Noncausal Vector Autoregression[†]

Markku Lanne[*]
Department of Economics,
and HECER,
University of Helsinki

Pentti Saikkonen[**]
Department of Mathematics and Statistics,
and HECER,
University of Helsinki

April 2009

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 importance in empirical economic research which currently uses 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. This is emphasized in the paper by noting that noncausality is closely related to the notion of nonfundamentalness, under which structural economic shocks cannot be recovered from an estimated causal VAR model. As detecting nonfundamentalness is therefore of great importance, we propose a procedure for discriminating between causality and noncausality that can be seen as a test of nonfundamentalness. The methods are illustrated with applications to fiscal foresight and the term structure of interest rates.

[†] We thank Martin Ellison, Juha Kilponen, and Antti Ripatti for useful comments. Financial support from the Academy of Finland and the Okobank 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.

* Department of Economics, University of Helsinki, P.O.Box 17 (Arkadiankatu 7), FIN–00014 University of Helsinki, Finland, e-mail: markku.lanne@helsinki.fi

** Department of Mathematics and Statistics, University of Helsinki, P.O.Box 68 (Gustaf Hällströmin katu 2b), FIN–00014 University of Helsinki, Finland, e-mail: pentti.saikkonen@helsinki.fi

1 Introduction

In economic and financial applications, the vector autoregressive (VAR) model is usually considered as an atheoretical summary of the dynamics of the included variables. Especially when the model is used for forecasting its error term is interpreted as a forecast error that should be an independent white noise process in order for the model to capture all relevant dynamic dependencies. Typically, 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 causal VAR models may lead to suboptimal forecasts and false economic interpretations.

Noncausality is closely related to nonfundamentalness that arises, in particular, in rational expectations models. The issue of nonfundamentalness was probably first pointed out by Hansen and Sargent (1980, 1991), who showed that in its presence structural economic shocks cannot be recovered from an estimated causal VAR model. Subsequently, a relatively large literature has explored nonfundamentalness in various applications; for a recent survey, see Alessi et al. (2008).

To define nonfundamentalness, let us consider a dynamic rational expectations model whose solution is typically a stationary stochastic vector process y_t that can be expressed as a vector autoregression. Thus, an econometrician considers the specification

$$D(B)y_t = \epsilon_t, \tag{1}$$

where the errors ϵ_t are interpreted as (functions of) the random shocks to agents' information set and $D(B) = \sum_{j=0}^{\infty} D_j B^j$ is a potentially infinite-order lag polynomial in the backward shift operator B (i.e., $B^k y_t = y_{t-k}$ for $k = 0, \pm 1, ...$). In the econometric analysis, ϵ_t is usually assumed to be a sequence of independent and identically distributed random vectors with zero mean and positive definite covariance matrix, and the roots of $\det D(z)$, the determinant of $D(z)$, are assumed to lie outside the unit disc. The latter condition implies that the process y_t can equivalently be written as

$$y_t = C(B)\epsilon_t, \tag{2}$$

where $C(B)$ is an infinite-order lag polynomial depending only on positive powers of B. In other words, y_t only depends on the past and present errors ϵ_{t-j}, $j \geq 0$, which can be recovered by the employed autoregression and interpreted as fundamental economic shocks. In this case the autoregression (1) is referred to as fundamental.

The autoregression (1) is nonfundamental when some of the roots of $\det D(z)$ lie inside the unit disc. As discussed by Hansen and Sargent (1991) and Alessi et al. (2008), this can happen because the underlying economic model simply leads to such a nonfundamental autoregression or because some relevant state variables are not observed by the econometrician and, therefore, not included in the analysis. However, even in this case the process y_t admits an infinite-order moving average representation of the type (2) but, unlike in the preceding fundamental case, the filter $C(B)$ now depends on negative powers of B, implying dependence of y_t on future errors ϵ_{t+j}, $j \geq 0$. A similar dependence on future errors also occurs in the noncausal VAR model to be introduced in Section 2 so that nonfundamentalness

shows up as noncausality in the VAR representation of y_t. However, in conventional causal VAR analysis the infinite-order moving average representation only depends on past and present errors. This means that, in the presence of noncausality, the analysis is based on a misspecified model and, consequently, the errors recovered from the employed VAR model cannot be interpreted as (functions of) the random shocks to agents' information set. Thus, checking for noncausality also serves as a check for nonfundamentalness. Although we have here only discussed rational expectations models, nonfundamentalness is also common in many other kinds of economic models; one example being models with heterogeneous information, exemplified in Section 4.2, and others can be found in Alessi et al. (2008) and the references therein.

The evaluation of dynamic stochastic general equilibrium (DSGE) models by means of structural vector autoregressions is an application where ensuring the fundamentalness (or causality) of the VAR representation is of great importance. If such a representation is falsely assumed, the structural shocks obtained have no economic meaning and validating a DSGE model based on the impulse responses implied by the structural VAR model is misleading. Therefore, it is in this field that some ways of checking for fundamentalness have been devised although they should be more generally applicable. Fernández-Villaverde et al. (2007) derived conditions under which the economic shocks of (a linearization of) a DSGE model match up with those associated with a fundamental VAR model. This approach, however, only works when there are as many economic shocks as there are observable variables, which restricts its applicability to relatively small systems. Giannone and Reichlin (2006) pointed out that nonfundamentalness can be detected by augmenting the VAR model with additional variables and checking whether they Granger cause the variables of interest. Under fundamentalness, there should be no such Granger causality. The additional variables should be "potentially relevant" and "likely to be driven by sources that are common with the variables of interest", but their selection seems, however, rather arbitrary. Hence, we argue that checking for noncausality provides a viable and potentially more general approach to detecting nonfundamentalness.

A concept closely related to nonfundamentalness is indeterminacy of equilibria in economic models, which is a highly topical issue in macroeconomics, especially in studying monetary policy. Indeterminacy allows structural shocks to be nonfundamental. Therefore, checking for causality facilitates checking for determinacy in that detecting a causal VAR representation of the data can be interpreted as evidence in favor of determinate equilibria. Some tests for indeterminacy have been presented in the previous literature, but it has turned out to be very difficult to discriminate empirically between determinacy and indeterminacy. In particular, Beyer and Farmer (2007) have shown that two DSGE models, one with a determinate and the other with an indeterminate equilibrium, may be observationally equivalent in that they generate the same likelihood function, rendering tests of parameter restrictions (e.g. Lubik and Schorfheide (2004)), in general, futile. Also, commonly used test procedures based on evaluating the amount of variation in the residuals of rational expectations models that is left unexplained by fundamentals (see, e.g. Salyer and Sheffrin (1998)) are rather arbitrary as they crucially depend on the variables that are included in the analysis. The approach based on checking for noncausality, in contrast, is quite general as it is based on unrestricted VAR models and there is no need to determine the suitable set of additional fundamental economic variables.

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 (references to previous univariate work can be found in Rosenblatt (2000), 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 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 a local ML estimator, and justify the applicability of usual likelihood based tests.

The remainder of the paper is structured as follows. Section 2 introduces the noncausal VAR model. Section 3 presents the likelihood function and properties of the ML estimator. Section 4 illustrates the use of the noncausal VAR model in detecting potential nonfundamentalness in the context of fiscal foresight and the term structure of interest rates. Section 5 concludes. A mathematical appendix contains proofs of the results 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

$$\Pi(B)\,\Phi\left(B^{-1}\right) y_t = \epsilon_t, \tag{3}$$

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

$$\det \Pi(z) \neq 0, \quad |z| \leq 1, \quad \text{and} \quad \det \Phi(z) \neq 0, \quad |z| \leq 1. \tag{4}$$

If $\Phi_j \neq 0$ for some $j \in \{1, .., s\}$, equation (3) 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 (4) guarantees the stationarity of the model. In the general set up of equation (3) 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

$$u_t = M(B)\,\epsilon_t = \sum_{j=0}^{\infty} M_j \epsilon_{t-j}. \tag{5}$$

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 $\Pi(z)$. 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 $w_t = \det(\Pi(B))\, y_t$. Note that $\Xi(z)$ is a matrix polynomial of degree at most $(n-1)\,r$ and, because $\Pi(0) = I_n$, we also have $\Xi(0) = I_n$. By the latter condition in (4) 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

$$w_t = \sum_{j=-(n-1)r}^{\infty} N_j \epsilon_{t+j}, \tag{6}$$

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

From (4) it follows that the process y_t itself has the representation

$$y_t = \sum_{j=-\infty}^{\infty} \Psi_j \epsilon_{t-j}, \tag{7}$$

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 $j \to \infty$. Clearly, the representation (7) can be obtained by multiplying both sides of (6) by $\det(\Pi(B))^{-1}$ so that we also have $\Psi(z) = \det(\Pi(z))^{-1} N(z^{-1})$. The representation (7) 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 (3). In the causal case $s = 0$, the conventional abbreviation VAR(r) is also used.

In the noncausal case, $\Psi_j \neq 0$ for some $j < 0$, which shows the connection of our noncausal VAR model and nonfundamentalness discussed in the Introduction. To see further implications of noncausality, let $\mathbb{E}_t(\cdot)$ stand for the conditional expectation operator with respect to the information set $\{y_t, y_{t-1}, ...\}$. From (3) and (7) it is seen 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 (5). However, in the noncausal case this does not happen. Then $\Psi_j \neq 0$ for some $j < 0$, which in conjunction with the representation (7) shows that y_t and ϵ_{t-j} are correlated and, consequently, $\mathbb{E}_t(\epsilon_{t-j}) \neq 0$ for some $j < 0$. Thus, future errors can be predicted by past values of the process y_t, which can be seen as an alternative characterization of nonfundamentalness.

In addition to depending on expected future errors, the process y_t can also be interpreted as being dependent on its expected future values. To see this, let us, for simplicity, concentrate on the purely noncausal model, where $\Pi(B) = I_n$. In this

case, model (3) can be written as

$$y_t = \Phi_1 y_{t+1} + \cdots + \Phi_s y_{t+s} + \epsilon_t,$$

and, taking conditional expectations with respect to the information set $\{y_t, y_{t-1}, ...\}$, one obtains

$$y_t = \Phi_1 \mathbb{E}_t \left(y_{t+1}\right) + \cdots + \Phi_s \mathbb{E}_t \left(y_{t+s}\right) + \mathbb{E}_t \left(\epsilon_t\right). \tag{8}$$

This shows that the elements of the coefficient matrix Φ_j give the effect of the expectation of y_{t+j} on y_t. In the general case ($\Pi(B) \neq I_n$), we obtain a similar expression for y_t with the exception that $\mathbb{E}_t(\epsilon_t)$ is replaced by $\mathbb{E}_t(u_t)$.

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 (3) 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

[1] A direct application of Hannan's (1970) Theorem 10' would give a representation with ω replaced

of y_t has the representation $(2\pi)^{-1}\left(\sum_{j=0}^{r+s}\mathcal{C}_j e^{ij\omega}\right)^{-1}\left(\sum_{j=0}^{r+s}\mathcal{C}_j e^{-ij\omega}\right)^{\prime-1}$, 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 (3) are statistically indistinguishable and nothing is lost by using a conventional causal representation. However, if the errors are non-Gaussian using a causal representation of a true noncausal process means using a VAR model whose errors are only guaranteed to be uncorrelated but not independent. Then the errors can be predicted by past values of the considered series and, as discussed above, one is faced with the problem of nonfundamentalness, implying that the errors of the employed causal VAR model do not match up with fundamental economic shocks. Thus, when fundamentalness is an issue, it is advisable to first fit an (adequate) causal autoregression to the observed series by standard least squares or Gaussian ML and check whether the residuals look non-Gaussian. If deviations from Gaussianity are detected it is reasonable to consider the noncausal VAR model (3) and check for nonfundamentalness by the procedures to be developed in subsequent sections.

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 noncausal VAR models 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

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)).

distribution belongs to the class of elliptical distributions but we will not rule out it at this point. Other examples of elliptical distributions are given in Fang et al. (1990, Chapter 3). Perhaps the best known non-Gaussian example is the multivariate t-distribution.

Assumption 1. The error process ϵ_t in (3) 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 $\varepsilon_t = \Sigma^{-1/2}\epsilon_t$. By Assumption 1, we have the representations

$$\epsilon_t \overset{d}{=} \rho_t \Sigma^{1/2} \upsilon_t \quad \text{and} \quad \varepsilon_t \overset{d}{=} \rho_t \upsilon_t, \tag{9}$$

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 (so that $\upsilon_t'\upsilon_t = 1$). The density of ϵ_t is of the form

$$f_\Sigma(x;\lambda) = \det(\Sigma)^{-1/2} f\left(x'\Sigma^{-1}x;\lambda\right) \tag{10}$$

for some nonnegative function $f(\cdot;\lambda)$ of a scalar variable. In addition to the positive definite parameter matrix Σ $(n \times n)$ the distribution of ϵ_t is allowed to depend on the parameter vector λ $(\mathrm{d} \times 1)$. The parameter matrix Σ is closely related to the covariance matrix of ϵ_t from which it only differs by a multiplicative scalar. 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 (9) that

$$\mathbb{C}(\epsilon_t) = \frac{\mathbb{E}(\rho_t^2)}{n}\Sigma. \tag{11}$$

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. Equality (11) already

illustrates this and for subsequent purposes we note that the density of ρ_t^2, denoted by $\varphi_{\rho^2}(\cdot;\lambda)$, is related to the function $f(\cdot;\lambda)$ in (10) via

$$\varphi_{\rho^2}(\zeta;\lambda) = \frac{\pi^{n/2}}{\Gamma(n/2)}\zeta^{n/2-1} f(\zeta;\lambda), \quad \zeta \geq 0, \tag{12}$$

where $\Gamma(\cdot)$ is the gamma function (see Fang et al. (1990, p. 36)). Assumptions 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. Note, however, that when our assumptions are specialized to the univariate case the first argument in the density function on the right hand side of (10) will be the square of that appearing in these previous papers.

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 $\lambda \in \Lambda$,

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

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 related to condition (A2) in the aforementioned papers. To see this, notice that, for all $\lambda \in \Lambda$,

$$\int_0^\infty \zeta^{n/2} f'(\zeta;\lambda)\, d\zeta = \zeta^{n/2} f(\zeta;\lambda)\,|_0^\infty - \frac{n}{2}\int_0^\infty \zeta^{n/2-1} f(\zeta;\lambda)\, d\zeta = -\frac{n\Gamma(n/2)}{2\pi^{n/2}}.$$

Here the latter equality follows because, by the latter part of Assumption 2(ii), the first term in the second expression is zero and

$$\int_0^\infty \zeta^{n/2-1} f(\zeta;\lambda)\, d\zeta = \frac{\Gamma(n/2)}{\pi^{n/2}}\int f(x'x;\lambda)\, dx = \frac{\Gamma(n/2)}{\pi^{n/2}} \tag{13}$$

as in Fang et al. (1990, p. 35). When $n = 1$ the last expression equals unity, showing the aforementioned connection. In Andrews et al. (2006) and Lanne and Saikkonen (2008) the values of the parameter λ are only assumed to belong to some (small) neighborhood of the true parameter value but we have preferred to be slightly less general here (this also applies to some subsequent assumptions).

The first condition in Assumption 2(iii) implies that $\mathbb{E}_\lambda(\rho_t^4)$ is finite (see (12)) and, taken together, this assumption guarantees finiteness of some expectations needed in subsequent developments. In particular, the latter condition in Assumption 2(iii) implies finiteness of the quantities

$$\boldsymbol{j}(\lambda) = \frac{4\pi^{n/2}}{n\Gamma(n/2)}\int_0^\infty \zeta^{n/2}\frac{(f'(\zeta;\lambda))^2}{f(\zeta;\lambda)}d\zeta = \frac{4}{n}\mathbb{E}_\lambda\left[\rho_t^2\left(\frac{f'(\rho_t^2;\lambda)}{f(\rho_t^2;\lambda)}\right)^2\right] \tag{14}$$

and

$$\boldsymbol{i}(\lambda) = \frac{\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty \zeta^{n/2+1}\frac{(f'(\zeta;\lambda))^2}{f(\zeta;\lambda)}d\zeta = \mathbb{E}_\lambda\left[\rho_t^4\left(\frac{f'(\rho_t^2;\lambda)}{f(\rho_t^2;\lambda)}\right)^2\right], \tag{15}$$

where the latter equalities follow from the expression of the density of ρ_t^2 (see (12)). 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* $\boldsymbol{i}(\lambda) \geq (n+2)^2\left[\mathbb{E}_\lambda(\rho_t^2)\right]^2/4\mathbb{E}_\lambda(\rho_t^4)$ *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

$$\boldsymbol{\tau}(\lambda) = \boldsymbol{j}(\lambda)\,\mathbb{E}_\lambda\left(\rho_t^2\right)/n. \tag{16}$$

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 (3) 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 (4) holds for all $\vartheta = (\vartheta_1, \vartheta_2) \in \Theta_1 \times \Theta_2$.

This is a standard assumption. The differentiability requirement guarantees that the likelihood function is twice continuously differentiable. We will continue to use to 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 (3). The parametrization (3) was employed by Lanne and Saikkonen (2008) whose results we here generalize. Unless otherwise stated, Assumptions 1-3 are supposed to hold.

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

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

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

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

From the definition of u_t and (3) it follows that $\Pi\left(B\right) u_t = \epsilon_t$ so that from the preceding equality we find

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

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 (5) and (6) 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 (3) 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

$$l_T(\theta) = \sum_{t=r+1}^{T-s-(n-1)r} g_t(\theta), \tag{17}$$

where the parameter vector θ contains the unknown parameters and (cf. (10))

$$g_t(\theta) = \log f\left(\epsilon_t(\vartheta)'\Sigma^{-1}\epsilon_t(\vartheta);\lambda\right) - \frac{1}{2}\log\det(\Sigma), \tag{18}$$

with

$$\epsilon_t(\vartheta) = u_t(\vartheta_2) - \sum_{j=1}^{r} \Pi_j(\vartheta_1)\, u_{t-j}(\vartheta_2) \tag{19}$$

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 (5), (6), and (7) 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 $\partial l_T(\theta_0)/\partial\theta$, that is, the score of θ evaluated at the true parameter value θ_0. Explicit expressions of the derivatives of the log-likelihood function are given in Appendix A. Here we only present the expression of the limit

$\lim_{T\to\infty} T^{-1}\mathbb{C}\left(\partial l_T\left(\theta_0\right)/\partial\theta\right)$. 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 $f_1\left(\zeta\right)$ such that $\int_0^\infty \zeta^{n/2-1} f_1\left(\zeta\right) d\zeta < \infty$ and, in some neighborhood of λ_0, $\left|\partial f\left(\zeta;\lambda\right)/\partial\lambda_i\right| \leq f_1\left(\zeta\right)$ for all $\zeta \geq 0$ and $i = 1, ..., d$.

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

The first condition is a standard dominance condition which is needed to guarantee 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 $\boldsymbol{\tau}\left(\lambda\right) = \boldsymbol{j}\left(\lambda\right)\mathbb{E}_\lambda\left(\rho_t^2\right)/n$ where $\boldsymbol{j}\left(\lambda\right)$ is defined in (14). In what follows, we denote $\boldsymbol{j}_0 = \boldsymbol{j}\left(\lambda_0\right)$ and $\boldsymbol{\tau}_0 = \boldsymbol{j}_0\mathbb{E}_{\lambda_0}\left(\rho_t^2\right)/n$. Define the $n \times n$ matrix

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

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

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

Notice that $\boldsymbol{j}_0^{-1} C_{11}\left(a, b\right) = \mathbb{E}_{\lambda_0}\left(u_{0,t-a} u'_{0,t-b}\right)$. As shown in Appendix B, $\mathcal{I}_{\vartheta_1\vartheta_1}\left(\theta_0\right)$ 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 $\partial l_T\left(\theta_0\right)/\partial\theta$. Thus, the true parameter value θ_0 as well as the standardization and (possible) limiting operation are not necessarily mentioned.

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

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

a square matrix of order $n(n+1)/2$. An explicit expression of the matrix J_0 can be obtained from Wong and Wang (1992, p. 274) or Fang et al. (1990, Theorem 3.3). For later purposes we note that

$$\mathbb{E}\left[\left(\text{vec}(\upsilon_t\upsilon_t')\right)\left(\text{vec}(\upsilon_t\upsilon_t')\right)'\right] = \frac{1}{n(n+2)}\left(I_{n^2} + K_{nn} + \text{vec}(I_n)\,\text{vec}(I_n)'\right) \tag{20}$$

We also denote $\Pi_{i0} = \Pi(\vartheta_{10})$, $i = 1, ..., r$, and $\Pi_{00} = -I_n$, and define the partitioned matrix $C_{22}(\theta_0) = [C_{22}(a,b;\theta_0)]_{a,b=1}^{s}$ $(n^2 s \times n^2 s)$ where the $n \times n$ matrix $C_{22}(a,b;\theta_0)$ is

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

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). Notice that in the scalar case $n = 1$ and in the purely noncausal case $r = 0$ the expression of $C_{22}(\theta_0)$ simplifies because the latter term in the definition of $C_{22}(a,b;\theta_0)$ vanishes (see equality (B.6) in Appendix B) and the former only depends on the coefficient matrices Ψ_{j0} with $j < 0$ (cf. Lanne and Saikkonen (2008)).

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

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

In the scalar case $n = 1$ and in the purely noncausal case $r = 0$ simplifications again result because in the expression of $C_{12}(a, b; \theta_0)$ the former term vanishes (see equality (B.6) in Appendix B). 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

$$\mathcal{I}_{\sigma\sigma}(\theta_0) = D_n'\left(\Sigma_0^{-1/2} \otimes \Sigma_0^{-1/2}\right) D_n J_0 D_n'\left(\Sigma_0^{-1/2} \otimes \Sigma_0^{-1/2}\right) D_n$$

and

$$\mathcal{I}_{\vartheta_2\sigma}(\theta_0) = -2\sum_{j=0}^{s} \frac{\partial}{\partial\vartheta_2}\phi_j(\vartheta_2) \sum_{i=0}^{r}\left(\Psi_{j-i,0}\Sigma_0^{1/2} \otimes \Pi_{i0}'\Sigma_0^{-1/2}\right) D_n J_0 D_n'\left(\Sigma_0^{-1/2} \otimes \Sigma_0^{-1/2}\right) D_n$$

with $\mathcal{I}_{\vartheta_2\sigma}(\theta)' = \mathcal{I}_{\sigma\vartheta_2}(\theta)$. Note that in the scalar case $n = 1$ and in the purely noncausal case $r = 0$ we have $\mathcal{I}_{\vartheta_2\sigma}(\theta_0) = 0$ (see equality (B.6) in Appendix B). 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

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

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.

Now we can define

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

the information matrix of the whole parameter vector θ. As already noted, 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. The latter fact means that then the parameters ϑ and (σ, λ) are orthogonal so that, asymptotically, their ML estimators are independent.

Before presenting the asymptotic distribution of the score of θ we introduce conditions which guarantee the positive definiteness of its covariance matrix. These include conventional rank conditions on the first derivatives of the functions in Assumption 3 and assumptions on the score of λ which are needed because of the general nature of the parameter vector λ. 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.

As already indicated, Assumption 5(i) is standard. 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 multivariate case the orthogonality of σ with respect to ϑ_2 fails but it is still possible to do without assuming more than assumed in the univariate case. We also note 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 that the matrix J_0 is positive definite. The latter fact is established in Lemma 6 in Appendix B even when the errors are Gaussian.

Now we can present the limiting distribution of the score.

Proposition 2 *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 the usefulness of this result it is crucial that ϵ_t is non-Gaussian because in the Gaussian case the information matrix $I_{\theta\theta}(\theta_0)$ is singular (see the proof of Proposition 2, 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 3 *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 3 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. The conditions in Assumption 7 are only slightly different from those used in Andrews et al. (2006) and Lanne and Saikkonen (2008).

Now we can state the main result of this section.

Theorem 4 *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 (17) 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* $-(T - s - nr)^{-1}\partial^2 l_T(\hat{\theta})/\partial\theta\partial\theta'$.

Thus, Theorem 4 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 $\partial g_t(\theta)/\partial\theta$ and $\partial g_k(\theta)/\partial\theta$, $k \neq t$, 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 2 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 applications

In this section, we consider two economic applications of the noncausal VAR model. In each case, discriminating between causality and noncausality is primarily seen as an indirect test of the economic hypothesis being considered. In other words, the presence of noncausality per se would be seen as evidence against the theory. Moreover, basing a test of the theory on the assumption of causality, as is typically done, would be incorrect in the presence of noncausality. It is only after ascertaining that the variables indeed have a causal VAR representation that the economic theory can be evaluated by testing restrictions on the parameters of such a VAR model. Hence, checking for noncausality can be considered a pretest validating conventional testing procedures.

In our applications, 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.

If a noncausal model is selected, it would often be of interest to proceed to impulse response analysis to fully understand the effects of economic shocks but, as the relevant methods are not readily available, that lies outside the scope of this paper. The main difficulty with impulse response analysis is that prediction in noncausal autoregressions is, in general, a nonlinear problem and no closed form of the forecast function is currently available (see Rosenblatt (2000, Chapter 5) and the discussion in Lanne and Saikkonen (2008)).

4.1 Fiscal foresight

Our first application is concerned with fiscal foresight, i.e., the phenomenon that due to lags in implementation, agents receive signals about a future change in the tax rate or government spending before they actually take place. It can be shown that the presence of foresight leads to time series with a non-invertible moving average component in equilibrium (for a survey of this literature, see Leeper et al. (2008)). In other words, if there is foresight, a VAR model incorporating the key variables of the economy (including taxes and government expenditure), is noncausal. Finding noncausality therefore provides evidence in favor of fiscal foresight, which invalidates analyses based on conventional causal VAR models common in the previous literature. This was illustrated by Yang (2005) who showed by simulations of a standard neoclassical growth model that relying on a causal VAR model in the presence of foresight of only one quarter can yield very misleading estimates of tax effects. This,

of course, follows from the fact that the errors of the identified VAR model are not the true fiscal shocks in this case.

The previous empirical evidence of fiscal foresight is mostly based on case studies around major fiscal policy changes and not closely connected to theory (see Poterba (1988), Auerbach and Slemrod (1997) Steigerwald and Stuart (1997), and House and Shapiro (2006, 2008), inter alia). Indirect evidence in favor of fiscal foresight not based on a single tax change was recently provided by Yang (2007) who showed that in a VAR model augmented by variables capturing expectations (such as prices and interest rates), the responses of labor, investment, and output to a tax shock become weaker. Our approach of checking for noncausality can be seen as a more direct test of fiscal foresight.

We consider a simple trivariate VAR model for the (demeaned) differences of U.S. GDP, total government expenditure, and total government revenue (all in real per capita terms). The quarterly data from 1955:1 to 2000:4 (184 observations) were previously used by Mountford and Uhlig (in press), who also provide a detailed description of the construction of the variables. We start the analysis by searching for an adequate Gaussian vector autoregression. The AIC and BIC select VAR(3) and VAR(2) models, respectively, and according to the diagnostic test results reported in Table 1, the second-order model is deemed sufficient in capturing autocorrelation.[2] However, the errors, especially those of the equation for the government expenditure, exhibit conditional heteroskedasticity. Also, the quantile-quantile (Q-Q) plots in the upper panel of Figure 1 suggest that the errors are not normally distributed[3] with

[2]Note that, when the orders of the model are misspecified, the Ljung-Box and McLeod-Li 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 used to check the error distribution.

[3]The p-values of the Shapiro-Wilk test are 0.005, 0.416 and 1.80e–6 for the residuals of the equations for the GDP, government expenditure and government revenue, respectively.

the greatest discrepancies at the tails, suggesting that a fat-tailed error distribution might be more appropriate.

Instead of the normal distribution, we consider the multivariate t-distribution for the errors. In this case, the second-order model that best fits the data in terms of the log likelihood function, is the VAR(1,1)-t model. It also seems to be the only model in Table 1 that produces well-behaved residuals, with the other specifications exhibiting autocorrelation or conditional heteroskedasticity in at least one of the equations. Hence, there is evidence in favor of a noncausal VAR representation of the data, indicating the presence of fiscal foresight. The Q-Q plots of the residuals of the preferred model in the lower panel of Figure 1 attest to the good fit of the multivariate t-distribution.[4] Also, the estimated value of the degrees-of-freedom parameter λ in Table 2 is relatively small (8.253), which lends further support to the need for a fat-tailed error distribution.

Because the noncausal model provides the best fit, analyses based on causal VAR models are expected to be misleading as they fail to extract the correct structural shocks. The presence of a noncausal VAR representation indicates the importance of expectations of future tax and government expenditure changes that the conventional causal VAR model does not take into account. The elements of the matrix Φ_1 give the effect of a change in expected next-period values of the variables on the current values, as discussed in Section 2.1 (see, in particular, Equation (8) and the ensuing discussion). In particular, the estimates in Table 2 suggest that expectations of future tax increases tend to increase the GDP and government revenue.

4.2 Term structure of interest rates

As another application, we consider the expectations hypothesis of the term structure of interest rates. According to this theory, the long-term interest rate is a weighted sum of present and expected future short-term interest rates. Campbell and Shiller

[4]The p-values of the Shapiro-Wilk test for the three residual series are 0.455, 0.295 and 0.186, respectively.

(1987, 1991) suggested testing the expectations hypothesis by testing the restrictions it imposes on the parameters of a VAR model for the change in the short-term interest rate and the spread between the long-term and short-term interest rates. Furthermore, they showed how the theoretical spread satisfying these restrictions can be computed based on the estimated VAR model. The expectations hypothesis is a special case of the present value model, and similar techniques have been widely employed in testing that model in the context of various applications, including stock returns (Campbell and Shiller (1987)) and the net present value budget balance (Roberds (1991)). Although this method is straightforward, it crucially depends on the existence of a causal VAR representation, suggesting that its validity can be assured by checking the causality of the related vector autoregression.

Finding noncausality indicates nonfundamentalness that can arise because the agents' information set is larger than the econometrician's. Although the discrepancy between the information sets poses no problem in testing the expectations hypothesis under the assumption of the existence of a causal VAR representation maintained in most of the previous literature, the conclusions can be misleading if this assumption is falsely imposed. One explanation for nonfundamentalness and, hence, noncausality, in asset pricing models recently put forth by Kasa et al. (2007) are heterogenous beliefs. Indeed, they show that if agents have different information, nonfundamental representations of the data correspond to nonrevealing equilibria where the agents "forecast the forecasts of others". So, detecting noncausality in the term structure may indicate that agents have heterogeneous information useful in predicting future interest rates.

We concentrate on a bivariate VAR model for the (demeaned) change in the three-month interest rate (Δr_t) and the spread between the ten-year and three-month interest rates (S_t) (quarter-end yields on U.S. zero-coupon bonds) from 1970:1–1998:4 (116 observations).[5] AIC and BIC select Gaussian VAR(3) and VAR(2) models, respec-

[5]The data were previously used by Duffee (2002). We thank Gregory Duffee for providing them on his website.

tively, but only the third-order model produces serially uncorrelated errors. However, the results in Table 3 show that the residuals are conditionally heteroskedastic and the Q-Q plots is the upper panel of Figure 2, indicate considerable deviations from normality.[6] 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 3. 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 2 lend support to the adequacy of the multivariate t distribution of the errors; the p-values of the Shapiro-Wilk test for the residuals of the equations for Δr_t and S_t equal 0.509 and 0.451, respectively. Moreover, the estimate of the degrees-of-freedom parameter λ reported in Table 4 is small (8.187) and accurate, suggesting inadequacy of the Gaussian error distribution. Thus, there is evidence of noncausality, but not pure noncausality, i.e., the term structure depends on expectations of future interest rates as well as past values.

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). This may also explain the common rejections of the hypothesis when testing is based on the assumption of a causal VAR model, which in view of our results is likely to be misspecified. Indirectly these findings lend support to the heterogeneous beliefs explanation of Kasa et al. (2007) discussed above although that is likely not to be the only possibility. The estimated Φ_1 matrix also seems to have an interpretation that goes contrary to the expectations hypothesis: an expected increase of the short-term rate tends to increase the current short-term rate while having no significant effect on

[6] The p-values of the Shapiro-Wilk test for the residuals of the equations of Δr_t and S_t equal 4.09e–8 and 0.001, respectively.

the spread. According to the expectations hypothesis, in contrast, an expected future increase in the short-term rate should have no effect on the current short-term rate, but it should increase the long-term rate and, therefore, the spread. Furthermore, here 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.

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. In the Gaussian case, causal and noncausal VAR models cannot be distinguished which underlines the importance of a careful specification of the error distribution of the model. This may also be important in causal VAR models because in the non-Gaussian case, absence of serial correlation does not necessarily guarantee nonpredictability of the errors. While the new model is likely to be useful in providing a more accurate description of the dynamics of economic time series than the causal model, it is probably in checking for nonfundamentalness that it is most valuable. Nonfundamentalness often invalidates the use of conventional econometric methods and it arises, in particular, in rational expectations models.

We have derived asymptotic properties of a (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 have been illustrated by means of two empirical applications. Evidence in favor of fiscal foresight in the U.S. was found, suggesting that shocks identified by imposing structural restrictions on causal VAR models to study the effects of fiscal policy, are not likely to carry any economic interpretation. Likewise, a noncausal VAR model for the U.S. term struc-

ture of interest rates turned out to be superior to the causal model, invalidating the commonly employed test procedures of the expectations hypothesis that explicitly assume causality.

While checking for nonfundamentalness is an important application of our methods, it can only be considered as the first step in the analysis of economic and financial data. Once noncausality is 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. Another issue of great interest is the use of noncausal VAR models for modeling expectations and the relation of noncausal VAR models to economic models involving expectations. 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 and, in particular, its performance in detecting indeterminacy in economic models could be examined by means of simulation experiments. We leave all of these issues for future research.

Mathematical Appendix

A Derivatives of the approximate log-likelihood function

To simplify subsequent derivations, we first introduce some notation. We set $h(\zeta;\lambda) = f'(\zeta;\lambda)/f(\zeta;\lambda)$ so that

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

Next, define

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

From (9) it is seen that

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

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

First derivatives of $l_T(\theta)$. It will be sufficient to consider the derivatives of $g_t(\theta)$. By straightforward differentiation one first obtains from (18)

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

where, from (19),

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

and

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

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 $Y_{t+1}(\vartheta_{10}) = Y_{0,t+1}$,

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

and

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

As for the parameters $\sigma=\text{vech}(\Sigma)$ and λ, straightforward differentiation yields

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

and

$$\frac{\partial}{\partial\lambda}g_t(\theta_0)=\frac{1}{f\left(\epsilon_t'\Sigma_0^{-1}\epsilon_t;\lambda_0\right)}\frac{\partial}{\partial\lambda}f\left(\epsilon_t'\Sigma_0^{-1}\epsilon_t;\lambda_0\right). \qquad (A.10)$$

Replacing θ_0 by θ gives the corresponding derivatives evaluated at an arbitrary θ.

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

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

By straightforward differentiation we now have

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

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

and

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

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

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

and furthermore that (see(A.4))

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

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

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

whereas

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

and

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

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

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

where the definition of the function h (see the beginning of Appendix A), density of ρ_t^2 (see (12)), 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 $\boldsymbol{j}\left(\lambda\right)$ (see (14)) yield

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

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 $\zeta^{n/4} f'\left(\zeta;\lambda\right)/\sqrt{f\left(\zeta;\lambda\right)}$ is proportional to $\zeta^{n/4}\sqrt{f\left(\zeta;\lambda\right)}$ or if and only if

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

This implies $f\left(\zeta;\lambda\right) = b\exp\left(-a\zeta\right)$ with $a > 0$ and $b > 0$. From the fact that $f\left(x'x;\lambda\right)$, $x \in \mathbb{R}^n$, is the density function of $\rho_t\upsilon_t$ (see (9) and (10)) it further follows that $b = \left(a/\pi\right)^{n/2}$ and that $\rho_t\upsilon_t$ has the normal density $\left(2\pi\right)^{-n/2}\exp\left(-x'x/2\right)$. Here the identity covariance matrix is obtained because $\rho_t^2 \sim \chi_n^2$, and hence from (11), $\mathbb{C}\left(\rho_t^2\upsilon_t\right) = I_n$ (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 $f'\left(\zeta;\lambda\right)/f\left(\zeta;\lambda\right) = c$ clearly holds with $c = -1/2$ and, because then $\rho_t^2 \sim \chi_n^2$, $\mathbb{E}_\lambda\left(\rho_t^2\right) = n$. This completes the proof for $\boldsymbol{j}\left(\lambda\right)$.

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 (12)). Proceeding as in the case of the first assertion yields

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

(see the definition of $\boldsymbol{i}(\lambda)$ in (15)). 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 $\boldsymbol{H}_1$. We have not found a simple way to show the nonsingularity of $\boldsymbol{H}_1$, so we demonstrate it 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

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

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

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

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 (4) 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 (4) the eigenvalues of $\mathbf{\Phi}$ are smaller than one in absolute value whereas the former condition in (4) 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 2 we present some auxiliary results. In the following lemmas, as well as in the proof of Proposition 2, 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 (9)) 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 5 *Under the conditions of Proposition 2,*

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

and

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

Proof of Lemma 5. By the definition of the function $h_0(\cdot)$ (see (A.3)) and the density of ρ_t^2 (see (12)) 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 (14). 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 (9) 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 6 . *Under the conditions of Proposition 2,*

$$\mathbb{C}(\varepsilon_{t-i} \otimes e_{0t}, \varepsilon_{k-j} \otimes e_{0k}) = \begin{cases} D_n J_0 D_n', & \textit{if } t = k,\ i = j = 0 \\ \frac{\boldsymbol{\tau}_0}{4} I_{n^2}, & \textit{if } t = k,\ i = j \neq 0 \\ \frac{1}{4} K_{nn}, & \textit{if } t \neq k,\ i = t - k,\ j = k - t \\ 0, & \textit{otherwise.} \end{cases}$$

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

Proof. First notice that (see (9) and (A.3))

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

Consider the case $t = k$ and $i = j = 0$. The preceding fact and independence of ρ_t and υ_t yield

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

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 (15)). 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 non-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 $\mathbb{E}(\upsilon_t\upsilon_t') = n^{-1}I_n$.

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, $\mathbb{E}(\varepsilon_k \otimes e_{0t}) = 0$. 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

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

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

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

Now we can prove Proposition 1.

Proof of Proposition 2. 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 $\mathcal{I}_{\theta\theta}(\theta_0)$. The positive definiteness of $\mathcal{I}_{\theta\theta}(\theta_0)$ 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 $\mathcal{I}_{\theta\theta}(\theta_0)$ 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 $\mathcal{I}_{\vartheta_1\vartheta_1}(\theta_0)$. From the definitions and (5) it can be seen that $U_{0,t-1}$ and e_{0t} are independent. Thus, (B.3), (A.7), and straightforward calculation give $\mathbb{E}(\partial g_t(\theta_0)/\partial\vartheta_1) = 0$ and, furthermore,

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

Block $\mathcal{I}_{\vartheta_2\vartheta_2}(\theta_0)$. Deriving $\mathcal{I}_{\vartheta_2\vartheta_2}(\theta_0)$ is somewhat complicated. From the expression of $\partial g_t(\theta_0)/\vartheta_2$ (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

$\Phi(z)^{-1} = L(z) = \sum_{j=0}^{\infty} L_j z^j$ 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 $L(z^{-1}) = \Psi(z)\Pi(z)$ (see the discussion below (7)) 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

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

Notice that from (B.6) we find that

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

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

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

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

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

This in conjunction with (17) and (A.8) shows that $\mathbb{E}(\partial l_T(\theta_0)/\partial\vartheta_2) = 0$, and we proceed to the covariance matrix of the score of ϑ_2.

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

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

Here the former equality is again obtained by using (7) and the definition of $A_0(k,i)$ whereas the latter is justified by Lemma 6. 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 $C_{22}(a,b;\theta_0)$ (see (A.8) and the definition of $\mathcal{I}_{\vartheta_2\vartheta_2}(\theta_0)$). Thus,

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

To see that the right hand side 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 $k = 0 \pm 1, \pm 2, ...$, 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 (5) and (7), and the previously introduced notation $A_0(k,i)$ and $B_0(k)$ ($B_0(k)=0$ for $k<0$) we consider

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

where the latter equality is based on Lemma 6. 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,

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

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 6 (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 (9) that

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

Here the second equality is based on the expression of the density function of ρ_t^2 (see (12)), the third one on Assumption 4(i), and the fourth one on equation (13).

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), (9), 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 (5)).

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

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

Here the former equality is based on (7), 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 6. 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 (7) whereas the remaining ones are obtained from (A.10), (9), (A.3), the independence of the processes ρ_t and υ_t, and the fact that

$\partial g_t(\theta_0)/\partial\lambda$ has zero mean. Thus, because $\mathbb{E}(\upsilon_{t+a-i-c}\otimes\upsilon_t) = n^{-1}\text{vec}(I_n)\,\mathbf{1}(c=a-i)$,

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

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 (9), (A.3), and (A.10))

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

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 (12)).

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

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

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

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

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

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

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 6 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

$$
\begin{aligned}
& [\underline{A}_0(-1) : \cdots : \underline{A}_0(-r-2)] \begin{bmatrix} \Sigma_0^{-1/2}\Pi_{00} \otimes \Sigma_0^{1/2} & 0 \\ \vdots & \Sigma_0^{-1/2}\Pi_{00} \otimes \Sigma_0^{1/2} \\ \Sigma_0^{-1/2}\Pi_{r0} \otimes \Sigma_0^{1/2} & \vdots \\ 0 & \Sigma_0^{-1/2}\Pi_{r0} \otimes \Sigma_0^{1/2} \end{bmatrix} \qquad \text{(B.10)} \\
= & \begin{bmatrix} \sum_{j=0}^{r}\left(\sum_{i=0}^{r}\Psi_{-j-i,0}\Pi_{i0} \otimes \Pi'_{j0}\right) & \sum_{j=0}^{r}\left(\sum_{i=0}^{r}\Psi_{-1-j-i,0}\Pi_{i0} \otimes \Pi'_{j0}\right) \\ \sum_{j=0}^{r}\left(\sum_{i=0}^{r}\Psi_{1-j-i,0}\Pi_{i0} \otimes \Pi'_{j0}\right) & \sum_{j=0}^{r}\left(\sum_{i=0}^{r}\Psi_{-j-i,0}\Pi_{i0} \otimes \Pi'_{j0}\right) \end{bmatrix} \\
= & \begin{bmatrix} \sum_{j=0}^{r}\left(-L_{j0} \otimes \Pi'_{j0}\right) & \sum_{j=0}^{r}\left(-L_{j+1,0} \otimes \Pi'_{j0}\right) \\ \sum_{j=0}^{r}\left(-L_{j-1,0} \otimes \Pi'_{j0}\right) & \sum_{j=0}^{r}\left(-L_{j0} \otimes \Pi'_{j0}\right) \end{bmatrix},
\end{aligned}
$$

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.10) 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 nonsin-

gularity of the matrix

$$\begin{aligned}
&\begin{bmatrix} \sum_{j=0}^{r}\left(-L_{j0}\otimes\Pi_{j0}'\right) & \sum_{j=0}^{r}\left(-L_{j+1,0}\otimes\Pi_{j0}'\right) \\ \sum_{j=0}^{r}\left(-L_{j-1,0}\otimes\Pi_{j0}'\right) & \sum_{j=0}^{r}\left(-L_{j0}\otimes\Pi_{j0}'\right) \end{bmatrix}\begin{bmatrix} I_{n^2} & -\Phi_{10}\otimes I_n \\ 0 & I_{n^2} \end{bmatrix} \\
= &\begin{bmatrix} I_n & L_{10}-\Phi_{10} \\ 0 & I_n \end{bmatrix}\otimes I_n - \sum_{j=1}^{r}\left(\begin{bmatrix} L_{j0} & L_{j+1,0}-L_{j0}\Phi_{10} \\ L_{j-1,0} & L_{j,0}-L_{j-1,0}\Phi_{10} \end{bmatrix}\otimes\Pi_{j0}'\right) \\
= &\begin{bmatrix} I_n & 0 \\ 0 & I_n \end{bmatrix}\otimes I_n - \sum_{j=1}^{r}\left(\begin{bmatrix} L_{j0} & L_{j-1,0}\Phi_{20} \\ L_{j-1,0} & L_{j-2,0}\Phi_{20} \end{bmatrix}\otimes\Pi_{j0}'\right).
\end{aligned}$$

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

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

where $\boldsymbol{\Phi}_0$ is the companion matrix corresponding the matrix polynomial $I_n-\Phi_{10}z-\Phi_{20}z^2$ and $\boldsymbol{\Phi}_0=\mathbf{P}_0\mathbf{D}_0\mathbf{P}_0^{-1}$ 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 $\boldsymbol{\Phi}_0$. These determinants are nonzero because, by the latter condition in (4), the eigenvalues of $\boldsymbol{\Phi}_0$ are smaller than one in absolute value whereas the former condition in (4) implies that the zeros of $\det\Pi(z)$ lie outside the unit disc. This completes the proof of the positive definiteness of $\mathcal{I}_{\theta\theta}(\theta_0)$.

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 (5) and (7) to approximate the processes $u_{t-i}(\vartheta_{10})$ and y_{t+j-i} $(i=1,...,r,\ j=1,...,s)$ in $\partial g_t(\theta_0)/\partial\vartheta_1$ and $\partial g_t(\theta_0)/\partial\vartheta_1$, respectively, by long moving averages. This amounts to replacing $\partial g_t(\theta_0)/\partial\theta$ 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 (5) and (7) decay to zero at a geometric. Details are omitted. □

Proof of Lemma 3. In the same way as in the proof of Step 1 of Proposition 2 we consider the different blocks of $\mathcal{I}_{\theta\theta}(\theta_0)$ separately. For simplicity, we again suppress the subscript from the expectation operator and denote $\mathbb{E}(\cdot)$ instead of $\mathbb{E}_{\theta_0}(\cdot)$.

Block $\mathcal{I}_{\vartheta_1\vartheta_1}(\theta_0)$. 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

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

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

and

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

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

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

We need to show that the last expression equals $-\mathcal{I}_{\vartheta_1\vartheta_1}(\theta_0)$, which follows if $\sum_{i=1}^{3}W^{(i)}_{\vartheta_1\vartheta_1}(a,b)=-C_{11}(a,b)\otimes\Sigma_0^{-1}$. To see this, conclude from the definitions, (9), and the fact

$\mathbb{C}(\upsilon_t) = n^{-1} I_n$ that

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

Using definitions and the expression of the density of ρ_t^2 (see (12)) yields

$$\begin{aligned} & \mathbb{E}\left(h_0\left(\rho_t^2\right)\right) + \frac{2}{n}\mathbb{E}\left(\rho_t^2 \frac{f_0''(\rho_t^2)}{f_0(\rho_t^2)}\right) \qquad \text{(B.11)} \\ = & \frac{\pi^{n/2}}{\Gamma(n/2)}\left(\int_0^\infty \zeta^{n/2-1} f_0'(\zeta)\, d\zeta + \frac{2}{n}\int_0^\infty \zeta^{n/2} f_0''(\zeta)\, d\zeta\right) \\ = & \frac{\pi^{n/2}}{\Gamma(n/2)}\left(\int_0^\infty \zeta^{n/2-1} f_0'(\zeta)\, d\zeta + \frac{2}{n}\zeta^{n/2} f_0'(\zeta)\,|_0^\infty - \int_0^\infty \zeta^{n/2-1} f_0'(\zeta)\, d\zeta\right) \\ = & 0, \end{aligned}$$

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

Regarding $W^{(3)}_{\vartheta_1\vartheta_1}(a,b)$, use again (9) 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 (14)). 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 2, Block $\mathcal{I}_{\vartheta_2\vartheta_2}(\theta_0)$). Thus, it suffices to consider the second term for which we first note that

$$\begin{aligned} \mathbb{E}\left(\rho_t^4 \frac{f_0''(\rho_t^2)}{f_0(\rho_t^2)}\right) &= \frac{\pi^{n/2}}{\Gamma(n/2)}\int_0^\infty \zeta^{n/2+1} f_0''(\zeta)\, d\zeta \\ &= \frac{\pi^{n/2}}{\Gamma(n/2)}\left(\zeta^{n/2+1} f_0'(\zeta)\Big|_0^\infty - \frac{n+2}{2}\int_0^\infty \zeta^{n/2} f_0'(\zeta)\, d\zeta\right) \\ &= n(n+2)/4, \qquad \text{(B.12)} \end{aligned}$$

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

Next define

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

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

and

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

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 (7),

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

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 (9), (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 (7)

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

where the latter equality is again obtained from (9) and the fact $\mathbb{C}(\upsilon_t) = n^{-1}I_n$. From (B.11) and (B.12) we can now conclude that

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

Next, arguments similar to those already used give

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

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

From the preceding derivations we find that

$$\begin{aligned}\sum_{i=1}^{3} W_{\vartheta_2\vartheta_2}^{(i)}(a,b) &= -\boldsymbol{\tau}_0 \sum_{i,j=0}^{r} \sum_{\substack{c=-\infty \\ c\neq 0}}^{\infty} A_0(c+a-i,i)\, A_0(c+b-j,j) \\ &\quad - \sum_{i,j=0}^{r} A_0(a-i,i)\left[4D_n J_0 D_n' + I_n - n(n+2)\mathbb{E}\left(\upsilon_t\upsilon_t' \otimes \upsilon_t\upsilon_t'\right)\right] A_0(b-j,j).\end{aligned}$$

That $\sum_{i=1}^{3} W_{\vartheta_2\vartheta_2}^{(i)}(a,b) = -C_{22}(a,b,;\theta_0)$ holds, can now be obtained from (20) by observing that $\mathbb{E}\left(\upsilon_t\upsilon_t' \otimes \upsilon_t\upsilon_t'\right) = \mathbb{E}\left[\left(\mathrm{vec}(\upsilon_t\upsilon_t')\right)\left(\mathrm{vec}(\upsilon_t\upsilon_t')\right)'\right]$ and that the impact of the term $\mathrm{vec}(I_n)\mathrm{vec}(I_n)'$ in (20) 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 (9) 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

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

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 (5), (7), (9), and the fact $\mathbb{C}(\upsilon_t) = n^{-1}I_n$, are similar to those used in the previous cases. First note that

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

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

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

and

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

From the preceding expressions and (B.11) it is seen that $W_{\vartheta_1\vartheta_2}^{(2)}(a,b) + W_{\vartheta_1\vartheta_2}^{(3)}(a,b) = 0$.

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

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

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

Block $\mathcal{I}_{\sigma\sigma}(\theta_0)$. From (A.15) and (9) we obtain

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

The first term on the right hand side consists of two additive terms. Using (9) and taking expectation the first one can be written as

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

Here the former equality is based on (B.1) and the fact $\mathbb{E}\left(\upsilon_t\upsilon_t'\right)=n^{-1}I_n$ 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

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

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 $\mathrm{vec}(ABC)=(C'\otimes A)\mathrm{vec}(B)$). 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

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

Using (9) and (A.1) one obtains

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

where the latter equality is based on (B.12) and the definition of $\boldsymbol{i}_0$ (see (15)). Thus,

$$\begin{aligned}\mathbb{E}\left(\frac{\partial^2}{\partial\sigma\partial\sigma'}g_t(\theta_0)\right) &= \frac{1}{4}D_n'(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2})\left[n(n+2)\mathbb{E}\left(\upsilon_t\upsilon_t'\otimes\upsilon_t\upsilon_t'\right)-2I_{n^2}\right](\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2})D_n \\ &\quad -\boldsymbol{i}_0D_n'(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2})\mathbb{E}\left(\upsilon_t\upsilon_t'\otimes\upsilon_t\upsilon_t'\right)(\Sigma_0^{-1/2}\otimes\Sigma_0^{-1/2})D_n.\end{aligned}$$

Because $\mathbb{E}\left(\upsilon_t\upsilon_t'\otimes\upsilon_t\upsilon_t'\right)=D_n\mathbb{E}\left((\mathrm{vech}(\upsilon_t\upsilon_t'))(\mathrm{vech}(\upsilon_t\upsilon_t'))D_n'\right.$ the right hand side equals $-\mathcal{I}_{\sigma\sigma}(\theta_0)$ if the expression in the brackets can be replaced by $\mathrm{vec}(I_n)\mathrm{vec}(I_n)'$. From (20) it is seen that this expression can be replaced by $\mathrm{vec}(I_n)\mathrm{vec}(I_n)'+K_{nn}-I_{n^2}$. Thus, the desired result follows because

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

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

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

where the former equality follows from (12) and the latter from Assumption 6(ii) (cf. the corresponding part of the proof of Proposition 2, 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 $\mathbb{E}\left(\partial\epsilon_t(\vartheta_0)/\partial\vartheta_1\right)=0$ (see (A.5)) which imply $\mathbb{E}\left(\partial^2 g_t(\theta_0)/\partial\vartheta_1\partial\sigma'\right)=0$.

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 (7), 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 (9) and (A.1), the expectation equals

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

Here we have used (B.12), the definition of $\boldsymbol{i}_0$ (see (15)), and straightforward calculation. Combining the preceding derivations shows that

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial^2}{\partial\vartheta_2\partial\sigma'} g_t\left(\theta_0\right)\right) &= 2\left(\boldsymbol{i}_0 - \frac{n\left(n+2\right)}{4}\right) \sum_{b=1}^{s} \frac{\partial}{\partial\vartheta_2}\phi_b\left(\vartheta_{20}\right) \sum_{a=0}^{r} A_0\left(b-a,i\right) \mathbb{E}\left(\upsilon_t\upsilon_t' \otimes \upsilon_t\upsilon_t'\right) \\
&\qquad \times \left(\Sigma_0^{-1/2} \otimes \Sigma_0^{-1/2}\right) D_n \\
&\quad + \sum_{b=1}^{s} \frac{\partial}{\partial\vartheta_2}\phi_b\left(\vartheta_{20}\right) \sum_{a=0}^{r} A_0\left(b-a,i\right) \left(\Sigma_0^{-1/2} \otimes \Sigma_0^{-1/2}\right) D_n \\
&= 2\sum_{b=1}^{s} \frac{\partial}{\partial\vartheta_2}\phi_b\left(\vartheta_{20}\right) \sum_{a=0}^{r} A_0\left(b-a,i\right) D_n J_0 D_n' \left(\Sigma_0^{-1/2} \otimes \Sigma_0^{-1/2}\right) D_n,
\end{aligned}$$

where the last expression equals $-\mathcal{I}_{\vartheta_2\sigma}\left(\theta_0\right)$ and the latter equality can be justified by using the definition of J_0, the identity (20), and arguments similar to those already used in the case of block $\mathcal{I}_{\sigma\sigma}\left(\theta_0\right)$ (see the end of that proof).

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

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

and

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

The argument is similar in both cases and also similar to that used in the proof of Proposition 2 (see Block $\mathcal{I}_{\vartheta_2\lambda}\left(\theta_0\right)$). For example, consider the former and use (7) and independence of ϵ_t to write the left hand side of the equality as

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

where that equality is due to (9). 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

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

where the equality is based on (9) and, using (12), 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 4. First note that our Proposition 2 and Lemma 3 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 their proof of Theorem 1, shows that we only need to show that the appropriately standardized Hessian of the log-likelihood function satisfies

$$\sup_{\theta\in\Theta_0}\left\| N^{-1}\sum_{t=r+1}^{T-s-(n-1)r}\left(\frac{\partial^2}{\partial\theta\partial\theta'}g_t(\theta) - \frac{\partial^2}{\partial\theta\partial\theta'}g_t(\theta_0)\right)\right\| \xrightarrow{p} 0, \tag{B.13}$$

where Θ_0 is some small enough compact neighborhood of θ_0 (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.13) 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

$$\mathbb{E}_{\theta_0}\left(\sup_{\theta\in\Theta_0}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}g_t(\theta)\right\|\right) < \infty \tag{B.14}$$

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

We demonstrate (B.14) 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

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

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,

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

where the last equality is obtained from the definition of $\epsilon_t(\vartheta)$ (see (19)) 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 (7) 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 (9) and the latter by Assumption 7.

From (19) 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,

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

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

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

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

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.

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.

Auerbach, A.J., and J. Slemrod (1997). The Economic Effects of the Tax Reform Act of 1986. Journal of Economic Literature 35, 589–632.

Beyer, A., and R.E.A. Farmer (2007). Testing for Indeterminacy: An Application to U.S. Monetary Policy: Comment. American Economic Review 97, 524–529.

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.

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.

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

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

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.

Cecchetti, S.G., and G. Debelle (2006). Has the inflation process changed? Economic Policy, April 2006, 311–352.

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

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

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

Fernández-Villaverde, J., J.F. Rubio Ramírez, T.J. Sargent, and M.W. Watson (2007). ABCs (and Ds) of Understanding VARs. American Economic Review 97, 1021–1026.

Giannone, D., and L. Reichlin (2006). Does Information Help Recovering Structural Shocks from Past Observations? Journal of the European Economic Association 4, 455–465.

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

Hansen, L.P., and T.J. Sargent (1980). Formulating and Estimating Dynamic Linear Rational Expectations Models. Journal of Economic Dynamics and Control 2, 7–46.

Hansen, L.P., and T.J. Sargent (1991). Two Difficulties in Interpreting Vector Autoregressions. In Rational Expectations Econometrics, ed. by L.P. Hansen, and T.J. Sargent, Westview Press, Inc., Boulder, CO, 77–119.

House, C.L., and M.D. Shapiro (2006). Phased-in Tax Cuts and Economic Activity, American Economic Review 96, 1835–1849.

House, C.L., and M.D. Shapiro (2008). Temporary Investment Tax Incentives: Theory with Evidence from Bonus Depreciation. American Economic Review 98, 737–768.

Kasa, K., T.B. Walker, and C.H. Whiteman (2007). Asset Prices in a Time Series Model with Perpetually Disparately Informed, Competitive Traders. Unpublished manuscript, Department of Economics, Simon Fraser University.

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

Leeper, E.M., T.B. Walker, and S.-C. S. Yang (2008). Fiscal Foresight: Analytics and Econometrics. NBER Working Paper 14028.

Lubik, T.A., and F. Schorfheide (2004). Testing for Indeterminacy: An Application to U.S. Monetary Policy. American Economic Review 94, 190–217.

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

Mountford, A., and H. Uhlig (in press). What are Effects of Fiscal Policy Shocks? Journal of Applied Econometrics.

Poterba, J.M. (1988). Are Consumers Forward Looking? Evidence from Fiscal Experiments. American Economic Review, Papers and Proceedings 78, 413–418.

Roberds, W. (1991). Implications of Expected Present Value Budget Balance: Application to Postwar U.S. Data. In Rational Expectations Econometrics, ed. by L.P. Hansen, and T.J. Sargent, Westview Press, Inc., Boulder, CO, 163–175.

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

Salyer, K.D., and S.M. Sheffrin (1998). Spotting Sunspots: Some Evidence in Support of Models with Self-Fulfilling Prophecies. Journal of Monetary Economics 42, 511–523.

Steigerwald, D.G., and C. Stuart (1997). Econometric Estimation of Foresight: Tax Policy and Investment in the United States. Review of Economics and Statistics 79, 32–40.

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

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

Yang, S.-C.S. (2005), Quantifying Tax Effects under Policy Foresight. Journal of Monetary Economics 52, 1557–1568.

Yang, S.-C.S. (2007), Tentative Evidence of Tax Foresight. Economics Letters 96, 30–37.

Figure 1: Quantile-quantile plots of the residuals of the VAR(2,0)-N (upper panel) and VAR(1,1)-t (lower panel) models for the changes in U.S. GDP, government expenditure and government revenue.

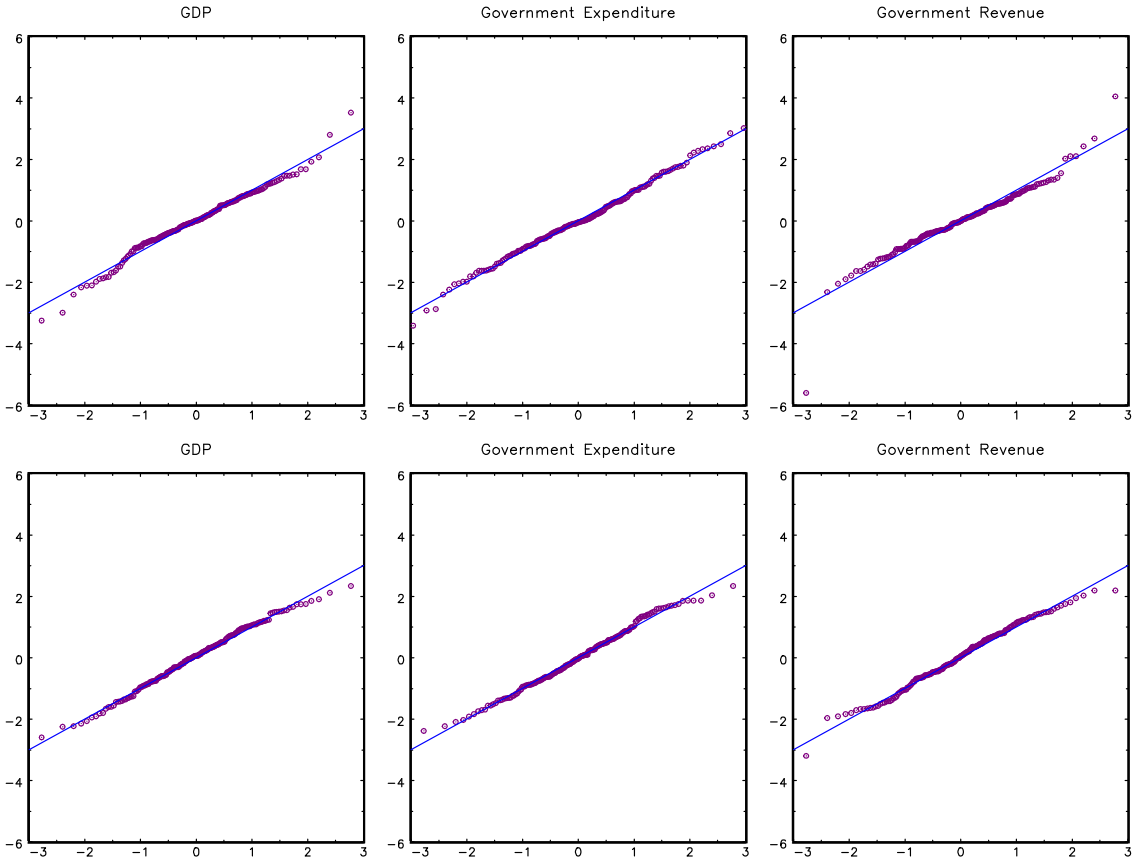

Figure 2: Quantile-quantile plots of the residuals of the VAR(3,0)-N (upper panel) and VAR(2,1)-t (lower panel) models for the U.S. term structure data.

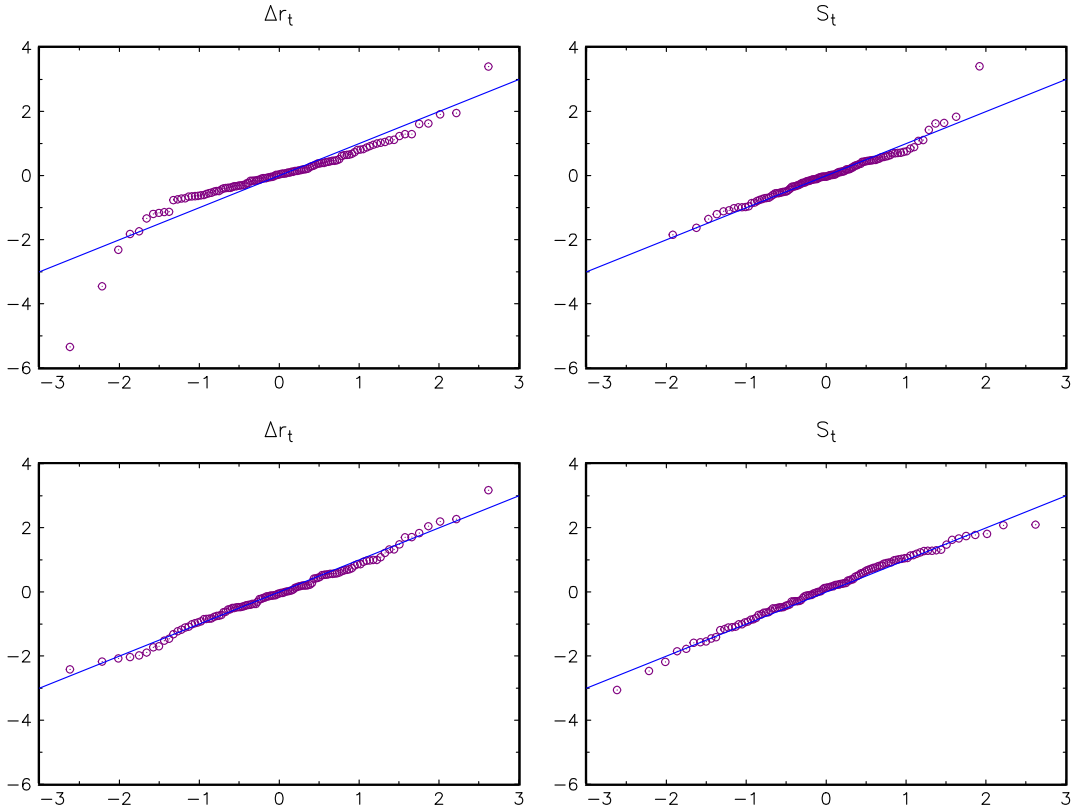

Table 1: Results of diagnostic checks of the second-order VAR models for examining the presence of fiscal foresight.

	Model			
	VAR(2,0)-N	VAR(2,0)-t	VAR(1,1)-t	VAR(0,2)-t
Ljung-Box (4)	0.994	0.712	0.733	0.843
	0.194	0.038	0.065	0.021
	0.633	0.272	0.160	0.281
McLeod-Li (4)	0.084	0.111	0.767	0.940
	0.103	0.085	0.762	0.687
	1.66e–7	1.47e–9	0.969	1.80e–8
Log-likelihood	–960.407	–967.489	–944.947	–949.472

VAR(r, s) denotes the vector autoregressive model for $(\Delta GDP, \Delta Government\ expenditure, \Delta Government\ revenue)'$ 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(1,1)-t model for $(\Delta GDP, \Delta Government\ expenditure, \Delta Government\ revenue)'$.

Π_1	–0.064	0.107	–0.049
	(0.091)	(0.075)	(0.016)
	–0.046	–0.093	0.012
	(0.111)	(0.104)	(0.024)
	0.192	–0.065	–0.333
	(0.315)	(0.205)	(0.071)
Φ_1	0.244	0.006	0.072
	(0.104)	(0.060)	(0.022)
	–0.320	0.268	0.066
	(0.167)	(0.093)	(0.045)
	0.607	–0.045	0.389
	(0.278)	(0.176)	(0.071)
Σ	0.165	0.071	0.346
	(0.024)	(0.017)	(0.055)
	0.071	0.305	0.156
	(0.017)	(0.040)	(0.053)
	0.346	0.156	2.141
	(0.055)	(0.053)	(0.295)
λ	8.253		
	(0.954)		

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

Table 3: 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)	0.710	0.005	0.134	4.58e–5	0.003
	0.725	0.014	0.976	0.002	0.155
McLeod-Li (4)	5.83e–5	0.011	0.478	6.07e–5	0.058
	3.85e–4	0.103	0.879	0.001	0.199
Log-likelihood	–261.443	–110.307	–93.960	–106.453	–106.012

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

Π_1	–0.532	0.464
	(0.107)	(0.138)
	0.248	0.257
	(0.156)	(0.143)
Π_2	–0.338	0.306
	(0.055)	(0.097)
	0.461	–0.018
	(0.094)	(0.129)
Φ_1	0.441	–0.136
	(0.076)	(0.041)
	–0.191	0.658
	(0.201)	(0.096)
Σ	0.123	–0.095
	(0.023)	(0.043)
	–0.095	0.184
	(0.043)	(0.074)
λ	8.187	
	(1.214)	

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