Ling, Shiqing; MacAleer, Michael

Working Paper

Asymptotic theory for a vector ARMA-GARCH model

ISER Discussion Paper, Institute of Social and Economic Research, Osaka University, No. 549

Provided in Cooperation with:
The Institute of Social and Economic Research (ISER), Osaka University

Suggested Citation: Ling, Shiqing; MacAleer, Michael (2001) : Asymptotic theory for a vector ARMA-GARCH model, ISER Discussion Paper, Institute of Social and Economic Research, Osaka University, No. 549, Osaka University, Institute of Social and Economic Research (ISER), Osaka

This Version is available at:
http://hdl.handle.net/10419/92700
ASYMPTOTIC THEORY FOR
A VECTOR ARMA-GARCH MODEL*

Shiqing Ling
and
Michael McAleer

July 2001

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

* To appear in Econometric Theory.
ASYMPTOTIC THEORY FOR A VECTOR ARMA-GARCH MODEL *

Shiqing Ling
Hong Kong University of Science and Technology

Michael McAleer
University of Western Australia

Abstract

This paper investigates the asymptotic theory for a vector ARMA-GARCH model. The conditions for the strict stationarity, ergodicity, and the higher-order moments of the model are established. Consistency of the quasi-maximum likelihood estimator (QMLE) is proved under only the second-order moment condition. This consistency result is new, even for the univariate ARCH and GARCH models. Moreover, the asymptotic normality of the QMLE for the vector ARCH model is obtained under only the second-order moment of the unconditional errors, and the finite fourth-order moment of the conditional errors. Under additional moment conditions, the asymptotic normality of the QMLE is also obtained for the vector ARMA-ARCH and ARMA-GARCH models, as well as a consistent estimator of the asymptotic covariance.

1 INTRODUCTION

The primary feature of the autoregressive conditional heteroskedasticity (ARCH) model, as proposed by Engle (1982), is that the conditional variance of the errors varies over time. Such conditional variances have been strongly supported by a huge body of empirical research, especially in stock returns, interest rates, and foreign exchange markets. Following Engle’s pathbreaking idea, many alternatives have

*The authors wish to thank the Co-Editor, Bruce Hansen, and two referees for very helpful comments and suggestions, and to acknowledge the financial support of the Australian Research Council. This paper was revised while the second author was visiting the Institute of Social and Economic Research at Osaka University.
been proposed to model conditional variances, forming an immense ARCH family; see, for example, the surveys of Bollerslev, Chou and Kroner (1992), Bollerslev, Engle and Nelson (1994), and Li, Ling and McAleer (1999). Of these models, the most popular is undoubtedly the generalised ARCH (GARCH) model of Bollerslev (1986). Some multivariate extensions of these models have been proposed; see, for example, Engle, Granger and Kraft (1984), Bollerslev, Engle and Wooldridge (1988), Engle and Rodrigues (1989), Ling and Deng (1993), Engle and Kroner (1995), Wong and Li (1997), and Li, Ling and Wong (1999), among others. However, apart from Ling and Deng (1993) and Li, Ling and Wong (1998), it seems that no asymptotic theory of the estimators has been established for these multivariate ARCH-type models. In most of these multivariate extensions, the primary purpose has been to investigate the structure of the model, as in Engle and Kroner (1995), and the reporting of empirical findings.

In this paper, we propose a vector ARMA-GARCH model which includes the multivariate GARCH model of Bollerslev (1990) as a special case. The sufficient conditions for the strict stationarity and ergodicity, and a causal representation of the vector ARMA-GARCH model, are obtained as extensions of Ling and Li (1997). Based on Tweedie (1988), a simple sufficient condition for the higher-order moments of the model is also obtained.

The main part of this paper investigates the asymptotic theory of the quasi-maximum likelihood estimator (QMLE) for the vector ARMA-GARCH model. Consistency of the QMLE is proved under only the second-order moment condition. Jeantheau (1998) proved consistency for the constant conditional mean drift model with vector GARCH errors. His result is based on a modified result in Pfanzagl (1969), in which it is assumed that the initial values consisting of the infinite past observations are known. In practice, of course, this is not possible.

In the univariate case, the QMLE based on any fixed initial values has been investigated by Weiss (1986), Pantula (1989), Lee and Hansen (1994), Lumsdaine
Weiss (1986) and Ling and Li (1997) use the conditions of Basawa, Feigin and Heyde (1976), whereby their consistency results rely on the assumption that the fourth-order moments exist. Lee and Hansen (1994) and Lumsdaine (1996) use the conditions of Amemiya (1985, pages 106-111), but their methods are only valid for the simple GARCH (1,1) model and cannot be extended to more general cases. Moreover, the conditional errors, that is, \( \eta_{lt} \) when \( m = 1 \) in equation (2.3) in the next section, are required to have the \( (2 + \kappa) \)th \( (\kappa > 0) \) finite moment by Lee and Hansen (1994), and the 32nd finite moment by Lumsdaine (1996).

The consistency result in this paper is based on a uniform convergence as a modification of a theorem in Amemiya (1985, page 116). Moreover, the consistency of the QMLE for the vector ARMA-GARCH model is obtained only under the second-order moment condition. This result is new, even for the univariate ARCH and GARCH models. For the univariate GARCH (1,1) model, our consistency result also avoids the requirement of the higher-order moment of the conditional errors, as in Lee and Hansen (1994) and Lumsdaine (1996).

This paper also investigates the asymptotic normality of the QMLE. For the ARCH model, asymptotic normality requires only the second-order moment of the unconditional errors, and the finite fourth-order moment of the conditional errors. The corresponding result for univariate ARCH requires the fourth-order moment, as in Weiss (1986) and Pantula (1989). The conditions for asymptotic normality of the GARCH (1,1) model in Lee and Hansen (1994) and Lumsdaine (1996) are quite weak. However, their GARCH(1,1) model explicitly excludes the special case of the ARCH(1) model because they assume that \( B_1 \neq 0 \) (see equation (2.7) in the next section) for purposes of identifiability. Under additional moment conditions, the asymptotic normality of the QMLE for the general vector ARMA-GARCH model is also obtained. Given the uniform convergence result, the proof of asymptotic normality does not need to explore the third-order derivative of the quasi-likelihood.
function. Hence, our method is simpler than those in Weiss (1986), Lee and Hansen (1994), Lumsdaine (1996), and Ling and Li (1997).

It is worth emphasizing that, unlike Lumsdaine (1996) and Ling and Li (1997), Lee and Hansen (1994) do not assume that the conditional errors \( \eta_{0t} \) are i.i.d instead of a series of strictly stationary and ergodic martingale difference. Although it is possible to use this weaker assumption for our model, for simplicity we use the i.i.d. assumption.

The paper is organized as follows. Section 2 defines the vector ARMA-GARCH model and investigates its properties. Section 3 presents the quasi-likelihood function and gives a uniform convergence result. Section 4 establishes the consistency of the QMLE and Section 5 develops its asymptotic normality. Concluding remarks are offered in Section 6. All proofs are given in Appendices A and B.

Throughout this paper, we use the following notation. \(|\cdot|\) denotes the absolute value of a univariate variable or the determinant of a matrix. \(\|\cdot\|\) denotes the Euclidean norm of a matrix or vector. \(A'\) denotes the transpose of the matrix or vector \(A\). \(O(1)\) (or \(o(1)\)) denotes a series of non-stochastic variables that are bounded (or converge to zero). \(O_p(1)\) (or \(o_p(1)\)) denotes a series of random variables that are bounded (or converge to zero) in probability. \(\rightarrow_p\) (or \(\rightarrow_L\)) denotes convergence in probability (or in distribution). \(\rho(A)\) denotes the eigenvalue of the matrix \(A\) with largest absolute value.

### 2 THE MODEL AND ITS PROPERTIES

Bollerslev (1990) presented an \(m\)-dimensional multivariate conditional covariance model, namely,

\[
Y_t = E(Y_t|\mathcal{F}_{t-1}) + \varepsilon_{0t}, \quad Var(\varepsilon_{0t}|\mathcal{F}_{t-1}) = D_{0t} \Gamma_0 D_{0t},
\]

where \(\mathcal{F}_t\) is the past information available up to time \(t\), \(D_{0t} = diag(h^{1/2}_{01t}, \ldots, h^{1/2}_{0mt})\), and
\[
\Gamma_0 = \begin{pmatrix}
1 & \sigma_{012} & \cdots & \sigma_{01m} \\
\sigma_{021} & 1 & \sigma_{023} & \cdots \\
\sigma_{0m,1} & \cdots & \sigma_{0m,m-1} & 1
\end{pmatrix},
\]
in which \(\sigma_{0ij} = \sigma_{0ji}\). The main feature of this model is that the conditional correlation
\[E(\varepsilon_{0it}\varepsilon_{0jt}|\mathcal{F}_{t-1})/\sqrt{E(\varepsilon_{0it}^2|\mathcal{F}_{t-1})E(\varepsilon_{0jt}^2|\mathcal{F}_{t-1})} = \sigma_{0ij}\]
is constant over time, where \(i \neq j\) and \(\varepsilon_{0it}\) is the \(i\)th element of \(\varepsilon_{0t}\). By assuming that
\[
h_{0it} = w_{0i} + \sum_{j=1}^{r} a_{0ij} \varepsilon_{0it-j}^2 + \sum_{j=1}^{s} b_{0ij} h_{0it-j}, \quad i = 1, \cdots, m,
\]
Bollerslev (1990) modeled the exchange rates of the German mark, French franc, and British pound against the U.S. dollar. His results provided evidence that the assumption of constant correlations was adequate. Tse (2000) developed the Lagrange multiplier test for the hypothesis of constant correlation in Bollerslev’s model, and provided evidence that the hypothesis was adequate for spot and futures prices, and foreign exchange rates.

It is possible to provide a straightforward explanation for the hypothesis of constant correlation. Suppose that \(h_{0it}\) captures completely the past information, with \(E h_{0it} = E \varepsilon_{0it}^2\). Then \(\eta_{0it} = \varepsilon_{0it} h_{0it}^{-1/2}\) will be independent of the past information. Thus, for each \(i\), \(\{\eta_{0it}, t = 0, \pm 1, \pm 2, \cdots\}\) will be a sequence of independently and identically distributed (i.i.d.) random variables, with zero mean and variance 1. In general, \(\eta_{0it}\) and \(\eta_{0jt}\) are correlated for \(i \neq j\), and hence it is natural to assume that \(\eta_{0t} = (\eta_{01t}, \cdots, \eta_{0mt})'\) is a sequence of i.i.d. random vectors, with zero mean and covariance \(\Gamma_0\). Thus, we can write
\[
\varepsilon_{0t} = D_{0t} \eta_{0t}.
\]
Obviously, \(\varepsilon_{0t}\) in (2.1) has the same conditional covariance matrix as that in (2.3).

Now, the remaining problem is how to define \(h_{0it}\) so that it can capture completely the past information. It is obvious that \(h_{0it}\) may have as many different forms as in the univariate case. In the multivariate case, \(h_{0it}\) should contain some
past information, not only from $\varepsilon_{it}$ but also from $\varepsilon_{0it}$. Hence, a simple specification such as (2.2) is likely to be inadequate. In particular, if it is desired to explain the relationships of the volatilities across different markets, it would be necessary to accommodate some interdependence of the $\varepsilon_{0it}$ or the $h_{0it}$ in the model. Note that $D_{0t}$ depends only on $(h_{01t}, \cdots, h_{0mt})'$, denoted by $H_{0t}$. It is natural to define $H_{0t}$ in the form of (2.5) below, which has also been used by Jeantheau (1998). Specifying the conditional mean part as the vector ARMA model, we define the vector ARMA-GARCH model as follows:

$$
\Phi_0(L)(Y_t - \mu_0) = \Psi_0(L)\varepsilon_{0t},
$$

$$
\varepsilon_{0t} = D_{0t}\eta_{0t}, \quad H_{0t} = W_0 + \sum_{i=1}^r A_{0i}\varepsilon_{0t-i} + \sum_{i=1}^s B_{0i}H_{0t-i},
$$

where $D_{0t}$ and $\eta_{0t}$ are defined as in (2.3), $\Phi_0(L) = I_m - \Phi_{01}L - \cdots - \Phi_{0p}L^p$ and $\Psi_0(L) = I_m + \Psi_{01}L + \cdots + \Psi_{0q}L^q$ are polynomials in $L$, $I_k$ is the $k \times k$ identity matrix, and $\varepsilon_{0t} = (\varepsilon_{01t}', \cdots, \varepsilon_{0mt}')'$. The true parameter vector is denoted by $\lambda_0 = (\varphi_0', \delta_0', \sigma_0')'$, where $\varphi_0 = vec(\mu_0, \Phi_{01}, \cdots, \Phi_{0p}, \Psi_{01}, \cdots, \Psi_{0q})$, $\delta_0 = vec(W_0, A_{01}, \cdots, A_{0r}, B_{01}, \cdots, B_{0s})$, and $\sigma_0 = (\sigma_{021}, \cdots, \sigma_{0m,1}, \sigma_{032}, \cdots, \sigma_{0m,2}, \cdots, \sigma_{0m,m-1})'$. This model was used to analyze the Hang Seng index and Standard and Poor’s 500 Composite index by Wong, Li and Ling (2000). They found that the off-diagonal elements in $A_{01}$ are significantly different from zero, and hence can be used to explain the volatility relationship between the two markets.

The model for the unknown parameter $\lambda = (\varphi', \delta', \sigma')'$, with $\varphi$, $\delta$, and $\sigma$ defined in a similar manner to $\varphi_0$, $\delta_0$, and $\sigma_0$, respectively, is

$$
\Phi(L)(Y_t - \mu) = \Psi(L)\varepsilon_t,
$$

$$
H_t = W + \sum_{i=1}^r A_i\varepsilon_{t-i} + \sum_{i=1}^s B_iH_{t-i},
$$

where $H_t = (h_{1t}, \cdots, h_{mt})'$, $\varepsilon_t = (\varepsilon_{1t}', \cdots, \varepsilon_{mt}')'$, and $\Phi(L)$ and $\Psi(L)$ are defined in a similar manner to $\Phi_0(L)$ and $\Psi_0(L)$, respectively. First, the $\varepsilon_t$ are computed from the observations $Y_1, \cdots, Y_n$, from (2.6), with initial value $\bar{Y}_0 = (Y_0, \cdots, Y_{1-p}, \varepsilon_0, \cdots, \varepsilon_{1-q})$. 
Then \( H_t \) can be calculated from (2.7), with initial values \( \bar{\varepsilon}_0 = (\tilde{\varepsilon}_0, \cdots, \tilde{\varepsilon}_{1-r}, H_0, \cdots, H_{1-s}) \). We assume that the parameter space \( \Theta \) is a compact subspace of Euclidean space, such that \( \lambda_0 \) is an interior point in \( \Theta \) and, for each \( \lambda \in \Theta \), it is assumed that:

Assumption 1. All the roots of \(|\Phi(L)| = 0 \) and all the roots of \(|\Psi(L)| = 0 \) are outside the unit circle.

Assumption 2. \( \Phi(L) \) and \( \Psi(L) \) are left coprime (i.e., if \( \Phi(L) = U(L)\Phi_1(L) \) and \( \Psi(L) = U(L)\Psi_1(L) \), then \( U(L) \) is unimodular with constant determinant), and satisfy other identifiability conditions given in Dunsmuir and Hannan (1976).

Assumption 3. \( \Gamma \) is a finite and positive definite symmetric matrix, with the elements on the diagonal being 1 and \( \rho(\Gamma) \) having a positive lower bound over \( \Theta \); all the elements of \( A_i \) and \( B_j \) are nonnegative, \( i = 1, \cdots, r \), \( j = 1, \cdots, s \); each element of \( W \) has positive lower and upper bounds over \( \Theta \); and all the roots of \(|I_m - \sum_{i=1}^{r} A_i L^i - \sum_{i=1}^{s} B_i L^i| = 0 \) are outside the unit circle.

Assumption 4. \( I_m - \sum_{r=1}^{r} A_i L^i \) and \( \sum_{i=1}^{s} B_i L^i \) are left coprime; and satisfy other identifiability conditions given in Jeantheau (1998) (see also Dunsmuir and Hannan (1976)).

In Assumptions 2 and 4, we use the identifiability conditions in Dunsmuir and Hannan (1976) and Jeantheau (1998). These conditions may be too strong. Alternatively, we can use other identifiability conditions, such as the final form or echelon form in Lütkepohl (1991, Chapter 7), under which the results in this paper for consistency and asymptotic normality will still hold with some minor modifications.

These identifiability conditions are sufficient for the proofs of (B.3) and (B.6) in Appendix B.

Note that, under Assumption 4, \( B_s \neq 0 \) and hence the ARCH and GARCH models are nonnested. We define the ARMA-ARCH model as follows:

\[
\Phi_0(L)(Y_t - \mu_0) = \Psi_0(L)\varepsilon_{0t}, \tag{2.8}
\]

\[
\varepsilon_{0t} = D_0 \eta_{0t}, \quad H_{0t} = W_0 + \sum_{i=1}^{r} A_{0i} \tilde{\varepsilon}_{0t-i}. \tag{2.9}
\]
Similarly, under Assumption 2, it is not allowed that all the coefficients in the ARMA model are zero, so that the ARMA-ARCH model does not include the following ARCH model as a special case:

\[
Y_t = \mu_0 + \varepsilon_{0t},
\]

\[
\varepsilon_{0t} = D_{0t} \eta_{0t}, \quad H_{0t} = W_0 + \sum_{i=1}^{r} A_{0i} \tilde{\varepsilon}_{0t-i}.
\]

In models (2.8)-(2.9) and (2.10)-(2.11), we assume that all the components of \(A_{0i}, i = 1, \cdots, r\), are positive. In practice, this assumption may be too strong. If the parameter matrices \(A_i\) are assumed to have the nested reduced-rank form, as in Ahn and Reinsel (1988), then the results in this and following sections will still hold with some minor modifications.

The unknown parameter ARCH and ARMA-ARCH models are similarly defined as models (2.6)-(2.7). The true parameter \(\lambda_0 = (\varphi_0', \delta_0', \sigma_0')', \) with \(\delta_0 = vec(W_0, A_{01}, \cdots, A_{0r}), \sigma_0\) being defined as in models (2.4)-(2.5), and \(\varphi_0\) being defined as in models (2.4)-(2.5) for models (2.8)-(2.9), and \(\varphi_0 = \mu_0\) for models (2.10)-(2.11). Similarly, define the unknown parameter \(\lambda\) and the parametric space \(\Theta\), with \(0 < a_{ijk}^l \leq a_{ijk} \leq a_{ijk}^w < \infty\), where \(a_{ijk}\) is the \((j,k)th\) component of \(A_{ij}\), \(a_{ijk}^l\) and \(a_{ijk}^w\) are independent of \(\lambda, i = 1, \cdots, r\), and \(j, k = 1, \cdots, m^1\).

The following theorem gives some basic properties of models (2.4)-(2.5). When \(m = 1\), the result in Theorem 2.1 reduces to that in Ling and Li (1997) and the result in Theorem 2.2 reduces to Theorem 6.2 in Ling (1999). When the ARMA model is replaced by a constant mean drift, the second-order stationarity and ergodicity condition in Theorem 2.1 appears to be the same as Proposition 3.1 in Jeantheau (1998). Our proof is different from that in his paper and provides a useful causal expansion. Also note that, in the following theorems, Assumptions 2 and 4 are not imposed and hence these results hold for models (2.8)-(2.9) and models (2.10)-(2.11). However, for these two special cases, the matrix \(\tilde{A}_{0t}\) below can simply be replaced by its \((1,1)\) block.
THEOREM 2.1. Under Assumptions 1 and 3, models (2.4)-(2.5) possess an \( \mathcal{F}_t \)-measurable second-order stationary solution \( \{Y_t, \varepsilon_{0t}, H_{0t}\} \), which is unique, given the \( \eta_{0t} \), where \( \mathcal{F}_t \) is a \( \sigma \)-field generated by \( \{\eta_{0k} : k \leq t\} \). The solutions \( \{Y_t\} \) and \( \{H_{0t}\} \) have the following causal representations:

\[
Y_t = \sum_{k=0}^{\infty} \Upsilon_{0k} \varepsilon_{0t-k}, \text{ a.s.,} \quad (2.12)
\]

\[
H_{0t} = W_0 + \sum_{j=1}^{\infty} c' \left( \prod_{i=1}^{j} \tilde{A}_{0t-i} \right) \xi_{t-j-1}, \text{ a.s.,} \quad (2.13)
\]

where \( \Phi^{-1}(L)\Psi_0(L) = \sum_{k=0}^{\infty} \Upsilon_{0k} L^k \), \( \xi_t = [(\tilde{\eta}_{0t} W_0)', 0, \ldots, 0, W_0', 0, \ldots, 0]'_{(r+s)m \times 1} \), that is, the subvector consisting of the first \( m \) components is \( \tilde{\eta}_{0t} W_0 \) and the subvector consisting of the \((rm+1)\)th to \((r+1)m\)th components is \( W_0 \); \( \tilde{\eta}_{0t} = \text{diag}(\eta_{01t}, \ldots, \eta_{0mt}) \), \( c' = (0, \ldots, 0, I_m, 0, \ldots, 0)_{m \times (r+s)m} \) with the subvector consisting of the \((rm+1)\)th to \((r+1)m\)th columns being \( I_m \), and

\[
\tilde{A}_{0t} = \begin{pmatrix}
\tilde{\eta}_{0t} A_{01} & \cdots & \tilde{\eta}_{0t} A_{0r} \\
I_{m(r-1)} & O_{m(r-1) \times m} & I_{m(r-1) \times m} & O_{m(r-1) \times ms} & \cdots & \tilde{\eta}_{0t} B_{0s} \\
O_{m(s-1) \times m} & \cdots & A_{0r} & B_{01} & \cdots & B_{0s} \\
\end{pmatrix}.
\]

Hence, \( \{Y_t, \varepsilon_{0t}, H_{0t}\} \) are strictly stationary and ergodic.

THEOREM 2.2. Suppose that the assumptions of Theorem 2.1 hold. If \( \rho[E(\tilde{A}_{0t}^k)] < 1 \), with \( k \) being a strictly positive integer, then the \( 2k \)th moments of \( \{Y_t, \varepsilon_{0t}\} \) are finite, where \( \tilde{A}_{0t} \) is defined as in Theorem 2.1, and \( A^\otimes k \) is the Kronecker product of the \( k \) matrices \( A \).

3 QUASI-MAXIMUM LIKELIHOOD ESTIMATOR

The estimators of the parameters in models (2.4)-(2.5) are obtained by maximizing, conditional on \( (\bar{Y}_0, \bar{\varepsilon}_0) \),

\[
L_n(\lambda) = \frac{1}{n} \sum_{t=1}^{n} l_t(\lambda), \quad l_t(\lambda) = -\frac{1}{2} \ln |D_t \Gamma D_t| - \frac{1}{2} \varepsilon'_t (D_t \Gamma D_t)^{-1} \varepsilon_t, \quad (3.1)
\]
where \( L_n(\lambda) \) takes the form of the Gaussian log-likelihood, and \( D_t = \text{diag}(h_{1t}^{1/2}, \ldots, h_{mt}^{1/2}) \). Since we do not assume that \( \eta_{0t} \) is normal, the estimators from (3.1) are the quasi-maximum likelihood estimators (QMLE). Note that the processes \( \varepsilon_i \) and \( D_i, i \leq 0 \), are unobserved, and hence they are only some chosen constant vectors. Thus, \( L_n(\lambda) \) is the likelihood function which is not conditional on the true \((\bar{Y}_0, \bar{\varepsilon}_0)\) and, in practice, we work with this likelihood function.

For convenience, we introduce the unobserved process \( \{(\varepsilon_t^\epsilon, H_t^\epsilon) : t = 0, \pm 1, \pm 2, \cdots\} \), which satisfies

\begin{equation}
\Phi(L)(Y_t - \mu) = \Psi(L)\varepsilon_t^\epsilon,
\end{equation}

\begin{equation}
H_t^\epsilon = W + \sum_{i=1}^r A_i \varepsilon_{t-i}^\epsilon + \sum_{i=1}^s B_i H_{t-i}^\epsilon,
\end{equation}

where \( \varepsilon_t^\epsilon = (\varepsilon_{1t}^\epsilon, \cdots, \varepsilon_{mt}^\epsilon)' \) and \( H_t^\epsilon = (h_{1t}^\epsilon, \cdots, h_{mt}^\epsilon)' \). Denote \( Y_0 = (Y_0, Y_{-1}, \cdots) \). The unobserved log-likelihood function conditional on \( Y_0 \) is

\begin{equation}
L_n^\epsilon(\lambda) = \frac{1}{n} \sum_{t=1}^n l_t^\epsilon(\lambda), \quad l_t^\epsilon(\lambda) = -\frac{1}{2} \ln |D_t^\epsilon \Gamma D_t^\epsilon| - \frac{1}{2} \varepsilon_t^\epsilon (D_t^\epsilon \Gamma D_t^\epsilon)^{-1} \varepsilon_t^\epsilon,
\end{equation}

where \( D_t^\epsilon = \text{diag}(h_{1t}^\epsilon, \cdots, h_{mt}^\epsilon) \). When \( \lambda = \lambda_0 \), we have \( \varepsilon_t^\epsilon = \varepsilon_{0t} \), \( H_t^\epsilon = H_{0t} \) and \( D_t^\epsilon = D_{0t} \). The primary difference in the likelihoods (3.1) and (3.4) is that (3.1) is conditional on any initial values, while (3.4) is conditional on the infinite past observations. In practice, the use of (3.4) is not possible. Jeantheau (1998) investigated the likelihood (3.4) for models (2.4)-(2.5) with \( p = q = 0 \), that is, with the conditional mean part identified as the constant drift. By modifying a result in Pfanzagl (1969), he proved the consistency of the QMLE for a special case of models (2.4)-(2.5). An improvement on his result requires only the second-order moment condition. However, the method of his proof is valid only for the log-likelihood function (3.4) and it is not clear whether his result also holds for the likelihood (3.1).

The likelihood function \( L_n(\lambda) \) and the unobserved log-likelihood function \( L_n^\epsilon(\lambda) \) for models (2.8)-(2.9) and models (2.10)-(2.11) are similarly defined as in (3.1) and (3.4).
The following uniform convergence theorem is a modification of Theorem 4.2.1 in Amemiya (1985). This theorem, as well as Lemma 4.5 in the next Section, makes it possible to prove the consistency of the QMLE from the likelihood (3.1) under only a second-order moment condition.

**Theorem 3.1.** Let \( g(y, \theta) \) be a measurable function of \( y \) in Euclidean space for each \( \theta \in \Theta \), a compact subset of \( \mathbb{R}^m \) (Euclidean \( m \)-space), and a continuous function of \( \theta \in \Theta \) for each \( y \). Suppose that \( y_t \) is a sequence of strictly stationary and ergodic time series, such that \( E g(y_t, \theta) = 0 \) and \( E \sup_{\theta \in \Theta} |g(y_t, \theta)| < \infty \). Then \( \sup_{\theta \in \Theta} |n^{-1} \sum_{t=1}^{n} g(y_t, \theta)| = o_p(1) \).

4 CONSISTENCY OF THE QMLE

In (3.4), \( D_t^e \) is evaluated by an infinite expansion of (3.3). We need to show that such an expansion is convergent. In general, all the roots of \( |I_m - \sum_{i=1}^{s} A_i L^i - \sum_{i=1}^{s} B_i L^i| = 0 \) lying outside the unit circle does not ensure that all the roots of \( |I_m - \sum_{i=1}^{s} B_i L^i| = 0 \) are outside the unit circle. However, since all the elements of \( A_i \) and \( B_i \) are negative, we have the following lemma.

**Lemma 4.1.** Under Assumption 3, all the roots of \( |I_m - \sum_{i=1}^{s} B_i L^i| = 0 \) are outside the unit circle.

We first present five lemmas. Lemma 4.2 ensures the identification of \( \lambda_0 \). Lemmas 4.3, 4.4 and 4.6 ensure that the likelihood \( L_n(\lambda) \) of the ARMA-GARCH, ARMA-ARCH and ARCH models converges uniformly in the whole parameter space, with its limit attaining a unique maximum at \( \lambda_0 \). Lemma 4.5 is important for the proof of Lemma 4.6 under the second-order moment condition.

**Lemma 4.2.** Suppose that \( Y_t \) is generated by models (2.4)-(2.5) satisfying Assumptions 1-4, or models (2.8)-(2.9) satisfying Assumptions 1-3, or models (2.10)-(2.11) satisfying Assumption 3. Let \( c_\varphi \) and \( c \) be constant vectors, with the same dimensions as \( \varphi \) and \( \delta \), respectively. Then \( c'_\varphi (\partial \varepsilon_t^e / \partial \varphi) = 0 \) a.s. only if \( c_\varphi = 0 \), and \( c' (\partial H^e_t / \partial \delta) = 0 \) a.s. only if \( c = 0 \).
Lemma 4.3. Define $L(\lambda) = E[l_t(\lambda)]$. Under the assumptions of Lemma 4.2, $L(\lambda)$ exists for all $\lambda \in \Theta$ and $\sup_{\lambda \in \Theta} |L_n(\lambda) - L(\lambda)| = o_p(1)$.

Lemma 4.4. Under the assumptions of Lemma 4.2, $L(\lambda)$ achieves a unique maximum at $\lambda_0$.

Lemma 4.5. Let $X_t$ be a strictly stationary and ergodic time series, with $E|X_t| < \infty$, and $\xi_t$ be a sequence of random variables such that

$$\sup_{1 \leq t \leq n} |\xi_t| \leq C \text{ and } n^{-1} \sum_{t=1}^{n} |\xi_t| = o_p(1).$$

Then $n^{-1} \sum_{t=1}^{n} X_t \xi_t = o_p(1)$.

Lemma 4.6. Under the assumptions of Lemma 4.2, $\sup_{\lambda \in \Theta} |L_n(\lambda) - L(\lambda)| = o_p(1)$.

Based on the above lemmas, we now have the following consistency theorem.

Theorem 4.1. Denote $\hat{\lambda}_n$ as the solution to $\max_{\lambda \in \Theta} L_n(\lambda)$. Under the assumptions of Lemma 4.2, $\hat{\lambda}_n \rightarrow_p \lambda_0$.

5 Asymptotic Normality of the QMLE

To prove the asymptotic normality of the QMLE, it is inevitable to explore the second derivative of the likelihood. The method adopted by Weiss (1986), Lee and Hansen (1994), Lumsdaine (1996) and Ling and Li (1997) uses the third derivative of the likelihood. By using Theorem 3.1, our method requires only the second derivative of the likelihood, which simplifies the proof and reduces the requirement for higher-order moments.

For the general models (2.4)-(2.5), the asymptotic normality of the QMLE would require the existence of the sixth moment. However, for models (2.8)-(2.9) or models (2.10)-(2.11), the moment requirements are weaker. Now we can state some basic results.

Lemma 5.1. Suppose that $Y_t$ is generated by models (2.4)-(2.5) satisfying Assumptions 1-4, or models (2.8)-(2.9) satisfying Assumptions 1-3, or models (2.10)-
(2.11) satisfying Assumption 3. Then, it follows that

\[ E \sup_{\lambda \in \Theta} \left\| \frac{\partial \xi_t}{\partial \phi} (D_t \Gamma D_t)^{-1} \frac{\partial \xi_t}{\partial \phi'} \right\| < \infty \quad \text{and} \quad E \left[ \frac{\partial \xi_t'}{\partial \phi} (D_t \Gamma D_t)^{-1} \frac{\partial \xi_t}{\partial \phi'} \right] > 0, \tag{5.1} \]

where a matrix \( A > 0 \) means that \( A \) is positive definite.

**LEMMA 5.2.** Suppose that \( Y_t \) is generated by models (2.4)-(2.5) satisfying Assumptions 1-4 and \( E ||Y_t||^4 < \infty \), or models (2.8)-(2.9) satisfying Assumptions 1-3 and \( E ||Y_t||^4 < \infty \), or models (2.10)-(2.11) satisfying Assumption 3 and \( E ||\eta_{0t}||^4 < \infty \). Then \( \Omega_0 = E[(\partial l_{0t} / \partial \lambda)(\partial l_{0t} / \partial \lambda')] \) is finite. Furthermore, if \( \Omega_0 > 0 \), then

\[ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_{0t}}{\partial \lambda} \rightarrow \mathcal{L} N(0, \Omega_0), \]

where \( \partial l_{0t} / \partial \lambda = \partial l_t / \partial \lambda |_{\lambda_0} \) and \( \partial l_{0t} / \partial \lambda = \partial l_t / \partial \lambda |_{\lambda_0} \).

**LEMMA 5.3.** Suppose that \( Y_t \) is generated by models (2.4)-(2.5) satisfying Assumptions 1-4 and \( E ||Y_t||^6 < \infty \), or models (2.8)-(2.9) satisfying Assumptions 1-3 and \( E ||Y_t||^4 < \infty \), or models (2.10)-(2.11) satisfying Assumption 3. Then,

\[ E \sup_{\lambda \in \Theta} \left\| \frac{\partial H_t'}{\partial \lambda} D_t^{-2} \Delta_t^{2} D_t^{-2} \frac{\partial H_t'}{\partial \lambda'} \right\| < \infty, \tag{5.2} \]

where \( \tilde{\lambda} = (\phi', \delta')', \Delta_t^c = \tilde{\eta}_t \Gamma^{-1} \tilde{\eta}_t + \tilde{\Delta}_t \tilde{\eta}_t, \tilde{\Delta}_t = diag(e_1 \Gamma^{-1} \eta_t, \cdots, e_m \Gamma^{-1} \eta_t), e_i = (0, \cdots, 0, 1, 0, \cdots, 0)' \) of which the \( i \)th element is 1, \( \eta_t = (\eta_{1t}, \cdots, \eta_{mt})' \) and \( \tilde{\eta}_t = diag(\eta_{1t}, \cdots, \eta_{mt}) \) with \( \eta_{it} = \varepsilon_{it} / h_{it}^{1/2}, i = 1, \cdots, m. \)

**LEMMA 5.4.** Under the assumptions of Lemma 5.3,

(a) \( \sup_{\lambda \in \Theta} \left\| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 l_t^c}{\partial \lambda \partial \lambda} - E \left[ \frac{\partial^2 l_t^c}{\partial \lambda \partial \lambda} \right] \right\| = o_p(1), \)

(b) \( \sup_{\lambda \in \Theta} \left\| \frac{1}{n} \sum_{t=1}^{n} \left[ \frac{\partial^2 l_t^c}{\partial \lambda \partial \lambda} - \frac{\partial^2 l_t}{\partial \lambda \partial \lambda} \right] \right\| = o_p(1). \)

By straightforward calculation, we can show that

\[ \Sigma_0 \equiv E \left[ \frac{\partial^2 l_t^c}{\partial \lambda \partial \lambda} \right] |_{\lambda_0} = \left( \frac{\Sigma_{\lambda_0}}{\Sigma_{\lambda_0}'} \frac{\Sigma_{\lambda_0}}{\Sigma_{\lambda_0}'}' \right), \]

where \( \Sigma_{\lambda_0} = E[(\partial \varepsilon_{0t}^c / \partial \lambda) (D_{0t} \Gamma_0 D_{0t})^{-1} (\partial \varepsilon_{0t}^c / \partial \lambda')] + E[(\partial H_{0t}^c / \partial \lambda) D_{0t}^{-2} C D_{0t}^{-2} (\partial H_{0t}^c / \partial \lambda')] / 4, \Sigma_{\lambda_0} = E \left[ (\partial H_{0t}^c / \partial \lambda) D_{0t}^{-2} \right] C_1 P / 2, \partial \varepsilon_{0t} / \partial \lambda' = \varepsilon_{0t} / \partial \lambda' |_{\lambda_0}, \partial H_{0t} / \partial \lambda' = \varepsilon_{0t} / \partial \lambda' |_{\lambda_0}. \)
\( \partial H_t^i / \partial \tilde{\lambda} |_{\lambda_0}, \ P = (I_m \otimes \Gamma_0^{-1})K, \ C_1 = (C_{11}, \ldots, C_{1m}), \ C_{1i} \) is an \( m \times m \) matrix with the \((i,i)\)th component being 1 and the other components zero, \( K = \partial \text{vec}(\Gamma) / \partial \sigma' \) is a constant matrix, and \( C = \Gamma_0^{-1} \otimes \Gamma_0 + I_m, \) where \( A \otimes B = (a_{ij}b_{ij}) \) for two matrices \( A = (a_{ij}) \) and \( B = (b_{ij}). \) In practice, \( \Sigma_0 \) is evaluated by

\[
\hat{\Sigma}_n = - \left( \hat{\Sigma}_\hat{\lambda} \hat{\Sigma}_\lambda' + \frac{1}{2} \hat{P}' \hat{P} \right),
\]

where \( \hat{\Gamma}_n = \Gamma |_{\hat{\lambda}_n}, \)

\[
\hat{\Sigma}_\hat{\lambda} = \frac{1}{n} \sum_{t=1}^{n} \left[ \frac{\partial e_t^i (D_t^i \Gamma D_t^i)^{-1} \partial \lambda'}{\partial \lambda'} \right]_{\hat{\lambda}_n} + \frac{1}{4n} \sum_{t=1}^{n} \left[ \frac{\partial H_t^i}{\partial \lambda} D_t^{-2} \hat{C}_n D_t^{-2} \frac{\partial H_t}{\partial \lambda'} \right]_{\hat{\lambda}_n},
\]

\[
\hat{\Sigma}_\lambda' = - \frac{1}{2n} \sum_{t=1}^{n} \left[ \frac{\partial H_t^i}{\partial \lambda} D_t^{-2} C_1 \hat{P}, \ \hat{P} = (I_m \otimes \hat{\Gamma}_n^{-1})K, \ \hat{C}_n = \hat{\Gamma}_n^{-1} \otimes \hat{\Gamma}_n + I_m. \]

**Lemma 5.5.** Under the assumptions of Lemma 5.3, \( ||\Sigma_0|| < \infty \) and \( \hat{\Sigma}_n = \Sigma_0 + o_P(1) \) for any sequence \( \lambda_n, \) such that \( \lambda_n - \lambda_0 = o_P(1). \) If \( \Gamma_0^{-1} \otimes \Gamma_0 \geq I_m, \) then \(-\Sigma_0 > 0.\)

From the proof, we can see that the sixth-order moment in models (2.4)-(2.5) is required only for Lemma 5.4(a), while the fourth-order moment is sufficient for Lemma 5.4(b). If we can show that the convergent rate of the QMLE is \( O_p(n^{-1/2}), \) then the fourth-order moment is sufficient for models (2.4)-(2.5). However, it would seem that proving the rate of convergence is quite difficult.

**Lemma 5.6.** Under the assumptions of Lemma 5.2, if \( \sqrt{n}(\lambda_n - \lambda_0) = O_p(1), \) then

\[
(a) \ \frac{1}{n} \sum_{t=1}^{n} \left[ \frac{\partial l_t^e}{\partial \lambda} \frac{\partial l_t^e}{\partial \lambda'} - \frac{\partial l_t^e}{\partial \lambda_t} \frac{\partial l_t^e}{\partial \lambda_t'} \right]_{\lambda_n} = o_P(1),
\]

\[
(b) \ \hat{\Omega}_n = \frac{1}{n} \sum_{t=1}^{n} \left[ \frac{\partial l_t^e}{\partial \lambda} \frac{\partial l_t^e}{\partial \lambda'} \right]_{\lambda_n} = \Omega_0 + o_P(1).
\]

**Theorem 5.1.** Suppose that \( Y_t \) is generated by models (2.4)-(2.5) satisfying Assumptions 1-4 and \( E||Y_t||^6 < \infty, \) or models (2.8)-(2.9) satisfying Assumptions 1-3 and \( E||Y_t||^4 < \infty, \) or models (2.10)-(2.11) satisfying Assumption 3 and \( E||\eta_0||^4 < \infty. \) If \( \Omega_0 > 0 \) and \( \Gamma_0^{-1} \otimes \Gamma_0 \geq I_m, \) then \( \sqrt{n}(\hat{\lambda}_n - \lambda_0) \to c N(0, \Sigma_0^{-1} \Omega_0 \Sigma_0^{-1}). \) Furthermore, \( \Sigma_0 \) and \( \Omega_0 \) can be estimated consistently by \( \hat{\Sigma}_n \) and \( \hat{\Omega}_n, \) respectively.
When \( m = 1 \) or \( 2 \), we can show that \( \Gamma_0^{-1} \otimes \Gamma_0 \geq I_m \), and hence, in this case, \(-\Sigma_0 > 0\). However, it is difficult to prove \( \Gamma_0^{-1} \otimes \Gamma_0 \geq I_m \) for the general case. When \( \Gamma_0 = I_m \), it is straightforward to show that \(-\Sigma_0 > 0\) and \( \Omega_0 \) are positive definite.

When \( \eta_{0t} \) follows a symmetric distribution, \[ \Omega_0 = E\left( \begin{array}{cc} \frac{\partial \varphi_{0t}}{\partial \varphi} & 0 \\ \frac{\partial \varphi_{0t}}{\partial \delta} & \frac{\partial \varphi_{0t}}{\partial \sigma} \end{array} \right) \] and \( \Sigma_0 = -\left( \begin{array}{cc} \Sigma_{\varphi 0} & 0 \\ 0 & \Sigma_{\delta 0} \end{array} \right) \),

in which \( \tilde{\delta} = (\delta', \sigma')' \),

\[ \Sigma_{\varphi 0} = E\left[ \frac{\partial \varphi_{0t}}{\partial \varphi} (D_{0t} \Gamma_0 D_{0t})^{-1} \frac{\partial \varphi_{0t}}{\partial \varphi} \right] + \frac{1}{4} E\left[ \frac{\partial H_{0t}^c}{\partial \varphi} D_{0t}^{-2} C D_{0t}^{-2} \frac{\partial H_{0t}^c}{\partial \varphi} \right], \]

\[ \Sigma_{\delta 0} = \left( \begin{array}{cc} \Sigma_{\delta 0} & \Sigma_{\delta \sigma 0} \\ \Sigma_{\delta \sigma 0} & \frac{1}{2} P' P \end{array} \right), \]

where \( \Sigma_{\delta 0} = E[\partial H_{0t}^c / \partial \delta D_{0t}^{-2} C D_{0t}^{-2} \partial H_{0t}^c / \delta]/4 \) and \( \Sigma_{\delta \sigma 0} = E[\partial H_{0t}^c / \partial \delta D_{0t}^{-2}] C_1 P/2 \).

Furthermore, if \( \eta_{0t} \) is normal, it follows that \(-\Sigma_0 = \Omega_0\). Note that the QMLE here is the global maximum over the whole parameter space. The requirement of the sixth-order moment is quite strong for models (2.4)-(2.5), and is used only because we need to verify the uniform convergence of the second derivative of the log-likelihood function, that is, Lemma 5.4(a). If we consider only the local QMLE, then the fourth-order moment is sufficient. For univariate cases, such proofs can be found in Ling and Li (1998) and Ling and McAleer (1999).

6 CONCLUSION

This paper presented the asymptotic theory for a vector ARMA-GARCH model. An explanation of the proposed model was offered. Using a similar idea, different multivariate models such as E-GARCH, threshold GARCH, and asymmetric GARCH can be proposed for modelling multivariate conditional heteroskedasticity. The conditions for the strict stationarity and ergodicity of the vector ARMA-GARCH model were obtained. A simple sufficient condition for the higher-order moments of the model was also provided. We established a uniform convergence result by modifying a theorem in Amemiya (1985). Based on the uniform convergence result, the
consistency of the QMLE was obtained under only the second-order moment condition. Unlike Weiss (1986) and Pantula (1989) for the univariate case, the asymptotic normality of the QMLE for the vector ARCH model requires only the second-order moment of the unconditional errors, and the finite fourth-order moment of the conditional errors. The asymptotic normality of the QMLE for the vector ARMA-ARCH model was proved using the fourth-order moment, which is an extension of Weiss (1986) and Pantula (1989). For the general vector ARMA-GARCH model, the asymptotic normality of the QMLE requires the assumption that the sixth-order moment exists. Whether this result will hold under some weaker moment conditions remains to be proved.

NOTES
1. For models (2.8)-(2.9) and (2.10)-(2.11), $B_i$ in Assumption 3 reduces to the zero matrix, where $i = 1, \ldots, s$.

2. The Co-editor has suggested that this theorem may not be new, consisting of Lemma 2.4 and footnote 18 of Newey and McFadden (1994).

REFERENCES


Bollerslev, T. (1990) Modelling the coherence in short-run nominal exchange rates:


Lumsdaine, R.L. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH (1,1) and covariance stationary GARCH(1,1) models. *Econometrica* 64, 575-596.


A \textbf{APPENDIX: PROOFS OF THEOREMS 2.1-2.2}

\textbf{Proof of Theorem 2.1.} Multiplying (2.5) by \( \tilde{\eta}_{0t} \) yields

\[ \tilde{\varepsilon}_{0t} = \tilde{\eta}_{0t}W_0 + \sum_{i=1}^{r} \tilde{\eta}_{0t}A_{0i}\tilde{\varepsilon}_{0t-i} + \sum_{i=1}^{s} \tilde{\eta}_{0t}B_{0i}H_{0t-i}. \]  
(A.1)

Now rewrite (A.1) in vector form as

\[ X_t = \tilde{A}_{0t}X_{t-1} + \xi_t, \]  
(A.2)

where \( X_t = (\tilde{\varepsilon}_{0t}, \cdots, \tilde{\varepsilon}_{0t-r+1}, H'_{0t}, \cdots, H'_{0t-s+1})' \) and \( \xi_t \) is defined as in (2.9). Let

\[ S_{n,t} = \xi_t + \sum_{j=1}^{n} \left( \prod_{i=1}^{j} \tilde{A}_{0t-i+1} \right) \xi_{t-j}, \]  
(A.3)

where \( n = 1, 2, \cdots \). Denote the \( k \)th element of \( (\prod_{i=1}^{j} \tilde{A}_{0t-i}) \xi_{t-j-1} \) by \( s_{n,t} \). We have

\[ E|s_{n,t}| = e'_k E(\prod_{i=1}^{j} \tilde{A}_{0t-i}) \xi_{t-j-1} = e'_k \tilde{A}^j c^*, \]  
(A.4)
Thus, \( \rho(\tilde{A}) < 1 \) and hence each component of \( \tilde{A}^i \) is \( O(\rho^i) \). Therefore, the right-hand side of (A.4) is equal to \( O(\rho^t) \). Note that \( \tilde{\eta}_{0t} \) is a sequence of i.i.d. random matrices, and each element of \( \tilde{A}_{0t} \) and \( \xi_t \) is non-negative. We know that each component of \( S_{n,t} \) converges almost surely (a.s.) as \( n \to \infty \), as does \( S_{n,t} \). Denote the limit of \( S_{n,t} \) by \( X_t \). We have

\[
X_t = \xi_t + \sum_{j=1}^{\infty} \left( \prod_{i=1}^{j} \tilde{A}_{0t-i} \right) \xi_{t-j-1},
\]

with the first-order moment being finite.

It is easy to verify that \( X_t \) satisfies (A.2). Hence, there exists an \( \mathcal{F}_t \)-measurable second-order solution \( \varepsilon_{0t} \) to (2.5) with \( i \)th element \( \varepsilon_{0it} = \eta_{0it} \sqrt{\tilde{h}_{0it}} = \eta_{0it}(\varepsilon'_{rm+i}X_t)^{1/2} \), with the representation (2.13).

Now we show that such a solution is unique to (2.5). Let \( \varepsilon_t^{(1)} \) be another \( \mathcal{F}_t \)-measurable second-order stationary solution of (2.5). As in (A.2), we have \( X_t^{(1)} = \tilde{A}_{0t}X_{t-1}^{(1)} + \xi_t \), where \( X_t^{(1)} = (\varepsilon_t^{(1)}, \ldots, \varepsilon_t^{(1)}, H_t^{(1)}, \ldots, H_{t-s+1}^{(1)}', \ldots, H_{t-s+1}^{(1)}')' \), and \( H_t^{(1)} = W_0 + \sum_{i=1}^{s} A_{0i} \varepsilon_{t-i}^{(1)} + \sum_{i=1}^{s} B_{0i} H_{t-i}^{(1)} \) with \( \varepsilon_t^{(1)} = (\varepsilon_t^{(1)}, \ldots, \varepsilon_t^{(1)})' \). Let \( U_t = X_t - X_t^{(1)} \). Then \( U_t \) is first-order stationary and, by (A.2), \( U_t = (\prod_{i=0}^{n-1} \tilde{A}_{0t-i})U_{t-n-1} \). Denote the \( k \)th component of \( U_t \) as \( u_{k,t} \). Then, as each element of \( \tilde{A} \) is nonnegative,

\[
|u_{k,t}| \leq |\varepsilon_k^{(1)}(\prod_{i=0}^{n} \tilde{A}_{0t-i})U_{t-n-1}| \leq \varepsilon_k^{(1)}(\prod_{i=0}^{n} \tilde{A}_{0t-i})|U_{t-n-1}|,
\]

where \( e_k = (0, \ldots, 0, 1, 0, \ldots, 0)'_{m(r+s) \times 1} \) and 1 appears in the \( k \)th position, \( e^* = E\xi_t \) is a constant vector, and

\[
\tilde{A} = \begin{pmatrix}
A_{01} & \cdots & A_{0r} \\
I_{m(r-1)} & O_{m(r-1) \times m} & B_{01} \\
A_{01} & \cdots & A_{0r} \\
O_{m(s-1) \times mr} & B_{01} & I_{m(s-1)} & O_{m(s-1) \times m}
\end{pmatrix}.
\]

By direct calculation, we know that the characteristic polynomial of \( \tilde{A} \) is

\[
f(z) = |z|(r+s)m |I_m - \sum_{i=1}^{n} A_i z^{-i} - \sum_{i=1}^{n} B_i z^{-i}|.
\]

(2.13)
where $e_k$ is defined as in (A.4), and $|U_t|$ is defined as $(|u_{1t}|, \ldots, |u_{(r+s)m,t}|)'$. As $U_t$ is first-order stationary and $\mathcal{F}_t$–measurable, by (A.8), we have

$$
E|u_{kt}| \leq e_k' E(\prod_{i=0}^{n} \tilde{A}_{0t-i}) E|U_{t-n-1}| = e_k' \tilde{A}^n c^*_1 \to 0 \tag{A.9}
$$

as $n \to \infty$, where $c^*_1 = E|U_t|$ is a constant vector. So $u_{kt} = 0$ a.s., that is, $X_t = X_t^{(1)}$ a.s.. Thus, $h_{it} = h_{it}^{(1)}$ a.s., and hence $\varepsilon_{0t} = \varepsilon_{0t}^{(1)} = \eta_{0it} h_{0it}^{1/2}$ a.s.. That is, $\varepsilon_{0t}$ satisfying (2.5) is unique.

For the unique solution $\varepsilon_{0t}$, by the usual method, we can show that there exists a unique $\mathcal{F}_t$–measurable second-order stationary solution $Y_t$ satisfying (2.4), with the expansion given by

$$
Y_t = \sum_{k=0}^{\infty} \Upsilon_{0k} \varepsilon_{0t-k}. \tag{A.10}
$$

Note that the solution $\{Y_t, \varepsilon_{0t}, H_{0t}\}$ is a fixed function of a sequence of i.i.d. random vectors $\eta_{0it}$, and hence is strictly stationary and ergodic. This completes the proof.

The proof of Theorem 2.2 first transforms models (2.4)-(2.5) into a Markov chain and then uses Tweedie’s criterion. Let $\{X_t; t = 1, 2, \ldots\}$ be a temporally homogeneous Markov chain with a locally compact completely separable metric state space $(S, \mathcal{B})$. The transition probability is $P(x, A) = Pr(X_n \in A|X_{n-1} = x)$, where $x \in S$ and $A \in \mathcal{B}$. Tweedie’s criterion is the following lemma.

**Lemma A.1.** (Tweedie, 1988, Theorem 2) Suppose that $\{X_t\}$ is a Feller chain. (1) If there exists, for some compact set $A \in \mathcal{B}$, a non-negative function $g$ and $\varepsilon > 0$ satisfying

$$
\int_{A^c} P(x, dy) g(y) \leq g(x) - \varepsilon, \quad x \in A^c, \tag{A.11}
$$

then there exists a $\sigma$-finite invariant measure $\mu$ for $P$ with $0 < \mu(A) < \infty$; (2) Furthermore, if

$$
\int_A \mu(dx) [\int_{A^c} P(x, dy) g(y)] < \infty, \tag{A.12}
$$

2
then \( \mu \) is finite, and hence \( \pi = \mu / \mu(S) \) is an invariant probability; (3) Furthermore, if

\[
\int_{A^c} P(x, dy) g(y) \leq g(x) - f(x), \quad x \in A^c,
\]

then \( \mu \) admits a finite \( f \)-moment, i.e.

\[
\int_S \mu(dy) f(y) < \infty.
\]  

The following two lemmas are preliminary results for the proof of Theorem 2.2.

**Lemma A.2.** Suppose that \( E(||\eta_0||^{2k}) < \infty \) and \( \rho[E(\tilde{A}_{0t}^{\otimes k})] < 1 \). Then there exists a vector \( M > 0 \) such that \([I_m - E(\tilde{A}_{0t}^{\otimes k})]M > 0\), where a vector \( B > 0 \) means that each element of \( B \) is positive.

**Proof.** From the condition given, \( I_m - E(\tilde{A}_{0t}^{\otimes k}) \) is invertible. Since each element of \( E(\tilde{A}_{0t}^{\otimes k}) \) is non-negative, we can choose a vector \( L_1 > 0 \) such that

\[
M := [I_m - E(\tilde{A}_{0t}^{\otimes k})]^{-1}L_1 = L_1 + \sum_{i=1}^{\infty}[E(\tilde{A}_{0t}^{\otimes k})]iL_1 > 0.
\]

Thus, \([I_m - E(\tilde{A}_{0t}^{\otimes k})]M = L_1 > 0\). This completes the proof.

**Lemma A.3.** Suppose that there is a vector \( M > 0 \) such that

\[
[I_m - E(\tilde{A}_{0t}^{\otimes k})]M > 0.
\]  

Then there exists a compact set \( A = \{x : \tilde{x}^k \equiv (\sum_{i=1}^{r+s}m x_i)^k \leq \Delta\} \subset R_0^{(r+s)m} \) with \( R_0 = (0, \infty) \), a function \( g_1(x) \), and \( \kappa > 0 \) such that the function \( g \), defined by \( g(x) = 1 + (x^{\otimes k})'M \), satisfies

\[
E(g(X_t)|X_{t-1} = x) \leq g(x) + g_1(x), \quad x \in R_0^{(r+s)m},
\]  

and

\[
E(g(X_t)|X_{t-1} = x) \leq (1 - \kappa)g(x), \quad x \in A^c,
\]  

where \( A^c = R^{(r+s)m} - A \), \( x_i \) is the \( i \)th component of \( x \), \( \max_{x \in A} g_1(x) < C_0 \), \( X_t \) is defined as in (A.2), and \( C_0, \kappa \) and \( \Delta \) are positive constants not depending on \( x \).
Proof. We illustrate the proof for \( k = 3 \). The technique for \( k \neq 3 \) is analogous.

For any \( x \in R_0^{(r+s)m} \), by straightforward algebra, we can show that

\[
E[(\xi_t + \hat{A}_{0t}x)^{\otimes 3}]M
\leq (x^{\otimes 3})'E(\hat{A}_{0t}^{\otimes 3})'M + c(1 + \tilde{x} + \tilde{x}^2),
\]

where \( C_1, C_2 \) and \( C_3 \) are some constant vectors or matrices with non-negative elements, which do not depend on \( x \), and \( c = \max_k \{ \text{all components of } C_1'M, C_2'M \text{ and } C_3'M \} \).

By (A.2) and (A.18), we have

\[
E[g(X_t)|X_{t-1} = x] = 1 + E[(\xi_t + \hat{A}_{0t}x)^{\otimes 3}]M
\leq 1 + (x^{\otimes 3})'E(\hat{A}_{0t}^{\otimes 3})'M + g_1(x)
= 1 + (x^{\otimes 3})'M - (x^{\otimes 3})'M^* + g_1(x)
= g(x)[1 - \left(\frac{(x^{\otimes 3})'M^*}{g(x)}\right) + \frac{g_1(x)}{g(x)}],
\]

where \( M^* = [I_m - E(\hat{A}_{0t}^{\otimes 3})']M \) and \( g_1(x) = c(1 + \tilde{x} + \tilde{x}^2) \).

Denote

\[
A = \{ x : \tilde{x}^3 \leq \Delta, x \in R_0^{(r+s)m} \},
\]

\[
c_1 = \min \{ \text{all components of } M^* \},
\]

\[
c_2 = \max \{ \text{all components of } M \},
\]

\[
c_3 = \min \{ \text{all components of } M \}.
\]

It is obvious that \( A \) is a compact set on \( R_0^{(r+s)m} \). Since \( M^*, M > 0 \), it follows that \( c_1, c_2, c_3 > 0 \). From (A.19), we can show that

\[
E[g(X_t)|X_{t-1} = x] \leq g(x) + g_1(x), \quad x \in R_0^{(r+s)m},
\]

where \( \max_{x \in A} g_1(x) < C_0(\Delta) \), and \( C_0(\Delta) \) is a constant not depending on \( x \).

Let \( \Delta > \max \{ 1/c_2, 1 \} \). When \( x \in A^c \),

\[
c_3\Delta < c_2\tilde{x}^3 \leq g(x) \leq 1 + c_2\tilde{x}^3 \leq 2c_2\tilde{x}^3.
\]
Thus,
\[
\frac{(x^{\otimes 3})^t M^t}{g(x)} \geq \frac{c_1 \tilde{x}^3}{2c_2 \tilde{x}^3} = \frac{c_1}{2c_2}, \tag{A.22}
\]
and furthermore, since \(1 + \tilde{x} \leq 2\tilde{x}\) as \(x \in A^c\), we can show that
\[
\frac{g_1(x)}{g(x)} \leq \frac{g_1(x)}{c_3 \tilde{x}^3} \leq \frac{C}{\Delta}, \tag{A.23}
\]
where \(C\) is a positive constant not depending on \(x\) and \(\Delta\). By (A.19), and (A.22)-(A.23), as \(x \in A^c\),
\[
E[g(X_t)|X_{t-1} = x] \leq g(x)(1 - \frac{c_1}{2c_2} + \frac{C}{\Delta}).
\]
Provided \(0 < c_1/4c_2 < \kappa < c_1/2c_2\) and \(\Delta > \max\{1, 1/c_2, C/(c_1/2c_2 - \kappa)\}\), then
\[
E[g(X_t)|X_{t-1} = x] \leq g(x)(1 - \kappa). \tag{A.24}
\]
This completes the proof. 2

Proof of Theorem 2.2. Obviously, \(X_t\) defined as in (A.2) is a Markov chain with state space \(R^{(r+s)m}_0\). It is straightforward to prove that, for each bounded
continuous function \(g\) on \(R^{(r+s)m}_0\), \(E[g(X_t)|X_{t-1} = x]\) is continuous in \(x\), that is, \(\{X_t\}\) is a Feller chain. In a similar manner to Lemma A.3, in the following we illustrate only that the conditions (A.11)-(A.13) are satisfied for \(k = 3\).

From Lemmas A.2 and A.3, we know that there exists a vector \(M > 0\), a compact
set \(A = \{x : \tilde{x}^3 = (\sum_{i=1}^{(r+s)m} x_i)^3 \leq \Delta\} \subset R^{(r+s)m}_0\) and \(\kappa > 0\) such that the function
defined by \(g(x) = 1 + (x^{\otimes 3})^t M\) satisfies
\[
E[g(X_t)|X_{t-1} = x] \leq g(x) + g_1(x), \ x \in R^{(r+s)m}_0 \tag{A.24}
\]
and
\[
E[g(X_t)|X_{t-1} = x] \leq (1 - \kappa)g(x), \ x \in A^c, \tag{A.25}
\]
where \(\max_{x \in A} g_1(x) < C_0\), and \(C_0, \kappa\) and \(\Delta\) are positive constants not depending on \(x\).

Since \(g(x) \geq 1\), it follows that \(E[g(X_t)|X_{t-1} = x] \leq g(x) - \kappa\). By Lemma A.1, there exists a \(\sigma\)-finite invariant measure \(\mu\) for \(P\) with \(0 < \mu(A) < \infty\).
Denote $c_2 = \max\{all\ components\ of\ M\}$ and $c_3 = \min\{all\ components\ of\ M\}$. From (A.24), as $x \in A$, it is easy to show that

$$E[g(X_t)|X_{t-1} = x] \leq 1 + c_2 \left( \sum_{i=1}^{(r+s)m} x_i \right)^3 + g_1(x) \leq \Delta_1 < \infty,$$

where $\Delta_1$ is a constant not depending on $x$. Hence,

$$\int_A \mu(dx) \left\{ \int_{A^c} P(x, dy)g(y) \right\} \leq \int_A \mu(dx)E[g(X_t)|X_{t-1} = x] \leq \Delta_1 \mu(A) < \infty.$$

This shows that $\{X_t\}$ has a finite invariant measure $\mu$, and hence $\pi = \mu/\mu(R_0^{(r+s)m})$ is an invariant probability measure of $\{X_t\}$, that is, there exists a strictly stationary solution satisfying (A.2), still denoted by $X_t$.

Let $f(x)$ be the function on $R_0^{(r+s)m}$ defined by $f(x) = c_3 \kappa \left( \sum_{i=1}^{(r+s)m} x_i \right)^3$. Then, by (A.25), as $x \in A^c$, we have

$$\int_{A^c} P(x, dy)g(y) \leq E[g(X_t)|X_{t-1} = x] \leq g(x) - \kappa g(x) \leq g(x) - f(x).$$

By Lemma A.1(3), we know that $E_\pi[f(X_t)] = c_3 \kappa E[\left( \sum_{i=1}^{(r+s)m} x_i \right)^3] < \infty$, where $\pi$ is the stationary distribution of $\{X_t\}$, where $x_{it}$ is the $i$-th component of $X_t$. Thus, $E_\pi_1||\varepsilon_{0t}||^6 < \infty$, where $\pi_1$ are the stationary distributions of $\{\varepsilon_{0t}\}$. Now, since $E_\pi_1||\varepsilon_{0t}||^6 < \infty$, it is easy to show that $E_{\pi_2}||Y_t||^6 < \infty$, where $\pi_2$ is the stationary distribution of $Y_t$.

By Hölder’s inequality, $E_\pi_1||\varepsilon_{0t}||^2 < (E_\pi_1||\varepsilon_{0t}||^{2k})^{\frac{1}{k}} < \infty$. Similarly, we have $E_{\pi_2}||Y_t||^2 < \infty$. Thus, $\{Y_t, \varepsilon_{0t}\}$ is a second-order stationary solution of models (2.4)-(2.5). Furthermore, by Theorem 2.1, the solution $\{Y_t, \varepsilon_{0t}\}$ is unique and ergodic. Thus, the process $\{Y_t, \varepsilon_{0t}\}$ satisfying models (2.4)-(2.5) has a finite $2k$th moment. This completes the proof.
BAPPENDIX: PROOFS OF RESULTS IN SECTIONS 3-5

Proof of Theorem 3.1. The proof is similar to that of Theorem 4.2.1 in Amemina (1985), except that the Kolmogorov law of large numbers is replaced by the ergodic theorem. This completes the proof. 2

Proof of Lemma 4.1. Note that

$$\tilde{A} \succeq \begin{pmatrix} O & O \\ O & \tilde{B} \end{pmatrix},$$

where $\tilde{A}$ is defined as in (A.5), $\tilde{B} = \begin{pmatrix} B_{01} & \cdots & B_{0s} \\ I_{m(s-1)} & O_{m(s-1) \times m} \end{pmatrix}$, and here "the matrix $A \succeq$ the matrix $B$" means that each component of $A$ is larger than or equal to the corresponding component of $B$. Thus, we have

$$\tilde{A}^i \succeq \begin{pmatrix} O & O \\ O & \tilde{B}^i \end{pmatrix}. \quad (B.1)$$

By Assumption 3, $\rho(\tilde{A}) < 1$, and hence $\sum_{i=0}^{k} \tilde{A}^i$ converges to a finite limit as $k \to \infty$. By (B.1), $\sum_{i=0}^{k} \tilde{B}^i$ also converges to a finite limit as $k \to \infty$, and hence $\rho(\tilde{B}) < 1$, which is equivalent to all the roots of $|I_m - \sum_{i=1}^{s} B_i L^i| = 0$ lying outside the unit circle. This completes the proof. 2

In the following, we prove Lemmas 4.2-4.4 and 4.6 and Theorem 4.1 only for models (2.4)-(2.5). The proofs for models (2.8)-(2.9) and (2.10)-(2.11) are similar and simpler, and hence are omitted.

Proof of Lemma 4.2. First, by (3.2),

$$\varepsilon_t^i = \Psi(L)^{-1} \Phi(L)(Y_t - \mu), \quad \frac{\partial \varepsilon_t^i}{\partial \varphi'} = \Psi^{-1}(L)[-\Phi(1), X_{t-1} \otimes I_m], \quad (B.2)$$

where $X_{t-1} = (Y_{t-1}^i - \mu^i, \cdots, Y_{t-p+1}^i - \mu^i, \varepsilon_{t-1}^i, \cdots, \varepsilon_{t-q+1}^i)$, and the above vector differentiation follows rules in Lütkepohl (1993, Appendix A). Denote $U_t = \partial \varepsilon_t^i / \partial \varphi'$ and $V_t = [-\Phi(1), X_{t-1} \otimes I_m]$. Then

$$U_t + \Psi_1 U_{t-1} + \cdots + \Psi_q U_{t-q} = V_t. \quad (B.3)$$
If \( U_t c_\varphi = 0 \) a.s., then \( V_t c_\varphi = 0 \) a.s.. Let \( c_1 \) be the vector consisting of the first \( m \) elements of \( c_\varphi \), while \( c_2 \) is the vector consisting of the remaining elements of \( c_\varphi \). Then \( -\Phi(1)c_1 + (X_{t-1} \otimes I_m)c_2 = 0 \). Since \( X_{t-1} \) is not degenerate, \( (X_{t-1} \otimes I_m)c_2 = 0 \) and \( \Phi(1)c_1 = 0 \). By Assumption 1, \( \Phi(1) \) is of full rank, and hence \( c_1 = 0 \). By Assumption 2, we can show that \( c_2 = 0 \). Thus, \( c_\varphi = 0 \).

Next, by (3.3),

\[
H_t^c = (I_m - \sum_{i=1}^{s} B_i L^i)^{-1}[W + (\sum_{i=1}^{r} A_i L^i)\varepsilon_t^i], \tag{B.4}
\]

\[
\frac{\partial H_t^c}{\partial \delta^i} = (I_m - \sum_{i=1}^{s} B_i L^i)^{-1}(I_m, \tilde{H}_{t-1}^c \otimes I_m), \tag{B.5}
\]

where \( \tilde{H}_{t-1}^c = (\varepsilon_{t-1}', \ldots, \varepsilon_{t-r}', H_{t-1}', \ldots, H_{t-s}') \). Denoting \( U_{1t} = \partial H_t^c / \partial \delta^i \) and \( V_{1t} = (I_m, \tilde{H}_{t-1}^c \otimes I_m) \), we have the following recursive equation:

\[
U_{1t} = B_1 U_{1t-1} + \cdots + B_s U_{1t-s} + V_{1t}. \tag{B.6}
\]

If \( U_{1t} c = 0 \) a.s., then \( V_{1t} c = 0 \) a.s.. By Assumptions 3-4, in a similar manner to \( V_t c_\varphi = 0 \), we can conclude \( c = 0 \) (also refer to the proof of Proposition 3.4 in Jeantheau (1998)). This completes the proof.

**Proof of Lemma 4.3.** As the parameter space \( \Theta \) is compact, all the roots of \( \Phi(L) \) lie outside the unit circle, and the roots of a polynomial are continuous functions of its coefficients, there exist constants \( c_0, c_1 > 0 \) and \( 0 < \rho < 1 \), independent of all \( \lambda \in \Theta \), such that

\[
||\varepsilon_t^c|| \leq c_0 + c_1 \sum_{i=0}^{\infty} \rho^i ||Y_{t-i}|| \equiv \varepsilon_t^h. \tag{B.7}
\]

Thus, \( E \sup_{\lambda \in \Theta} ||\varepsilon_t^c||^2 < \infty \) by Theorem 2.1. Note that, by Assumption 3, \( |D_t^c \Gamma D_t^c| \) has a lower bound uniformly over \( \Theta \). We have \( E \sup_{\lambda \in \Theta} [\varepsilon_t^c (D_t^c \Gamma D_t^c)^{-1} \varepsilon_t^c] < \infty \). By Assumption 3 and Lemma 4.1, we can show that

\[
||H_t^c|| \leq c_2 + c_3 \sum_{i=1}^{\infty} \rho_1^i ||Y_{t-i}||^2 \equiv \varepsilon_{ht}^c, \tag{B.8}
\]

where \( c_2, c_3 > 0 \) and \( 0 < \rho_1 < 1 \) are constants independent of all \( \lambda \in \Theta \). Thus, \( E \sup_{\lambda \in \Theta} ||H_t^c|| < \infty \), and hence \( E \sup_{\lambda \in \Theta} |D_t^c \Gamma D_t^c| < \infty \). By Jensen’s inequality,
\( E \sup_{\lambda \in \Theta} |\ln |D_1^t \Gamma D_1^t || < \infty \). Thus, \( E|l_t^t(\lambda)| < \infty \) for all \( \lambda \in \Theta \). Let \( g(Y_t, \lambda) = l_t^t - E l_t^t \), where \( Y_t = (Y_t, Y_{t-1}, \cdots) \). Then \( E \sup_{\lambda \in \Theta} |g(Y_t, \lambda)| < \infty \). Furthermore, since \( g(Y_t, \lambda) \) is strictly stationary with \( g(Y_t, \lambda) = 0 \), by Theorem 3.1, \( \sup_{\lambda \in \Theta} |n^{-1} \sum_{t=1}^{n} g(Y_t, \lambda)| = o_p(1) \). This completes the proof.

**Proof of Lemma 4.4.** First,

\[
-E \ln |D_1^t \Gamma D_1^t| - E[\varepsilon_t^t (D_1^t \Gamma D_1^t)^{-1} \varepsilon_t^t] \\
= -E \ln |D_1^t \Gamma D_1^t| - E[(\varepsilon_t^t - \varepsilon_{0t} + \varepsilon_{0t})'(D_1^t \Gamma D_1^t)^{-1}(\varepsilon_t^t - \varepsilon_{0t} + \varepsilon_{0t})] \\
= \{-E \ln |D_1^t \Gamma D_1^t| - E[\varepsilon_t^t (D_1^t \Gamma D_1^t)^{-1} \varepsilon_t^t] \} \\
L_1(\lambda) + L_2(\lambda). \tag{B.9}
\]

\( L_2(\lambda) \) obtains its maximum at zero, and this occurs if and only if \( \varepsilon_t^t = \varepsilon_{0t} \). Thus,

\[
\varepsilon_t^t - \varepsilon_{0t} = \frac{\partial \varepsilon_t^t}{\partial \varphi} \bigg|_{\varphi^*} (\varphi - \varphi_0) = 0. \tag{B.10}
\]

By Lemma 4.2, we know that equation (B.10) holds if and only if \( \varphi = \varphi_0 \).

\[
L_1(\lambda) = -E \ln |D_1^t \Gamma D_1^t| - E \text{tr}(M_t) \\
= -[-E \ln |M_t| + E \text{tr}(M_t)] - E \ln |D_{0t} \Gamma_0 D_{0t}|. \tag{B.11}
\]

where \( M_t = (D_1^t \Gamma D_1^t)^{-1/2}(D_{0t} \Gamma_0 D_{0t})(D_1^t \Gamma D_1^t)^{-1/2} \). Note that, for any positive definite matrix \( M \), \(-f(M) \equiv -\ln |M| + tr M \geq m \) (see Lemma A.6 in Johansen (1995)), and hence

\[
-E \ln |M_t| + E \text{tr}(M_t) \geq m. \tag{B.12}
\]

When \( M_t = I_m \), we have \( f(M_t) = f(I_m) = -m \). If \( M_t \neq I_m \), then \( f(M_t) < f(I_m) \), so that \( Ef(M_t) \leq Ef(I_m) \) with equality only if \( M_t = I_m \) with probability one. Thus, \( L_1(\lambda) \) reaches its maximum \(-m - E \ln |D_{0t} \Gamma_0 D_{0t}|\), and this occurs if and only if \( D_1^t \Gamma D_1^t = D_{0t} \Gamma_0 D_{0t} \). From the definition of \( \Gamma \), we have \( h_{it} = h_{0it} \), and hence \( \Gamma = \Gamma_0 \). Note that

\[
\max_{\lambda \in \Theta} L(\lambda) \leq \max_{\lambda \in \Theta} L_1(\lambda) + \max_{\lambda \in \Theta} L_2(\lambda).
\]
max_{\lambda \in \Theta} L(\lambda) = -m - E \ln(D_0 \Gamma_0 D_0) if and only if max_{\lambda \in \Theta} L_2(\lambda) = 0 and max_{\lambda \in \Theta} L_1(\lambda) = -m - E \ln(D_0 \Gamma_0 D_0), which occurs if and only if \varphi = \varphi_0, \Gamma = \Gamma_0 and h_{it} = h_{0it}. From \varphi = \varphi_0 and h_{it} = h_{0it}, we have

\[ (H^e_t - H_{0it})|_{\varphi = \varphi_0} = \frac{\partial H^e_t}{\partial \varphi^{\delta^*}}|_{(\varphi_0, \delta^*)}(\delta - \delta_0) = 0 \] (B.13)

with probability one, where \delta^* lies between \delta and \delta_0. By Lemma 4.2, (B.13) holds if and only if \delta = \delta_0. Thus, L(\lambda) is uniquely maximised at \lambda_0. This completes the proof.

Proof of Lemma 4.5. First, for any positive constant M,

\[ \left| \frac{1}{n} \sum_{t=1}^{n} X_t \xi_t I(|X_t| > M) \right| \leq C \frac{1}{n} \sum_{t=1}^{n} |X_t|I(|X_t| > M), \] (B.14)

where I(\cdot) is the indicator function. For any small \epsilon, \kappa > 0, since \( E|X_t| < \infty \), there exists a constant M_0 such that

\[
P\left( \left| \frac{1}{n} \sum_{t=1}^{n} X_t \xi_t I(|X_t| > M_0) \right| > \kappa \right) \\
\leq \frac{1}{\kappa} E \left( \frac{1}{n} \sum_{t=1}^{n} X_t \xi_t I(|X_t| > M_0) \right) \\
\leq \frac{C}{\kappa} \int_{|x| > M_0} |x| dF(x) < \frac{\epsilon}{2}, \] (B.15)

where F(x) is the distribution of X_t. For such a constant M_0, by the given condition, there exists a positive integer N such that, when \( n > N \),

\[ P \left( \left| \frac{1}{n} \sum_{t=1}^{n} X_t \xi_t I(|X_t| \leq M_0) \right| > \kappa \right) \leq P \left( \frac{1}{n} \sum_{t=1}^{n} |\xi_t| > \kappa / M_0 \right) < \frac{\epsilon}{2}. \] (B.16)

By (B.15) and (B.16), as \( n > N \), \( P \left( |n^{-1} \sum_{t=1}^{n} X_t \xi_t| > 2 \kappa \right) < \epsilon \), that is, \( n^{-1} \sum_{t=1}^{n} X_t \xi_t = o_p(1) \). This completes the proof.

Proof of Lemma 4.6. For convenience, let the initial values be \( \bar{Y}_0 = 0 \) and \( \bar{\varepsilon}_0 = 0 \). When the initial values are not equal to zero, the proof is similar. By Assumption 1, \( \varepsilon^e_t \) and \( \varepsilon_t \) have the expansions:

\[ \varepsilon^e_t = \sum_{k=0}^{\infty} \Upsilon_k (Y_{t-k} - \mu), \quad \varepsilon_t = \sum_{k=0}^{t-1} \Upsilon_k (Y_{t-k} - \mu), \] (B.17)
where $\Phi^{-1}(L)\Psi(L) = \sum_{k=0}^{\infty} \gamma_k L^k$. By (B.17),

$$
||\varepsilon^\epsilon_t - \varepsilon_t|| \leq c_1 \sum_{k=t}^{\infty} q_1^k ||Y_{t-k} - \mu||, \tag{B.18}
$$

where $0 < \varrho_1 < 1$, and $c_1$ and $\varrho_1$ are constants independent of the parameter $\lambda$. By Assumption 3 and Lemma 4.1, we have

$$
H^\epsilon_t = \sum_{k=0}^{\infty} \Gamma_k [W + (\sum_{i=1}^{\rho} A_i L^i)\varepsilon^\epsilon_{t-k}], \quad H_t = \sum_{k=0}^{t-1} \Gamma_k [W + (\sum_{i=1}^{\rho} A_i L^i)\varepsilon_{t-k}], \tag{B.19}
$$

where $(I_m - \sum_{i=1}^{\rho} B_i L^i)^{-1} = \sum_{k=0}^{\infty} \Gamma_k L^k$. By (B.19)

$$
||H^\epsilon_t - H_t|| \leq \sum_{k=t}^{\infty} q_2^k (c_2 + c_3 ||\varepsilon^\epsilon_{t-k} - \varepsilon_{t-k}||), \tag{B.20}
$$

where $0 < \varrho_2 < 1$, and $c_2$, $c_3$ and $\varrho_2$ are constants independent of the parameter $\lambda$. By (B.18) and (B.20), we have

$$
E \sup_{\lambda \in \Theta} (\varepsilon^\epsilon_{it} - \varepsilon_{it})^2 = O(\varrho^t) \quad \text{and} \quad E \sup_{\lambda \in \Theta} |h^\epsilon_{it} - h_{it}| = O(\varrho^t), \tag{B.21}
$$

where $i = 1, \ldots, m$, $0 < \varrho < 1$, and $O(\cdot)$ holds uniformly in all $t$. Since $h_{it}$ has a lower bound, by (B.21), it follows that

$$
\frac{1}{n} \sum_{i=1}^{n} E \sup_{\lambda \in \Theta} |\ln |D_t^\epsilon \Gamma D_t|| - \ln |D_t \Gamma D_t||
\leq \sum_{i=1}^{m} \left[ \frac{1}{n} \sum_{t=1}^{n} E \sup_{\lambda \in \Theta} |\ln \left( \frac{h^\epsilon_{it}}{h_{it}} \right) || \right]
\leq \sum_{i=1}^{m} \left[ \frac{1}{n} \sum_{t=1}^{n} E \sup_{\lambda \in \Theta} |h^\epsilon_{it} - h_{it}| |h_{it}| \right]
= O(1) \sum_{i=1}^{m} \left[ \frac{1}{n} \sum_{t=1}^{n} E \sup_{\lambda \in \Theta} |h^\epsilon_{it} - h_{it}| \right]
= O(1) \sum_{i=1}^{m} \left[ \frac{1}{n} \sum_{t=1}^{n} O(\varrho^t) = o(1). \tag{B.22} \right]
$$

Again, since $h^\epsilon_{it}$ and $h_{it}$ have a lower bound uniformly in all $t$, $i$ and $\lambda$,

$$
\sum_{i=1}^{m} \left\{ \frac{\varepsilon^\epsilon_{it}}{\sqrt{h^\epsilon_{it}}} - \frac{\varepsilon_{it}}{\sqrt{h_{it}}} \right\}^2 \leq \sum_{i=1}^{m} \left[ \varepsilon^\epsilon_{it} \left( \frac{1}{\sqrt{h^\epsilon_{it}}} - \frac{1}{\sqrt{h_{it}}} \right)^2 + (\varepsilon^\epsilon_{it} - \varepsilon_{it})^2 \right] O(1), \tag{B.23} \right.$$
where \(O(1)\) holds uniformly in all \(t\). We have

\[
\left| \varepsilon_t' (D_t^r D_t^r)^{-1} \varepsilon_t' - \varepsilon_t' (D_t^r D_t^r)^{-1} \varepsilon_t \right| \\
= \left| 2 \varepsilon_t' D_t^r \Gamma^{-1} (D_t^r \varepsilon_t' - D_t^r \varepsilon_t) - (\varepsilon_t' D_t^r \varepsilon_t' - \varepsilon_t' D_t^r \varepsilon_t) \Gamma^{-1} (D_t^r \varepsilon_t' - D_t^r \varepsilon_t) \right| \\
\leq \left( \sum_{i=1}^{m} \left( \frac{\varepsilon_{it}'}{\sqrt{h_{it}}} - \frac{\varepsilon_{it}}{\sqrt{h_{it}}} \right)^2 \right)^{1/2} \| \varepsilon_t' \| O(1) + \left( \sum_{i=1}^{m} \left( \frac{\varepsilon_{it}'}{\sqrt{h_{it}}} - \frac{\varepsilon_{it}}{\sqrt{h_{it}}} \right)^2 \right) O(1) \\
\leq \sum_{i=1}^{m} \left[ \| \varepsilon_t' \| \| \varepsilon_{it}' \| \frac{1}{\sqrt{h_{it}}} - \frac{1}{\sqrt{h_{it}}} \right] + \varepsilon_{it}^2 \left( \frac{1}{\sqrt{h_{it}}} - \frac{1}{\sqrt{h_{it}}} \right)^2 O(1) \\
+ \sum_{i=1}^{m} \left[ \| \varepsilon_t' \| \| \varepsilon_{it}' - \varepsilon_{it} \| + (\varepsilon_{it}' - \varepsilon_{it})^2 \right] O(1) \\
= O(1) R_{1t} + O(1) R_{2t}, \tag{B.24}
\]

where \(O(1)\) holds uniformly in all \(t\) and the second inequality comes from (B.23).

By (B.7) and (B.21), it is easy to show that \(n^{-1} \sum_{t=1}^{n} \sup_{\lambda \in \Theta} R_{2t} = o_p(1)\). Thus, it is sufficient to show that \(n^{-1} \sum_{t=1}^{n} \sup_{\lambda \in \Theta} R_{1t} = o_p(1)\). Let \(X_t = \varepsilon_t^2\) and \(\xi_t = \sup_{\lambda \in \Theta} |h_{it}^{\epsilon-1/2} - h_{it}^{-1/2}|^2\), where \(\varepsilon_t^*\) is defined by (B.7). Then, \(X_t\) is a strictly stationary and ergodic time series, with \(EX_t < \infty\) and \(|\xi_t| \leq C\), a constant. Furthermore, by (B.21),

\[
\frac{1}{n} \sum_{t=1}^{n} \xi_t = \frac{1}{n} \sum_{t=1}^{n} \sup_{\lambda \in \Theta} \left| \frac{h_{it}^\epsilon - h_{it}}{\sqrt{h_{it}^\epsilon h_{it} (\sqrt{h_{it}^\epsilon} + \sqrt{h_{it}})}} \right|^2 \\
\leq \frac{1}{n} \sum_{t=1}^{n} \sup_{\lambda \in \Theta} \left| \frac{h_{it}^\epsilon - h_{it}}{h_{it}^\epsilon h_{it} (\sqrt{h_{it}^\epsilon} + \sqrt{h_{it}})} \right|^2 \\
\leq O(1) \frac{1}{n} \sum_{t=1}^{n} \sup_{\lambda \in \Theta} \left| h_{it}^\epsilon - h_{it} \right| \\
= O(1) \frac{1}{n} \sum_{t=1}^{n} O_p(\phi^t) = o_p(1).
\]

By Lemma 4.5, \(n^{-1} \sum_{t=1}^{n} X_t \sup_{\lambda \in \Theta} |h_{it}^{\epsilon-1/2} - h_{it}^{-1/2}|^2 = o_p(1)\). Similarly, we can show that \(n^{-1} \sum_{t=1}^{n} X_t \sup_{\lambda \in \Theta} |h_{it}^{\epsilon-1/2} - h_{it}^{-1/2}| = o_p(1)\). Thus,

\[
\frac{1}{n} \sum_{t=1}^{n} \sup_{\lambda \in \Theta} R_{1t} \leq \sum_{i=1}^{m} \left\{ \frac{1}{n} \sum_{t=1}^{n} \left[ X_t \sup_{\lambda \in \Theta} \left| h_{it}^{\epsilon-1/2} - h_{it}^{-1/2} \right|^2 \\
+ X_t \sup_{\lambda \in \Theta} \left| h_{it}^{\epsilon-1/2} - h_{it}^{-1/2} \right| \right] \right\} = o_p(1).
\]

This completes the proof.
Proof of Theorem 4.1. First, the space $\Theta$ is compact and $\lambda_0$ is an interior point in $\Theta$. Second, $L_n(\lambda)$ is continuous in $\lambda \in \Theta$ and is a measurable function of $Y_t$, $t = 1, \cdots, n$ for all $\lambda \in \Theta$. Third, by Lemmas 4.3-4.4, $L_n^*(\lambda) \to_p L(\lambda)$ uniformly in $\Theta$. From Lemma 4.6, we have

$$\sup_{\lambda \in \Theta} |L_n(\lambda) - L(\lambda)| \leq \sup_{\lambda \in \Theta} |L_n^*(\lambda) - L(\lambda)| + \sup_{\lambda \in \Theta} |L_n^*(\lambda) - L_n(\lambda)| = o_p(1).$$

Fourth, Lemma 4.4 showed that $L(\lambda)$ has a unique maximum at $\lambda_0$. Thus, we have established all the conditions for consistency in Theorem 4.1.1 in Amemiya (1985). This completes the proof.

Proof of Lemma 5.1. In the proof of Lemma 4.3, we have shown that $E \sup_{\lambda \in \Theta} ||\epsilon_t||^2 < \infty$. With the same argument, it can be shown that $E \sup_{\lambda \in \Theta} |(\partial \epsilon'_t / \partial \varphi)|^2 < \infty$. Since $D_t \Gamma D_t'$ has a lower bound uniformly for all $\lambda \in \Theta$, we have $E \sup_{\lambda \in \Theta} |(\partial \epsilon'_t / \partial \varphi)(D_t \Gamma D_t')^{-1}(\partial \epsilon'_t / \partial \varphi')| < \infty$. Let $c$ be any constant vector with the same dimension as $\varphi$. If $c'E[(\partial \epsilon'_t / \partial \varphi)(D_t \Gamma D_t')^{-1}(\partial \epsilon'_t / \partial \varphi')]c = 0$, then $c'(\partial \epsilon'_t / \partial \varphi)(D_t \Gamma D_t')^{-1/2} = 0$ a.s., and hence $c'(\partial \epsilon'_t / \partial \varphi) = 0$ a.s.. By Lemma 4.2, $c = 0$. Thus $E[(\partial \epsilon'_t / \partial \varphi)(D_t \Gamma D_t')^{-1}(\partial \epsilon'_t / \partial \varphi')] > 0$. This completes the proof.

Proof of Lemma 5.2. First,

\begin{align*}
\frac{\partial h_t^c}{\partial \varphi} &= -\frac{\partial \epsilon'_t}{\partial \varphi}(D_t \Gamma D_t')^{-1}\epsilon_t - \frac{1}{2} \frac{\partial H_t^c}{\partial \varphi} D_t^{-2} \zeta_t, \quad \text{(B.25)} \\
\frac{\partial H_t^c}{\partial \varphi'} &= (I_m - \sum_{i=1}^s B_i L^i)^{-1} (\sum_{i=1}^r A_i L^i) (2 \epsilon_t^* \frac{\partial \epsilon_t}{\partial \varphi'}), \quad \text{(B.26)} \\
\frac{\partial h_t^c}{\partial \delta} &= -\frac{1}{2} \frac{\partial H_t^c}{\partial \delta} D_t^{-2} \zeta_t, \quad \text{(B.27)} \\
\frac{\partial h_t^c}{\partial \sigma} &= -\frac{1}{2} \frac{\partial \text{vec}'(\Gamma)}{\partial \sigma} \text{vec}(\Gamma^{-1} - \Gamma^{-1} D_t^{-1} \epsilon_t^* \epsilon_t D_t^{-1} \Gamma^{-1}), \quad \text{(B.28)}
\end{align*}

where $\epsilon_t^* = \text{diag}(\epsilon^*_{1t}, \cdots, \epsilon^*_{mt})$, $\zeta_t = \Pi - \eta_t^* \Gamma^{-1} \eta_t$, $\Pi = (1, \cdots, 1)'_{m \times 1}$, and $\eta_t$ and $\eta_t^*$ are defined as in Lemma 5.3. When $\lambda = \lambda_0$, $\eta_t = \eta_0$, and, in this case, we denote $\zeta_t$ and $\tilde{\eta}_t$ by $\zeta_0$ and $\tilde{\eta}_0$, respectively.

For models (2.10)-(2.11),

$$\frac{\partial H_t^c}{\partial \mu'} = -2 \sum_{i=1}^r A_i \epsilon_{t-i}^*. \quad \text{(B.29)}$$

32
Since $|e_{jt-i}^2| \leq h_{jt}^e/\alpha_{ijj}$ and $\alpha_{ijj} \geq a_{ijj}^l > 0$, $j = 1, \ldots, m$ and $i = 1, \ldots, r$, we have
\[
\left| \frac{\partial H_t^e}{\partial \mu} D_t^{-2} \right| \leq \kappa_1 \sum_{j=1}^{m} \sum_{i=1}^{r} \frac{|e_{jt-i}|}{h_{jt}^e} < \kappa_2,
\]
(B.30)
where $\kappa_1$ and $\kappa_2$ are some constants independent of $\lambda$. Furthermore, since all the terms in $\partial h_{it}/\partial \delta$ appear in $h_{jt}^e$, $\|(\partial H_t^e/\partial \delta) D_t^{-2}\| < M$, a constant independent of $\lambda$. Since $E\eta_{0it}^4 < \infty$ and $E||\zeta_{it}||^2 < \infty$, it follows that $\Omega_0 < \infty$.

For models (2.4)-(2.5), since (B.25)-(B.26), $E||\zeta_{it}||^2 < \infty$, $E||Y_t||^4 < \infty$ and $D_{0t}$ has a lower bound, we have
\[
E\left| \frac{\partial l_{0t}^e}{\partial \varphi} \right|^2 \leq 2E|\frac{\partial e_{jt}^e}{\partial \varphi}|^2 + \frac{1}{2} E\left| \frac{\partial H_t^e}{\partial \varphi} D_{0t}^{-2} \right|^2 E||\zeta_{it}||^2 < \infty.
\]
Similarly, we can show that $E\left| \frac{\partial l_{0t}^e}{\partial \sigma} \right|^2$ is finite. It is obvious that $E\left| \frac{\partial l_{0t}^e}{\partial \lambda} \right|^2 < \infty$. Thus, we also have $\Omega_0 < \infty$. In a similar manner, it can be shown that $\Omega_0 < \infty$ for models (2.8)-(2.9).

Let $S_t = \sum_{i=1}^{n} c^i \partial l_{0t}^e/\partial \lambda$, where $c$ is a constant vector with the same dimension as $\lambda$. Then $S_n$ is a martingale array with respect to $\mathcal{F}_t$. By the given assumptions, $E S_n/n = c^T E[\partial l_{0t}^e/\partial \lambda \partial l_{0t}^e/\partial \lambda^o]c > 0$. Using the central limit theorem of Stout (1974), $n^{-1/2} S_n$ converges to $N(0, c^T \Omega_0 c)$ in distribution. Finally, by the Cramér-Wold device, $n^{-1/2} \sum_{t=1}^{n} \partial l_{0t}^e/\partial \lambda$ converges to $N(0, \Omega_0)$ in distribution.

In a similar manner to the proof of Lemma 4.6, we can show that
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left| \frac{\partial l_{0t}^e}{\partial \lambda} - \frac{\partial l_{0t}}{\partial \lambda} \right| = o_p(1).
\]
Thus, $n^{-1/2} \sum_{t=1}^{n} \partial l_{0t}^e/\partial \lambda$ converges to $N(0, \Omega_0)$ in distribution. This completes the proof.

2 Proof of Lemma 5.3. For models (2.10)-(2.11), from the proof of Lemma 5.2, we have shown that
\[
\sup_{\lambda \in \Theta} \left| (\partial H_t^e/\partial \lambda) L_t^{-2} \right| < C < \infty \quad \text{with probability one},
\]
where $C$ is a nonrandom constant. Furthermore,
\[
\sup_{\lambda \in \Theta} \left| \Delta_t^e \right| \leq \kappa_1 ||\eta_t^e||^2 \leq \kappa_1 ||\varepsilon_t^e||^2 \leq \kappa_3 \varepsilon_t^{e2},
\]
where $\varepsilon_t^*$ is defined as in (B.7). Thus, $E \sup_{\lambda \in \Theta} ||(\partial H_t^\epsilon / \partial \lambda) D_t^{\epsilon-2} \Delta_t^\epsilon D_t^{\epsilon-2} (\partial H_t^\epsilon / \partial \lambda')|| < \infty$.

For models (2.8)-(2.9),

$$\frac{\partial H_t^\epsilon}{\partial \varphi} = 2 \sum_{i=1}^r A_i \varepsilon_{t-i} \frac{\partial \varepsilon_t^*}{\partial \varphi},$$

where $\varepsilon_t^*$ is defined as in (B.26). Thus, with probability one,

$$||\frac{\partial H_t^\epsilon}{\partial \varphi} D_t^{\epsilon-2}|| \leq \kappa_1 \sum_{j=1}^m \sum_{i=1}^r \frac{\varepsilon_{t-i}}{h_{jt}} ||\frac{\partial \varepsilon_t^*}{\partial \varphi}|| \leq \kappa_2 \sum_{j=1}^m \sum_{i=1}^r ||\frac{\partial \varepsilon_t^*}{\partial \varphi}||,$$

(B.31)

where $\kappa_1$ and $\kappa_2$ are constants independent of $t$. Since all the components in $\partial H_t^\epsilon / \partial \delta$ also appear in $D_t^{\epsilon-2}$, we have

$$\sup_{\lambda \in \Theta} ||\frac{\partial H_t^\epsilon}{\partial \delta} D_t^{\epsilon-2}|| < C < \infty,$$

(B.32)

where $C$ is a nonrandom constant independent of $\lambda$. By (B.31)-(B.32), it is easy to show that, if $E||Y_t||^4 < \infty$, $E \sup_{\lambda \in \Theta} ||(\partial H_t^\epsilon / \partial \lambda) D_t^{\epsilon-2} \Delta_t^\epsilon D_t^{\epsilon-2} (\partial H_t^\epsilon / \partial \lambda')|| < \infty$.

For models (2.4)-(2.5), since $E||Y_t||^6 < \infty$,

$$E \sup_{\lambda \in \Theta} ||\frac{\partial H_t^\epsilon}{\partial \lambda} D_t^{\epsilon-2} \Delta_t^\epsilon D_t^{\epsilon-2} (\partial H_t^\epsilon / \partial \lambda')|| \leq C E \sup_{\lambda \in \Theta} ||\frac{\partial H_t^\epsilon}{\partial \lambda} \frac{\partial H_t^\epsilon}{\partial \lambda'}|| < \infty,$$

where $C$ is a nonrandom constant independent of $\lambda$. This completes the proof.

Proof of Lemma 5.4. By direct differentiation of (B.25) and (B.27)-(B.28), we have

$$\frac{\partial^2 t_t^\epsilon}{\partial \lambda \partial \lambda'} = -R_t^{(1)} - R_t^{(2)} - R_t^{(3)},$$

(B.33)

where

$$R_t^{(1)} = \frac{\partial \varepsilon_t^*}{\partial \lambda} (D_t^\epsilon \Gamma D_t^\epsilon)^{-1} \frac{\partial \varepsilon_t^*}{\partial \lambda'}, \quad R_t^{(2)} = \frac{1}{4} \frac{\partial H_t^\epsilon}{\partial \lambda} D_t^{\epsilon-2} \Delta_t^\epsilon D_t^{\epsilon-2} \frac{\partial H_t^\epsilon}{\partial \lambda'},$$

$$R_t^{(3)} = (\varepsilon_t^* \otimes I_m) \frac{\partial}{\partial \lambda'} vec \left[ \frac{\partial \varepsilon_t^*}{\partial \lambda} (D_t^\epsilon \Gamma D_t^\epsilon)^{-1} \right] + (\zeta_t^* \otimes I_m) \frac{\partial}{\partial \lambda'} vec \left[ \frac{1}{2} \frac{\partial H_t^\epsilon}{\partial \lambda} D_t^{\epsilon-2} - \frac{1}{2} \frac{\partial H_t^\epsilon}{\partial \lambda} D_t^{\epsilon-2} \left[ \tilde{\eta}_t \Gamma^{-1} \Delta_t^{-1} + \tilde{\Delta}_t D_t^{-1} \right] \frac{\partial \varepsilon_t^*}{\partial \lambda} \right].$$
and $\Delta^\epsilon_t$, $\tilde{\Delta}^\epsilon_t$ and $\tilde{\eta}^\epsilon_t$ are defined as in Lemma 5.3. By Lemmas 5.1 and 5.3, we have $E\sup_{\lambda \in \Theta} R_t^{(1)} < \infty$ and $E\sup_{\lambda \in \Theta} R_t^{(2)} < \infty$. Similarly, we can show that $E\sup_{\lambda \in \Theta} R_t^{(3)} < \infty$. Thus, by (B.33), $E\sup_{\lambda \in \Theta} ||\partial^2 l^\epsilon_t / \partial \delta \partial \sigma'|| < \infty$. Furthermore,

$$\frac{\partial^2 l^\epsilon_t}{\partial \varphi \partial \sigma'} = \frac{\partial \delta}{\partial \varphi} (c')^t D^{-1}_t \Gamma^{-1} \otimes D^{-1}_t \Gamma^{-1})K - \frac{1}{2} \frac{\partial H^\epsilon_t}{\partial \varphi} D^{-2}_t \partial \zeta_t / \partial \sigma',$$

$$\frac{\partial^2 l^\epsilon_t}{\partial \delta \partial \sigma'} = -\frac{1}{2} \frac{\partial H^\epsilon_t}{\partial \delta} D^{-2}_t \partial \zeta_t / \partial \sigma',$$

$$\frac{\partial^2 l^\epsilon_t}{\partial \sigma \partial \sigma'} = K^t (I_m \otimes \Gamma^{-1})K,$$

$$\frac{\partial^2 l^\epsilon_t}{\partial \sigma \partial \sigma'} = (I_m \otimes \Gamma^{-1})[I_m - (\Gamma^{-1} D^{-1}_t c' t D^{-1}_t \otimes I_m) - (I_m \otimes \Gamma^{-1} D^{-1}_t c' t D^{-1}_m)][(I_m \otimes \Gamma^{-1})K].$$

In a similar manner, it is straightforward to show that $E\sup_{\lambda \in \Theta} ||\partial^2 l^\epsilon_t / \partial \delta \partial \sigma'|| < \infty$, $E\sup_{\lambda \in \Theta} ||\partial^2 l^\epsilon_t / \partial \delta \partial \sigma'|| < \infty$ and $E\sup_{\lambda \in \Theta} ||\partial^2 l^\epsilon_t / \partial \sigma \partial \sigma'|| < \infty$. Finally, by the triangle inequality, we can show that $E\sup_{\lambda \in \Theta} ||\partial^2 l^\epsilon_t / \partial \lambda \partial \lambda'|| < \infty$. By Theorem 3.1, (a) holds. The proof of (b) is similar to that of Lemma 4.6, and hence the details are omitted. This completes the proof.

**Proof of Lemma 5.5.** By Lemmas 5.1 and 5.3, we know $||\Sigma_0|| < \infty$. By Lemma 5.4, we have $\Sigma_n = \Sigma_0 + \sigma_P(1)$.

Let $c$ be a constant vector with the same dimension as $\delta$. If $c'[\partial H^\epsilon_{ot}/\partial \delta D^{-1}_t \partial H^\epsilon_{ot}/\partial \delta']c = 0$, then $c'[\partial H^\epsilon_{ot}/\partial \delta] D^{-2}_t = 0$ and hence $c'[\partial H^\epsilon_{ot}/\partial \delta] = 0$. By Lemma 4.2, $c = 0$. Thus, $E[\partial H^\epsilon_{ot}/\partial \delta D^{-1}_t \partial H^\epsilon_{ot}/\partial \delta'] > 0$.

Denote

$$\Sigma_{\delta_0} = E \left[ \begin{pmatrix} \frac{1}{2} \frac{\partial H^\epsilon_{ot}}{\partial \delta} D^{-2}_t & 0 \\ 0 & P' \end{pmatrix} \begin{pmatrix} C \\ C_1 \\ I_m \end{pmatrix} \right] \begin{pmatrix} 0 \\ \frac{1}{2} \frac{\partial H^\epsilon_{ot}}{\partial \sigma} D^{-2}_t \\ 0 \\ P \end{pmatrix}. $$

By the condition given, $C \geq 2I_m$. Thus, it is easy to show that $C \geq 2I_m$. Thus, it is easy to show that $C$ is positive by Theorem 14.8.5 in Harville (1997). Since $P'P = K^t (\Gamma^{-1}_0 \otimes \Gamma^{-1}_m)K$ and $E[\partial H^\epsilon_{ot}/\partial \delta D^{-1}_t \partial H^\epsilon_{ot}/\partial \delta']$ are positive, we know that $\Sigma_{\delta_0}$ is positive.

$$-\Sigma_0 = \begin{pmatrix} E[\frac{\partial^2 l^\epsilon_t}{\partial \varphi \partial \sigma'} (D_0 \Gamma_0 D_0)^{-1} \frac{\partial \delta t}{\partial \varphi} ] & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \Sigma_{\rho_0} & \Sigma_{\rho \delta \sigma_0} \\ \Sigma_{\rho \delta \sigma_0} & \Sigma_{\delta_0} \end{pmatrix},$$

\[ 35 \]
where $\Sigma_{\varphi} = E \left[ (\partial H_{0t}^\epsilon/\partial \varphi) D_{0t}^{-2} C D_{0t}^{-2} (\partial H_{0t}^\epsilon/\partial \varphi') \right] / 4$, $\Sigma_{\varphi} = (\Sigma_{\varphi}, \Sigma_{\varphi})$, $\Sigma_{\varphi} = E \left[ (\partial H_{0t}^\epsilon/\partial \varphi) D_{0t}^{-2} C D_{0t}^{-2} (\partial H_{0t}^\epsilon/\partial \varphi') \right] / 4$, $\Sigma_{\varphi} = E \left[ (\partial H_{0t}^\epsilon/\partial \varphi) D_{0t}^{-2} C D_{0t}^{-2} \right] C_1 P/2$. Let $c = (c_1, c_2)'$ be any constant vector with the same dimension as $\lambda$, and let $c_1$ have the same dimension as $\varphi$, i.e. $m + (p + q)m^2$ for models (2.4)-(2.5) and (2.8)-(2.9), and $m$ for models (2.10)-(2.11). If $-c'\Sigma_0c = 0$, then $c_1 E[(\varphi_{0t}^\epsilon/\partial \varphi)(D_{0t}^\epsilon \Gamma_0 D_{0t}^\epsilon)^{-1} (\varphi_{0t}^\epsilon/\partial \varphi')]c_1 = 0$. By Lemma 5.1, $c_1 = 0$. Thus, $c_2 \Sigma_0 c_2 = 0$. As we have shown that $\Sigma_0$ is positive definite, $c_2 = 0$. Thus, $-\Sigma_0$ is positive definite. This completes the proof.  

**Proof of Lemma 5.6.** We only present the proof for models (2.4)-(2.5). The proofs for models (2.8)-(2.9) and models (2.10)-(2.11) are similar, except that (B.29)-(B.30) are used to avoid the requirement of moments. In the following, $c_i$ and $\rho_i$ are some constants independent of $\lambda$, with $0 < \rho_i < 1$. By (B.2), we can show that

$$||\frac{\partial \varphi}{\partial \varphi}|| \leq c_2 + c_3 \sum_{i=1}^{\infty} \rho_i ||Y_{t-i}|| \equiv X_{1t}. \quad (B.34)$$

Since $X_{1t}$ is a strictly stationary time series with $EX_{1t}^2 < \infty$, we have (see Chung, 1968, p.93)

$$\frac{1}{\sqrt{n}} \max_{1 \leq t \leq n} \sup_{\lambda \in \Theta} ||\frac{\partial \varphi}{\partial \varphi}|| = o_p(1). \quad (B.35)$$

By (B.5), (B.7),(B.8) and (B.26), it follows that

$$\sup_{\lambda \in \Theta} ||\frac{\partial H_{\lambda}^\epsilon}{\partial \lambda}|| \leq c_4 + c_5 \sum_{i=1}^{\infty} \rho_i ||Y_{t-i}||^2 \equiv X_{2t}. \quad (B.36)$$

Since $X_{2t}$ is a strictly stationary time series with $EX_{2t}^2 < \infty$, we have

$$\frac{1}{\sqrt{n}} \max_{1 \leq t \leq n} \sup_{\lambda \in \Theta} ||\frac{\partial H_{\lambda}^\epsilon}{\partial \lambda}|| = o_p(1). \quad (B.37)$$

In the following, $\zeta_t$ is defined as in (B.27) and $\tilde{\eta}_t^\epsilon$ and $\eta_t^\epsilon$ are defined as in Lemma 5.3. Denote $\eta_t^\epsilon$, $\tilde{\eta}_t^\epsilon$, $\zeta_t$ and $D_t^\epsilon$ by $\eta_{nt}^\epsilon$, $\tilde{\eta}_{nt}^\epsilon$, $\zeta_{nt}$ and $D_{nt}^\epsilon$, respectively, when $\lambda = \lambda_n$.

By (B.35) and (B.37),

$$|\eta_{nt}^\epsilon - \eta_{0it}^\epsilon| \leq |\epsilon_{nt}^\epsilon - \epsilon_{0it}^\epsilon| \frac{1}{h_{nt}^{1/2}} + |h_{nt}^{1/2} - h_{0it}^{1/2}| \frac{\epsilon_{0it}^\epsilon}{h_{0it}^{1/2} h_{nt}^{1/2}}.$$
By (B.41),

where \( o_p(1) \) holds uniformly in all \( t, i = 1, \ldots, m \), and \( \lambda_1^* \) and \( \lambda_2^* \) lie in between \( \lambda_0 \) and \( \lambda_n \). From (B.38), we have

\[
|\zeta_{nt} - \zeta_{0t}| = |\tilde{\eta}_{nt}^\epsilon \Gamma_n^{-1} \eta_{nt} - \tilde{\eta}_{0t} \Gamma_0^{-1} \eta_{0t}|
\leq |\tilde{\eta}_{nt}^\epsilon|||\tilde{\eta}_{0t}|||\Gamma_n^{-1} - \Gamma_0^{-1}|
+ 2||\tilde{\eta}_{nt}^\epsilon - \tilde{\eta}_{0t}|||\Gamma_0^{-1} \eta_{0t}| + ||\tilde{\eta}_{nt}^\epsilon - \tilde{\eta}_{0t}||^2 ||\Gamma_0^{-1}|
= o_p(1) + o_p(1)||\eta_{0t}||^2,
\]

where \( o_p(1) \) holds uniformly in all \( t \). By (B.37),

\[
\max_{1 \leq i \leq n} |h_{nit}^{-1} - h_{0it}^{-1}| = |\sqrt{n}(\tilde{\lambda}_n - \tilde{\lambda}_0)| \frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} \left| \frac{1}{h_{it}^{\epsilon/2}} \frac{\partial h_{it}^\epsilon}{\partial \lambda} \right| \lambda_{3n}
= o_p(1),
\]

where \( \lambda_{3n} \) lies in between \( \lambda_0 \) and \( \lambda_n \). By (B.39)-(B.40),

\[
||D_{nt}^{\epsilon-2} \zeta_{nt} - D_{0t}^{\epsilon-2} \zeta_{0t}|| \leq ||D_{nt}^{\epsilon-2} - D_{0t}^{\epsilon-2}|| ||\zeta_{0t}|| + ||D_{nt}^{\epsilon-2}|| ||\zeta_{nt} - \zeta_{0t}||
= o_p(1) + o_p(1)||\eta_{0t}||^2.
\]

By (B.41),

\[
||D_{nt}^{\epsilon-2} \zeta_{nt} \zeta_{nt} D_{nt}^{\epsilon-2} - D_{0t}^{\epsilon-2} \zeta_{0t} \zeta_{0t} D_{0t}^{\epsilon-2}||
\leq 2||D_{nt}^{\epsilon-2} \zeta_{nt} - D_{0t}^{\epsilon-2} \zeta_{0t}|| ||D_{nt}^{\epsilon-2} \zeta_{0t}|| + ||D_{nt}^{\epsilon-2} \zeta_{nt} - D_{0t}^{\epsilon-2} \zeta_{0t}||^2
= o_p(1) + o_p(1)||\eta_{0t}||^2.
\]

In a similar manner to (B.37), we can show that

\[
\sup_{\lambda \in \Theta} \left| \frac{\partial^2 h_{it}^\epsilon}{\partial \lambda \partial \lambda} \right| \leq c_6 + c_7 \sum_{j=1}^{\infty} \rho_j^2 ||Y_{1-j}||^2 \equiv X_{3it},
\]

(B.43)
where \( i = 1, \ldots, m \). By (B.42)-(B.43), we can show that

\[
\left\| \frac{\partial H^e_{nt}}{\partial \lambda} D_{nt}^{-2} \zeta_{nt} \xi_{nt} D_{nt}^{-2} \frac{\partial H^e_{nt}}{\partial \lambda} - D_{nt}^{-2} \zeta_{nt} \xi_{nt} D_{nt}^{-2} \frac{\partial H^e_{nt}}{\partial \lambda} \right\| \\
\leq \left\| \frac{\partial H^e_{nt}}{\partial \lambda} D_{nt}^{-2} \zeta_{nt} \xi_{nt} D_{nt}^{-2} \frac{\partial H^e_{nt}}{\partial \lambda} \right\| \\
+ \left\| \frac{\partial H^e_{nt}}{\partial \lambda} \right\|^2 [o_p(1) + o_p(1)||\eta_0||^4] \\
\leq 2 \left\| \frac{\partial H^e_{nt}}{\partial \lambda} \right\| |\eta_0||\zeta_{nt}|^2 \\
+ \frac{1}{\sqrt{n}} O_p(1)(\sqrt{n}(\lambda_n - \lambda_0) \sum_{i=1}^{n} X_{3it}|\frac{\partial H^e_{nt}}{\partial \lambda}||\zeta_{nt}|^2 \\
+ \frac{1}{\sqrt{n}} |\eta_0||\zeta_{nt}|^2 |\sum_{i=1}^{n} X_{3it}|^2 + X_{2t}(1 + ||\eta_0||^4) o_p(1) \\
= o_p(1) \sum_{i=1}^{n} X_{3it}|\frac{\partial H^e_{nt}}{\partial \lambda}|| + \left( \sum_{i=1}^{n} X_{3it}^2 + X_{2t}^2 \right) (1 + ||\eta_0||^4) \\
\equiv o_p(1) X^*_t (1 + ||\eta_0||^4), \\
\tag{B.44}
\]

where \( O_p(1) \) and \( o_p(1) \) hold uniformly in all \( t \). Note that \( X^*_t (1 + ||\eta_0||^4) \) is strictly stationary, with \( E[X^*_t (1 + ||\eta_0||^4)] = EX^*_t E(1 + ||\eta_0||^4) < \infty \). By the ergodic theorem, we have \( n^{-1} \sum_{t=1}^{n} X^*_t (1 + ||\eta_0||^4) = O_p(1) \). Thus, by (B.44), we have

\[
\frac{1}{n} \sum_{t=1}^{n} \left\| \frac{\partial H^e_{nt}}{\partial \lambda} D_{nt}^{-2} \zeta_{nt} \xi_{nt} D_{nt}^{-2} \frac{\partial H^e_{nt}}{\partial \lambda} \right\| = o_p(1). \\
\tag{B.45}
\]

Similarly, we can show that

\[
\frac{1}{n} \sum_{t=1}^{n} \left\| \frac{\partial \epsilon^e_{nt}}{\partial \varphi} (D_{nt}^e \Gamma D_{nt}^e)^{-1} \epsilon^e_{nt} \xi^e_{nt} D_{nt}^e \Gamma D_{nt}^e)^{-1} \frac{\partial \epsilon^e_{nt}}{\partial \varphi} \\
- \frac{\partial \epsilon^e_{nt}}{\partial \varphi} (D_{nt}^e \Gamma D_{nt}^e)^{-1} \epsilon^e_{nt} \xi^e_{nt} (D_{nt}^e \Gamma D_{nt}^e)^{-1} \frac{\partial \epsilon^e_{nt}}{\partial \varphi} \right\| = o_p(1), \\
\tag{B.46}
\]

and

\[
\frac{1}{n} \sum_{t=1}^{n} \left\| \frac{\partial l^e_{nt}}{\partial \sigma^t} \frac{\partial l^e_{nt}}{\partial \sigma^t} \right\| = o_p(1). \\
\tag{B.47}
\]

Thus, by (B.45)-(B.47) and the triangle inequality, we can show that

\[
\frac{1}{n} \sum_{t=1}^{n} \left\| \frac{\partial l^e_{nt}}{\partial \lambda} \frac{\partial l^e_{nt}}{\partial \lambda} \right\| = o_p(1). \\
\tag{B.48}
\]
Thus, (a) holds. In a similar manner to the proof of Lemma 4.6, we can show that

\[ \frac{1}{n} \sum_{t=1}^{n} \left\| \frac{\partial l_{nt}}{\partial \lambda} \frac{\partial l_{nt}}{\partial \lambda'} - \frac{\partial l_{nt}}{\partial \lambda} \frac{\partial l_{nt}}{\partial \lambda'} \right\| = o_p(1). \]  

(B.49)

Note that \((\partial l_{0t}/\partial \lambda)(\partial l_{0t}/\partial \lambda')\) is strictly stationary and ergodic with \(E||\partial l_{0t}/\partial \lambda(\partial l_{0t}/\partial \lambda')|| < \infty\). By the ergodic theorem, we have \(n^{-1} \sum_{t=1}^{n} ||(\partial l_{0t}/\partial \lambda)(\partial l_{0t}/\partial \lambda')|| = \Omega_0 + o_p(1)\). Furthermore, by (B.48)-(B.49), (b) holds. This completes the proof.

2

Proof of Theorem 5.1. We need only to verify the conditions of Theorem 4.1.3 in Amemiya (1985). First, by Theorem 4.1, the QMLE \( \hat{\lambda}_n \) of \( \lambda_0 \) is consistent. Second, \( n^{-1} \sum_{t=1}^{n} (\partial l^2_t/\partial \lambda \partial \lambda') \) exists and is continuous in \( \Theta \). Third, by Lemmas 5.4-5.5, we can immediately obtain that \( n^{-1} \sum_{t=1}^{n} (\partial l^2_{nt}/\partial \lambda \partial \lambda') \) converges to \( \Sigma_0 > 0 \) for any sequence \( \lambda_n \), such that \( \lambda_n \to \lambda_0 \) in probability. Fourth, by Lemma 5.2, \( n^{-1/2} \sum_{t=1}^{n} (\partial l_{0t}/\partial \lambda) \) converges to \( N(0, \Omega_0) \) in distribution. Thus, we have established all the conditions in Theorem 4.1.3 in Amemiya (1985), and hence \( n^{1/2}(\hat{\lambda}_n - \lambda_0) \) converges to \( N(0, \Sigma_0^{-1} \Omega_0 \Sigma_0^{-1}) \). Finally, by Lemmas 5.5-5.6, \( \Sigma_0 \) and \( \Omega_0 \) can be estimated consistently by \( \hat{\Sigma}_n \) and \( \hat{\Omega}_n \), respectively. This completes the proof. 2