_{i-j} A_j, \quad i = 1, 2, \dots,$$

starting with $\Phi_0 = I_K$ and setting $A_j = 0$ for $j > p$. Furthermore, the Θ_i's are obtained from $A_1, \dots, A_p$, and Σ_u as $\Theta_i = \Phi_i P$, where P is as specified in the foregoing. In addition, the asymptotic distributions of the corresponding accumulated responses

$$\begin{aligned}
\Psi_n &= \sum_{i=0}^{n} \Phi_i, \quad \Psi_\infty = \sum_{i=0}^{\infty} \Phi_i = (I_K - A_1 - \cdots - A_p)^{-1} \quad \text{(if it exists)},\\
\Xi_n &= \sum_{i=0}^{n} \Theta_i, \quad \Xi_\infty = \sum_{i=0}^{\infty} \Theta_i = (I_K - A_1 - \cdots - A_p)^{-1} P \quad \text{(if it exists)},
\end{aligned}$$

and the forecast error variance components,

$$\omega_{jk,h} = \sum_{i=0}^{h-1} (e_j' \Theta_i e_k)^2 / \mathrm{MSE}_j(h), \tag{3.7.3}$$

will be given. Here e_k is the k-th column of I_K and

$$\mathrm{MSE}_j(h) = \sum_{i=0}^{h-1} e_j' \Phi_i \Sigma_u \Phi_i' e_j$$

is the j-th diagonal element of the MSE matrix $\Sigma_y(h)$ of an h-step forecast (see Chapter 2, Section 2.2.2).

The derivation of the asymptotic distributions is based on the following result from Appendix C, Proposition C.15(3). Suppose $\boldsymbol{\beta}$ is an $(n \times 1)$ vector of parameters and $\widehat{\boldsymbol{\beta}}$ is an estimator such that

$$\sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\widehat{\boldsymbol{\beta}}}),$$

where T, as usual, denotes the sample size (time series length) used for estimation. Let $g(\boldsymbol{\beta})$ be a continuously differentiable function with values in the m-dimensional Euclidean space and suppose that $\partial g_i/\partial \boldsymbol{\beta}' = (\partial g_i/\partial \beta_j)$ is nonzero at the true vector $\boldsymbol{\beta}$, for $i = 1, \dots, m$. Then,

$$\sqrt{T}\left[g(\widehat{\boldsymbol{\beta}}) - g(\boldsymbol{\beta})\right] \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial g}{\partial \boldsymbol{\beta}'} \Sigma_{\widehat{\boldsymbol{\beta}}} \frac{\partial g'}{\partial \boldsymbol{\beta}}\right).$$

In writing down the asymptotic distributions formally, we use the notation

$$\begin{aligned}
\boldsymbol{\alpha} &:= \mathrm{vec}(A_1, \dots, A_p) && (K^2 p \times 1), \\
\mathbf{A} &:= \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_K & 0 & \dots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_K & 0 \end{bmatrix} && (Kp \times Kp), \\
\boldsymbol{\sigma} &:= \mathrm{vech}(\Sigma_u) && (\tfrac{1}{2}K(K+1) \times 1)
\end{aligned}$$

and the corresponding estimators are furnished with a hat. As before, vec denotes the column stacking operator and vech is the corresponding operator that stacks the elements on and below the main diagonal only. We also use the commutation matrix $\mathbf{K}_{mn}$, defined such that, for any $(m \times n)$ matrix G, $\mathbf{K}_{mn}\mathrm{vec}(G) = \mathrm{vec}(G')$, the $(m^2 \times \frac{1}{2}m(m+1))$ duplication matrix $\mathbf{D}_m$, defined such that $\mathbf{D}_m\mathrm{vech}(F) = \mathrm{vec}(F)$, for any symmetric $(m \times m)$ matrix F, and the $(\frac{1}{2}m(m+1) \times m^2)$ elimination matrix $\mathbf{L}_m$, defined such that, for any $(m \times m)$ matrix F, $\mathrm{vech}(F) = \mathbf{L}_m\mathrm{vec}(F)$ (see Appendix A.12.2). Furthermore, $J := [I_K : 0 : \cdots : 0]$ is a $(K \times Kp)$ matrix. With this notation, the following proposition from Lütkepohl (1990) can be stated.

Proposition 3.6 (*Asymptotic Distributions of Impulse Responses*)
Suppose

$$\sqrt{T}\begin{bmatrix} \widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \\ \widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma} \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} \Sigma_{\widehat{\boldsymbol{\alpha}}} & 0 \\ 0 & \Sigma_{\widehat{\boldsymbol{\sigma}}} \end{bmatrix}\right). \tag{3.7.4}$$

Then

$$\sqrt{T}\,\mathrm{vec}(\widehat{\Phi}_i - \Phi_i) \xrightarrow{d} \mathcal{N}(0, G_i \Sigma_{\widehat{\boldsymbol{\alpha}}} G_i'), \quad i = 1, 2, \dots, \tag{3.7.5}$$

where

$$G_i := \frac{\partial\,\mathrm{vec}(\Phi_i)}{\partial \boldsymbol{\alpha}'} = \sum_{m=0}^{i-1} J(\mathbf{A}')^{i-1-m} \otimes \Phi_m.$$

$$\sqrt{T}\,\mathrm{vec}(\widehat{\Psi}_n - \Psi_n) \xrightarrow{d} \mathcal{N}(0, F_n \Sigma_{\widehat{\boldsymbol{\alpha}}} F_n'), \quad n = 1, 2, \dots, \tag{3.7.6}$$

where $F_n := G_1 + \cdots + G_n$.

If $(I_K - A_1 - \cdots - A_p)$ is nonsingular,

$$\sqrt{T}\,\mathrm{vec}(\widehat{\Psi}_\infty - \Psi_\infty) \xrightarrow{d} \mathcal{N}(0, F_\infty \Sigma_{\widehat{\boldsymbol{\alpha}}} F_\infty'), \tag{3.7.7}$$

where $F_\infty := \underbrace{(\Psi_\infty', \dots, \Psi_\infty')}_{p \text{ times}} \otimes \Psi_\infty$.

$$\sqrt{T}\,\mathrm{vec}(\widehat{\Theta}_i - \Theta_i) \xrightarrow{d} \mathcal{N}(0, C_i \Sigma_{\widehat{\boldsymbol{\alpha}}} C_i' + \bar{C}_i \Sigma_{\widehat{\boldsymbol{\alpha}}} \bar{C}_i'), \quad i = 0, 1, 2, \dots, \tag{3.7.8}$$

where

$$C_0 := 0,\ C_i := (P' \otimes I_K) G_i, i = 1, 2, \dots,\ \bar{C}_i := (I_K \otimes \Phi_i) H,\ i = 0, 1, \dots,$$

and

$$\begin{aligned} H &:= \frac{\partial\,\mathrm{vec}(P)}{\partial \boldsymbol{\sigma}'} = \mathbf{L}_K' \{\mathbf{L}_K [(I_K \otimes P)\mathbf{K}_{KK} + (P \otimes I_K)] \mathbf{L}_K'\}^{-1} \\ &= \mathbf{L}_K' \{\mathbf{L}_K (I_{K^2} + \mathbf{K}_{KK})(P \otimes I_K) \mathbf{L}_K'\}^{-1}. \end{aligned}$$

$$\sqrt{T}\ \mathrm{vec}(\widehat{\Xi}_n - \Xi_n) \xrightarrow{d} \mathcal{N}(0, B_n \Sigma_{\widehat{\boldsymbol{\alpha}}} B_n' + \bar{B}_n \Sigma_{\widehat{\boldsymbol{\sigma}}} \bar{B}_n'), \tag{3.7.9}$$

where $B_n := (P' \otimes I_K) F_n$ and $\bar{B}_n := (I_K \otimes \Psi_n) H$.

If $(I_K - A_1 - \cdots - A_p)$ is nonsingular,

$$\sqrt{T}\ \mathrm{vec}(\widehat{\Xi}_\infty - \Xi_\infty) \xrightarrow{d} \mathcal{N}(0, B_\infty \Sigma_{\widehat{\boldsymbol{\alpha}}} B_\infty' + \bar{B}_\infty \Sigma_{\widehat{\boldsymbol{\sigma}}} \bar{B}_\infty'), \tag{3.7.10}$$

where $B_\infty := (P' \otimes I_K) F_\infty$ and $\bar{B}_\infty := (I_K \otimes \Psi_\infty) H$.

Finally,

$$\begin{aligned} \sqrt{T}(\widehat{\omega}_{jk,h} - \omega_{jk,h}) &\xrightarrow{d} \mathcal{N}(0, d_{jk,h} \Sigma_{\widehat{\boldsymbol{\alpha}}} d_{jk,h}' + \overline{d}_{jk,h} \Sigma_{\widehat{\boldsymbol{\sigma}}} \overline{d}_{jk,h}') \\ & j, k = 1, \dots, K, \quad h = 1, 2, \dots, \end{aligned} \tag{3.7.11}$$

where

$$\begin{aligned} d_{jk,h} \ &:= \frac{2}{\mathrm{MSE}_j(h)^2} \sum_{i=0}^{h-1} \Big[\mathrm{MSE}_j(h)(e_j' \Phi_i P e_k)(e_k' P' \otimes e_j') G_i \\ &\qquad - (e_j' \Phi_i P e_k)^2 \sum_{m=0}^{h-1} (e_j' \Phi_m \Sigma_u \otimes e_j') G_m \Big] \end{aligned}$$

with $G_0 := 0$ and

$$\overline{d}_{jk,h} := \sum_{i=0}^{h-1} \Big[2\,\mathrm{MSE}_j(h)(e_j'\Phi_i P e_k)(e_k' \otimes e_j'\Phi_i)H \\ -(e_j'\Phi_i P e_k)^2 \sum_{m=0}^{h-1} (e_j'\Phi_m \otimes e_j'\Phi_m)\mathbf{D}_K \Big] \Big/ \mathrm{MSE}_j(h)^2.$$

■

In the next subsection, the proof of the proposition is indicated. Some remarks are worthwhile now.

Remark 1 In the proposition, some matrices of partial derivatives may be zero. For instance, if a VAR(1) model is fitted although the true order is zero, that is, y_t is white noise, then

$$G_2 = J\mathbf{A}' \otimes I_K + JI_K \otimes \Phi_1 = 0$$

because $\mathbf{A} = A_1 = 0$ and $\Phi_1 = A_1 = 0$. Hence, a degenerate asymptotic distribution with zero covariance matrix is obtained for $\sqrt{T}\,\mathrm{vec}(\widehat{\Phi}_2 - \Phi_2)$. As explained in Appendix B, we call such a distribution also multivariate normal. Otherwise it would be necessary to distinguish between cases with zero and nonzero partial derivatives or we have to assume that all partial derivatives are such that the covariance matrices have no zeros on the diagonal. Note that estimators of the covariance matrices obtained by replacing unknown quantities by their usual estimators may be problematic when the asymptotic distribution is degenerate. In that case, the usual t-ratios and confidence intervals may not be appropriate.

To illustrate the potential problems resulting from a degenerate asymptotic distribution, we follow Benkwitz, Lütkepohl & Neumann (2000) and consider a univariate AR(1) process $y_t = \alpha y_{t-1} + u_t$. In this case, $\Phi_i = \alpha^i$. Suppose that $\widehat{\alpha}$ is an estimator of α satisfying $\sqrt{T}(\widehat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\widehat{\alpha}})$ with $\sigma^2_{\widehat{\alpha}} \neq 0$. For instance, $\widehat{\alpha}$ may be the LS estimator of α. Then

$$\sqrt{T}(\widehat{\alpha}^2 - \alpha^2) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\widehat{\alpha}^2})$$

with $\sigma^2_{\widehat{\alpha}^2} = 4\alpha^2\sigma^2_{\widehat{\alpha}}$. This quantity is, of course, zero if $\alpha = 0$. In the latter case, $\sqrt{T}\widehat{\alpha}/\sigma_{\widehat{\alpha}}$ has an asymptotic standard normal distribution and, hence, $T\widehat{\alpha}^2/\sigma^2_{\widehat{\alpha}}$ has an asymptotic $\chi^2(1)$-distribution. Thus, it is clear that in this case $\sqrt{T}\widehat{\alpha}^2$ is asymptotically degenerate.

Because the estimated $\sigma^2_{\widehat{\alpha}^2}$ obtained by replacing α and $\sigma^2_{\widehat{\alpha}}$ by their usual LS estimators is nonzero almost surely, it is tempting to use the quantity $\sqrt{T}(\widehat{\alpha}^2 - \alpha^2)/2\widehat{\alpha}\widehat{\sigma}_{\widehat{\alpha}}$ for constructing a confidence interval, say, for Φ_2. However, for $\alpha = 0$, the t-ratio becomes $\sqrt{T}\widehat{\alpha}/2\widehat{\sigma}_{\widehat{\alpha}}$ which converges to $\mathcal{N}(0, 1/4)$ asymptotically, because $\sqrt{T}\widehat{\alpha}/\widehat{\sigma}_{\widehat{\alpha}} \xrightarrow{d} \mathcal{N}(0, 1)$. A confidence interval constructed on the basis of the asymptotic standard normal distribution would therefore be a conservative one. In other words, asymptotic inference which ignores the possible singularity in the asymptotic distribution of the impulse responses may be misleading (see Benkwitz et al. (2000) for further discussion). ■

Remark 2 In the proposition, it is not explicitly assumed that y_t is stable. While the stability condition is partly introduced in (3.7.7) and (3.7.10) by requiring that $(I_K - A_1 - \cdots - A_p)$ be nonsingular so that

$$\det(I_K - A_1 z - \cdots - A_p z^p) \neq 0 \quad \text{for } z = 1,$$

it is not needed for the other results to hold. The crucial condition is the asymptotic distribution of the process parameters in (3.7.4). Although we have used the stationarity and stability assumptions in Sections 3.2–3.4 in order to derive the asymptotic distribution of the process parameters, we will see in later chapters that asymptotic normality is also obtained for certain nonstationary, unstable processes. Therefore, at least parts of Proposition 3.6 will be useful in a nonstationary environment. ■

Remark 3 The block-diagonal structure of the covariance matrix of the asymptotic distribution in (3.7.4) is in no way essential for the asymptotic normality of the impulse responses. In fact, the asymptotic distributions in (3.7.5)–(3.7.7) remain unchanged if the asymptotic covariance matrix of the parameter estimators is not block-diagonal. On the other hand, without the block-diagonal structure, the simple additive structure of the asymptotic covariance matrices in (3.7.8)–(3.7.11) is lost. Although these asymptotic distributions are easily generalizable to the case of a general asymptotic covariance matrix of the VAR coefficients in (3.7.4), we have not stated the more general result here because it is not needed in subsequent chapters of this text. ■

Remark 4 Under the conditions of Proposition 3.4, the covariance matrix of the asymptotic distribution of the parameters has precisely the block-diagonal structure assumed in (3.7.4) with

$$\Sigma_{\widehat{\boldsymbol{\alpha}}} = \Gamma_Y(0)^{-1} \otimes \Sigma_u$$

and

$$\Sigma_{\widehat{\boldsymbol{\sigma}}} = 2\mathbf{D}_K^+(\Sigma_u \otimes \Sigma_u)\mathbf{D}_K^{+\prime},$$

where $\mathbf{D}_K^+ = (\mathbf{D}_K'\mathbf{D}_K)^{-1}\mathbf{D}_K'$ is the Moore-Penrose inverse of the duplication matrix $\mathbf{D}_K$. Using these expressions in the proposition, some simplifications of the covariance matrices can be obtained. For instance, the covariance matrix in (3.7.5) becomes

$$G_i \Sigma_{\widehat{\boldsymbol{\alpha}}} G_i'$$

$$= \left[\sum_{m=0}^{i-1} J(\mathbf{A}')^{i-1-m} \otimes \Phi_m \right] (\Gamma_Y(0)^{-1} \otimes \Sigma_u) \left[\sum_{n=0}^{i-1} J(\mathbf{A}')^{i-1-n} \otimes \Phi_n \right]'$$

$$= \sum_{m=0}^{i-1} \sum_{n=0}^{i-1} \left[J(\mathbf{A}')^{i-1-m} \Gamma_Y(0)^{-1} \mathbf{A}^{i-1-n} J' \right] \otimes (\Phi_m \Sigma_u \Phi_n')$$

which is computationally convenient because all matrices involved are of a relatively small size. The advantage of the general formulation is that it can be used with other $\Sigma_{\widehat{\boldsymbol{\alpha}}}$ matrices as well. We will see examples in subsequent chapters. ■

Remark 5 In practice, to use the asymptotic distributions for inference, the unknown quantities in the covariance matrices in Proposition 3.6 may be replaced by their usual estimators given in Sections 3.2–3.4 for the case of a stationary, stable process y_t (see, however, Remark 1). ■

Remark 6 Summing the forecast error variance components over k,

$$\sum_{k=1}^{K} \omega_{jk,h} = \sum_{k=1}^{K} \widehat{\omega}_{jk,h} = 1$$

for each j and h. These restrictions are not taken into account in the derivation of the asymptotic distributions in (3.7.11). It is easily checked, however, that for dimension $K = 1$ the standard errors obtained from Proposition 3.6 are zero as they should be, because all forecast error variance components are 1 in that case. A problem in this context is that the asymptotic distribution of $\widehat{\omega}_{jk,h}$ cannot be used in the usual way for tests of significance and setting up confidence intervals if $\omega_{jk,h} = 0$. In that case, from the definitions of $d_{jk,h}$ and $\overline{d}_{jk,h}$, the variance of the asymptotic distribution is easily seen to be zero and, hence, estimating this quantity by replacing unknown parameters by their usual estimators may lead to t-ratios that are not standard normal asymptotically and, hence, cannot be used in the usual way for inference (see Remark 1). This state of affairs is unfortunate from a practical point of view because testing the significance of forecast error variance components is of particular interest in practice. Note, however, that

$$\omega_{jk,h} = 0 \iff \theta_{jk,i} = 0 \quad \text{for} \quad i = 0, \dots, h.$$

A test of the latter hypothesis may be possible. ■

Remark 7 Joint confidence regions and test statistics for testing hypotheses that involve several of the response coefficients can be obtained from Proposition 3.6 in the usual way. However, it has to be taken into account that, for instance, the elements of $\widehat{\Phi}_i$ and $\widehat{\Phi}_j$ will not be independent asymptotically. If elements from two or more MA matrices are involved the joint distribution of all the matrices must be determined. This distribution can be derived easily from the results given in the proposition. For instance, the covariance matrix of the joint asymptotic distribution of $\mathrm{vec}(\widehat{\Phi}_i, \widehat{\Phi}_j)$ is

$$\frac{\partial \, \mathrm{vec}(\Phi_i, \Phi_j)}{\partial \boldsymbol{\alpha}'} \Sigma_{\widehat{\boldsymbol{\alpha}}} \frac{\partial \, \mathrm{vec}(\Phi_i, \Phi_j)'}{\partial \boldsymbol{\alpha}},$$

where

$$\frac{\partial \, \mathrm{vec}(\Phi_i, \Phi_j)}{\partial \boldsymbol{\alpha}'} = \begin{bmatrix} \dfrac{\partial \, \mathrm{vec}(\Phi_i)}{\partial \boldsymbol{\alpha}'} \\ \dfrac{\partial \, \mathrm{vec}(\Phi_j)}{\partial \boldsymbol{\alpha}'} \end{bmatrix}$$

etc. We have chosen to state the proposition for individual MA coefficient matrices because thereby all required matrices have relatively small dimensions and, hence, are easy to compute. ■

Remark 8 Denoting the jk-th elements of Φ_i and Θ_i by $\phi_{jk,i}$ and $\theta_{jk,i}$, respectively, hypotheses of obvious interest, for $j \neq k$, are

$$H_0 : \phi_{jk,i} = 0 \quad \text{for} \quad i = 1, 2, \ldots \tag{3.7.12}$$

and

$$H_0 : \theta_{jk,i} = 0 \quad \text{for} \quad i = 0, 1, 2, \ldots \tag{3.7.13}$$

because they can be interpreted as hypotheses on noncausality from variable k to variable j, that is, an impulse in variable k does not induce any response of variable j. From Chapter 2, Propositions 2.4 and 2.5, we know that (3.7.12) is equivalent to

$$H_0 : \phi_{jk,i} = 0 \quad \text{for} \quad i = 1, 2, \ldots, p(K-1) \tag{3.7.14}$$

and (3.7.13) is equivalent to

$$H_0 : \theta_{jk,i} = 0 \quad \text{for} \quad i = 0, 1, \ldots, p(K-1). \tag{3.7.15}$$

Using Bonferroni's inequality (see Chapter 2, Section 2.2.3), a test of (3.7.14) with significance level at most $100\gamma\%$ is obtained by rejecting H_0 if

$$|\sqrt{T}\widehat{\phi}_{jk,i}/\widehat{\sigma}_{\phi_{jk}}(i)| > z_{(\gamma/2p(K-1))} \tag{3.7.16}$$

for at least one $i \in \{1, 2, \ldots, p(K-1)\}$. Here $z_{(\gamma)}$ is the upper 100γ percentage point of the standard normal distribution and $\widehat{\sigma}_{\phi_{jk}}(i)$ is an estimate of the asymptotic standard deviation $\sigma_{\phi_{jk}}(i)$ of $\sqrt{T}\widehat{\phi}_{jk,i}$ obtained via Proposition 3.6. In order to obtain an asymptotic standard normal distribution of the t-ratio $\sqrt{T}\widehat{\phi}_{jk,i}/\widehat{\sigma}_{\phi_{jk}}(i)$, the variance $\sigma^2_{\phi_{jk}}(i)$ must be nonzero, however.

A test of (3.7.15) with significance level at most γ is obtained by rejecting H_0 if

$$|\sqrt{T}\widehat{\theta}_{jk,i}/\widehat{\sigma}_{\theta_{jk}}(i)| \begin{cases} > z_{(\gamma/2(pK-p+1))} & \text{for at least one} \\ & i \in \{0,1,2,\dots,p(K-1)\} \text{ if } j > k \\ > z_{(\gamma/2(pK-p))} & \text{for at least one} \\ & i \in \{1,2,\dots,p(K-1)\} \text{ if } j < k. \end{cases} \tag{3.7.17}$$

Here $\widehat{\sigma}_{\theta_{jk}}(i)$ is a consistent estimator of the standard deviation of the asymptotic distribution of $\sqrt{T}\widehat{\theta}_{jk,i}$ obtained from Proposition 3.6 and that standard deviation is assumed to be nonzero.

A test based on Bonferroni's principle may have quite low power because the actual significance level may be much smaller than the given upper bound. Therefore a test based on some χ^2- or F-statistic would be preferable. Unfortunately, such tests are not easily available for the present situation. The problem is similar to the one discussed in Section 3.6.4 in the context of testing for multi-step causality. For more discussion of this point see also Lütkepohl (1990) and for a different approach of representing the uncertainty in estimated impulse responses see Sims & Zha (1999). ■

3.7.2 Proof of Proposition 3.6

The proof of Proposition 3.6 is a straightforward application of the matrix differentiation rules given in Appendix A.13. It is sketched here for completeness and because it is spread out over a number of publications in the literature. Readers mainly interested in applying the proposition may skip this section without loss of continuity.

To prove (3.7.5), note that $\Phi_i = J\mathbf{A}^i J'$ (see Chapter 2, Section 2.1.2) and apply Rule (8) of Appendix A.13. The expression for F_n in (3.7.6) follows because

$$\frac{\partial \operatorname{vec}(\Psi_n)}{\partial \boldsymbol{\alpha}'} = \sum_{i=1}^{n} \frac{\partial \operatorname{vec}(\Phi_i)}{\partial \boldsymbol{\alpha}'}$$

and

$$\begin{aligned} F_\infty &= \frac{\partial \operatorname{vec}(\Psi_\infty)}{\partial \boldsymbol{\alpha}'} = \frac{\partial \operatorname{vec}(\Psi_\infty)}{\partial \operatorname{vec}(\Psi_\infty^{-1})'} \frac{\partial \operatorname{vec}(\Psi_\infty^{-1})}{\partial \boldsymbol{\alpha}'} \\ &= -(\Psi_\infty' \otimes \Psi_\infty) \frac{\partial \operatorname{vec}(I_K - A_1 - \cdots - A_p)}{\partial \boldsymbol{\alpha}'}. \end{aligned}$$

Furthermore,

$$C_i = \frac{\partial \operatorname{vec}(\Theta_i)}{\partial \boldsymbol{\alpha}'} = \frac{\partial \operatorname{vec}(\Phi_i P)}{\partial \boldsymbol{\alpha}'} = (P' \otimes I_K) \frac{\partial \operatorname{vec}(\Phi_i)}{\partial \boldsymbol{\alpha}'}$$

and

$$\bar{C}_i = \frac{\partial \operatorname{vec}(\Theta_i)}{\partial \boldsymbol{\sigma}'} = (I_K \otimes \Phi_i)\frac{\partial \operatorname{vec}(P)}{\partial \boldsymbol{\sigma}'},$$

where

$$\frac{\partial \operatorname{vec}(P)}{\partial \boldsymbol{\sigma}'} = \mathbf{L}_K' \frac{\partial \operatorname{vech}(P)}{\partial \boldsymbol{\sigma}'} = H,$$

follows from Appendix A.13, Rule (10). The matrices $B_n, \bar{B}_n, B_\infty$, and $\bar{B}_\infty$ are obtained in a similar manner, using the relations $\Xi_n = \Psi_n P$ and $\Xi_\infty = \Psi_\infty P$.

Finally, in (3.7.11),

$$\begin{aligned} d_{jk,h} &= \frac{\partial \omega_{jk,h}}{\partial \boldsymbol{\alpha}'} \\ &= \left[2\sum_{i=0}^{h-1}(e_j'\Phi_i P e_k)(e_k'P' \otimes e_j')\frac{\partial \operatorname{vec}(\Phi_i)}{\partial \boldsymbol{\alpha}'}\operatorname{MSE}_j(h)\right. \\ &\qquad \left. -\sum_{i=0}^{h-1}(e_j'\Phi_i P e_k)^2\frac{\partial \operatorname{MSE}_j(h)}{\partial \boldsymbol{\alpha}'}\right] \Big/ \operatorname{MSE}_j(h)^2, \end{aligned}$$

$$\begin{aligned} \frac{\partial \operatorname{MSE}_j(h)}{\partial \boldsymbol{\alpha}'} &= \sum_{m=0}^{h-1}\left[(e_j'\Phi_m\Sigma_u \otimes e_j')\frac{\partial \operatorname{vec}(\Phi_m)}{\partial \boldsymbol{\alpha}'}\right. \\ &\qquad \left. +(e_j' \otimes e_j'\Phi_m\Sigma_u)\frac{\partial \operatorname{vec}(\Phi_m')}{\partial \boldsymbol{\alpha}'}\right] \\ &= \sum_{m=0}^{h-1}\left[(e_j'\Phi_m\Sigma_u \otimes e_j') + (e_j' \otimes e_j'\Phi_m\Sigma_u)\mathbf{K}_{KK}\right]\frac{\partial \operatorname{vec}(\Phi_m)}{\partial \boldsymbol{\alpha}'} \\ &= \sum_{m=0}^{h-1}\left[(e_j'\Phi_m\Sigma_u \otimes e_j') + \mathbf{K}_{11}(e_j'\Phi_m\Sigma_u \otimes e_j')\right]G_m \\ &= 2\sum_{m=0}^{h-1}(e_j'\Phi_m\Sigma_u \otimes e_j')G_m, \end{aligned}$$

(see Appendix A.12.2, Rule (23))

$$\begin{aligned} \bar{d}_{jk,h} &= \frac{\partial \omega_{jk,h}}{\partial \boldsymbol{\sigma}'} \\ &= \sum_{i=0}^{h-1}\left[2(e_j'\Phi_i P e_k)(e_k' \otimes e_j'\Phi_i)\frac{\partial \operatorname{vec}(P)}{\partial \boldsymbol{\sigma}'}\operatorname{MSE}_j(h)\right. \\ &\qquad \left. -(e_j'\Phi_i P e_k)^2\frac{\partial \operatorname{MSE}_j(h)}{\partial \boldsymbol{\sigma}'}\right] \Big/ \operatorname{MSE}_j(h)^2, \end{aligned}$$

and

$$\begin{aligned}\frac{\partial \mathrm{MSE}_j(h)}{\partial \boldsymbol{\sigma}'} &= \sum_{m=0}^{h-1} (e_j'\Phi_m \otimes e_j'\Phi_m) \frac{\partial \operatorname{vec}(\Sigma_u)}{\partial \boldsymbol{\sigma}'} \\ &= \sum_{m=0}^{h-1} (e_j'\Phi_m \otimes e_j'\Phi_m) \mathbf{D}_K \frac{\partial \operatorname{vech}(\Sigma_u)}{\partial \boldsymbol{\sigma}'}.\end{aligned}$$

Thereby Proposition 3.6 is proven. In the next section an example is discussed.

3.7.3 An Example

To illustrate the results of Section 3.7.1, we use again the investment/income/consumption example from Section 3.2.3. Because

$$\widehat{\Phi}_1 = \widehat{A}_1 = \begin{bmatrix} -.320 & .146 & .961 \\ .044 & -.153 & .289 \\ -.002 & .225 & -.264 \end{bmatrix},$$

the elements of $\widehat{\Phi}_1$ must have the same standard errors as the elements of $\widehat{A}_1$. Checking the covariance matrix in (3.7.5), it is seen that the asymptotic covariance matrix of $\widehat{\Phi}_1$ is indeed the upper left-hand $(K^2 \times K^2)$ block of $\Sigma_{\widehat{\boldsymbol{\alpha}}}$ because

$$G_1 = J \otimes I_K = [I_{K^2} : 0 : \cdots : 0].$$

Thus, the square roots of the diagonal elements of

$$G_1 \widehat{\Sigma}_{\widehat{\boldsymbol{\alpha}}} G_1'/T = \frac{1}{T}[I_9 : 0 : \cdots : 0](\widehat{\Gamma}_Y(0)^{-1} \otimes \widehat{\Sigma}_u) \begin{bmatrix} I_9 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

are estimates of the asymptotic standard errors of $\widehat{\Phi}_1$. Note that here and in the following we use the LS estimators from the standard form of the VAR process (see Section 3.2) and not the mean-adjusted form. Accordingly, the estimate $\widehat{\Gamma}_Y(0)^{-1}$ is obtained from $(ZZ'/T)^{-1}$ by deleting the first row and column.

From (2.1.22) we get

$$\widehat{\Phi}_2 = \widehat{\Phi}_1 \widehat{A}_1 + \widehat{A}_2 = \begin{bmatrix} -.054 & .262 & .416 \\ .029 & .114 & -.088 \\ .045 & .261 & .110 \end{bmatrix}.$$

To estimate the corresponding standard errors, we note that

$$G_2 = J\mathbf{A}' \otimes I_K + J \otimes \Phi_1.$$

Replacing the unknown quantities by the usual estimates gives

$$\begin{aligned}\frac{1}{T}\widehat{G}_2\widehat{\Sigma}_{\widehat{\boldsymbol{\alpha}}}\widehat{G}_2' &= \frac{1}{T}\big[J\widehat{\mathbf{A}}'\widehat{\Gamma}_Y(0)^{-1}\widehat{\mathbf{A}}J' \otimes \widehat{\Sigma}_u + J\widehat{\mathbf{A}}'\widehat{\Gamma}_Y(0)^{-1}J' \otimes \widehat{\Sigma}_u\widehat{\Phi}_1' \\ &\quad + J\widehat{\Gamma}_Y(0)^{-1}\widehat{\mathbf{A}}J' \otimes \widehat{\Phi}_1\widehat{\Sigma}_u + J\widehat{\Gamma}_Y(0)^{-1}J' \otimes \widehat{\Phi}_1\widehat{\Sigma}_u\widehat{\Phi}_1'\big].\end{aligned}$$

The square roots of the diagonal elements of this matrix are estimates of the standard deviations of the elements of $\widehat{\Phi}_2$ and so on. Some $\widehat{\Phi}_i$ matrices together with estimated standard errors are given in Table 3.3. In Figures 3.4 and 3.5, some impulse responses are depicted graphically along with two-standard error bounds.

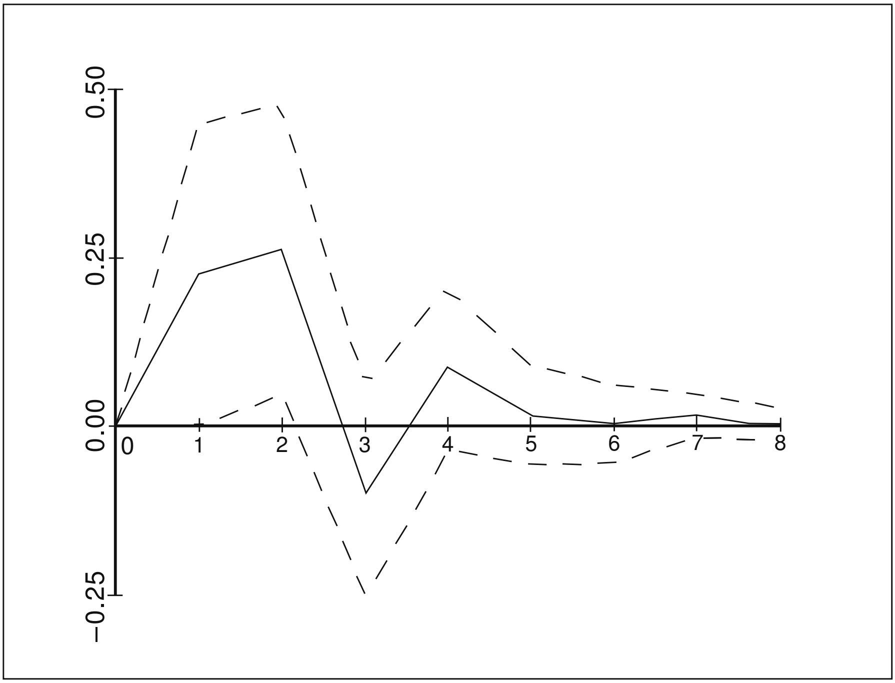

Fig. 3.4. Estimated responses of consumption to a forecast error impulse in income with estimated asymptotic two-standard error bounds.

In Figure 3.4, consumption is seen to increase in response to a unit shock in income. However, under a two-standard error criterion (approximate 95% confidence bounds) only the second response coefficient is significantly different from zero. Of course, the large standard errors of the impulse response coefficients reflect the substantial estimation uncertainty in the VAR coefficient matrices A_1 and A_2.

To check the overall significance of the response coefficients of consumption to an income impulse, we may use the procedure described in Remark 8 of

Table 3.3. Estimates of impulse responses for the investment/income/consumption system with estimated asymptotic standard errors in parentheses

i	$\widehat{\Phi}_i$	$\widehat{\Psi}_i$
1	$\begin{bmatrix} -0.320 & 0.146 & 0.961 \\ (0.125) & (0.562) & (0.657) \\ 0.044 & -0.153 & 0.289 \\ (0.032) & (0.143) & (0.167) \\ -0.002 & 0.225 & -0.264 \\ (0.025) & (0.115) & (0.134) \end{bmatrix}$	$\begin{bmatrix} 0.680 & 0.146 & 0.961 \\ (0.125) & (0.562) & (0.657) \\ 0.044 & 0.847 & 0.289 \\ (0.032) & (0.143) & (0.167) \\ -0.002 & 0.225 & 0.736 \\ (0.025) & (0.115) & (0.134) \end{bmatrix}$
2	$\begin{bmatrix} -0.054 & 0.262 & 0.416 \\ (0.129) & (0.546) & (0.663) \\ 0.029 & 0.114 & -0.088 \\ (0.032) & (0.135) & (0.162) \\ 0.045 & 0.261 & 0.110 \\ (0.026) & (0.108) & (0.131) \end{bmatrix}$	$\begin{bmatrix} 0.626 & 0.408 & 1.377 \\ (0.148) & (0.651) & (0.755) \\ 0.073 & 0.961 & 0.200 \\ (0.043) & (0.192) & (0.222) \\ 0.043 & 0.486 & 0.846 \\ (0.033) & (0.144) & (0.167) \end{bmatrix}$
3	$\begin{bmatrix} 0.119 & 0.353 & -0.408 \\ (0.084) & (0.384) & (0.476) \\ -0.009 & 0.071 & 0.120 \\ (0.016) & (0.078) & (0.094) \\ -0.001 & -0.098 & 0.091 \\ (0.017) & (0.078) & (0.102) \end{bmatrix}$	$\begin{bmatrix} 0.745 & 0.761 & 0.969 \\ (0.099) & (0.483) & (0.550) \\ 0.064 & 1.033 & 0.320 \\ (0.037) & (0.176) & (0.203) \\ 0.042 & 0.388 & 0.937 \\ (0.033) & (0.156) & (0.183) \end{bmatrix}$
∞	0	$\begin{bmatrix} 0.756 & 0.836 & 1.295 \\ (0.133) & (0.661) & (0.798) \\ 0.076 & 1.076 & 0.344 \\ (0.048) & (0.236) & (0.285) \\ 0.053 & 0.505 & 0.964 \\ (0.043) & (0.213) & (0.257) \end{bmatrix}$

Section 3.7.1. That is, we have to check the significance of the first $p(K-1) = 4$ response coefficients. Because one of them is individually significant at an asymptotic 5% level we may reject the null hypothesis of no response of consumption to income impulses at a significance level not greater than $4 \times 5\% = 20\%$. Of course, this is not a significance level we are used to in applied work. However, it becomes clear from Table 3.3 that the second response coefficient $\widehat{\phi}_{32,2}$ is still significant if the individual significance levels are reduced to 2.5%. Note that the upper 1.25 percentage point of the standard normal distribution is $c_{0.0125} = 2.24$. Thus, we may reject the no-response

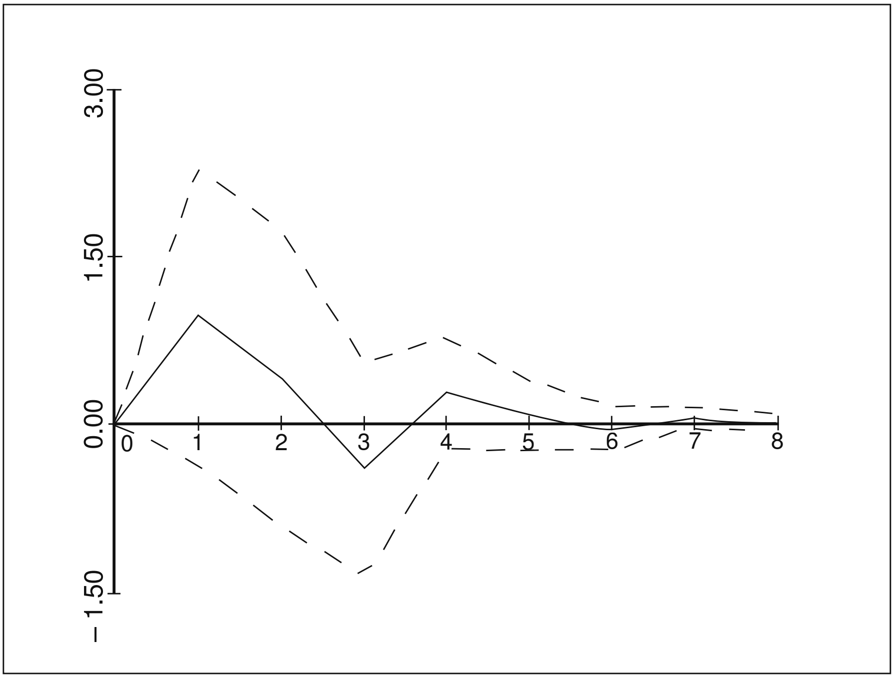

Fig. 3.5. Estimated responses of investment to a forecast error impulse in consumption with estimated asymptotic two-standard error bounds.

hypothesis at an overall $4 \times 2.5\% = 10\%$ level which is clearly a more common size for a test in applied work. Still, in this exercise, the data do not reveal strong evidence for the intuitively appealing hypothesis that consumption responds to income impulses. In later chapters, we will see how the coefficients can potentially be estimated with more precision.

In Figure 3.5, the responses of investment to consumption impulses are depicted. None of them is significant under a two-standard error criterion. This result is in line with the Granger-causality analysis in Section 3.6. In that section, we did not find evidence for Granger-causality from income/consumption to investment. Assuming that the test result describes the actual situation, the $\phi_{13,i}$ must be zero for $i = 1, 2, \ldots$ (see also Chapter 2, Section 2.3.1).

The covariance matrix of

$$\widehat{\Psi}_1 = I_3 + \widehat{\Phi}_1 = \begin{bmatrix} .680 & .146 & .961 \\ .044 & .847 & .289 \\ -.002 & .225 & .736 \end{bmatrix}$$

is, of course, the same as that of $\widehat{\Phi}_1$ and an estimate of the covariance matrix of the elements of

$$\widehat{\Psi}_2 = I_3 + \widehat{\Phi}_1 + \widehat{\Phi}_2 = \begin{bmatrix} .626 & .408 & 1.377 \\ .073 & .961 & .200 \\ .043 & .486 & .846 \end{bmatrix}$$

is obtained as $(G_1+\widehat{G}_2)\widehat{\Sigma}_{\widehat{\boldsymbol{\alpha}}}(G_1+\widehat{G}_2)'/T$. Some accumulated impulse responses together with estimated standard errors are also given in Table 3.3 and accumulated responses of consumption to income impulses and of investment to consumption impulses are shown in Figures 3.6 and 3.7, respectively. They reinforce the findings for the individual impulse responses in Figures 3.4 and 3.5.

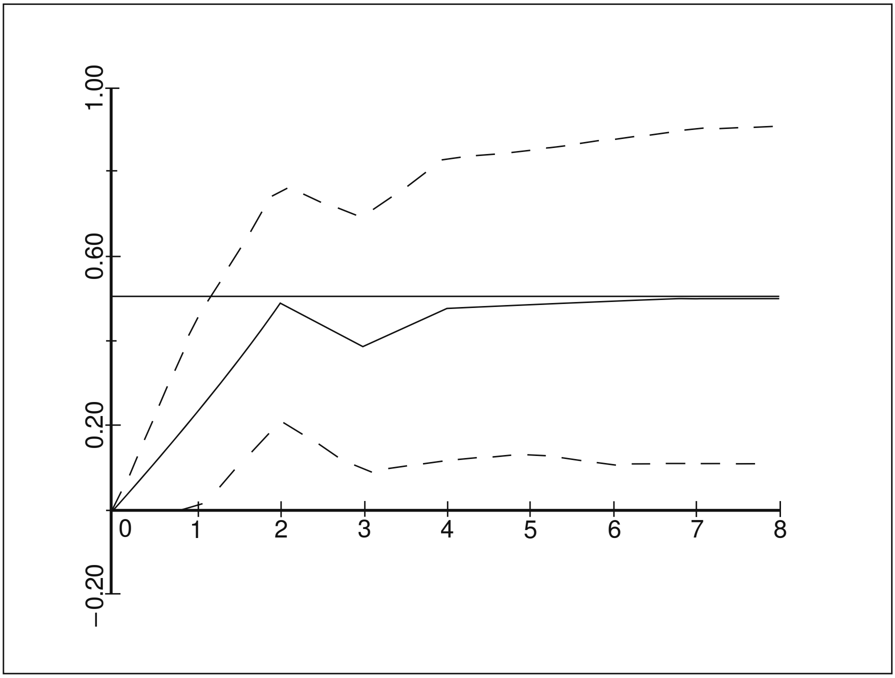

Fig. 3.6. Accumulated and long-run responses of consumption to a forecast error impulse in income with estimated asymptotic two-standard error bounds.

An estimate of the asymptotic covariance matrix of the estimated long-run responses $\widehat{\Psi}_\infty = (I_3 - \widehat{A}_1 - \widehat{A}_2)^{-1}$ is

$$\frac{1}{T}([\widehat{\Psi}'_\infty : \widehat{\Psi}'_\infty] \otimes \widehat{\Psi}_\infty)\widehat{\Sigma}_{\widehat{\boldsymbol{\alpha}}}\left(\begin{bmatrix} \widehat{\Psi}_\infty \\ \widehat{\Psi}_\infty \end{bmatrix} \otimes \widehat{\Psi}'_\infty\right).$$

The matrix $\widehat{\Psi}_\infty$ together with the resulting standard errors is also given in Table 3.3. For instance, the total long-run effect $\widehat{\psi}_{13,\infty}$ of a consumption impulse

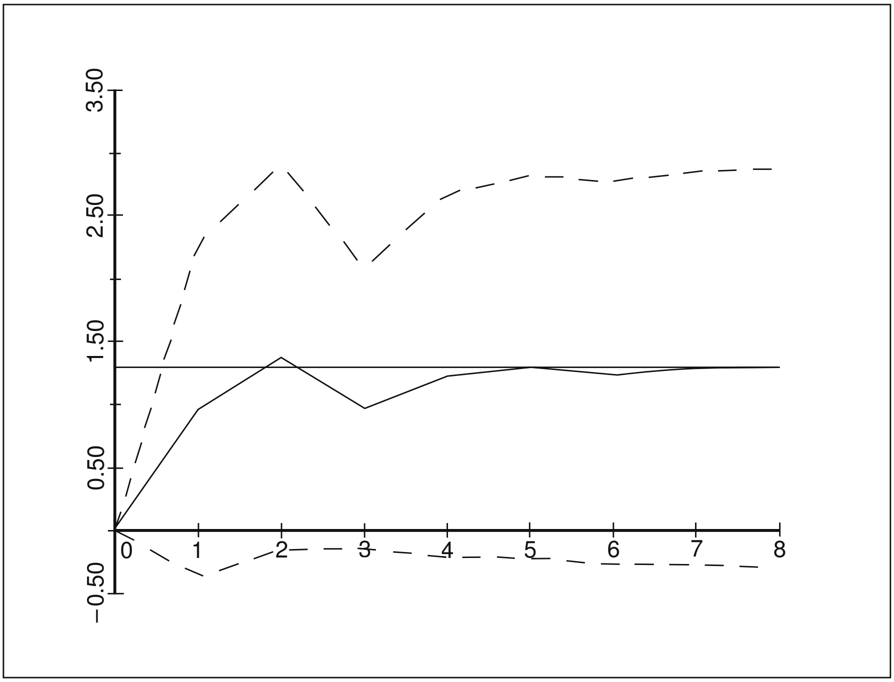

Fig. 3.7. Accumulated and long-run responses of investment to a forecast error impulse in consumption with estimated asymptotic two-standard error bounds.

on investment is 1.295 and its estimated asymptotic standard error is .798. Not surprisingly, $\widehat{\psi}_{13,\infty}$ is not significantly different from zero for any common level of significance (e.g., 10%). On the other hand, $\widehat{\psi}_{32,\infty}$, the long-run effect on consumption due to an impulse in income, is significant at an asymptotic 5% level.

For the interpretation of the $\widehat{\Phi}_i$'s, the critical remarks at the end of Chapter 2 must be kept in mind. As explained there, the $\widehat{\Phi}_i$ and $\widehat{\Psi}_n$ coefficients may not reflect the actual responses of the variables in the system. As an alternative, one may want to determine the responses to orthogonal residuals. In order to obtain the asymptotic covariance matrices of the $\widehat{\Theta}_i$ and $\widehat{\Xi}_i$, a decomposition of $\widehat{\Sigma}_u$ is needed. For our example,

$$\widehat{P} = \begin{bmatrix} 4.61 & 0 & 0 \\ .16 & 1.16 & 0 \\ .27 & .49 & .76 \end{bmatrix} \times 10^{-2}$$

is the lower triangular matrix with positive diagonal elements satisfying $\widehat{P}\widehat{P}' = \widehat{\Sigma}_u$ (Choleski decomposition). The asymptotic covariance matrix of $\text{vec}(\widehat{P}) = \text{vec}(\widehat{\Theta}_0)$ is a (9×9) matrix which is estimated as

$$\frac{1}{T}\widehat{\bar{C}}_0 \widehat{\Sigma}_{\widehat{\boldsymbol{\sigma}}} \widehat{\bar{C}}_0' = \frac{2}{T}\widehat{H}\mathbf{D}_K^+(\widehat{\Sigma}_u \otimes \widehat{\Sigma}_u)\mathbf{D}_K^{+\prime}\widehat{H}',$$

where, as usual, $\mathbf{D}_K^+ = (\mathbf{D}_K'\mathbf{D}_K)^{-1}\mathbf{D}_K'$ and

$$\widehat{H} = \mathbf{L}_3' \left\{ \mathbf{L}_3 \left[(I_3 \otimes \widehat{P})\mathbf{K}_{33} + (\widehat{P} \otimes I_3) \right] \mathbf{L}_3' \right\}^{-1}.$$

The resulting estimated asymptotic standard errors of the elements of $\widehat{P}$ are given in Table 3.4. Note that the variances corresponding to elements above the main diagonal of $\widehat{P}$ are all zero because these elements are zero by definition and are not estimated.

The asymptotic covariance matrix of the elements of

$$\widehat{\Theta}_1 = \begin{bmatrix} -1.196 & .644 & .730 \\ .256 & -.035 & .219 \\ -.047 & .131 & -.201 \end{bmatrix} \times 10^{-2}$$

is obtained as the sum of the two matrices

$$\widehat{C}_1 \widehat{\Sigma}_{\widehat{\boldsymbol{\alpha}}} \widehat{C}_1'/T = \left[(\widehat{P}' \otimes I_3) G_1 \widehat{\Sigma}_{\widehat{\boldsymbol{\alpha}}} G_1' (\widehat{P} \otimes I_3) \right] \Big/ T$$

and

$$\widehat{\bar{C}}_1 \widehat{\Sigma}_{\widehat{\boldsymbol{\sigma}}} \widehat{\bar{C}}_1'/T = (I_3 \otimes \widehat{\Phi}_1)\widehat{H}\widehat{\Sigma}_{\widehat{\boldsymbol{\sigma}}}\widehat{H}'(I_3 \otimes \widehat{\Phi}_1')/T.$$

The resulting standard errors for the elements of $\widehat{\Theta}_1$ are given in Table 3.4 along with some more $\widehat{\Theta}_i$ and $\widehat{\bar{\Xi}}_n$ matrices.

Some responses and accumulated responses of consumption to income innovations with two-standard error bounds are depicted in Figures 3.8 and 3.9. The responses in Figures 3.4 and 3.8 are obviously a bit different. Note the (significant) immediate reaction of consumption in Figure 3.8. However, from period 1 onwards the response of consumption in both figures is qualitatively similar. The difference of scales is due to the different sizes of the shocks traced through the system. For instance, Figure 3.4 is based on a unit shock in income while Figure 3.8 is based on an innovation of size one standard deviation due to the transformation of the white noise residuals.

Again, a test of overall significance of the impulse responses in Figure 3.8 could be performed using Bonferroni's principle. Now we have to check the significance of the $\widehat{\theta}_{32,i}$'s for $i = 0, 1, \ldots, 4 = p(K-1)$. We reject the null hypothesis of no response if at least one of the coefficients is significantly different from zero. In this case, we can reject at an asymptotic 5% level of significance because $\widehat{\theta}_{32,0}$ is significant at the 1% level (see Table 3.4). Thus, we may choose individual significance levels of 1% for each of the 5 coefficients and obtain 5% as an upper bound for the overall level. Of course, all these interpretations are based on the assumption that the actual asymptotic standard errors of the impulse responses are nonzero (see Section 3.7.1, Remark 1).

Table 3.4. Estimates of responses to orthogonal innovations for the investment/income/consumption system with estimated asymptotic standard errors in parentheses

i	$\widehat{\Theta}_i$	$\widehat{\Xi}_i$
0	$\begin{bmatrix} \underset{(.38)}{4.61} & 0 & 0 \\ \underset{(.14)}{.16} & \underset{(.10)}{1.16} & 0 \\ \underset{(.11)}{.27} & \underset{(.10)}{.49} & \underset{(.06)}{.76} \end{bmatrix} \times 10^{-2}$	$\begin{bmatrix} \underset{(.38)}{4.61} & 0 & 0 \\ \underset{(.14)}{.16} & \underset{(.10)}{1.16} & 0 \\ \underset{(.11)}{.27} & \underset{(.10)}{.49} & \underset{(.06)}{.76} \end{bmatrix} \times 10^{-2}$
1	$\begin{bmatrix} \underset{(.57)}{-1.20} & \underset{(.56)}{.64} & \underset{(.50)}{.73} \\ \underset{(.14)}{.26} & \underset{(.14)}{-.04} & \underset{(.13)}{.22} \\ \underset{(.12)}{-.05} & \underset{(.12)}{.13} & \underset{(.10)}{-.20} \end{bmatrix} \times 10^{-2}$	$\begin{bmatrix} \underset{(.63)}{3.46} & \underset{(.56)}{.64} & \underset{(.50)}{.73} \\ \underset{(.20)}{.41} & \underset{(.17)}{1.13} & \underset{(.13)}{.22} \\ \underset{(.15)}{.22} & \underset{(.14)}{.62} & \underset{(.11)}{.56} \end{bmatrix} \times 10^{-2}$
2	$\begin{bmatrix} \underset{(.58)}{-.10} & \underset{(.57)}{.51} & \underset{(.50)}{.32} \\ \underset{(.14)}{.13} & \underset{(.14)}{.09} & \underset{(.12)}{-.07} \\ \underset{(.12)}{.28} & \underset{(.12)}{.36} & \underset{(.10)}{.08} \end{bmatrix} \times 10^{-2}$	$\begin{bmatrix} \underset{(.74)}{3.32} & \underset{(.69)}{1.15} & \underset{(.58)}{1.05} \\ \underset{(.24)}{.54} & \underset{(.22)}{1.22} & \underset{(.17)}{.15} \\ \underset{(.20)}{.50} & \underset{(.18)}{.98} & \underset{(.14)}{.64} \end{bmatrix} \times 10^{-2}$
∞	0	$\begin{bmatrix} \underset{(.82)}{3.97} & \underset{(.92)}{1.61} & \underset{(.61)}{.98} \\ \underset{(.31)}{.61} & \underset{(.34)}{1.42} & \underset{(.22)}{.26} \\ \underset{(.28)}{.58} & \underset{(.32)}{1.06} & \underset{(.20)}{.73} \end{bmatrix} \times 10^{-2}$

We have also performed forecast error variance decompositions and we have computed the standard errors on the basis of the results given in Proposition 3.6. For some forecast horizons the decompositions are given in Table 3.5. The standard errors may be regarded as rough indications of the sampling uncertainty. It must be kept in mind, however, that they may be quite misleading if the true forecast error variance components are zero, as explained in Remark 6 of Section 3.7.1. Obviously, this qualification limits their value in

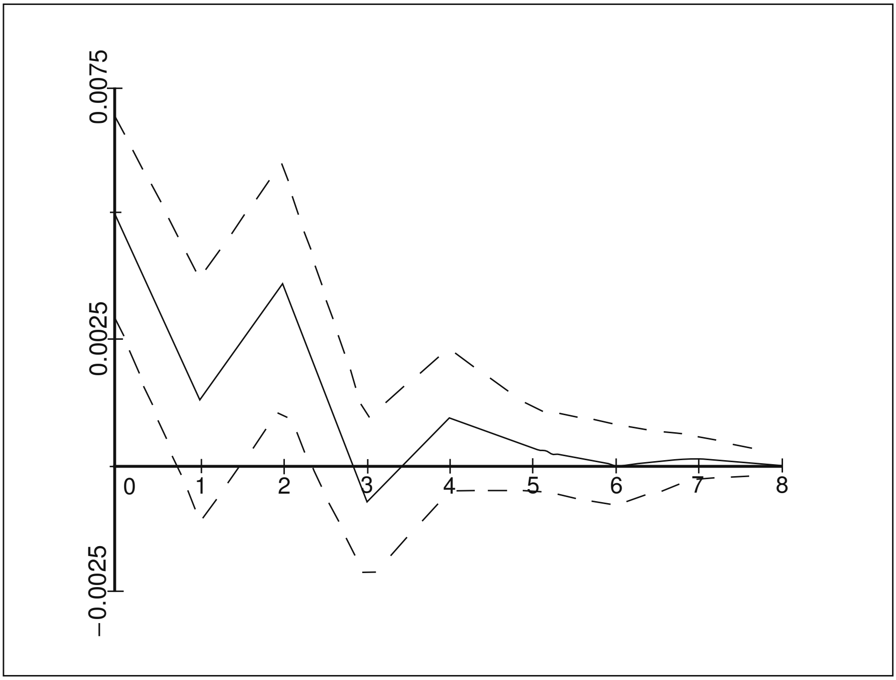

Fig. 3.8. Estimated responses of consumption to an orthogonalized impulse in income with estimated asymptotic two-standard error bounds.

the present example. Students are invited to reproduce the numbers in Table 3.5 and the previous tables of this section.

3.7.4 Investigating the Distributions of the Impulse Responses by Simulation Techniques

In the previous subsections, it was indicated repeatedly that in some cases the small sample validity of the asymptotic results is problematic. In that situation, one possibility is to use Monte Carlo or bootstrapping methods for investigating the sampling properties of the quantities of interest. Although these methods are quite expensive in terms of computer time, they were used in the past for evaluating the properties of impulse response functions (see, e.g., Runkle (1987) and Kilian (1998, 1999)). The general methodology is described in Appendix D.

In the present situation, there are different approaches to simulation. One possibility is to assume a specific distribution of the white noise process, e.g., $u_t \sim \mathcal{N}(0, \widehat{\Sigma}_u)$, and generate a large number of time series realizations based on the estimated VAR coefficients. From these time series, new sets of coefficients are then estimated and the corresponding impulse responses and/or

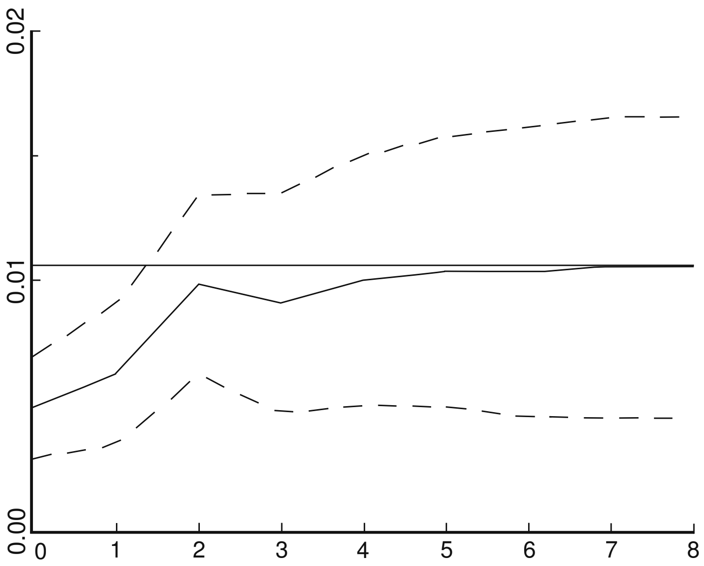

Fig. 3.9. Estimated accumulated and long-run responses of consumption to an orthogonalized impulse in income with estimated asymptotic two-standard error bounds.

forecast error variance components are computed. The empirical distributions obtained in this way may be used to investigate the actual distributions of the quantities of interest.

Alternatively, if an assumption regarding the white noise distribution cannot be made, bootstrap methods may be used and new sets of residuals may be drawn from the estimation residuals. A large number of y_t time series is generated on the basis of these sets of disturbances. The bootstrap multiple time series obtained in this way are then used to compute estimates of the quantities of interest and study their properties. Three different methods for computing bootstrap confidence intervals in the present context are described in Appendix D.3. We have used the standard and the Hall percentile methods to compute confidence intervals for the response of consumption to a forecast error impulse and an orthogonalized impulse in income for our example system. The results are shown in Figures 3.10 and 3.11, respectively.

Some interesting observations can be made. First, for the forecast error impulse responses, the two different methods for establishing confidence intervals produce quite similar results which are also at least qualitatively similar to the asymptotic confidence intervals in Figure 3.4. Second, the situation is a

Table 3.5. Forecast error variance decomposition of the investment/income/consumption system with estimated asymptotic standard errors in parentheses

forecast error in	forecast horizon h	proportions of forecast error variance, h periods ahead, accounted for by innovations in investment $\widehat{\omega}_{j1,h}$	income $\widehat{\omega}_{j2,h}$	consumption $\widehat{\omega}_{j3,h}$
investment	1	1.00(.00)	.00(.00)	.00(.00)
($j = 1$)	2	.96(.04)	.02(.03)	.02(.03)
	3	.95(.04)	.03(.03)	.03(.03)
	4	.94(.05)	.03(.03)	.03(.03)
	8	.94(.05)	.03(.03)	.03(.04)
income	1	.02(.03)	.98(.03)	.00(.00)
($j = 2$)	2	.06(.05)	.91(.06)	.03(.04)
	3	.07(.06)	.90(.07)	.03(.04)
	4	.07(.06)	.89(.07)	.04(.04)
	8	.07(.06)	.89(.07)	.04(.04)
consumption	1	.08(.06)	.27(.09)	.65(.09)
($j = 3$)	2	.08(.06)	.27(.08)	.65(.09)
	3	.13(.08)	.33(.09)	.54(.09)
	4	.13(.08)	.34(.09)	.54(.09)
	8	.13(.08)	.34(.09)	.53(.09)

bit different for the orthogonalized impulse responses in Figure 3.11. Here the two different bootstrap methods produce rather different confidence intervals. These intervals are quite asymmetric in the sense that the estimated impulse responses are not in the middle between the lower and upper bound of the intervals. Thereby they also look quite differently from the asymptotic intervals shown in Figure 3.8. The latter intervals are symmetric around the estimated impulse response coefficients by construction. Again, the qualitative interpretation does not change, however. In other words, the instantaneous and the second coefficient are significantly different from zero, as before. Moreover, the confidence intervals in Figure 3.11 are consistent with a rapidly declining effect of an impulse in income.

It must be emphasized, however, that the bootstrap generally does not solve the problem of a singular asymptotic distribution of the impulse responses and the resulting potentially invalid inference. If the asymptotic distribution is singular, the bootstrap may fail to produce meaningful confidence intervals, for example. Again it may be worth considering a univariate AR(1) process $y_t = \alpha y_{t-1} + u_t$ for illustrative purposes. The second forecast error impulse response coefficient is $\Phi_2 = \alpha^2$. The corresponding estimator $\widehat{\Phi}_2 = \widehat{\alpha}^2$ was found to have a singular asymptotic distribution if $\alpha = 0$ (see Remark 1 in Section 3.7.1). Suppose a bootstrap is used to produce N bootstrap estimates of α, $\widehat{\alpha}^*_{(n)}$, $n = 1, \dots, N$. Clearly, the corresponding bootstrap estimates

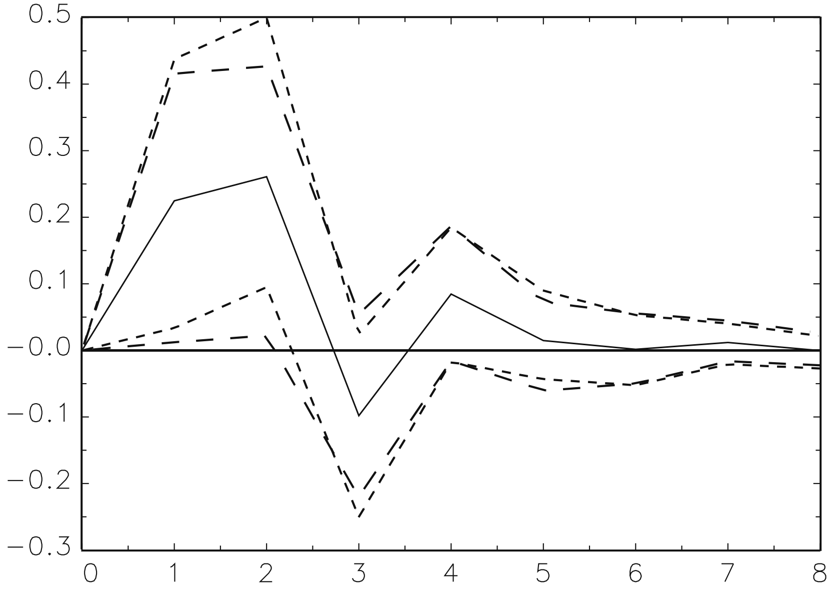

Fig. 3.10. Estimated responses (——) of consumption to a forecast error impulse in income with 95% bootstrap confidence bounds based on 2000 bootstrap replications (— — standard intervals, - - - Hall's percentile intervals).

$\widehat{\Phi}^*_{2(n)} = \widehat{\alpha}^{*2}_{(n)}$ will all be positive with probability one because they are squares. Thus, if the standard $(1-\gamma)100\%$ bootstrap confidence interval is constructed in the usual way by choosing $\widehat{\Phi}^*_{2(N\gamma/2)}$ and $\widehat{\Phi}^*_{2(N(1-\gamma)/2)}$ as lower and upper bound, respectively, the true value of zero will never be within the confidence interval. Hence, in this case the actual confidence level will be zero. Although the Hall confidence intervals may be a bit better in this case, they will also not provide the desired coverage level even in large samples. A more detailed discussion of this problem is given by Benkwitz et al. (2000), where also methods for correct asymptotic inference are considered. One possible solution is to eliminate all points where nonsingularities of the asymptotic distribution may occur by fitting subset models (see Chapter 5). Another possibility to circumvent the problem is to allow the VAR process to be of infinite order and increase the order with growing sample size. This possibility will be discussed in detail in Chapter 15.

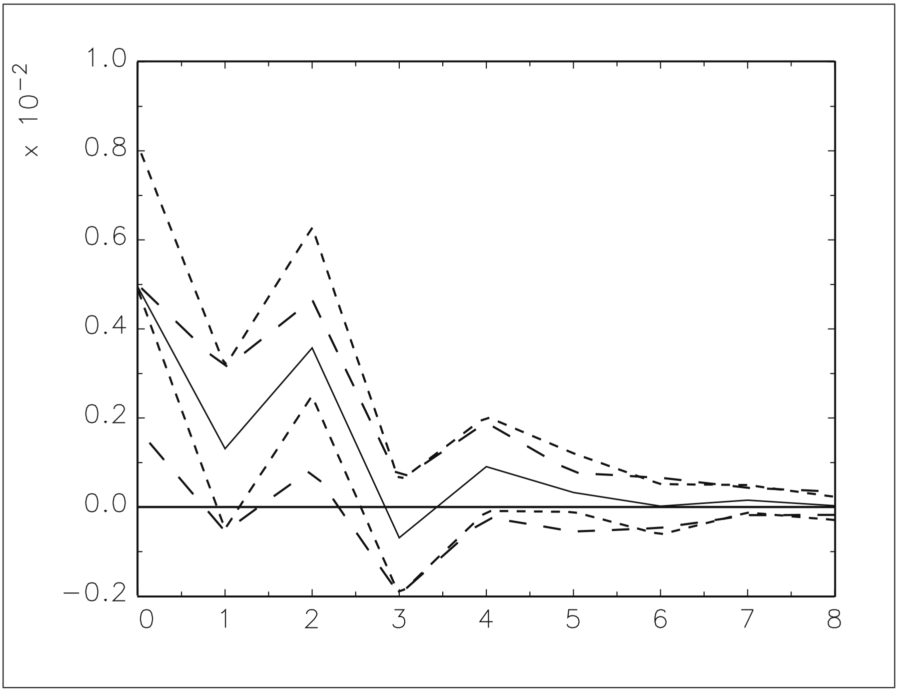

Fig. 3.11. Estimated responses (——) of consumption to an orthogonalized impulse in income with 95% bootstrap confidence bounds based on 2000 bootstrap replications (— — standard intervals, - - - Hall's percentile intervals).

3.8 Exercises

3.8.1 Algebraic Problems

The notation of Sections 3.2–3.5 is used in the following problems.

Problem 3.1
Show that $\widehat{\boldsymbol{\beta}} = ((ZZ')^{-1}Z \otimes I_K)\mathbf{y}$ minimizes

$$\bar{S}(\boldsymbol{\beta}) = \mathbf{u}'\mathbf{u} = [\mathbf{y} - (Z' \otimes I_K)\boldsymbol{\beta}]'[\mathbf{y} - (Z' \otimes I_K)\boldsymbol{\beta}].$$

Problem 3.2
Prove that

$$\sqrt{T}(\widehat{\mathbf{b}} - \mathbf{b}) \xrightarrow{d} \mathcal{N}(0, \Sigma_u \otimes \Gamma^{-1}),$$

if y_t is stable and

$$\frac{1}{\sqrt{T}} \operatorname{vec}(ZU') = \frac{1}{\sqrt{T}}(I_K \otimes Z)\operatorname{vec}(U') \xrightarrow{d} \mathcal{N}(0, \Sigma_u \otimes \Gamma).$$

Problem 3.3
Show (3.4.17). (Hint: Use the product rule for matrix differentiation and $\partial \operatorname{vec}(\Sigma_u^{-1})/\partial \operatorname{vec}(\Sigma_u)' = -\Sigma_u^{-1} \otimes \Sigma_u^{-1}$.)

Problem 3.4
Derive (3.4.18). (Hint: Use the last expression given in (3.4.6).)

Problem 3.5
Show (3.4.19).

Problem 3.6
Derive (3.4.20).

Problem 3.7
Prove that plim $\widetilde{z}_T/\sqrt{T} = 0$, where

$$\widetilde{z}_T = \sum_{i=1}^{p} \widetilde{A}_i \sum_{j=0}^{i-1} (y_{-j} - y_{T-j}).$$

(Hint: Show that $E(\widetilde{z}_T/\sqrt{T}) \to 0$ and $\operatorname{Var}(\widetilde{z}_T/\sqrt{T}) \to 0$.)

Problem 3.8
Show that Equation (3.5.10) holds.
(Hint: Define

$$Z_t(h) := \begin{bmatrix} 1 \\ y_t(h) \\ \vdots \\ y_t(h-p+1) \end{bmatrix}$$

and show $Z_t(h) = \mathbf{B} Z_t(h-1)$ by induction.)

Problem 3.9
In the context of Section 3.5, suppose that y_t is a stable Gaussian VAR(p) process which is estimated by ML in mean-adjusted form. Show that the forecast MSE correction term has the form

$$\Omega(h) = E\left(\frac{\partial y_t(h)}{\partial \mu'} \Sigma_{\widetilde{\mu}} \frac{\partial y_t(h)'}{\partial \mu}\right) + E\left(\frac{\partial y_t(h)}{\partial \boldsymbol{\alpha}'} \Sigma_{\widetilde{\boldsymbol{\alpha}}} \frac{\partial y_t(h)'}{\partial \boldsymbol{\alpha}}\right),$$

with

$$\frac{\partial y_t(h)}{\partial \mu'} = I_K - J\mathbf{A}^h \underset{(Kp\times K)}{\begin{bmatrix} I_K \\ \vdots \\ I_K \end{bmatrix}}$$

and

$$\frac{\partial y_t(h)}{\partial \boldsymbol{\alpha}'} = \sum_{i=0}^{h-1} (Y_t - \boldsymbol{\mu})'(\mathbf{A}')^{h-1-i} \otimes \Phi_i.$$

Here $\boldsymbol{\mu} := (\mu', \dots, \mu')'$ is a $(Kp \times 1)$ vector, Y_t and $\mathbf{A}$ are as defined in (2.1.8), $J := [I_K : 0 : \cdots : 0]$ is a $(K \times Kp)$ matrix, and Φ_i is the i-th coefficient matrix of the prediction error MA representation (2.1.17).

Problem 3.10
Derive the ML estimator and its asymptotic distribution for the parameter of a stable AR(1) process, $y_t = \alpha y_{t-1} + u_t$, $u_t \sim$ i.i.d. $\mathcal{N}(0, \sigma_u^2)$.

3.8.2 Numerical Problems

The following problems require the use of a computer. They are based on the two quarterly, seasonally adjusted U.S. investment series given in File E2. Consider the variables

y_1 – first differences of fixed investment,
y_2 – first differences of change in business inventories,

in the following problems. Use the data from 1947 to 1968 only.

Problem 3.11
Plot the two time series y_{1t} and y_{2t} and comment on the stationarity and stability of the series.

Problem 3.12
Estimate the parameters of a VAR(1) model for $(y_{1t}, y_{2t})'$ using multivariate LS, that is, compute $\widehat{B}$ and $\widehat{\Sigma}_u$. Comment on the stability of the estimated process.

Problem 3.13
Use the mean-adjusted form of a VAR(1) model and estimate the coefficients. Assume that the data generation process is Gaussian and estimate the covariance matrix of the asymptotic distribution of the ML estimators.

Problem 3.14
Determine the Yule-Walker estimate of the VAR(1) coefficient matrix and compare it to the LS estimate.

Problem 3.15
Use the LS estimate and compute point forecasts $\widehat{y}_{86}(1)$, $\widehat{y}_{86}(2)$ (that is, the forecast origin is the last quarter of 1968) and the corresponding MSE matrices $\widehat{\Sigma}_y(1), \widehat{\Sigma}_y(2), \widehat{\Sigma}_{\widehat{y}}(1)$, and $\widehat{\Sigma}_{\widehat{y}}(2)$. Use these estimates to set up approximate 95% interval forecasts assuming that the process y_t is Gaussian.

Problem 3.16
Test the hypothesis that y_2 does not Granger-cause y_1.

Problem 3.17
Estimate the coefficient matrices Φ_1 and Φ_2 from the LS estimates of the VAR(1) model for y_t and determine approximate standard errors of the estimates.

Problem 3.18
Determine the *upper* triangular matrix $\widehat{P}$ with positive diagonal for which $\widehat{P}\widehat{P}' = \widehat{\Sigma}_u$. Estimate the covariance matrix of the asymptotic distribution of $\widehat{P}$ under the assumption that y_t is Gaussian. Test the hypothesis that the upper right-hand corner element of the underlying matrix P is zero.

Problem 3.19
Use the results of the previous problems to compute $\widehat{\Theta}_0$, $\widehat{\Theta}_1$, and $\widehat{\Theta}_2$. Determine also estimates of the asymptotic standard errors of the elements of these three matrices.