Noncausal Vector Autoregression

Lanne, Markku and Saikkonen, Pentti

University of Helsinki, HECER

April 2010

Helsinki Center of Economic Research

Discussion Papers

Noncausal Vector Autoregression

Markku Lanne
University of Helsinki and HECER

and

Pentti Saikkonen
University of Helsinki and HECER

Discussion Paper No. 293
April 2010

ISSN 1795-0562

HECER – Helsinki Center of Economic Research, P.O. Box 17 (Arkadiankatu 7), FI-00014 University of Helsinki, FINLAND, Tel +358-9-191-28780, Fax +358-9-191-28781,
E-mail info-hecer@helsinki.fi, Internet www.hecer.fi

HECER
Discussion Paper No. 293

Noncausal Vector Autoregression*

Abstract

In this paper, we propose a new noncausal vector autoregressive (VAR) model for non-Gaussian time series. The assumption of non-Gaussianity is needed for reasons of identifiability. Assuming that the error distribution belongs to a fairly general class of elliptical distributions, we develop an asymptotic theory of maximum likelihood estimation and statistical inference. We argue that allowing for noncausality is of particular importance in economic applications which currently use only conventional causal VAR models. Indeed, if noncausality is incorrectly ignored, the use of a causal VAR model may yield suboptimal forecasts and misleading economic interpretations. Therefore, we propose a procedure for discriminating between causality and noncausality. The methods are illustrated with an application to interest rate data.

JEL Classification: C32, C52, E43

Keywords: Vector autoregression, Noncausal time series, Non-Gaussian time series.

Markku Lanne

Department of Political and Economic Studies
University of Helsinki
P.O. Box 17 (Arkadiankatu 7)
FI-00014 University of Helsinki
FINLAND

e-mail: markku.lanne@helsinki.fi

Pentti Saikkonen

Department of Mathematics and Statistics
University of Helsinki
P.O. Box 68 (Gustaf Hällströmn katu 2b)
FI-00014 University of Helsinki
FINLAND

e-mail: pentti.saikkonen@helsinki.fi

* We thank Martin Ellison, Juha Kilponen, and Antti Ripatti for useful comments on an earlier version of this paper. Financial support from the Academy of Finland and the OP-Pohjola Group Research Foundation is gratefully acknowledged. The paper was written while the second author worked at the Bank of Finland whose hospitality is gratefully acknowledged.

1 Introduction

The vector autoregressive (VAR) model is widely applied in various fields of application to summarize the joint dynamics of a number of time series and to obtain forecasts. Especially in economics and finance the model is also employed in structural analyses, and it often provides a suitable framework for conducting tests of theoretical interest. Typically, the error term of a VAR model is interpreted as a forecast error that should be an independent white noise process for the model to capture all relevant dynamic dependencies. Hence, the model is deemed adequate if its errors are not serially correlated. However, unless the errors are Gaussian, this is not sufficient to guarantee independence and, even in the absence of serial correlation, it may be possible to predict the error term by lagged values of the considered variables. This is a relevant point because diagnostic checks in empirical analyses often suggest non-Gaussian residuals and the use of a Gaussian likelihood has been justified by properties of quasi maximum likelihood (ML) estimation. A further point is that, to the best of our knowledge, only causal VAR models have previously been considered although noncausal autoregressions, which explicitly allow for the aforementioned predictability of the error term, might provide a correct VAR specification (for noncausal (univariate) autoregressions, see, e.g., Brockwell and Davis (1987, Chapter 3) or Rosenblatt (2000)). These two issues are actually connected as distinguishing between causality and noncausality is not possible under Gaussianity. Hence, in order to assess the nature of causality, allowance must be made for deviations from Gaussianity when they are backed up by the data. If noncausality indeed is present, confining to (misspecified) causal VAR models may lead to suboptimal forecasts and false conclusions.

The statistical literature on noncausal univariate time series models is relatively small, and, to our knowledge, noncausal VAR models have not been considered at all prior to this study (for available work on noncausal autoregressions and their applications, see Rosenblatt (2000), Andrews, Davis, and Breidt (2006), Lanne and Saikkonen (2008), and the references therein). In this paper, the previous statistical theory of univariate noncausal autoregressive models is extended to the vector case. Our formulation of the noncausal VAR model is a direct extension of that used by Lanne and Saikkonen (2008) in the univariate case. To obtain a feasible approximation for the non-Gaussian likelihood function, the distribution of the error term is assumed to belong to a fairly general class of elliptical distributions. Using this assumption, we can show the consistency and asymptotic normality of an approximate (local) ML estimator, and justify the applicability of usual likelihood based tests.

As already indicated, the noncausal VAR model can be used to check the validity of statistical analyses based on a causal VAR model. This is important, for instance, in economic applications where VAR models are commonly applied to test for economic theories. Typically such tests assume the existence of a causal VAR representation whose errors are not predictable by lagged values of the considered time series. If this is not the case, the employed tests based on a causal VAR model are not valid and the resulting conclusions may be misleading. We provide an illustration of this with interest rate data.

The remainder of the paper is structured as follows. Section 2 introduces the noncausal VAR model. Section 3 derives an approximation for the likelihood function and properties of the related approximate ML estimator. Section 4 provides our empirical illustration. Section 5 concludes. An appendix contains proofs and some technical derivations.

The following notation is used throughout. The expectation operator and the covariance operator are denoted by $\mathbb{E}(\cdot)$ and $\mathbb{C}(\cdot)$ or $\mathbb{C}(\cdot,\cdot)$, respectively, whereas $x \overset{d}{=} y$ means that the random quantities x and y have the same distribution. By $\text{vec}(A)$ we denote a column vector obtained by stacking the columns of the matrix A one below another. If A is a square matrix then $\text{vech}(A)$ is a column vector obtained by stacking the columns of A from the principal diagonal downwards (including elements on the diagonal). The usual notation $A \otimes B$ is used for the Kronecker product of the matrices A and B. The $mn \times mn$ commutation matrix and the $n^2 \times n(n+1)/2$ duplication matrix are denoted by K_{mn} and D_n, respectively. Both of them are of full column rank. The former is defined by the relation $K_{mn}\text{vec}(A) = \text{vec}(A')$, where A is any $m \times n$ matrix, and the latter by the relation $\text{vec}(B) = D_n\text{vech}(B)$, where B is any symmetric $n \times n$ matrix.

2 Model

2.1 Definition and basic properties

Consider the n-dimensional stochastic process y_t $(t = 0, \pm 1, \pm 2, ...)$ generated by

$$\Pi(B)\,\Phi\left(B^{-1}\right) y_t = \epsilon_t, \tag{1}$$

where $\Pi(B) = I_n - \Pi_1 B - \cdots - \Pi_r B^r$ $(n \times n)$ and $\Phi\left(B^{-1}\right) = I_n - \Phi_1 B^{-1} - \cdots - \Phi_s B^{-s}$ $(n \times n)$ are matrix polynomials in the backward shift operator B, and ϵ_t $(n \times 1)$ is a sequence of independent, identically distributed (continuous) random vectors with zero mean and finite positive definite covariance matrix. Moreover, the matrix polynomials $\Pi(z)$ and $\Phi(z)$ $(z \in \mathbb{C})$ have their zeros outside the unit disc so that

$$\det \Pi(z) \neq 0, \quad |z| \leq 1, \quad \text{and} \quad \det \Phi(z) \neq 0, \quad |z| \leq 1. \tag{2}$$

If $\Phi_j \neq 0$ for some $j \in \{1, .., s\}$, equation (1) defines a noncausal vector autoregression referred to as purely noncausal when $\Pi_1 = \cdots = \Pi_r = 0$. The corresponding conventional causal model is obtained when $\Phi_1 = \cdots = \Phi_s = 0$. Then the former condition in (2) guarantees the stationarity of the model. In the general set up of equation (1) the same is true for the process

$$u_t = \Phi\left(B^{-1}\right) y_t.$$

Specifically, there exists a $\delta_1 > 0$ such that $\Pi(z)^{-1}$ has a well defined power series representation $\Pi(z)^{-1} = \sum_{j=0}^{\infty} M_j z^j = M(z)$ for $|z| < 1 + \delta_1$. Consequently, the process u_t has the causal moving average representation

$$u_t = M(B)\,\epsilon_t = \sum_{j=0}^{\infty} M_j \epsilon_{t-j}. \tag{3}$$

Notice that $M_0 = I_n$ and that the coefficient matrices M_j decay to zero at a geometric rate as $j \to \infty$. When convenient, $M_j = 0$, $j < 0$, will be assumed.

Write $\Pi(z)^{-1} = (\det \Pi(z))^{-1}\,\Xi(z) = M(z)$, where $\Xi(z)$ is the adjoint polynomial matrix of $\Pi(z)$ with degree at most $(n-1)r$. Then, $\det \Pi(B)\, u_t = \Xi(B)\,\epsilon_t$ and, by the definition of u_t,

$$\Phi\left(B^{-1}\right) w_t = \Xi(B)\,\epsilon_t,$$

where $w_t = (\det \Pi(B))y_t$. By the latter condition in (2) one can find a $0 < \delta_2 < 1$ such that $\Phi(z^{-1})^{-1} \Xi(z)$ has a well defined power series representation $\Phi(z^{-1})^{-1} \Xi(z) = \sum_{j=-(n-1)r}^{\infty} N_j z^{-j} = N(z^{-1})$ for $|z| > 1 - \delta_2$. Thus, the process w_t has the representation

$$w_t = \sum_{j=-(n-1)r}^{\infty} N_j \epsilon_{t+j}, \tag{4}$$

where the coefficient matrices N_j decay to zero at a geometric rate as $j \to \infty$.

From (2) it follows that the process y_t itself has the representation

$$y_t = \sum_{j=-\infty}^{\infty} \Psi_j \epsilon_{t-j}, \tag{5}$$

where Ψ_j $(n \times n)$ is the coefficient matrix of z^j in the Laurent series expansion of $\Psi(z) \overset{def}{=} \Phi(z^{-1})^{-1} \Pi(z)^{-1}$ which exists for $1 - \delta_2 < |z| < 1 + \delta_1$ with Ψ_j decaying to zero at a geometric rate as $|j| \to \infty$. The representation (5) implies that y_t is a stationary and ergodic process with finite second moments. We use the abbreviation VAR(r, s) for the model defined by (1). In the causal case $s = 0$, the conventional abbreviation VAR(r) is also used.

Denote by $\mathbb{E}_t(\cdot)$ the conditional expectation operator with respect to the information set $\{y_t, y_{t-1}, ...\}$ and conclude from (1) and (5) that

$$y_t = \sum_{j=-\infty}^{s-1} \Psi_j \mathbb{E}_t(\epsilon_{t-j}) + \sum_{j=s}^{\infty} \Psi_j \epsilon_{t-j}.$$

In the conventional causal case, $s = 0$ and $\mathbb{E}_t(\epsilon_{t-j}) = 0$, $j \leq -1$, so that the right hand side reduces to the moving average representation (3). However, in the noncausal case this does not happen. Then $\Psi_j \neq 0$ for some $j < 0$, which in conjunction with the representation (5) shows that y_t and ϵ_{t-j} are correlated. Consequently, $\mathbb{E}_t(\epsilon_{t-j}) \neq 0$ for some $j < 0$, implying that future errors can be predicted by past values of the process y_t. A possible interpretation of this predictability is that the errors contain factors which are not included in the model and can be predicted by the time series selected in the model. This seems quite plausible, for instance, in economic applications where time series are typically interrelated and only a few time series out of a larger selection are used in the analysis. The reason why some variables are excluded may be that data are not available or the underlying economic model only contains the variables for which hypotheses of interest are formulated.

A practical complication with noncausal autoregressive models is that they cannot be identified by second order properties or Gaussian likelihood. In the univariate case this is explained, for example, in Brockwell and Davis (1987, p. 124-125)). To demonstrate the same in the multivariate case described above, note first that, by well-known results on linear filters (cf. Hannan (1970, p. 67)), the spectral density matrix of the process y_t defined by (1) is given by

$$\begin{aligned} & (2\pi)^{-1} \Phi\left(e^{-i\omega}\right)^{-1} \Pi\left(e^{i\omega}\right)^{-1} \mathbb{C}\left(\epsilon_t\right) \Pi\left(e^{-i\omega}\right)'^{-1} \Phi\left(e^{i\omega}\right)'^{-1} \\ = \; & (2\pi)^{-1} \left[\Phi\left(e^{i\omega}\right)' \Pi\left(e^{-i\omega}\right)' \mathbb{C}\left(\epsilon_t\right)^{-1} \Pi\left(e^{i\omega}\right) \Phi\left(e^{-i\omega}\right)\right]^{-1}. \end{aligned}$$

In the latter expression, the matrix in the brackets is 2π times the spectral density matrix of a second order stationary process whose autocovariances are zero at lags larger than $r+s$. As is well known, this process can be represented as an invertible moving average of order $r+s$. Specifically, by a slight modification of Theorem 10' of Hannan (1970), we get the unique representation

$$\Phi\left(e^{i\omega}\right)' \Pi\left(e^{-i\omega}\right)' \mathbb{C}\left(\epsilon_t\right)^{-1} \Pi\left(e^{i\omega}\right) \Phi\left(e^{-i\omega}\right) = \left(\sum_{j=0}^{r+s} \mathcal{C}_j e^{-i\omega}\right)' \left(\sum_{j=0}^{r+s} \mathcal{C}_j e^{i\omega}\right),$$

where the $n \times n$ matrixes $\mathcal{C}_0, ..., \mathcal{C}_{r+s}$ are real with $\mathcal{C}_0$ positive definite, and the zeros of $\det\left(\sum_{j=0}^{r+s} \mathcal{C}_j e^{i\omega}\right)$ lie outside the unique disc.[1] Thus, the spectral density matrix of y_t has the representation $(2\pi)^{-1}\left(\sum_{j=0}^{r+s} \mathcal{C}_j e^{ij\omega}\right)^{-1}\left(\sum_{j=0}^{r+s} \mathcal{C}_j e^{-ij\omega}\right)'^{-1}$, which is the spectral density matrix of a causal VAR($r+s$) process.

The preceding discussion means that, even if y_t is noncausal, its spectral density and, hence, autocovariance function cannot be distinguished from those of a causal VAR($r+s$) process. If y_t or, equivalently, the error term ϵ_t is Gaussian this means that causal and noncausal representations of (1) are statistically indistinguishable and nothing is lost by using a conventional causal representation. However, if the errors are non-Gaussian using

[1] A direct application of Hannan's (1970) Theorem 10' would give a representation with ω replaced by $-\omega$. That this modification is possible can be seen from the proof of the mentioned theorem (see the discussion starting in the middle of p. 64 of Hannan (1970)).

a causal representation of a true noncausal process means using a VAR model whose errors can be predicted by past values of the considered series and potentially better fit and forecasts could be obtained by using the correctly specified noncausal model.

2.2 Assumptions

In this section, we introduce assumptions that enable us to derive the likelihood function and its derivatives. Further assumptions, needed for the asymptotic analysis of the ML estimator and related tests, will be introduced in subsequent sections.

As already discussed, meaningful application of the noncausal VAR model requires that the distribution of ϵ_t is non-Gaussian. In the following assumption the distribution of ϵ_t is restricted to a general elliptical form. As is well known, the normal distribution belongs to the class of elliptical distributions but we will not rule out it at this point. Other examples of elliptical distributions are discussed in Fang, Kotz, and Ng (1990, Chapter 3). Perhaps the best known non-Gaussian example is the multivariate t-distribution.

Assumption 1. The error process ϵ_t in (1) is independent and identically distributed with zero mean, finite and positive definite covariance matrix, and an elliptical distribution possessing a density.

Results on elliptical distributions needed in our subsequent developments can be found in Fang et al. (1990, Chapter 2) on which the following discussion is based. To simplify notation in subsequent derivations, we define $\varepsilon_t = \Sigma^{-1/2}\epsilon_t$ where Σ $(n \times n)$ is a positive definite parameter matrix. By Assumption 1, we have the representations

$$\epsilon_t \overset{d}{=} \rho_t \Sigma^{1/2} \upsilon_t \quad \text{and} \quad \varepsilon_t \overset{d}{=} \rho_t \upsilon_t, \tag{6}$$

where (ρ_t, υ_t) is an independent and identically distributed sequence such that ρ_t (scalar) and υ_t $(n \times 1)$ are independent, ρ_t is nonnegative, and υ_t is uniformly distributed on the unit ball (and hence $\upsilon_t'\upsilon_t = 1$).

The density of ϵ_t is of the form

$$f_\Sigma(x; \lambda) = \frac{1}{\sqrt{\det(\Sigma)}} f\left(x'\Sigma^{-1}x; \lambda\right) \tag{7}$$

for some nonnegative function $f(\cdot;\lambda)$ of a scalar variable. In addition to the positive definite parameter matrix Σ the distribution of ϵ_t is allowed to depend on the parameter vector λ (d $\times$ 1). The parameter matrix Σ is closely related to the covariance matrix of ϵ_t. Specifically, because $\mathbb{E}(\upsilon_t) = 0$ and $\mathbb{C}(\upsilon_t) = n^{-1}I_n$ (see Fang et al. (1990, Theorem 2.7)) one obtains from (6) that

$$\mathbb{C}(\epsilon_t) = \frac{\mathbb{E}(\rho_t^2)}{n}\Sigma. \tag{8}$$

Note that the finiteness of the covariance matrix $\mathbb{C}(\epsilon_t)$ is equivalent to $\mathbb{E}(\rho_t^2) < \infty$.

A convenient feature of elliptical distributions is that we can often work with the scalar random variable ρ_t instead of the random vector ϵ_t. For subsequent purposes we therefore note that the density of ρ_t^2, denoted by $\varphi_{\rho^2}(\cdot;\lambda)$, is related to the function $f(\cdot;\lambda)$ in (7) via

$$\varphi_{\rho^2}(\zeta;\lambda) = \frac{\pi^{n/2}}{\Gamma(n/2)}\zeta^{n/2-1}f(\zeta;\lambda), \quad \zeta \geq 0, \tag{9}$$

where $\Gamma(\cdot)$ is the gamma function (see Fang et al. (1990, p. 36)). Assumptions to be imposed on the density of ϵ_t can be expressed by using the function $f(\zeta;\lambda)$ $(\zeta \geq 0)$. These assumptions are similar to those previously used by Andrews et al. (2006) and Lanne and Saikkonen (2008) in so-called all-pass models and univariate noncausal autoregressive models, respectively.

We denote by Λ the permissible parameter space of λ and use $f'(\zeta;\lambda)$ to signify the partial derivative $\partial f(\zeta,\lambda)/\partial\zeta$ with a similar definition for $f''(\zeta;\lambda)$. Also, we include a subscript (typically λ) in the expectation operator or covariance operator when it seems reasonable to emphasize the parameter value assumed in the calculations. Our second assumption is as follows.

Assumption 2. (i) The parameter space Λ is an open subset of $\mathbb{R}^{\mathrm{d}}$ and that of the parameter matrix Σ is the set of positive definite $n \times n$ matrices.
(ii) The function $f(\zeta;\lambda)$ is positive and twice continuously differentiable on $(0,\infty) \times \Lambda$. Furthermore, for all $\lambda \in \Lambda$, $\lim_{\zeta\to\infty} \zeta^{n/2} f(\zeta;\lambda) = 0$, and a finite and positive right limit $\lim_{\zeta\to 0+} f(\zeta;\lambda)$ exists.
(iii) For all $\lambda \in \Lambda$,

$$\int_0^\infty \zeta^{n/2+1} f(\zeta;\lambda)\, d\zeta < \infty \quad \text{and} \quad \int_0^\infty \zeta^{n/2}(1+\zeta)\frac{(f'(\zeta;\lambda))^2}{f(\zeta;\lambda)} d\zeta < \infty.$$

Assuming that the parameter space Λ is open is not restrictive and facilitates exposition. The former part of Assumption 2(ii) is similar to condition (A1) in Andrews et al. (2006) and Lanne and Saikkonen (2008) although in these papers the domain of the first argument of the function f is the whole real line. The latter part of Assumption 2(ii) is technical and needed in some proofs. The first condition in Assumption 2(iii) implies that $\mathbb{E}_\lambda(\rho_t^4)$ is finite (see (9)) and altogether this assumption guarantees finiteness of some expectations needed in subsequent developments. In particular, the latter condition implies finiteness of the quantities

$$\boldsymbol{j}(\lambda) = \frac{4\pi^{n/2}}{n\Gamma(n/2)} \int_0^\infty \zeta^{n/2} \frac{(f'(\zeta;\lambda))^2}{f(\zeta;\lambda)} d\zeta = \frac{4}{n}\mathbb{E}_\lambda \left[\rho_t^2 \left(\frac{f'(\rho_t^2;\lambda)}{f(\rho_t^2;\lambda)}\right)^2\right] \tag{10}$$

and

$$\boldsymbol{i}(\lambda) = \frac{\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty \zeta^{n/2+1} \frac{(f'(\zeta;\lambda))^2}{f(\zeta;\lambda)} d\zeta = \mathbb{E}_\lambda \left[\rho_t^4 \left(\frac{f'(\rho_t^2;\lambda)}{f(\rho_t^2;\lambda)}\right)^2\right], \tag{11}$$

where the latter equalities are obtained by using the density of ρ_t^2 (see (9)). The quantities $\boldsymbol{j}(\lambda)$ and $\boldsymbol{i}(\lambda)$ can be used to characterize non-Gaussianity of the error term ϵ_t. Specifically we can prove the following.

Lemma 1. . *Suppose that Assumptions 1-3 hold. Then,* $\boldsymbol{j}(\lambda) \geq n/\mathbb{E}_\lambda(\rho_t^2)$ *and* $\boldsymbol{i}(\lambda) \geq (n+2)^2 [\mathbb{E}_\lambda(\rho_t^2)]^2 / 4\mathbb{E}_\lambda(\rho_t^4)$ *where equalities hold if and only if* ϵ_t *is Gaussian. If* ϵ_t *is Gaussian,* $\boldsymbol{j}(\lambda) = 1$ *and* $\boldsymbol{i}(\lambda) = n(n+2)/4$.

Lemma 1 shows that assuming $\boldsymbol{j}(\lambda) > n/\mathbb{E}_\lambda(\rho_t^2)$ gives a counterpart of condition (A5) in Andrews et al. (2006) and Lanne and Saikkonen (2008). A difference is, however, that in these papers the variance of the error term is scaled so that the lower part of the inequality does not involve a counterpart of the expectation $\mathbb{E}_\lambda(\rho_t^2)$. For later purposes it is convenient to introduce a scaled version of $\boldsymbol{j}(\lambda)$ given by

$$\boldsymbol{\tau}(\lambda) = \boldsymbol{j}(\lambda)\,\mathbb{E}_\lambda\left(\rho_t^2\right)/n. \tag{12}$$

Clearly, $\boldsymbol{\tau}(\lambda) \geq 1$ with equality if and only if ϵ_t is Gaussian.

It appears useful to generalize the model defined in equation (1) by allowing the coefficient matrices Π_j $(j = 1, ..., r)$ and Φ_j $(j = 1, ..., s)$ to depend on smaller dimensional parameter vectors. We make the following assumption.

Assumption 3. The parameter matrices $\Pi_j = \Pi_j(\vartheta_1)$ $(j = 1, ..., r)$ and $\Phi_j(\vartheta_2)$ $(j = 1, ..., s)$ are twice continuously differentiable functions of the parameter vectors $\vartheta_1 \in \Theta_1 \subseteq \mathbb{R}^{m_1}$ and $\vartheta_2 \in \Theta_2 \subseteq \mathbb{R}^{m_2}$, where the permissible parameter spaces Θ_1 and Θ_2 are open and such that condition (2) holds for all $\vartheta = (\vartheta_1, \vartheta_2) \in \Theta_1 \times \Theta_2$.

This is a standard assumption which guarantees that the likelihood function is twice continuously differentiable. We will continue to use the notation Π_j and Φ_j when there is no need to make the dependence on the underlying parameter vectors explicit.

3 Parameter estimation

3.1 Likelihood function

ML estimation of the parameters of a univariate noncausal autoregression was studied by Breidt et al. (1991) by using a parametrization different from that in (1). The parametrization (1) was employed by Lanne and Saikkonen (2008) whose results we here extend. Unless otherwise stated, Assumptions 1-3 are supposed to hold.

Suppose we have an observed time series $y_1, ..., y_T$. Denote

$$\det \Pi(z) = a(z) = 1 - a_1 z - \cdots - a_{nr} z^{nr}.$$

Then, $w_t = a(B) y_t$ which in conjunction with the definition $u_t = \Phi(B^{-1}) y_t$ yields

$$\begin{bmatrix} u_1 \\ \vdots \\ u_{T-s} \\ w_{T-s+1} \\ \vdots \\ w_T \end{bmatrix} = \begin{bmatrix} y_1 - \Phi_1 y_2 - \cdots - \Phi_s y_{s+1} \\ \vdots \\ y_{T-s} - \Phi_1 y_{T-s+1} - \cdots - \Phi_s y_T \\ y_{T-s+1} - a_1 y_{T-s} - \cdots - a_{nr} y_{T-s-nr+1} \\ \vdots \\ y_T - a_1 y_{T-1} - \cdots - a_{nr} y_{T-nr} \end{bmatrix} = \boldsymbol{H}_1 \begin{bmatrix} y_1 \\ \vdots \\ y_{T-s} \\ y_{T-s+1} \\ \vdots \\ y_T \end{bmatrix}$$

or briefly

$$\boldsymbol{x} = \boldsymbol{H}_1 \boldsymbol{y}.$$

The definition of u_t and (1) yield $\Pi(B)\, u_t = \epsilon_t$ so that, by the preceding equality,

$$\begin{bmatrix} u_1 \\ \vdots \\ u_r \\ \epsilon_{r+1} \\ \vdots \\ \epsilon_{T-s} \\ w_{T-s+1} \\ \vdots \\ w_T \end{bmatrix} = \begin{bmatrix} u_1 \\ \vdots \\ u_r \\ u_{r+1} - \Pi_1 u_r - \cdots - \Pi_r u_1 \\ \vdots \\ u_{T-s} - \Pi_1 u_{T-s-1} - \cdots - \Pi_r u_{T-s-r} \\ w_{T-s+1} \\ \vdots \\ w_T \end{bmatrix} = \boldsymbol{H}_2 \begin{bmatrix} u_1 \\ \vdots \\ u_r \\ u_{r+1} \\ \vdots \\ u_{T-s} \\ w_{T-s+1} \\ \vdots \\ w_T \end{bmatrix}$$

or

$$\boldsymbol{z} = \boldsymbol{H}_2 \boldsymbol{x}.$$

Hence, we get the equation

$$\boldsymbol{z} = \boldsymbol{H}_2 \boldsymbol{H}_1 \boldsymbol{y},$$

where the (nonstochastic) matrices $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are nonsingular. The nonsingularity of $\boldsymbol{H}_2$ follows from the fact that $\det(\boldsymbol{H}_2) = 1$, as can be easily checked. Justifying the nonsingularity of $\boldsymbol{H}_1$ is somewhat more complicated, and will be demonstrated in Appendix B.

From (3) and (4) it can be seen that the components of $\boldsymbol{z}$ given by $\boldsymbol{z}_1 = (u_1, ..., u_r)$, $\boldsymbol{z}_2 = \left(\epsilon_{r+1}, ..., \epsilon_{T-s-(n-1)r}\right)$, and $\boldsymbol{z}_3 = (\epsilon_{T-s-(n-1)r+1}, ..., \epsilon_{T-s}, w_{T-s+1}, ..., w_T)$ are independent. Thus, (under true parameter values) the joint density function of $\boldsymbol{z}$ can be expressed as

$$h_{\boldsymbol{z}_1}(\boldsymbol{z}_1) \left(\prod_{t=r+1}^{T-s-(n-1)r} f_\Sigma(\epsilon_t; \lambda) \right) h_{\boldsymbol{z}_3}(\boldsymbol{z}_3),$$

where $h_{\boldsymbol{z}_1}(\cdot)$ and $h_{\boldsymbol{z}_3}(\cdot)$ signify the joint density functions of $\boldsymbol{z}_1$ and $\boldsymbol{z}_3$, respectively. Using (1) and the fact that the determinant of $\boldsymbol{H}_2$ is unity we can write the joint density function of the data vector $\mathbf{y}$ as

$$h_{\boldsymbol{z}_1}(\boldsymbol{z}_1(\vartheta)) \left(\prod_{t=r+1}^{T-s-(n-1)r} f_\Sigma\left(\Pi(B)\, \Phi\left(B^{-1}\right) y_t; \lambda\right) \right) h_{\boldsymbol{z}_3}(\boldsymbol{z}_3(\vartheta)) \left|\det(\boldsymbol{H}_1)\right|,$$

where the arguments $\boldsymbol{z}_1(\vartheta)$ and $\boldsymbol{z}_3(\vartheta)$ are defined by replacing u_t, ϵ_t, and w_t in the definitions of $\boldsymbol{z}_1$ and $\boldsymbol{z}_3$ by $\Phi(B^{-1})\,y_t$, $\Pi(B)\,\Phi(B^{-1})\,y_t$, and $a(B)\,y_t$, respectively.

It is easy to check that the determinant of the $(T-s)\,n\times(T-s)\,n$ block in the upper left hand corner of $\boldsymbol{H}_1$ is unity and, using the well-known formula for the determinant of a partitioned matrix, it can furthermore be seen that the determinant of $\boldsymbol{H}_1$ is independent of the sample size T. This suggests approximating the joint density of $\boldsymbol{y}$ by the second factor in the preceding expression, giving rise to the approximate log-likelihood function

$$l_T(\theta)=\sum_{t=r+1}^{T-s-(n-1)r} g_t(\theta)\,, \tag{13}$$

where the parameter vector θ contains the unknown parameters and (cf. (7))

$$g_t(\theta)=\log f\left(\epsilon_t(\vartheta)'\,\Sigma^{-1}\epsilon_t(\vartheta)\,;\lambda\right)-\frac{1}{2}\log\det(\Sigma)\,, \tag{14}$$

with

$$\epsilon_t(\vartheta)=u_t(\vartheta_2)-\sum_{j=1}^{r}\Pi_j(\vartheta_1)\,u_{t-j}(\vartheta_2) \tag{15}$$

and $u_t(\vartheta_2)=I_n-\Phi_1(\vartheta_2)\,y_{t+1}-\cdots-\Phi_s(\vartheta_2)\,y_{t+s}$. In addition to ϑ and λ the parameter vector θ also contains the different elements of the matrix Σ, that is, the vector $\sigma=\text{vech}(\Sigma)$. For simplicity, we shall usually drop the word 'approximate' and speak about likelihood function. The same convention is used for related quantities such as the ML estimator of the parameter θ or its score and Hessian.

Maximizing $l_T(\theta)$ over permissible values of θ (see Assumptions 2(i) and 3) gives an approximate ML estimator of θ. Note that here, as well as in the next section, the orders r and s are assumed known. Procedures to specify these quantities will be discussed later.

3.2 Score vector

At this point we introduce the notation θ_0 for the true value of the parameter θ and similarly for its components. Note that our assumptions imply that θ_0 is an interior point of the parameter space of θ. To simplify notation we write $\epsilon_t(\vartheta_0)=\epsilon_t$ and $u_t(\vartheta_{20})=u_{0t}$ when convenient. The subscript '0' will similarly be included in the coefficient matrices of the infinite moving average representations (3), (4), and (5) to emphasize that they are related to the data generation process (i.e. M_{j0}, N_{j0}, and Ψ_{j0}). We also denote $\pi_j(\vartheta_1) = \text{vec}(\Pi_j(\vartheta_1))$ $(j = 1, ..., r)$ and $\phi_j(\vartheta_2) = \text{vec}(\Phi_j(\vartheta_2))$ $(j = 1, ..., s)$, and set

$$\nabla_1(\vartheta_1) = \left[\frac{\partial}{\partial \vartheta_1}\pi_1(\vartheta_1) : \cdots : \frac{\partial}{\partial \vartheta_1}\pi_r(\vartheta_1)\right]'$$

and

$$\nabla_2(\vartheta_2) = \left[\frac{\partial}{\partial \vartheta_2}\phi_1(\vartheta_2) : \cdots : \frac{\partial}{\partial \vartheta_2}\pi_s(\vartheta_2)\right]'.$$

In this section, we consider $\partial l_T(\theta_0)/\partial\theta$, the score of θ evaluated at the true parameter value θ_0. Explicit expressions of the components of the score vector are given in Appendix A. Here we only present the expression of the limit $\lim_{T\to\infty} T^{-1}\mathbb{C}(\partial l_T(\theta_0)/\partial\theta)$. The asymptotic distribution of the score is presented in the following proposition for which additional assumptions and notation are needed. For the treatment of the score of λ we impose the following assumption.

Assumption 4. (i) There exists a function $f_1(\zeta)$ such that $\int_0^\infty \zeta^{n/2-1} f_1(\zeta)\, d\zeta < \infty$ and, in some neighborhood of λ_0, $|\partial f(\zeta;\lambda)/\partial\lambda_i| \leq f_1(\zeta)$ for all $\zeta \geq 0$ and $i = 1, ..., d$.

(ii) $$\left|\int_0^\infty \frac{\zeta^{n/2-1}}{f(\zeta;\lambda_0)}\frac{\partial}{\partial\lambda_i} f(\zeta;\lambda_0)\frac{\partial}{\partial\lambda_j}\partial f(\zeta;\lambda_0)\, d\zeta\right| < \infty, \quad i, j = 1, ..., d.$$

The first condition is a standard dominance condition which guarantees that the score of λ (evaluated at θ_0) has zero mean. The second condition simply assumes that the covariance matrix of the score of λ (evaluated at θ_0) is finite. For other scores the corresponding properties are obtained from the assumptions made in the previous section.

Recall the definition $\boldsymbol{\tau}(\lambda) = \boldsymbol{j}(\lambda)\mathbb{E}_\lambda(\rho_t^2)/n$ where $\boldsymbol{j}(\lambda)$ is defined in (10). In what follows, we denote $\boldsymbol{j}_0 = \boldsymbol{j}(\lambda_0)$ and $\boldsymbol{\tau}_0 = \boldsymbol{j}_0\mathbb{E}_{\lambda_0}(\rho_t^2)/n$. Define the $n \times n$ matrix

$$C_{11}(a,b) = \boldsymbol{\tau}_0 \sum_{k=0}^{\infty} M_{k-a,0}\Sigma_0 M'_{k-b,0}$$

and set $C_{11}(\theta_0) = \left[C_{11}(a,b) \otimes \Sigma_0^{-1}\right]_{a,b=1}^r$ $(n^2 r \times n^2 r)$ and, furthermore,

$$\mathcal{I}_{\vartheta_1\vartheta_1}(\theta_0) = \nabla_1(\vartheta_{10})' C_{11}(\theta_0)\nabla_1(\vartheta_{10}).$$

Notice that $\boldsymbol{j}_0^{-1}C_{11}(a,b) = \mathbb{E}_{\lambda_0}\left(u_{0,t-a}u'_{0,t-b}\right)$. As shown in Appendix B, $\mathcal{I}_{\vartheta_1\vartheta_1}(\theta_0)$ is the standardized covariance matrix of the score of ϑ_1 or the (Fisher) information matrix of

ϑ_1 evaluated at θ_0. In what follows, the term information matrix will be used to refer to the covariance matrix of the asymptotic distribution of the score vector $\partial l_T(\theta_0)/\partial\theta$.

Presenting the information matrix of ϑ_2 is somewhat complicated. First define

$$J_0 = \boldsymbol{i}_0\mathbb{E}\left[(\text{vech}(\upsilon_t\upsilon_t'))(\text{vech}(\upsilon_t\upsilon_t'))'\right] - \tfrac{1}{4}\text{vech}(I_n)\,\text{vech}(I_n)',$$

a square matrix of order $n(n+1)/2$. An explicit expression of the expectation on the right hand side can be obtained from Wong and Wang (1992, p. 274). We also denote $\Pi_{i0} = \Pi(\vartheta_{10})$, $i = 1, ..., r$, and $\Pi_{00} = -I_n$, and define the partitioned matrix $C_{22}(\theta_0) = [C_{22}(a, b; \theta_0)]_{a,b=1}^{s}$ $(n^2s \times n^2s)$ where the $n \times n$ matrix $C_{22}(a, b; \theta_0)$ is

$$\begin{aligned} C_{22}(a,b;\theta_0) &= \boldsymbol{\tau}_0 \sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} \sum_{i,j=0}^{r} \left(\Psi_{k+a-i,0}\Sigma_0\Psi_{k+b-j,0}' \otimes \Pi_{i0}'\Sigma_0^{-1}\Pi_{j0}\right) \\ &\quad + \sum_{i,j=0}^{r} \left(\Psi_{a-i,0}\Sigma_0^{1/2} \otimes \Pi_{i0}'\Sigma_0^{-1/2}\right)(4D_nJ_0D_n' - K_{nn})\left(\Sigma_0^{1/2}\Psi_{b-j,0}' \otimes \Sigma_0^{-1/2}\Pi_{j0}\right). \end{aligned}$$

Now set

$$\mathcal{I}_{\vartheta_2\vartheta_2}(\theta_0) = \nabla_2(\vartheta_{20})'\, C_{22}(\theta_0)\, \nabla_2(\vartheta_{20}),$$

which is the (limiting) information matrix of ϑ_2 (see Appendix B).

To be able to present the information matrix of the whole parameter vector ϑ we define the $n^2 \times n^2$ matrix

$$C_{12}(a,b;\theta_0) = -\boldsymbol{\tau}_0 \sum_{k=a}^{\infty}\sum_{i=0}^{r} \left(M_{k-a,0}\Sigma_0\Psi_{k+b-i,0}' \otimes \Sigma_0^{-1}\Pi_{i0}\right) + K_{nn}\left(\Psi_{b-a,0}' \otimes I_n\right)$$

and the $n^2r \times n^2s$ matrix $C_{12}(\theta_0) = [C_{12}(a, b; \theta_0)] = C_{21}(\theta_0)'$ $(a = 1, ..., r,\ b = 1, ..., s)$. Then the off-diagonal blocks of the (limiting) information matrix of ϑ are given by

$$\mathcal{I}_{\vartheta_1\vartheta_2}(\theta_0) = \nabla_1(\vartheta_{10})'\, C_{12}(\theta_0)\, \nabla_2(\vartheta_{20}) = \mathcal{I}_{\vartheta_2\vartheta_1}(\theta_0)'.$$

Combining the preceding definitions we now define the matrix

$$\mathcal{I}_{\vartheta\vartheta}(\theta) = \left[\mathcal{I}_{\vartheta_i\vartheta_j}(\theta)\right]_{i,j=1,2}.$$

For the remaining blocks of the information matrix of θ, we first define

$$\mathcal{I}_{\sigma\sigma}(\theta_0) = D_n'\left(\Sigma_0^{-1/2} \otimes \Sigma_0^{-1/2}\right) D_nJ_0D_n' \left(\Sigma_0^{-1/2} \otimes \Sigma_0^{-1/2}\right) D_n$$

and

$$\mathcal{I}_{\vartheta_2\sigma}(\theta_0) = -2\sum_{j=1}^{s}\frac{\partial}{\partial\vartheta_2}\phi_j(\vartheta_2)\sum_{i=0}^{r}\left(\Psi_{j-i,0}\Sigma_0^{1/2}\otimes\Pi_{i0}'\Sigma_0^{-1/2}\right)D_nJ_0D_n'\left(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2}\right)D_n$$

with $\mathcal{I}_{\vartheta_2\sigma}(\theta)' = \mathcal{I}_{\sigma\vartheta_2}(\theta)$. Finally, define

$$\mathcal{I}_{\lambda\lambda}(\theta_0) = \frac{\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty \frac{\zeta^{n/2-1}}{f(\zeta;\lambda_0)}\left(\frac{\partial}{\partial\lambda}f(\zeta;\lambda_0)\right)\left(\frac{\partial}{\partial\lambda}f(\zeta;\lambda_0)\right)' d\zeta$$

and

$$\mathcal{I}_{\sigma\lambda}(\theta_0) = -D_n'\left(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2}\right)D_n\text{vech}(I_n)\frac{\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty \zeta^{n/2}\frac{f'(\zeta;\lambda_0)}{f(\zeta;\lambda_0)}\frac{\partial}{\partial\lambda'}f(\zeta;\lambda_0)\,d\zeta$$

with $\mathcal{I}_{\sigma\lambda}(\theta_0)' = \mathcal{I}_{\lambda\sigma}(\theta_0)$. Here the integrals are finite by Assumptions 2(iii) and 4(ii), and the Cauchy-Schwarz inequality.

The information matrix of the whole parameter vector θ is given by

$$\mathcal{I}_{\theta\theta}(\theta_0) = \begin{bmatrix} \mathcal{I}_{\vartheta_1\vartheta_1}(\theta_0) & \mathcal{I}_{\vartheta_1\vartheta_2}(\theta_0) & 0 & 0 \\ \mathcal{I}_{\vartheta_2\vartheta_1}(\theta_0) & \mathcal{I}_{\vartheta_2\vartheta_2}(\theta_0) & \mathcal{I}_{\vartheta_2\sigma}(\theta_0) & 0 \\ 0 & \mathcal{I}_{\sigma\vartheta_2}(\theta_0) & \mathcal{I}_{\sigma\sigma}(\theta_0) & \mathcal{I}_{\sigma\lambda}(\theta_0) \\ 0 & 0 & \mathcal{I}_{\lambda\sigma}(\theta_0) & \mathcal{I}_{\lambda\lambda}(\theta_0) \end{bmatrix}.$$

Note that in the scalar case $n = 1$ and in the purely noncausal case $r = 0$ the expressions of $\mathcal{I}_{\vartheta_2\vartheta_2}(\theta_0)$ and $\mathcal{I}_{\vartheta_1\vartheta_2}(\theta_0)$ simplify and $\mathcal{I}_{\vartheta_2\sigma}(\theta_0)$ becomes zero (see equality (B.6) in Appendix B). The latter fact means that in these special cases the parameters ϑ and (σ, λ) are orthogonal so that their ML estimators are asymptotically independent.

Before presenting the limiting distribution of the score of θ we introduce conditions which guarantee the positive definiteness of its covariance matrix. Specifically, we assume the following.

Assumption 5. (i) The matrices $\nabla_1(\vartheta_{10})$ $(rn^2 \times m_1)$ and $\nabla_2(\vartheta_{10})$ $(sn^2 \times m_2)$ are of full column rank.

(ii) The matrix $\begin{bmatrix} \mathcal{I}_{\sigma\sigma}(\theta_0) & \mathcal{I}_{\sigma\lambda}(\theta_0) \\ \mathcal{I}_{\lambda\sigma}(\theta_0) & \mathcal{I}_{\lambda\lambda}(\theta_0) \end{bmatrix}$ is positive definite.

Assumption 5(i) imposes conventional rank conditions on the first derivatives of the functions in Assumption 3. Assumption 5(ii) is analogous to what has been assumed in previous univariate models (see Andrews et al. (2006) and Lanne and Saikkonen (2008)). Note, however, that unlike in the univariate case it is here less obvious that this assumption is sufficient for the positive definiteness of the whole information matrix $\mathcal{I}_{\theta\theta}(\theta_0)$. The reason is that in the univariate case the situation is simpler in that the parameters λ and σ are orthogonal to the autoregressive parameters (here ϑ_1 and ϑ_2). In the present case the orthogonality of σ with respect to ϑ_2 generally fails but it is still possible to do without assuming more than assumed in the univariate case. Note also that, similarly to the aforementioned univariate cases, Assumption 5(ii) is not needed to guarantee the positive definiteness of $\mathcal{I}_{\sigma\sigma}(\theta_0)$. This follows from the definition of $\mathcal{I}_{\sigma\sigma}(\theta_0)$ and the facts that duplication matrices are of full column rank and the matrix J_0 is positive definite even in the Gaussian case (see Lemma 4 in Appendix B).

Now we can present the limiting distribution of the score.

Proposition 1. *Suppose that Assumptions 1–5 hold and that ϵ_t is non-Gaussian. Then,*

$$(T-s-nr)^{-1/2}\sum_{t=r+1}^{T-s-(n-1)r} g_t(\theta_0) \xrightarrow{d} N(0, \mathcal{I}_{\theta\theta}(\theta_0)),$$

where the matrix $\mathcal{I}_{\theta\theta}(\theta_0)$ is positive definite.

This result generalizes the corresponding univariate result given in Breidt et al. (1991) and Lanne and Saikkonen (2008). In the following section we generalize the work of these authors further by deriving the limiting distribution of the (approximate) ML estimator of θ. Note that for this result it is crucial that ϵ_t is non-Gaussian because in the Gaussian case the information matrix $\mathcal{I}_{\theta\theta}(\theta_0)$ is singular (see the proof of Proposition 1, Step 2).

3.3 Limiting distribution of the approximate ML estimator

The expressions of the second partial derivatives of the log-likelihood function can be found in Appendix A. The following lemma shows that the expectations of these derivatives evaluated at the true parameter value agree with the corresponding elements of $-\mathcal{I}_{\theta\theta}(\theta_0)$. For this lemma we need the following assumption.

Assumption 6.(i) The integral $\int_0^\infty \zeta^{n/2-1} f'(\zeta;\lambda_0)\,d\zeta$ is finite, $\lim_{\zeta\to\infty}\zeta^{n/2+1}f'(\zeta;\lambda_0)$ $=0$, and a finite right limit $\lim_{\zeta\to 0+} f'(\zeta;\lambda_0)$ exists.

(ii) There exists a function $f_2(\zeta)$ such that $\int_0^\infty \zeta^{n/2-1} f_2(\zeta)\, d\zeta < \infty$ and, in some neighborhood of λ_0, $\zeta \left|\partial f'(\zeta;\lambda)/\partial\lambda_i\right| \leq f_2(\zeta)$ and $\left|\partial^2 f(\zeta;\lambda)/\partial\lambda_i\partial\lambda_j\right| \leq f_2(\zeta)$ for all $\zeta \geq 0$ and $i, j = 1, ..., \mathrm{d}$.

Assumption 6(i) is similar to the latter part of Assumption 2(ii) except that it is formulated for the derivative $f'(\zeta;\lambda_0)$. Assumption 6(ii) imposes a standard dominance condition which guarantees that the expectation of $\partial g_t(\theta_0)/\partial\lambda\partial\lambda'$ behaves in the desired fashion. It complements Assumption 4(i) which is formulated similarly to deal with the expectation of $\partial g_t(\theta_0)/\partial\lambda$. Now we can formulate the following lemma.

Lemma 2. *If Assumptions 1-6 hold then* $-T^{-1}\mathbb{E}_{\theta_0}\left[\partial^2 l_T(\theta_0)/\partial\theta\partial\theta'\right] = \mathcal{I}_{\theta\theta}(\theta_0)$.

Lemma 2 shows that the Hessian of the log-likelihood function evaluated at the true parameter value is related to the information matrix in the standard way, implying that $\partial g_t(\theta_0)/\partial\theta\partial\theta'$ obeys a desired law of large numbers. However, to establish the asymptotic normality of the ML estimator more is needed, namely the applicability of a uniform law of large numbers in some neighborhood of θ_0, and for that additional assumptions are required. As usual, it suffices to impose appropriate dominance conditions such as those given in the following assumption.

Assumption 7. For all $\zeta \geq 0$ and all λ in some neighborhood of λ_0, the functions

$$\left(\frac{f'(\zeta;\lambda)}{f(\zeta;\lambda)}\right)^2, \quad \left|\frac{f''(\zeta;\lambda)}{f(\zeta;\lambda)}\right|, \quad \frac{1}{f(\zeta;\lambda)^2}\left(\frac{\partial}{\partial\lambda_j} f(\zeta;\lambda)\right)^2$$

$$\frac{1}{f(\zeta;\lambda)}\left|\frac{\partial}{\partial\lambda_j} f'(\zeta;\lambda)\right|, \quad \frac{1}{f(\zeta;\lambda)}\left|\frac{\partial^2}{\partial\lambda_j\partial\lambda_k} f(\zeta;\lambda)\right|, \quad j,k = 1,...,\mathrm{d},$$

are dominated by $a_1 + a_2\zeta^{a_3}$ with a_1, a_2, and a_3 nonnegative constants and $\int_0^\infty \zeta^{n/2+1+a_3} f(\zeta;\lambda_0)\, d\zeta < \infty$.

The dominance means that, for example, $\left(f'(\zeta;\lambda)/f(\zeta;\lambda)\right)^2 \leq a_1 + a_2\zeta^{a_3}$ for ζ and λ as specified. These dominance conditions are very similar to those required in condition (A7) of Andrews et al. (2006) and Lanne and Saikkonen (2008).

Now we can state the main result of this section.

Theorem 1. *Suppose that Assumptions 1–7 of hold and that ϵ_t is non-Gaussian. Then there exists a sequence of (local) maximizers $\hat{\theta}$ of $l_T(\theta)$ in (13) such that*

$$(T - s - nr)^{1/2} (\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, \mathcal{I}_{\theta\theta}(\theta_0)^{-1}\right).$$

Furthermore, $\mathcal{I}_{\theta\theta}(\theta_0)$ can consistently be estimated by $-(T - s - nr)^{-1} \partial^2 l_T(\hat{\theta})/\partial\theta\partial\theta'$.

Theorem 1 shows that the usual result on asymptotic normality holds for a local maximizer of the likelihood function and that the limiting covariance matrix can consistently be estimated with the Hessian of the log-likelihood function. Based on these results and arguments used in their proof, conventional likelihood based tests with limiting chi-square distribution can be obtained. It is worth noting, however, that consistent estimation of the limiting covariance matrix cannot be based on the outer product of the first derivatives of the log-likelihood function. Specifically, $(T - s - nr)^{-1} \sum_{t=r+1}^{T-s-(n-1)r} (\partial g_t(\hat{\theta})/\partial\theta)(\partial g_t(\hat{\theta})/\partial\theta')$ is, in general, not a consistent estimator of $\mathcal{I}_{\theta\theta}(\theta_0)$. The reason is that this estimator does not take nonzero covariances between $\partial g_t(\theta_0)/\partial\theta$ and $\partial g_k(\theta_0)/\partial\theta$, $k \neq t$, into account. Such covariances are, for example, responsible for the term $K_{nn}\left(\Psi'_{b-a} \otimes I_n\right)$ in $\mathcal{I}_{\vartheta_1\vartheta_2}(\theta_0)$ (see the definition of $C_{12}(a, b; \theta_0)$ and the related proof of Proposition 1 in Appendix B). For instance, in the scalar case $n = 1$ this estimator would be consistent only when the ML estimators of ϑ_1 and ϑ_2 are asymptotically independent which only holds in special cases.

4 Empirical application

We illustrate the use of the noncausal VAR model with an application to U.S. interest rate data. Specifically, we consider the so-called expectations hypothesis of the term structure of interest rates, according to which the long-term interest rate is a weighted sum of present and expected future short-term interest rates. Campbell and Shiller (1987, 1991) suggested testing the expectations hypothesis by testing the restrictions it imposes on the parameters of a bivariate VAR model for the change in the short-term interest rate and the spread between the long-term and short-term interest rates. The general idea is that a causal VAR model captures the dynamics of interest rates, and therefore, its forecasts can be considered as investors' expectations. If these expectations are rational, i.e., they do not systematically deviate from the observed values, this together with the expectations hypothesis imposes testable restrictions on the parameters of the VAR model. This method, already proposed by Sargent (1979), is straightforward to implement and widely applied in economics besides this particular application. However, it crucially depends on the causality of the employed VAR model, suggesting that the validity of this assumption should be checked to avoid potentially misleading conclusions. If the selected VAR model turns out to be noncausal, the estimates may yield evidence in favor of or against the expectations hypothesis. In particular, according to the expectations hypothesis, the expected changes in the short rate drive the term structure, and therefore, their coefficients in the Φ matrices should be significant in the equation of the spread.

The specification of a potentially noncausal VAR model is carried out along the same lines as in the univariate case in Breidt et al. (1991) and Lanne and Saikkonen (2008). The first step is to fit a conventional causal VAR model by least squares or Gaussian ML and determine its order by using conventional procedures such as diagnostic checks and model selection criteria. Once an adequate causal model is found, we check its residuals for Gaussianity. As already discussed, it makes sense to proceed to noncausal models only if deviations from Gaussianity are detected. If this happens, a non-Gaussian error distribution is adopted and all causal and noncausal models of the selected order are estimated. Of these models the one that maximizes the log-likelihood function is selected and its adequacy is checked by diagnostic tests.

We use the Ljung-Box and McLeod-Li tests to check for error autocorrelation and conditional heteroskedasticity, respectively. Note, however, that when the orders of the model are misspecified, these tests are not exactly valid as they do not take estimation errors correctly into account. The reason is that a misspecification of the model orders makes the errors dependent. Nevertheless, p-values of these tests can be seen as convenient summary measures of the autocorrelation remaining in the residuals and their squares. A similar remark applies to the Shapiro-Wilk test we use to check the error distribution.

Our data set comprises the (demeaned) change in the six-month interest rate (Δr_t) and the spread between the five-year and six-month interest rates (S_t) (quarter-end yields on U.S. zero-coupon bonds) from the thirty-year period 1967:1–1996:4 (120 observations) previously used in Duffee (2002). The AIC and BIC select Gaussian VAR(3) and VAR(1) models, respectively, but only the third-order model produces serially uncorrelated errors. However, the results in Table 1 show that its residuals are conditionally heteroskedastic and the Q-Q plots is the upper panel of Figure 1, indicate considerable deviations from normality. The p-values of the Shapiro-Wilk test for the residuals of the equations of Δr_t and S_t equal 5.06e–9 and 7.23e–7, respectively. Because the most severe violations of normality occur at the tails, a more leptokurtic distribution, such as the multivariate t-distribution, might prove suitable for these data.

The estimation results of all four third-order VAR models with t-distributed errors are summarized in Table 1. By a wide margin, the specification maximizing the log-likelihood function is the VAR(2,1)-t model. It also turns out to be the only one of the estimated models that shows no signs of remaining autocorrelation or conditional heteroskedasticity in the residuals. The Q-Q plots of the residuals in the lower panel of Figure 1 lend support to the adequacy of the multivariate t-distribution of the errors. In particular, the t-distribution seems to capture the tails reasonably well. Moreover, the estimate of the degrees-of-freedom parameter λ turned out to be small (4.085), suggesting inadequacy of the Gaussian error distribution. Thus, there is evidence of noncausality.

The estimates of the preferred model are presented in Table 2. The estimated Φ_1 matrix seems to have an interpretation that goes contrary to the implications of the expectations hypothesis discussed above: an expected increase of the short-term rate has no significant effect on the spread. Furthermore, an expected future increase of the spread tends to decrease the short-term rate and increase the spread. This might be interpreted in favor of (expected) time-varying term premia driving the term structure instead of expectations of future short-term rates as implied by the expectations hypothesis.

The presence of a noncausal VAR representation of Δr_t and S_t invalidates the test of the expectations hypothesis suggested by Campbell and Shiller (1987, 1991). If noncausality prevails more generally in interest rates this might also explain the common rejections of the expectations hypothesis when testing is based on the assumption of a causal VAR model.

5 Conclusion

In this paper, we have proposed a new noncausal VAR model that contains the commonly used causal VAR model as a special case. Under Gaussianity, causal and noncausal VAR models cannot be distinguished which underlines the importance of careful specification of the error distribution of the model. We have derived asymptotic properties of an approximate (local) ML estimator and related tests in the noncausal VAR model, and we have successfully employed an extension of the model selection procedure presented by Breidt et al. (1991) and Lanne and Saikkonen (2008) in the corresponding univariate case. The methods were illustrated by means of an empirical application to the U.S. term structure of interest rates. In that case, evidence of noncausality was found, invalidating the previously employed test of the expectations hypothesis of the term structure of interest rates explicitly based on a causal VAR model.

While the new model appears useful in providing a more accurate description of time series dynamics and checking for the validity of a causal VAR representation, it may also have other uses. For instance, in economic applications noncausal VAR models are expected to be valuable in checking for so-called nonfundamentalness. In economics, a model is said to exhibit nonfundamentalness if its solution explicitly depends on the future so that it does not have a causal VAR representation (for a recent survey of the relevant literature, see Alessi, Barigozzi, and Capasso (2008)). Hence, nonfundamentalness is closely related to noncausality, and checking for noncausality can be seen as a way of testing for nonfundamentalness. Because nonfundamentalness often invalidates the use of conventional econometric methods, being able to detect it in advance is important. However, the test procedures suggested in the previous literature are not very convenient and have not been much applied in practice.

Checking for causality (or fundamentalness) is an important application of our methods, but it can only be considered as the first step in the empirical analysis of time series data. Once noncausality has been detected, it would be natural to use the noncausal VAR model for forecasting and structural analysis. These, however, require methods that are not readily available. Because the prediction problem in noncausal VAR models is generally nonlinear (see Rosenblatt (2000, Chapter 5)) methods used in the causal case are not applicable and, due the explicit dependence on the future, the same is true for conventional simulation-based methods. In the univariate case, Lanne, Luoto, and Saikkonen (2010) have proposed a forecasting method that could plausibly be extended to the noncausal VAR model.

Regarding statistical aspects, the theory presented in this paper is confined to the class of elliptical distributions. Even though the multivariate t-distribution belonging to this class seemed adequate in our empirical applications, it would be desirable to make extensions to other relevant classes of distributions. Also, the finite-sample properties of the proposed model selection procedure could be examined by means of simulation experiments. We leave all of these issues for future research.

Mathematical Appendix

A Derivatives of the log-likelihood function

It will be sufficient to consider the derivatives of $g_t(\theta)$ which can be obtained by straightforward differentiation. To simplify notation we set $h(\zeta;\lambda) = f'(\zeta;\lambda)/f(\zeta;\lambda)$ so that

$$h'\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right) = \frac{f''\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)}{f\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)} - \left(\frac{f'\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)}{f\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)}\right)^2. \quad \text{(A.1)}$$

Next, define

$$e_t(\theta) = h\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)\Sigma^{-1/2}\epsilon_t(\vartheta) \quad \text{and} \quad e_{0t} = e_t(\theta_0). \quad \text{(A.2)}$$

From (6) it is seen that

$$e_{0t} \overset{d}{=} \rho_t h\left(\rho_t^2;\lambda_0\right)\upsilon_t = \rho_t h_0\left(\rho_t^2\right)\upsilon_t, \quad \text{(A.3)}$$

where the latter equality defines the notation $h_0(\cdot) = h(\cdot;\lambda_0)$.

First derivatives of $l_T(\theta)$. From (14) we first obtain

$$\frac{\partial}{\partial\vartheta_i}g_t(\theta) = 2h\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)\frac{\partial}{\partial\vartheta_i}\epsilon_t(\vartheta)\Sigma^{-1}\epsilon_t(\vartheta), \quad i = 1, 2, \quad \text{(A.4)}$$

where, from (15),

$$\frac{\partial}{\partial \vartheta_1}\epsilon_t(\vartheta) = -\sum_{i=1}^{r}\frac{\partial}{\partial \vartheta_1}\pi_i(\vartheta_1)(u_{t-i}(\vartheta_2)\otimes I_n) \tag{A.5}$$

and

$$\frac{\partial}{\partial \vartheta_2}\epsilon_t(\vartheta) = \sum_{i=0}^{r}\sum_{j=1}^{s}\frac{\partial}{\partial \vartheta_2}\phi_j(\vartheta_2)(y_{t+j-i}\otimes \Pi_i'), \tag{A.6}$$

with $\Pi_0 = -I_n = \Pi_{00}$. We also set $U_{t-1}(\vartheta_2) = \left[(u_{t-1}(\vartheta_2)\otimes I_n)' \ \cdots \ (u_{t-r}(\vartheta_2)\otimes I_n)'\right]'$ and $Y_{t+1}(\vartheta_1) = \left[\sum_{i=0}^{r}(y_{t+1-i}\otimes \Pi_i')' \ \cdots \ \sum_{i=0}^{r}(y_{t+s-i}\otimes \Pi_i')'\right]'$. Then, using the notation $U_{t-1}(\vartheta_{20}) = U_{0,t-1}$ and $Y_{t+1}(\vartheta_{10}) = Y_{0,t+1}$,

$$\begin{aligned}\frac{\partial}{\partial \vartheta_1}g_t(\theta_0) &= -2\sum_{i=1}^{r}\frac{\partial}{\partial \vartheta_1}\pi_i(\vartheta_{10})(u_{0,t-i}\otimes I_n)\Sigma_0^{-1/2}e_{0t} \\ &= -2\nabla_1(\vartheta_{10})'U_{0,t-1}\Sigma_0^{-1/2}e_{0t}\end{aligned} \tag{A.7}$$

and

$$\begin{aligned}\frac{\partial}{\partial \vartheta_2}g_t(\theta_0) &= 2\sum_{j=1}^{s}\frac{\partial}{\partial \vartheta_2}\phi_j(\vartheta_{20})\sum_{i=0}^{r}(y_{t+j-i}\otimes \Pi_{i0}')\Sigma_0^{-1/2}e_{0t} \\ &= 2\nabla_2(\vartheta_{20})'Y_{0,t+1}\Sigma_0^{-1/2}e_{0t}.\end{aligned} \tag{A.8}$$

As for the parameters $\sigma = \mathrm{vech}(\Sigma)$ and λ,

$$\begin{aligned}\frac{\partial}{\partial \sigma}g_t(\theta) &= -h\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)D_n'\left(\Sigma^{-1}\otimes\Sigma^{-1}\right)(\epsilon_t(\vartheta)\otimes\epsilon_t(\vartheta)) - \frac{1}{2}D_n'\mathrm{vec}\left(\Sigma^{-1}\right) \\ &= -D_n'\left(\Sigma_0^{-1}\otimes\Sigma_0^{-1}\right)\left(\epsilon_t\otimes\Sigma_0^{1/2}e_{0t} + \frac{1}{2}\mathrm{vec}(\Sigma_0)\right), \quad \text{as } \theta = \theta_0,\end{aligned} \tag{A.9}$$

and

$$\begin{aligned}\frac{\partial}{\partial \lambda}g_t(\theta) &= \frac{1}{f\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)}\frac{\partial}{\partial \lambda}f\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right) \\ &= \frac{1}{f\left(\epsilon_t'\Sigma_0^{-1}\epsilon_t;\lambda_0\right)}\frac{\partial}{\partial \lambda}f\left(\epsilon_t'\Sigma_0^{-1}\epsilon_t;\lambda_0\right) \quad \text{as } \theta = \theta_0.\end{aligned} \tag{A.10}$$

Second derivatives of $l_T(\theta)$. First note that

$$\begin{aligned}\frac{\partial}{\partial \vartheta_i'}e_t(\theta) &= h\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)\Sigma^{-1/2}\frac{\partial}{\partial \vartheta_i'}\epsilon_t(\vartheta) \\ &\quad +2h'\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)\Sigma^{-1/2}\epsilon_t(\vartheta)\epsilon_t(\vartheta)'\Sigma^{-1}\frac{\partial}{\partial \vartheta_i'}\epsilon_t(\vartheta), \quad i = 1, 2.\end{aligned} \tag{A.11}$$

Using these expressions we now have

$$\frac{\partial^2}{\partial\vartheta_1\partial\vartheta_1'}g_t(\theta) = -2\sum_{i=1}^{r}\left(u_{t-i}(\vartheta_2)'\otimes e_t(\theta)'\Sigma^{-1/2}\otimes I_{m_1}\right)\frac{\partial}{\partial\vartheta_1'}\text{vec}\left(\frac{\partial}{\partial\vartheta_1}\pi_i(\vartheta_1)\right)$$
$$-2\sum_{i=1}^{r}\frac{\partial}{\partial\vartheta_1}\pi_i(\vartheta_1)\left(u_{t-i}(\vartheta_2)\otimes I_n\right)\Sigma^{-1/2}\frac{\partial}{\partial\vartheta_1'}e_t(\theta), \quad \text{(A.12)}$$

$$\frac{\partial^2}{\partial\vartheta_2\partial\vartheta_2'}g_t(\theta) = 2\sum_{j=1}^{s}\sum_{i=0}^{r}\left(y_{t+j-i}'\otimes e_t(\theta)'\Sigma^{-1/2}\Pi_i\otimes I_{m_2}\right)\frac{\partial}{\partial\vartheta_2'}\text{vec}\left(\frac{\partial}{\partial\vartheta_2}\phi_j(\vartheta_2)\right)$$
$$+2\sum_{j=1}^{s}\frac{\partial}{\partial\vartheta_2}\phi_j(\vartheta_2)\sum_{i=0}^{r}\left(y_{t+j-i}\otimes\Pi_i'\right)\Sigma^{-1/2}\frac{\partial}{\partial\vartheta_2'}e_t(\theta), \quad \text{(A.13)}$$

and

$$\frac{\partial^2}{\partial\vartheta_1\partial\vartheta_2'}g_t(\theta) = -2\sum_{i=1}^{r}\frac{\partial}{\partial\vartheta_1}\pi_i(\vartheta_1)\left(I_n\otimes\Sigma^{-1/2}e_t(\theta)\right)\frac{\partial}{\partial\vartheta_2'}u_{t-i}(\vartheta_2)$$
$$-2\sum_{i=1}^{r}\frac{\partial}{\partial\vartheta_1}\pi_i(\vartheta_1)\left(u_{t-i}(\vartheta_2)\otimes I_n\right)\Sigma^{-1/2}\frac{\partial}{\partial\vartheta_2'}e_t(\theta), \quad \text{(A.14)}$$

where $\partial u_{t-i}(\vartheta_2)/\partial\vartheta_2' = -\sum_{j=1}^{s}(y_{t+j-i}'\otimes I_n)\partial\phi_j(\vartheta_2)/\partial\vartheta_2'$.

Next consider $\partial^2 g_t(\theta)/\partial\sigma\partial\sigma'$ and conclude from (A.9) that

$$\frac{\partial^2}{\partial\sigma\partial\sigma'}g_t(\theta) = h\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)\left(\epsilon_t(\vartheta)'\otimes\epsilon_t(\vartheta)'\otimes D_n'\right)\left(I_n\otimes K_{nn}\otimes I_n\right)$$
$$\times\left[\Sigma^{-1}\otimes\Sigma^{-1}\otimes\text{vec}\left(\Sigma^{-1}\right)+\text{vec}\left(\Sigma^{-1}\right)\otimes\Sigma^{-1}\otimes\Sigma^{-1}\right]D_n$$
$$+h'\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)D_n'\left(\Sigma^{-1}\otimes\Sigma^{-1}\right)\left(\epsilon_t(\vartheta)\epsilon_t(\vartheta)'\otimes\epsilon_t(\vartheta)\epsilon_t(\vartheta)'\right)$$
$$\times\left(\Sigma^{-1}\otimes\Sigma^{-1}\right)D_n+\frac{1}{2}D_n'\left(\Sigma^{-1}\otimes\Sigma^{-1}\right)D_n, \quad \text{(A.15)}$$

and furthermore that (see(A.4))

$$\frac{\partial^2}{\partial\vartheta_i\partial\sigma'}g_t(\theta) = -2h\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)\left(\epsilon_t(\vartheta)'\otimes\frac{\partial}{\partial\vartheta_i}\epsilon_t(\vartheta)\right) \quad \text{(A.16)}$$
$$\times\left(\Sigma^{-1}\otimes\Sigma^{-1}\right)D_n$$
$$-2h'\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)\frac{\partial}{\partial\vartheta_i}\epsilon_t(\vartheta)\Sigma^{-1}\epsilon_t(\vartheta)\left(\epsilon_t(\vartheta)'\otimes\epsilon_t(\vartheta)'\right)$$
$$\times\left(\Sigma^{-1}\otimes\Sigma^{-1}\right)D_n, \quad i=1,2.$$

For $\partial^2 g_t(\theta)/\partial\lambda\partial\lambda'$ it suffices to note that

$$\begin{aligned}\frac{\partial^2}{\partial\lambda\partial\lambda'}g_t(\theta) &= -\frac{1}{f\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)^2}\frac{\partial}{\partial\lambda}f\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)\\ &\qquad\times\frac{\partial}{\partial\lambda'}f\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)\\ &\quad+\frac{1}{f\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)}\frac{\partial^2}{\partial\lambda\partial\lambda'}f\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)\end{aligned}\tag{A.17}$$

whereas

$$\frac{\partial^2}{\partial\vartheta_i\partial\lambda'}g_t(\theta)=2\frac{\partial}{\partial\vartheta_i}\epsilon_t(\vartheta)\Sigma^{-1}\epsilon_t(\vartheta)\frac{\partial}{\partial\lambda'}h\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right),\quad i=1,2,\tag{A.18}$$

and

$$\frac{\partial^2}{\partial\sigma\partial\lambda'}g_t(\theta)=-D_n'\left(\Sigma^{-1}\otimes\Sigma^{-1}\right)\left(\epsilon_t(\vartheta)\otimes\epsilon_t(\vartheta)\right)\frac{\partial}{\partial\lambda'}h\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right),\tag{A.19}$$

where

$$\begin{aligned}\frac{\partial}{\partial\lambda}h\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right) &= \frac{1}{f\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)}\frac{\partial}{\partial\lambda}f'\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)\\ &\quad-\frac{f'\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)}{\left(f\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)\right)^2}\frac{\partial}{\partial\lambda'}f\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right).\end{aligned}$$

B Proofs for Sections 2 and 3

Proof of Lemma 1. For the former inequality, first consider the expectation

$$\mathbb{E}_\lambda\left[\rho_t^2 h\left(\rho_t^2;\lambda\right)\right]=\frac{\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty\zeta^{n/2}f'(\zeta;\lambda)\,d\zeta=-\frac{n}{2},\tag{B.1}$$

where the definition of the function h (see the beginning of Appendix A), density of ρ_t^2 (see (9)), and Assumption 2(ii) have been used (see the discussion after Assumption 2). The same arguments combined with the Cauchy-Schwarz inequality and the definition of $\boldsymbol{j}(\lambda)$ (see (10)) yield

$$\begin{aligned}1 &= \left\{\frac{2\pi^{n/2}}{n\Gamma(n/2)}\int_0^\infty\zeta^{n/4}\frac{f'(\zeta;\lambda)}{\sqrt{f(\zeta;\lambda)}}\zeta^{n/4}\sqrt{f(\zeta;\lambda)}d\zeta\right\}^2\\ &\le \frac{4\pi^{n/2}}{n\Gamma(n/2)}\int_0^\infty\zeta^{n/2}\frac{\left(f'(\zeta;\lambda)\right)^2}{f(\zeta;\lambda)}d\zeta\cdot\frac{\pi^{n/2}}{n\Gamma(n/2)}\int_0^\infty\zeta^{n/2}f(\zeta;\lambda)\,d\zeta\\ &= \boldsymbol{j}(\lambda)\cdot\mathbb{E}_\lambda\left(\rho_t^2\right)/n.\end{aligned}\tag{B.2}$$

Thus, we have shown the claimed inequality.

From the preceding proof it is seen that equality holds if and only if there is equality in (B.2). As is well known, this happens if and only if $\zeta^{n/4} f'(\zeta;\lambda)/\sqrt{f(\zeta;\lambda)}$ is proportional to $\zeta^{n/4}\sqrt{f(\zeta;\lambda)}$ or if and only if

$$\frac{f'(\zeta;\lambda)}{f(\zeta;\lambda)} = \frac{\partial}{\partial\zeta}\log f(\zeta;\lambda) = c \quad \text{for some } c.$$

This implies $f(\zeta;\lambda) = b\exp(-a\zeta)$ with $a > 0$ and $b > 0$. From the fact that $f(x'x;\lambda)$, $x \in \mathbb{R}^n$, is the density function of $\rho_t \upsilon_t$ (see (6) and (7)) it further follows that $b = (a/\pi)^{n/2}$ and that $\rho_t\upsilon_t$ has the normal density $(2\pi)^{-n/2}\exp(-x'x/2)$. Here the identity covariance matrix is obtained because $\rho_t^2 \sim \chi_n^2$, and hence from (8), $\mathbb{C}(\rho_t^2\upsilon_t) = I_n$ (cf. the corollary to Lemma 1.4 and Example 1.3 of Fang et al. (1990), p. 23). Thus, ϵ_t is Gaussian as a linear transformation of $\rho_t\upsilon_t$. On the other hand, if ϵ_t is Gaussian the equality $f'(\zeta;\lambda)/f(\zeta;\lambda) = c$ clearly holds with $c = -1/2$ and, because then $\rho_t^2 \sim \chi_n^2$, $\mathbb{E}_\lambda(\rho_t^2) = n$ and $\boldsymbol{j}(\lambda) = 1$. This completes the proof for $\boldsymbol{j}(\lambda)$.

Regarding $\boldsymbol{i}(\lambda)$, first notice that

$$\begin{aligned}\int_0^\infty \zeta^{n/2+1} f'(\zeta;\lambda_0)\,d\zeta &= \left(\zeta^{n/2+1} f(\zeta;\lambda)\,|_0^\infty - \frac{n+2}{2}\int_0^\infty \zeta^{n/2} f(\zeta;\lambda)\,d\zeta\right) \\ &= -\frac{n+2}{2}\cdot\frac{\Gamma(n/2)}{\pi^{n/2}}\mathbb{E}_\lambda\left(\rho_t^2\right),\end{aligned}$$

where we have used Assumptions 2(ii) and (iii), and the expression of the density of ρ_t^2 (see (9)). Proceeding as in the case of the first assertion yields

$$\begin{aligned}1 &= \left(\frac{2}{(n+2)\,\mathbb{E}_\lambda(\rho_t^2)}\cdot\frac{\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty \zeta^{n/4+1/2}\frac{f'(\zeta;\lambda)}{\sqrt{f(\zeta;\lambda)}}\zeta^{n/4+1/2}\sqrt{f(\zeta;\lambda)}d\zeta\right)^2 \\ &\leq \left(\frac{2}{(n+2)\,\mathbb{E}_\lambda(\rho_t^2)}\right)^2\cdot\frac{\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty \zeta^{n/2+1}\left(\frac{f'(\zeta;\lambda)}{f(\zeta;\lambda)}\right)^2 f(\zeta;\lambda)\,d\zeta \\ &\qquad\qquad\times\frac{\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty \zeta^{n/2+1} f(\zeta;\lambda)\,d\zeta \\ &= \left(\frac{2}{(n+2)\,\mathbb{E}_\lambda(\rho_t^2)}\right)^2\cdot\boldsymbol{i}(\lambda)\cdot\mathbb{E}_\lambda\left(\rho_t^4\right)\end{aligned}$$

(see the definition of $\boldsymbol{i}(\lambda)$ in (11)). This shows the stated inequality and the condition for equality leads to the same condition as in the case of $\boldsymbol{j}(\lambda)$. Finally, in the Gaussian case, $\mathbb{E}_\lambda(\rho_t^2) = n$ and $\mathbb{E}_\lambda(\rho_t^4) = 2n + n^2$, implying $\boldsymbol{i}(\lambda) = n(n+2)/4$. □

Proof of the nonsingularity of the matrix $\boldsymbol{H}_1$. To simplify notation we demonstrate the nonsingularity of $\boldsymbol{H}_1$ when $s = 2$. From the definition of $\boldsymbol{H}_1$ it is not difficult to see that the possible singularity of $\boldsymbol{H}_1$ can only be due to a linear dependence of its last $n(r+2)$ rows and, furthermore, that it suffices to show the nonsingularity of the lower right hand corner $\boldsymbol{H}_1$ of order $n(r+2) \times n(r+2)$. This matrix reads as

$$\boldsymbol{H}_1^{(2,2)} = \left[\begin{array}{cccccc|cc} I_n & -\Phi_1 & -\Phi_2 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & \cdots & 0 & I_n & -\Phi_1 & -\Phi_2 & 0 & 0 \\ \vdots & \cdots & 0 & 0 & I_n & -\Phi_1 & -\Phi_2 & 0 \\ 0 & \cdots & 0 & 0 & 0 & I_n & -\Phi_1 & -\Phi_2 \\ \hline -a_{nr}I_n & \cdots & \cdots & \cdots & \cdots & -a_1 I_n & I_n & 0 \\ 0 & -a_{nr}I_n & \cdots & \cdots & \cdots & \cdots & -a_1 I_n & I_n \end{array}\right]$$
$$\overset{def}{=} \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{bmatrix},$$

where the partition is as indicated. The determinant of $\boldsymbol{B}_{11}$ is evidently unity so that from the well-known formula for the determinant of a partitioned matrix it follows that we need to show the nonsingularity of the matrix $\boldsymbol{B}_{11\cdot 2} = \boldsymbol{B}_{22} - \boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}$. The inverse of $\boldsymbol{B}_{11}$ depends on coefficients of the power series representation of $L(z) = \Phi(z)^{-1}$ given by $L(z) = \sum_{j=0}^{\infty} L_j z^j$ where $L_0 = I_n$ and, when convenient, $L_j = 0$, $j < 0$, will be used. Equating the coefficient matrices of z on both sides of the identity $L(z)\Phi(z) = I_n$ yields $L_j = L_{j-1}\Phi_1 + L_{j-2}\Phi_2$. Using this identity it is readily seen that $\boldsymbol{B}_{11}^{-1}$ is an upper triangular matrix with I_n on the diagonal and L_j, $j = 1, ..., nr-1$, on the diagonals above the main diagonal. This fact and straightforward but tedious calculations further show that

$$\begin{aligned} \boldsymbol{B}_{11\cdot 2} &= \begin{bmatrix} I_n - \sum_{j=1}^{nr} a_j L_j & -\sum_{j=1}^{nr} a_j L_{j-1}\Phi_2 \\ -\sum_{j=1}^{nr} a_j L_{j-1} & I_n - \sum_{j=2}^{nr} a_j L_{j-2}\Phi_2 \end{bmatrix} \\ &= \begin{bmatrix} I_n & 0 \\ 0 & I_n \end{bmatrix} - \sum_{j=1}^{nr} a_j \begin{bmatrix} L_j & L_{j-1}\Phi_2 \\ L_{j-1} & L_{j-2}\Phi_2 \end{bmatrix}. \end{aligned}$$

Next define the companion matrix

$$\mathbf{\Phi} = \begin{bmatrix} \Phi_1 & \Phi_2 \\ I_n & 0 \end{bmatrix}$$

and note that the latter condition in (2) implies that the eigenvalues of $\mathbf{\Phi}$ are smaller than one in absolute value. Also, the matrices L_j and L_{j-1} $(j \geq 0)$ can be obtained from the upper and lower left hand corners of the matrix $\mathbf{\Phi}^j$, respectively. Using these facts, the identity $L_j = L_{j-1}\Phi_1 + L_{j-2}\Phi_2$, and properties of the powers $\mathbf{\Phi}^j$ it can further be seen that

$$\boldsymbol{B}_{11\cdot 2} = I_{2n} - \sum_{j=1}^{nr} a_j \mathbf{\Phi}^j = \mathbf{P}\left(I_{2n} - \sum_{j=1}^{nr} a_j \mathbf{D}^j\right)\mathbf{P}^{-1},$$

where the latter equality is based on the Jordan decomposition of $\mathbf{\Phi}$ so that $\mathbf{\Phi} = \mathbf{PDP}^{-1}$. Thus, the determinant of $\boldsymbol{B}_{11\cdot 2}$ equals the determinant of the matrix in parentheses in its latter expression. Because $\mathbf{D}^j$ is an upper triangular matrix having the jth powers of the eigenvalues of $\mathbf{\Phi}$ on the diagonal this determinant is a product of quantities of the form $1 - \sum_{j=1}^{nr} a_j \nu^j$ where ν signifies an eigenvalue of $\mathbf{\Phi}$. By the latter condition in (2) the eigenvalues of $\mathbf{\Phi}$ are smaller than one in absolute value whereas the former condition in (2) implies that the zeros of $a(z)$ lie outside the unit disc. Thus, the nonsingularity of $\boldsymbol{B}_{11\cdot 2}$, and hence that of $\boldsymbol{H}_1^{(2,2)}$ and $\boldsymbol{H}_1$ follow.

We note that in the case $s = 1$ the preceding proof simplifies because then we need to show the nonsingularity of the matrix obtained from $\boldsymbol{H}_1^{(2,2)}$ by deleting its last n rows and columns and setting $\Phi_2 = 0$. In place of $\boldsymbol{B}_{11\cdot 2}$ we then have $I_n - \sum_{j=1}^{nr} a_j \Phi_1^j$ and, because now the eigenvalues of Φ_1 are smaller than one in modulus, the preceding argument applies without the need to use a companion matrix. $\square$

Before proving Proposition 1 we present some auxiliary results. In the following lemmas, as well as in the proof of Proposition 1, the true parameter value is assumed, so the notation $\mathbb{E}(\cdot)$ will be used instead of $\mathbb{E}_{\lambda_0}(\cdot)$ and similarly for $\mathbb{C}(\cdot)$. In these proofs frequent use will be made of the facts that the processes ρ_t and υ_t are independent and that $\mathbb{E}(\upsilon_t) = 0$ and $\mathbb{E}(\upsilon_t \upsilon_k')$ equals 0 if $t \neq k$ and $n^{-1} I_n$ if $t = k$. The same can be said about well-known properties of the Kronecker product and vec operator, especially the result $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$ which holds for any conformable matrices A, B, and

C. This and other results of matrix algebra to be employed can be found in Lütkepohl (1996). We also recall the definition $\varepsilon_t = \Sigma_0^{-1/2}\epsilon_t$ (see (6)) and, to simplify notation, we will frequently write $f(\cdot;\lambda_0) = f_0(\cdot)$ and similarly for $f_0'(\cdot)$ and $f_0''(\cdot)$.

Lemma 3. *Under the conditions of Proposition 1,*

$$\mathbb{E}(e_{0t}) = 0 \quad and \quad \mathbb{C}(e_{0t}) = \frac{\boldsymbol{j}_0}{4} I_n, \tag{B.3}$$

and

$$\mathbb{C}(\varepsilon_t, e_{0k}) = \begin{cases} 0, & if \;\; t \neq k \\ -\frac{1}{2} I_n, & if \;\; t = k \end{cases} \tag{B.4}$$

Proof of Lemma 3. By the definition of the function $h_0(\cdot)$ (see (A.3)) and the density of ρ_t^2 (see (9)) we have

$$\mathbb{E}\left[\rho_t^2 \left(h_0\left(\rho_t^2\right)\right)^2\right] = \frac{\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty \zeta^{n/2} \frac{\left(f_0'(\zeta)\right)^2}{f_0(\zeta)} d\zeta = \frac{n}{4} \boldsymbol{j}_0,$$

where the latter equality is due to (10). Thus, because $\mathbb{E}(\upsilon_t) = 0$ and $\mathbb{C}(\upsilon_t) = n^{-1} I_n$, the independence of the processes ρ_t and υ_t in conjunction with (A.3) proves (B.3). The same arguments and (6) yield

$$\mathbb{E}(\varepsilon_t e_{0k}') = \mathbb{E}\left[\rho_t \rho_k h_0\left(\rho_k^2\right)\right] \mathbb{E}(\upsilon_t \upsilon_k'),$$

where $\mathbb{E}(\upsilon_t \upsilon_k') = 0$ for $t \neq k$. Thus, one obtains (B.4) from this and (B.1). □

Lemma 4. . *Under the conditions of Proposition 1,*

$$\mathbb{C}(\varepsilon_{t-i} \otimes e_{0t}, \varepsilon_{k-j} \otimes e_{0k}) = \begin{cases} D_n J_0 D_n', & if \;\; t = k, \; i = j = 0 \\ \frac{\tau_0}{4} I_{n^2}, & if \;\; t = k, \; i = j \neq 0 \\ \frac{1}{4} K_{nn}, & if \;\; t \neq k, \; i = t - k, \; j = k - t \\ 0, & otherwise. \end{cases}$$

Moreover, the matrix J_0 is positive definite even when ϵ_t is Gaussian.

Proof. First notice that (see (6) and (A.3))

$$\varepsilon_{t-i} \otimes e_{0t} \overset{d}{=} \rho_{t-i} \rho_t h_0\left(\rho_t^2\right) (\upsilon_{t-i} \otimes \upsilon_t). \tag{B.5}$$

Consider the case $t = k$ and $i = j = 0$. Using (B.5) and independence of ρ_t and υ_t yields

$$\mathbb{E}\left(\varepsilon_t \otimes e_{0t}\right) = \mathbb{E}\left[\rho_t^2 h_0\left(\rho_t^2\right)\right]\mathbb{E}\left(\upsilon_t \otimes \upsilon_t\right) = -\frac{1}{2}D_n \text{vech}\left(I_n\right),$$

where the latter equality is due to (B.1) and $\mathbb{E}\left(\upsilon_t \otimes \upsilon_t\right) = \text{vec}(\mathbb{E}\left(\upsilon_t \upsilon_t'\right)) = n^{-1}\text{vec}(I_n)$. By the same arguments we also find that

$$\mathbb{E}\left[(\varepsilon_t \otimes e_{0t})(\varepsilon_t \otimes e_{0t})'\right] = \mathbb{E}\left[\rho_t^4\left(h_0\left(\rho_t^2\right)\right)^2\right]\mathbb{E}\left(\upsilon_t\upsilon_t' \otimes \upsilon_t\upsilon_t'\right) = \boldsymbol{i}_0\mathbb{E}\left(\upsilon_t\upsilon_t' \otimes \upsilon_t\upsilon_t'\right),$$

where the latter equality follows from the definition of $\boldsymbol{i}_0$ (see (11)). Because

$$\mathbb{E}\left(\upsilon_t\upsilon_t' \otimes \upsilon_t\upsilon_t'\right) = \mathbb{E}\left[\left(\upsilon_t \otimes \upsilon_t\right)\left(\upsilon_t' \otimes \upsilon_t'\right)\right] = D_n\mathbb{E}\left[\left(\text{vech}(\upsilon_t\upsilon_t')\right)\left(\text{vech}(\upsilon_t\upsilon_t')\right)'\right]D_n',$$

the stated result is obtained from the preceding calculations and the definition of the matrix J_0.

To show the positive definiteness of the matrix J_0, note first that J_0 is clearly symmetric. From the definition of $\boldsymbol{i}_0$ and (B.1) we find that, even when ϵ_t is Gaussian, $\boldsymbol{i}_0 > \left\{\mathbb{E}\left[\rho_t^2 h_0\left(\rho_t^2\right)\right]\right\}^2 = n^2/4$ where the inequality is strict because ρ_t^2 has positive density. Now, let x be a nonzero $n \times 1$ vector and conclude from the preceding inequality and the definition of J_0 that

$$\begin{aligned} 4x'J_0x &> n^2x'\mathbb{E}\left[\left(\text{vech}(\upsilon_t\upsilon_t')\right)\left(\text{vech}(\upsilon_t\upsilon_t')\right)'\right]x - x'\text{vech}\left(I_n\right)\text{vech}\left(I_n\right)'x \\ &= n^2x'\mathbb{C}\left(\text{vech}(\upsilon_t\upsilon_t')\right)x, \end{aligned}$$

where the equality is justified by $\mathbb{E}\left[\text{vech}(\upsilon_t\upsilon_t')\right] = n^{-1}\text{vech}(I_n)$. Because the last quadratic form is clearly nonnegative, the positive definiteness of J_0 follows.

For the case $t = k$, $i = j \neq 0$ we have by independence $\mathbb{E}\left(\varepsilon_{t-i} \otimes e_{0t}\right) = \mathbb{E}\left(\varepsilon_{t-i}\right) \otimes \mathbb{E}\left(e_{0t}\right) = 0$. Thus, by (B.5) and arguments already used,

$$\mathbb{C}\left(\varepsilon_{t-i} \otimes e_{0t}, \varepsilon_{t-i} \otimes e_{0t}\right) = \mathbb{E}\left(\rho_{t-i}^2\right)\mathbb{E}\left[\rho_t^2\left(h_0\left(\rho_t^2\right)\right)^2\right]\left[\mathbb{E}\left(\upsilon_{t-i}\upsilon_{t-i}'\right) \otimes \mathbb{E}\left(\upsilon_t\upsilon_t'\right)\right].$$

The stated result is obtained from this by using definitions and $\mathbb{E}\left(\upsilon_t\upsilon_t'\right) = n^{-1}I_n$.

In the case $t \neq k$, $i = t - k$, and $j = k - t$ we have $i \neq 0 \neq j$ and, as in the preceding case, $\mathbb{E}\left(\varepsilon_k \otimes e_{0t}\right) = 0$. We also note that $\varepsilon_t \otimes e_{0k} = K_{nn}(e_{0k} \otimes \varepsilon_t)$ (see Result 9.2.2(3) in

Lütkepohl (1996)). As before, we now obtain

$$\begin{aligned}
\mathbb{C}\left(\varepsilon_k \otimes e_{0t}, \varepsilon_t \otimes e_{0k}\right) &= \mathbb{C}\left(\varepsilon_k \otimes e_{0t}, K_{nn}\left(e_{0k} \otimes \varepsilon_t\right)\right) \\
&= \mathbb{E}\left[\left(\rho_k \upsilon_k \otimes \rho_t h_0\left(\rho_t^2\right)\upsilon_t\right)\left(\rho_k h_0\left(\rho_k^2\right)\upsilon_k' \otimes \rho_t \upsilon_t'\right)\right] K_{nn}' \\
&= \left\{\mathbb{E}\left[\rho_t^2 h_0\left(\rho_t^2\right)\right]\right\}^2 \left\{\mathbb{E}\left(\upsilon_k \upsilon_k'\right) \otimes \mathbb{E}\left(\upsilon_t \upsilon_t'\right)\right\} K_{nn}' \\
&= \frac{1}{4} K_{nn},
\end{aligned}$$

where the last equality follows from (B.1), the symmetry of the commutation matrix K_{nn}, and the fact $\mathbb{E}\left(\upsilon_t \upsilon_t'\right) = n^{-1} I_n$.

Finally, in the last case the stated results follows from independence. □

Now we can prove Proposition 1.

Proof of Proposition 1. The proof consists of three steps. In the first one we show that the expectation of the score of θ at the true parameter value is zero and its limiting covariance matrix is $\mathcal{I}_{\theta\theta}\left(\theta_0\right)$. The positive definiteness of $\mathcal{I}_{\theta\theta}\left(\theta_0\right)$ is established in the second step and the third step proves the asymptotic normality of the score.

Step 1. We consider the different blocks of $\mathcal{I}_{\theta\theta}\left(\theta_0\right)$ separately and, to simplify notation, we set $N = T - s - nr$. In what follows, frequent use will be made of the identity $\left(f'\left(\epsilon_t'\Sigma_0^{-1}\epsilon_t;\lambda_0\right)/f\left(\epsilon_t'\Sigma_0^{-1}\epsilon_t;\lambda_0\right)\right)\Sigma_0^{-1}\epsilon_t = \Sigma_0^{-1/2} e_{0t}$ (see (A.2)).

Block $\mathcal{I}_{\vartheta_1\vartheta_1}\left(\theta_0\right)$. From the definitions and (3) it can be seen that $U_{0,t-1}$ and e_{0t} are independent. Thus, (B.3), (A.7), and straightforward calculation give $\mathbb{E}\left(\partial g_t\left(\theta_0\right)/\partial\vartheta_1\right) = 0$ and, furthermore,

$$\mathbb{C}\left(N^{-1/2}\sum_{t=r+1}^{T-s-(n-1)r}\frac{\partial}{\partial\vartheta_1}g_t\left(\theta_0\right)\right) = \nabla_1\left(\vartheta_{10}\right)' C_{11}\left(\theta_0\right)\nabla_1\left(\vartheta_{10}\right) = \mathcal{I}_{\vartheta_1\vartheta_1}\left(\theta_0\right).$$

Block $\mathcal{I}_{\vartheta_2\vartheta_2}\left(\theta_0\right)$. Deriving $\mathcal{I}_{\vartheta_2\vartheta_2}\left(\theta_0\right)$ is somewhat complicated. From the expression of $\partial g_t\left(\theta_0\right)/\vartheta_2$ (see (A.8)) it may not be quite immediate that the expectation of the score of ϑ_2 is zero so that we shall first demonstrate this. Recall that $\Phi\left(z\right)^{-1} = L\left(z\right) = \sum_{j=0}^{\infty} L_j z^j$ with $L_0 = I_n$ and, $L_j = 0$, $j < 0$. Similarly to the notation M_{j0}, N_{j0}, and Ψ_{j0} we shall also write L_{j0} when L_j is based on true parameter values. Equating the coefficient matrices related to the same powers of z in the identity $L\left(z^{-1}\right) = \Psi\left(z\right)\Pi\left(z\right)$ (see the discussion below (5)) one readily obtains

$$-\sum_{i=0}^{r}\Psi_{j-i,0}\Pi_{i0}=\begin{cases}0, & j>0\\ I_n, & j=0\\ L_{-j0}, & j<0,\end{cases}\tag{B.6}$$

where, as before, $\Pi_{00}=-I_n$. To simplify notation we also denote

$$A_0\left(k,i\right)=\Psi_{k0}\Sigma_0^{1/2}\otimes\Pi_{i0}'\Sigma_0^{-1/2}\quad\text{and}\quad B_0\left(d\right)=M_{d0}\Sigma_0^{1/2}\otimes\Sigma_0^{-1/2}.$$

Notice that from (B.6) we find that

$$\sum_{i=0}^{r}A_0\left(a-i,i\right)\text{vec}\left(I_n\right)=\text{vec}\left(\sum_{i=0}^{r}\Pi_{i0}'\Psi_{a-i,0}'\right)=0,\quad a\in\left\{1,...,s\right\}.\tag{B.7}$$

Now recall that the matrix $Y_{0,t+1}$ consists of the blocks $\sum_{i=0}^{r}\left(y_{t+a-i}\otimes\Pi_{i0}'\right)$, $a\in\{1,...,s\}$, and consider the expectation

$$\begin{aligned}\mathbb{E}\left(\sum_{i=0}^{r}\left(y_{t+a-i}\otimes\Pi_{i0}'\right)\Sigma_0^{-1/2}e_{0t}\right)&=\sum_{i=0}^{r}\sum_{k=-\infty}^{\infty}\mathbb{E}\left(\left(\Psi_{k0}\epsilon_{t+a-i-k}\otimes\Pi_{i0}'\Sigma_0^{-1/2}\right)e_{0t}\right)\\&=\sum_{i=0}^{r}\sum_{k=-\infty}^{\infty}A_0\left(k,i\right)\mathbb{E}\left(\varepsilon_{t+a-i-k}\otimes e_{0t}\right),\end{aligned}$$

where the former equality is based on (5) and the latter on the definition of $A_0\left(k,i\right)$ and the definition $\varepsilon_t=\Sigma_0^{-1/2}\epsilon_t$. By Lemma 3, the expectation in the last expression equals zero if $k\neq a-i$ and $-\frac{1}{2}\text{vec}(I_n)$ if $k=a-i$. From this and (B.7) we find that

$$\mathbb{E}\left(\sum_{i=0}^{r}\left(y_{t+a-i}\otimes\Pi_{i0}'\right)\Sigma_0^{-1/2}e_{0t}\right)=-\frac{1}{2}\sum_{i=0}^{r}A_0\left(a-i,i\right)\text{vec}\left(I_n\right)=0.$$

This in conjunction with (13) and (A.8) shows that $\mathbb{E}\left(\partial l_T\left(\theta_0\right)/\partial\vartheta_2\right)=0$, and we proceed to the covariance matrix of the score of ϑ_2.

Let $\mathbf{1}\left(\cdot\right)$ stand for the indicator function and, for $a,b\in\{1,...,s\}$, consider the covariance matrix

$$\begin{aligned}
&\mathbb{C}\left(\sum_{i=0}^{r}\left(y_{t+a-i}\otimes\Pi_{i0}'\right)\Sigma_0^{-1/2}e_{0t},\sum_{j=0}^{r}\left(y_{k+b-j}\otimes\Pi_{i0}'\right)\Sigma_0^{-1/2}e_{0k}\right)\\
=&\sum_{c,d=-\infty}^{\infty}\sum_{i,j=0}^{r}A_0\left(c,i\right)\mathbb{C}\left(\left(\varepsilon_{t+a-i-c}\otimes e_{0t}\right),\left(\varepsilon_{k+b-j-d}\otimes e_{0k}\right)\right)A_0\left(d,j\right)'\\
=&\frac{\boldsymbol{\tau}_0}{4}\sum_{\substack{c=-\infty\\ c\neq 0}}^{\infty}\sum_{i,j=0}^{r}A_0\left(c+a-i,i\right)A_0\left(c+b-j,j\right)'\mathbf{1}\left(t=k\right)\\
&+\frac{1}{4}\sum_{i,j=0}^{r}A_0\left(t-k+a-i,i\right)K_{nn}A_0\left(k-t+b-j,j\right)'\mathbf{1}\left(t\neq k\right)\\
&+\sum_{i,j=0}^{r}A_0\left(a-i,i\right)D_nJ_0D_n'A_0\left(b-j,j\right)'\mathbf{1}\left(t=k\right).
\end{aligned}$$

Here the former equality is again obtained by using (5) and the definition of $A_0\left(k,i\right)$ whereas the latter is justified by Lemma 4. Summing the last expression over $t,k=r+1,...,T-s-(n-1)\,r$, multiplying by $4/N$, and letting T tend to infinity yields the matrix $C_{22}\left(a,b;\theta_0\right)$ (see (A.8) and the definition of $\mathcal{I}_{\vartheta_2\vartheta_2}\left(\theta_0\right)$). Thus,

$$\begin{aligned}
C_{22}\left(a,b;\theta_0\right)=&\ \boldsymbol{\tau}_0\sum_{\substack{k=-\infty\\ k\neq 0}}^{\infty}\sum_{i=0}^{r}A_0\left(k+a-i,i\right)\sum_{j=0}^{r}A_0\left(k+b-j,j\right)'\\
&+\sum_{\substack{k=-\infty\\ k\neq 0}}^{\infty}\sum_{i=0}^{r}A_0\left(k+a-i,i\right)K_{nn}\sum_{j=0}^{r}A_0\left(-k+b-j,j\right)'\\
&+4\sum_{i=0}^{r}A_0\left(a-i,i\right)D_nJ_0D_n'\sum_{j=0}^{r}A_0\left(b-j,j\right)'.
\end{aligned}\tag{B.8}$$

To see that the right hand side really equals the expression given in the main text, we have to show that the second term vanishes when the range of summation is changed to $k=0,\pm1,\pm2,...$, or that

$$\sum_{k=-\infty}^{\infty}\sum_{i,j=0}^{r}\left(\Psi_{k+a-i,0}\Sigma_0^{1/2}\otimes\Pi_{i0}'\Sigma_0^{-1/2}\right)K_{nn}\left(\Sigma_0^{1/2}\Psi_{-k+b-j,0}'\otimes\Sigma_0^{-1/2}\Pi_{j0}\right)=0.$$

To see this, notice that $(\Psi_{k+a-i,0}\Sigma_0^{1/2}\otimes\Pi_{i0}'\Sigma_0^{-1/2})K_{nn}=K_{nn}(\Pi_{i0}'\Sigma_0^{-1/2}\otimes\Psi_{k+a-i,0}\Sigma_0^{1/2})$ (see Lütkepohl (1996), Result 9.2.2 (5)(a)). Thus, the left hand side of the preceding equality can be written as

$$
\begin{aligned}
K_{nn}\sum_{k=-\infty}^{\infty}\sum_{i,j=0}^{r}\left(\Pi_{i0}'\Psi_{-k+b-j,0}'\otimes\Psi_{k+a-i,0}\Pi_{j0}\right) &= K_{nn}\sum_{l=-\infty}^{\infty}\sum_{j=0}^{r}\left(\sum_{i=0}^{r}\Pi_{i0}'\Psi_{-l+a+b-j-i,0}'\otimes\Psi_{l,0}\Pi_{j0}\right)\\
&= K_{nn}\sum_{l=-\infty}^{\infty}\sum_{j=0}^{r}\left(L_{l-a-b+j,0}'\otimes\Psi_{l,0}\Pi_{j0}\right)\\
&= K_{nn}\sum_{k=0}^{\infty}\left(L_{k,0}'\otimes\sum_{j=0}^{r}\Psi_{k+a+b-j,0}\Pi_{j0}\right)\\
&= 0.
\end{aligned}
$$

Here the second and fourth equalities are obtained from (B.6) (because $a,\, b > 0$).

From (A.8), the definition of $A_0(c,i)$, and the preceding derivations it follows that the covariance matrix of the score of ϑ_2 divided by N converges to $\mathcal{I}_{\vartheta_2\vartheta_2}(\theta_0)$.

Block $\mathcal{I}_{\vartheta_1\vartheta_2}(\theta_0)$. Let $a \in \{1,...,r\}$ and $b \in \{1,...,s\}$. Using (3) and (5), and the previously introduced notation $A_0(k,i)$ and $B_0(k)$ ($B_0(k) = 0$ for $k < 0$) we consider

$$
\begin{aligned}
&\mathbb{C}\left((u_{0,t-a}\otimes I_n)\,\Sigma_0^{-1/2}e_{0t},\sum_{i=0}^{r}\left(y_{k+b-i}\otimes\Pi_{i0}'\right)\Sigma_0^{-1/2}e_{0k}\right)\\
=\;&\sum_{c=0}^{\infty}\sum_{d=-\infty}^{\infty}\sum_{i=0}^{r}B_0(c)\,\mathbb{C}\left((\varepsilon_{t-a-c}\otimes e_{0t}),(\varepsilon_{k+b-i-d}\otimes e_{0k})\right)A_0(d,i)'\\
=\;&\frac{\boldsymbol{\tau}_0}{4}\sum_{c=a}^{\infty}\sum_{i=0}^{r}B_0(c-a)\,A_0(c+b-i,i)'\,\mathbf{1}(t=k)\\
&+\frac{1}{4}\sum_{i=0}^{r}B_0(t-k-a)\,K_{nn}A_0(k-t+b-i,i)'\,\mathbf{1}(t\neq k),
\end{aligned}
$$

where the latter equality is based on Lemma 4. Summing over $t,k = r+1,...,T-s-(n-1)\,r$, multiplying by $-4/N$, and letting T tend to infinity yields the matrix $C_{12}(a,b;\theta_0)$ (see (A.7), (A.8) and the definition of $\mathcal{I}_{\vartheta_1\vartheta_2}(\theta_0)$). Thus,

$$
\begin{aligned}
C_{12}(a,b;\theta_0) &= -\boldsymbol{\tau}_0\sum_{c=a}^{\infty}\sum_{i=0}^{r}B_0(c-a)\,A_0(c+b-i,i)'\\
&\quad-\sum_{c=a}^{\infty}\sum_{i=0}^{r}B_0(c-a)\,K_{nn}A_0(-c+b-i,i)'. \qquad \text{(B.9)}
\end{aligned}
$$

It is easy to see that the first term on the right hand side equals the the first term on the right hand side of the defining equation of $C_{12}(a,b;\theta_0)$. To show the same for the second term, we need to show that

$$-K_{nn}\left(\Psi'_{b-a,0}\otimes I_n\right)=-\sum_{c=a}^{\infty}\sum_{i=0}^{r}\left(M_{c-a,0}\Sigma_0^{1/2}\otimes\Sigma_0^{-1/2}\right)K_{nn}\left(\Sigma_0^{1/2}\Psi'_{-c+b-i,0}\otimes\Sigma_0^{-1/2}\Pi_{i0}\right).$$

Using again Result 9.2.2 (5)(a) in Lütkepohl (1996) and the convention $M_{j0}=0$, $j<0$, we can write the right hand side as

$$\begin{aligned}-K_{nn}\sum_{c=-\infty}^{\infty}\sum_{i=0}^{r}\left(\Psi'_{-c+b-i,0}\otimes M_{c-a,0}\Pi_{i0}\right) &= -K_{nn}\sum_{k=-\infty}^{\infty}\left(\Psi'_{k0}\otimes\sum_{i=0}^{r}\Pi_{i0}M_{-k-a+b-i,0}\right)\\ &= K_{nn}\left(\Psi'_{b-a,0}\otimes I_n\right).\end{aligned}$$

Here the latter equality can be justified by using the identity $\Pi(z)M(z)=I_n$ to obtain an analog of (B.6) with $\Psi_{j-i,0}$ and L_{-j0} replaced by $M_{j-i,0}$ and 0, respectively.

The preceding derivations and the definitions (see (A.7) and (A.8)) show that the covariance matrix of the scores of ϑ_1 and ϑ_2 divided by N converges to $\mathcal{I}_{\vartheta_2\vartheta_1}(\theta_0)$.

Block $\mathcal{I}_{\sigma\sigma}(\theta_0)$. First note that, by (A.9) and independence of ϵ_t, we only need to show that $\mathbb{E}\left(\partial g_t(\theta_0)/\partial\sigma\right)=0$ and $\mathbb{C}\left(\partial g_t(\theta_0)/\partial\sigma\right)=\mathcal{I}_{\sigma\sigma}(\theta_0)$. These facts can be established by writing equation (A.9) as

$$\frac{\partial}{\partial\sigma}g_t(\theta_0)=-D_n'(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2})(\varepsilon_t\otimes e_{0t}+\frac{1}{2}\mathrm{vec}(I_n)),$$

using Lemma 4 (case $t=k$ and $i=j=0$), and arguments in its proof.

Block $\mathcal{I}_{\lambda\lambda}(\theta_0)$. As in the preceding case, it suffices to show that $\mathbb{E}\left(\partial g_t(\theta_0)/\partial\lambda\right)=0$ and $\mathbb{C}\left(\partial g_t(\theta_0)/\partial\lambda\right)=\mathcal{I}_{\lambda\lambda}(\theta_0)$. For the former, conclude from (A.10) and (6) that

$$\begin{aligned}\mathbb{E}_{\lambda_0}\left(\frac{\partial}{\partial\lambda}g_t(\theta_0)\right) &= \mathbb{E}_{\lambda_0}\left(\frac{1}{f(\rho_t^2;\lambda_0)}\cdot\frac{\partial}{\partial\lambda}f(\rho_t^2;\lambda)\Big|_{\lambda=\lambda_0}\right)\\ &= \frac{\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty\zeta^{n/2-1}\frac{\partial}{\partial\lambda}f(\zeta;\lambda)\Big|_{\lambda=\lambda_0}d\zeta\\ &= \frac{\pi^{n/2}}{\Gamma(n/2)}\frac{\partial}{\partial\lambda}\int_0^\infty\zeta^{n/2-1}f(\zeta;\lambda)\,d\zeta|_{\lambda=\lambda_0}\\ &= 0.\end{aligned}$$

Here the second equality is based on the expression of the density function of ρ_t^2 (see (9)), the third one on Assumption 4(i), and the fourth one on the identity

$$\int_0^\infty\zeta^{n/2-1}f(\zeta;\lambda)\,d\zeta=\frac{\Gamma(n/2)}{\pi^{n/2}}\int f(x'x;\lambda)\,dx=\frac{\Gamma(n/2)}{\pi^{n/2}},\qquad\text{(B.10)}$$

which can be obtained as in Fang et al. (1990, p. 35).

That $\mathbb{C}\left(\partial g_t\left(\theta_0\right)/\partial\lambda\right) = \mathcal{I}_{\lambda\lambda}\left(\theta_0\right)$ is an immediate consequence of Assumption 4(ii), (A.10), (6), and the expression of the density function of ρ_t^2.

Blocks $\mathcal{I}_{\vartheta_1\sigma}\left(\theta_0\right)$ *and* $\mathcal{I}_{\vartheta_1\lambda}\left(\theta_0\right)$. That these blocks are zero follows from (A.7), (A.9), (A.10), independence of ϵ_t, and the fact that $U_{0,t-1}$ is independent of ϵ_t and has zero mean (see (3)).

Block $\mathcal{I}_{\vartheta_2\sigma}\left(\theta_0\right)$. Consider the covariance matrix (cf. the derivation of $\mathcal{I}_{\vartheta_2\vartheta_2}\left(\theta_0\right)$)

$$\begin{aligned}
&\mathbb{C}\left(\sum_{i=0}^{r}\left(y_{t+a-i}\otimes\Pi_{i0}'\right)\Sigma_0^{-1/2}e_{0t},\frac{\partial}{\partial\sigma}g_k\left(\theta_0\right)\right)\\
&= -\sum_{c=-\infty}^{\infty}\sum_{i=0}^{r}A_0\left(c,i\right)\mathbb{C}\left(\left(\varepsilon_{t+a-i-c}\otimes e_{0t}\right),\left(\varepsilon_k\otimes e_{0k}\right)\right)\left(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2}\right)D_n\\
&= -\sum_{i=0}^{r}A_0\left(a-i,i\right)D_nJ_0D_n'\left(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2}\right)D_n\mathbf{1}\left(t=k\right).
\end{aligned}$$

Here the former equality is based on (5), the definition on $A_0\left(c,i\right)$, and the expression of $\partial g_t\left(\theta_0\right)/\partial\sigma$ given in the case of block $\mathcal{I}_{\sigma\sigma}\left(\theta_0\right)$. The latter equality is due to Lemma 4. The stated expression of $\mathcal{I}_{\vartheta_2\sigma}\left(\theta_0\right)$ is a simple consequence of this, (A.8), and (A.9).

Block $\mathcal{I}_{\vartheta_2\lambda}\left(\theta_0\right)$. Similarly to the preceding case we consider the covariance matrix

$$\begin{aligned}
&\mathbb{C}\left(\sum_{i=0}^{r}\left(y_{t+a-i}\otimes\Pi_{i0}'\right)\Sigma_0^{-1/2}e_{0t},\frac{\partial}{\partial\lambda}g_k\left(\theta_0\right)\right)\\
&= \sum_{c=-\infty}^{\infty}\sum_{i=0}^{r}A_0\left(c,i\right)\mathbb{C}\left(\left(\varepsilon_{t+a-i-c}\otimes e_{0t}\right),\frac{\partial}{\partial\lambda}g_k\left(\theta_0\right)\right)\\
&= \sum_{c=-\infty}^{\infty}\sum_{i=0}^{r}A_0\left(c,i\right)\mathbb{E}\left[\left(\rho_{t+a-i-c}\upsilon_{t+a-i-c}\otimes\rho_th_0\left(\rho_t^2\right)\upsilon_t\right)\frac{1}{f_0\left(\rho_k^2\right)}\frac{\partial}{\partial\lambda'}f\left(\rho_k^2;\lambda_0\right)\right]\\
&= \sum_{c=-\infty}^{\infty}\sum_{i=0}^{r}A_0\left(c,i\right)\mathbb{E}\left(\upsilon_{t+a-i-c}\otimes\upsilon_t\right)\mathbb{E}\left[\rho_{t+a-i-c}\rho_th_0\left(\rho_t^2\right)\frac{1}{f_0\left(\rho_{k0}^2\right)}\frac{\partial}{\partial\lambda'}f\left(\rho_k^2;\lambda_0\right)\right].
\end{aligned}$$

Here the first equality is justified by (5) whereas the remaining ones are obtained from (A.10), (6), (A.3), the independence of the processes ρ_t and υ_t, and the fact that $\partial g_t\left(\theta_0\right)/\partial\lambda$ has zero mean. Thus, because $\mathbb{E}\left(\upsilon_{t+a-i-c}\otimes\upsilon_t\right) = n^{-1}\text{vec}(I_n)\,\mathbf{1}\left(c=a-i\right)$,

$$\begin{aligned}
&\mathbb{C}\left(\sum_{i=0}^{r}\left(y_{t+a-i}\otimes\Pi_{i0}'\right)\Sigma_0^{-1/2}e_{0t},\frac{\partial}{\partial\lambda}g_k\left(\theta_0\right)\right)\\
&= \frac{1}{n}\sum_{i=0}^{r}A_0\left(a-i,i\right)\text{vec}\left(I_n\right)\mathbb{E}\left(\rho_t^2\frac{h_0\left(\rho_t^2\right)}{f_0\left(\rho_t^2;\right)}\frac{\partial}{\partial\lambda'}f\left(\rho_t^2;\lambda_0\right)\right)\mathbf{1}\left(t=k\right),
\end{aligned}$$

which in conjunction with (B.7) gives the desired result $\mathcal{I}_{\vartheta_2\lambda}(\theta_0) = 0$.

Block $\mathcal{I}_{\sigma\lambda}(\theta_0)$. The employed arguments are similar to those in the cases of blocks $\mathcal{I}_{\sigma\sigma}(\theta_0)$ and $\mathcal{I}_{\lambda\lambda}(\theta_0)$. By the independence of ϵ_t it suffices to consider

$$\mathbb{C}\left(\frac{\partial}{\partial\sigma}g_t(\theta_0), \frac{\partial}{\partial\lambda}g_t(\theta_0)\right) = -D_n'\left(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2}\right)\mathbb{E}\left[(\varepsilon_t\otimes e_{0t})\frac{\partial}{\partial\lambda'}g_t(\theta_0)\right],$$

where the expectation equals (see (6), (A.3), and (A.10))

$$\mathbb{E}\left[\left(\rho_t\upsilon_t\otimes\rho_t h_0\left(\rho_t^2\right)\upsilon_t\right)\frac{1}{f_0\left(\rho_t^2\right)}\frac{\partial}{\partial\lambda'}f\left(\rho_t^2;\lambda_0\right)\right] = \mathbb{E}\left(\upsilon_t\otimes\upsilon_t\right)\mathbb{E}\left[\rho_t^2\frac{h_0\left(\rho_t^2\right)}{f_0\left(\rho_t^2;\right)}\frac{\partial}{\partial\lambda'}f\left(\rho_t^2;\lambda_0\right)\right].$$

Because $\mathbb{E}(\upsilon_t\otimes\upsilon_t) = n^{-1}\text{vec}(I_n) = n^{-1}D_n\text{vech}(I_n)$, the stated expression of $\mathcal{I}_{\sigma\lambda}(\theta_0)$ follows from the definitions and the expression of the density function of ρ_t^2 (see (9)).

Thus, we have completed the derivation of $\mathcal{I}_{\theta\theta}(\theta_0)$.

Step 2. From Assumption 5(i) it readily follows that it suffices to prove the positive definiteness of $\mathcal{I}_{\theta\theta}(\theta_0)$ when $\nabla_1(\vartheta_{10}) = I_{rn^2}$ and $\nabla_2(\vartheta_{20}) = I_{sn^2}$. First we introduce some notation. Define the $sn^2\times n^2$ and $rn^2\times n^2$ matrices

$$\underline{A}_0(k) = \left[\sum_{i=0}^{r}A_0(k+j-i,i)\right]_{j=1}^{s} \quad \text{and} \quad \underline{B}_0(k) = \left[B_0(k-i)\right]_{i=1}^{r},$$

where, as before, $A_0(k+j-i,i) = \Psi_{k+j-i,0}\Sigma_0^{1/2}\otimes\Pi_{i0}'\Sigma_0^{-1/2}$, $j = 1,...,s$, and $B_0(k-i) = M_{k-i,0}\Sigma_0^{1/2}\otimes\Sigma_0^{-1/2}$, $i = 1,...,r$. We also set

$$F_0 = \frac{\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty\zeta^{n/2}\frac{f'(\zeta;\lambda_0)}{f(\zeta;\lambda_0)}\frac{\partial}{\partial\lambda}f(\zeta;\lambda_0)\,d\zeta\cdot\text{vech}(I_n)'J_0^{-1} \quad \left(\text{d}\times\tfrac{1}{2}n(n+1)\right)$$

Let $\eta_t = [\eta_{1t}' \;\; \eta_{2t}' \;\; \eta_{3t}' \;\; \eta_{4t}']'$ be a sequence of independent and identically distributed random vectors with zero mean. The covariance matrix of η_t as well as the dimensions of its components will be specified shortly. We consider the linear process

$$x_t = \sum_{k=1}^{\infty}\underline{G}_0(k)\,\eta_t,$$

where $x_t = [x_{1t}' \;\; x_{2t}' \;\; x_{3t}' \;\; x_{4t}']'$ and

$$\underline{G}_0(k) = \begin{bmatrix} -\underline{B}_0(k) & 0 & 0 & 0 \\ \underline{A}_0(k) & \underline{A}_0(-k) & \mathbf{2}\mathbf{1}(k=1)\,\underline{A}_0(k-1)\,D_n & 0 \\ 0 & 0 & -\mathbf{1}(k=1)\,D_n'(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2})D_n & 0 \\ 0 & 0 & \mathbf{1}(k=1)\,F_0 & \mathbf{1}(k=1)\,I_{\text{d}} \end{bmatrix}$$

With an appropriate definition of the covariance matrix of η_t we have $\mathbb{C}(x_t) = \mathcal{I}_{\theta\theta}(\theta_0)$. This is achieved by assuming

$$\mathbb{C}(\eta_t) = \text{diag}\left(\begin{bmatrix} \boldsymbol{\tau}_0 I_{n^2} & K_{nn} \\ K'_{nn} & \boldsymbol{\tau}_0 I_{n^2} \end{bmatrix} \;:\; J_0 \;:\; \mathcal{I}_{\lambda\lambda}(\theta_0) - F_0 J_0 F'_0 \right),$$

where the first block defines the covariance matrix of $[\eta'_{1t} \;\; \eta'_{2t}]'$. Thus, $[\eta'_{1t} \;\; \eta'_{2t}]'$, η_{3t}, and η_{4t} are uncorrelated and the dimension of both η_{1t} and η_{2t} is $n^2 \times 1$ whereas the dimensions of η_{3t} and η_{4t} are $(n(n+1)/2) \times 1$ and $\mathrm{d} \times 1$, respectively. The dimensions of x_{it} agree with those of η_{it} $(i = 1, ..., 4)$. By straightforward calculations one can check that the equality $\mathbb{C}(x_t) = \mathcal{I}_{\theta\theta}(\theta_0)$ really holds (with $\nabla_1(\vartheta_{10}) = I_{rn^2}$ and $\nabla_2(\vartheta_{20}) = I_{sn^2}$). Here we only note that for $\mathcal{I}_{\vartheta\vartheta}(\theta_0)$ the calculations yield the expressions given for $C_{22}(a, b; \theta_0)$ and $C_{21}(a, b; \theta_0)$ in the derivation of $\mathcal{I}_{\vartheta_2\vartheta_2}(\theta_0)$ and $\mathcal{I}_{\vartheta_2\vartheta_1}(\theta_0)$ (see (B.8) and (B.9)) and that for $\mathcal{I}_{\vartheta_2\lambda}(\theta_0)$ equation (B.7) can be used.

From Lemma 1 and the fact that K_{nn} is a permutation matrix it follows that the first block of $\mathbb{C}(\eta_t)$ is positive definite. Indeed, this is implied by the positive definiteness of $\boldsymbol{\tau}_0 I_{n^2} - \boldsymbol{\tau}_0^{-1} K'_{nn} K_{nn} = \boldsymbol{\tau}_0 I_{n^2} - \boldsymbol{\tau}_0^{-1} I_{n^2}$, which clearly holds because $\boldsymbol{\tau}_0 > 1$. That J_0 is positive definite follows from Lemma 4 whereas the positive definiteness of the third block of $\mathbb{C}(\eta_t)$ holds in view of Assumption 5(ii) and the identity $\mathcal{I}_{\lambda\lambda}(\theta_0) - F_0 J_0 F'_0 = \mathcal{I}_{\lambda\lambda}(\theta_0) - \mathcal{I}_{\lambda\sigma}(\theta_0) \mathcal{I}_{\sigma\sigma}(\theta_0)^{-1} \mathcal{I}_{\sigma\lambda}(\theta_0)$, which can be checked by direct calculation. Thus, the whole covariance matrix $\mathbb{C}(\eta_t)$ is positive definite.

The preceding discussion implies that we need to show that the covariance matrix $\mathbb{C}(x_t)$ is positive definite. This holds if the infinite dimensional matrix $[\underline{G}_0(1) : \underline{G}_0(2) : \cdots]$ is of full row rank. First note that the first block of rows is readily seen to be of full row rank. Indeed, using the definition of $\underline{B}_0(k)$ it is straightforward to see that the matrix $[\underline{B}_0(1) : \cdots : \underline{B}_0(r)]$ $(rn^2 \times rn^2)$ is upper triangular with diagonal blocks $\Sigma_0^{-1/2} \otimes \Sigma_0^{-1/2}$ and, therefore, of full row rank. The last two block of rows are also linearly independent because the covariance matrix of $[x'_{3t} \;\; x'_{4t}]'$ equals that of the scores of σ and λ, which is positive definite by Assumption 5(ii). It is furthermore obvious that these two block of rows are linearly independent of the first block of rows. Thus, from the definition of $\underline{G}_0(k)$ it can be seen that it suffices to show that the infinite dimensional matrix $[\underline{A}_0(-1) : \underline{A}_0(-2) : \cdots]$ is of full row rank. We shall demonstrate that the matrix

$[\underline{A}_0(-1):\cdots:\underline{A}_0(-r-s)]$ $(sn^2\times s(s+r)n^2)$ is of full row rank. For simplicity, we do this in the special case $s=2$.

Consider the matrix product

$$[\underline{A}_0(-1):\cdots:\underline{A}_0(-r-2)]\begin{bmatrix}\Sigma_0^{-1/2}\Pi_{00}\otimes\Sigma_0^{1/2} & 0\\ \vdots & \Sigma_0^{-1/2}\Pi_{00}\otimes\Sigma_0^{1/2}\\ \Sigma_0^{-1/2}\Pi_{r0}\otimes\Sigma_0^{1/2} & \vdots\\ 0 & \Sigma_0^{-1/2}\Pi_{r0}\otimes\Sigma_0^{1/2}\end{bmatrix} \tag{B.11}$$

$$= \begin{bmatrix}\sum_{j=0}^r\left(\sum_{i=0}^r\Psi_{-j-i,0}\Pi_{i0}\otimes\Pi_{j0}'\right) & \sum_{j=0}^r\left(\sum_{i=0}^r\Psi_{-1-j-i,0}\Pi_{i0}\otimes\Pi_{j0}'\right)\\ \sum_{j=0}^r\left(\sum_{i=0}^r\Psi_{1-j-i,0}\Pi_{i0}\otimes\Pi_{j0}'\right) & \sum_{j=0}^r\left(\sum_{i=0}^r\Psi_{-j-i,0}\Pi_{i0}\otimes\Pi_{j0}'\right)\end{bmatrix}$$

$$= \begin{bmatrix}\sum_{j=0}^r\left(-L_{j0}\otimes\Pi_{j0}'\right) & \sum_{j=0}^r\left(-L_{j+1,0}\otimes\Pi_{j0}'\right)\\ \sum_{j=0}^r\left(-L_{j-1,0}\otimes\Pi_{j0}'\right) & \sum_{j=0}^r\left(-L_{j0}\otimes\Pi_{j0}'\right)\end{bmatrix},$$

where the equalities follow from the definitions and from (B.6) by direct calculation. We shall show below that the last expression, a square matrix of order $2n^2\times 2n^2$, is nonsingular. Assume this for the moment and note that the latter matrix in the product (B.11) is of full column rank $2n^2$ (because $\Pi_{00}=-I_n$). Thus, as the rank of a matrix product cannot exceed the ranks of the factors of the product, it follows that the matrix $[\underline{A}_0(-1):\cdots:\underline{A}_0(-r-2)]$ has to be of full row rank $2n^2$.

To show the aforementioned nonsingularity, it clearly suffices to show the nonsingularity of the matrix

$$\begin{bmatrix}\sum_{j=0}^r\left(-L_{j0}\otimes\Pi_{j0}'\right) & \sum_{j=0}^r\left(-L_{j+1,0}\otimes\Pi_{j0}'\right)\\ \sum_{j=0}^r\left(-L_{j-1,0}\otimes\Pi_{j0}'\right) & \sum_{j=0}^r\left(-L_{j0}\otimes\Pi_{j0}'\right)\end{bmatrix}\begin{bmatrix}I_{n^2} & -\Phi_{10}\otimes I_n\\ 0 & I_{n^2}\end{bmatrix}$$

$$= \begin{bmatrix}I_n & L_{10}-\Phi_{10}\\ 0 & I_n\end{bmatrix}\otimes I_n-\sum_{j=1}^r\left(\begin{bmatrix}L_{j0} & L_{j+1,0}-L_{j0}\Phi_{10}\\ L_{j-1,0} & L_{j,0}-L_{j-1,0}\Phi_{10}\end{bmatrix}\otimes\Pi_{j0}'\right)$$

$$= \begin{bmatrix}I_n & 0\\ 0 & I_n\end{bmatrix}\otimes I_n-\sum_{j=1}^r\left(\begin{bmatrix}L_{j0} & L_{j-1,0}\Phi_{20}\\ L_{j-1,0} & L_{j-2,0}\Phi_{20}\end{bmatrix}\otimes\Pi_{j0}'\right).$$

As in the proof of proof of the nonsingularity of the matrix $\boldsymbol{H}_1$, we have here used the identity $L_{j0}=L_{j-1,0}\Phi_{10}+L_{j-2,0}\Phi_{20}$ with $L_{00}=I_n$ and $L_{j0}=0$, $j<0$, as well as direct calculation. In the same way as in that proof, we can now show the nonsingularity of the last matrix by using the fact that this matrix can be expressed as

$$I_{n^2}\otimes I_n-\sum_{j=1}^{r}\left(\mathbf{\Phi}_0^j\otimes\Pi_{j0}'\right)=\left(\mathbf{P}_0\otimes I_n\right)\left(I_{n^2}\otimes I_n-\sum_{j=1}^{r}\left(\mathbf{D}_0^j\otimes\Pi_{j0}'\right)\right)\left(\mathbf{P}_0^{-1}\otimes I_n\right),$$

where $\mathbf{\Phi}_0$ is the companion matrix corresponding the matrix polynomial $I_n-\Phi_{10}z-\Phi_{20}z^2$ and $\mathbf{\Phi}_0$= $\mathbf{P}_0\mathbf{D}_0\mathbf{P}_0^{-1}$ is its Jordan decomposition (cf. the aforementioned previous proof). The determinant of the matrix on the right hand side of the preceding equation is a product of determinants of the form $\det\left(I_n-\sum_{j=1}^{r}\Pi_{j0}'\nu^j\right)$ where ν signifies an eigenvalue of $\mathbf{\Phi}_0$. These determinants are nonzero because, by the latter condition in (2), the eigenvalues of $\mathbf{\Phi}_0$ are smaller than one in absolute value whereas the former condition in (2) implies that the zeros of $\det\Pi\left(z\right)$ lie outside the unit disc. This completes the proof of the positive definiteness of $\mathcal{I}_{\theta\theta}\left(\theta_0\right)$.

Step 3. The asymptotic normality can be proved in the same way as in previous univariate models (see Proposition 2 of Breidt et al. (1991)). The idea is to use (3) and (5) to approximate the processes $u_{t-i}\left(\vartheta_{10}\right)$ and y_{t+j-i} $(i=1,...,r,\ j=1,...,s)$ in $\partial g_t\left(\theta_0\right)/\partial\vartheta_1$ and $\partial g_t\left(\theta_0\right)/\partial\vartheta_1$, respectively, by long moving averages. This amounts to replacing $\partial g_t\left(\theta_0\right)/\partial\theta$ by a finitely dependent stationary and ergodic process with finite second moments. As is well known, a central limit theorem holds for such a process. The stated asymptotic normality can then be established by using a standard result to deal with the approximation error (see, e.g., Hannan (p. 242)). As in the aforementioned univariate case, one can here make use of the fact that the coefficient matrices in (3) and (5) decay to zero at a geometric. Details are omitted. □

Proof of Lemma 2. In the same way as in the proof of Step 1 of Proposition 1 we consider the different blocks of $\mathcal{I}_{\theta\theta}\left(\theta_0\right)$ separately. For simplicity, we again suppress the subscript from the expectation operator and denote $\mathbb{E}\left(\cdot\right)$ instead of $\mathbb{E}_{\theta_0}\left(\cdot\right)$.

Block $\mathcal{I}_{\vartheta_1\vartheta_1}\left(\theta_0\right)$. Using the independence of $u_{0,t-i}$ $(i>0)$ and e_{0t} along with (B.3) it can be seen that the first term on the right hand side of (A.12) evaluated at $\theta=\theta_0$ has zero expectation. Thus, it suffices to consider the expectation of the second term. To this end, recall the notation $\varepsilon_t=\Sigma_0^{-1/2}\epsilon_t$ and define

$$W_{\vartheta_1\vartheta_1}^{(1)}\left(a,b\right)=2\mathbb{E}\left[h_0\left(\varepsilon_t'\varepsilon_t\right)\left(u_{0,t-a}u_{0,t-b}'\otimes\Sigma_0^{-1}\right)\right],$$

$$W_{\vartheta_1\vartheta_1}^{(2)}(a,b) = 4\mathbb{E}\left[\frac{f_0''(\varepsilon_t'\varepsilon_t)}{f_0(\varepsilon_t'\varepsilon_t)}\left(u_{0,t-a}u_{0,t-b}'\otimes\Sigma_0^{-1}\epsilon_t\epsilon_t'\Sigma_0^{-1}\right)\right],$$

and

$$W_{\vartheta_1\vartheta_1}^{(3)}(a,b) = -4\mathbb{E}\left[\left(h_0(\varepsilon_t'\varepsilon_t)\right)^2\left(u_{0,t-a}u_{0,t-b}'\otimes\Sigma_0^{-1}\epsilon_t\epsilon_t'\Sigma_0^{-1}\right)\right].$$

Using these definitions in conjunction with (A.11), (A.1), and (A.5) we can write the aforementioned expectation (see (A.12)) as

$$\begin{aligned}
& -2\sum_{a=1}^{r}\frac{\partial}{\partial\vartheta_1}\pi_a(\vartheta_{10})\,\mathbb{E}\left[(u_{0,t-a}\otimes I_n)\,\Sigma_0^{-1/2}\frac{\partial}{\partial\vartheta_1'}e_t(\theta_0)\right]\\
= & -2\sum_{a=1}^{r}\frac{\partial}{\partial\vartheta_1}\pi_a(\vartheta_{10})\,\mathbb{E}\left[h_0(\varepsilon_t'\varepsilon_t)(u_{0,t-a}\otimes I_n)\,\Sigma_0^{-1}\frac{\partial}{\partial\vartheta_1'}\epsilon_t(\vartheta_0)\right]\\
& -4\sum_{a=1}^{r}\frac{\partial}{\partial\vartheta_1}\pi_a(\vartheta_{10})\,\mathbb{E}\left[\frac{f_0''(\varepsilon_t'\varepsilon_t)}{f_0(\varepsilon_t'\varepsilon_t)}(u_{0,t-a}\otimes I_n)\,\Sigma_0^{-1}\epsilon_t\epsilon_t'\Sigma_0^{-1}\frac{\partial}{\partial\vartheta_1'}\epsilon_t(\vartheta_0)\right]\\
& +4\sum_{a=1}^{r}\frac{\partial}{\partial\vartheta_1}\pi_a(\vartheta_{10})\,\mathbb{E}\left[\left(h_0(\varepsilon_t'\varepsilon_t)\right)^2(u_{0,t-a}\otimes I_n)\,\Sigma_0^{-1}\epsilon_t\epsilon_t'\Sigma_0^{-1}\frac{\partial}{\partial\vartheta_1'}\epsilon_t(\vartheta_0)\right]\\
= & \sum_{a,b=1}^{r}\frac{\partial}{\partial\vartheta_1}\pi_a(\vartheta_{10})\left[W_{\vartheta_1\vartheta_1}^{(1)}(a,b)+W_{\vartheta_1\vartheta_1}^{(2)}(a,b)+W_{\vartheta_1\vartheta_1}^{(3)}(a,b)\right]\frac{\partial}{\partial\vartheta_1'}\pi_b(\vartheta_{10}).
\end{aligned}$$

We need to show that the last expression equals $-\mathcal{I}_{\vartheta_1\vartheta_1}(\theta_0)$, which follows if $\sum_{i=1}^{3}W_{\vartheta_1\vartheta_1}^{(i)}(a,b) = -C_{11}(a,b)\otimes\Sigma_0^{-1}$. To see this, conclude from the definitions, (6), and the fact $\mathbb{C}(\upsilon_t) = n^{-1}I_n$ that

$$W_{\vartheta_1\vartheta_1}^{(1)}(a,b)+W_{\vartheta_1\vartheta_1}^{(2)}(a,b) = 2\left[\mathbb{E}\left(h_0\left(\rho_t^2\right)\right)+\frac{2}{n}\mathbb{E}\left(\rho_t^2\frac{f_0''\left(\rho_t^2\right)}{f_0\left(\rho_t^2\right)}\right)\right]\left(\mathbb{E}\left(u_{0,t-a}u_{0,t-b}'\right)\otimes\Sigma_0^{-1}\right).$$

Using definitions and the expression of the density of ρ_t^2 (see (9)) yields

$$\begin{aligned}
& \mathbb{E}\left(h_0\left(\rho_t^2\right)\right)+\frac{2}{n}\mathbb{E}\left(\rho_t^2\frac{f_0''\left(\rho_t^2\right)}{f_0\left(\rho_t^2\right)}\right) \qquad \text{(B.12)}\\
= & \frac{\pi^{n/2}}{\Gamma(n/2)}\left(\int_0^\infty\zeta^{n/2-1}f_0'(\zeta)\,d\zeta+\frac{2}{n}\int_0^\infty\zeta^{n/2}f_0''(\zeta)\,d\zeta\right)\\
= & \frac{\pi^{n/2}}{\Gamma(n/2)}\left(\int_0^\infty\zeta^{n/2-1}f_0'(\zeta)\,d\zeta+\frac{2}{n}\zeta^{n/2}f_0'(\zeta)\,|_0^\infty-\int_0^\infty\zeta^{n/2-1}f_0'(\zeta)\,d\zeta\right)\\
= & 0,
\end{aligned}$$

where the last two equalities are justified by Assumption 6(i). Thus, we can conclude that $W_{\vartheta_1\vartheta_1}^{(1)}(a,b)+W_{\vartheta_1\vartheta_1}^{(2)}(a,b)=0$.

Regarding $W^{(3)}_{\vartheta_1\vartheta_1}(a,b)$, use again (6) and the fact $\mathbb{C}(\upsilon_t) = n^{-1}I_n$ to obtain

$$\begin{aligned} W^{(3)}_{\vartheta_1\vartheta_1}(a,b) &= -\frac{4}{n}\mathbb{E}\left[\rho_t^2\left(h_0\left(\rho_t^2\right)\right)^2\right]\mathbb{E}\left(u_{0,t-a}u'_{0,t-b}\right)\otimes\Sigma_0^{-1} \\ &= -\boldsymbol{j}_0\mathbb{E}\left(u_{0,t-a}u'_{0,t-b}\right)\otimes\Sigma_0^{-1}, \end{aligned}$$

by the definitions of $h_0(\cdot)$ and $\boldsymbol{j}_0$ (see (10)). Thus, because $\boldsymbol{j}_0\mathbb{E}\left(u_{0,t-a}u'_{0,t-b}\right) = C_{11}(a,b)$, we have $\sum_{i=1}^3 W^{(i)}_{\vartheta_1\vartheta_1}(a,b) = C_{11}(a,b)\otimes\Sigma_0^{-1}$, as desired.

Block $\mathcal{I}_{\vartheta_2\vartheta_2}(\theta_0)$. The first term on the right hand side of (A.13) evaluated at $\theta=\theta_0$ has zero expectation by arguments entirely similar to those used to show that the expectation of $\partial g_t(\theta_0)/\partial\vartheta_2$ is zero (see the proof of Proposition 1, Block $\mathcal{I}_{\vartheta_2\vartheta_2}(\theta_0)$). Thus, it suffices to consider the second term for which we first note that

$$\begin{aligned} \mathbb{E}\left(\rho_t^4\frac{f''_0\left(\rho_t^2\right)}{f_0\left(\rho_t^2\right)}\right) &= \frac{\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty \zeta^{n/2+1}f''_0(\zeta)\,d\zeta \\ &= \frac{\pi^{n/2}}{\Gamma(n/2)}\left(\zeta^{n/2+1}f'_0(\zeta)\Big|_0^\infty - \frac{n+2}{2}\int_0^\infty \zeta^{n/2}f'_0(\zeta)\,d\zeta\right) \\ &= n(n+2)/4, \end{aligned} \tag{B.13}$$

where the last equality is justified by Assumption 6(i) and (B.1).

Next define

$$W^{(1)}_{\vartheta_2\vartheta_2}(a,b) = 2\mathbb{E}\left[h_0\left(\varepsilon'_t\varepsilon_t\right)\sum_{i,j=0}^{r}\left(y_{t+a-i}y'_{t+b-j}\otimes\Pi'_{i0}\Sigma_0^{-1}\Pi_{j0}\right)\right],$$

$$W^{(2)}_{\vartheta_2\vartheta_2}(a,b) = 4\mathbb{E}\left[\frac{f''_0\left(\varepsilon'_t\varepsilon_t\right)}{f_0\left(\varepsilon'_t\varepsilon_t\right)}\sum_{i,j=0}^{r}\left(y_{t+a-i}y'_{t+b-j}\otimes\Pi'_{i0}\Sigma_0^{-1}\epsilon_t\epsilon'_t\Sigma_0^{-1}\Pi_{j0}\right)\right]$$

and

$$W^{(3)}_{\vartheta_2\vartheta_2}(a,b) = -4\mathbb{E}\left[\left(h_0\left(\varepsilon'_t\varepsilon_t\right)\right)^2\sum_{i,j=0}^{r}\left(y_{t+a-i}y'_{t+b-j}\otimes\Pi'_{i0}\Sigma_0^{-1}\epsilon_t\epsilon'_t\Sigma_0^{-1}\Pi_{j0}\right)\right].$$

Using these definitions in conjunction with (A.11) and (A.6) the expectation of the second term on the right hand side of (A.13) evaluated at $\theta = \theta_0$ can be written as

$$
\begin{aligned}
& 2\sum_{a=1}^{s} \frac{\partial}{\partial \vartheta_2} \phi_a(\vartheta_{20}) \mathbb{E}\left[\sum_{i=0}^{r} (y_{t+a-i} \otimes \Pi'_{i0}) \Sigma_0^{-1/2} \frac{\partial}{\partial \vartheta'_2} e_t(\theta_0)\right] \\
= \; & 2\sum_{a,b=1}^{s} \frac{\partial}{\partial \vartheta_2} \phi_a(\vartheta_{20}) \mathbb{E}\left[\frac{f'_0(\varepsilon'_t\varepsilon_t)}{f_0(\varepsilon'_t\varepsilon_t)} \sum_{i,j=0}^{r} \left(y_{t+a-i} y'_{t+b-j} \otimes \Pi'_{i0}\Sigma_0^{-1}\Pi_{j0}\right)\right] \frac{\partial}{\partial \vartheta'_2} \phi_b(\vartheta_{20}) \\
& +4\sum_{a,b=1}^{s} \frac{\partial}{\partial \vartheta_2} \phi_a(\vartheta_{20}) \mathbb{E}\left[\frac{f''_0(\varepsilon'_t\varepsilon_t)}{f_0(\varepsilon'_t\varepsilon_t)} \sum_{i,j=0}^{r} \left(y_{t+a-i} y'_{t+b-j} \otimes \Pi'_{i0}\Sigma_0^{-1}\epsilon_t\epsilon'_t\Sigma_0^{-1}\Pi_{j0}\right)\right] \frac{\partial}{\partial \vartheta'_2} \phi_b(\vartheta_{20}) \\
& -4\sum_{a,b=1}^{s} \frac{\partial}{\partial \vartheta_2} \phi_a(\vartheta_{20}) \mathbb{E}\left[\left(\frac{f'_0(\varepsilon'_t\varepsilon_t)}{f_0(\varepsilon'_t\varepsilon_t)}\right)^2 \sum_{i,j=0}^{r} \left(y_{t+a-i} y'_{t+b-j} \otimes \Pi'_{i0}\Sigma_0^{-1}\epsilon_t\epsilon'_t\Sigma_0^{-1}\Pi_{j0}\right)\right] \frac{\partial}{\partial \vartheta'_2} \phi_b(\vartheta_{20}) \\
= \; & \sum_{a,b=1}^{s} \frac{\partial}{\partial \vartheta_2} \phi_a(\vartheta_{20}) \left[W^{(1)}_{\vartheta_2\vartheta_2}(a,b) + W^{(2)}_{\vartheta_2\vartheta_2}(a,b) + W^{(3)}_{\vartheta_2\vartheta_2}(a,b)\right] \frac{\partial}{\partial \vartheta'_2} \phi_b(\vartheta_{20}).
\end{aligned}
$$

Thus, to show that the last expression equals $-\mathcal{I}_{\vartheta_2\vartheta_2}(\theta_0)$ it suffices to show that $\sum_{i=1}^{3} W^{(i)}_{\vartheta_2\vartheta_2}(a,b) = -C_{22}(a,b,;\theta_0)$. To this end, first note that, by (5),

$$
\begin{aligned}
W^{(1)}_{\vartheta_2\vartheta_2}(a,b) &= 2\sum_{i,j=0}^{r} \sum_{c,d=-\infty}^{\infty} \mathbb{E}\left[h_0(\varepsilon'_t\varepsilon_t)\left(\Psi_{c0}\epsilon_{t+a-i-c}\epsilon'_{t+b-j-d}\Psi'_{d0} \otimes \Pi'_{i0}\Sigma_0^{-1}\Pi_{j0}\right)\right] \\
&= \frac{2}{n}\mathbb{E}\left(\rho_t^2\right)\mathbb{E}\left(h_0(\varepsilon'_t\varepsilon_t)\right) \sum_{i,j=0}^{r} \sum_{\substack{c=-\infty \\ c\neq 0}}^{\infty} A_0(c+a-i,i) A_0(c+b-j,j) \\
&\qquad - \sum_{i,j=0}^{r} A_0(a-i,i) A_0(b-j,j),
\end{aligned}
$$

where, as before, $\Psi_{k0}\Sigma_0^{1/2} \otimes \Pi'_{i0}\Sigma_0^{-1/2} = A_0(k,i)$. The latter equality is a straightforward consequence of (6), (B.1), and the fact $\mathbb{C}(\upsilon_t) = n^{-1}I_n$.

For $W^{(2)}_{\vartheta_2\vartheta_2}(a,b)$ one obtains from (5)

$$
\begin{aligned}
W^{(2)}_{\vartheta_2\vartheta_2}(a,b) &= 4\sum_{i,j=0}^{r} \sum_{c,d=-\infty}^{\infty} \mathbb{E}\left[\frac{f''_0(\varepsilon'_t\varepsilon_t)}{f_0(\varepsilon'_t\varepsilon_t)}\left(\Psi_{c0}\epsilon_{t+a-i-c}\epsilon'_{t+b-j-d}\Psi'_{d0} \otimes \Pi'_{i0}\Sigma_0^{-1}\epsilon_t\epsilon'_t\Sigma_0^{-1}\Pi_{j0}\right)\right] \\
&= \frac{4}{n^2}\mathbb{E}\left(\rho_t^2\right)\mathbb{E}\left(\rho_t^2\frac{f''_0(\rho_t^2)}{f_0(\rho_t^2)}\right) \sum_{i,j=0}^{r} \sum_{\substack{c=-\infty \\ c\neq 0}}^{\infty} A_0(c+a-i,i) A_0(c+b-j,j) \\
&\quad +4\mathbb{E}\left(\rho_t^4\frac{f''_0(\rho_t^2)}{f_0(\rho_t^2)}\right) \sum_{i,j=0}^{r} A_0(a-i,i)\,\mathbb{E}\left(\upsilon_t\upsilon'_t \otimes \upsilon_t\upsilon'_t\right) A_0(b-j,j),
\end{aligned}
$$

where the latter equality is again obtained from (6) and the fact $\mathbb{C}(\upsilon_t) = n^{-1}I_n$. From

(B.12) and (B.13) we can now conclude that

$$W^{(1)}_{\vartheta_2\vartheta_2}(a,b)+W^{(2)}_{\vartheta_2\vartheta_2}(a,b) = -\sum_{i=0}^{r}\sum_{j=0}^{r}A_0(a-i,i)\,A_0(b-j,j)$$
$$+n(n+2)\sum_{i=0}^{r}\sum_{j=0}^{r}A_0(a-i,i)\,\mathbb{E}\left(\upsilon_t\upsilon_t'\otimes\upsilon_t\upsilon_t'\right)A_0(b-j,j)\,.$$

Next, arguments similar to those already used give

$$\begin{aligned}
W^{(3)}_{\vartheta_2\vartheta_2}(a,b) &= -4\sum_{i,j=0}^{r}\sum_{c,d=-\infty}^{\infty}\mathbb{E}\left[\left(h_0\left(\varepsilon_t'\varepsilon_t\right)\right)^2\left(\Psi_{c0}\epsilon_{t+a-i-c}\epsilon_{t+b-j-d}'\Psi_{d0}'\otimes\Pi_{i0}'\Sigma_0^{-1}\epsilon_t\epsilon_t'\Sigma_0^{-1}\Pi_{j0}\right)\right]\\
&= -\frac{4}{n^2}\mathbb{E}\left(\rho_t^2\right)\mathbb{E}\left[\rho_t^2\left(h_0\left(\rho_t^2\right)\right)^2\right]\sum_{i,j=0}^{r}\sum_{\substack{c=-\infty\\c\neq0}}^{\infty}A_0(c+a-i,i)\,A_0(c+b-j,j)\\
&\quad -4\mathbb{E}\left[\rho_t^4\left(h_0\left(\rho_t^2\right)\right)^2\right]\sum_{i,j=0}^{r}A_0(a-i,i)\,\mathbb{E}\left(\upsilon_t\upsilon_t'\otimes\upsilon_t\upsilon_t'\right)A_0(b-j,j)\\
&= -\boldsymbol{\tau}_0\sum_{i,j=0}^{r}\sum_{\substack{c=-\infty\\c\neq0}}^{\infty}A_0(c+a-i,i)\,A_0(c+b-j,j)\\
&\quad -4\sum_{i,j=0}^{r}A_0(a-i,i)\,D_nJ_0D_n'A_0(b-j,j)\,.
\end{aligned}$$

Here the last equality follows from the definitions of $\boldsymbol{\tau}_0$, $\boldsymbol{\imath}_0$, and J_0 (in the term involving J_0 (B.7) has also been used).

From the preceding derivations we find that

$$\begin{aligned}
\sum_{i=1}^{3}W^{(i)}_{\vartheta_2\vartheta_2}(a,b) &= -\boldsymbol{\tau}_0\sum_{i,j=0}^{r}\sum_{\substack{c=-\infty\\c\neq0}}^{\infty}A_0(c+a-i,i)\,A_0(c+b-j,j)\\
&\quad -\sum_{i,j=0}^{r}A_0(a-i,i)\left[4D_nJ_0D_n'+I_n-n(n+2)\mathbb{E}\left(\upsilon_t\upsilon_t'\otimes\upsilon_t\upsilon_t'\right)\right]A_0(b-j,j)\,.
\end{aligned}$$

That $\sum_{i=1}^{3}W^{(i)}_{\vartheta_2\vartheta_2}(a,b)=-C_{22}(a,b,;\theta_0)$ holds, can now be seen by using the identity

$$\mathbb{E}\left[\left(\mathrm{vec}(\upsilon_t\upsilon_t')\right)\left(\mathrm{vec}(\upsilon_t\upsilon_t')\right)'\right]=\frac{1}{n(n+2)}\left(I_{n^2}+K_{nn}+\mathrm{vec}(I_n)\,\mathrm{vec}(I_n)'\right) \qquad \text{(B.14)}$$

(see Wong and Wang (1992, p. 274)) and observing that the left hand side equals $\mathbb{E}\left(\upsilon_t\upsilon_t'\otimes\upsilon_t\upsilon_t'\right)$ and the impact of the term $\mathrm{vec}(I_n)\mathrm{vec}(I_n)'$ on the right hand side cancels by equality (B.7) (see the definition of $C_{22}(a,b,;\theta_0)$).

Block $\mathcal{I}_{\vartheta_1\vartheta_2}(\theta_0)$. First conclude from (A.14), (A.11), (A.6), and (6) that

$$\begin{aligned}\frac{\partial^2}{\partial\vartheta_1\partial\vartheta_2'}g_t(\theta_0) &= 2\sum_{a=1}^{r}\sum_{b=1}^{s}\frac{\partial}{\partial\vartheta_1}\pi_a(\vartheta_{10})\left(I_n\otimes\Sigma_0^{-1/2}e_t(\theta_0)\right)\left(y_{t+b-a}'\otimes I_n\right)\frac{\partial}{\partial\vartheta_2'}\phi_b(\vartheta_{20})\\ &\quad -2\sum_{a=1}^{r}\sum_{b=1}^{s}\frac{\partial}{\partial\vartheta_1}\pi_a(\vartheta_{10})h_0(\varepsilon_t'\varepsilon_t)\sum_{i=0}^{r}\left(u_{0,t-a}y_{t+b-i}'\otimes\Sigma_0^{-1}\Pi_{i0}\right)\frac{\partial}{\partial\vartheta_2'}\phi_b(\vartheta_{20})\\ &\quad -4\sum_{a=1}^{r}\sum_{b=1}^{s}\frac{\partial}{\partial\vartheta_1}\pi_a(\vartheta_{10})h_0'(\varepsilon_t'\varepsilon_t)\sum_{i=0}^{r}\left(u_{0,t-a}y_{t+b-i}'\otimes\Sigma_0^{-1}\epsilon_t\epsilon_t'\Sigma_0^{-1}\Pi_{i0}\right)\frac{\partial}{\partial\vartheta_2'}\phi_b(\vartheta_{20}).\end{aligned}$$

In the first expression on the right hand side,

$$\left(I_n\otimes\Sigma_0^{-1/2}e_t(\theta_0)\right)\left(y_{t+b-a}'\otimes I_n\right)=h_0(\varepsilon_t'\varepsilon_t)K_{nn}\left(\Sigma_0^{-1}\epsilon_t y_{t+b-a}'\otimes I_n\right)$$

by the definition of $e_t(\theta_0)$ and Result 9.2.2(3) in Lütkepohl (1996). Define

$$W_{\vartheta_1\vartheta_2}^{(1)}(a,b)=2K_{nn}\mathbb{E}\left[h_0(\varepsilon_t'\varepsilon_t)\left(\Sigma_0^{-1}\epsilon_t y_{t+b-a}'\otimes I_n\right)\right],$$

$$W_{\vartheta_1\vartheta_2}^{(2)}(a,b)=-2\mathbb{E}\left[h_0(\varepsilon_t'\varepsilon_t)\sum_{i=0}^{r}\left(u_{0,t-a}y_{t+b-i}'\otimes\Sigma_0^{-1}\Pi_{i0}\right)\right]$$

$$W_{\vartheta_1\vartheta_2}^{(3)}(a,b)=-4\mathbb{E}\left[\frac{f_0''(\varepsilon_t'\varepsilon_t)}{f_0(\varepsilon_t'\varepsilon_t)}\sum_{i=0}^{r}\left(u_{0,t-a}y_{t+b-i}'\otimes\Sigma_0^{-1}\epsilon_t\epsilon_t'\Sigma_0^{-1}\Pi_{i0}\right)\right]$$

and

$$W_{\vartheta_1\vartheta_2}^{(4)}(a,b)=4\mathbb{E}\left[(h_0(\varepsilon_t'\varepsilon_t))^2\sum_{i=0}^{r}\left(u_{0,t-a}y_{t+b-i}'\otimes\Sigma_0^{-1}\epsilon_t\epsilon_t'\Sigma_0^{-1}\Pi_{i0}\right)\right].$$

We need to show that $\sum_{i=1}^4 W_{\vartheta_1\vartheta_2}^{(i)}(a,b)=-C_{12}(a,b;\theta_0)$. The employed arguments, based mostly on (3), (5), (6), and the fact $\mathbb{C}(\upsilon_t)=n^{-1}I_n$, are similar to those used in the previous cases. First note that

$$\begin{aligned}W_{\vartheta_1\vartheta_2}^{(1)}(a,b) &= 2K_{nn}\sum_{c=-\infty}^{\infty}\mathbb{E}\left[h_0(\varepsilon_t'\varepsilon_t)\left(\Sigma_0^{-1}\epsilon_t\epsilon_{t+b-a-c}'\Psi_{c0}'\otimes I_n\right)\right]\\ &= \frac{2}{n}\mathbb{E}\left[\rho_t^2h_0(\rho_t^2)\right]K_{nn}\left(\Psi_{b-a,0}'\otimes I_n\right)\\ &= -K_{nn}\left(\Psi_{b-a,0}'\otimes I_n\right),\end{aligned}$$

where the last equality is due to (B.1). Next,

$$\begin{aligned}W_{\vartheta_1\vartheta_2}^{(2)}(a,b) &= -2\sum_{c=0}^{\infty}\sum_{d=-\infty}^{\infty}\sum_{i=0}^{r}\mathbb{E}\left[h_0(\varepsilon_t'\varepsilon_t)\left(M_{c0}\epsilon_{t-a-c}\epsilon_{t+b-i-d}'\Psi_{d0}'\otimes\Sigma_0^{-1}\Pi_{i0}\right)\right]\\ &= -\frac{2}{n}\mathbb{E}\left(\rho_t^2\right)\mathbb{E}\left(h_0(\rho_t^2)\right)\sum_{c=0}^{\infty}\sum_{i=0}^{r}\left(M_{c0}\Sigma_0\Psi_{c+a+b-i,0}'\otimes\Sigma_0^{-1}\Pi_{i0}\right)\end{aligned}$$

and

$$
\begin{aligned}
W^{(3)}_{\vartheta_1\vartheta_2}(a,b) &= -4\sum_{c=0}^{\infty}\sum_{d=-\infty}^{\infty}\sum_{i=0}^{r}\mathbb{E}\left[\frac{f_0''(\varepsilon_t'\varepsilon_t)}{f_0(\varepsilon_t'\varepsilon_t)}\left(M_{c0}\epsilon_{t-a-c}\epsilon_{t+b-i-d}'\Psi_{d0}'\otimes\Sigma_0^{-1}\epsilon_t\epsilon_t'\Sigma_0^{-1}\Pi_{i0}\right)\right] \\
&= -\frac{4}{n^2}\mathbb{E}\left(\rho_t^2\right)\mathbb{E}\left(\rho_t^2\frac{f_0''(\rho_t^2)}{f_0(\rho_t^2)}\right)\sum_{c=0}^{\infty}\sum_{i=0}^{r}\left(M_{c0}\Sigma_0\Psi_{c+a+b-i,0}'\otimes\Sigma_0^{-1}\Pi_{i0}\right).
\end{aligned}
$$

From the preceding expressions and (B.12) it is seen that $W^{(2)}_{\vartheta_1\vartheta_2}(a,b)+W^{(3)}_{\vartheta_1\vartheta_2}(a,b)=0$.

Regarding $W^{(4)}_{\vartheta_1\vartheta_2}(a,b)$, we have

$$
\begin{aligned}
W^{(4)}_{\vartheta_1\vartheta_2}(a,b) &= 4\sum_{c=0}^{\infty}\sum_{d=-\infty}^{\infty}\sum_{i=0}^{r}\mathbb{E}\left[\left(h_0(\varepsilon_t'\varepsilon_t)\right)^2\left(M_{c0}\epsilon_{t-a-c}\epsilon_{t+b-i-d}'\Psi_{d0}'\otimes\Sigma_0^{-1}\epsilon_t\epsilon_t'\Sigma_0^{-1}\Pi_{i0}\right)\right] \\
&= \frac{4}{n^2}\mathbb{E}\left(\rho_t^2\right)\mathbb{E}\left[\rho_t^2\left(h_0(\rho_t^2)\right)^2\right]\sum_{c=0}^{\infty}\sum_{i=0}^{r}\left(M_{c0}\Sigma_0\Psi_{c+a+b-i,0}'\otimes\Sigma_0^{-1}\Pi_{i0}\right) \\
&= \boldsymbol{\tau}_0\sum_{c=a}^{\infty}\sum_{i=0}^{r}\left(M_{c-a,0}\Sigma_0\Psi_{c+b-i,0}'\otimes\Sigma_0^{-1}\Pi_{i0}\right),
\end{aligned}
$$

where the last equality holds by the definitions of $h_0(\cdot)$ and $\boldsymbol{\tau}_0$. Combining the preceding derivations yields $\sum_{i=1}^{4}W^{(i)}_{\vartheta_1\vartheta_2}(a,b)=-C_{12}(a,b;\theta_0)$, as desired.

Block $\mathcal{I}_{\sigma\sigma}(\theta_0)$. From (A.15) and (6) we obtain

$$
\begin{aligned}
\frac{\partial^2}{\partial\sigma\partial\sigma'}g_t(\theta_0) &= h_0(\varepsilon_t'\varepsilon_t)\left(\epsilon_t'\otimes\epsilon_t'\otimes D_n'\right)\left(I_n\otimes K_{nn}\otimes I_n\right) \\
&\quad\times\left[\Sigma_0^{-1}\otimes\Sigma_0^{-1}\otimes\operatorname{vec}\left(\Sigma_0^{-1}\right)+\operatorname{vec}\left(\Sigma_0^{-1}\right)\otimes\Sigma_0^{-1}\otimes\Sigma_0^{-1}\right]D_n \\
&\quad+h_0'(\varepsilon_t'\varepsilon_t)D_n'(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2})\left(\varepsilon_t\varepsilon_t'\otimes\varepsilon_t\varepsilon_t'\right)(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2})D_n \\
&\qquad+\frac{1}{2}D_n'\left(\Sigma_0^{-1}\otimes\Sigma_0^{-1}\right)D_n.
\end{aligned}
$$

The first term on the right hand side consists of two additive terms. Using (6) and taking expectation the first one can be written as

$$
\begin{aligned}
&\mathbb{E}\left(\rho_t^2h_0\left(\rho_t^2\right)\right)\left(\operatorname{vec}\left(\Sigma_0^{1/2}\mathbb{E}\left(\upsilon_t\upsilon_t'\right)\Sigma_0^{1/2}\right)'\otimes D_n'\right)\left(I_n\otimes K_{nn}\otimes I_n\right) \\
&\qquad\times\left(\Sigma_0^{-1}\otimes\Sigma_0^{-1}\otimes\operatorname{vec}\left(\Sigma_0^{-1}\right)\right)D_n \\
&= -\frac{1}{2}D_n'\left(\operatorname{vec}(\Sigma_0)'\otimes I_{n^2}\right)\left(I_n\otimes K_{nn}\otimes I_n\right)\left(\Sigma_0^{-1}\otimes\Sigma_0^{-1}\otimes\operatorname{vec}\left(\Sigma_0^{-1}\right)\right)D_n \\
&= -\frac{1}{2}D_n'\left(\Sigma^{-1}\otimes\Sigma^{-1}\right)D_n.
\end{aligned}
$$

Here the former equality is based on (B.1) and the fact $\mathbb{E}(\upsilon_t\upsilon_t')=n^{-1}I_n$ whereas the latter can be seen as follows. Let B_1 and B_2 be arbitrary symmetric $(n\times n)$ matrices and consider the quantity

$$\begin{aligned}
& \operatorname{vech}(B_1)' D_n' \left(\operatorname{vec}(\Sigma_0)' \otimes I_{n^2}\right)(I_n \otimes K_{nn} \otimes I_n)\left(\Sigma_0^{-1} \otimes \Sigma_0^{-1} \otimes \operatorname{vec}\left(\Sigma_0^{-1}\right)\right) D_n \operatorname{vech}(B_2) \\
= \; & \operatorname{vec}(B_1)' \left(\operatorname{vec}(\Sigma_0)' \otimes I_{n^2}\right)(I_n \otimes K_{nn} \otimes I_n)\left(\left(\Sigma_0^{-1} \otimes \Sigma_0^{-1}\right) \operatorname{vec}(B_2) \otimes \operatorname{vec}\left(\Sigma_0^{-1}\right)\right) \\
= \; & \operatorname{vec}(B_1)' \left(\operatorname{vec}(\Sigma_0)' \otimes I_{n^2}\right)(I_n \otimes K_{nn} \otimes I_n)\left(\operatorname{vec}\left(\Sigma_0^{-1} B_2 \Sigma_0^{-1}\right) \otimes \operatorname{vec}\left(\Sigma_0^{-1}\right)\right) \\
= \; & \operatorname{vec}(B_1)' \left(\operatorname{vec}(\Sigma_0)' \otimes I_{n^2}\right) \operatorname{vec}\left(\Sigma_0^{-1} B_2 \Sigma_0^{-1} \otimes \Sigma_0^{-1}\right) \\
= \; & \operatorname{vec}(B_1)' \left(\Sigma_0^{-1} B_2 \Sigma_0^{-1} \otimes \Sigma_0^{-1}\right) \operatorname{vec}(\Sigma_0) \\
= \; & \operatorname{vec}(B_1)' \operatorname{vec}\left(\Sigma_0^{-1} B_2 \Sigma_0^{-1}\right) \\
= \; & \operatorname{vech}(B_1)' D_n' \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1}\right) D_n \operatorname{vech}(B_2).
\end{aligned}$$

Here the third equality follows from Lütkepohl (1996, Result 9.2.2(5)(c)) whereas the other equalities are due to definitions and well-known properties of the Kronecker product and vec operator (especially the result $\operatorname{vec}(ABC) = (C' \otimes A)\operatorname{vec}(B)$). Because B_1 and B_2 are arbitrary symmetric $(n \times n)$ matrices the stated result follows and in the same way it can be seen that a similar result holds for the second additive component obtained from the first term of the preceding expression of $\partial^2 g_t(\theta_0) / \partial\sigma\partial\sigma'$. Thus, we can conclude that

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial^2}{\partial\sigma\partial\sigma'} g_t(\theta_0)\right) = \; & D_n'(\Sigma_0^{-1/2} \otimes \Sigma_0^{-1/2})\mathbb{E}\left[h_0'(\varepsilon_t'\varepsilon_t)(\varepsilon_t\varepsilon_t' \otimes \varepsilon_t\varepsilon_t')\right]\left(\Sigma^{-1/2} \otimes \Sigma^{-1/2}\right) D_n \\
& - \frac{1}{2} D_n' \left(\Sigma^{-1} \otimes \Sigma^{-1}\right) D_n.
\end{aligned}$$

Using (6) and (A.1) one obtains

$$\begin{aligned}
\mathbb{E}\left[h_0'(\varepsilon_t'\varepsilon_t)(\varepsilon_t\varepsilon_t' \otimes \varepsilon_t\varepsilon_t')\right] & = \left[\mathbb{E}\left(\rho_t^4 \frac{f_0''(\rho_t^2)}{f_0(\rho_t^2)}\right) - \mathbb{E}\left(\rho_t^4 \left(h_0(\rho_t^2)\right)^2\right)\right] \mathbb{E}(\upsilon_t\upsilon_t' \otimes \upsilon_t\upsilon_t') \\
& = \frac{n(n+2)}{4} \mathbb{E}(\upsilon_t\upsilon_t' \otimes \upsilon_t\upsilon_t') - \boldsymbol{i}_0 \mathbb{E}(\upsilon_t\upsilon_t' \otimes \upsilon_t\upsilon_t'),
\end{aligned}$$

where the latter equality is based on (B.13) and the definition of $\boldsymbol{i}_0$ (see (11)). Thus,

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial^2}{\partial\sigma\partial\sigma'} g_t(\theta_0)\right) = \; & \frac{1}{4} D_n'(\Sigma_0^{-1/2} \otimes \Sigma_0^{-1/2})\left[n(n+2)\mathbb{E}(\upsilon_t\upsilon_t' \otimes \upsilon_t\upsilon_t') - 2I_{n^2}\right](\Sigma_0^{-1/2} \otimes \Sigma_0^{-1/2}) D_n \\
& - \boldsymbol{i}_0 D_n'(\Sigma_0^{-1/2} \otimes \Sigma_0^{-1/2})\mathbb{E}(\upsilon_t\upsilon_t' \otimes \upsilon_t\upsilon_t')(\Sigma_0^{-1/2} \otimes \Sigma_0^{-1/2}) D_n.
\end{aligned}$$

Because $\mathbb{E}(\upsilon_t\upsilon_t' \otimes \upsilon_t\upsilon_t') = D_n\mathbb{E}((\operatorname{vech}(\upsilon_t\upsilon_t'))(\operatorname{vech}(\upsilon_t\upsilon_t'))D_n'$ the right hand side equals $-\mathcal{I}_{\sigma\sigma}(\theta_0)$ if the expression in the brackets can be replaced by $\operatorname{vec}(I_n)\operatorname{vec}(I_n)'$. From

(B.14) it is seen that this expression can be replaced by $\text{vec}(I_n)\text{vec}(I_n)' + K_{nn} - I_{n^2}$. Thus, the desired result follows because

$$(K_{nn} - I_{n^2})\,(\Sigma_0^{-1/2} \otimes \Sigma_0^{-1/2})D_n = (\Sigma_0^{-1/2} \otimes \Sigma_0^{-1/2})\,(K_{nn} - I_{n^2})\,D_n = 0$$

by Results 9.2.2(2)(b) and 9.2.3(2) in Lütkepohl (1996).

Block $\mathcal{I}_{\lambda\lambda}(\theta_0)$. By the definition of $\mathcal{I}_{\lambda\lambda}(\theta_0)$ and (A.17) it suffices to note that

$$\mathbb{E}\left[\frac{1}{f(\rho_t^2;\lambda_0)}\frac{\partial^2}{\partial\lambda\partial\lambda'}f\left(\rho_t^2;\lambda_0\right)\right] = \frac{\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty \zeta^{n/2-1}\frac{\partial^2}{\partial\lambda\partial\lambda'}f(\zeta;\lambda_0)\,d\zeta = 0,$$

where the former equality follows from (9) and the latter from Assumption 6(ii) (cf. the corresponding part of the proof of Proposition 1, Block $\mathcal{I}_{\lambda\lambda}(\theta_0)$).

Blocks $\mathcal{I}_{\vartheta_1\sigma}(\theta_0)$ and $\mathcal{I}_{\vartheta_1\lambda}(\theta_0)$. The former is an immediate consequence of (A.16), the independence of ϵ_t and $\partial\epsilon_t(\vartheta_0)/\partial\vartheta_1$, and the fact $\mathbb{E}(\partial\epsilon_t(\vartheta_0)/\partial\vartheta_1) = 0$ (see (A.5)) which imply $\mathbb{E}(\partial^2 g_t(\theta_0)/\partial\vartheta_1\partial\sigma') = 0$.

As for $\mathcal{I}_{\vartheta_1\lambda}(\theta_0)$, it is seen from (A.18), (A.1), and (A.5) that we need to show that

$$\mathbb{E}\left[\frac{1}{f_0(\varepsilon_t'\varepsilon_t)}(u_{0,t-a}\otimes I_n)\,\Sigma_0^{-1}\epsilon_t\frac{\partial}{\partial\lambda'}f'(\varepsilon_t'\varepsilon_t;\lambda_0)\right] = 0, \quad a = 1,...,r,$$

and similarly when $1/f_0(\varepsilon_t'\varepsilon_t)$ is replaced by $f_0'(\varepsilon_t'\varepsilon_t)/(f_0(\varepsilon_t'\varepsilon_t))^2$. These facts follow from the independence of $u_{0,t-a}$ and ϵ_t and $\mathbb{E}(u_{0,t-a}) = 0$.

Block $\mathcal{I}_{\vartheta_2\sigma}(\theta_0)$. From (A.16) and (A.6) we find that

$$\begin{aligned}&\frac{\partial^2}{\partial\vartheta_2\partial\sigma'}g_t(\theta_0)\\ =\; & -2h_0(\varepsilon_t'\varepsilon_t)\sum_{b=1}^{s}\frac{\partial}{\partial\vartheta_2}\phi_b(\vartheta_{20})\sum_{a=0}^{r}(\epsilon_t'\otimes y_{t+b-a}\otimes\Pi_{a0}')\left(\Sigma_0^{-1}\otimes\Sigma_0^{-1}\right)D_n\\ & -2h_0'(\varepsilon_t'\varepsilon_t)\sum_{b=1}^{s}\frac{\partial}{\partial\vartheta_2}\phi_b(\vartheta_{20})\sum_{a=0}^{r}(y_{t+b-a}\otimes\Pi_{a0}')\,\Sigma_0^{-1}\epsilon_t(\epsilon_t'\otimes\epsilon_t')\left(\Sigma^{-1}\otimes\Sigma^{-1}\right)D_n.\end{aligned}$$

By independence of ϵ_t and equation (5), y_{t+b-a} on the right hand side can be replaced by $\Psi_{b-a,0}\epsilon_t$ when expectation is taken. Thus, using the definition of e_{t0} (see (A.2)) and straightforward calculation the expectation of the first term on the right hand side becomes

$$\begin{aligned}
& -2\sum_{b=1}^{s}\frac{\partial}{\partial\vartheta_2}\phi_b(\vartheta_{20})\sum_{a=0}^{r}\mathbb{E}\left[e_{0t}'\otimes\Psi_{b-a,0}\epsilon_t\otimes\Pi_{a0}'\Sigma_0^{-1/2}\right](\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2})D_n \\
= & -2\sum_{b=1}^{s}\frac{\partial}{\partial\vartheta_2}\phi_b(\vartheta_{20})\sum_{a=0}^{r}A_0(b-a,i)\,\mathbb{E}\left[(e_{0t}'\otimes\varepsilon_t\otimes I_n)\right](\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2})D_n \\
= & \sum_{b=1}^{s}\frac{\partial}{\partial\vartheta_2}\phi_b(\vartheta_{20})\sum_{a=0}^{r}A_0(b-a,i)(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2})D_n,
\end{aligned}$$

where, again, $A_0(b-a,i)=\Psi_{b-a0}\Sigma_0^{1/2}\otimes\Pi_{a0}'\Sigma_0^{-1/2}$ and the latter equality is due to $\mathbb{E}(e_{0t}'\otimes\varepsilon_t\otimes I_n)=\mathbb{E}(\varepsilon_t e_{0t}'\otimes I_n)=-2^{-1}I_{n^2}$ (see (B.4)).

The expectation of the second term in the preceding expression of $\partial^2 g_t(\theta_0)/\partial\vartheta_2\partial\sigma'$ can similarly be written as

$$-2\sum_{b=1}^{s}\frac{\partial}{\partial\vartheta_2}\phi_b(\vartheta_{20})\,\mathbb{E}\left[h_0'(\varepsilon_t'\varepsilon_t)\sum_{a=0}^{r}(\Psi_{b-a,0}\epsilon_t\otimes\Pi_{a0}')\Sigma_0^{-1}\epsilon_t(\varepsilon_t'\otimes\varepsilon_t')\right](\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2})D_n,$$

where, by (6) and (A.1), the expectation equals

$$\begin{aligned}
& \left\{\mathbb{E}\left[\rho_t^4\frac{f_0''(\rho_t^2)}{f_0(\rho_t^2)}\right]-\mathbb{E}\left[\rho_t^4\left(h_0(\rho_t^2)\right)^2\right]\right\}\sum_{a=0}^{r}A_0(b-a,i)\,\mathbb{E}(\upsilon_t\upsilon_t'\otimes\upsilon_t\upsilon_t') \\
= & \left(\frac{n(n+2)}{4}-\boldsymbol{i}_0\right)\sum_{a=0}^{r}A_0(b-a,i)\,\mathbb{E}(\upsilon_t\upsilon_t'\otimes\upsilon_t\upsilon_t').
\end{aligned}$$

Here we have used (B.13), the definition of $\boldsymbol{i}_0$ (see (11)), and straightforward calculation. Combining the preceding derivations shows that

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial^2}{\partial\vartheta_2\partial\sigma'}g_t(\theta_0)\right) = & \; 2\left(\boldsymbol{i}_0-\frac{n(n+2)}{4}\right)\sum_{b=1}^{s}\frac{\partial}{\partial\vartheta_2}\phi_b(\vartheta_{20})\sum_{a=0}^{r}A_0(b-a,i)\,\mathbb{E}(\upsilon_t\upsilon_t'\otimes\upsilon_t\upsilon_t') \\
& \times(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2})D_n \\
& +\sum_{b=1}^{s}\frac{\partial}{\partial\vartheta_2}\phi_b(\vartheta_{20})\sum_{a=0}^{r}A_0(b-a,i)(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2})D_n \\
= & \; 2\sum_{b=1}^{s}\frac{\partial}{\partial\vartheta_2}\phi_b(\vartheta_{20})\sum_{a=0}^{r}A_0(b-a,i)D_nJ_0D_n'(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2})D_n,
\end{aligned}$$

where the last expression equals $-\mathcal{I}_{\vartheta_2\sigma}(\theta_0)$ and the latter equality can be justified by using the definition of J_0, the identity (B.14), and arguments similar to those already used in the case of block $\mathcal{I}_{\sigma\sigma}(\theta_0)$ (see the end of that proof).

Block $\mathcal{I}_{\vartheta_2\lambda}(\theta_0)$. From (A.18) and (A.6) it is seen that we need to show that

$$\sum_{i=0}^{r}\mathbb{E}\left[\frac{1}{f_0(\varepsilon_t'\varepsilon_t)}(y_{t+a-i}\otimes\Pi_{i0}')\Sigma_0^{-1}\epsilon_t\frac{\partial}{\partial\lambda'}f'(\varepsilon_t'\varepsilon_t;\lambda_0)\right]=0,\quad a=1,...,r,$$

and

$$\sum_{i=0}^{r}\mathbb{E}\left[\frac{f_0'(\varepsilon_t'\varepsilon_t)}{(f_0(\varepsilon_t'\varepsilon_t))^2}(y_{t+a-i}\otimes\Pi_{i0}')\Sigma_0^{-1}\epsilon_t\frac{\partial}{\partial\lambda'}f(\varepsilon_t'\varepsilon_t;\lambda_0)\right]=0,\quad a=1,...,r.$$

The argument is similar in both cases and also similar to that used in the proof of Proposition 1 (see Block $\mathcal{I}_{\vartheta_2\lambda}(\theta_0)$). For example, consider the former and use (5) and independence of ϵ_t to write the left hand side of the equality as

$$\begin{aligned}&\sum_{i=0}^{r}\mathbb{E}\left[\frac{1}{f_0(\varepsilon_t'\varepsilon_t)}(\Psi_{a-i,0}\epsilon_t\otimes\Pi_{i0}')\Sigma_0^{-1}\epsilon_t\frac{\partial}{\partial\lambda'}f'(\varepsilon_t'\varepsilon_t;\lambda_0)\right]\\ =\;&\sum_{i=0}^{r}A_0(a-i,i)\,\mathbb{E}(\upsilon_t\otimes\upsilon_t)\,\mathbb{E}\left[\frac{\rho_t^2}{f_0(\rho_t^2)}\frac{\partial}{\partial\lambda'}f'(\rho_t^2;\lambda_0)\right],\end{aligned}$$

where that equality is due to (6). Because $\mathbb{E}(\upsilon_t\otimes\upsilon_t)=\text{vec}(\mathbb{E}(\upsilon_t\upsilon_t'))=n^{-1}\text{vec}(I_n)$ the last expression is zero by (B.7). A similar proof applies to the other expectation.

Block $\mathcal{I}_{\sigma\lambda}(\theta_0)$. One obtains from (A.19) that $\mathbb{E}(\partial^2 g_t(\theta_0)/\partial\sigma\partial\lambda)$ is a sum of two terms. One is

$$\begin{aligned}-D_n'(\Sigma_0^{-1}\otimes\Sigma_0^{-1})\mathbb{E}\left[\frac{1}{f_0(\varepsilon_t'\varepsilon_t)}\frac{\partial}{\partial\lambda'}f'(\varepsilon_t'\varepsilon_t;\lambda_0)\right]&=-D_n'(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2})\mathbb{E}(\upsilon_t\otimes\upsilon_t)\\&\quad\times\mathbb{E}\left[\frac{\rho_t^2}{f_0(\rho_t^2)}\frac{\partial}{\partial\lambda'}f'(\rho_t^2;\lambda_0)\right],\end{aligned}$$

where the equality is based on (6) and, using (9), the last expectation can be written as

$$\frac{\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty\zeta^{n/2}\left.\frac{\partial}{\partial\lambda'}f'(\zeta;\lambda)\right|_{\lambda=\lambda_0}d\zeta=\frac{\pi^{n/2}}{\Gamma(n/2)}\frac{\partial}{\partial\lambda'}\int_0^\infty\zeta^{n/2}f'(\zeta;\lambda)\,d\zeta|_{\lambda=\lambda_0}=0.$$

Here the former equality is justified by Assumption 6(ii) and the latter by (B.1). By similar arguments it is seen that the second term of $\mathbb{E}(\partial^2 g_t(\theta_0)/\partial\sigma\partial\lambda)$ becomes $-\mathcal{I}_{\sigma\lambda}(\theta_0)$. □

Proof of Theorem 1. First note that our Proposition 1 and Lemma 2 are analogous to Lemmas 1 and 2 of Andrews et al. (2006) so that the method of proof used in that paper also applies here. That method is based on a standard Taylor expansion and, an inspection of the arguments used by Andrews et al. (2006) in the proof of their Theorem

1, shows that we only need to show that the appropriately standardized Hessian of the log-likelihood function satisfies

$$\sup_{\theta\in\Theta_0}\left\|N^{-1}\sum_{t=r+1}^{T-s-(n-1)r}\left(\frac{\partial^2}{\partial\theta\partial\theta'}g_t(\theta)-\frac{\partial^2}{\partial\theta\partial\theta'}g_t(\theta_0)\right)\right\|\xrightarrow{p}0, \tag{B.15}$$

where Θ_0 is a small compact neighborhood of θ_0 with non-empty interior (cf. Lanne and Saikkonen (2008)). From the expressions of the components of $\partial^2 g_t(\theta)/\partial\theta\partial\theta'$ it can be checked that $\partial^2 g_t(\theta)/\partial\theta\partial\theta'$ is stationary and ergodic, and, as a function of θ, continuous. Hence, a sufficient condition for (B.15) to hold is that $\partial^2 g_t(\theta)/\partial\theta\partial\theta'$ obeys a uniform law of large numbers over Θ_0, which is turn is implied by

$$\mathbb{E}_{\theta_0}\left(\sup_{\theta\in\Theta_0}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}g_t(\theta)\right\|\right)<\infty \tag{B.16}$$

(see Theorem A.2.2 in White (1994)).

We demonstrate (B.16) for some typical components of $\partial^2 g_t(\theta)/\partial\theta\partial\theta'$ and note that the remaining components can be handled along similar lines. Of $\partial^2 g_t(\theta)/\partial\vartheta_i\partial\vartheta_j'$ $i,j\in\{1,2\}$ we only consider $\partial^2 g_t(\theta)/\partial\vartheta_1\partial\vartheta_2'$. In what follows, c_1, c_2, ... will denote positive constants. From (A.14), Assumption 3, and the definitions of the quantities involved (see (A.2), (A.11), (A.6)) it can be seen that

$$\begin{aligned}\mathbb{E}_{\theta_0}\left(\sup_{\theta\in\Theta_0}\left\|\frac{\partial^2}{\partial\vartheta_1\partial\vartheta_2'}g_t(\theta)\right\|\right) &\leq c_1\mathbb{E}_{\theta_0}\left(\sup_{\theta\in\Theta_0}\left\|e_t(\theta)\right\|\sum_{i=1}^{r}\left\|\frac{\partial}{\partial\vartheta_2}u_{t-i}(\vartheta_2)\right\|\right)\\ &\quad+c_2\mathbb{E}_{\theta_0}\left(\sup_{\theta\in\Theta_0}\sum_{i=1}^{r}\left\|u_{t-i}(\vartheta_2)\right\|\left\|\frac{\partial}{\partial\vartheta_2}e_t(\theta)\right\|\right)\\ &\leq c_3\mathbb{E}_{\theta_0}\left(\left\|y_t\right\|^2\sup_{\theta\in\Theta_0}\left|h\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)\right|\right)\\ &\quad+c_4\mathbb{E}_{\theta_0}\left(\left\|y_t\right\|^4\sup_{\theta\in\Theta_0}\left|h'\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)\right|\right).\end{aligned}$$

The finiteness of the last two expectations can be established similarly, so we only show the latter. First conclude from (A.1) and Assumption 7 that, with Θ_0 small enough,

$$\begin{aligned}\sup_{\theta\in\Theta_0}\left|h'\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right)\right| &\leq 2a_1+2a_2\left(\sup_{\theta\in\Theta_0}\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta)\right)^{a_3}\\ &\leq c_5\left(1+\sup_{\theta\in\Theta_0}\left\|\epsilon_t(\vartheta)\right\|^{2a_3}\right)\\ &\leq c_6\left(1+\left\|y_t\right\|^{2a_3}\right),\end{aligned}$$

where the last equality is obtained from the definition of $\epsilon_t(\vartheta)$ (see (15)) and Loeve's c_r–inequality (see Davidson (1994), p. 140). Thus, it follows that we need to show the finiteness of $\mathbb{E}_{\theta_0}\left(\|y_t\|^{4+2a_3}\right)$ or, by (5) and Minkowski's inequality, the finiteness of

$$\mathbb{E}_{\theta_0}\left(\|\epsilon_t\|^{4+2a_3}\right) \leq c_7 \mathbb{E}_{\lambda_0}\left(\rho_t^{4+2a_3}\right) = \frac{\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty \zeta^{n/2+1+2a_3} f(\zeta; \lambda_0)\, d\zeta < \infty,$$

where the former inequality is justified by (6) and the latter by Assumption 7.

From (15) and (A.15) it can be seen that the treatment of $\partial^2 g_t(\theta)/\partial\sigma\partial\sigma'$ is very similar to that of $\partial^2 g_t(\theta)/\partial\vartheta_1\partial\vartheta_2'$ and the same is true for $\partial^2 g_t(\theta)/\partial\vartheta_i\partial\sigma'$ $(i = 1, 2)$ (see (A.16), (A.5), and (A.6)). Next consider $\partial^2 g_t(\theta)/\partial\lambda\partial\lambda'$. The dominance assumptions imposed on the third and fifth functions in Assumption 7 together with the triangular inequality and the Cauchy-Schwarz inequality imply that, with Θ_0 small enough,

$$\mathbb{E}_{\theta_0}\left(\sup_{\theta\in\Theta_0}\left\|\frac{\partial^2}{\partial\lambda\partial\lambda'} g_t(\theta)\right\|\right) \leq 2a_1 + 2a_2 \mathbb{E}_{\theta_0}\left(\left(\sup_{\theta\in\Theta_0} \epsilon_t(\vartheta)' \Sigma^{-1} \epsilon_t(\vartheta)\right)^{a_3}\right),$$

where the finiteness of the right hand side was established in the case of $\partial^2 g_t(\theta)/\partial\vartheta_1\partial\vartheta_2'$. The treatment of the remaining components, $\partial^2 g_t(\theta)/\partial\vartheta_i\partial\lambda'$ and $\partial^2 g_t(\theta)/\partial\sigma\partial\lambda'$, involve no new features, so details are omitted.

Finally, because

$$-(T-s-nr)^{-1} \partial^2 l_T(\hat{\theta})/\partial\theta\partial\theta' = -(T-s-nr)^{-1} \sum_{t=r+1}^{T-s-(n-1)r} \partial^2 g_t(\hat{\theta})/\partial\theta\partial\theta',$$

the consistency claim is a straightforward consequence of the fact that $\partial^2 g_t(\theta)/\partial\theta\partial\theta'$ obeys a uniform law of large numbers. This completes the proof. □

References

[1] Alessi, L., M. Barigozzi, and M. Capasso (2008). A Review of Nonfundamentalness and Identification in Structural VAR Models. Working Paper Series 922, European Central Bank.

[2] Andrews, B. R.A. Davis, and F.J. Breidt (2006). Maximum likelihood estimation for all-pass time series models. Journal of Multivariate Analysis 97, 1638-1659.

[3] Breidt, J., R.A. Davis, K.S. Lii, and M. Rosenblatt (1991). Maximum likelihood estimation for noncausal autoregressive processes. Journal of Multivariate Analysis 36, 175-198.

[4] Breidt, J., R.A. Davis, and A.A. Trindade (2001). Least absolute deviation estimation for all-pass time series models. The Annals of Statistics 29, 919-946.

[5] Brockwell , P.J. and R.A. Davis (1987). Time Series: Theory and Methods. Springer-Verlag. New York.

[6] Campbell, J.Y., and R.J. Shiller (1987). Cointegration and Tests of Present Value Models. Journal of Political Economy 95, 1062–1088.

[7] Campbell, J.Y., and R.J. Shiller (1991). Yield Spreads and Interest Rate Movements: A Bird's Eye View. Review of Economic Studies 58, 495–514.

[8] Davidson, J. (1994). Stochastic Limit Theory. Oxford University Press, Oxford.

[9] Duffee, G. (2002). Term Premia and Interest Rate Forecasts in Affine Models. Journal of Finance 57, 405–443.

[10] Fang, K.T., S. Kotz, S., and K.W. Ng (1990). Symmetric Multivariate and Related Distributions. Chapman and Hall. London.

[11] Hannan, E.J. (1970). Multiple Time Series. John Wiley and sons. New York

[12] Lanne, M., J. Luoto, and P. Saikkonen (2010). Optimal Forecasting of Noncausal Autoregressive Time Series. HECER Discussion Paper No. 286.

[13] Lanne, M., and P. Saikkonen (2008). Modeling Expectations with Noncausal Autoregressions. HECER Discussion Paper No. 212.

[14] Lütkepohl, H. (1996). Handbook of Matrices. John Wiley & Sons,New York.

[15] Rosenblatt, M. (2000). Gaussian and Non-Gaussian Linear Time Series and Random Fields. Springer-Verlag, New York.

[16] Sargent, T.J. (1979). A Note on Maximum Likelihood Estimation of the Rational Expectations Model of the Term Structure. Journal of Monetary Economics 5, 133–143.

[17] White, H. (1994). Estimation, Inference and Specification Analysis. Cambridge University Press. New York.

[18] Wong, C.H. and T. Wang (1992). Moments for elliptically countered random matrices. Sankhyā 54, 265-277.

Figure 1: Quantile-quantile plots of the residuals of the VAR(3,0)-N (upper panel) and VAR(2,1)-t (lower panel) models for the U.S. term structure data.

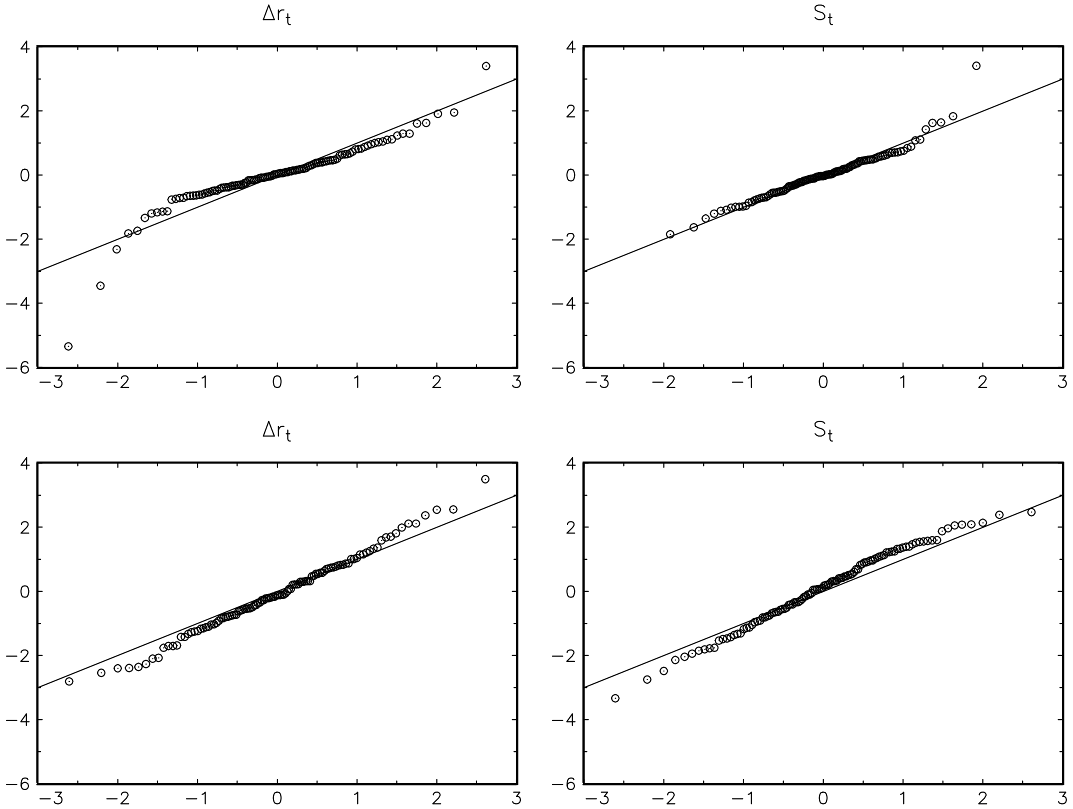

Table 1: Results of diagnostic checks of the third-order VAR models for the term structure.

	Model				
	VAR(3,0)-N	VAR(3,0)-t	VAR(2,1)-t	VAR(1,2)-t	VAR(0,3)-t
Ljung-Box (4)	0.172	0.014	0.094	9.4e–5	0.003
	0.118	0.069	0.063	3.2e–5	0.027
McLeod-Li (4)	0.4.2e–4	0.023	0.896	5.2e–5	0.101
	0.002	0.183	0.930	0.018	0.003
Log-likelihood	–258.510	–229.985	–222.953	-227.454	–231.252

VAR(r, s) denotes the vector autoregressive model for $(\Delta r_t, S_t)'$ with the rth and sth order polynomials $\Pi(B)$ and $\Phi(B^{-1})$, respectively. N and t refer to Gaussian and t-distributed errors, respectively. Marginal significance levels of the Ljung-Box and McLeod-Li tests with 4 lags are reported for each equation.

Table 2: Estimation results of the VAR(2,1)-t model for $(\Delta r_t, S_t)'$.

Π_1	–0.458	0.782
	(0.156)	(0.189)
	0.138	0.075
	(0.143)	(0.183)
Π_2	–0.241	0.298
	(0.090)	(0.184)
	0.320	–0.006
	(0.097)	(0.164)
Φ_1	0.399	–0.210
	(0.126)	(0.067)
	–0.240	0.673
	(0.260)	(0.144)
Σ	0.296	–0.167
	(0.096)	(0.106)
	–0.167	0.312
	(0.106)	(0.189)
λ	4.085	
	(1.210)	

The figures in parentheses are standard errors based on the Hessian of the log-likelihood function.