Meitz, Mika; Saikkonen, Pentti

Working Paper

Maximum Likelihood Estimation of a Noninvertible ARMA Model with Autoregressive Conditional Heteroskedasticity

Working Paper, No. 1226

Provided in Cooperation with:
Koç University - TÜSİAD Economic Research Forum, Istanbul

Suggested Citation: Meitz, Mika; Saikkonen, Pentti (2012) : Maximum Likelihood Estimation of a Noninvertible ARMA Model with Autoregressive Conditional Heteroskedasticity, Working Paper, No. 1226, Koç University-TÜSİAD Economic Research Forum (ERF), Istanbul

This Version is available at:
https://hdl.handle.net/10419/108605

KOÇ UNIVERSITY-TÜSİAD ECONOMIC RESEARCH FORUM
WORKING PAPER SERIES

MAXIMUM LIKELIHOOD ESTIMATION OF A NONINVERTIBLE ARMA MODEL WITH AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY

Mika Meitz
Pentti Saikkonen

Working Paper 1226
September 2012

KOÇ UNIVERSITY-TÜSİAD ECONOMIC RESEARCH FORUM
Rumelifeneri Yolu 34450 Sarıyer/Istanbul

Maximum likelihood estimation of a noninvertible ARMA model with autoregressive conditional heteroskedasticity*

Mika Meitz
Koç University

Pentti Saikkonen
University of Helsinki

June 2011

Abstract

We consider maximum likelihood estimation of a particular noninvertible ARMA model with autoregressive conditionally heteroskedastic (ARCH) errors. The model can be seen as an extension to so-called all-pass models in that it allows for autocorrelation and for more flexible forms of conditional heteroskedasticity. These features may be attractive especially in economic and financial applications. Unlike in previous literature on maximum likelihood estimation of noncausal and/or noninvertible ARMA models and all-pass models, our estimation theory does allow for Gaussian innovations. We give conditions under which a strongly consistent and asymptotically normally distributed solution to the likelihood equations exists, and we also provide a consistent estimator of the limiting covariance matrix.

Keywords: maximum likelihood estimation, autoregressive moving average, ARMA, autoregressive conditional heteroskedasticity, ARCH, noninvertible, noncausal, all-pass, nonminimum phase

JEL Codes: C22, C51

*The second author acknowledges financial support from the Academy of Finland and the OP-Pohjola Group Research Foundation. Address correspondence to: Mika Meitz, Department of Economics, Koç University, Rumelifeneri Yolu, 34450 Sariyer, Istanbul, Turkey; e-mail: mmeitz@ku.edu.tr; or to: Pentti Saikkonen, Department of Mathematics and Statistics, University of Helsinki, P. O. Box 68, FIN–00014 University of Helsinki, Finland; e-mail: pentti.saikkonen@helsinki.fi.

1 Introduction

Autoregressive moving average (ARMA) models are commonly used when modelling the conditional mean of a (strictly) stationary time series. In conventional terminology (see, e.g., Brockwell and Davis (2006)), an ARMA process is called causal if, at each point in time, its components can be expressed as a weighted sum of present and past error terms. On the other hand, it is called invertible if these error terms can be represented as a weighted sum of the present and past components of the process. Stationarity and invertibility are typically expressed by requiring the autoregressive and moving average polynomials to have their roots outside the unit circle. If causality (invertibility) does not hold, the model is called noncausal (noninvertible); see Rosenblatt (2000) or the other references listed below (in some of these references noncausal and/or noninvertible ARMA models are also called 'nonminimum phase').

Much of the literature on ARMA models considers only the conventional stationary and invertible case. A reason for this is that if the error terms are independent and identically distributed (IID) with a Gaussian distribution, also the observed process forms a Gaussian sequence, and in this case a noncausal and/or noninvertible ARMA model will be statistically indistinguishable from a particular causal and invertible ARMA model (see, e.g., Rosenblatt (2000, pp. 10–11)). Therefore, in the Gaussian case causality and invertibility are often imposed to ensure identification. However, in many applications, it seems more reasonable to allow the observed process to be potentially non-Gaussian. Alternatively, after fitting a causal and invertible ARMA model to an observed time series one may find that the residuals appear non-Gaussian. In such cases, noncausal and/or noninvertible ARMA models may be more appropriate and can be distinguished from their conventional causal and invertible counterparts (see op. cit.). Allowing for the possibility of noncausality or noninvertibility can in such cases also lead to a better fit and increased forecast accuracy (see Breidt and Hsu (2005) and Lanne, Luoto, and Saikkonen (2010)). Noncausal and/or noninvertible ARMA models have found applications in various fields. Many of the early applications were in natural sciences or engineering, but recently there have also been applications to economic and financial time series (for such applications, see Huang and Pawitan (2000), Breidt, Davis, and Trindade (2001), Breidt and Hsu (2005), Wu and Davis (2010), and Lanne and Saikkonen (in press)).

Maximum likelihood (ML) estimation in noncausal and/or noninvertible ARMA models has been studied in a number of papers. Breidt, Davis, Lii, and Rosenblatt (1991) discussed the case of noncausal AR models, Lii and Rosenblatt (1992) noninvertible MA models, and Lii and Rosenblatt (1996) noncausal and/or noninvertible ARMA models. Andrews, Davis, and Breidt (2006) consider so called all-pass models, which are noncausal and/or noninvertible ARMA models in which all the roots of the autoregressive polynomial are reciprocals of the roots of the moving average polynomial and vice versa. Estimation of all-pass models based on the least absolute deviation criterion and rank-based methods

are considered in Breidt, Davis, and Trindade (2001) and Andrews, Davis, and Breidt (2007). Other relevant references include Huang and Pawitan (2000), Hsu and Breidt (2009), Lanne and Saikkonen (2009), Wu and Davis (2010), Lanne and Saikkonen (in press), and the monograph Rosenblatt (2000).

All of the above-mentioned literature on noncausal and/or noninvertible ARMA models considers the case in which the errors are IID. Unlike in the causal and invertible case, the observed process will nevertheless be conditionally heteroskedastic (for details, see the discussion in Section 2 below). Indeed, Breidt, Davis, and Trindade (2001) (partially) motivate (linear) all-pass models as alternatives to nonlinear models with time varying conditional variances, such as Autoregressive Conditionally Heteroskedastic (ARCH) models. However, they note that "While all-pass models can generate examples of linear time series with 'nonlinear' behavior, their dependence structure is highly constrained, limiting their ability to compete with ARCH". It is therefore of interest to consider noncausal and/or noninvertible ARMA models with errors that are not IID but themselves conditionally heteroskedastic, as such models may be more appropriate in many applications, especially those in economics and finance. This paper is a first attempt in combining noncausal and/or noninvertible ARMA models and ARCH-type models.

In this paper, we consider a particular noninvertible ARMA model with errors that are not IID, but dependent, following a standard ARCH model. As discussed above, such models may be particularly appealing in economic and financial applications. A typical feature of many financial time series is that they are only mildly autocorrelated and, quite commonly, they are treated as uncorrelated. The particular ARMA structure assumed in our model readily accommodates to such cases. With a simple (linear) parameter restriction the ARMA structure of our model reduces to that assumed in (causal) all-pass models, extending these models to allow for ARCH errors and thereby also addressing the above-mentioned statement of Breidt, Davis, and Trindade (2001). However, a special feature of all-pass models is that they assume uncorrelated data. In this respect, our model is more general and can allow for (potentially mild) autocorrelation, which may be useful in some applications. On the other hand, compared to fully general noninvertible (and possibly noncausal) ARMA models our model is more restricted because, similarly to the previously considered all-pass models, it assumes that all roots of the moving average polynomial lie inside the unit circle, hence excluding the case with roots both inside and outside the unit circle.

As a preliminary step for our developments we give conditions for stationarity, ergodicity, and existence of moments of the data generation process. Theory of ML estimation can then be developed by extending the ideas put forward in the case of noncausal and noninvertible ARMA models and all-pass models with IID errors (see Breidt, Davis, Lii, and Rosenblatt (1991), Lii and Rosenblatt (1996), and Andrews, Davis, and Breidt (2006)). Similarly to Lii and Rosenblatt (1996) we first derive an infeasible likelihood-like function that assumes knowledge of an infinite number of observations and thereafter we show how to obtain a feasible approximate likelihood function that only involves observed

data. The former provides a useful theoretical tool which can be used to obtain results for the latter. We give conditions under which a strongly consistent and asymptotically normally distributed solution to the (approximate) likelihood equations exists, and we also provide a consistent estimator of the limiting covariance matrix. The techniques used in the proofs also resemble those employed in the estimation theory of conventional (causal and invertible) ARMA–ARCH models (see, e.g., Francq and Zakoïan (2004) and Meitz and Saikkonen (in press), and also Berkes and Horváth (2004) in which estimation in ARCH models based on non-Gaussian likelihoods is considered). As already indicated, our results can be specialized to (causal) all-pass models so that we also extend the work of Andrews, Davis, and Breidt (2006) by allowing for ARCH type conditional heteroskedasticity.

In addition to allowing for ARCH errors this paper also differs in another important way from the previous literature on noncausal and/or noninvertible ARMA models. In all previous papers on ML estimation of noncausal and/or noninvertible ARMA models it has been necessary to constrain the IID error sequence to be non-Gaussian. This has been due to the above-mentioned fact that Gaussianity of the errors, or equivalently Gaussianity of the observed time series, makes it impossible for the likelihood function to distinguish the considered noncausal and/or noninvertible ARMA model from the corresponding causal and invertible counterpart with the same autocovariance function. A related consequence is that the (limiting) information matrix will then be singular, and the usual theory of ML estimation breaks down. In our noninvertible ARMA model the errors are dependent and follow an ARCH process. The rescaled innovations (i.e., the process obtained by dividing the errors by their conditional standard deviation) are still assumed to be IID but they are not required to be non-Gaussian. The reason is that, even if the rescaled innovations are Gaussian, the error terms will be non-Gaussian (although conditionally Gaussian) and, consequently, the observed noninvertible ARMA–ARCH process will also be non-Gaussian. Therefore the above-mentioned complications with Gaussian errors vanish, providing an intuition why conventional results on ML estimation are obtained even if the rescaled innovations are Gaussian.

The plan of the paper is as follows. Section 2 introduces the model and the basic assumptions employed. Section 3 first shows how to approximate the likelihood function and then obtains results for the score vector and the Hessian matrix needed to prove the main results presented at the end of the section. Section 4 concludes. All proofs along with auxiliary results are presented in Appendices (further details of the proofs are provided in a Supplementary Appendix that is available from the authors upon request).

Finally, a few notational conventions. Unless otherwise indicated, all vectors will be treated as column vectors. For the sake of uncluttered notation, we shall write $x = (x_1, \ldots, x_n)$ for the (column) vector x where the components x_i may be either scalars or vectors (or both). For any scalar, vector, or matrix x, the Euclidean norm is denoted by $|x|$. For a random variable (scalar, vector, or matrix), the L_p–norm is denoted by $\|X\|_p = (E[|X|^p])^{1/p}$, where $p > 0$ (note that this is a vector norm only when

$p \geq 1$). The indicator function will be denoted $\mathbf{1}(\cdot)$. We use $\mathbf{1}$ also to signify the vector $(1, 0, \ldots, 0)$ whose dimension will be clear from the context. An identity matrix of order n will be denoted by $\mathbf{I}_n$.

2 Model

Let y_t ($t = 0, \pm 1, \pm 2, \ldots$) be a stochastic process generated by

$$a_0(B)\, y_t = b_0(B^{-1})\varepsilon_t, \tag{1}$$

where $a_0(B) = 1 - a_{0,1}B - \cdots - a_{0,P}B^P$, $b_0(B^{-1}) = 1 - b_{0,1}B^{-1} - \cdots - b_{0,Q}B^{-Q}$, and ε_t is a zero mean error term allowed to be conditionally heteroskedastic with the conditional heteroskedasticity modeled by a standard stationary ARCH(R) process (see below). Moreover, B is the usual backward shift operator (e.g., $B^k y_t = y_{t-k}$ for $k = 0, \pm 1, ...$), and the polynomials $a_0(z)$ and $b_0(z)$ have their zeros outside the unit circle so that

$$a_0(z) \neq 0 \quad \text{for} \quad |z| \leq 1 \quad \text{and} \quad b_0(z) \neq 0 \quad \text{for} \quad |z| \leq 1. \tag{2}$$

The former condition in (2) is the usual stationarity condition of an ARMA model. It implies that we have the moving average representation

$$y_t = a_0(B)^{-1} b_0(B^{-1})\varepsilon_t = \sum_{j=-Q}^{\infty} \psi_{0,j}\varepsilon_{t-j} \tag{3}$$

in terms of ε_{t+Q}, ε_{t+Q-1}, ... In this representation, $\psi_{0,j}$ is the coefficient of z^j in the Laurent series expansion of $a_0(z)^{-1} b_0(z^{-1}) \stackrel{def}{=} \psi_0(z)$, which is well defined for $|z| \leq 1 + \delta_a$ with some positive δ_a. Moreover, the coefficients $\psi_{0,j}$ decay to zero at a geometric rate as $j \to \infty$. Because the argument of the polynomial $b_0(\cdot)$ in (1) is B^{-1} and not B, the moving average representation (3) is not in terms of past and present ε_t only but also involves ε_{t+Q}, ε_{t+Q-1}, ..., ε_{t+1}. For the same reason, the latter condition in (2) means that the moving average part of the model is not invertible in the conventional sense. Instead, we have an AR(∞) representation

$$\varepsilon_t = b_0(B^{-1})^{-1} a_0(B)\, y_t = \sum_{j=-P}^{\infty} \pi_{0,j} y_{t+j} \tag{4}$$

in terms of $y_{t-P}, \ldots, y_t, y_{t+1}, \ldots$, so that ε_t is expressed in terms of the future of the process y_t. In this representation, $\pi_{0,j}$ is the coefficient of z^j in the Laurent series expansion of $b_0(z^{-1})^{-1} a_0(z) \stackrel{def}{=} \pi_0(z)$, which is well defined for $|z| \geq 1 - \delta_b$ with some positive δ_b, and the coefficients $\pi_{0,j}$ decay to zero at a geometric rate as $j \to \infty$.

As for the conditionally heteroskedastic error term ε_t, we assume that

$$\varepsilon_t = \sigma_t \eta_t, \tag{5}$$

where η_t is a sequence of continuous IID random variables with zero mean and unit variance and the square of σ_t follows a conventional ARCH(R) process. Specifically,

$$\sigma_t^2 = \omega_0 + \alpha_{0,1}\varepsilon_{t-1}^2 + \cdots + \alpha_{0,R}\varepsilon_{t-R}^2, \tag{6}$$

where the parameters are assumed to satisfy the usual conditions $\omega_0 > 0$, $\alpha_{0,1}, \ldots, \alpha_{0,R} \geq 0$. With suitable further conditions to be discussed shortly the error term ε_t is stationary with $E[\varepsilon_t^2] < \infty$. In the following discussion this will be assumed.

Consider the relation of this model to those discussed in earlier literature. In the special case $P = Q$ and $a_0(z) = b_0(z)$ the observed process y_t exhibits no autocorrelation (as ε_t is clearly an uncorrelated sequence, this follows by observing that in this special case the spectral density of y_t is constant, cf. Breidt, Davis, and Trindade (2001)). In the case of a homoskedastic error term the model is then similar to the (causal) all-pass model studied by Breidt, Davis, and Trindade (2001) and Andrews, Davis, and Breidt (2006, 2007). This model in turn is a special case of the general (possibly) noncausal and noninvertible ARMA model considered by Lii and Rosenblatt (1996) and Wu and Davis (2010). A slight difference in formulation is, however, that in these previous papers the counterpart of the operator $b_0(B^{-1})$ in (1) is replaced by $b_0(B)$ and, correspondingly, the inequality in the latter condition in (2) is reversed. Our formulation is similar to that used in noncausal autoregressive models by Lanne and Saikkonen (in press) and, in the same way as in that paper, it appears convenient in terms of statistical inference (see Section 3.1).

It may be worth noting that in our model the squared volatility process σ_t^2 is, in general, not the conditional variance of y_t given the past history of the process. Using equation (1) and denoting expectations conditional on the past history of y_t's with $E_{t-1}[\cdot]$, it is easy to see that

$$E_{t-1}[y_t] = a_{0,1}y_{t-1} + \cdots + a_{0,P}y_{t-p} + E_{t-1}[\varepsilon_t] - b_{0,1}E_{t-1}[\varepsilon_{t+1}] - \cdots - b_{0,Q}E_{t-1}[\varepsilon_{t+Q}].$$

As equation (4) makes clear, the error term ε_t is correlated with lagged and future values of y_t and, therefore, the conditional expectations on the right hand side of this equation are, in general, nonzero and nonlinear functions of the variables y_{t-j}, $j \geq 1$. Thus, the conditional mean of y_t, and hence also its conditional variance, is not obtained in the same way as in previous (invertible) ARMA models with ARCH errors. Even when the error term ε_t is homoskedastic, that is, $\alpha_{0,1} = \cdots = \alpha_{0,R}$, the conditional mean is, in general, a nonlinear function of past values of the process (see Rosenblatt (2000), Section 5.4). This implies that even without an ARCH term the model exhibits conditional heteroskedasticity albeit of a rather limited type, as already mentioned in the introduction. Although σ_t^2 does not have an interpretation as the conditional variance of y_t, it is still the conditional variance of the error term ε_t given the past history of the error terms; in our context homoskedasticity and conditional heteroskedasticity will refer to properties of the error term ε_t. Specifying σ_t^2 as in (6) also appears to be a natural way to introduce conditionally heteroskedastic errors in the model (1), and it also agrees with the formulation used in conventional causal and invertible ARMA–ARCH models.

Forecasting in the context of noncausal and/or noninvertible models has only recently been considered in Breidt and Hsu (2005) and Lanne, Luoto, and Saikkonen (2010). As their work and the discussion in the preceding paragraph indicates, obtaining forecasts from our noninvertible ARMA–ARCH model is not as straightforward as in the conventional invertible case. Even when the errors are homoskedastic, explicit expressions for the conditional mean and variance of y_t appear difficult to obtain. We leave this for future research.

We now discuss assumptions which, among other things, imply that the preceding infinite sums are well defined. Of the following three assumptions, the first one presents conditions imposed on the innovation η_t, the second one specifies the parameter space, and third will ensure the existence of certain moments.

Assumption 1. *The innovation process η_t is a sequence of IID random variables with $E[\eta_t] = 0$ and $E[\eta_t^2] = 1$. The distribution of η_t is symmetric, and has a (Lebesgue) density $f_\eta(x;\lambda_0)$ which (possibly) depends on a parameter vector λ_0 $(d \times 1)$.*

The conditions imposed on the density of the innovation in Assumption 1 are fairly mild and similar to those used by Andrews, Davis, and Breidt (2006) in all-pass models and Lanne and Saikkonen (in press) in noncausal autoregressive models. Requiring a symmetric distribution is only for simplicity because otherwise the needed calculations and the expression of the limiting covariance matrix of the ML estimator would become extremely involved. Further conditions on the density of the innovation will be imposed later.

Our aim is to estimate the true but unknown parameter value θ_0 that characterizes the data generation process and is assumed lie in the permissible parameter space Θ defined by the following assumption. Denote $a(z) = 1 - a_1 z - \cdots - a_P z^P$ and $b(z) = 1 - b_1 z - \cdots - b_Q z^Q$. Decompose the parameter vector θ as $\theta = (\theta_a, \theta_b, \theta_c, \theta_d)$ where $\theta_a = (a_1, \ldots, a_P)$, $\theta_b = (b_1, \ldots, b_Q)$, $\theta_c = (\omega, \alpha_1, \ldots, \alpha_R)$, and $\theta_d = \lambda \in \mathbb{R}^d$ contain the parameters for the AR-part, MA-part, ARCH-part, and the innovation density, respectively. Now we can formulate the following assumption.

Assumption 2. *The true parameter value θ_0 is an interior point of the permissible parameter space $\Theta = \Theta_a \times \Theta_b \times \Theta_c \times \Theta_d$, where*

$$
\begin{aligned}
&\Theta_a = \left\{(a_1, \ldots, a_P) \in \mathbb{R}^P : a(z) \neq 0 \text{ for } |z| \leq 1\right\}, \\
&\Theta_b = \left\{(b_1, \ldots, b_Q) \in \mathbb{R}^Q : b(z) \neq 0 \text{ for } |z| \leq 1\right\}, \\
&\Theta_c = \left\{(\omega, \alpha_1, \ldots, \alpha_R) \in [0, \infty)^{R+1} : \omega > \underline{\omega} \text{ and } \textstyle\sum_{i=1}^R \alpha_i < 1\right\} \text{ with some } \underline{\omega} > 0, \\
&\Theta_d \subseteq \mathbb{R}^d.
\end{aligned}
$$

The assumption that the true parameter value θ_0 is an interior point of the parameter space is standard and required to establish the asymptotic normality of the ML estimator. One particular

consequence of this is that the true values of the parameters $\alpha_1, \ldots, \alpha_R$ are all positive, which in turn implies that we necessarily have conditional heteroskedasticity.[1]

As will be seen shortly in Lemma 1 below, Assumptions 1 and 2 ensure that the data generation process is well defined with finite second moments. However, for establishing asymptotic normality of the ML estimator, finiteness of fourth order moments of the observed process will be needed. To formulate an assumption ensuring this, we first introduce the matrix

$$\Pi_t = \begin{bmatrix} \alpha_{0,1}\eta_t^2 & \alpha_{0,2} & \cdots & \alpha_{0,R-1} & \alpha_{0,R} \\ \eta_t^2 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}. \tag{7}$$

This matrix can be used to write the ARCH–process (6) in companion form which will be used to prove our results (see the proof of Lemma 1 in Appendix B). Now we can formulate a condition that guarantees finiteness of the needed moments.

Assumption 3. *The matrix $E[\Pi_t \otimes \Pi_t]$ has spectral radius (i.e., largest absolute eigenvalue) strictly less than one.*

Note that this assumption implies that $E\left[\eta_t^4\right] < \infty$. Combined with the previous assumptions it enables us to establish the basic properties of the data generation process needed in subsequent developments. We denote by $\mathcal{F}_t^\eta$ the σ–algebra generated by $\{\eta_{t-j},\ j \geq 0\}$ and state the following lemma whose proof can be found in Appendix B.

Lemma 1. *Suppose Assumptions 1 and 2 hold. Then the process $(y_t, \varepsilon_t, \sigma_t)$ defined by equations (1), (5), and (6) is stationary and ergodic with $E\left[y_t^2\right]$, $E\left[\varepsilon_t^2\right]$, and $E\left[\sigma_t^2\right]$ finite and σ_t $\mathcal{F}_{t-1}^\eta$–measurable, ε_t $\mathcal{F}_t^\eta$–measurable, and y_t $\mathcal{F}_{t+Q}^\eta$–measurable. If Assumption 3 also holds, then $E\left[y_t^4\right]$, $E\left[\varepsilon_t^4\right]$, and $E\left[\sigma_t^4\right]$ are finite.*

As already indicated, Lemma 1 (together with Lemmas A.1 and A.2 in Appendix A) ensures that the infinite sums in (3) and (4) are well defined. Stationarity and ergodicity facilitate the use of conventional limit theorems to prove asymptotic normality of the ML estimator. As already mentioned, finiteness of second moments that follows from Assumptions 1 and 2 is not sufficient for this. Existence of finite fourth moments of y_t guaranteed by Assumption 3 is required. This, however, is not surprising because in this respect the situation has been similar in the previous estimation theory of stationary and invertible ARMA–ARCH models where known proofs also assume finite fourth moments (see Francq and Zakoïan (2004)).

[1] Results for the homoskedastic special case could be obtained by dropping the α–parameters from θ and making similar omissions throughout the paper. The assumptions would also require some adjustments, in particular the distribution of η_t would need to be assumed to be non-Gaussian. To conserve space, we do not pursue this further.

3 Parameter estimation and statistical inference

3.1 Approximate likelihood function

Suppose we have an observed time series $y_{1-P}, \ldots, y_0, y_1, \ldots, y_T$ generated by the process described in the previous section, and our aim is to estimate the unknown parameter θ_0 using these observations. ML estimation can be carried out by extending the ideas put forward in the case of homoskedastic noncausal and noninvertible ARMA models and all-pass models; see Breidt, Davis, Lii, and Rosenblatt (1991), Lii and Rosenblatt (1996), and Andrews, Davis, and Breidt (2006). However, as in these papers, deriving a likelihood-like function to be used for estimation requires some care (the main reason for this being that now y_t depends on both the future and past of η_t). We will initially discuss how to derive a likelihood-like function assuming that y_t is available for all t, and subsequently provide an approximation using only the observed data.

First we introduce counterparts of the processes ε_t and σ_t^2 defined for $\theta \neq \theta_0$. In analogue with (4), set

$$u_t(\theta) = b(B^{-1})^{-1} a(B)\, y_t = \sum_{j=-P}^{\infty} \pi_j y_{t+j},$$

where π_j is the coefficient of z^j in the Laurent series expansion of $b(z^{-1})^{-1}a(z) \stackrel{def}{=} \pi(z)$. Our definition of the permissible parameter space makes it clear that $u_t(\theta)$ is well-defined for all $\theta \in \Theta$. Moreover, $u_t(\theta_0) = \varepsilon_t$. To define the counterpart of σ_t^2, set

$$h_t(\theta) = \omega + \alpha_1 u_{t-1}^2(\theta) + \cdots + \alpha_R u_{t-R}^2(\theta) \tag{8}$$

and notice that $h_t(\theta_0) = \sigma_t^2$. (Note also that $u_t(\theta)$ and $h_t(\theta)$ require the knowledge of infinite future of y_t's so that the likelihood we first derive will not be feasible in practice.)

Now, to obtain an approximation of the likelihood, we first derive the joint density of an augmented data vector using a change of variables argument. To this end, notice that

$$\begin{bmatrix} -a_{0,P} & \cdots & -a_{0,1} & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & -a_{0,P} & \cdots & -a_{0,1} & 1 \end{bmatrix} \begin{bmatrix} y_{1-R-P} \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} 1 & -b_{0,1} & \cdots & -b_{0,Q} & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & -b_{0,1} & \cdots & -b_{0,Q} \end{bmatrix} \begin{bmatrix} \varepsilon_{1-R} \\ \vdots \\ \varepsilon_{T+Q} \end{bmatrix},$$

which can be obtained by using (1) with $t = 1-R, \ldots, T$. More briefly, this relation can be written as

$$\mathbf{M}_a \begin{bmatrix} y_{1-R-P} \\ \vdots \\ y_T \end{bmatrix} = \mathbf{M}_b \begin{bmatrix} \varepsilon_{1-R} \\ \vdots \\ \varepsilon_{T+Q} \end{bmatrix}$$

with obvious definitions of the $(T+R) \times (T+P+R)$ matrix $\mathbf{M}_a$ and the $(T+R) \times (T+Q+R)$

matrix $\mathbf{M}_b$. With further augmenting we obtain the relation

$$\begin{bmatrix} \mathbf{I}_P : \mathbf{0}_{P\times(T+R)} & \\ \mathbf{M}_a & \\ & \mathbf{I}_Q \end{bmatrix} \begin{bmatrix} y_{1-R-P} \\ \vdots \\ y_T \\ \varepsilon_{T+1} \\ \vdots \\ \varepsilon_{T+Q} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_P & \\ & \mathbf{M}_b \\ \mathbf{0}_{Q\times(T+R)} : \mathbf{I}_Q & \end{bmatrix} \begin{bmatrix} y_{1-R-P} \\ \vdots \\ y_{-R} \\ \varepsilon_{1-R} \\ \vdots \\ \varepsilon_{T+Q} \end{bmatrix},$$

where the (square and $(T+P+Q+R)$-dimensional) coefficient matrices have determinants equal to unity.

Using a standard sequential conditioning argument, we can write the joint density function of $(\varepsilon_{T+Q},\ldots,\varepsilon_{1-R},y_{-R},\ldots,y_{1-R-P})$ as

$$\begin{aligned} &f(\varepsilon_{T+Q},\ldots,\varepsilon_{T+1}\mid\varepsilon_T,\ldots,\varepsilon_{1-R},y_{-R},\ldots,y_{1-R-P}) \\ &\cdot\prod_{t=1}^{T} f(\varepsilon_t\mid\varepsilon_{t-1},\ldots,\varepsilon_{1-R},y_{-R},\ldots,y_{1-R-P})\cdot f(\varepsilon_0,\ldots,\varepsilon_{1-R},y_{-R},\ldots,y_{1-R-P}), \end{aligned}$$

where $f(\cdot)$ is a generic notation for a (joint and/or conditional) density function indicated by its arguments. In the homoskedastic case, the variables $\varepsilon_{T+Q},\ldots,\varepsilon_{T+1}$ in the first factor are independent of the conditioning information (for sufficiently large T) but this is no longer the case in the presence of heteroskedasticity. Instead, the first factor can be written as (again using a sequential conditioning argument)

$$\prod_{t=T+1}^{T+Q}\sigma_t^{-1}f_\eta(\varepsilon_t\sigma_t^{-1};\lambda_0)=\prod_{t=T+1}^{T+Q}\frac{1}{h_t^{1/2}(\theta_0)}f_\eta\left(\frac{u_t(\theta_0)}{h_t^{1/2}(\theta_0)};\lambda_0\right),$$

which, among other things, depends on the variables $\varepsilon_T,\ldots,\varepsilon_{T-R+1}$. Similarly, the middle term in the preceding expression equals

$$\prod_{t=1}^{T}\sigma_t^{-1}f_\eta(\varepsilon_t\sigma_t^{-1};\lambda_0)=\prod_{t=1}^{T}\frac{1}{h_t^{1/2}(\theta_0)}f_\eta\left(\frac{u_t(\theta_0)}{h_t^{1/2}(\theta_0)};\lambda_0\right).$$

Using these expressions we can write the logarithm of the joint density function of the augmented data vector $(y_{1-R-P},\ldots,y_T,\varepsilon_{T+1},\ldots,\varepsilon_{T+Q})$ as

$$\begin{aligned} &\sum_{t=T+1}^{T+Q}\left[\log f_\eta\left(\frac{u_t(\theta)}{h_t^{1/2}(\theta)};\lambda\right)-\log h_t^{1/2}(\theta)\right] \\ &+\sum_{t=1}^{T}\left[\log f_\eta\left(\frac{u_t(\theta)}{h_t^{1/2}(\theta)};\lambda\right)-\log h_t^{1/2}(\theta)\right]+\log f(\varepsilon_0,\ldots,\varepsilon_{1-R},y_{-R},\ldots,y_{1-R-P}). \end{aligned}$$

Using Assumptions 1 and 2 and the assumptions to be imposed in subsequent sections it is not difficult to see that the first and the third term in the above expression are stochastically bounded and therefore

asymptotically negligible. This suggests using

$$L_T(\theta) = T^{-1}\sum_{t=1}^{T} l_t(\theta) \quad \text{where} \quad l_t(\theta) = \log f_\eta\left(\frac{u_t(\theta)}{h_t^{1/2}(\theta)};\lambda\right) - \frac{1}{2}\log h_t(\theta)$$

as an approximation to the log-likelihood of the observed data vector $(y_1,\ldots,y_T)$ (conditional on initial values). However, as computing $u_t(\theta)$ and $h_t(\theta)$ for $t = 1,\ldots,T$ is not feasible in terms of the available data, a further approximation is needed.

To obtain a likelihood feasible in practice we need approximations to the sequences $u_t(\theta)$ and $h_t(\theta)$ for $t = 1,\ldots,T$ that are expressible in terms of the observations $y_{1-P},\ldots,y_0,\ y_1,\ldots,y_T$ and the parameters. We first define an approximation $\tilde{u}_t(\theta)$ to the sequence $u_t(\theta)$. To this end, set $\tilde{u}_{T+1}(\theta) = \cdots = \tilde{u}_{T+Q}(\theta) = 0$ and recursively solve for $\tilde{u}_T(\theta),\ldots,\tilde{u}_1(\theta)$ by using the backward recursion

$$\tilde{u}_t(\theta) = y_t - a_1 y_{t-1} - \cdots - a_P y_{t-P} + b_1\tilde{u}_{t+1}(\theta) + \cdots + b_Q\tilde{u}_{t+Q}(\theta), \qquad t = T,\ldots,1.$$

To obtain an approximation $\tilde{h}_t(\theta)$ to the sequence $h_t(\theta)$ we set $\tilde{u}_0(\theta) = u_0,\ldots,\tilde{u}_{1-R}(\theta) = u_{1-R}$, where $u_0,\ldots,u_{1-R}$ are real-valued constants independent of θ.[2] Then we compute $\tilde{h}_1(\theta),\ldots,\tilde{h}_T(\theta)$ as

$$\tilde{h}_t(\theta) = \omega + \alpha_1\tilde{u}_{t-1}^2(\theta) + \cdots + \alpha_R\tilde{u}_{t-R}^2(\theta), \qquad t = 1,\ldots,T. \tag{9}$$

The resulting approximate log-likelihood takes the form

$$\tilde{L}_T(\theta) = T^{-1}\sum_{t=1}^{T}\tilde{l}_t(\theta) \quad \text{where} \quad \tilde{l}_t(\theta) = \log f_\eta\left(\frac{\tilde{u}_t(\theta)}{\tilde{h}_t^{1/2}(\theta)};\lambda\right) - \frac{1}{2}\log\tilde{h}_t(\theta).$$

In practice, estimation is carried out by maximizing $\tilde{L}_T(\theta)$, whereas its infeasible counterpart $L_T(\theta)$ is useful in the subsequent theoretical results.

As mentioned in Section 2, our formulation of the model differs slightly from that used in related previous papers where the counterpart of the operator $b_0(B^{-1})$ in (1) is replaced by $b_0(B)$ and the inequality in the latter condition in (2) is reversed. When this alternative formulation is used the approximate log-likelihood function also involves the term $\log|b_Q|$ (cf. Lii and Rosenblatt (1996) and Andrews, Davis, and Breidt (2006)). This term is absent from our approximate log-likelihood function which makes constructing statistical tests for hypotheses on the unknown order of the polynomial $b_0(z)$ straightforward. Indeed, such hypothesis imply $b_{0,Q} = 0$ which makes the term $\log|b_Q|$ undefined. Dealing with this feature therefore calls for additional explanations (see Andrews, Davis, and Breidt (2006)) not needed when our formulation is used.

[2]The choice $\tilde{u}_{T+1}(\theta) = \cdots = \tilde{u}_{T+Q}(\theta) = 0$ for the end values is a counterpart of the common practice of setting initial values to zero when estimating conventional invertible MA models by conditional maximum likelihood. On the other hand, when the estimation of conventional ARCH models is considered, it is common to set the required initial values to some positive constants, and our choice $\tilde{u}_0(\theta) = u_0,\ldots,\tilde{u}_{1-R}(\theta) = u_{1-R}$ reflects this. These assumptions could be relaxed so that these initializations would become dependent on the observed data and θ, but we do not pursue this further.

3.2 Score Vector

In this and the following subsection, we first consider the infeasible approximate log-likelihood $L_T(\theta)$. Due to stationarity, the function $L_T(\theta)$ is easier to work with than its feasible counterpart $\tilde{L}_T(\theta)$ and, using assumptions to be made subsequently, it can be shown that the score vectors obtained from $L_T(\theta)$ and $\tilde{L}_T(\theta)$ are asymptotically equivalent and so are the corresponding ML estimators (see Section 3.4 and Appendix E).

As a first step, we obtain the asymptotic distribution of the score vector associated with $L_T(\theta)$, evaluated at the true parameter value θ_0. The first partial derivatives of $l_t(\theta)$ are derived in Appendix C (an assumption which guarantees the existence of these partial derivatives will be given shortly). Here we only give the explicit expression of the score evaluated at the true parameter value. First some notation. It will be convenient to decompose the score vector conformably with the decomposition of the parameter vector θ as $\theta = (\theta_a, \theta_b, \theta_c, \theta_d)$. In what follows, we will use a subscript to signify a partial derivative indicated by the subscript, for instance $l_{\theta,t}(\theta) = \frac{\partial}{\partial\theta} l_t(\theta)$, $f_{\eta,x}(x;\lambda) = \frac{\partial}{\partial x} f_\eta(x;\lambda)$, and $f_{\eta,\lambda}(x;\lambda) = \frac{\partial}{\partial\lambda} f_\eta(x;\lambda)$. To make the notation lighter, when taking derivatives with respect to the subvectors θ_a, θ_b, θ_c, or θ_d, we drop the θ and only write the letter a, b, c or d (for instance, we write $l_{a,t}(\theta)$, $h_{a,t}(\theta) = \frac{\partial}{\partial\theta_a} h_t(\theta)$ and $u_{a_p,t}(\theta)$ instead of $l_{\theta_a,t}(\theta)$, $h_{\theta_a,t}(\theta)$ and $\frac{\partial}{\partial a_p} u_t(\theta)$).

From Appendix C we find the score vector of a single observation evaluated at the true parameter value as

$$l_{\theta,t}(\theta_0) = \begin{bmatrix} e_{x,t}\frac{u_{a,t}(\theta_0)}{\sigma_t} - \frac{1}{2}\frac{h_{a,t}(\theta_0)}{\sigma_t^2}\left(e_{x,t}\eta_t + 1\right) \\ e_{x,t}\frac{u_{b,t}(\theta_0)}{\sigma_t} - \frac{1}{2}\frac{h_{b,t}(\theta_0)}{\sigma_t^2}\left(e_{x,t}\eta_t + 1\right) \\ -\frac{1}{2}\frac{h_{c,t}(\theta_0)}{\sigma_t^2}\left(e_{x,t}\eta_t + 1\right) \\ e_{\lambda,t} \end{bmatrix},$$

where $e_{x,t} = \frac{f_{\eta,x}(\eta_t;\lambda_0)}{f_\eta(\eta_t;\lambda_0)}$ and $e_{\lambda,t} = \frac{f_{\eta,\lambda}(\eta_t;\lambda_0)}{f_\eta(\eta_t;\lambda_0)}$, and the components of the vectors $u_{a,t}(\theta_0)$ and $u_{b,t}(\theta_0)$ are given by

$$\begin{aligned} u_{a_p,t}(\theta_0) &= -a_0(B)^{-1}\varepsilon_{t-p} \quad (p = 1,\ldots,P), \\ u_{b_q,t}(\theta_0) &= b_0(B^{-1})^{-1}\varepsilon_{t+q} \quad (q = 1,\ldots,Q), \end{aligned}$$

whereas

$$h_{a,t}(\theta_0) = 2\sum_{r=1}^{R}\alpha_{0,r}\varepsilon_{t-r}u_{a,t-r}(\theta_0), \quad h_{b,t}(\theta_0) = 2\sum_{r=1}^{R}\alpha_{0,r}\varepsilon_{t-r}u_{b,t-r}(\theta_0),$$

and $h_{c,t}(\theta_0) = \left(1, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-R}^2\right)$.

We can now formulate an assumption that ensures that the score vector is well defined and asymptotically normally distributed. Let Θ_0 be a compact convex set contained in the interior of Θ that has θ_0 as an interior point, and partition Θ_0 as $\Theta_0 = \Theta_{0a} \times \Theta_{0b} \times \Theta_{0c} \times \Theta_{0d}$.

Assumption 4.

(i) For all $x \in \mathbb{R}$ *and* $\lambda \in \Theta_{0d}$, $f_\eta(x;\lambda) > 0$ *and* $f_\eta(x;\lambda)$ *is twice continuously differentiable with respect to* $(x;\lambda)$.

(ii) For all $\lambda \in \Theta_{0d}$, $\int x f_\eta(x;\lambda)\,dx = 0$ *and* $\int x^2 f_\eta(x;\lambda)\,dx = 1$.

(iii) The matrix $E[e_{\lambda,t}e'_{\lambda,t}]$ *is positive definite.*

(iv) For all $x \in \mathbb{R}$, *the functions*

$$x^4 \frac{f^2_{\eta,x}(x;\lambda_0)}{f^2_\eta(x;\lambda_0)} \quad \text{and} \quad \frac{f^2_{\eta,\lambda_i}(x;\lambda_0)}{f^2_\eta(x;\lambda_0)}$$

are dominated by $d_1(1+|x|^{d_2})$ *with* $d_1, d_2 \geq 0$ *and* $\int |x|^{d_2} f_\eta(x;\lambda_0)\,dx < \infty$.

(v) For all $x \in \mathbb{R}$ *and* $\lambda \in \Theta_{0d}$, *the function* $|x^2 f_{\eta,\lambda}(x;\lambda)|$ *is dominated by a function* $\overline{f}(x)$ *such that* $\int \overline{f}(x)\,dx < \infty$.

Overall, Assumption 4 requires that the density function $f_\eta(x;\lambda)$ satisfies regularity conditions similar to those in Andrews, Davis, and Breidt (2006). Assumption 4(i) imposes a fairly conventional differentiability condition that ensures the partial derivatives presented above exist for all $x \in \mathbb{R}$ and $\lambda \in \Theta_{0d}$. In Assumption 1 we already required the innovations η_t to have a distribution with mean zero and unit variance for the true parameter value λ_0. Assumption 4(ii) extends this requirement to a neighborhood of λ_0. Milder analogues of the dominance conditions assumed in Assumptions 4(iv) and (v) are also used in Andrews, Davis, and Breidt (2006). Being forced to strengthen these conditions in the present context is a direct consequence of the necessity to have finite fourth moments not needed in the homoskedastic case.

Assumption 4(iii) is similar to condition (A6) of Andrews, Davis, and Breidt (2006) and is needed to show the positive definiteness of the information matrix, that is, the limiting covariance matrix of the rescaled score vector. It is trivially satisfied by distributions such as the normal distribution which are free of the parameter λ (for such distributions the following lemma and results based on it need obvious modifications). Assumption 4(iii) also holds for other commonly used distributions, including the rescaled t–distribution and weighted averages of the normal distribution (cf. Remarks 4 and 5 of Andrews, Davis, and Breidt (2006)).

As already discussed in the Introduction, a notable feature of our estimation theory is that, due to the presence of conditional heteroskedasticity, it also works when the innovation sequence η_t is Gaussian. Without conditional heteroskedasticity Gaussian innovations have to be excluded, as the previous work on ML estimation of homoskedastic noninvertible (and noncausal) ARMA models shows. In the homoskedastic case Gaussian innovations imply Gaussianity of the observed process, and invertible and noninvertible ARMA models become indistinguishable by the autocovariance function and, hence, the likelihood function. Thus, non-Gaussianity is needed to achieve identifiability and also a positive definite information matrix (see, e.g., Lii and Rosenblatt (1996) and Andrews, Davis,

and Breidt (2006)). As our model always contains an ARCH component (see Assumption 2 and the following discussion), the observed process is not Gaussian even if the innovation sequence η_t is Gaussian. Because the likelihood function is determined by the distribution of the observed process this explains why we do not need to rule out Gaussian innovations.

To establish the asymptotic distribution of the score vector evaluated at the true parameter value θ_0 we first derive an expression for the limiting covariance matrix of the rescaled score vector, $\mathrm{Cov}\left[T^{1/2}L_{\theta,T}\left(\theta_0\right)\right]$, and establish its positive definiteness. This is rather tedious and is given in the following lemma. For this lemma, we need additional notation. We set $c_j = \mathbf{1}'\Pi^j\mathbf{1}$ with $\Pi = E\left[\Pi_t\right]$ (see (7)), and let $\psi_{0,j}^{(a)}$ and $\psi_{0,j}^{(b)}$ stand for the coefficients in the power series expansions $a_0\left(z\right)^{-1} = \sum_{j=0}^{\infty}\psi_{0,j}^{(a)}z^j$ and $b_0(z^{-1})^{-1} = \sum_{j=0}^{\infty}\psi_{0,j}^{(b)}z^{-j}$ (these expansions are well defined by Assumption 2; see Appendix A). For $j<0$ the conventions $\psi_{0,j}^{(b)} = \psi_{0,j}^{(a)} = 0$ and $c_j = 0$ will be assumed.

Lemma 2. *If Assumptions 1–4 hold,*

$$\mathrm{Cov}\left[T^{1/2}L_{\theta,T}\left(\theta_0\right)\right] \to \mathcal{I}\left(\theta_0\right) \quad \text{as } T\to\infty,$$

where $\mathcal{I}\left(\theta_0\right)$ is finite, positive definite, and can be expressed as

$$\mathcal{I}\left(\theta_0\right) = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{21}'+\mathbf{B}_{21}' & \mathbf{0}_{P\times(R+1)} & \mathbf{0}_{P\times d} \\ \mathbf{A}_{21}+\mathbf{B}_{21} & \mathbf{A}_{22}+\mathbf{B}_{22} & \mathbf{0}_{Q\times(R+1)} & \mathbf{0}_{Q\times d} \\ \mathbf{0}_{(R+1)\times P} & \mathbf{0}_{(R+1)\times Q} & \mathbf{A}_{33} & \mathbf{A}_{43}' \\ \mathbf{0}_{d\times P} & \mathbf{0}_{d\times Q} & \mathbf{A}_{43} & \mathbf{A}_{44} \end{bmatrix}$$

where

$$\begin{aligned}
\mathbf{A}_{11} &= E\left[e_{x,t}^2\right]E\left[\frac{u_{a,t}(\theta_0)}{\sigma_t}\frac{u_{a,t}'(\theta_0)}{\sigma_t}\right] + \frac{1}{4}E[(e_{x,t}\eta_t+1)^2]E\left[\frac{h_{a,t}(\theta_0)}{\sigma_t^2}\frac{h_{a,t}'(\theta_0)}{\sigma_t^2}\right] \quad (P\times P)\\
\mathbf{A}_{21} &= \frac{1}{4}E\left[\left(e_{x,t}\eta_t+1\right)^2\frac{h_{b,t}(\theta_0)}{\sigma_t^2}\frac{h_{a,t}'(\theta_0)}{\sigma_t^2}\right] - \frac{1}{2}E\left[e_{x,t}\left(e_{x,t}\eta_t+1\right)\frac{h_{b,t}(\theta_0)}{\sigma_t^2}\frac{u_{a,t}'(\theta_0)}{\sigma_t}\right] \quad (Q\times P)\\
\mathbf{A}_{22} &= E\left[e_{x,t}^2\frac{u_{b,t}(\theta_0)}{\sigma_t}\frac{u_{b,t}'(\theta_0)}{\sigma_t}\right] + \frac{1}{4}E\left[\left(e_{x,t}\eta_t+1\right)^2\frac{h_{b,t}(\theta_0)}{\sigma_t^2}\frac{h_{b,t}'(\theta_0)}{\sigma_t^2}\right] \quad (Q\times Q)\\
\mathbf{A}_{33} &= \frac{1}{4}E[(e_{x,t}\eta_t+1)^2]E\left[\frac{h_{c,t}(\theta_0)}{\sigma_t^2}\frac{h_{c,t}'(\theta_0)}{\sigma_t^2}\right] \quad (R\times R)\\
\mathbf{A}_{43} &= -\frac{1}{2}E\left[e_{x,t}\eta_t e_{\lambda,t}\right]E\left[\frac{h_{c,t}'(\theta_0)}{\sigma_t^2}\right] \quad (d\times R)\\
\mathbf{A}_{44} &= E[e_{\lambda,t}e_{\lambda,t}'] \quad (d\times d)
\end{aligned}$$

and typical elements of the matrices $\mathbf{B}_{21}$ $(Q\times P)$ and $\mathbf{B}_{22}$ $(Q\times Q)$ are given by

$$\left(\mathbf{B}_{21}\right)_{q,p} = -\sum_{j=0}^{\infty}\psi_{0,j-q}^{(b)}\psi_{0,j-p}^{(a)} \quad (q=1,\ldots,Q,\ p=1,\ldots,P)$$

and

$$\left(\mathbf{B}_{22}\right)_{q,\tilde{q}} = -4\sum_{j=0}^{\infty}\psi_{0,j-q}^{(b)}\psi_{0,j-\tilde{q}}^{(b)}c_j \quad (q,\tilde{q}=1,\ldots,Q).$$

The expression for the limiting covariance matrix of the score is rather involved. The matrices $\mathbf{A}_{ij}$ $(i,j=1,...,4)$ are obtained from the contemporaneous covariance matrix $\text{Cov}\left[l_{\theta,t}(\theta_0)\right]$, whereas the matrices $\mathbf{B}_{21}$ and $\mathbf{B}_{22}$ are due to the serial correlation in the process $l_{\theta,t}(\theta_0)$.

In the previous literature on noncausal and noninvertible ARMA models, it has been common to use an approximation argument and reduce the proof of the asymptotic normality of the score vector to that of a finite-dependent process for which a suitable central limit theorem is available (see, e.g., Rosenblatt (2000, Ch. 8)). This approach appears tedious in the case with conditional heteroskedasticity. It is shown in the proof of Lemma 2 (see Appendix C) that $E\left[l_{\theta,t}(\theta_0) \mid \mathcal{F}_{t-1}^{\eta}\right]=0$ but, as the process $l_{\theta,t}(\theta_0)$ is serially correlated, a martingale central limit theorem is not applicable (this also follows from the fact that $l_{\theta,t}(\theta_0)$ is not $\mathcal{F}_t^{\eta}$–measurable the reason for this being that $u_{b,t}\left(\theta_0\right)$ depends on future innovations η_{t+j}, $j>0$). However, from Lemma 1 and the expression of $l_{\theta,t}(\theta_0)$ it is readily seen that $l_{\theta,t}(\theta_0)$ is stationary and ergodic and, as the following lemma shows, it is a mixingale. This will allow us to use a central limit theorem due to Scott (1973). The definition of the L_2–mixingale and its size used in the following lemma can be found in McLeish (1975) or Davidson (1994, p. 247) (see also the proof of Lemma 3 in Appendix C).

Lemma 3. *If Assumptions 1–4 hold, the sequence* $\{a'l_{\theta,t}(\theta_0),\mathcal{F}_t^{\eta}\}$ *is an* L_2*–mixingale of size* -1 *for all conformable nonrandom vectors* $a\neq 0$.

Using the preceding lemmas and a mixingale central limit theorem based on the aforementioned theorem of Scott (1973) (see Lemma A.4 in Appendix A) we obtain the asymptotic distribution of the score vector as follows.

Lemma 4. *If Assumptions 1–4 hold,*

$$T^{1/2}L_{\theta,T}\left(\theta_0\right)=T^{-1/2}\sum_{t=1}^{T}l_{\theta,t}\left(\theta_0\right)\xrightarrow{d}N\left(0,\mathcal{I}\left(\theta_0\right)\right),$$

where the positive definite matrix $\mathcal{I}\left(\theta_0\right)$ *is given in Lemma 2.*

3.3 Hessian Matrix

We next consider the Hessian matrix associated with the infeasible approximate log-likelihood function $L_T(\theta)$. Expressions for the required second partial derivatives are given in Appendix D. Similarly to the first partial derivatives we use notations such as $l_{\theta\theta',t}(\theta)=\frac{\partial^2}{\partial\theta\partial\theta'}l_t(\theta)$ and $f_{\eta,xx}(x;\lambda)=\frac{\partial^2}{\partial x^2}f_{\eta}(x;\lambda)$. Our first result shows that the expectation of the Hessian evaluated at the true parameter value coincides with the negative of the information matrix. For this we need the following assumption.

Assumption 5.

(i) For all $x\in\mathbb{R}$ *and all* $\lambda\in\Theta_{0d}$, *the function* $|f_{\eta,\lambda\lambda}\left(x;\lambda\right)|$ *is dominated by a function* $\overline{f}\left(x\right)$ *such that* $\int\overline{f}\left(x\right)dx<\infty$.

(ii) $\int f_{\eta,xx}(x;\lambda_0)\,dx = 0$

(iii) $\int x^2 f_{\eta,xx}(x;\lambda_0)\,dx = 2$

Assumption 5(i) is similar to Assumption 4(v) and imposes a conventional dominance condition which ensures that a certain function that appears in the proof of the following lemma can be differentiated under the integral sign. Assumptions 5(ii) and (iii) coincide with assumptions A3 and A4 used by Andrews, Davis, and Breidt (2006) in a similar context.

Lemma 5. *If Assumptions 1–5 hold,* $E[l_{\theta\theta,t}(\theta_0)] = -\mathcal{I}(\theta_0)$.

To be able to prove the asymptotic normality of the infeasible approximate ML estimator we need uniform convergence of the Hessian matrix in some neighborhood of the true parameter value. Our next assumption is needed to establish this.

Assumption 6. *For all* $x \in \mathbb{R}$ *and all* $\lambda \in \Theta_{0d}$*, the functions*

$$x^4 \frac{f^4_{\eta,x}(x;\lambda)}{f^4_{\eta}(x;\lambda)}, \quad \frac{f^4_{\eta,\lambda}(x;\lambda)}{f^4_{\eta}(x;\lambda)}, \quad x^4 \frac{f^2_{\eta,xx}(x;\lambda)}{f^2_{\eta}(x;\lambda)}, \quad \frac{f^2_{\eta,\lambda x}(x;\lambda)}{f^2_{\eta}(x;\lambda)}, \quad \textit{and} \quad \left|\frac{f_{\eta,\lambda\lambda}(x;\lambda)}{f_{\eta}(x;\lambda)}\right|$$

are dominated by $d_1(1+|x|^{d_2})$ *with* $d_1, d_2 \geq 0$ *and* $\int |x|^{d_2} f_\eta(x;\lambda_0)\,dx < \infty$.

These dominance conditions are very similar to those assumed in condition (A7) of Andrews, Davis, and Breidt (2006). There are some differences, however. As with the moment conditions, the allowance of conditionally heteroskedastic errors makes some of our dominance conditions more stringent than their counterparts in Andrews, Davis, and Breidt (2006). Together with our previous assumptions, Assumption 6 ensures that the Hessian matrix has a finite expectation uniformly over Θ_0. Formally, we can establish the following result.

Lemma 6. *If Assumptions 1–6 hold,*

$$\sup_{\theta\in\Theta_0} |L_{\theta\theta,T}(\theta) - \mathcal{J}(\theta)| \to 0 \quad \textit{a.s.},$$

where $\mathcal{J}(\theta) = E[l_{\theta\theta,t}(\theta)]$ *is continuous at* θ_0.

3.4 Main Results

The preceding Lemmas 4–6 are the key ingredients required in proving that the infeasible approximate log-likelihood equations $L_{\theta,T}(\theta) = 0$ have consistent and asymptotically normally distributed solutions. To ensure that these results carry over to the feasible log-likelihood function $\tilde{L}_T(\theta)$, we have to show that the feasible likelihood is asymptotically 'close' to its infeasible counterpart. The following assumption is sufficient for this.

Assumption 7. *For all $x \in \mathbb{R}$, $\triangle x \in \mathbb{R}$, and $\lambda \in \Theta_{0d}$, and for some $C < \infty$ and $d_1, d_2 > 0$,*

$$|v(x + \triangle x; \lambda) - v(x; \lambda)| \leq C\big((1 + |x|^{d_1})\,|\triangle x| + |\triangle x|^{d_2}\big)$$

for the following choices of the function $v(x; \lambda)$:
(a) *(i)* $v(x;\lambda) = \frac{f_{\eta,x}(x;\lambda)}{f_\eta(x;\lambda)}$, *(ii)* $v(x;\lambda) = \frac{f_{\eta,\lambda}(x;\lambda)}{f_\eta(x;\lambda)}$.
(b) *(i)* $v(x;\lambda) = \frac{f_{\eta,xx}(x;\lambda)}{f_\eta(x;\lambda)}$, *(ii)* $v(x;\lambda) = \frac{f_{\eta,\lambda x}(x;\lambda)}{f_\eta(x;\lambda)}$, *(iii)* $v(x;\lambda) = \frac{f_{\eta,\lambda\lambda}(x;\lambda)}{f_\eta(x;\lambda)}$.

Assumption 7 is an analogue of Assumption B in Lii and Rosenblatt (1992, 1996). In these papers, the innovation density is not allowed to depend on an unknown parameter λ, and for this reason their conditions only include counterparts of conditions a(i) and b(i). Of the conditions in Assumption 7, part (a) together with the earlier Assumptions 1–6 will suffice to prove the following result.

Lemma 7. *If Assumptions 1–6 and 7(a) hold,*
(i) $\sup_{\theta\in\Theta_0}\big|L_T(\theta) - \tilde{L}_T(\theta)\big| \to 0$ *a.s. as* $T \to \infty$,
(ii) $T^{1/2}\sup_{\theta\in\Theta_0}\big|L_{\theta,T}(\theta) - \tilde{L}_{\theta,T}(\theta)\big| \to 0$ *a.s. as* $T \to \infty$.

Lemma 7(i) shows that the feasible log-likelihood and its infeasible counterpart converge to each other uniformly over Θ_0. This fact will enable us to deduce the existence of a consistent solution to $\tilde{L}_{\theta,T}(\theta) = 0$ from the existence of a consistent solution to $L_{\theta,T}(\theta) = 0$ which is a convenient first step for the former result. Part (ii) of this Lemma will be used to show that these consistent solutions are asymptotically equivalent.

We can now state the main result of the paper. Of the conditions presented above, Assumptions 1–6 and 7(a) are enough to ensure the existence of a consistent and asymptotically normal root of the likelihood equations, whereas the additional Assumption 7(b) is required for consistent estimation of the limiting covariance matrix.

Theorem 1. *If Assumptions 1–6 and 7(a) hold, there exists a sequence of solutions $\tilde{\theta}_T$ to the (feasible) likelihood equations $\tilde{L}_{\theta,T}(\theta) = 0$ such that $T^{1/2}(\tilde{\theta}_T - \theta_0) \xrightarrow{d} N(0, \mathcal{I}(\theta_0)^{-1})$ as $T \to \infty$. If Assumption 7(b) also holds, a consistent estimator for the limiting covariance matrix is given by $\tilde{L}^{-1}_{\theta\theta,T}(\tilde{\theta}_T)$, that is, $\tilde{L}^{-1}_{\theta\theta,T}(\tilde{\theta}_T) \to \mathcal{I}(\theta_0)^{-1}$ a.s. as $T \to \infty$.*

Theorem 1 establishes the usual result on consistency and asymptotic normality of local maximizers of the feasible approximate log-likelihood $\tilde{L}_T(\theta)$. Without further assumptions one can straightforwardly extend this result to allow for linear restrictions on the parameter vector θ_0. In particular, using the restriction $\theta_{0,a} = \theta_{0,b}$ we can extend the work of Andrews, Davis, and Breidt (2006) to (causal) all-pass models with ARCH-errors. The consistent estimator of the limiting covariance matrix makes possible to apply conventional test procedures for testing hypotheses on the parameter vector θ_0. For instance, it is straightforward to show that Wald tests and likelihood ratio tests based on the feasible approximate likelihood function have the usual limiting χ^2–distribution under the null hypothesis.

4 Conclusion

In this paper we have developed an asymptotic estimation theory for a particular noninvertible ARMA model with errors that are dependent and follow a standard ARCH model. We give conditions under which a strongly consistent and asymptotically normally distributed solution to the (approximate) likelihood equations exists, and for statistical inference we also provide a consistent estimator of the limiting covariance matrix. The assumptions required to obtain these results are similar to those previously used in ML estimation of noncausal and/or noninvertible ARMA models and all-pass models with IID errors. An important exception is that we do not need to assume the innovations to be non-Gaussian.

This paper also extends previous work on causal all-pass models to allow for ARCH-type conditional heteroskedasticity. Although our ARMA specification is more general than that assumed in all-pass models and allows for autocorrelation, it is still more restricted than in fully general noninvertible (and possibly noncausal) ARMA models because, similarly to previously considered all-pass models, it assumes that all roots of the moving average polynomial lie inside the unit circle. One could consider a more general specification in which roots of the moving average polynomial both inside and outside the unit circle are allowed. Another topic for potential future work is to extend our results to the case where the errors follow a Generalized ARCH (GARCH) model. Finally, empirical applications, especially to economic and financial time series, are of interest and will be considered in future work.

Appendix A: Auxiliary Results

This appendix contains four auxiliary lemmas that will be used to prove our main results. Their proofs are given in the Supplementary Appendix. The first of these lemmas makes precise in what sense the Laurent series expansions used in this paper are well-defined and satisfy the conditions mentioned in the text.

Lemma A.1. *(i) Suppose the polynomial* $a(z;\theta_a) = 1 - a_1(\theta_a) z - \cdots - a_P(\theta_a) z^P$ *satisfies* $a(z;\theta_a) \neq 0$ *for* $|z| \leq 1$ *and* $\theta_a \in \Theta_a$ *and that the functions* $a_j(\theta_a)$ $(j = 1, \ldots, P)$ *are continuous. Then, for each* $\theta_a^\bullet \in \Theta_a$ *there exists a neighborhood* $N(\theta_a^\bullet)$ *of* $\theta_a^\bullet$ *such that for all* $\theta_a \in N(\theta_a^\bullet)$, $a(z;\theta_a)$ *has an inverse* $a(z;\theta_a)^{-1} = \sum_{j=0}^{\infty} \psi_j^{(a)}(\theta_a) z^j$ *defined by a Laurent series expansion that is absolutely convergent for* $|z| \leq 1 + \delta(\theta_a^\bullet)$ *with some positive* $\delta(\theta_a^\bullet)$. *Moreover, the coefficients in this expansion satisfy* $\sup_{\theta_a \in N(\theta_a^\bullet)} |\psi_j^{(a)}(\theta_a)| \leq C\rho_a^j$, $j = 0, 1, 2, \ldots$, *with some* $C < \infty$ *and* $\rho_a < 1$ *(that both depend on* $\theta_a^\bullet$*).*

(ii) Suppose the polynomial $b(z;\theta_b) = 1 - b_1(\theta_b) z - \cdots - b_Q(\theta_b) z^Q$ *satisfies* $b(z;\theta_b) \neq 0$ *for* $|z| \leq 1$ *and* $\theta_b \in \Theta_b$ *and that the functions* $b_j(\theta_b)$ $(j = 1, \ldots, Q)$ *are continuous. Then, for each* $\theta_b^\bullet \in \Theta_b$ *there exists a neighborhood* $N(\theta_b^\bullet)$ *of* $\theta_b^\bullet$ *such that for all* $\theta_b \in N(\theta_b^\bullet)$, $b(z^{-1};\theta_b)$ *has an inverse* $b(z^{-1};\theta_b)^{-1} = \sum_{j=0}^{\infty} \psi_j^{(b)}(\theta_b) z^{-j}$ *defined by a Laurent series expansion that is absolutely convergent for* $1 - \delta(\theta_b^\bullet) \leq |z|$ *with some positive* $\delta(\theta_b^\bullet)$. *Moreover, the coefficients in this expansion satisfy* $\sup_{\theta_b \in N(\theta_b^\bullet)} |\psi_j^{(b)}(\theta_b)| \leq C\rho_b^j$, $j = 0, 1, 2, \ldots$, *with some* $C < \infty$ *and* $\rho_b < 1$ *(that both depend on* $\theta_b^\bullet$*).*

This result makes clear that the Laurent series expansions considered in this paper for functions of the type $a(z;\theta_a)^{-1}$, $b\left(z^{-1};\theta_b\right)^{-1}$, $a(z;\theta_a)^{-1} b\left(z^{-1};\theta_b\right)$, $a(z;\theta_a) b\left(z^{-1};\theta_b\right)^{-1}$, etc., are well-defined (at least) in some annulus $1 - \delta(\theta_b^\bullet) \leq |z| \leq 1 + \delta(\theta_a^\bullet)$ containing the unit circle and that the coefficients, say $\psi_j(\theta_a, \theta_b)$, in those expansions always satisfy $\sup_{(\theta_a,\theta_b) \in N(\theta_a^\bullet, \theta_b^\bullet)} |\psi_j(\theta_a, \theta_b)| \leq C\rho^{|j|}$, $j = 0, \pm 1, \pm 2, \ldots$, with some $C < \infty$ and $\rho < 1$ and some neighborhood $N(\theta_a^\bullet, \theta_b^\bullet)$ of a point of interest $(\theta_a^\bullet, \theta_b^\bullet)$.

The next lemma shows that random processes defined via suitable convergent series expansions (e.g., $y_t = \sum_{j=-Q}^{\infty} \psi_{0,j} \varepsilon_{t-j}$ in (3)) are well-defined, and also gives conditions for existence of moments.

Lemma A.2. *Consider a stationary process* $X_t(\phi)$ *depending on a parameter* ϕ, $\phi \in \Phi$, *and satisfying* $\left\| \sup_{\phi \in \Phi} |X_t(\phi)| \right\|_r < \infty$ *with some* $r > 0$. *Moreover, let the sequence of constants* $\kappa_j(\phi)$, *also depending on* ϕ, *satisfy* $\sup_{\phi \in \Phi} |\kappa_j(\phi)| \leq C\rho^{|j|}$, $j = 0, \pm 1, \pm 2, \ldots$, *with some* $C < \infty$ *and* $\rho < 1$. *Then for each* $\phi \in \Phi$, *the series* $\sum_{j=-\infty}^{\infty} \kappa_j(\phi) X_{t-j}(\phi)$ *converges with probability one, and if one defines* $Y_t(\phi) = \sum_{j=-\infty}^{\infty} \kappa_j(\phi) X_{t-j}(\phi)$, *the process* $Y_t(\phi)$ *satisfies* $\left\| \sup_{\phi \in \Phi} |Y_t(\phi)| \right\|_r < \infty$.

Our third auxiliary lemma, which is similar to Lemma 4.1 of Francq and Zakoïan (2004), concerns the expectations of transformations of symmetric random variables, and will be used repeatedly to prove the main results of the paper.

Lemma A.3. *Let* $\{Z_i\}_{i=-\infty}^{\infty}$ *be a sequence of independent and identically distributed symmetric random variables and let* $Y = h(\ldots, Z_{i-1}, Z_i, Z_{i+1}, \ldots)$ *with a measurable function* $h(\ldots, z_{i-1}, z_i, z_{i+1}, \ldots)$.

Suppose the function $h(\ldots, z_{i-1}, z_i, z_{i+1}, \ldots)$ *is an odd function of* z_i*, that is,* $h(\ldots, z_{i-1}, z_i, z_{i+1}, \ldots) = -h(\ldots, z_{i-1}, -z_i, z_{i+1}, \ldots)$*, and that* $E[|Y|] < \infty$*. Then* $E[Y] = 0$*.*

The final lemma of this appendix presents a mixingale central limit theorem that can be applied to establish asymptotic normality of the score vector. The definition of an L_2–mixingale and its size can be found in McLeish (1975) or Davidson (1994, p. 247).

Lemma A.4. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space,* $(Z_t)_{t=-\infty}^{\infty}$ *and* $(\epsilon_t)_{t=-\infty}^{\infty}$ *two doubly-infinite sequences of stationary ergodic random variables defined on* $(\Omega, \mathcal{F}, P)$*, and* $\mathcal{F}_t^{\epsilon}$*,* $t \in \mathbb{Z}$*, an increasing sequence of sigma-algebras with* $\mathcal{F}_t^{\epsilon} = \sigma(\epsilon_t, \epsilon_{t-1}, \ldots)$ *the sigma-algebra generated by present and past random variables* ϵ_t*. Suppose* $E[Z_t] = 0$*,* $E[Z_t^2] < \infty$*, and that* $\{Z_t, \mathcal{F}_t^{\epsilon}\}$ *is an* L_2*–mixingale with size* -1*. Denote* $S_n = \sum_{t=1}^{n} Z_t$*. Then* $Var\left(n^{-1/2} S_n\right) \to \sigma^2$ *with* $0 \leq \sigma^2 < \infty$*, and if* $0 < \sigma^2$*, then* $n^{-1/2} S_n \xrightarrow{d} N(0, \sigma^2)$*.*

Appendix B: Data Generating Process

Proof of Lemma 1. Using the definition of the matrix Π_t in (7) we can write the ARCH(R) model in companion form as $X_t = \Pi_{t-1} X_{t-1} + \varpi$, where the R–dimensional vectors X_t and ϖ are defined as $X_t = \left(\sigma_t^2, \varepsilon_{t-1}^2, ..., \varepsilon_{t-R+1}^2\right)$ and $\varpi = (\omega_0, 0, ..., 0)$. (In the case $R = 1$, define them as the scalars $X_t = \sigma_t^2$ and $\varpi = \omega_0$.) Also define

$$\Pi = E[\Pi_t] = \begin{bmatrix} \alpha_{0,1} & \cdots & \alpha_{0,R-1} & \alpha_{0,R} \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}.$$

By the constraints imposed on Θ_c in Assumption 2, we have $\sum_{i=1}^{R} \alpha_{0,i} < 1$ or, equivalently, the spectral radius of the matrix Π is strictly less than one (see, e.g., Proposition 1 in Francq and Zakoïan (2004)). Now, proceeding in the same way as in the proof of Lemma 2.1 of Chen and An (1998) we can conclude that the process σ_t^2 is stationary and ergodic with the almost sure representation

$$\sigma_t^2 = \mathbf{1}' \left(\mathbf{I}_R + \sum_{k=1}^{\infty} \prod_{l=1}^{k} \Pi_{t-l} \right) \varpi.$$

(The companion form used by Chen and An (1998) is slightly different from ours, but this has no impact on the employed arguments.) Moreover, $E\left[\sigma_t^2\right] < \infty$ and σ_t is $\mathcal{F}_{t-1}^{\eta}$–measurable. By (5) and Assumption 1, $E\left[\varepsilon_t^2\right] < \infty$ and ε_t is $\mathcal{F}_t^{\eta}$–measurable. The results concerning y_t follow from the representation (3) and Lemmas A.1 and A.2. If Assumption 3 also holds, we can repeat the arguments used in the proof of Theorem 3.1(i) of Chen and An (1998) by using our companion form and conclude that finiteness of the fourth moments mentioned in the lemma follow. ■

We next present a lemma concerning expectations containing the process σ_t^2 that will repeatedly be used in subsequent proofs. Recall the notation $c_j = \mathbf{1}'\Pi^j\mathbf{1}$ for $j \geq 0$, and $c_j = 0$ for $j < 0$.

Lemma B.1. *Suppose Assumptions 1–3 hold. Let $g(\cdot)$ be a (measurable) function such that $E[|g(\eta_t)|] < \infty$ and $E\left[|g(\eta_t)|\,\eta_t^2\right] < \infty$, and let H_{t-1} be an $\mathcal{F}_{t-1}^{\eta}$–measurable random variable with $E[|H_{t-1}|] < \infty$ and $E[|H_{t-1}X_t|] < \infty$. Let a be a nonnegative integer. Then the expectation $E\left[H_{t-1}g(\eta_t)\frac{\sigma_{t+a}^2}{\sigma_t^2}\eta_{t+a}^2\right]$ is finite and equals $c_aE[H_{t-1}]\,E\left[g(\eta_t)\eta_t^2\right]$ if $E[g(\eta_t)] = 0$ whereas in general it equals*

$$\omega_0\sum_{k=0}^{a-1}c_kE\left[\frac{H_{t-1}}{\sigma_t^2}\right]E[g(\eta_t)] + c_aE[H_{t-1}]\,E\left[g(\eta_t)\eta_t^2\right] + E[g(\eta_t)]\left(\mathbf{1}'\Pi^aE\left[\frac{H_{t-1}X_t}{\sigma_t^2}\right] - c_aE[H_{t-1}]\right).$$

The proof of this lemma is given in the Supplementary Appendix. For later purposes we also note that the constants c_j satisfy the difference equation

$$c_j = \sum_{r=1}^{R}\alpha_{0,j}c_{j-r}, \quad j > 0. \tag{10}$$

This can be justified as follows. Consider the stationary autoregressive process $x_t = \alpha_{0,1}x_{t-1} + \cdots + \alpha_{0,R}x_{t-R} + \xi_t$ where ξ_t is a white noise sequence. Define $\alpha(z) = 1 - \alpha_{0,1}z - \cdots - \alpha_{0,R}z^R$ and write x_t in the moving average form $x_t = \alpha(B)^{-1}\xi_t$. Observing that Π is the coefficient matrix of the companion form of the process x_t we also have $x_t = \sum_{j=0}^{\infty}\mathbf{1}'\Pi^j\mathbf{1}\xi_{t-j} = \sum_{j=0}^{\infty}c_j\xi_{t-j}$. Thus, $\alpha(z)^{-1} = \sum_{j=0}^{\infty}c_jz^j$ from which equation (10) can be deduced.

Appendix C: Score Vector

Expression for the score vector. As in Section 3.2, we use a subscript to denote a partial derivative indicated by the subscript. For notational brevity, denote

$$e_{x,t}(\theta) = \frac{f_{\eta,x}(h_t^{-1/2}(\theta)u_t(\theta);\lambda)}{f_\eta(h_t^{-1/2}(\theta)u_t(\theta);\lambda)} \quad \text{and} \quad e_{\lambda,t}(\theta) = \frac{f_{\eta,\lambda}(h_t^{-1/2}(\theta)u_t(\theta);\lambda)}{f_\eta(h_t^{-1/2}(\theta)u_t(\theta);\lambda)}.$$

Then, with straightforward differentiation one obtains

$$l_{\theta,t}(\theta) = \begin{bmatrix} e_{x,t}(\theta)\frac{u_{a,t}(\theta)}{h_t^{1/2}(\theta)} - \frac{1}{2}\frac{h_{a,t}(\theta)}{h_t(\theta)}\left(e_{x,t}(\theta)\frac{u_t(\theta)}{h_t^{1/2}(\theta)} + 1\right) \\ e_{x,t}(\theta)\frac{u_{b,t}(\theta)}{h_t^{1/2}(\theta)} - \frac{1}{2}\frac{h_{b,t}(\theta)}{h_t(\theta)}\left(e_{x,t}(\theta)\frac{u_t(\theta)}{h_t^{1/2}(\theta)} + 1\right) \\ -\frac{1}{2}\frac{h_{c,t}(\theta)}{h_t(\theta)}\left(e_{x,t}(\theta)\frac{u_t(\theta)}{h_t^{1/2}(\theta)} + 1\right) \\ e_{\lambda,t}(\theta) \end{bmatrix}.$$

Expressions for the partial derivatives of $u_t(\theta)$ and $h_t(\theta)$. Next, we derive expressions for the quantities $h_{a,t}(\theta)$, $h_{b,t}(\theta)$, $h_{c,t}(\theta)$, $u_{a,t}(\theta)$, and $u_{b,t}(\theta)$ that appear in the score vector. As $h_t(\theta) = \omega + \alpha_1u_{t-1}^2(\theta) + \cdots + \alpha_Ru_{t-R}^2(\theta)$, straightforward computation gives

$$h_{a,t}(\theta) = 2\sum_{r=1}^{R}\alpha_r u_{t-r}(\theta)\,u_{a,t-r}(\theta), \qquad h_{b,t}(\theta) = 2\sum_{r=1}^{R}\alpha_r u_{t-r}(\theta)\,u_{b,t-r}(\theta),$$

and $h_{c,t}(\theta) = \left(1, u_{t-1}^2(\theta), \ldots, u_{t-R}^2(\theta)\right)$. As for the partial derivatives of $u_t(\theta)$, using the representation $u_t(\theta) = b(B^{-1})^{-1} a(B) y_t$ and noting that $\frac{\partial}{\partial a_p} a(B) y_t = -y_{t-p}$ gives

$$u_{a_p,t}(\theta) = -b\left(B^{-1}\right)^{-1} y_{t-p} = -a(B)^{-1} u_{t-p}(\theta) \quad (p = 1, \ldots, P).$$

Next note that $\frac{\partial}{\partial b_q} b(B^{-1}) u_t(\theta) = -B^{-q} u_t(\theta) + b(B^{-1}) u_{b_q,t}(\theta)$. From the relation $b(B^{-1}) u_t(\theta) = a(B) y_t$ it follows that the left hand side is zero, and hence

$$u_{b_q,t}(\theta) = b\left(B^{-1}\right)^{-1} u_{t+q}(\theta) \quad (q = 1, \ldots, Q).$$

This completes the derivation of the score vector.

An auxiliary lemma. The following lemma contains results needed in subsequent derivations. Its proof is straightforward and is given in the Supplementary Appendix.

Lemma C.1. *If Assumptions 1–4 hold, then (i)* $E[e_{x,t}^2 \eta_t^4] < \infty$, *(ii)* $E[e_{x,t}^2] < \infty$, *(iii)* $E[e_{x,t}^2 \eta_t^2] < \infty$, *(iv)* $E[|e_{x,t}\eta_t^3|] < \infty$, *(v)* $E[e_{\lambda,t} e'_{\lambda,t}] < \infty$, *(vi)* $E[|e_{\lambda,t} e_{x,t} \eta_t|] < \infty$, *(vii)* $E[e_{x,t}] = 0$, *(viii)* $E[e_{x,t}\eta_t + 1] = 0$, *(ix)* $E\left[e_{x,t}\eta_t^2\right] = 0$, *(x)* $E\left[e_{x,t}\eta_t^3\right] = -3$, *(xi)* $E[e_{\lambda,t}] = 0$, *(xii)* $E\left[\eta_t^2 e_{\lambda,t}\right] = 0$.

Proof of Lemma 2. We present the long proof in several steps. In Step 1, we show that $E\left[l_{\theta,t}(\theta_0) \mid \mathcal{F}_{t-1}^{\eta}\right] = 0$, and hence also that $E[l_{\theta,t}(\theta_0)] = 0$. In Step 2, we derive the expressions of the matrices $\mathbf{A}_{ij}$ and $\mathbf{B}_{ij}$, whereas in Step 3 we establish that these matrices are finite. Step 4 shows that $\mathcal{I}(\theta_0)$ is positive definite. In what follows, we will repeatedly make use of the following expansions for the components of $u_{a,t}(\theta_0)$, $u_{b,t}(\theta_0)$, $h_{a,t}(\theta_0)$, and $h_{b,t}(\theta_0)$ (as before, p will range over the values $1, \ldots, P$, and q over the values $1, \ldots, Q$):

$$u_{a_p,t}(\theta_0) = -a_0(B)^{-1}[\sigma_{t-p}\eta_{t-p}] = -\sum_{i=0}^{\infty} \psi_{0,i}^{(a)} \sigma_{t-p-i}\eta_{t-p-i} \tag{11a}$$

$$u_{b_q,t}(\theta_0) = b_0\left(B^{-1}\right)^{-1}[\sigma_{t+q}\eta_{t+q}] = \sum_{j=0}^{\infty} \psi_{0,j}^{(b)} \sigma_{t+q+j}\eta_{t+q+j} \tag{11b}$$

$$h_{a_p,t}(\theta_0) = 2\sum_{r=1}^{R} \alpha_{0,r}\varepsilon_{t-r} u_{a_p,t-r}(\theta_0) = -2\sum_{r=1}^{R}\sum_{i=0}^{\infty} \alpha_{0,r}\psi_{0,i}^{(a)} \sigma_{t-r}\eta_{t-r}\sigma_{t-r-p-i}\eta_{t-r-p-i} \tag{11c}$$

$$h_{b_q,t}(\theta_0) = 2\sum_{r=1}^{R} \alpha_{0,r}\varepsilon_{t-r} u_{b_q,t-r}(\theta_0) = 2\sum_{r=1}^{R}\sum_{j=0}^{\infty} \alpha_{0,r}\psi_{0,j}^{(b)} \sigma_{t-r}\eta_{t-r}\sigma_{t-r+q+j}\eta_{t-r+q+j} \tag{11d}$$

Step 1. First note that $E[e_{x,t}] = 0$, $E[e_{x,t}\eta_t + 1] = 0$, $E\left[e_{x,t}\eta_t^2\right] = 0$, and $E[e_{\lambda,t}] = 0$ (see Lemma C.1), and that $e_{x,t}$ is independent of $\mathcal{F}_{t-1}^{\eta}$. Also note that $u_{a,t}(\theta_0)$, $h_{a,t}(\theta_0)$, $h_{c,t}(\theta_0)$, and σ_t are all $\mathcal{F}_{t-1}^{\eta}$–measurable. Therefore, $E\left[l_{\theta,t}(\theta_0) \mid \mathcal{F}_{t-1}^{\eta}\right] = 0$ clearly holds for the components corresponding to the sub-vectors $l_{a,t}(\theta_0)$, $l_{c,t}(\theta_0)$, and $l_{d,t}(\theta_0)$. Now consider the sub-vector $l_{b,t}(\theta_0) = e_{x,t}\frac{u_{b,t}(\theta_0)}{\sigma_t} - \frac{1}{2}\frac{h_{b,t}(\theta_0)}{\sigma_t^2}(e_{x,t}\eta_t + 1)$, where the terms $u_{b,t}(\theta_0)$ and $h_{b,t}(\theta_0)$ are not $\mathcal{F}_{t-1}^{\eta}$–measurable. As σ_t is $\mathcal{F}_{t-1}^{\eta}$–measurable, it suffices to establish that $E\left[e_{x,t} u_{b,t}(\theta_0) \mid \mathcal{F}_{t-1}^{\eta}\right] = 0$ and $E\left[(e_{x,t}\eta_t + 1) h_{b,t}(\theta_0) \mid \mathcal{F}_{t-1}^{\eta}\right] =$

0. The former result is obtained straightforwardly by using the expansion of $u_{b_q,t}(\theta_0)$ in (11b) and the law of iterated expectations conditioning on $\mathcal{F}^{\eta}_{t+q+j-1}$. For the latter result, note that in the expansion of $h_{b_q,t}(\theta_0)$ in (11d) the terms σ_{t-r} and η_{t-r} are $\mathcal{F}^{\eta}_{t-1}$–measurable. Therefore it suffices to show that $E\left[(e_{x,t}\eta_t+1)\,\sigma_{t-r+q+j}\eta_{t-r+q+j}\mid\mathcal{F}^{\eta}_{t-1}\right]=0$ for all $r=1,\ldots,R$, $j\geq 0$, and $q=1,\ldots,Q$. If $t-r+q+j<t$, this follows from the $\mathcal{F}^{\eta}_{t-1}$–measurability of $\sigma_{t-r+q+j}\eta_{t-r+q+j}$ and the fact that $(e_{x,t}\eta_t+1)$ is independent of $\mathcal{F}^{\eta}_{t-1}$ and has expectation zero. If $t-r+q+j>t$, one can apply the law of iterated expectations conditioning on $\mathcal{F}^{\eta}_{t-r+q+j-1}$, and again show that the expectation is zero making use of the fact that $\sigma_{t-r+q+j}$ is $\mathcal{F}^{\eta}_{t-r+q+j-1}$–measurable. Finally, if $t-r+q+j=t$, we need to show that $E\left[(e_{x,t}\eta_t+1)\,\sigma_t\eta_t\mid\mathcal{F}^{\eta}_{t-1}\right]=\sigma_t E\left[(e_{x,t}\eta_t+1)\,\eta_t\mid\mathcal{F}^{\eta}_{t-1}\right]=0$ which follows because $E\left[e_{x,t}\eta_t^2\right]=E\left[\eta_t\right]=0$. This completes the proof of Step 1.

Step 2. Our next task is to derive the limit of $\mathrm{Cov}\left[T^{1/2}L_{\theta,T}\left(\theta_0\right)\right]$ as $T\to\infty$. To this end, first note that $l_{\theta,t}\left(\theta_0\right)$ forms a stationary and ergodic process (this holds because it can be expressed in terms of convergent power series expansions of stationary and ergodic processes). Moreover, as was shown in Step 1, $E\left[l_{\theta,t}\left(\theta_0\right)\right]=0$. Thus, as $L_{\theta,T}\left(\theta_0\right)=T^{-1}\sum_{t=1}^{T}l_{\theta,t}\left(\theta_0\right)$,

$$\mathrm{Cov}\left[T^{1/2}L_{\theta,T}\left(\theta_0\right)\right]=E\left[l_{\theta,t}\left(\theta_0\right)l'_{\theta,t}\left(\theta_0\right)\right]+\sum_{i=1}^{T-1}\frac{T-i}{T}\left\{E\left[l_{\theta,t}\left(\theta_0\right)l'_{\theta,t-i}\left(\theta_0\right)\right]+E\left[l_{\theta,t}\left(\theta_0\right)l'_{\theta,t+i}\left(\theta_0\right)\right]\right\}$$

so that, given that the expectations and limits exist,

$$\mathrm{Cov}\left[T^{1/2}L_{\theta,T}\left(\theta_0\right)\right]\to\sum_{s=-\infty}^{\infty}E\left[l_{\theta,t}\left(\theta_0\right)l'_{\theta,s}\left(\theta_0\right)\right]\quad\text{as}\quad T\to\infty.$$

Thus, our aim is to compute $\sum_{s=-\infty}^{\infty}E\big[l_{\theta,t}\left(\theta_0\right)l'_{\theta,s}\left(\theta_0\right)\big]$.

We will show that (i) the off-diagonal block consisting of $E\big[l_{c,t}(\theta_0)l'_{a,s}(\theta_0)\big]$, $E\big[l_{d,t}(\theta_0)l'_{a,s}(\theta_0)\big]$, $E\big[l_{c,t}(\theta_0)l'_{b,s}(\theta_0)\big]$, and $E\big[l_{d,t}(\theta_0)l'_{b,s}(\theta_0)\big]$ is zero for all t and s, (ii) the block in the lower-right-hand corner consisting of $E\big[l_{c,t}(\theta_0)l'_{c,s}(\theta_0)\big]$, $E\big[l_{d,t}(\theta_0)l'_{c,s}(\theta_0)\big]$, and $E\big[l_{d,t}(\theta_0)l'_{d,s}(\theta_0)\big]$ has the form shown in Lemma 2, (iii) the blocks consisting of $E\big[l_{a,t}(\theta_0)l'_{a,s}(\theta_0)\big]$ and $E\big[l_{b,t}(\theta_0)l'_{a,s}(\theta_0)\big]$ yield the terms in $\mathbf{A}_{11}$, $\mathbf{A}_{21}$, and $\mathbf{B}_{21}$, and (iv) the block consisting of $E\big[l_{b,t}(\theta_0)l'_{b,s}(\theta_0)\big]$ yields the terms in $\mathbf{A}_{22}$ and $\mathbf{B}_{22}$.

Step 2(i). Because $(l_{a,t}(\theta_0),l_{c,t}(\theta_0),l_{d,t}(\theta_0);\mathcal{F}^{\eta}_t)$ forms a martingale difference sequence, both $E\left[l_{c,t}(\theta_0)l'_{a,t}(\theta_0)\right]$ and $E\left[l_{d,t}(\theta_0)l'_{a,t}(\theta_0)\right]$ are zero for $t\neq s$. To see that the same holds for $t=s$, write

$$\begin{aligned}E\left[l_{c,t}(\theta_0)l'_{a,t}(\theta_0)\right]&=E\left[-\frac{1}{2}e_{x,t}\left(e_{x,t}\eta_t+1\right)\frac{h_{c,t}(\theta_0)}{\sigma_t^2}\frac{u'_{a,t}(\theta_0)}{\sigma_t}+\frac{1}{4}\left(e_{x,t}\eta_t+1\right)^2\frac{h_{c,t}(\theta_0)}{\sigma_t^2}\frac{h'_{a,t}(\theta_0)}{\sigma_t^2}\right]\\&=\frac{1}{4}E[(e_{x,t}\eta_t+1)^2]E\left[\frac{h_{c,t}(\theta_0)}{\sigma_t^2}\frac{h'_{a,t}(\theta_0)}{\sigma_t^2}\right],\end{aligned}$$

where the latter equality holds because $h_{a,t}(\theta_0)$, $h_{c,t}(\theta_0)$, $u_{a,t}(\theta_0)$, and σ_t are $\mathcal{F}^{\eta}_{t-1}$–measurable and $E[e_{x,t}]=E[e_{x,t}^2\eta_t]=0$ (see Lemma C.1). To see that the latter expectation in the last expression is zero, note that the terms $h_{c,t}(\theta_0)$ and σ_t are even functions of η_τ for all τ, and conclude from the expansion of $h_{a_p,t}(\theta_0)$ in (11c) that each summand therein is an odd function of η_{t-r}. Therefore, it follows

from Lemma A.3 that the expectation $E[\sigma_t^{-4} h_{c,t}(\theta_0) h'_{a,t}(\theta_0)]$ is zero, and hence $E\left[l_{c,t}(\theta_0) l'_{a,t}(\theta_0)\right] = 0$. Similar arguments show that $E\left[l_{d,t}(\theta_0) l'_{a,t}(\theta_0)\right] = 0$; details are given in the Supplementary Appendix.

Now consider the expectations $E\big[l_{c,t}(\theta_0) l'_{b,s}(\theta_0)\big]$ and $E\big[l_{d,t}(\theta_0) l'_{b,s}(\theta_0)\big]$. By direct calculation,

$$\begin{aligned} l_{c,t}(\theta_0) l'_{b,s}(\theta_0) &= -\frac{1}{2} e_{x,s} \left(e_{x,t}\eta_t + 1\right) \frac{h_{c,t}(\theta_0)}{\sigma_t^2} \frac{u'_{b,s}(\theta_0)}{\sigma_s} + \frac{1}{4} \left(e_{x,t}\eta_t + 1\right) \left(e_{x,s}\eta_s + 1\right) \frac{h_{c,t}(\theta_0)}{\sigma_t^2} \frac{h'_{b,s}(\theta_0)}{\sigma_s^2}, \\ l_{d,t}(\theta_0) l'_{b,s}(\theta_0) &= e_{x,s} e_{\lambda,t} \frac{u'_{b,s}(\theta_0)}{\sigma_s} - \frac{1}{2} \left(e_{x,s}\eta_s + 1\right) e_{\lambda,t} \frac{h'_{b,s}(\theta_0)}{\sigma_s^2}, \end{aligned}$$

and we show that each of the four terms appearing on the right hand sides of these equations has expectation zero for all t and s. For the first term of $l_{c,t}(\theta_0) l'_{b,s}(\theta_0)$, use the expansion of $u_{b_q,s}(\theta_0)$ in (11b) and conclude that we need to show that the Q expressions

$$-\frac{1}{2} \frac{h_{c,t}(\theta_0)}{\sigma_t^2 \sigma_s} \sum_{j=0}^{\infty} \psi_{0,j}^{(b)} \sigma_{s+q+j} \eta_{s+q+j} e_{x,s} \left(e_{x,t}\eta_t + 1\right), \quad q = 1, ..., Q,$$

have expectation zero for all t and s. Here we can consider each term in the summation separately and omit constant multipliers. Thus, it suffices to consider terms of the form

$$\frac{h_{c,t}(\theta_0) \sigma_{s+q+j}}{\sigma_t^2 \sigma_s} \eta_{s+q+j} e_{x,s} \left(e_{x,t}\eta_t + 1\right).$$

If $t \neq s+q+j$, the expression is an odd function of η_{s+q+j}, and hence by Lemma A.3 its expectation is zero. If $t = s+q+j$, the variable $\eta_t \left(e_{x,t}\eta_t + 1\right)$ is independent of the other variables and has expectation zero by Lemma C.1. Hence, the first term in the preceding expression of $l_{c,t}(\theta_0) l'_{b,s}(\theta_0)$ has expectation zero. Similar arguments show that also the second term of $l_{c,t}(\theta_0) l'_{b,s}(\theta_0)$ and the two terms of $l_{d,t}(\theta_0) l'_{b,s}(\theta_0)$ have expectation zero; details are given in the Supplementary Appendix.

Step 2(ii). Because $(l_{c,t}(\theta_0), l_{d,t}(\theta_0); \mathcal{F}_t^{\eta})$ forms a martingale difference sequence, $E\big[l_{c,t}(\theta_0) l'_{c,s}(\theta_0)\big]$, $E\big[l_{d,t}(\theta_0) l'_{c,s}(\theta_0)\big]$, and $E\big[l_{d,t}(\theta_0) l'_{d,s}(\theta_0)\big]$ are all zero when $t \neq s$. When $t = s$, simple calculations making use of the $\mathcal{F}_{t-1}^{\eta}$–measurability of $h_{c,t}(\theta_0)$ and σ_t and Lemma C.1 show that these expectations yield the expressions of $\mathbf{A}_{33}$, $\mathbf{A}_{43}$, and $\mathbf{A}_{44}$ in Lemma 2.

Step 2(iii). Now consider the blocks involving $E\big[l_{a,t}(\theta_0) l'_{a,s}(\theta_0)\big]$ and $E\big[l_{b,t}(\theta_0) l'_{a,s}(\theta_0)\big]$. For the former one, note that $(l_{a,t}(\theta_0); \mathcal{F}_t^{\eta})$ forms a martingale difference sequence so that $E\left[l_{a,t}(\theta_0) l'_{a,s}(\theta_0)\right] = 0$ for all $t \neq s$. When $t = s$, simple calculations making use of the $\mathcal{F}_{t-1}^{\eta}$–measurability of $h_{a,t}(\theta_0)$, $u_{a,t}(\theta_0)$, and σ_t and Lemma C.1 show that $E\left[l_{a,t}(\theta_0) l'_{a,t}(\theta_0)\right]$ equals the expression of $\mathbf{A}_{11}$ in Lemma 2. As for $E\left[l_{b,t}(\theta_0) l'_{a,s}(\theta_0)\right]$, arguments very similar to those already used in Step 2(i) can be used to obtain the expressions of $\mathbf{A}_{21}$ and $\mathbf{B}_{21}$ in Lemma 2; details are given in the Supplementary Appendix.

Step 2(iv). Finally, consider the block consisting of $E\big[l_{b,t}(\theta_0) l'_{b,s}(\theta_0)\big]$. To this end, write

$$\begin{aligned} l_{b,t}(\theta_0) l'_{b,s}(\theta_0) &= e_{x,t} e_{x,s} \frac{u_{b,t}(\theta_0)}{\sigma_t} \frac{u'_{b,s}(\theta_0)}{\sigma_s} + \frac{1}{4} \left(e_{x,t}\eta_t + 1\right) \left(e_{x,s}\eta_s + 1\right) \frac{h_{b,t}(\theta_0)}{\sigma_t^2} \frac{h'_{b,s}(\theta_0)}{\sigma_s^2} \\ &\quad - \frac{1}{2} e_{x,t} \left(e_{x,s}\eta_s + 1\right) \frac{u_{b,t}(\theta_0)}{\sigma_t} \frac{h'_{b,s}(\theta_0)}{\sigma_s^2} - \frac{1}{2} e_{x,s} \left(e_{x,t}\eta_t + 1\right) \frac{h_{b,t}(\theta_0)}{\sigma_t^2} \frac{u'_{b,s}(\theta_0)}{\sigma_s}. \end{aligned}$$

We begin with the first two terms of $l_{b,t}(\theta_0)l'_{b,s}(\theta_0)$. For $t = s$, these two terms have non-zero expectations that yield the expression of $\mathbf{A}_{22}$ in Lemma 2. When $t \neq s$, they have zero expectation. To show this, we proceed as in Step 2(i) and conclude from the expansions of $u_{b_q,t}(\theta_0)$ and $h_{b_q,t}(\theta_0)$ in (11b) and (11d) that it suffices to show that the two expressions

$$\frac{\sigma_{t+q+j}\sigma_{s+\tilde{q}+\tilde{j}}}{\sigma_t\sigma_s}\eta_{t+q+j}\eta_{s+\tilde{q}+\tilde{j}}e_{x,t}e_{x,s}$$
$$\frac{\sigma_{t-r}\sigma_{t-r+q+j}\sigma_{s-\tilde{r}}\sigma_{s-\tilde{r}+\tilde{q}+\tilde{j}}}{\sigma_t^2\sigma_s^2}\eta_{t-r}\eta_{t-r+q+j}\eta_{s-\tilde{r}}\eta_{s-\tilde{r}+\tilde{q}+\tilde{j}}\left(e_{x,t}\eta_t+1\right)\left(e_{x,s}\eta_s+1\right)$$

have expectation zero whenever $t \neq s$. In the first expression, as $t \neq s$, the only possibility for a non-odd expression is $t = s + \tilde{q} + \tilde{j}$ and $s = t + q + j$, but these cannot hold at the same time. As for the second expression, assume that $s > t$ (it is clear from the expression that the case $s < t$ can be treated in a completely analogous manner). Then the largest index can be any of the three indices s, $t - r + q + j$, or $s - \tilde{r} + \tilde{q} + \tilde{j}$. If any of these three indices is alone the largest, the expression will have expectation zero (with similar reasoning as before). The same holds if one of the indices $t - r + q + j$ and $s - \tilde{r} + \tilde{q} + \tilde{j}$ is equal to s and the other is smaller than s (in this case the term $\eta_s\left(e_{x,s}\eta_s + 1\right)$ has expectation zero and is independent of the other terms). Thus, we must have $t - r + q + j = s - \tilde{r} + \tilde{q} + \tilde{j} \overset{def}{=} s + a \geq s$ in order to have a (possibly) non-zero expectation. In this case we must also have $t - r = s - \tilde{r}$ in order to avoid an odd expression as a function of η_{t-r}. With these restrictions, the considered expression simplifies to

$$\frac{\sigma_{s-\tilde{r}}^2\sigma_{s+a}^2}{\sigma_t^2\sigma_s^2}\eta_{s-\tilde{r}}^2\left(e_{x,t}\eta_t+1\right)\left(e_{x,s}\eta_s+1\right)\eta_{s+a}^2.$$

Making use of Lemma B.1 and the fact $E\left[\left(e_{x,t}\eta_t+1\right)\eta_{s-\tilde{r}}^2\sigma_{s-\tilde{r}}^2/\sigma_t^2\right] = E\left[e_{x,t}\eta_t+1\right]E\left[\eta_{s-\tilde{r}}^2\sigma_{s-\tilde{r}}^2/\sigma_t^2\right] = 0$ we conclude that this expression has expectation zero.

Now consider the two last terms in $l_{b,t}(\theta_0)l'_{b,s}(\theta_0)$. Note that due to stationarity, the expectation of the former with any $s = t + x$ equals the transpose of the expectation of the latter with $s = t - x$. Therefore, if these expectations are summed over all $s \neq t$, the sum of the expectations of the latter terms will simply be the transpose of the sum of the expectations of the former terms. Moreover, as the derivations below will demonstrate, this sum will be symmetric. Thus, it suffices to consider the third term in $l_{b,t}(\theta_0)l'_{b,s}(\theta_0)$ only, and multiply the result with 2. Making use of (11b) and (11d), the element $(q, \tilde{q})$ of this term can be expressed as

$$-\sum_{r=1}^{R}\sum_{j=0}^{\infty}\sum_{\tilde{j}=0}^{\infty}\alpha_{0,r}\psi_{0,j}^{(b)}\psi_{0,\tilde{j}}^{(b)}\frac{\sigma_{t+q+j}\sigma_{s-r}\sigma_{s-r+\tilde{q}+\tilde{j}}}{\sigma_t\sigma_s^2}\eta_{t+q+j}\eta_{s-r}\eta_{s-r+\tilde{q}+\tilde{j}}e_{x,t}\left(e_{x,s}\eta_s+1\right).$$

In order to see which terms in this summation have non-zero expectation, note that the largest time index in the summands is either $t + q + j$, s, or $s - r + \tilde{q} + \tilde{j}$. If any of these three indices is alone the largest, the term will have zero expectation. If s equals one of the other two indices while the

other is smaller, then the summand contains the variable $\eta_s\left(e_{x,s}\eta_s+1\right)$ that is independent of the other variables involved and has zero expectation. Thus, in order to have nonzero expectation, we must have $t+q+j=s-r+\tilde{q}+\tilde{j}\overset{def}{=}s+a\geq s$. In this case, we must also have $t=s-r$ $(<s)$ to avoid an odd expression as a function of η_{s-r}. Thus, we can assume these restrictions under which the considered expression simplifies to

$$-\sum_{r=1}^{R}\sum_{j=0}^{\infty}\sum_{\tilde{j}=0}^{\infty}\alpha_{0,r}\psi_{0,j}^{(b)}\psi_{0,\tilde{j}}^{(b)}\frac{\sigma_{s+a}^{2}}{\sigma_{s}^{2}}\eta_{s-r}e_{x,s-r}\left(e_{x,s}\eta_s+1\right)\eta_{s+a}^{2}.$$

Making use of Lemma B.1 and the facts $E\left[\left(e_{x,s}\eta_s+1\right)\eta_s^2\right]=-2$ and $E[\eta_{s-r}e_{x,s-r}]=-1$ (see Lemma C.1), the expression has expectation equal to

$$-2\sum_{r=1}^{R}\sum_{j=0}^{\infty}\sum_{\tilde{j}=0}^{\infty}\alpha_{0,r}\psi_{0,j}^{(b)}\psi_{0,\tilde{j}}^{(b)}c_a\mathbf{1}\left(q+j-r=\tilde{q}+\tilde{j}-r=a\geq 0\right).$$

As $c_a=0$ for $a<0$, we may equally well consider the sum without the restriction $a\geq 0$. Also note that this expression is symmetric in q and $\tilde{q}$, and thus (as was noted above) the matrix $\mathbf{B}_{22}$ is obtained by multiplying the above expression with 2, yielding

$$-4\sum_{r=1}^{R}\sum_{j=0}^{\infty}\sum_{\tilde{j}=0}^{\infty}\alpha_{0,r}\psi_{0,j}^{(b)}\psi_{0,\tilde{j}}^{(b)}c_a\mathbf{1}\left(q+j-r=\tilde{q}+\tilde{j}-r=a\right).$$

As the expression is symmetric, it suffices to consider the case $q\geq\tilde{q}$. Solving for $\tilde{j}$ and a as $\tilde{j}=j+q-\tilde{q}$ and $a=q+j-r$ and substituting to the preceding expression yields

$$-4\sum_{r=1}^{R}\sum_{j=0}^{\infty}\alpha_{0,r}\psi_{0,j}^{(b)}\psi_{0,j+q-\tilde{q}}^{(b)}c_{q+j-r}=-4\sum_{j=0}^{\infty}\psi_{0,j-q}^{(b)}\psi_{0,j-\tilde{q}}^{(b)}c_j,$$

where the equality follows from (10) and the convention $\psi_{0,j}^{(b)}=0$, $j<0$. The last expression equals that given for $\left(\mathbf{B}_{22}\right)_{q,\tilde{q}}$ in Lemma 2. Thus, we have established (iv), completing the proof of Step 2.

Step 3. We now show that our assumptions guarantee that the expression derived for $\mathcal{I}\left(\theta_0\right)$ is finite, thereby also guaranteeing the validity of the employed arguments. As the elements of $\mathbf{B}_{21}$ and $\mathbf{B}_{22}$ are defined through convergent series, their finiteness is immediate. Hence it suffices to show that the matrix $\mathbf{A}=E\left[l_{\theta,t}(\theta_0)l_{\theta,t}'(\theta_0)\right]$ is finite. First consider the blocks $\mathbf{A}_{11}$, $\mathbf{A}_{33}$, $\mathbf{A}_{43}$, and $\mathbf{A}_{44}$ and note that $\sigma_t^2\geq\omega_0>0$ (see equation (6) and Assumption 2). Making use of the Cauchy-Schwarz inequality it is therefore easy to see that these blocks are finite if

$$E[e_{x,t}^2\eta_t^2],\quad E\left[e_{x,t}^2\right],\quad E\left[|e_{\lambda,t}e_{x,t}\eta_t|\right],\quad E[e_{\lambda,t}e_{\lambda,t}'],\quad \left\|u_{a,t}(\theta_0)\right\|_2,\quad \left\|h_{a,t}(\theta_0)\right\|_2,\quad \left\|h_{c,t}(\theta_0)\right\|_2$$

are all finite. The four first terms are finite by Lemma C.1. As for the fifth term, making use of the expansion (11a) and Lemmas A.1 and A.2 it is seen that even $\left\|u_{a,t}(\theta_0)\right\|_4$ is finite because $E\left[\varepsilon_t^4\right]<\infty$

by Lemma 1. That $\|h_{a,t}(\theta_0)\|_2 < \infty$ holds can now be seen by using the expansion (11c) and the Cauchy-Schwarz inequality whereas the finiteness of $\|h_{c,t}(\theta_0)\|_2$ follows from the fact that $E\left[\varepsilon_t^4\right] < \infty$.

The finiteness of the blocks $\mathbf{A}_{21}$ and $\mathbf{A}_{22}$ requires a somewhat more detailed investigation. (A direct application of Hölder's inequality would lead to unnecessarily strong conditions.) Consider the first of the four expectations appearing in the expressions of $\mathbf{A}_{21}$ and $\mathbf{A}_{22}$. Using the expansions in (11c) and (11d) it is seen that we need to consider the expectations of ($p = 1, .., P$, $q = 1, ..., Q$)

$$\begin{aligned} & \frac{h_{b_q,t}(\theta_0)}{\sigma_t^2}\frac{h'_{a_p,t}(\theta_0)}{\sigma_t^2}\left(e_{x,t}\eta_t + 1\right)^2 \\ = \; & -4\sum_{r=1}^{R}\sum_{\tilde{r}=1}^{R}\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\alpha_{0,r}\alpha_{0,\tilde{r}}\psi_{0,j}^{(b)}\psi_{0,i}^{(a)}\frac{\sigma_{t-r}\sigma_{t-r+q+j}\sigma_{t-\tilde{r}}\sigma_{t-\tilde{r}-p-i}}{\sigma_t^4}\eta_{t-r}\eta_{t-r+q+j}\eta_{t-\tilde{r}}\eta_{t-\tilde{r}-p-i}\left(e_{x,t}\eta_t + 1\right)^2. \end{aligned}$$

The only terms in this summation that have nonzero expectation are those in which $t-r = t-\tilde{r}-p-i$ and $t-\tilde{r} = t-r+q+j$. Therefore it suffices to consider these index combinations and show that

$$\frac{\sigma_{t-r}^2\sigma_{t-r+q+j}^2}{\sigma_t^4}\eta_{t-r}^2\eta_{t-r+q+j}^2\left(e_{x,t}\eta_t + 1\right)^2$$

has an expectation bounded by a finite constant (independent of the indices). As the indices satisfy $t-r < t-r+q+j < t$, arguments already used in similar previous calculations and the Cauchy-Schwarz inequality give

$$E\left[\frac{\sigma_{t-r}^2\sigma_{t-r+q+j}^2}{\sigma_t^4}\eta_{t-r}^2\eta_{t-r+q+j}^2\left(e_{x,t}\eta_t + 1\right)^2\right] \leq C\left(E[\varepsilon_{t-r}^4]E[\sigma_{t-r+q+j}^4]\right)^{1/2}E[\eta_{t-r+q+j}^2]E\left[\left(e_{x,t}\eta_t + 1\right)^2\right]$$

for some finite C. The expectations on the dominant side are finite by Lemma 1 and Lemma C.1.

The other three expectations appearing in the expressions of $\mathbf{A}_{21}$ and $\mathbf{A}_{22}$ can be handled in a similar manner making use of the expansions (11). Details are given in the Supplementary Appendix.

Step 4. As the matrix $\mathcal{I}(\theta_0)$ is block diagonal, what needs to be shown is that the blocks

$$\mathcal{I}_1(\theta_0) \overset{def}{=} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}'_{21} + \mathbf{B}'_{21} \\ \mathbf{A}_{21} + \mathbf{B}_{21} & \mathbf{A}_{22} + \mathbf{B}_{22} \end{bmatrix} \quad \text{and} \quad \mathcal{I}_2(\theta_0) \overset{def}{=} \begin{bmatrix} \mathbf{A}_{33} & \mathbf{A}'_{43} \\ \mathbf{A}_{43} & \mathbf{A}_{44} \end{bmatrix}$$

are positive definite. We begin with the latter. Note that $\mathcal{I}_2(\theta_0)$ is the covariance matrix of the vector $\left(-\frac{1}{2}(e_{x,t}\eta_t + 1)\frac{h_{c,t}(\theta_0)}{\sigma_t^2}, e_{\lambda,t}\right)$. Therefore, what needs to be proven is that

$$a'\left(-\frac{1}{2}(e_{x,t}\eta_t + 1)\frac{h_{c,t}(\theta_0)}{\sigma_t^2}\right) + b'e_{\lambda,t} = 0 \text{ a.s.} \tag{12}$$

only if $a = 0$ and $b = 0$ ($a \in \mathbb{R}^{R+1}$, $b \in \mathbb{R}^d$). Multiplying (12) with $\sigma_t^2\eta_t^2$ and taking expectations conditional on $\mathcal{F}_{t-1}^{\eta}$ yields (by Lemma C.1(xii), $E\left[\eta_t^2 e_{\lambda,t}\right] = 0$) $a'h_{c,t}(\theta_0) = 0$ or, written out, $a_1 + a_2\sigma_{t-1}^2\eta_{t-1}^2 + \cdots + a_{R+1}\sigma_{t-R}^2\eta_{t-R}^2 = 0$. If $a_2 \neq 0$, we obtain a contradiction, hence $a_2 = 0$. Similarly, $a_3 = \cdots = a_{R+1} = 0$, and thus also $a_1 = 0$. Therefore, (12) becomes $b'e_{\lambda,t} = 0$. By Assumption 4(iii), $E[e_{\lambda,t}e'_{\lambda,t}]$ is positive definite, and hence necessarily $b = 0$. Thus $\mathcal{I}_2(\theta_0)$ is positive definite.

To prove that $\mathcal{I}_1(\theta_0)$ is positive definite, define the processes $x_{a,t} = x_{a,1,t} + x_{a,2,t}$ $(P \times 1)$ and $x_{b,t} = x_{b,1,t} + x_{b,2,t} + x_{b,3,t}$ $(Q \times 1)$, where the vectors on the right-hand sides have the components $(p = 1, \ldots, P,\ q = 1, \ldots, Q)$

$$\begin{aligned}
x_{a_p,1,t} &= -\sum_{i=0}^{\infty} \psi_{0,i}^{(a)} \sigma_{t-p-i} \eta_{t-p-i} \frac{e_{x,t}}{\sigma_t}, \\
x_{a_p,2,t} &= \sum_{r=1}^{R} \sum_{i=0}^{\infty} \alpha_{0,r} \psi_{0,i}^{(a)} \sigma_{t-r} \eta_{t-r} \sigma_{t-r-p-i} \eta_{t-r-p-i} \frac{e_{x,t}\eta_t + 1}{\sigma_t^2}, \\
x_{b_q,1,t} &= \sum_{j=0}^{\infty} \psi_{0,j}^{(b)} \frac{e_{x,t-q-j}}{\sigma_{t-q-j}} \sigma_t \eta_t, \\
x_{b_q,2,t} &= -\sum_{r=1}^{R} \sum_{j=0}^{\infty} \alpha_{0,r} \psi_{0,j}^{(b)} \sigma_{t-r} \eta_{t-r} \sigma_{t-r+q+j} \eta_{t-r+q+j} \frac{e_{x,t}\eta_t + 1}{\sigma_t^2} \mathbf{1}(q + j < r), \\
x_{b_q,3,t} &= -\sum_{r=1}^{R} \sum_{j=0}^{\infty} \alpha_{0,r} \psi_{0,j}^{(b)} \sigma_{t-q-j} \eta_{t-q-j} \frac{e_{x,t+r-q-j}\eta_{t+r-q-j} + 1}{\sigma_{t+r-q-j}^2} \sigma_t \eta_t \mathbf{1}(q + j \geq r).
\end{aligned}$$

It can now be shown that the process $x_t = (x_{a,t}, x_{b,t})$ has the covariance matrix $\mathrm{Cov}[x_t] = \mathcal{I}_1(\theta_0)$. The lengthy arguments required to establish this are similar to those already used in Step 2 of the proof and rely on the expansions (11a)–(11d). The details are available in the Supplementary Appendix.

Making use of the way x_t is constructed we now show that $\mathrm{Cov}[x_t]$ is positive definite. First decompose x_t as $x_t = (x_{a,1,t}, x_{b,1,t} + x_{b,3,t}) + (x_{a,2,t}, x_{b,2,t}) \overset{def}{=} z_{1,t} + z_{2,t}$ where, according to the calculations given in the Supplementary Appendix, $z_{1,t}$ and $z_{2,t}$ are uncorrelated. To show that $\mathrm{Cov}[x_t]$ is positive definite, it thus suffices to show that $\mathrm{Cov}[z_{1,t}]$ is positive definite. To do this, we further decompose $z_{1,t}$ into a sum of two uncorrelated components. To this end, define the processes

$$\begin{aligned}
\zeta_{t,j}^{(a)} &= -\sigma_{t-j}\eta_{t-j}\frac{e_{x,t}}{\sigma_t}, \quad j \geq 1 \\
\zeta_{t,j}^{(b1)} &= \left(\frac{e_{x,t-j}}{\sigma_{t-j}} - \sum_{r=1}^{R} \alpha_{0,r} \frac{(e_{x,t+r-j}\eta_{t+r-j} + 1)}{\sigma_{t+r-j}^2} \sigma_{t-j}\eta_{t-j} \mathbf{1}(j > r) \right) \sigma_t \eta_t, \quad j \geq 1 \\
\zeta_{t,j}^{(b2)} &= \begin{cases} -\alpha_{0,j}\sigma_{t-j}\eta_{t-j}\dfrac{(e_{x,t}\eta_t + 1)\eta_t}{\sigma_t}, & j = 1, \ldots, R \\ 0, & j > R. \end{cases}
\end{aligned}$$

Note that these three series are serially uncorrelated ($E[\zeta_{t,j}^{(a)}\zeta_{t,\tilde{j}}^{(a)}] = E[\zeta_{t,j}^{(b1)}\zeta_{t,\tilde{j}}^{(b1)}] = E[\zeta_{t,j}^{(b2)}\zeta_{t,\tilde{j}}^{(b2)}] = 0$ for $j \neq \tilde{j}$). Moreover, non-contemporaneous elements of the different series also have zero correlation ($E[\zeta_{t,j}^{(a)}\zeta_{t,\tilde{j}}^{(b1)}] = E[\zeta_{t,j}^{(b1)}\zeta_{t,\tilde{j}}^{(b2)}] = E[\zeta_{t,j}^{(b2)}\zeta_{t,\tilde{j}}^{(a)}] = 0$ for $j \neq \tilde{j}$). Now write $z_{1,t} = (z_{a,1,t}, z_{b,1,t}) = (x_{a,1,t}, x_{b,1,t} + x_{b,3,t})$. The components of $z_{1,t}$ can be expressed as

$$\begin{aligned}
z_{a_p,1,t} &= x_{a_p,1,t} = \sum_{i=0}^{\infty} \psi_{0,i}^{(a)} \zeta_{t,i+p}^{(a)} \\
z_{b_q,1,t} &= x_{b_q,1,t} + x_{b_q,3,t} = \sum_{j=0}^{\infty} \psi_{0,j}^{(b)} \left(\zeta_{t,j+q}^{(b1)} + \zeta_{t,j+q}^{(b2)} \right).
\end{aligned}$$

To decompose $z_{1,t}$ into two uncorrelated vectors, define $K = \max\{P, Q\}$, and write

$$\begin{aligned} z_{a_p,1,t} &= \sum_{i=0}^{K-p} \psi_{0,i}^{(a)} \zeta_{t,i+p}^{(a)} + \sum_{i=K-p+1}^{\infty} \psi_{0,i}^{(a)} \zeta_{t,i+p}^{(a)} = z_{a_p,1,t}^{(1)} + z_{a_p,1,t}^{(2)} \\ z_{b_q,1,t} &= \sum_{j=0}^{K-q} \psi_{0,j}^{(b)} \left(\zeta_{t,j+q}^{(b1)} + \zeta_{t,j+q}^{(b2)} \right) + \sum_{j=K-q+1}^{\infty} \psi_{0,j}^{(b)} \left(\zeta_{t,j+q}^{(b1)} + \zeta_{t,j+q}^{(b2)} \right) = z_{b_q,1,t}^{(1)} + z_{b_q,1,t}^{(2)}, \end{aligned}$$

where the latter equalities define the vectors $z_{1,t}^{(1)} = (z_{a,1,t}^{(1)}, z_{b,1,t}^{(1)})$ and $z_{1,t}^{(2)} = (z_{a,1,t}^{(2)}, z_{b,1,t}^{(2)})$ that satisfy $z_{1,t} = z_{1,t}^{(1)} + z_{1,t}^{(2)}$. Note that $z_{1,t}^{(1)}$ depends on $\zeta_{t,j}^{(a)}$, $\zeta_{t,j}^{(b1)}$, and $\zeta_{t,j}^{(b2)}$ only for $j = 1, \ldots, K$, whereas $z_{1,t}^{(2)}$ depends on these processes only for $j > K$. Therefore $z_{1,t}^{(1)}$ and $z_{1,t}^{(2)}$ are uncorrelated, and thus it suffices to prove that $\text{Cov}[z_{1,t}^{(1)}]$ is positive definite.

To show this, denote $\zeta = (\zeta_{t,1}^{(a)}, \ldots, \zeta_{t,K}^{(a)}, \zeta_{t,1}^{(b1)} + \zeta_{t,1}^{(b2)}, \ldots, \zeta_{t,K}^{(b1)} + \zeta_{t,K}^{(b2)})$, and note that $z_{1,t}^{(1)}$ can be written as $z_{1,t}^{(1)} = \Psi\zeta$, where the $(P+Q)\times 2K$ constant matrix Ψ has full row rank and whose expression is given in the Supplementary Appendix. To prove that $\text{Cov}[z_{1,t}^{(1)}]$ is positive definite, it thus suffices to prove that $\text{Cov}[\zeta]$ is positive definite. Equivalently, we can show that the covariance matrix of the vector $(\zeta_{t,1}^{(a)}, \zeta_{t,1}^{(b1)} + \zeta_{t,1}^{(b2)}, \ldots, \zeta_{t,K}^{(a)}, \zeta_{t,K}^{(b1)} + \zeta_{t,K}^{(b2)})$ is positive definite. This vector has a block-diagonal covariance matrix with the 2-by-2 diagonal blocks given by the expectations of $(k = 1, \ldots, K)$

$$\begin{bmatrix} \zeta_{t,k}^{(a)2} & \zeta_{t,k}^{(a)} \left(\zeta_{t,k}^{(b1)} + \zeta_{t,k}^{(b2)} \right) \\ \zeta_{t,k}^{(a)} \left(\zeta_{t,k}^{(b1)} + \zeta_{t,k}^{(b2)} \right) & \left(\zeta_{t,k}^{(b1)} + \zeta_{t,k}^{(b2)} \right)^2 \end{bmatrix} \quad (k = 1, \ldots, K).$$

It thus suffices to show that the covariance matrices of the vectors $(\zeta_{t,k}^{(a)}, \zeta_{t,k}^{(b1)} + \zeta_{t,k}^{(b2)})$, $k = 1, \ldots, K$, are positive definite. This requires a careful argument, and for clarity, we consider the two cases $1 \leq k \leq R$ and $k > R$ separately.

In the latter case, $\zeta_{t,k}^{(b2)} = 0$, and what needs to be proven is that the equality $a_1\zeta_{t,k}^{(a)} + a_2\zeta_{t,k}^{(b1)} = 0$ (a.s.) $(k > R)$ implies $a_1 = a_2 = 0$. Slightly reorganizing, this equality can be written as (note that $\mathbf{1}(k > r)$ can be omitted in this case)

$$a_1 \eta_{t-k} e_{x,t} = a_2 \left(\frac{e_{x,t-k}}{\sigma_{t-k}^2} - \sum_{r=1}^{R} \alpha_{0,r} \frac{(e_{x,t+r-k}\eta_{t+r-k} + 1)}{\sigma_{t+r-k}^2} \eta_{t-k} \right) \sigma_t^2 \eta_t. \tag{13}$$

Squaring and taking expectations conditional on $\mathcal{F}_{t-1}^{\eta}$ yields

$$a_1^2 \eta_{t-k}^2 E\left[e_{x,t}^2\right] = a_2^2 \left(\frac{e_{x,t-k}}{\sigma_{t-k}^2} - \sum_{r=1}^{R} \alpha_{0,r} \frac{(e_{x,t+r-k}\eta_{t+r-k} + 1)}{\sigma_{t+r-k}^2} \eta_{t-k} \right)^2 \sigma_t^4.$$

As $\alpha_{0,r} > 0$ for all $r = 1, ..., R$ (see Assumption 2) the difference in parentheses on the right-hand side cannot be equal to zero a.s. (if it were, its expectation conditional on $\mathcal{F}_{t-k}^{\eta}$ would also equal zero a.s., but this expectation equals $e_{x,t-k}/\sigma_{t-k}^2$). Therefore, if $a_1 = 0$, then $a_2 = 0$, and vice versa.

On the other hand, if $a_1 \neq 0 \neq a_2$ (and $k > R$) multiply (13) with $\eta_{t-k}(e_{x,t} + \eta_t)$, divide by $a_1 \neq 0$, and take expectations conditional on $\mathcal{F}_{t-1}^{\eta}$ to obtain $\eta_{t-k}^2 \left(E[e_{x,t}^2] - 1 \right) = 0$. In the non-Gaussian case,

$E[e_{x,t}^2] > 1$ (see Remark 2 in Andrews, Davis, and Breidt (2006)), giving a contradiction. However, in the Gaussian case $e_{x,t} = -\eta_t$ so that $E[e_{x,t}^2] = 1$ and we need to use a different argument.

Multiplying (13) with $\eta_{t-k}\eta_t$, dividing by $a_2\sigma_t^2 \neq 0$, taking expectations conditional on $\mathcal{F}_{t-1}^{\eta}$, substituting $e_{x,t} = -\eta_t$, and reorganizing yields (note that $\alpha_{0,R} > 0$ by Assumption 2)

$$(1-\eta_{t+R-k}^2)\eta_{t-k}^2 = \frac{\sigma_{t+R-k}^2}{\alpha_{0,R}}\left(\frac{a_1}{a_2}\frac{\eta_{t-k}^2}{\sigma_t^2} - \frac{\eta_{t-k}^2}{\sigma_{t-k}^2} + \sum_{r=1}^{R-1}\alpha_{0,r}\frac{(\eta_{t+r-k}^2-1)}{\sigma_{t+r-k}^2}\eta_{t-k}^2\right). \tag{14}$$

Next consider the event

$$|\eta_{t+R-k}| > M_1, \quad 1 \leq |\eta_{t-k}| \leq 2, \quad \sigma_{t+R-k}^2 \leq M_2, \quad |\eta_\tau| \leq 1 \text{ for } \tau = t-k+1,\ldots,t-k+R-1,$$

which has positive probability for any fixed positive M_1 and M_2 (where M_2 is large enough). (To see this, it suffices to note that η_t has an everywhere positive density, and that σ_t^2 is stationary with finite mean.) On this event (for any fixed M_1 and M_2), the right-hand side of (14) is bounded in absolute value by a constant independent of M_1, say C (note that σ_t^{-2} is bounded). On the other, by choosing M_1 large enough, the left-hand side of (14) will only attain values that are smaller than $-C$. Thus, we have a contradiction.

Now consider the case $1 \leq k \leq R$. We need to show that the equality $a_1\zeta_{t,k}^{(a)} + a_2(\zeta_{t,k}^{(b1)} + \zeta_{t,k}^{(b2)}) = 0$ (a.s.) $(1 \leq k \leq R)$ implies $a_1 = a_2 = 0$. If $a_2 = 0$, we clearly have $a_1 = 0$ also. Now suppose $a_1 = 0$, but $a_2 \neq 0$. In this case, $\zeta_{t,k}^{(b1)} + \zeta_{t,k}^{(b2)} = 0$ (a.s.) must hold $(1 \leq k \leq R)$, so that

$$\left(\frac{e_{x,t-k}}{\sigma_{t-k}} - \sum_{r=1}^{k-1}\alpha_{0,r}\frac{(e_{x,t+r-k}\eta_{t+r-k}+1)}{\sigma_{t+r-k}^2}\sigma_{t-k}\eta_{t-k}\right)\sigma_t\eta_t - \alpha_{0,k}\sigma_{t-k}\eta_{t-k}\frac{(e_{x,t}\eta_t+1)\eta_t}{\sigma_t} = 0. \tag{15}$$

Multiplying (15) with $\sigma_t\eta_{t-k}\eta_t$, dividing by $\sigma_{t-k} \neq 0$, taking expectations conditional on $\mathcal{F}_{t-1}^{\eta}$ (recall that $E\left[(e_{x,t}\eta_t+1)\eta_t^2\right] = -2$), and reorganizing yields

$$e_{x,t-k}\eta_{t-k}\frac{\sigma_t^2}{\sigma_{t-k}^2} - \sum_{r=1}^{k-1}\alpha_{0,r}\frac{\sigma_t^2}{\sigma_{t+r-k}^2}(e_{x,t+r-k}\eta_{t+r-k}+1)\eta_{t-k}^2 + 2\alpha_{0,k}\eta_{t-k}^2 = 0. \tag{16}$$

Adding and subtracting $\sigma_t^2/\sigma_{t-k}^2$, taking expectations, and using Lemma B.1 we obtain

$$-2c_k - E\left[\sigma_t^2/\sigma_{t-k}^2\right] + 2\sum_{r=1}^{k-1}\alpha_{0,r}c_{k-r} + 2\alpha_{0,k} = 0.$$

Because $c_0 = 1$ and $c_j = 0$ for $j < 0$, the identity (10) leads to the contradiction $E\left[\sigma_t^2/\sigma_{t-k}^2\right] = 0$ (this holds even in the Gaussian case which requires no special treatment).

Now consider the case $a_1 \neq 0 \neq a_2$ whereupon the equality $a_1\zeta_{t,k}^{(a)} + a_2(\zeta_{t,k}^{(b1)} + \zeta_{t,k}^{(b2)}) = 0$ (a.s.) $(1 \leq k \leq R)$, multiplied by σ_t and divided by $\sigma_{t-k}a_2 \neq 0$, becomes

$$-\frac{a_1}{a_2}\eta_{t-k}e_{x,t} + \left(e_{x,t-k}\frac{\sigma_t^2}{\sigma_{t-k}^2} - \sum_{r=1}^{k-1}\alpha_{0,r}(e_{x,t+r-k}\eta_{t+r-k}+1)\eta_{t-k}\frac{\sigma_t^2}{\sigma_{t+r-k}^2}\right)\eta_t - \alpha_{0,k}\eta_{t-k}(e_{x,t}\eta_t+1)\eta_t = 0.$$

For brevity, denote the difference in parentheses in the middle term by κ_{t-1}. Next, let b_1 and b_2 be constants such that the variables $(e_{x,t}\eta_t+1)\eta_t-b_1\eta_t-b_2(e_{x,t}+\eta_t)$ and $(\eta_t,(e_{x,t}+\eta_t))$ are uncorrelated. The constants b_1 and b_2 are determined by the linear regression of $(e_{x,t}\eta_t+1)\eta_t$ on the (uncorrelated) regressors η_t and $(e_{x,t}+\eta_t)$. In the Gaussian case $e_{x,t}=-\eta_t$ so that $b_1=-2$ and b_2 is undefined. In the non-Gaussian case, $b_1=-2$ and $b_2=(E[e_{x,t}^2\eta_t^2]-3)/(E[e_{x,t}^2]-1)$ $(E[e_{x,t}^2]>1)$. Defining $\kappa_{1,t-1}=(-\frac{a_1}{a_2}-\alpha_{0,k}b_2)\eta_{t-k}$, $\kappa_{2,t-1}=\kappa_{t-1}+\frac{a_1}{a_2}\eta_{t-k}-\alpha_{0,k}b_1\eta_{t-k}$, and $\kappa_{3,t-1}=-\alpha_{0,k}\eta_{t-k}$, the preceding equation can be written as

$$\kappa_{1,t-1}(e_{x,t}+\eta_t)+\kappa_{2,t-1}\eta_t+\kappa_{3,t-1}[(e_{x,t}\eta_t+1)\eta_t-b_1\eta_t-b_2(e_{x,t}+\eta_t)]=0.$$

By construction, the variables η_t, $(e_{x,t}+\eta_t)$, and $(e_{x,t}\eta_t+1)\,\eta_t-b_1\eta_t-b_2(e_{x,t}+\eta_t)$ are uncorrelated and $\kappa_{i,t-1}$ $(i=1,2,3)$ are $\mathcal{F}_{t-1}^{\eta}$–measurable. Squaring and taking expectations conditional on $\mathcal{F}_{t-1}^{\eta}$,

$$\kappa_{1,t-1}^2(E\left[e_{x,t}^2\right]-1)+\kappa_{2,t-1}^2+\kappa_{3,t-1}^2E[((e_{x,t}\eta_t+1)\eta_t-b_1\eta_t-b_2(e_{x,t}+\eta_t))^2]=0. \tag{17}$$

Each one of the three terms on the left hand side must be zero (a.s.). As $\kappa_{3,t-1}=-\alpha_{0,k}\eta_{t-k}\neq 0$ (a.s.), the expectation in the third term is zero and therefore

$$(e_{x,t}\eta_t+1)\eta_t-b_1\eta_t-b_2(e_{x,t}+\eta_t)=0 \quad \text{(a.s.)}. \tag{18}$$

If η_t is Gaussian we have $e_{x,t}=-\eta_t$ and this equality becomes $-\eta_t^3+(1-b_1)\,\eta_t=0$. This is clearly a contradiction so that we can continue by assuming that η_t is non-Gaussian and that (18) holds.

In the non-Gaussian case $E[e_{x,t}^2]-1>0$ and on the left hand side of (17) we must have $\kappa_{1,t-1}^2=(-\frac{a_1}{a_2}-\alpha_{0,k}b_2)^2\eta_{t-k}^2=0$ (a.s.). This implies $\frac{a_1}{a_2}=-\alpha_{0,k}b_2$, and hence $\kappa_{2,t-1}=\kappa_{t-1}-\alpha_{0,k}b_2\eta_{t-k}-\alpha_{0,k}b_1\eta_{t-k}=0$ can be written as

$$e_{x,t-k}\frac{\sigma_t^2}{\sigma_{t-k}^2}-\sum_{r=1}^{k-1}\alpha_{0,r}(e_{x,t+r-k}\eta_{t+r-k}+1)\eta_{t-k}\frac{\sigma_t^2}{\sigma_{t+r-k}^2}-\alpha_{0,k}\,(b_2+b_1)\,\eta_{t-k}=0 \quad \text{(a.s.)}. \tag{19}$$

Multiplying with η_{t-k} and taking expectations yields (cf. the steps following equation (16))

$$-E\left[\sigma_t^2/\sigma_{t-k}^2\right]-2\alpha_{0,k}-\alpha_{0,k}\,(b_2+b_1)=0$$

and as $b_1=-2$, a contradiction is obtained unless $b_2<0$ (in which case $E[e_{x,t}^2\eta_t^2]<3$).

Now consider equation (18). Substituting the definition of $e_{x,t}$ and $b_1=-2$ and rearranging equation (18) can be written as

$$\frac{f_{\eta,x}\,(\eta_t;\lambda_0)}{f_{\eta}\,(\eta_t;\lambda_0)}=\frac{(b_2-3)\eta_t}{\eta_t^2-b_2}=-\eta_t\left(\frac{b_2}{b_2-3}-\frac{1}{b_2-3}\eta_t^2\right)^{-1}.$$

By definition, the density functions satisfying this differential equation (with $b_2<0$) are members of the Pearson type VII distribution family (see Johnson, Kotz, and Balakrishnan (1994, Sec. 12.4.1)).

Given that we also assume that $E[\eta_t] = 0$, $E[\eta_t^2] = 1$, and $E[\eta_t^4] < \infty$, the only distribution not contradicting (18) is the rescaled t–distribution with density

$$f_\eta(x;\lambda_0) = C(\lambda_0)\left(1+\frac{x^2}{\lambda_0-2}\right)^{-(\lambda_0+1)/2}, \quad C(\lambda_0) = (\pi(\lambda_0-2))^{-1/2}\,\Gamma\left(\frac{\lambda_0+1}{2}\right)\Big/\Gamma\left(\frac{\lambda_0}{2}\right),$$

where the parameter $\lambda_0 > 4$ and $\Gamma(\cdot)$ signifies the Gamma function. If we can show that rescaled t–densities violate equation (19) the proof is complete.

To this end, note that for the rescaled t–distribution $e_{x,t} = -\frac{(\lambda_0+1)\eta_t}{\eta_t^2+\lambda_0-2}$ so that $b_2 = 2-\lambda_0$, and hence $b_2 + b_1 = -\lambda_0$. Equation (19) can now be slightly rewritten as

$$e_{x,t-k}\frac{\sigma_t^2}{\sigma_{t-k}^2} - \sum_{r=1}^{k-1}\alpha_{0,r}(e_{x,t+r-k}\eta_{t+r-k}+1)\eta_{t-k}\frac{\sigma_t^2}{\sigma_{t+r-k}^2} + \alpha_{0,k}\lambda_0\eta_{t-k} = 0 \quad \text{(a.s.)}. \tag{20}$$

We now proceed iteratively. First suppose that $k = 1$. Then (20) becomes

$$e_{x,t-1}\sigma_t^2/\sigma_{t-1}^2 + \alpha_{0,1}\lambda_0\eta_{t-1} = 0. \tag{21}$$

Substituting $e_{x,t-1} = -\frac{(\lambda_0+1)\eta_{t-1}}{\eta_{t-1}^2+\lambda_0-2}$, $\sigma_t^2 = \alpha_{0,1}\sigma_{t-1}^2\eta_{t-1}^2 + \upsilon_{t-2}$, where $\upsilon_{t-2} = \omega_0 + \alpha_{0,2}\sigma_{t-2}^2\eta_{t-2}^2 + \cdots + \alpha_{0,R}\sigma_{t-R}^2\eta_{t-R}^2$ is $\mathcal{F}_{t-2}^\eta$–measurable, multiplying by $\sigma_{t-1}^2(\eta_{t-1}^2+\lambda_0-2)$, and reorganizing results in

$$\left[-\alpha_{0,1}\sigma_{t-1}^2\right]\eta_{t-1}^3 + \left[-(\lambda_0+1)\upsilon_{t-2} + \alpha_{0,1}\lambda_0\sigma_{t-1}^2(\lambda_0-2)\right]\eta_{t-1} = 0.$$

The coefficients in square brackets are $\mathcal{F}_{t-2}^\eta$–measurable and independent of η_{t-1} that follows a rescaled t–distribution, and thus the only way to avoid a contradiction is that the coefficients are zero a.s. (if they were not, the polynomial could be solved and η_{t-1} expressed as a $\mathcal{F}_{t-2}^\eta$–measurable function). However, the coefficient of η_{t-1}^3 is $-\alpha_{0,1}\sigma_{t-1}^2$ which is nonzero, a contradiction.

Now suppose that $k > 1$. Reorganizing, (20) can be written as

$$-\alpha_{0,k-1}(e_{x,t-1}\eta_{t-1}+1)\eta_{t-k}\frac{\sigma_t^2}{\sigma_{t-1}^2} + \left(\frac{e_{x,t-k}}{\sigma_{t-k}^2} - \sum_{r=1}^{k-2}\alpha_{0,r}(e_{x,t+r-k}\eta_{t+r-k}+1)\frac{\eta_{t-k}}{\sigma_{t+r-k}^2}\right)\sigma_t^2 + \alpha_{0,k}\lambda_0\eta_{t-k} = 0$$

or, with obvious definitions of the $\mathcal{F}_{t-2}^\eta$–measurable variables $\mu_{1,t-2}$ and $\nu_{1,t-2}$, as

$$-\alpha_{0,k-1}\eta_{t-k}\frac{\sigma_t^2}{\sigma_{t-1}^2}(e_{x,t-1}\eta_{t-1}+1) + \mu_{1,t-2}\sigma_t^2 + \nu_{1,t-2} = 0.$$

Substituting $e_{x,t-1} = -\frac{(\lambda_0+1)\eta_{t-1}}{\eta_{t-1}^2+\lambda_0-2}$ and $\sigma_t^2 = \alpha_{0,1}\sigma_{t-1}^2\eta_{t-1}^2 + \upsilon_{t-2}$, multiplying with $\sigma_{t-1}^2(\eta_{t-1}^2+\lambda_0-2)$, and reorganizing we get

$$\left[\alpha_{0,k-1}\eta_{t-k}\alpha_{0,1}\sigma_{t-1}^2\lambda_0 + \mu_{1,t-2}\alpha_{0,1}\sigma_{t-1}^4\right]\eta_{t-1}^4 + \kappa_{4,t-2}\eta_{t-1}^2 + \kappa_{5,t-2} = 0$$

where the coefficients $\kappa_{4,t-2}$ and $\kappa_{5,t-2}$ are $\mathcal{F}_{t-2}^\eta$–measurable (we omit their exact expressions for brevity). As above, this equation can be seen as a polynomial in η_{t-1} with coefficients that are $\mathcal{F}_{t-2}^\eta$–measurable and independent of η_{t-1} (which follows a rescaled t–distribution). Thus the only way to

avoid a contradiction is that the coefficients are zero. In particular, the coefficient of η_{t-1}^4 needs to be zero so that (dividing with $\alpha_{0,1}\sigma_{t-1}^2 \neq 0$) $\alpha_{0,k-1}\eta_{t-k}\lambda_0 + \mu_{1,t-2}\sigma_{t-1}^2 = 0$. Substituting the definition of $\mu_{1,t-2}$ into this expression and reorganizing yields

$$e_{x,t-k}\frac{\sigma_{t-1}^2}{\sigma_{t-k}^2} - \sum_{r=1}^{k-2}\alpha_{0,r}(e_{x,t+r-k}\eta_{t+r-k}+1)\eta_{t-k}\frac{\sigma_{t-1}^2}{\sigma_{t+r-k}^2} + \alpha_{0,k-1}\eta_{t-k}\lambda_0 = 0. \tag{22}$$

Note that this is very close to equation (20). If $k = 2$, we get $e_{x,t-2}\sigma_{t-1}^2/\sigma_{t-2}^2 + \alpha_{0,1}\eta_{t-2}\lambda_0 = 0$, which is exactly the same as equation (21) obtained above in the case $k = 1$ except for a shift in the time index. Hence, we can derive a contradiction.

For $2 < k \leq R$, the proof proceeds iteratively using arguments analogous to those used in the two previous paragraphs. For such values of k we arrive at a contradiction similarly as above. All the details are available in the Supplementary Appendix. ■

Proof of Lemma 3. Let $l_{\theta,i,t}(\theta_0)$ $i = 1,\ldots,P+Q+R+d+1$, denote the components of the vector $l_{\theta,t}(\theta_0)$. To establish the result, we need to show that the following conditions hold for all $i = 1,\ldots,P+Q+R+d+1$ (cf. Davidson 1994, Definition 16.1): i) $E\left[|l_{\theta,i,t}(\theta_0)|\right] < \infty$, ii) $\|E\left[l_{\theta,i,t}(\theta_0) \mid \mathcal{F}_t^{\eta}\right]\|_2 < \infty$, iii) $E\left[l_{\theta,i,t}(\theta_0) \mid \mathcal{F}_{t-1}^{\eta}\right] = 0$, and iv) $\left\|l_{\theta,i,t}(\theta_0) - E\left[l_{\theta,i,t}(\theta_0) \mid \mathcal{F}_{t+m}^{\eta}\right]\right\|_2 \leq c\rho^{m+1}$ for all $m \geq 0$ with some $c < \infty$ and $\rho < 1$. Conditions i) and ii) hold because $E\big[l_{\theta,t}(\theta_0)l_{\theta,t}'(\theta_0)\big] < \infty$, as shown in Step 3 of the proof of Lemma 2 whereas condition iii) was shown in Step 1 of the same proof .

Condition iv) clearly holds for the components of the sub-vectors $l_{a,t}(\theta_0)$, $l_{c,t}(\theta_0)$, and $l_{d,t}(\theta_0)$ because they are $\mathcal{F}_t^{\eta}$–measurable. Concerning the sub-vector $l_{b,t}(\theta_0) = e_{x,t}\frac{u_{b,t}(\theta_0)}{\sigma_t} - \frac{1}{2}\frac{h_{b,t}(\theta_0)}{\sigma_t^2}(e_{x,t}\eta_t + 1)$, note that $e_{x,t}$, σ_t, and η_t are $\mathcal{F}_{t+m}^{\eta}$–measurable for all $m \geq 0$ so that

$$\begin{aligned} l_{b,t}(\theta_0) - E\left[l_{b,t}(\theta_0) \mid \mathcal{F}_{t+m}^{\eta}\right] &= \frac{e_{x,t}}{\sigma_t}\left\{u_{b,t}(\theta_0) - E\left[u_{b,t}(\theta_0) \mid \mathcal{F}_{t+m}^{\eta}\right]\right\} \\ &\quad -\frac{1}{2}\frac{e_{x,t}\eta_t + 1}{\sigma_t^2}\left\{h_{b,t}(\theta_0) - E\left[h_{b,t}(\theta_0) \mid \mathcal{F}_{t+m}^{\eta}\right]\right\}. \end{aligned}$$

Recalling the expansions of $u_{b_q,t}(\theta_0)$ and $h_{b_q,t}(\theta_0)$ in (11), we obtain ($q = 1,\ldots,Q$)

$$\begin{aligned} E\left[u_{b_q,t}(\theta_0) \mid \mathcal{F}_{t+m}^{\eta}\right] &= \sum_{j=0}^{m-q}\psi_{0,j}^{(b)}\sigma_{t+q+j}\eta_{t+q+j} \\ E\left[h_{b_q,t}(\theta_0) \mid \mathcal{F}_{t+m}^{\eta}\right] &= 2\sum_{r=1}^{R}\sum_{j=0}^{m-q+r}\alpha_{0,r}\psi_{0,j}^{(b)}\sigma_{t-r}\eta_{t-r}\sigma_{t-r+q+j}\eta_{t-r+q+j}. \end{aligned}$$

Therefore, to establish iv), it suffices to show that

$$\sum_{j=\max\{0,m-q+1\}}^{\infty}\left|\psi_{0,j}^{(b)}\right|\left\|\frac{e_{x,t}}{\sigma_t}\sigma_{t+q+j}\eta_{t+q+j}\right\|_2 \leq c\rho^{m+1}$$

and

$$\sum_{r=1}^{R}\sum_{j=\max\{0,m-q+r+1\}}^{\infty}\alpha_{0,r}\left|\psi_{0,j}^{(b)}\right|\left\|\frac{e_{x,t}\eta_t + 1}{\sigma_t^2}\sigma_{t-r}\eta_{t-r}\sigma_{t-r+q+j}\eta_{t-r+q+j}\right\|_2 \leq c\rho^{m+1}$$

for all $q = 1, ..., Q$ and $m \geq 0$ for some $c < \infty$ and $\rho < 1$.

In light of Lemma A.1, all that is required is to show that

$$E\left[e_{x,t}^2 \frac{\sigma_{t+q+j}^2}{\sigma_t^2}\eta_{t+q+j}^2\right] \quad \text{and} \quad E\left[\frac{\sigma_{t-r}^2\eta_{t-r}^2}{\sigma_t^2}\left(e_{x,t}\eta_t+1\right)^2\frac{\sigma_{t-r+q+j}^2}{\sigma_t^2}\eta_{t-r+q+j}^2\right]$$

are dominated by a finite constant (independent of q, j, and r). For the former this is an immediate consequence of Lemma B.1 and Lemma C.1. Concerning the latter, if $-r+q+j \geq 0$ the desired result is obtained from Lemma B.1 and Lemma C.1. If $-r+q+j < 0$, the term $\left(e_{x,t}\eta_t+1\right)^2$ is independent of the other terms involved and has finite expectation. Thus, using also the bound $\sigma_t^2 \geq \omega_0 > 0$ we can find a finite constant C such that

$$\begin{aligned} E\left[\frac{\sigma_{t-r}^2\eta_{t-r}^2}{\sigma_t^2}\left(e_{x,t}\eta_t+1\right)^2\frac{\sigma_{t-r+q+j}^2}{\sigma_t^2}\eta_{t-r+q+j}^2\right] &\leq CE\left[\sigma_{t-r}^2\eta_{t-r}^2\sigma_{t-r+q+j}^2\right]E\left[\eta_{t-r+q+j}^2\right] \\ &\leq CE\left[\varepsilon_{t-r}^4\right]^{1/2}E\left[\sigma_{t-r+q+j}^4\right]^{1/2}E\left[\eta_{t-r+q+j}^2\right]. \end{aligned}$$

Here the latter bound is based on the Cauchy-Schwarz inequality and the expectations therein are finite by Lemma 1. ■

Proof of Lemma 4. As was noted in the beginning of Step 2 in the proof of Lemma 2, $l_{\theta,t}(\theta_0)$ forms a stationary and ergodic process with $E\left[l_{\theta,t}(\theta_0)\right] = 0$. In Step 3 of the same proof it was shown that $E\left[l_{\theta,t}(\theta_0)l_{\theta,t}'(\theta_0)\right] < \infty$, and hence $E\left[\left|l_{\theta,t}(\theta_0)l_{\theta,t}'(\theta_0)\right|\right] < \infty$. By Lemma 3, the sequence $\{a'l_{\theta,t}(\theta_0), \mathcal{F}_t^\eta\}$ is an L_2–mixingale of size -1 for all conformable fixed vectors $a \neq 0$. By Lemma 2, the matrix $\mathcal{I}(\theta_0)$ is positive definite and an application of Lemma A.4 yields the result $T^{-1/2}\sum_{t=1}^T a'l_{\theta,t}(\theta_0) \xrightarrow{d} N\left(0, a'\mathcal{I}(\theta_0)a\right)$. The stated result follows from this by the Cramér-Wold device. ■

Appendix D: Hessian Matrix

Expression for the Hessian matrix. In accordance with the partition of θ as $\theta = (\theta_a, \theta_b, \theta_c, \theta_d)$, we will denote the 16 blocks of the Hessian matrix with $l_{aa,t}(\theta) = \frac{\partial^2 l_t(\theta)}{\partial\theta_a\partial\theta_a'}$, $l_{ba,t}(\theta) = \frac{\partial^2 l_t(\theta)}{\partial\theta_b\partial\theta_a'}$, etc. In what follows, we will also denote $\theta_{abc} = (\theta_a, \theta_b, \theta_c)$.

Let us summarize what form the 16 blocks of the Hessian $l_{\theta\theta,t}(\theta)$ take. To simplify notation, define

$$\begin{aligned} e_{xx,t}(\theta) &= \frac{f_{\eta,xx}(h_t^{-1/2}(\theta)u_t(\theta);\lambda)}{f_\eta(h_t^{-1/2}(\theta)u_t(\theta);\lambda)} - \left(\frac{f_{\eta,x}(h_t^{-1/2}(\theta)u_t(\theta);\lambda)}{f_\eta(h_t^{-1/2}(\theta)u_t(\theta);\lambda)}\right)^2 \\ e_{\lambda x,t}(\theta) &= \frac{f_{\eta,\lambda x}(h_t^{-1/2}(\theta)u_t(\theta);\lambda)}{f_\eta(h_t^{-1/2}(\theta)u_t(\theta);\lambda)} - \frac{f_{\eta,\lambda}(h_t^{-1/2}(\theta)u_t(\theta);\lambda)}{f_\eta(h_t^{-1/2}(\theta)u_t(\theta);\lambda)}\frac{f_{\eta,x}(h_t^{-1/2}(\theta)u_t(\theta);\lambda)}{f_\eta(h_t^{-1/2}(\theta)u_t(\theta);\lambda)} \\ e_{\lambda\lambda,t}(\theta) &= \frac{f_{\eta,\lambda\lambda}(h_t^{-1/2}(\theta)u_t(\theta);\lambda)}{f_\eta(h_t^{-1/2}(\theta)u_t(\theta);\lambda)} - \frac{f_{\eta,\lambda}(h_t^{-1/2}(\theta)u_t(\theta);\lambda)}{f_\eta(h_t^{-1/2}(\theta)u_t(\theta);\lambda)}\frac{f_{\eta,\lambda}'(h_t^{-1/2}(\theta)u_t(\theta);\lambda)}{f_\eta(h_t^{-1/2}(\theta)u_t(\theta);\lambda)} \end{aligned}$$

and also

$$\begin{aligned}
\mathsf{E}_{1,t}(\theta) &= \frac{1}{2}\left(\frac{1}{2}e_{xx,t}(\theta)\frac{u_t^2(\theta)}{h_t(\theta)}+\frac{3}{2}e_{x,t}(\theta)\frac{u_t(\theta)}{h_t^{1/2}(\theta)}+1\right)\\
\mathsf{E}_{2,t}(\theta) &= -\frac{1}{2}\left(e_{xx,t}(\theta)\frac{u_t(\theta)}{h_t^{1/2}(\theta)}+e_{x,t}(\theta)\right)\\
\mathsf{E}_{3,t}(\theta) &= -\frac{1}{2}\left(e_{x,t}(\theta)\frac{u_t(\theta)}{h_t^{1/2}(\theta)}+1\right).
\end{aligned}$$

Now, straightforward differentiation yields the different blocks of $l_{\theta\theta,t}(\theta)$ as

$$\begin{aligned}
l_{aa,t}(\theta) &= e_{xx,t}(\theta)\frac{u_{a,t}(\theta)}{h_t^{1/2}(\theta)}\frac{u'_{a,t}(\theta)}{h_t^{1/2}(\theta)}+\mathsf{E}_{1,t}(\theta)\frac{h_{a,t}(\theta)}{h_t(\theta)}\frac{h'_{a,t}(\theta)}{h_t(\theta)}\\
&\quad+\mathsf{E}_{2,t}(\theta)\left(\frac{u_{a,t}(\theta)}{h_t^{1/2}(\theta)}\frac{h'_{a,t}(\theta)}{h_t(\theta)}+\frac{h_{a,t}(\theta)}{h_t(\theta)}\frac{u'_{a,t}(\theta)}{h_t^{1/2}(\theta)}\right)+e_{x,t}(\theta)\frac{u_{aa,t}(\theta)}{h_t^{1/2}(\theta)}+\mathsf{E}_{3,t}(\theta)\frac{h_{aa,t}(\theta)}{h_t(\theta)},\\
l_{ba,t}(\theta) &= e_{xx,t}(\theta)\frac{u_{b,t}(\theta)}{h_t^{1/2}(\theta)}\frac{u'_{a,t}(\theta)}{h_t^{1/2}(\theta)}+\mathsf{E}_{1,t}(\theta)\frac{h_{b,t}(\theta)}{h_t(\theta)}\frac{h'_{a,t}(\theta)}{h_t(\theta)}\\
&\quad+\mathsf{E}_{2,t}(\theta)\left(\frac{u_{b,t}(\theta)}{h_t^{1/2}(\theta)}\frac{h'_{a,t}(\theta)}{h_t(\theta)}+\frac{h_{b,t}(\theta)}{h_t(\theta)}\frac{u'_{a,t}(\theta)}{h_t^{1/2}(\theta)}\right)+e_{x,t}(\theta)\frac{u_{ba,t}(\theta)}{h_t^{1/2}(\theta)}+\mathsf{E}_{3,t}(\theta)\frac{h_{ba,t}(\theta)}{h_t(\theta)},\\
l_{bb,t}(\theta) &= e_{xx,t}(\theta)\frac{u_{b,t}(\theta)}{h_t^{1/2}(\theta)}\frac{u'_{b,t}(\theta)}{h_t^{1/2}(\theta)}+\mathsf{E}_{1,t}(\theta)\frac{h_{b,t}(\theta)}{h_t(\theta)}\frac{h'_{b,t}(\theta)}{h_t(\theta)}\\
&\quad+\mathsf{E}_{2,t}(\theta)\left(\frac{u_{b,t}(\theta)}{h_t^{1/2}(\theta)}\frac{h'_{b,t}(\theta)}{h_t(\theta)}+\frac{h_{b,t}(\theta)}{h_t(\theta)}\frac{u'_{b,t}(\theta)}{h_t^{1/2}(\theta)}\right)+e_{x,t}(\theta)\frac{u_{bb,t}(\theta)}{h_t^{1/2}(\theta)}+\mathsf{E}_{3,t}(\theta)\frac{h_{bb,t}(\theta)}{h_t(\theta)},\\
l_{ca,t}(\theta) &= \mathsf{E}_{1,t}(\theta)\frac{h_{c,t}(\theta)}{h_t(\theta)}\frac{h'_{a,t}(\theta)}{h_t(\theta)}+\mathsf{E}_{2,t}(\theta)\frac{h_{c,t}(\theta)}{h_t(\theta)}\frac{u'_{a,t}(\theta)}{h_t^{1/2}(\theta)}+\mathsf{E}_{3,t}(\theta)\frac{h_{ca,t}(\theta)}{h_t(\theta)},\\
l_{cb,t}(\theta) &= \mathsf{E}_{1,t}(\theta)\frac{h_{c,t}(\theta)}{h_t(\theta)}\frac{h'_{b,t}(\theta)}{h_t(\theta)}+\mathsf{E}_{2,t}(\theta)\frac{h_{c,t}(\theta)}{h_t(\theta)}\frac{u'_{b,t}(\theta)}{h_t^{1/2}(\theta)}+\mathsf{E}_{3,t}(\theta)\frac{h_{cb,t}(\theta)}{h_t(\theta)},\\
l_{cc,t}(\theta) &= \mathsf{E}_{1,t}(\theta)\frac{h_{c,t}(\theta)}{h_t(\theta)}\frac{h'_{c,t}(\theta)}{h_t(\theta)}+\mathsf{E}_{3,t}(\theta)\frac{h_{cc,t}(\theta)}{h_t(\theta)},\\
l_{da,t}(\theta) &= e_{\lambda x,t}(\theta)\left(\frac{u'_{a,t}(\theta)}{h_t^{1/2}(\theta)}-\frac{1}{2}\frac{u_t(\theta)}{h_t^{1/2}(\theta)}\frac{h'_{a,t}(\theta)}{h_t(\theta)}\right),\ l_{db,t}(\theta)=e_{\lambda x,t}(\theta)\left(\frac{u'_{b,t}(\theta)}{h_t^{1/2}(\theta)}-\frac{1}{2}\frac{u_t(\theta)}{h_t^{1/2}(\theta)}\frac{h'_{b,t}(\theta)}{h_t(\theta)}\right),\\
l_{dc,t}(\theta) &= e_{\lambda x,t}(\theta)\left(-\frac{1}{2}\frac{u_t(\theta)}{h_t^{1/2}(\theta)}\frac{h'_{c,t}(\theta)}{h_t(\theta)}\right), \qquad l_{dd,t}(\theta)=e_{\lambda\lambda,t}(\theta).
\end{aligned}$$

Expressions for the second partial derivatives of $u_t(\theta)$ and $h_t(\theta)$. To complete the derivation of the Hessian, we need expressions for $u_{aa,t}(\theta)$, $u_{ba,t}(\theta)$, $u_{bb,t}(\theta)$, and $h_{\theta_{abc}\theta_{abc},t}(\theta)$. Concerning

$h_{\theta_{abc}\theta_{abc},t}(\theta)$, with straightforward differentiation we obtain

$$\begin{aligned}
h_{aa,t}(\theta) &= 2\sum_{r=1}^{R}\alpha_r\left(u_{a,t-r}(\theta)u'_{a,t-r}(\theta)+u_{t-r}(\theta)u_{aa,t-r}(\theta)\right)\\
h_{ba,t}(\theta) &= 2\sum_{r=1}^{R}\alpha_r\left(u_{b,t-r}(\theta)u'_{a,t-r}(\theta)+u_{t-r}(\theta)u_{ba,t-r}(\theta)\right)\\
h_{bb,t}(\theta) &= 2\sum_{r=1}^{R}\alpha_r\left(u_{b,t-r}(\theta)u'_{b,t-r}(\theta)+u_{t-r}(\theta)u_{bb,t-r}(\theta)\right)\\
h_{ca,t}(\theta) &= \begin{bmatrix}0\\ 2u_{t-1}(\theta)u'_{a,t-1}(\theta)\\ \vdots\\ 2u_{t-R}(\theta)u'_{a,t-R}(\theta)\end{bmatrix}, \quad h_{cb,t}(\theta)=\begin{bmatrix}0\\ 2u_{t-1}(\theta)u'_{b,t-1}(\theta)\\ \vdots\\ 2u_{t-R}(\theta)u'_{b,t-R}(\theta)\end{bmatrix}
\end{aligned}$$

whereas $h_{cc,t}(\theta)$ is a matrix of zeros. What remains is to compute $u_{aa,t}(\theta)$, $u_{ba,t}(\theta)$, and $u_{bb,t}(\theta)$. Because $u_{a_p,t}(\theta) = -b(B^{-1})^{-1}y_{t-p}$ $(p=1,\ldots,P)$, we have $u_{aa,t}(\theta)=0$. For the remaining two terms, recall that from the relation $b(B^{-1})u_t(\theta) = a(B)\,y_t$ we obtain $\frac{\partial b(B^{-1})u_t(\theta)}{\partial b_q}=0$ for $q=1,\ldots,Q$. On the other hand,

$$0=\frac{\partial b(B^{-1})u_t(\theta)}{\partial b_q}=-B^{-q}u_t(\theta)+b(B^{-1})\frac{\partial u_t(\theta)}{\partial b_q}.$$

Taking partial derivatives with respect to a_p $(p=1,\ldots,P)$ or $b_{\tilde{q}}$ $(\tilde{q}=1,\ldots,Q)$ yields

$$\begin{aligned}
0 &= \frac{\partial^2 b(B^{-1})u_t(\theta)}{\partial b_q\partial a_p}=-B^{-q}\frac{\partial u_t(\theta)}{\partial a_p}+b(B^{-1})\frac{\partial^2 u_t(\theta)}{\partial b_q\partial a_p}\\
0 &= \frac{\partial^2 b(B^{-1})u_t(\theta)}{\partial b_q\partial b_{\tilde{q}}}=-B^{-q}\frac{\partial u_t(\theta)}{\partial b_{\tilde{q}}}-B^{-\tilde{q}}\frac{\partial u_t(\theta)}{\partial b_q}+b(B^{-1})\frac{\partial^2 u_t(\theta)}{\partial b_q\partial b_{\tilde{q}}}
\end{aligned}$$

so that

$$\begin{aligned}
\frac{\partial^2 u_t(\theta)}{\partial b_q\partial a_p} &= b(B^{-1})^{-1}\frac{\partial u_{t+q}(\theta)}{\partial a_p}=-b(B^{-1})^{-2}y_{t+q-p}=-b(B^{-1})^{-1}a(B)^{-1}u_{t+q-p}(\theta)\\
\frac{\partial^2 u_t(\theta)}{\partial b_q\partial b_{\tilde{q}}} &= b(B^{-1})^{-1}\left(\frac{\partial u_{t+q}(\theta)}{\partial b_{\tilde{q}}}+\frac{\partial u_{t+\tilde{q}}(\theta)}{\partial b_q}\right)=2b(B^{-1})^{-2}u_{t+q+\tilde{q}}(\theta)
\end{aligned}$$

This completes the calculation of the Hessian.

Proof of Lemma 5. The arguments used in the proof are analogous to those used in the proof of Lemma 2, Step 2. For the sake of brevity, we only present a short outline of the required steps. All the details are given in the Supplementary Appendix. There we first present an explicit expression for the Hessian matrix evaluated at the true parameter value, $l_{\theta\theta,t}(\theta_0)$. Then we show that the four blocks in the lower left-hand corner of this matrix ($l_{ca,t}(\theta_0)$, $l_{cb,t}(\theta_0)$, $l_{da,t}(\theta_0)$, and $l_{db,t}(\theta_0)$) all have expectation zero. Finally, a tedious argument shows that the remaining blocks have expectations that equal -1 times the corresponding term in the covariance matrix of the score. ■

Proof of Lemma 6. From Lemma 1 and the expressions of the components of $l_{\theta\theta,t}(\theta)$ at the beginning of this Appendix it follows that $l_{\theta\theta,t}(\theta)$ forms a stationary ergodic sequence of random variables that are continuous in θ over Θ_0. The desired result thus follows from Theorem 2.7 in Straumann and Mikosch (2006) if we establish that $E\left[\sup_{\theta\in\Theta_0}|l_{\theta\theta,t}(\theta)|\right]$ is finite. In light of the expression of $l_{\theta\theta,t}(\theta)$, definition of Θ_c in Assumption 2 (ensuring $h_t(\theta)\geq\underline{\omega}$), and Hölder's inequality, it suffices to show that

$$\left\|\sup_{\theta\in\Theta_0}|e_{x,t}(\theta)|\right\|_2,\quad \left\|\sup_{\theta\in\Theta_0}|e_{xx,t}(\theta)|\right\|_2,\quad \left\|\sup_{\theta\in\Theta_0}|e_{\lambda x,t}(\theta)|\right\|_2,\quad \left\|\sup_{\theta\in\Theta_0}|e_{\lambda\lambda,t}(\theta)|\right\|_1,$$
$$\left\|\sup_{\theta\in\Theta_0}|\mathsf{E}_{1,t}(\theta)|\right\|_2,\quad \left\|\sup_{\theta\in\Theta_0}|\mathsf{E}_{2,t}(\theta)|\right\|_2,\quad \left\|\sup_{\theta\in\Theta_0}|\mathsf{E}_{3,t}(\theta)|\right\|_2,$$
$$\left\|\sup_{\theta\in\Theta_0}|u_{a,t}(\theta)|\right\|_4,\quad \left\|\sup_{\theta\in\Theta_0}|u_{b,t}(\theta)|\right\|_4,\quad \left\|\sup_{\theta\in\Theta_0}|u_{aa,t}(\theta)|\right\|_2,\quad \left\|\sup_{\theta\in\Theta_0}|u_{ba,t}(\theta)|\right\|_2,\quad \left\|\sup_{\theta\in\Theta_0}|u_{bb,t}(\theta)|\right\|_2,$$
$$\left\|\sup_{\theta\in\Theta_0}\left|\frac{h_{\theta_{abc},t}(\theta)}{h_t(\theta)}\right|\right\|_4,\quad \left\|\sup_{\theta\in\Theta_0}\left|\frac{h_{\theta_{abc}\theta_{abc},t}(\theta)}{h_t(\theta)}\right|\right\|_2$$

are all finite. Recalling the definitions of the terms appearing in the first seven expressions, it is straightforward to see that the first seven norms are finite by Assumptions 1–6.

Now consider the moment conditions required for the derivatives of $u_t(\theta)$. Recall from Appendix C the expressions $u_{a_p,t}(\theta)=-a(B)^{-1}u_{t-p}(\theta)$ and $u_{b_q,t}(\theta)=b(B^{-1})^{-1}u_{t+q}(\theta)$, and from the beginning of this appendix the expressions $u_{aa,t}(\theta)=0$, $u_{b_qa_p,t}(\theta)=-b(B^{-1})^{-1}a(B)^{-1}u_{t+q-p}(\theta)$, and $u_{b_qb_{\tilde{q}},t}(\theta)=2b(B^{-1})^{-2}u_{t+q+\tilde{q}}(\theta)$. In light of these and of Lemmas A.1 and A.2, the required moment conditions are satisfied as long as $\left\|\sup_{\theta\in\Theta_0}|u_t(\theta)|\right\|_4$ is finite. Recalling that $u_t(\theta)=b(B^{-1})^{-1}a(B)\,y_t$, this in turn follows (due to Lemmas A.1 and A.2) because $E[y_t^4]<\infty$ by Lemma 1.

To establish the moment conditions required for the derivatives of $h_t(\theta)$, we first consider the components of $h_{a,t}(\theta)$. Making use of the expression of $h_{a,t}(\theta)$ (see Appendix C), the Cauchy-Schwarz inequality, and the facts that $\omega>0$ and $0<\alpha_r<1$ (see Assumption 2), we obtain ($p=1,\ldots,P$)

$$h_{a_p,t}(\theta)\leq 2\left(\omega+\sum_{r=1}^{R}\alpha_r u_{t-r}^2(\theta)\right)^{1/2}\left(\sum_{r=1}^{R}\alpha_r u_{a_p,t-r}^2(\theta)\right)^{1/2}=2h_t^{1/2}(\theta)\left(\sum_{r=1}^{R}\alpha_r u_{a_p,t-r}^2(\theta)\right)^{1/2}.$$

Therefore, as $h_t(\theta)\geq\omega\geq\underline{\omega}>0$ for all $\theta\in\Theta_0$,

$$\sup_{\theta\in\Theta_0}\left|\frac{h_{a_p,t}(\theta)}{h_t(\theta)}\right|\leq C\sup_{\theta\in\Theta_0}\left(\sum_{r=1}^{R}\alpha_r u_{a_p,t-r}^2(\theta)\right)^{1/2}$$

for some finite C. Thus, it follows that $\left\|\sup_{\theta\in\Theta_0}\left|\frac{h_{a_p,t}(\theta)}{h_t(\theta)}\right|\right\|_4$ is finite if $\left\|\sup_{\theta\in\Theta_0}\left|u_{a_p,t}(\theta)\right|\right\|_4$ is finite – but this has already been shown. With analogous reasoning, $\left\|\sup_{\theta\in\Theta_0}\left|\frac{h_{b_q,t}(\theta)}{h_t(\theta)}\right|\right\|_4$ is finite for $q=1,\ldots,Q$ (for the expressions of $h_{b_q,t}(\theta)$ and the components of $h_{c,t}(\theta)$, see Appendix C). Concerning the vector $h_{c,t}(\theta)$, the required moment condition is clearly satisfied for the first component. For the remaining components, notice that $\alpha_r u_{t-r}^2(\theta)\leq h_t(\theta)$ ($r=1,\ldots,R$), so that the components of $h_{c,t}(\theta)$

satisfy $\frac{h_{c_r,t}(\theta)}{h_t(\theta)} \leq \frac{1}{\alpha_r}$. The definition of the set Θ_0 implies that α_r is bounded away from zero on Θ_0, and thus the required moment condition holds.

Finally, we show the moment conditions required for the second partial derivatives of $h_t(\theta)$, and start with $h_{aa,t}(\theta)$ (for the expressions of these derivatives, see the beginning of this appendix). From the expression of $h_{aa,t}(\theta)$ and the already shown fact that $\left\|\sup_{\theta\in\Theta_0}|u_{a,t}(\theta)|\right\|_4$ is finite, it follows that it suffices to consider the sum $\sum_{r=1}^R \alpha_r u_{t-r}(\theta)u_{aa,t-r}(\theta)$. Using the Cauchy-Schwarz inequality and the fact that $h_t(\theta) \geq \omega \geq \underline{\omega} > 0$ for all $\theta \in \Theta_0$ it is seen that each element of this matrix satisfies ($p = 1, \ldots, P$, $\tilde{p} = 1, \ldots, P$)

$$\sum_{r=1}^R \alpha_r u_{t-r}(\theta)u_{a_p a_{\tilde{p}},t-r}(\theta) \leq \left(\omega + \sum_{r=1}^R \alpha_r u_{t-r}^2(\theta)\right)^{1/2} \left(\sum_{r=1}^R \alpha_r u_{a_p a_{\tilde{p}},t-r}^2(\theta)\right)^{1/2},$$

and hence

$$\sup_{\theta\in\Theta_0}\left|\frac{1}{h_t(\theta)}\sum_{r=1}^R \alpha_r u_{t-r}(\theta)u_{a_p a_{\tilde{p}},t-r}(\theta)\right| \leq C \sup_{\theta\in\Theta_0}\left(\sum_{r=1}^R \alpha_r u_{a_p a_{\tilde{p}},t-r}^2(\theta)\right)^{1/2}.$$

As we have already shown that $\left\|\sup_{\theta\in\Theta_0}|u_{aa,t}(\theta)|\right\|_2$ is finite, we can conclude that $\left\|\sup_{\theta\in\Theta_0}\left|\frac{h_{aa,t}(\theta)}{h_t(\theta)}\right|\right\|_2$ is finite. With similar reasoning the corresponding result for $h_{ba,t}(\theta)$ and $h_{bb,t}(\theta)$ is obtained. Finally, the moment results for the terms involving $h_{ca,t}(\theta)$ and $h_{cb,t}(\theta)$ follow from the results $\left\|\sup_{\theta\in\Theta_0}|u_t(\theta)|\right\|_4 < \infty$, $\left\|\sup_{\theta\in\Theta_0}|u_{a,t}(\theta)|\right\|_4 < \infty$, and $\left\|\sup_{\theta\in\Theta_0}|u_{b,t}(\theta)|\right\|_4 < \infty$ that have already been proven. This completes the proof of the moment conditions. ■

Appendix E: Main Results

Expressions for the feasible log-likelihood $\tilde{l}_t(\theta)$ and the score vector $\tilde{l}_{\theta,t}(\theta)$. Recall that for each fixed T, the quantities $\tilde{u}_t(\theta)$ were defined through the initial and end conditions $\tilde{u}_{T+1}(\theta) = \cdots = \tilde{u}_{T+Q}(\theta) = 0$ and $\tilde{u}_0(\theta) = u_0, \ldots, \tilde{u}_{1-R}(\theta) = u_{1-R}$, and the backward recursion $a(B)y_t = b(B^{-1})\tilde{u}_t(\theta)$ for $t = T, \ldots, 1$. In other words, for $t = T, \ldots, 1$, the $\tilde{u}_t(\theta)$ can be solved from the equations

$$\begin{aligned}
a(B)y_T &= \tilde{u}_T(\theta) \\
a(B)y_{T-1} &= \tilde{u}_{T-1}(\theta) - b_1\tilde{u}_T(\theta) \\
&\vdots \\
a(B)y_{T-Q} &= \tilde{u}_{T-Q}(\theta) - b_1\tilde{u}_{T-Q+1}(\theta) - \cdots - b_Q\tilde{u}_T(\theta) = b(B^{-1})\tilde{u}_{T-Q}(\theta) \\
&\vdots \\
a(B)y_1 &= b(B^{-1})\tilde{u}_1(\theta)
\end{aligned}$$

from which the $\tilde{u}_t(\theta)$ can be solved recursively and are seen to satisfy the relation

$$\tilde{u}_t(\theta) = \sum_{j=0}^{T-t} \psi_j^{(b)} a(B)y_{t+j}, \qquad t = T, \ldots, 1, \tag{23}$$

cf. Andrews, Davis, and Breidt (2006, p. 1656). In contrast, from the relation $a(B)y_t = b(B^{-1})u_t(\theta)$ one obtains the counterpart

$$u_t(\theta) = \sum_{j=0}^{\infty} \psi_j^{(b)} a(B) y_{t+j}. \tag{24}$$

Once the quantities $\tilde{u}_t(\theta)$ are available, we can form $\tilde{h}_t(\theta)$, $t = 1, ..., T$, by using equation (9) and, furthermore, the feasible log-likelihood $\tilde{l}_t(\theta)$.

It is clear from (23) that $\tilde{u}_t(\theta)$ also depends on the sample size T, although we suppress this dependence from the notation. Consequently, the feasible log-likelihood $\tilde{l}_t(\theta)$ and its components and derivatives also depend on T. As a convention, relations involving variables associated with the feasible log-likelihood ($\tilde{u}_t(\theta)$, $\tilde{h}_t(\theta)$, etc.) are understood to hold only for $t = 1, \ldots, T$ unless otherwise stated.

Now we can derive the score vector. Exactly the same calculations that lead to the infeasible score vector give

$$\tilde{l}_{\theta,t}(\theta) = \begin{bmatrix} \tilde{e}_{x,t}(\theta)\frac{\tilde{u}_{a,t}(\theta)}{\tilde{h}_t^{1/2}(\theta)} - \frac{1}{2}\frac{\tilde{h}_{a,t}(\theta)}{\tilde{h}_t(\theta)}\Big(\tilde{e}_{x,t}(\theta)\frac{\tilde{u}_t(\theta)}{\tilde{h}_t^{1/2}(\theta)} + 1\Big) \\ \tilde{e}_{x,t}(\theta)\frac{\tilde{u}_{b,t}(\theta)}{\tilde{h}_t^{1/2}(\theta)} - \frac{1}{2}\frac{\tilde{h}_{b,t}(\theta)}{\tilde{h}_t(\theta)}\Big(\tilde{e}_{x,t}(\theta)\frac{\tilde{u}_t(\theta)}{\tilde{h}_t^{1/2}(\theta)} + 1\Big) \\ -\frac{1}{2}\frac{\tilde{h}_{c,t}(\theta)}{\tilde{h}_t(\theta)}\Big(\tilde{e}_{x,t}(\theta)\frac{\tilde{u}_t(\theta)}{\tilde{h}_t^{1/2}(\theta)} + 1\Big) \\ \tilde{e}_{\lambda,t}(\theta) \end{bmatrix},$$

where

$$\tilde{e}_{x,t}(\theta) = \frac{f_{\eta,x}(\tilde{h}_t^{-1/2}(\theta)\tilde{u}_t(\theta);\lambda)}{f_\eta(\tilde{h}_t^{-1/2}(\theta)\tilde{u}_t(\theta);\lambda)} \quad \text{and} \quad \tilde{e}_{\lambda,t}(\theta) = \frac{f_{\eta,\lambda}(\tilde{h}_t^{-1/2}(\theta)\tilde{u}_t(\theta);\lambda)}{f_\eta(\tilde{h}_t^{-1/2}(\theta)\tilde{u}_t(\theta);\lambda)},$$

and $\tilde{u}_{a,t}(\theta)$, $\tilde{u}_{b,t}(\theta)$, $\tilde{h}_{a,t}(\theta)$, $\tilde{h}_{b,t}(\theta)$, and $\tilde{h}_{c,t}(\theta)$ are obtained next.

Expressions for the partial derivatives of $\tilde{u}_t(\theta)$ and $\tilde{h}_t(\theta)$. From (9) we immediately find that

$$\begin{aligned} \tilde{h}_{a,t}(\theta) &= 2\alpha_1\tilde{u}_{t-1}(\theta)\tilde{u}_{a,t-1}(\theta) + \cdots + 2\alpha_R\tilde{u}_{t-R}(\theta)\tilde{u}_{a,t-R}(\theta) \\ \tilde{h}_{b,t}(\theta) &= 2\alpha_1\tilde{u}_{t-1}(\theta)\tilde{u}_{b,t-1}(\theta) + \cdots + 2\alpha_R\tilde{u}_{t-R}(\theta)\tilde{u}_{b,t-R}(\theta) \\ \tilde{h}_{c,t}(\theta) &= \big(1, \tilde{u}_{t-1}^2(\theta), \ldots, \tilde{u}_{t-R}^2(\theta)\big). \end{aligned}$$

Now consider the partial derivatives $\tilde{u}_{a,t}(\theta)$ and $\tilde{u}_{b,t}(\theta)$ which, due to the initializations $\tilde{u}_{T+1}(\theta) = \cdots = \tilde{u}_{T+Q}(\theta) = 0$ and $\tilde{u}_0(\theta) = u_0, \ldots, \tilde{u}_{1-R}(\theta) = u_{1-R}$, are zero for $t = 0, \ldots, 1-R$ and $t = T+1, \ldots, T+Q$. For $t = 1, \ldots, T$, from (23) the partial derivatives with respect to a_p ($p = 1, \ldots, P$) are obtained as

$$\tilde{u}_{a_p,t}(\theta) = -\sum_{j=0}^{T-t} \psi_j^{(b)} y_{t-p+j}. \tag{25}$$

To derive the partial derivatives with respect to b_q ($q = 1, \ldots, Q$), first note that the relation $a(B)y_t = b(B^{-1})\tilde{u}_t(\theta)$ implies $\partial b(B^{-1})\tilde{u}_t(\theta)/\partial b_q = 0$. On the other hand, $\partial b(B^{-1})\tilde{u}_t(\theta)/\partial b_q = -B^{-q}\tilde{u}_t(\theta) + b(B^{-1})\partial\tilde{u}_t(\theta)/\partial b_q$, so that one obtains the relation $b(B^{-1})\partial\tilde{u}_t(\theta)/\partial b_q = \tilde{u}_{t+q}(\theta)$. Due to the initialization (not satisfying this relation) the recursive argument used for $\tilde{u}_t(\theta)$ yields

$$\tilde{u}_{b_q,t}(\theta) = \sum_{j=0}^{T-t} \psi_j^{(b)} \tilde{u}_{t+q+j}(\theta). \tag{26}$$

In contrast, the corresponding partial derivatives of $u_t(\theta)$ with respect to a_p and b_q were given by

$$u_{a_p,t}(\theta) = -\sum_{j=0}^{\infty} \psi_j^{(b)} y_{t-p+j} \quad \text{and} \quad u_{b_q,t}(\theta) = \sum_{j=0}^{\infty} \psi_j^{(b)} u_{t+q+j}(\theta). \tag{27}$$

This completes the computation of the derivatives of $\tilde{l}_t(\theta)$.

An auxiliary lemma. The following lemma whose proof is given in the Supplementary Appendix concerns differences between various feasible and infeasible quantities needed later. Denote

$$e_t(\theta) = \log f_\eta\left(\frac{u_t(\theta)}{h_t^{1/2}(\theta)}; \lambda\right) \quad \text{and} \quad \tilde{e}_t(\theta) = \log f_\eta\left(\frac{\tilde{u}_t(\theta)}{\tilde{h}_t^{1/2}(\theta)}; \lambda\right).$$

To express the results in a reasonably compact form, we also define the following sequences of constants. For $t = 1, \ldots, T$, let $U_{1,t} = \rho^{T+1-t}$,

$$U_{2,t} = \begin{cases} 1, & t = 1, \ldots, R \\ \rho^{T+1-t}, & t = R+1, \ldots, T \end{cases} \quad \text{and} \quad U_{3,t} = \begin{cases} 1 + (T+1-t)\,\rho^{T+1-t}, & t = 1, \ldots, R \\ (T+1-t)\,\rho^{T+1-t}, & t = R+1, \ldots, T \end{cases}$$

where $\rho \in (0,1)$ (cf. the discussion after Lemma A.1).

Lemma E.1. *If Assumptions 1–6 and 7(a) hold, then, for $t = 1, \ldots, T$,*

(i)	$\left\|\sup_{\theta\in\Theta_0} \lvert u_t(\theta) - \tilde{u}_t(\theta)\rvert\right\|_4 \le CU_{1,t}$,	*(ii)*	$\left\|\sup_{\theta\in\Theta_0} \lvert u_t^2(\theta) - \tilde{u}_t^2(\theta)\rvert\right\|_2 \le CU_{1,t}$,
(iii)	$\left\|\sup_{\theta\in\Theta_0} \lvert h_t(\theta) - \tilde{h}_t(\theta)\rvert\right\|_2 \le CU_{2,t}$,	*(iv)*	$\left\|\sup_{\theta\in\Theta_0} \lvert \log h_t(\theta) - \log \tilde{h}_t(\theta)\rvert\right\|_2 \le CU_{2,t}$,
(v)	$\left\|\sup_{\theta\in\Theta_0} \left\lvert \frac{u_t(\theta)}{h_t^{1/2}(\theta)} - \frac{\tilde{u}_t(\theta)}{\tilde{h}_t^{1/2}(\theta)}\right\rvert\right\|_{4/3} \le CU_{2,t}$,	*(vi)*	$\left\|\sup_{\theta\in\Theta_0} \left\lvert \frac{u_{a,t}(\theta)}{h_t^{1/2}(\theta)} - \frac{\tilde{u}_{a,t}(\theta)}{\tilde{h}_t^{1/2}(\theta)}\right\rvert\right\|_{4/3} \le CU_{2,t}$,
(vii)	$\left\|\sup_{\theta\in\Theta_0} \left\lvert \frac{h_{a,t}(\theta)}{h_t(\theta)} - \frac{\tilde{h}_{a,t}(\theta)}{\tilde{h}_t(\theta)}\right\rvert\right\|_1 \le CU_{2,t}$,	*(viii)*	$\left\|\sup_{\theta\in\Theta_0} \left\lvert \frac{u_{b,t}(\theta)}{h_t^{1/2}(\theta)} - \frac{\tilde{u}_{b,t}(\theta)}{\tilde{h}_t^{1/2}(\theta)}\right\rvert\right\|_{4/3} \le CU_{3,t}$,
(ix)	$\left\|\sup_{\theta\in\Theta_0} \left\lvert \frac{h_{b,t}(\theta)}{h_t(\theta)} - \frac{\tilde{h}_{b,t}(\theta)}{\tilde{h}_t(\theta)}\right\rvert\right\|_1 \le CU_{3,t}$,	*(x)*	$\left\|\sup_{\theta\in\Theta_0} \left\lvert \frac{h_{c,t}(\theta)}{h_t(\theta)} - \frac{\tilde{h}_{c,t}(\theta)}{\tilde{h}_t(\theta)}\right\rvert\right\|_1 \le CU_{2,t}$,
(xi)	$\left\|\sup_{\theta\in\Theta_0} \lvert e_{x,t}(\theta) - \tilde{e}_{x,t}(\theta)\rvert\right\|_{r_1} \le CU_{2,t}$,	*(xii)*	$\left\|\sup_{\theta\in\Theta_0} \lvert e_{\lambda,t}(\theta) - \tilde{e}_{\lambda,t}(\theta)\rvert\right\|_{r_2} \le CU_{2,t}$,
(xiii)	$\left\|\sup_{\theta\in\Theta_0} \lvert e_t(\theta) - \tilde{e}_t(\theta)\rvert\right\|_{r_3} \le CU_{2,t}$,		

where the constant $C < \infty$ varies from part to part (but is independent of t, T, and θ), and where parts (xi)–(xiii) hold for some $r_1, r_2, r_3 > 0$.

Note that due to the initializations for $t = 0, \ldots, 1-R$ and $t = T+1, \ldots, T+Q$, the differences in Lemma E.1 are non-negligible for t 'close' to 1 and T, and diminish as T increases for 'intermediate' values of t.

Proof of Lemma 7. (i) Note that $|L_T(\theta) - \tilde{L}_T(\theta)| \le T^{-1}\sum_{t=1}^{T} |l_t(\theta) - \tilde{l}_t(\theta)|$, where $|l_t(\theta) - \tilde{l}_t(\theta)| \le |e_t(\theta) - \tilde{e}_t(\theta)| + \frac{1}{2}|\log h_t(\theta) - \log \tilde{h}_t(\theta)|$, because $l_t(\theta) = e_t(\theta) - \frac{1}{2}\log h_t(\theta)$ and similarly for $\tilde{l}_t(\theta)$. By Loève's c_r–inequality (see Davidson (1994), p. 140) and Lemma E.1(iv) and (xiii) we thus obtain

$\left\|\sup_{\theta\in\Theta_0}|l_t(\theta)-\tilde{l}_t(\theta)|\right\|_p \leq CU_{2,t}$ for some finite constant C and a small enough positive exponent p (the exact value of it does not matter). Using this result we can justify (in a moment) that

$$\lim_{T\to\infty}\sum_{t=1}^{T}\sup_{\theta\in\Theta_0}|l_t(\theta)-\tilde{l}_t(\theta)|<\infty \quad \text{a.s.}, \tag{28}$$

which implies $\sup_{\theta\in\Theta_0}|L_T(\theta)-\tilde{L}_T(\theta)|\leq T^{-1}\sum_{t=1}^{T}\sup_{\theta\in\Theta_0}|l_t(\theta)-\tilde{l}_t(\theta)|\to 0$ a.s. as $T\to\infty$, proving the desired result.

To justify (28), denote $l_t^{\bullet}=\sup_{\theta\in\Theta_0}|l_t(\theta)-\tilde{l}_t(\theta)|$ for $t=1,\ldots,T$, and for every fixed (sufficiently large) T define $l_t^{\bullet\bullet}=l_t^{\bullet}$ for $t=1,\ldots,R$ and $l_t^{\bullet\bullet}=l_{T-(t-(R+1))}^{\bullet}$ for $t=R+1,\ldots,T$. Obviously $\sum_{t=1}^{T}l_t^{\bullet\bullet}=\sum_{t=1}^{T}l_t^{\bullet}$, so that proving $\lim_{T\to\infty}\sum_{t=1}^{T}l_t^{\bullet\bullet}<\infty$ a.s. will establish (28). To this end, notice that, for $t=R+1,\ldots,T$,

$$\|l_t^{\bullet\bullet}\|_p=\left\|\sup_{\theta\in\Theta_0}\left|l_{T-(t-(R+1))}(\theta)-\tilde{l}_{T-(t-(R+1))}(\theta)\right|\right\|_p\leq C\rho^{T+1-(T-(t-(R+1)))}=C\rho^{t-R}$$

so that for a suitably defined new C we have $\|l_t^{\bullet\bullet}\|_p\leq C\rho^t$ for all $t=1,\ldots,T$. The result $\lim_{T\to\infty}\sum_{t=1}^{T}l_t^{\bullet\bullet}<\infty$ a.s. now follows from Lemma A.2 of Meitz and Saikkonen (in press), and proof of part (i) is complete.

(ii) Proof of part (ii) is similar and is given in the Supplementary Appendix. ■

Proof of Theorem 1. The proof makes use of standard arguments, and hence we only present an outline of the required steps (additional details are available in the Supplementary Appendix).

Existence of a consistent root. We first show that there exists a sequence of solutions $\hat{\theta}_T$ to the infeasible likelihood equations $L_{\theta,T}(\theta)=0$ that are strongly consistent for θ_0, and then that the same holds for the solutions $\tilde{\theta}_T$ to the feasible likelihood equations $\tilde{L}_{\theta,T}(\theta)=0$. To this end, choose a small fixed $\epsilon>0$ such that the sphere $\Theta_\epsilon=\{\theta:|\theta-\theta_0|=\epsilon\}$ is contained in Θ_0. We will compare the values attained by $L_T(\theta)$ on this sphere with $L_T(\theta_0)$. For an arbitrary point $\theta\in\Theta_\epsilon$, using a second-order Taylor expansion around θ_0 and adding and subtracting terms yields

$$\begin{aligned}L_T(\theta)-L_T(\theta_0) &= (\theta-\theta_0)'L_{\theta,T}(\theta_0)+\frac{1}{2}(\theta-\theta_0)'\left[L_{\theta\theta,T}(\theta_\bullet)-\mathcal{J}(\theta_\bullet)\right](\theta-\theta_0)\\ &\quad+\frac{1}{2}(\theta-\theta_0)'\left[\mathcal{J}(\theta_\bullet)-\mathcal{J}(\theta_0)\right](\theta-\theta_0)+\frac{1}{2}(\theta-\theta_0)'\mathcal{J}(\theta_0)(\theta-\theta_0)\\ &= S_1+S_2+S_3+S_4,\end{aligned}$$

where $\theta_\bullet$ lies on the line segment between θ and θ_0, and the latter equality defines the terms S_i, $i=1,\ldots,4$. We show in the Supplementary Appendix that (a) for any sufficiently small fixed ϵ, $\sup_{\theta\in\Theta_\epsilon}(S_1+S_2)\to 0$ a.s. as $T\to\infty$. The terms S_3 and S_4 do not depend on T, and we show in the Supplementary Appendix that (b) there exists a positive δ such that for each sufficiently small ϵ, $\sup_{\theta\in\Theta_\epsilon}(S_3+S_4)<-\delta\epsilon^2$. Therefore, for each sufficiently small ϵ,

$$\sup_{\theta\in\Theta_\epsilon}L_T(\theta)<L_T(\theta_0) \text{ a.s. as } T\to\infty. \tag{29}$$

As a consequence, for each fixed sufficiently small ϵ, and for all T sufficiently large, $L_T(\theta)$ must have a local maximum, and hence a root of the likelihood equation $L_{\theta,T}(\theta) = 0$, in the interior of Θ_ϵ with probability one. Having established this, the existence of a sequence $\hat{\theta}_T$, independent of ϵ, such that the $\hat{\theta}_T$ are solutions of the likelihood equations $L_{\theta,T}(\theta) = 0$ for all sufficiently large T and that $\hat{\theta}_T \to \theta_0$ a.s. as $T \to \infty$ can be shown as in Serfling (1980, pp. 147–148).

Now consider the feasible likelihood, and first note that

$$\begin{aligned}\sup_{\theta\in\Theta_\epsilon}\left[\tilde{L}_T(\theta) - \tilde{L}_T(\theta_0)\right] &\leq \sup_{\theta\in\Theta_\epsilon}\left|\tilde{L}_T(\theta) - L_T(\theta)\right| + \sup_{\theta\in\Theta_\epsilon}\left[L_T(\theta) - L_T(\theta_0)\right] + \left|L_T(\theta_0) - \tilde{L}_T(\theta_0)\right| \\ &\leq 2\sup_{\theta\in\Theta_0}\left|L_T(\theta) - \tilde{L}_T(\theta)\right| + \sup_{\theta\in\Theta_\epsilon}\left[L_T(\theta) - L_T(\theta_0)\right].\end{aligned}$$

By Lemma 7, the first term on the majorant side converges to zero a.s. as $T \to \infty$, whereas by (29), for each sufficiently small ϵ, $\sup_{\theta\in\Theta_\epsilon} L_T(\theta) < L_T(\theta_0)$ a.s. as $T \to \infty$. Therefore, for each sufficiently small ϵ, $\sup_{\theta\in\Theta_\epsilon} \tilde{L}_T(\theta) < \tilde{L}_T(\theta_0)$ a.s. as $T \to \infty$. The existence of a sequence $\tilde{\theta}_T$ such that the $\tilde{\theta}_T$ are solutions of the feasible likelihood equations $\tilde{L}_{\theta,T}(\theta) = 0$ for all sufficiently large T and $\tilde{\theta}_T \to \theta_0$ a.s. as $T \to \infty$ can be deduced as in the case of the infeasible likelihood.

Asymptotic Normality. Using Lemmas 4–6 in conjunction with standard arguments it can be shown that $T^{1/2}(\hat{\theta}_T - \theta_0) \to N(0, \mathcal{I}(\theta_0)^{-1})$ as $T \to \infty$ (see, e.g., Lemma D.4 in Meitz and Saikkonen (in press)). Moreover, exactly as in the proof of Lemma D.6 in Meitz and Saikkonen (in press) it can be shown that $T^{1/2}(\hat{\theta}_T - \tilde{\theta}_T) \to 0$ a.s. as $T \to \infty$, from which the desired result follows.

Consistent estimation of the limiting covariance matrix. In light of the strong consistency of $\tilde{\theta}_T$, the uniform convergence of $L_{\theta\theta,T}(\theta)$ (Lemma 6), and the fact that $E[l_{\theta\theta,t}(\theta_0)] = -\mathcal{I}(\theta_0)$ with $\mathcal{I}(\theta_0)$ positive definite (Lemmas 2 and 5), it is immediate that $L^{-1}_{\theta\theta,T}(\tilde{\theta}_T) \to \mathcal{I}(\theta_0)^{-1}$ a.s. as $T \to \infty$ (cf. Lemma A.1 of Pötscher and Prucha (1991)). For the same conclusion to hold for $\tilde{L}^{-1}_{\theta\theta,T}(\tilde{\theta}_T)$ we need to show that $\sup_{\theta\in\Theta_0}\left|L_{\theta\theta,T}(\theta) - \tilde{L}_{\theta\theta,T}(\theta)\right| \to 0$ a.s. as $T \to \infty$. That this holds under the additional Assumption 7(b) is shown in the Supplementary Appendix. ■

References

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.

Andrews, B., R. A. Davis, and F. J. Breidt (2007): "Rank-based estimation for all-pass time series models," *Annals of Statistics*, 35, 844–869.

Berkes, I., and L. Horváth (2004): "The efficiency of the estimators of the parameters in GARCH processes," *Annals of Statistics*, 32, 633–655.

Breidt, F. 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, F. J., R. A. Davis, and A. A. Trindade (2001): "Least absolute deviation estimation for all-pass time series models," *Annals of Statistics*, 29, 919–946.

Breidt, F. J., and N.-J. Hsu (2005): "Best mean square prediction for moving averages," *Statistica Sinica*, 15, 427–446.

Brockwell, P. J., and R. A. Davis (2006): *Time Series: Theory and Methods*, 2nd edn. Springer-Verlag, New York.

Chen, M., and H. Z. An (1998): "A note on the stationarity and the existence of moments of the GARCH model," *Statistica Sinica*, 8, 505–510.

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

Francq, C., and J.-M. Zakoïan (2004): "Maximum likelihood estimation of pure GARCH and ARMA–GARCH processes," *Bernoulli*, 10, 605–637.

Hsu, N.-J., and F. J. Breidt (2009): "Exact maximum likelihood estimation for non-Gaussian moving averages," *Statistica Sinica*, 19, 545–560.

Huang, J., and Y. Pawitan (2000): "Quasi-likelihood estimation of non-invertible moving average processes," *Scandinavian Journal of Statistics*, 27, 689–702.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1994): *Continuous Univariate Distributions, Volume 1*, 2nd edn. Wiley, New York.

Lanne, M., J. Luoto, and P. Saikkonen (2010): "Optimal forecasting of noncausal autoregressive time series," HECER Discussion Paper No. 286.

Lanne, M., and P. Saikkonen (2009): "Noncausal vector autoregression," HECER Discussion Paper No. 293.

Lanne, M., and P. Saikkonen (in press): "Noncausal autoregressions for economic time series," *Journal of Time Series Econometrics*, forthcoming.

Lii, K.-S., and M. Rosenblatt (1992): "An approximate maximum likelihood estimation for non-Gaussian non-minimum phase moving average processes," *Journal of Multivariate Analysis*, 43, 272–299.

Lii, K.-S., and M. Rosenblatt (1996): "Maximum likelihood estimation for nonGaussian nonminimum phase ARMA sequences," *Statistica Sinica*, 6, 1–22.

McLeish, D. L. (1975): "A maximal inequality and dependent strong laws," *Annals of Probability*, 3, 829–839.

Meitz, M., and P. Saikkonen (in press): "Parameter estimation in nonlinear AR–GARCH models," *Econometric Theory*, forthcoming.

Pötscher, B. M., and I. R. Prucha (1991): "Basic structure of the asymptotic theory in dynamic nonlinear econometric models, Part II: Asymptotic normality," *Econometric Reviews*, 10, 253–325.

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

Scott, D. J. (1973): "Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach," *Advances in Applied Probability*, 5, 119–137.

Serfling, R. J. (1980): *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Straumann, D., and T. Mikosch (2006): "Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach," *Annals of Statistics*, 34, 2449–2495.

Wu, R., and R. A. Davis (2010): "Least absolute deviation estimation for general autoregressive moving average time-series models," *Journal of Time Series Analysis*, 31, 98–112.