# ASYMPTOTIC THEORY FOR A VECTOR ARMA-GARCH MODEL 

Shioing Ling<br>Hong Kong University of Science and Technology<br>Michael McAleer<br>University of Western Australia


#### Abstract

This paper investigates the asymptotic theory for a vector autoregressive moving average-generalized autoregressive conditional heteroskedasticity (ARMAGARCH) model. The conditions for the strict stationarity, the 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 autoregressive conditional heteroskedasticity (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 and also 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 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 (2002). Of these models, the most popular is undoubtedly the generalized autoregressive conditional heteroskedasticity (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),

[^0]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 to report of empirical findings.

In this paper, we propose a vector autoregressive moving average-GARCH (ARMA-GARCH) model that 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) proves 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 (1996), and Ling and Li (1997). Weiss (1986) and Ling and Li (1997) use the conditions of Basawa, Feign, and Heyde (1976), whereby their consis- Q2 tency results rely on the assumption that the fourth-order moments exist. Lee and Hansen (1994) and Lumsdaine (1996) use the conditions of Amemiya (1985, pp. 106-111), but their methods are only valid for the simple $\operatorname{GARCH}(1,1)$ model and cannot be extended to more general cases. Moreover, the conditional errors, that is, $\eta_{0 t}$ 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, p. 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 $\operatorname{GARCH}(1,1)$ 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 vector 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 $\operatorname{GARCH}(1,1)$ model in Lee and Hansen (1994) and Lumsdaine (1996) are quite weak. However, their $\operatorname{GARCH}(1,1)$ model explicitly excludes the special case of the $\mathrm{ARCH}(1)$ model because they assume that $B_{1} \neq 0$ (see equation (2.7) in Section 2) 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 $\eta_{0 t}$ are independently and identically distributed (i.i.d) instead of a series of strictly stationary and ergodic martingale differences. 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 ARMAGARCH model and investigates its properties. Section 3 presents the quasilikelihood 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 Appendixes A and B.

Throughout this paper, we use the following notation. The term $|\cdot|$ denotes the absolute value of a univariate variable or the determinant of a matrix; $\|\cdot\|$ denotes the Euclidean norm of a matrix or vector; $A^{\prime}$ denotes the transpose of the matrix or vector $A ; O(1)$ (or $o(1))$ denotes a series of nonstochastic variables that are bounded (or converge to zero); $O_{p}(1)$ (or $\left.o_{p}(1)\right)$ denotes a series of random variables that are bounded (or converge to zero) in probability; $\rightarrow_{p}$ ( or $\rightarrow_{\mathcal{L}}$ ) 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) presents an $m$-dimensional multivariate conditional covariance model, namely,
$Y_{t}=E\left(Y_{t} \mid \mathcal{F}_{t-1}\right)+\varepsilon_{0 t}, \quad \operatorname{Var}\left(\varepsilon_{0 t} \mid \mathcal{F}_{t-1}\right)=D_{0 t} \Gamma_{0} D_{0 t}$,
where $\mathcal{F}_{t}$ is the past information available up to time $t, D_{0 t}=$ $\operatorname{diag}\left(h_{01 t}^{1 / 2}, \ldots, h_{0 m t}^{1 / 2}\right)$, and
$\Gamma_{0}=\left(\begin{array}{cccc}1 & \sigma_{012} & \ldots & \sigma_{01 m} \\ \sigma_{021} & 1 & \sigma_{023} & \cdots \\ & & \ldots & \\ \sigma_{0 m, 1} & \ldots & \sigma_{0 m, m-1} & 1\end{array}\right)$,
in which $\sigma_{0 i j}=\sigma_{0 j i}$. The main feature of this model is that the conditional correlation $E\left(\varepsilon_{0 i t} \varepsilon_{0 j t} \mid \mathcal{F}_{t-1}\right) / \sqrt{E\left(\varepsilon_{0 i t}^{2} \mid \mathcal{F}_{t-1}\right) E\left(\varepsilon_{0 j t}^{2} \mid \mathcal{F}_{t-1}\right)}=\sigma_{0 i j}$ is constant over time, where $i \neq j$ and $\varepsilon_{0 i t}$ is the $i$ th element of $\varepsilon_{0 t}$. By assuming
$h_{0 i t}=w_{0 i}+\sum_{j=1}^{r} a_{0 i j} \varepsilon_{0 i t-j}^{2}+\sum_{j=1}^{s} b_{0 i j} h_{0 i t-j}, \quad i=1, \ldots, m$,
that is, with only $i$-specific effects, Bollerslev (1990) models the exchange rates of the German mark, French franc, and British pound against the U.S. dollar. His results provide evidence that the assumption of constant correlations is adequate. Tse (2000) has developed the Lagrange multiplier test for the hypothesis of constant correlation in Bollerslev's model and provides evidence that the hypothesis is adequate for spot and futures prices and for foreign exchange rates.
It is possible to provide a straightforward explanation for the hypothesis of constant correlation. Suppose that $h_{0 i t}$ captures completely the past information, with $E h_{0 i t}=E \varepsilon_{0 i t}^{2}$. Then $\eta_{0 i t}=\varepsilon_{0 i t} h_{0 i t}^{-1 / 2}$ will be independent of the past information. Thus, for each $i,\left\{\eta_{0 i t}, t=0, \pm 1, \pm 2, \ldots\right\}$ will be a sequence of i.i.d. random variables, with zero mean and variance one. In general, $\eta_{0 i t}$ and $\eta_{0 j t}$ are correlated for $i \neq j$, and hence it is natural to assume that $\eta_{0 t}=$ $\left(\eta_{01 t}, \ldots, \eta_{0 m t}\right)^{\prime}$ is a sequence of i.i.d. random vectors, with zero mean and covariance $\Gamma_{0}$. Thus, we can write
$\varepsilon_{0 t}=D_{0 t} \eta_{0 t}$.
Obviously, $\varepsilon_{0 t}$ in (2.1) has the same conditional covariance matrix as that in (2.3).
Now, the remaining problem is how to define $h_{0 i t}$ so that it can capture completely the past information. It is obvious that $h_{0 i t}$ may have as many different forms as in the univariate case. In the multivariate case, $h_{0 i t}$ should contain some past information, not only from $\varepsilon_{0 i t}$ but also from $\varepsilon_{0 j t}$. 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 $\varepsilon_{0 i t}$, $\varepsilon_{0 j t}, h_{0 i t}$, and $h_{0 j t}$ in the model. Note that $D_{0 t}$ depends only on $\left(h_{01 t}, \ldots, h_{0 m t}\right)^{\prime}$, denoted by $H_{0 t}$. It is natural to define $H_{0 t}$ in the form of (2.5), which follows, 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:

$$
\begin{align*}
\Phi_{0}(L)\left(Y_{t}-\mu_{0}\right) & =\Psi_{0}(L) \varepsilon_{0 t},  \tag{2.4}\\
\varepsilon_{0 t} & =D_{0 t} \eta_{0 t}, \quad H_{0 t}=W_{0}+\sum_{i=1}^{r} A_{0 i} \vec{\varepsilon}_{0 t-i}+\sum_{i=1}^{s} B_{0 i} H_{0 t-i}, \tag{2.5}
\end{align*}
$$

where $D_{0 t}$ and $\eta_{0 t}$ are defined as in (2.3), $\Phi_{0}(L)=I_{m}-\Phi_{01} L-\cdots-\Phi_{0 p} L^{p}$ and $\Psi_{0}(L)=I_{m}+\Psi_{01} L+\cdots+\Psi_{0 q} L^{q}$ are polynomials in $L, I_{k}$ is the $k \times k$
identity matrix, and $\vec{\varepsilon}_{0 t}=\left(\varepsilon_{01 t}^{2}, \ldots, \varepsilon_{0 m t}^{2}\right)^{\prime}$. The true parameter vector is denoted by $\lambda_{0}=\left(\varphi_{0}^{\prime}, \delta_{0}^{\prime}, \sigma_{0}^{\prime}\right)^{\prime}$, where $\varphi_{0}=\operatorname{vec}\left(\mu_{0}, \Phi_{01}, \ldots, \Phi_{0 p}, \Psi_{01}, \ldots, \Psi_{0 q}\right)$, $\delta_{0}=\operatorname{vec}\left(W_{0}, A_{01}, \ldots, A_{0 r}, B_{01}, \ldots, B_{0 s}\right)$, and $\sigma_{0}=\left(\sigma_{021}, \ldots, \sigma_{0 m, 1}, \sigma_{032}, \ldots\right.$, $\left.\sigma_{0 m, 2}, \ldots, \sigma_{0 m, m-1}\right)^{\prime}$. 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 $\lambda=\left(\varphi^{\prime}, \delta^{\prime}, \sigma^{\prime}\right)^{\prime}$, with $\varphi, \delta$, and $\sigma$ defined in a similar manner to $\varphi_{0}, \delta_{0}$, and $\sigma_{0}$, respectively, is

$$
\begin{align*}
\Phi(L)\left(Y_{t}-\mu\right) & =\Psi(L) \varepsilon_{t},  \tag{2.6}\\
H_{t} & =W+\sum_{i=1}^{r} A_{i} \vec{\varepsilon}_{t-i}+\sum_{i=1}^{s} B_{i} H_{t-i}, \tag{2.7}
\end{align*}
$$

where $H_{t}=\left(h_{1 t}, \ldots, h_{m t}\right)^{\prime}, \vec{\varepsilon}_{t}=\left(\varepsilon_{1 t}^{2}, \ldots, \varepsilon_{m t}^{2}\right)^{\prime}$, 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 $Y_{1}, \ldots, Y_{n}$, from (2.6), with initial value $\bar{Y}_{0}=$ $\left(Y_{0}, \ldots, Y_{1-p}, \varepsilon_{0}, \ldots, \varepsilon_{1-q}\right)$. Then $H_{t}$ can be calculated from (2.7), with initial values $\bar{\varepsilon}_{0}=\left(\vec{\varepsilon}_{0}, \ldots, \vec{\varepsilon}_{1-r}, H_{0}, \ldots, H_{1-s}\right)$. 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 $\lambda \in \Theta$, we make the following assumptions.

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

Assumption 2. The terms $\Phi(L)$ 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. The term $\Gamma$ 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, $i=1, \ldots, r$, $j=1, \ldots, s$; each element of $W$ has positive lower and upper bounds over $\Theta$; and all the roots of $\left|I_{m}-\sum_{i=1}^{r} A_{i} L^{i}-\sum_{i=1}^{s} B_{i} L^{i}\right|=0$ are outside the unit circle.

Assumption 4. The expressions $I_{m}-\sum_{r=1}^{r} A_{i} L^{i}$ 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, Ch. 7), under which the results in this Q3 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:

$$
\begin{align*}
\Phi_{0}(L)\left(Y_{t}-\mu_{0}\right) & =\Psi_{0}(L) \varepsilon_{0 t},  \tag{2.8}\\
\varepsilon_{0 t} & =D_{0 t} \eta_{0 t}, \quad H_{0 t}=W_{0}+\sum_{i=1}^{r} A_{0 i} \vec{\varepsilon}_{0 t-i} \tag{2.9}
\end{align*}
$$

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:

$$
\begin{align*}
Y_{t} & =\mu_{0}+\varepsilon_{0 t},  \tag{2.10}\\
\varepsilon_{0 t} & =D_{0 t} \eta_{0 t}, \quad H_{0 t}=W_{0}+\sum_{i=1}^{r} A_{0 i} \vec{\varepsilon}_{0 t-i} . \tag{2.11}
\end{align*}
$$

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

The unknown parameter ARCH and ARMA-ARCH models are defined similarly to models (2.6) and (2.7). The true parameter $\lambda_{0}=\left(\varphi_{0}^{\prime}, \delta_{0}^{\prime}, \sigma_{0}^{\prime}\right)^{\prime}$, with $\delta_{0}=\operatorname{vec}\left(W_{0}, A_{01}, \ldots, A_{0 r}\right), \sigma_{0}$ being defined as in models (2.4) and (2.5), and $\varphi_{0}$ being defined as in models (2.4) and (2.5) for models (2.8) and (2.9), and $\varphi_{0}=\mu_{0}$ for models (2.10) and (2.11). Similarly, define the unknown parameter $\lambda$ and the parametric space $\Theta$, with $0<a_{i j k}^{l} \leq a_{i j k} \leq a_{i j k}^{u}<\infty$, where $a_{i j k}$ is the $(j, k)$ th component of $A_{i}, a_{i j k}^{l}$ and $a_{i j k}^{u}$ are independent of $\lambda$, $i=1, \ldots, r$, and $j, k=1, \ldots, m .{ }^{1}$

The following theorem gives some basic properties of models (2.4) and (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) (see also Ling and McAleer, 2002a, 2002b). 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) and (2.9) and models (2.10) and (2.11). However, for these two special cases, the matrix $\tilde{A}_{0 t}$, which follows, can simply be replaced by its $(1,1)$ block.

THEOREM 2.1. Under Assumptions 1 and 3, models (2.4) and (2.5) possess an $\mathcal{F}_{t}$-measurable second-order stationary solution $\left\{Y_{t}, \varepsilon_{0 t}, H_{0 t}\right\}$, which is unique, given the $\eta_{0 t}$, where $\mathcal{F}_{t}$ is a $\sigma$-field generated by $\left\{\eta_{0 k}: k \leq t\right\}$. The solutions $\left\{Y_{t}\right\}$ and $\left\{H_{0 t}\right\}$ have the following causal representations:

$$
\begin{equation*}
Y_{t}=\sum_{k=0}^{\infty} \Upsilon_{0 k} \varepsilon_{0 t-k}, \quad \text { a.s. } \tag{2.12}
\end{equation*}
$$

$H_{0 t}=W_{0}+\sum_{j=1}^{\infty} c^{\prime}\left(\prod_{i=1}^{j} \tilde{A}_{0 t-i}\right) \xi_{t-j-1}, \quad$ a.s.,
where $\Phi_{0}^{-1}(L) \Psi_{0}(L)=\sum_{k=0}^{\infty} \Upsilon_{0 k} L^{k}, \xi_{t}=\left[\left(\tilde{\eta}_{0 t} W_{0}\right)^{\prime}, 0, \ldots, 0, W_{0}^{\prime}, 0, \ldots, 0\right]_{(r+s) m \times 1}^{\prime}$, that is, the subvector consisting of the first $m$ components is $\tilde{\eta}_{0 t} W_{0}$ and the subvector consisting of the $(r m+1)$ th to $(r+1) m t h$ components is $W_{0} ; \tilde{\eta}_{0 t}=$ $\operatorname{diag}\left(\eta_{01 t}^{2}, \ldots, \eta_{0 m t}^{2}\right), c^{\prime}=\left(0, \ldots, 0, I_{m}, 0, \ldots, 0\right)_{m \times(r+s) m}$ with the subvector consisting of the $(r m+1)$ th to $(r+1) m$ th columns being $I_{m}$; and
$\tilde{A}_{0 t}=\left(\begin{array}{ccc|ccc}\tilde{\eta}_{0 t} A_{01} & \ldots & \tilde{\eta}_{0 t} A_{0 r} & \tilde{\eta}_{0 t} B_{01} & \ldots & \tilde{\eta}_{0 t} B_{0 s} \\ & I_{m(r-1)} & O_{m(r-1) \times m} & & O_{m(r-1) \times m s} & \\ \hline A_{01} & \ldots & A_{0 r} & B_{01} & \ldots & B_{0 s} \\ & O_{m(s-1) \times m r} & & & I_{m(s-1)} & O_{m(s-1) \times m}\end{array}\right)$.
Hence, $\left\{Y_{t}, \varepsilon_{0 t}, H_{0 t}\right\}$ are strictly stationary and ergodic.
THEOREM 2.2. Suppose that the assumptions of Theorem 2.1 hold. If $\rho\left[E\left(\tilde{A}_{0 t}^{\otimes k}\right)\right]<1$, with $k$ being a strictly positive integer, then the $2 k t h$ moments of $\left\{Y_{t}, \varepsilon_{0 t}\right\}$ are finite, where $\tilde{A}_{0 t}$ is defined as in Theorem 2.1 and $A^{\otimes k}$ is the Krönecker product of the $k$ matrices $A$.

## 3. QUASI-MAXIMUM-LIKELIHOOD ESTIMATOR

The estimators of the parameters in models (2.4) and (2.5) are obtained by maximizing, conditional on ( $\bar{Y}_{0}, \bar{\varepsilon}_{0}$ ),
$L_{n}(\lambda)=\frac{1}{n} \sum_{t=1}^{n} l_{t}(\lambda), \quad l_{t}(\lambda)=-\frac{1}{2} \ln \left|D_{t} \Gamma D_{t}\right|-\frac{1}{2} \varepsilon_{t}^{\prime}\left(D_{t} \Gamma D_{t}\right)^{-1} \varepsilon_{t}$,
where $L_{n}(\lambda)$ takes the form of the Gaussian $\log$-likelihood and $D_{t}=\operatorname{diag}\left(h_{1 t}^{1 / 2}\right.$, $\left.\ldots, h_{m t}^{1 / 2}\right)$. Because we do not assume that $\eta_{0 t}$ is normal, the estimators from (3.1) are the QMLEs. 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 that is not conditional on the true $\left(\bar{Y}_{0}, \bar{\varepsilon}_{0}\right)$ and, in practice, we work with this likelihood function.

For convenience, we introduce the unobserved process $\left\{\left(\varepsilon_{t}^{\epsilon}, H_{t}^{\epsilon}\right): t=0, \pm 1\right.$, $\pm 2, \ldots\}$, which satisfies

$$
\begin{align*}
\Phi(L)\left(Y_{t}-\mu\right) & =\Psi(L) \varepsilon_{t}^{\epsilon},  \tag{3.2}\\
H_{t}^{\epsilon} & =W+\sum_{i=1}^{r} A_{i} \vec{\varepsilon}_{t-i}^{\epsilon}+\sum_{i=1}^{s} B_{i} H_{t-i}^{\epsilon}, \tag{3.3}
\end{align*}
$$

where $\vec{\varepsilon}_{t}^{\epsilon}=\left(\varepsilon_{1 t}^{\epsilon 2}, \ldots, \varepsilon_{m t}^{\epsilon 2}\right)^{\prime}$ and $H_{t}^{\epsilon}=\left(h_{1 t}^{\epsilon}, \ldots, h_{m t}^{\epsilon}\right)^{\prime}$. Denote $\underline{Y}_{0}=\left(Y_{0}, Y_{-1}, \ldots\right)$. The unobserved log-likelihood function conditional on $\underline{Y}_{0}$ is
$L_{n}^{\epsilon}(\lambda)=\frac{1}{n} \sum_{t=1}^{n} l_{t}^{\epsilon}(\lambda), \quad l_{t}^{\epsilon}(\lambda)=-\frac{1}{2} \ln \left|D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right|-\frac{1}{2} \varepsilon_{t}^{\epsilon^{\prime}}\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1} \varepsilon_{t}^{\epsilon}$,
where $D_{t}^{\epsilon}=\operatorname{diag}\left(h_{1 t}^{\epsilon}, \ldots, h_{m t}^{\epsilon}\right)$. When $\lambda=\lambda_{0}$, we have $\varepsilon_{t}^{\epsilon}=\varepsilon_{0 t}, H_{t}^{\epsilon}=H_{0 t}$, and $D_{t}^{\epsilon}=D_{0 t}$. The primary difference in the likelihoods (3.1) and (3.4) is that (3.1) is conditional on any initial values, whereas (3.4) is conditional on the infinite past observations. In practice, the use of (3.4) is not possible. Jeantheau (1998) investigates the likelihood (3.4) for models (2.4) and (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 proves the consistency of the QMLE for a special case of models (2.4) and (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 $L_{n}^{\varepsilon}(\lambda)$ for models (2.8) and (2.9) and models (2.10) and (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, and also 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. ${ }^{2}$ Let $g(y, \theta)$ be a measurable function of $y$ in Euclidean space for each $\theta \in \Theta$, a compact subset of $R^{m}$ (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 $\operatorname{Eg}\left(y_{t}, \theta\right)=0$ and $E \sup _{\theta \in \Theta}\left|g\left(y_{t}, \theta\right)\right|<\infty$. Then $\sup _{\theta \in \Theta}\left|n^{-1} \sum_{t=1}^{n} g\left(y_{t}, \theta\right)\right|=o_{p}(1)$.

## 4. CONSISTENCY OF THE OMLE

In (3.4), $D_{t}^{\epsilon}$ 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 $\mid I_{m}-\sum_{i=1}^{r} A_{i} L^{i}-$ $\sum_{i=1}^{s} B_{i} L^{i} \mid=0$ lying outside the unit circle does not ensure that all the roots of $\left|I_{m}-\sum_{i=1}^{s} B_{i} L^{i}\right|=0$ are outside the unit circle. However, because 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 $\left|I_{m}-\sum_{i=1}^{s} B_{i} L^{i}\right|=0$ 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 ARMAGARCH, 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) and (2.5) satisfying Assumptions 1-4, or models (2.8) and (2.9) satisfying Assumptions 1-3, or models (2.10) and (2.11) satisfying Assumption 3. Let $c_{\varphi}$ and $c$ be constant vectors, with the same dimensions as $\varphi$ and $\delta$, respectively. Then $c_{\varphi}^{\prime}\left(\partial \varepsilon_{t}^{\epsilon^{\prime}} / \partial \varphi\right)=0$ a.s. only if $c_{\varphi}=0$, and $c^{\prime}\left(\partial H_{t}^{\epsilon^{\prime}} / \partial \delta\right)=0$ a.s. only if $c=0$.

LEMMA 4.3. Define $L(\lambda)=E\left[l_{t}^{\epsilon}(\lambda)\right]$. Under the assumptions of Lemma 4.2, $L(\lambda)$ exists for all $\lambda \in \Theta$ and $\sup _{\lambda \in \Theta}\left|L_{n}^{\epsilon}(\lambda)-L(\lambda)\right|=o_{p}(1)$.

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 $E\left|X_{t}\right|<\infty$, and $\xi_{t}$ be a sequence of random variables such that
$\sup _{1 \leq t \leq n}\left|\xi_{t}\right| \leq C \quad$ and $\quad n^{-1} \sum_{t=1}^{n}\left|\xi_{t}\right|=o_{p}(1)$.
Then $n^{-1} \sum_{t=1}^{n} X_{t} \xi_{t}=o_{p}(1)$.
LEMMA 4.6. Under the assumptions of Lemma 4.2, $\sup _{\lambda \in \Theta} \mid L_{n}^{\epsilon}(\lambda)-$ $L_{n}(\lambda) \mid=o_{p}(1)$.

Based on the preceding 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 OMLE

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) and (2.5), the asymptotic normality of the QMLE would require the existence of the sixth moment. However, for models (2.8) and (2.9) or models (2.10) and (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) and (2.5) satisfying Assumptions 1-4, or models (2.8) and (2.9) satisfying Assumptions 1-3, or models (2.10) and (2.11) satisfying Assumption 3. Then, it follows that

$$
\begin{equation*}
E \sup _{\lambda \in \Theta}\left\|\frac{\partial \varepsilon_{t}^{\epsilon^{\prime}}}{\partial \varphi}\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1} \frac{\partial \varepsilon_{t}^{\epsilon}}{\partial \varphi^{\prime}}\right\|<\infty \quad \text { and } \quad E\left[\frac{\partial \varepsilon_{t}^{\epsilon^{\prime}}}{\partial \varphi}\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1} \frac{\partial \varepsilon_{t}^{\epsilon}}{\partial \varphi^{\prime}}\right]>0 \tag{5.1}
\end{equation*}
$$

where a matrix $A>0$ means that $A$ is positive definite.
LEMMA 5.2. Suppose that $Y_{t}$ is generated by models (2.4) and (2.5) satisfying Assumptions $1-4$ and $E\left\|Y_{t}\right\|^{4}<\infty$, or models (2.8) and (2.9) satisfying Assumptions 1-3 and $E\left\|Y_{t}\right\|^{4}<\infty$, or models (2.10) and (2.11) satisfying Assumption 3 and $E\left\|\eta_{0 t}\right\|^{4}<\infty$. Then $\Omega_{0}=E\left[\left(\partial l_{0 t}^{\epsilon} / \partial \lambda\right)\left(\partial l_{0 t}^{\epsilon} / \partial \lambda^{\prime}\right)\right]$ is finite. Furthermore, if $\Omega_{0}>0$, then
$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_{0 t}}{\partial \lambda} \rightarrow_{\mathcal{L}} N\left(0, \Omega_{0}\right)$,
where $\partial l_{0 t}^{\varepsilon} / \partial \lambda=\partial l_{t}^{\varepsilon} /\left.\partial \lambda\right|_{\lambda_{0}}$ and $\partial l_{0 t} / \partial \lambda=\partial l_{t} /\left.\partial \lambda\right|_{\lambda_{0}}$.
LEMMA 5.3. Suppose that $Y_{t}$ is generated by models (2.4) and (2.5) satisfying Assumptions 1-4 and $E\left\|Y_{t}\right\|^{6}<\infty$, or models (2.8) and (2.9) satisfying Assumptions 1-3 and $E\left\|Y_{t}\right\|^{4}<\infty$, or models (2.10) and (2.11) satisfying Assumption 3. Then,
$E \sup _{\lambda \in \Theta}\left\|\frac{\partial H_{t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}} D_{t}^{\epsilon-2} \Delta_{t}^{\epsilon} D_{t}^{\epsilon-2} \frac{\partial H_{t}^{\epsilon}}{\partial \tilde{\lambda}^{\prime}}\right\|<\infty$,
where $\tilde{\lambda}=\left(\varphi^{\prime}, \delta^{\prime}\right)^{\prime}, \Delta_{t}^{\epsilon}=\tilde{\eta}_{t}^{\epsilon} \Gamma^{-1} \tilde{\eta}_{t}^{\epsilon}+\tilde{\Delta}_{t}^{\epsilon} \tilde{\eta}_{t}^{\epsilon}, \tilde{\Delta}_{t}^{\epsilon}=\operatorname{diag}\left(e_{1} \Gamma^{-1} \eta_{t}^{\epsilon}, \ldots, e_{m} \Gamma^{-1} \eta_{t}^{\epsilon}\right)$, $e_{i}=(0, \ldots, 0,1,0, \ldots, 0)^{\prime}$ of which the ith element is $1, \eta_{t}^{\epsilon}=\left(\eta_{1 t}^{\epsilon}, \ldots, \eta_{m t}^{\epsilon}\right)^{\prime}$, and $\tilde{\eta}_{t}^{\epsilon}=\operatorname{diag}\left(\eta_{1 t}^{\epsilon}, \ldots, \eta_{m t}^{\epsilon}\right)$ with $\eta_{i t}^{\epsilon}=\varepsilon_{i t}^{\epsilon} / h_{i t}^{\epsilon 1 / 2}, i=1, \ldots, m$.

LEMMA 5.4. Under the assumptions of Lemma 5.3,
(a) $\sup _{\lambda \in \Theta}\left\|\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^{2} l_{t}^{\epsilon}}{\partial \lambda \partial \lambda^{\prime}}-E\left[\frac{\partial^{2} l_{t}^{\epsilon}}{\partial \lambda \partial \lambda^{\prime}}\right]\right\|=o_{p}(1)$,
(b) $\sup _{\lambda \in \Theta}\left\|\frac{1}{n} \sum_{t=1}^{n}\left[\frac{\partial^{2} l_{t}^{\epsilon}}{\partial \lambda \partial \lambda^{\prime}}-\frac{\partial^{2} l_{t}}{\partial \lambda \partial \lambda^{\prime}}\right]\right\|=o_{p}(1)$.

By straightforward calculation, we can show that
$\Sigma_{0} \equiv E\left[\frac{\partial^{2} l_{t}^{\epsilon}}{\partial \lambda \partial \lambda^{\prime}}\right]_{\lambda_{0}}=-\left(\begin{array}{cc}\Sigma_{\tilde{\lambda} 0} & \Sigma_{\tilde{\lambda} \sigma 0} \\ \Sigma_{\tilde{\lambda} \sigma 0}^{\prime} & \frac{1}{2} P^{\prime} P\end{array}\right)$,
where $\Sigma_{\tilde{\lambda} 0}=E\left[\left(\partial \varepsilon_{0 t}^{\epsilon^{\prime}} / \partial \tilde{\lambda}\right)\left(D_{0 t} \Gamma_{0} D_{0 t}\right)^{-1}\left(\partial \varepsilon_{0 t}^{\epsilon} / \partial \tilde{\lambda}^{\prime}\right)\right]+E\left[\left(\partial H_{0 t}^{\epsilon^{\prime}} / \partial \tilde{\lambda}\right) D_{0 t}^{-2} \times\right.$ $\left.C D_{0 t}^{-2}\left(\partial H_{0 t}^{\epsilon} / \partial \tilde{\lambda}^{\prime}\right)\right] / 4, \Sigma_{\tilde{\lambda} \sigma 0}=E\left[\left(\partial H_{0 t}^{\epsilon^{\prime}} / \partial \tilde{\lambda}\right) D_{0 t}^{-2}\right] C_{1} P / 2, \partial \varepsilon_{0 t}^{\epsilon} / \partial \tilde{\lambda}^{\prime}=\partial \varepsilon_{t}^{\epsilon} /\left.\partial \tilde{\lambda}^{\prime}\right|_{\lambda_{0}}$, $\partial H_{0 t}^{\epsilon} / \partial \tilde{\lambda}^{\prime}=\partial H_{t}^{\epsilon} /\left.\partial \tilde{\lambda}^{\prime}\right|_{\lambda_{0}}, P=\left(I_{m} \otimes \Gamma_{0}^{-1}\right) \mathcal{K}, C_{1}=\left(C_{11}, \ldots, C_{1 m}\right), C_{1 i}$ is an $m \times m$ matrix with the $(i, i)$ th component being 1 and the other components zero, $\mathcal{K}=\partial \operatorname{vec}(\Gamma) / \partial \sigma^{\prime}$ is a constant matrix, and $C=\Gamma_{0}^{-1} \odot \Gamma_{0}+I_{m}$, where
$A \odot B=\left(a_{i j} b_{i j}\right)$ for two matrices $A=\left(a_{i j}\right)$ and $B=\left(b_{i j}\right)$. In practice, $\Sigma_{0}$ is evaluated by
$\hat{\Sigma}_{n}=-\left(\begin{array}{cc}\hat{\Sigma}_{\tilde{\lambda}} & \hat{\Sigma}_{\tilde{\lambda} \sigma} \\ \hat{\Sigma}_{\tilde{\lambda} \sigma}^{\prime} & \frac{1}{2} \hat{P}^{\prime} \hat{P}\end{array}\right)$,
where $\hat{\Gamma}_{n}=\left.\Gamma\right|_{\hat{\lambda}_{n}}$,

$$
\begin{aligned}
\hat{\Sigma}_{\tilde{\lambda}} & =\frac{1}{n} \sum_{t=1}^{n}\left[\frac{\partial \varepsilon_{t}^{\prime}}{\partial \tilde{\lambda}}\left(D_{t} \Gamma D_{t}\right)^{-1} \frac{\partial \varepsilon_{t}}{\partial \tilde{\lambda}^{\prime}}\right]_{\hat{\lambda}_{n}}+\frac{1}{4 n} \sum_{t=1}^{n}\left[\frac{\partial H_{t}^{\prime}}{\partial \tilde{\lambda}} D_{t}^{-2} \hat{C}_{n} D_{t}^{-2} \frac{\partial H_{t}}{\partial \tilde{\lambda}^{\prime}}\right]_{\hat{\lambda}_{n}} \\
\hat{\Sigma}_{\tilde{\lambda} \sigma} & =\frac{1}{2 n} \sum_{t=1}^{n}\left[\frac{\partial H_{t}^{\prime}}{\partial \tilde{\lambda}} D_{t}^{-2}\right]_{\hat{\lambda}_{n}} C_{1} \hat{P}, \quad \hat{P}=\left(I_{m} \otimes \hat{\Gamma}_{n}^{-1}\right) \mathcal{K} \\
\hat{C}_{n} & =\hat{\Gamma}_{n}^{-1} \odot \hat{\Gamma}_{n}+I_{m}
\end{aligned}
$$

LEMMA 5.5. Under the assumptions of Lemma 5.3, $\left\|\Sigma_{0}\right\|<\infty$ and $\hat{\Sigma}_{n}=$ $\Sigma_{0}+o_{p}(1)$ for any sequence $\lambda_{n}$, such that $\lambda_{n}-\lambda_{0}=o_{p}(1)$. If $\Gamma_{0}^{-1} \odot \Gamma_{0} \geq I_{m}$, then $-\Sigma_{0}>0$.

From the proof, we can see that the sixth-order moment in models (2.4) and (2.5) is required only for Lemma 5.4(a), whereas the fourth-order moment is sufficient for Lemma 5.4(b). If we can show that the convergent rate of the QMLE is $O_{p}\left(n^{-1 / 2}\right)$, then the fourth-order moment is sufficient for models (2.4) and (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 $\sqrt{n}\left(\lambda_{n}-\lambda_{0}\right)=$ $O_{p}(1)$, then
(a) $\frac{1}{n} \sum_{t=1}^{n}\left[\frac{\partial l_{t}^{\epsilon}}{\partial \lambda} \frac{\partial l_{t}^{\epsilon}}{\partial \lambda^{\prime}}-\frac{\partial l_{0 t}^{\epsilon}}{\partial \lambda} \frac{\partial l_{0 t}^{\epsilon}}{\partial \lambda^{\prime}}\right]_{\lambda_{n}}=o_{p}(1)$,
(b) $\hat{\Omega}_{n} \equiv \frac{1}{n} \sum_{t=1}^{n}\left[\frac{\partial l_{t}}{\partial \lambda} \frac{\partial l_{t}}{\partial \lambda^{\prime}}\right]_{\lambda_{n}}=\Omega_{0}+o_{p}(1)$.

THEOREM 5.1. Suppose that $Y_{t}$ is generated by models (2.4) and (2.5) satisfying Assumptions $1-4$ and $E\left\|Y_{t}\right\|^{6}<\infty$, or models (2.8) and (2.9) satisfying Assumptions $1-3$ and $E\left\|Y_{t}\right\|^{4}<\infty$, or models (2.10) and (2.11) satisfying Assumption 3 and $E\left\|\eta_{0 t}\right\|^{4}<\infty$. If $\Omega_{0}>0$ and $\Gamma_{0}^{-1} \odot \Gamma_{0} \geq I_{m}$, then $\sqrt{n}\left(\hat{\lambda}_{n}-\lambda_{0}\right) \rightarrow_{\mathcal{L}} N\left(0, \Sigma_{0}^{-1} \Omega_{0} \Sigma_{0}^{-1}\right)$. 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 $\Gamma_{0}^{-1} \odot \Gamma_{0} \geq I_{m}$, and hence, in this case, $-\Sigma_{0}>0$. However, it is difficult to prove $\Gamma_{0}^{-1} \odot \Gamma_{0} \geq I_{m}$ 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 $\eta_{0 t}$ follows a symmetric distribution,

$$
\Omega_{0}=E\left(\begin{array}{cc}
\frac{\partial l_{0 t}^{\epsilon}}{\partial \varphi} \frac{\partial l_{0 t}^{\epsilon}}{\partial \varphi^{\prime}} & 0 \\
0 & \frac{\partial l_{0 t}^{\epsilon}}{\partial \tilde{\delta}} \frac{\partial l_{0 t}^{\epsilon}}{\partial \tilde{\delta}^{\prime}}
\end{array}\right) \quad \text { and } \quad \Sigma_{0}=-\left(\begin{array}{cc}
\Sigma_{\varphi 0} & 0 \\
0 & \Sigma_{\tilde{\delta} 0}
\end{array}\right)
$$

in which $\tilde{\delta}=\left(\delta^{\prime}, \sigma^{\prime}\right)^{\prime}$,
$\Sigma_{\varphi 0}=E\left[\frac{\partial \varepsilon_{0 t}^{\epsilon^{\prime}}}{\partial \varphi}\left(D_{0 t} \Gamma_{0} D_{0 t}\right)^{-1} \frac{\partial \varepsilon_{0 t}^{\epsilon}}{\partial \varphi^{\prime}}\right]+\frac{1}{4} E\left[\frac{\partial H_{0 t}^{\epsilon^{\prime}}}{\partial \varphi} D_{0 t}^{-2} C D_{0 t}^{-2} \frac{\partial H_{0 t}^{\epsilon}}{\partial \varphi^{\prime}}\right]$,
$\Sigma_{\tilde{\delta} 0}=\left(\begin{array}{cc}\Sigma_{\delta 0} & \Sigma_{\delta \sigma 0} \\ \Sigma_{\delta \sigma 0}^{\prime} & \frac{1}{2} P^{\prime} P\end{array}\right)$,
where $\Sigma_{\delta 0}=E\left[\partial H_{0 t}^{\epsilon^{\prime}} / \partial \delta D_{0 t}^{-2} C D_{0 t}^{-2} \partial H_{0 t}^{\epsilon} / \partial \delta^{\prime}\right] / 4$ and $\Sigma_{\delta \sigma 0}=E\left[\partial H_{0 t}^{\epsilon^{\prime}} /\right.$ $\left.\partial \delta D_{0 t}^{-2}\right] C_{1} P / 2$. Furthermore, if $\eta_{0 t}$ 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) and (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 (2002c).

## 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 modeling 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 ARMAGARCH 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) and (2.9) and (2.10) and (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

Ahn, S.K. \& Reinsel, G.C. (1988) Nested reduced-rank autoregressive models for multiple time series. Journal of the American Statistical Association 83, 849-856.
Amemiya, T. (1985) Advanced Econometrics. Cambridge: Harvard University Press.
Basawa, I.V., P.D. Feign, \& C.C. Heyde (1976) Asymptotic properties of maximum likelihood estimators for stochastic processes. Sankhya A 38, 259-270.
Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307-327.
Bollerslev, T. (1990) Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach. Review of Economics and Statistics 72, 498-505.
Bollerslev, T., R.Y. Chou, \& K.F. Kroner (1992) ARCH modelling in finance: A review of the theory and empirical evidence. Journal of Econometrics 52, 5-59.
Bollerslev, T., R.F. Engle, \& D.B. Nelson (1994) ARCH models. In R.F. Engle and D.L. McFadden (eds.), Handbook of Econometrics, vol. 4, pp. 2961-3038. Amsterdam: North-Holland.
Bollerslev, T., R.F. Engle, \& J.M. Wooldridge (1988) A capital asset pricing model with time varying covariance. Journal of Political Economy 96, 116-131.
Chung, K.L. (1968) A Course in Probability Theory. New York: Harcourt Brace.
Dunsmuir, W. \& E.J. Hannan (1976) Vector linear time series models. Advances in Applied Probability 8, 339-364.
Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of variance of U.K. inflation. Econometrica 50, 987-1008.
Engle, R.F., C.W.J. Granger, \& D.F. Kraft (1984) Combing competing forecasts of inflation using a bivariate ARCH model. Journal of Economic Dynamics and Control 8, 151-165.
Engle, R.F. \& K.F. Kroner (1995). Multivariate simultaneous generalized ARCH. Econometric Theory 11, 122-150.
Engle, R.F. \& A.P. Rodrigues (1989) Tests of international CAPM with time varying covariances. Journal of Applied Econometrics 4, 119-128.
Harville, D.A. (1997) Matrix Algebra from a Statistician's Perspective. New York: Springer-Verlag.
Jeantheau, T. (1998) Strong consistency of estimators for multivariate ARCH models. Econometric Theory 14, 70-86.
Johansen, S. (1995) Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. New York: Oxford University Press.
Lee, S.-W. \& B.E. Hansen (1994) Asymptotic theory for the GARCH $(1,1)$ quasi-maximum likelihood estimator. Econometric Theory 10, 29-52.
Li, W.K., S. Ling, \& M. McAleer (2002) Recent theoretical results for time series models with GARCH errors. Journal of Economic Surveys 16, 245-269. Reprinted in M. McAleer and L. Oxley (eds.), Contributions to Financial Econometrics: Theoretical and Practical Issues (Oxford: Blackwell, 2002), pp. 9-33.
Li, W.K., S. Ling, \& H. Wong (1998) Estimation for Partially Nonstationary Multivariate Autoregressive Models with Conditional Heteroskedasticity. Technical report, Department of Statistics, University of Hong Kong.
Ling, S. (1999) On the probabilistic properties of a double threshold ARMA conditional heteroskedasticity model. Journal of Applied Probability 3, 1-18.
Ling, S. \& W.C. Deng (1993) Parametric estimate of multivariate autoregressive models with conditional heterocovariance matrix errors. Acta Mathematicae Applicatae Sinica 16, 517-533 (in Chinese).

Ling, S. \& W.K. Li (1997) On fractionally integrated autoregressive moving-average time series models with conditional heteroskedasticity. Journal of the American Statistical Association 92, 1184-1194.
Ling, S. \& W.K. Li (1998) Limiting distributions of maximum likelihood estimators for unstable ARMA models with GARCH errors. Annals of Statistics 26, 84-125.
Ling, S. \& M. McAleer (2002a) Necessary and sufficient moment conditions for the $\operatorname{GARCH}(r, s)$ and asymmetric power $\operatorname{GARCH}(r, s)$ models. Econometric Theory 18, 722-729.
Ling, S. \& M. McAleer (2002b) Stationarity and the existence of moments of a family of GARCH processes. Journal of Econometrics 106, 109-117.
Ling, S. \& M. McAleer (2002c) On adaptive estimation in nonstationary ARMA models with GARCH errors. Annals of Statistics, forthcoming.
Lumsdaine, R.L. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH $(1,1)$ and covariance stationary $\operatorname{GARCH}(1,1)$ models. Econometrica 64 , 575-596.
Lütkepohl, H. (1993) Introduction to Multiple Time Series Analysis, 2nd ed. Berlin: Springer-Verlag. Q
Newey, W.K. \& D.L. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle and D.L. McFadden (eds.), Handbook of Econometrics, vol. 4, pp. 2111-2245, Amsterdam: Elsevier.
Pantula, S.G. (1989) Estimation of autoregressive models with ARCH errors. Sankhya B 50, 119-138.
Pfanzagl, J. (1969) On the measurability and consistency of minimum contrast estimates. Metrika 14, 249-272.
Stout, W.F. (1974) Almost Sure Convergence. New York: Academic Press.
Tse, Y.K. (2000) A test for constant correlations in a multivariate GARCH model. Journal of Econometrics 98, 107-127.
Tweedie, R.L. (1988) Invariant measure for Markov chains with no irreducibility assumptions. Journal of Applied Probability 25A, 275-285.
Weiss, A.A. (1986) Asymptotic theory for ARCH models: Estimation and testing. Econometric Theory 2, 107-131.
Wong, H. \& W.K. Li (1997) On a multivariate conditional heteroscedasticity model. Biometrika 4, 111-123.
Wong, H., W.K. Li, \& S. Ling (2000) A Cointegrated Conditional Heteroscedastic Model with Financial Applications, Technical report, Department of Applied Mathematics, Hong Kong Polytechnic University.

## APPENDIX A: PROOFS OF THEOREMS 2.1 <br> AND 2.2

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

$$
\begin{equation*}
\vec{\varepsilon}_{0 t}=\tilde{\eta}_{0 t} W_{0}+\sum_{i=1}^{r} \tilde{\eta}_{0 t} A_{0 i} \vec{\varepsilon}_{0 t-i}+\sum_{i=1}^{s} \tilde{\eta}_{0 t} B_{0 i} H_{0 t-i} \tag{A.1}
\end{equation*}
$$

Now rewrite (A.1) in vector form as
$X_{t}=\tilde{A}_{0 t} X_{t-1}+\xi_{t}$,
where $X_{t}=\left(\vec{\varepsilon}_{0 t}^{\prime}, \ldots, \vec{\varepsilon}_{0 t-r+1}^{\prime}, H_{0 t}^{\prime}, \ldots, H_{0 t-s+1}^{\prime}\right)^{\prime}$ and $\xi_{t}$ is defined as in (2.9). Let
$S_{n, t}=\xi_{t}+\sum_{j=1}^{n}\left(\prod_{i=1}^{j} \tilde{A}_{0 t-i+1}\right) \xi_{t-j}$,
where $n=1,2, \ldots$. Denote the $k$ th element of $\left(\prod_{i=1}^{j} \tilde{A}_{0 t-i}\right) \xi_{t-j-1}$ by $s_{n, t}$. We have

$$
\begin{align*}
E\left|s_{n, t}\right| & =e_{k}^{\prime} E\left(\prod_{i=1}^{j} \tilde{A}_{0 t-i}\right) \xi_{t-j-1} \\
& =e_{k}^{\prime}\left(\prod_{i=1}^{j} E \tilde{A}_{0 t-i}\right) E \xi_{t-j-1}=e_{k}^{\prime} \tilde{A}^{j} c^{*}, \tag{A.4}
\end{align*}
$$

where $e_{k}=(0, \ldots, 0,1,0, \ldots, 0)_{m(r+s) \times 1}^{\prime}$ and 1 appears in the $k$ th position, $c^{*}=E \xi_{t}$ is a constant vector, and
$\tilde{A}=\left(\begin{array}{ccc|ccc}A_{01} & \ldots & A_{0 r} & B_{01} & \ldots & B_{0 s} \\ & I_{m(r-1)} & O_{m(r-1) \times m} & & O_{m(r-1) \times m s} & \\ \hline A_{01} & \ldots & A_{0 r} & B_{01} & \ldots & B_{0 s} \\ & O_{m(s-1) \times m r} & & & I_{m(s-1)} & O_{m(s-1) \times m}\end{array}\right)$.
By direct calculation, we know that the characteristic polynomial of $\tilde{A}$ is
$f(z)=|z|^{(r+s) m}\left|I_{m}-\sum_{i=1}^{r} A_{i} z^{-i}-\sum_{i=1}^{s} B_{i} z^{-i}\right|$.
By Assumption 3, it is obvious that all the roots of $f(z)$ lie inside the unit circle. Thus, $\rho(\tilde{A})<1$, and hence each component of $\tilde{A}^{i}$ is $O\left(\rho^{i}\right)$. Therefore, the right-hand side of (A.4) is equal to $O\left(\rho^{j}\right)$. Note that $\tilde{\eta}_{0 t}$ is a sequence of i.i.d. random matrices and each element of $\tilde{A}_{0 t}$ and $\xi_{t}$ is nonnegative. We know that each component of $S_{n, t}$ converges almost surely (a.s.) as $n \rightarrow \infty$, as does $S_{n, t}$. Denote the limit of $S_{n, t}$ by $X_{t}$. We have
$X_{t}=\xi_{t}+\sum_{j=1}^{\infty}\left(\prod_{i=1}^{j} \tilde{A}_{0 t-i}\right) \xi_{t-j-1}$,
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 secondorder solution $\varepsilon_{0 t}$ to (2.5) with $i$ th element $\varepsilon_{0 i t}=\eta_{0 i t} \sqrt{h_{0 i t}}=\eta_{0 i t}\left(e_{r m+i}^{\prime} X_{t}\right)^{1 / 2}$, 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 $X_{t}^{(1)}=\tilde{A}_{0 t} X_{t-1}^{(1)}+$ $\xi_{t}$, where $X_{t}^{(1)}=\left(\vec{\varepsilon}_{t}^{(1)^{\prime}}, \ldots, \vec{\varepsilon}_{t-r+1}^{(1)^{\prime}}, H_{t}^{(1)^{\prime}}, \ldots, H_{t-s+1}^{(1)^{\prime}}\right)^{\prime}$ and $H_{t}^{(1)}=W_{0}+\sum_{i=1}^{r} A_{0 i} \vec{\varepsilon}_{t-i}^{(1)}+$ $\sum_{i=1}^{s} B_{0 i} H_{t-i}^{(t)}$ with $\vec{\varepsilon}_{t}^{(1)}=\left(\varepsilon_{1 t}^{(1) 2}, \ldots, \varepsilon_{m t}^{(1) 2}\right)^{\prime}$. Let $U_{t}=X_{t}-X_{t}^{(1)}$. Then $U_{t}$ is first-order stationary and, by (A.2), $U_{t}=\left(\prod_{i=0}^{n} \tilde{A}_{0 t-i}\right) U_{t-n-1}$. Denote the $k$ th component of $U_{t}$ as $u_{k, t}$. Then, as each element of $\tilde{A}_{t}$ is nonnegative,

$$
\begin{equation*}
\left|u_{k t}\right| \leq\left|e_{k}^{\prime}\left(\prod_{i=0}^{n} \tilde{A}_{0 t-i}\right) U_{t-n-1}\right| \leq e_{k}^{\prime}\left(\prod_{i=0}^{n} \tilde{A}_{0 t-i}\right)\left|U_{t-n-1}\right| \tag{A.8}
\end{equation*}
$$

where $e_{k}$ is defined as in (A.4) and $\left|U_{t}\right|$ is defined as $\left(\left|u_{1 t}\right|, \ldots,\left|u_{(r+s) m, t}\right|\right)^{\prime}$. As $U_{t}$ is first-order stationary and $\mathcal{F}_{t}$-measurable, by (A.8), we have
$E\left|u_{k t}\right| \leq e_{k}^{\prime} E\left(\prod_{i=0}^{n} \tilde{A}_{0 t-i}\right) E\left|U_{t-n-1}\right|=e_{k}^{\prime} \tilde{A}^{n} c_{1}^{*} \rightarrow 0$
as $n \rightarrow \infty$, where $c_{1}^{*}=E\left|U_{t}\right|$ is a constant vector. So $u_{k t}=0$ a.s., that is, $X_{t}=X_{t}^{(1)}$ a.s. Thus, $h_{i t}=h_{i t}^{(1)}$ a.s., and hence $\varepsilon_{0 t}=\varepsilon_{0 t}^{(1)}=\eta_{0 i t} h_{0 i t}^{1 / 2}$ a.s. That is, $\varepsilon_{0 t}$ satisfying (2.5) is unique.
For the unique solution $\varepsilon_{0 t}$, 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
$Y_{t}=\sum_{k=0}^{\infty} \mathrm{Y}_{0 k} \varepsilon_{0 t-k}$.
Note that the solution $\left\{Y_{t}, \varepsilon_{0 t}, H_{0 t}\right\}$ is a fixed function of a sequence of i.i.d. random vectors $\eta_{0 t}$ and hence is strictly stationary and ergodic. This completes the proof.

The proof of Theorem 2.2 first transforms models (2.4) and (2.5) into a Markov chain and then uses Tweedie's criterion. Let $\left\{X_{t} ; t=1,2, \ldots\right\}$ be a temporally homogeneous Markov chain with a locally compact completely separable metric state space ( $S, \mathcal{B}$ ). The transition probability is $P(x, A)=\operatorname{Pr}\left(X_{n} \in A \mid X_{n-1}=x\right)$, 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 $\left\{X_{t}\right\}$ is a Feller chain.
(1) If there exist, for some compact set $A \in \mathcal{B}$, a nonnegative function $g$ and $\varepsilon>0$ satisfying

$$
\begin{equation*}
\int_{A^{c}} P(x, d y) g(y) \leq g(x)-\varepsilon, \quad x \in A^{c}, \tag{A.11}
\end{equation*}
$$

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

$$
\begin{equation*}
\int_{A} \mu(d x)\left[\int_{A^{c}} P(x, d y) g(y)\right]<\infty, \tag{A.12}
\end{equation*}
$$

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

$$
\begin{equation*}
\int_{A^{c}} P(x, d y) g(y) \leq g(x)-f(x), \quad x \in A^{c}, \tag{A.13}
\end{equation*}
$$

then $\mu$ admits a finite $f$-moment, that is,

$$
\begin{equation*}
\int_{S} \mu(d y) f(y)<\infty . \tag{A.14}
\end{equation*}
$$

The following two lemmas are preliminary results for the proof of Theorem 2.2.
LEMMA A.2. Suppose that $E\left(\left\|\eta_{0 t}\right\|^{2 k}\right)<\infty$ and $\rho\left[E\left(\tilde{A}_{0 t}^{\otimes k}\right)\right]<1$. Then there exists a vector $M>0$ such that $\left[I_{m}-E\left(\tilde{A}_{0 t}^{\otimes k}\right)^{\prime}\right] M>0$, where a vector $B>0$ means that each element of $B$ is positive.

Proof. From the condition given, $I_{m}-E\left(\tilde{A}_{0 t}^{\otimes k}\right)$ is invertible. Because each element of $E\left(\tilde{A}_{0 t}^{\otimes k}\right)$ is nonnegative, we can choose a vector $L_{1}>0$ such that
$M:=\left[I_{m}-E\left(\tilde{A}_{0 t}^{\otimes k}\right)^{\prime}\right]^{-1} L_{1}=L_{1}+\sum_{i=1}^{\infty}\left[E\left(\tilde{A}_{0 t}^{\otimes k}\right)^{\prime}\right]^{i} L_{1}>0$.
Thus, $\left[I_{m}-E\left(\tilde{A}_{0 t}^{\otimes k}\right)^{\prime}\right] M=L_{1}>0$. This completes the proof.
LEMMA A.3. Suppose that there is a vector $M>0$ such that
$\left[I_{m}-E\left(\tilde{A}_{0 t}^{\otimes k}\right)^{\prime}\right] M>0$.
Then there exists a compact set $A=\left\{x: \tilde{x}^{k} \equiv\left(\sum_{i=1}^{(r+s) m} x_{i}\right)^{k} \leq \Delta\right\} \subset R_{0}^{(r+s) m}$ with $R_{0}=$ $(0, \infty)$, a function $g_{1}(x)$, and $\kappa>0$ such that the function $g$, defined by $g(x)=1+$ $\left(x^{\otimes k}\right)^{\prime} M$, satisfies
$E\left(g\left(X_{t}\right) \mid X_{t-1}=x\right) \leq g(x)+g_{1}(x), \quad x \in R_{0}^{(r+s) m}$,
(A.16)
and
$E\left(g\left(X_{t}\right) \mid X_{t-1}=x\right) \leq(1-\kappa) g(x), \quad x \in A^{c}$,
where $A^{c}=R^{(r+s) m}-A, x_{i}$ is the ith 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

$$
\begin{align*}
& E\left[\left(\xi_{t}+\tilde{A}_{0 t} x\right)^{\otimes 3}\right]^{\prime} M \\
& \quad=\left(x^{\otimes 3}\right)^{\prime} E\left(\tilde{A}_{0 t}^{\otimes 3}\right)^{\prime} M+C_{1}^{\prime} M+x^{\prime} C_{2}^{\prime} M+\left(x^{\otimes 2}\right)^{\prime} C_{3}^{\prime} M \\
& \quad \leq\left(x^{\otimes 3}\right)^{\prime} E\left(\tilde{A}_{0 t}^{\otimes 3}\right)^{\prime} M+c\left(1+\tilde{x}+\tilde{x}^{2}\right), \tag{A.18}
\end{align*}
$$

where $C_{1}, C_{2}$, and $C_{3}$ are some constant vectors or matrices with nonnegative elements, which do not depend on $x$, and $c=\max _{k}\left\{\right.$ all components of $C_{1}^{\prime} M, C_{2}^{\prime} M$, and $\left.C_{3}^{\prime} M\right\}$.

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

$$
\begin{align*}
E\left[g\left(X_{t}\right) \mid X_{t-1}=x\right] & =1+E\left[\left(\xi_{t}+\tilde{A}_{0 t} x\right)^{\otimes 3}\right]^{\prime} M \\
& \leq 1+\left(x^{\otimes 3}\right)^{\prime} E\left(\tilde{A}_{0 t}^{\otimes 3}\right)^{\prime} M+g_{1}(x) \\
& =1+\left(x^{\otimes 3}\right)^{\prime} M-\left(x^{\otimes 3}\right)^{\prime} M^{*}+g_{1}(x) \\
& =g(x)\left[1-\frac{\left(x^{\otimes 3}\right)^{\prime} M^{*}}{g(x)}+\frac{g_{1}(x)}{g(x)}\right], \tag{A.19}
\end{align*}
$$

where $M^{*}=\left[I_{m}-E\left(\tilde{A}_{0 t}^{\otimes 3}\right)^{\prime}\right] M$ and $g_{1}(x)=c\left(1+\tilde{x}+\tilde{x}^{2}\right)$.
Denote
$A=\left\{x: \tilde{x}^{3} \leq \Delta, x \in R_{0}^{(r+s) m}\right\}, \quad c_{1}=\min \left\{\right.$ all components of $\left.M^{*}\right\}$,
$c_{2}=\max \{$ all components of $M\}, \quad c_{3}=\min \{$ all components of $M\}$.

It is obvious that $A$ is a compact set on $R_{0}^{(r+s) m}$. Because $M^{*}, M>0$, it follows that $c_{1}, c_{2}, c_{3}>0$. From (A.19), we can show that
$E\left[g\left(X_{t}\right) \mid X_{t-1}=x\right] \leq g(x)+g_{1}(x), \quad x \in R_{0}^{(r+s) m}$,
where $\max _{x \in A} g_{1}(x)<C_{0}(\Delta)$ and $C_{0}(\Delta)$ is a constant not depending on $x$.
Let $\Delta>\max \left\{1 / c_{2}, 1\right\}$. When $x \in A^{c}$,
$c_{3} \Delta<c_{3} \tilde{x}^{3} \leq g(x) \leq 1+c_{2} \tilde{x}^{3} \leq 2 c_{2} \tilde{x}^{3}$.

Thus,
$\frac{\left(x^{\otimes 3}\right)^{\prime} M^{*}}{g(x)} \geq \frac{c_{1} \tilde{x}^{3}}{2 c_{2} \tilde{x}^{3}}=\frac{c_{1}}{2 c_{2}}$,
and furthermore, because $1+\tilde{x} \leq 2 \tilde{x}$ as $x \in A^{c}$, we can show that
$\frac{g_{1}(x)}{g(x)} \leq \frac{g_{1}(x)}{c_{3} \tilde{x}^{3}} \leq \frac{C}{\Delta}$,
where $C$ is a positive constant not depending on $x$ and $\Delta$. By (A.19), (A.22), and (A.23), as $x \in A^{c}$,
$E\left[g\left(X_{t}\right) \mid X_{t-1}=x\right] \leq g(x)\left(1-\frac{c_{1}}{2 c_{2}}+\frac{C}{\Delta}\right)$.
Provided $0<c_{1} / 4 c_{2}<\kappa<c_{1} / 2 c_{2}$ and $\Delta>\max \left\{1,1 / c_{2}, C /\left(c_{1} / 2 c_{2}-\kappa\right)\right\}$, then $E\left[g\left(X_{t}\right) \mid X_{t-1}=x\right] \leq g(x)(1-\kappa)$. This completes the proof.

Proof of Theorem 2.2. Obviously, $X_{t}$ defined as in (A.2) is a Markov chain with state space $R_{0}^{(r+s) m}$. It is straightforward to prove that, for each bounded continuous function $g$ on $R_{0}^{(r+s) m}, E\left[g\left(X_{t}\right) \mid X_{t-1}=x\right]$ is continuous in $x$, that is, $\left\{X_{t}\right\}$ is a Feller chain. In a similar manner to Lemma A.3, in the following discussion 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 $A=\left\{x: \tilde{x}^{3}=\left(\sum_{i=1}^{(r+s) m} x_{i}\right)^{3} \leq \Delta\right\} \subset R_{0}^{(r+s) m}$, and $\kappa>0$ such that the function defined by $g(x)=1+\left(x^{\otimes 3}\right)^{\prime} M$ satisfies
$E\left[g\left(X_{t}\right) \mid X_{t-1}=x\right] \leq g(x)+g_{1}(x), \quad x \in R_{0}^{(r+s) m}$
and
$E\left[g\left(X_{t}\right) \mid X_{t-1}=x\right] \leq(1-\kappa) g(x), \quad x \in A^{c}$,
where $\max _{x \in A} g_{1}(x)<C_{0}$ and $C_{0}, \kappa$, and $\Delta$ are positive constants not depending on $x$.
Because $g(x) \geq 1$, it follows that $E\left[g\left(X_{t}\right) \mid X_{t-1}=x\right) \leq g(x)-\kappa$. By Lemma A.1, Q9 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

$$
\begin{aligned}
E\left[g\left(X_{t}\right) \mid X_{t-1}=x\right] & \leq 1+c_{2}\left(\sum_{i=1}^{(r+s) m} x_{i}\right)^{3}+g_{1}(x) \\
& \leq \Delta_{1}<\infty
\end{aligned}
$$

where $\Delta_{1}$ is a constant not depending on $x$. Hence,

$$
\begin{aligned}
& \int_{A} \mu(d x)\left\{\int_{A^{c}} P(x, d y) g(y)\right\} \\
& \quad \leq \int_{A} \mu(d x) E\left[g\left(X_{t}\right) \mid X_{t-1}=x\right] \leq \Delta_{1} \mu(A)<\infty .
\end{aligned}
$$

This shows that $\left\{X_{t}\right\}$ has a finite invariant measure $\mu$ and hence $\pi=\mu / \mu\left(R_{0}^{(r+s) m}\right)$ is an invariant probability measure of $\left\{X_{t}\right\}$; 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

$$
\begin{aligned}
\int_{A^{c}} P(x, d y) g(y) & \leq E\left[g\left(X_{t}\right) \mid X_{t-1}=x\right] \\
& \leq g(x)-\kappa g(x) \leq g(x)-f(x) .
\end{aligned}
$$

By Lemma A.1(3), we know that $E_{\pi}\left[f\left(X_{t}\right)\right]=c_{3} \kappa E\left[\left(\sum_{i=1}^{(r+s) m} x_{i t}\right)^{3}\right]<\infty$, where $\pi$ is the stationary distribution of $\left\{X_{t}\right\}$, where $x_{i t}$ is the $i$ th component of $X_{t}$. Thus, $E_{\pi_{1}}\left\|\varepsilon_{0 t}\right\|^{6}<\infty$, where $\pi_{1}$ are the stationary distributions of $\left\{\varepsilon_{0 t}\right\}$. Now, because $E_{\pi_{1}}\left\|\varepsilon_{0 t}\right\|^{6}<\infty$, it is easy to show that $E_{\pi_{2}}\left\|Y_{t}\right\|^{6}<\infty$, where $\pi_{2}$ is the stationary distribution of $Y_{t}$.

By Hölder's inequality, $E_{\pi_{1}}\left\|\varepsilon_{0 t}\right\|^{2}<\left(E_{\pi_{1}}\left\|\varepsilon_{0 t}\right\|^{2 k}\right)^{1 / k}<\infty$. Similarly, we have $E_{\pi_{2}}\left\|Y_{t}\right\|^{2}<\infty$. Thus, $\left\{Y_{t}, \varepsilon_{0 t}\right\}$ is a second-order stationary solution of models (2.4) and (2.5). Furthermore, by Theorem 2.1, the solution $\left\{Y_{t}, \varepsilon_{0 t}\right\}$ is unique and ergodic. Thus, the process $\left\{Y_{t}, \varepsilon_{0 t}\right\}$ satisfying models (2.4) and (2.5) has a finite $2 k$ th moment. This completes the proof.

## APPENDIX B: PROOFS OF RESULTS IN SECTIONS 3-5

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

Proof of Lemma 4.1. Note that
$\tilde{A} \geqslant\left(\begin{array}{cc}O & O \\ O & \tilde{B}\end{array}\right)$,
where $\tilde{A}$ is defined as in (A.5), $\tilde{B}=\left(\begin{array}{ccc}B_{01} & \ldots & B_{0 s} \\ & I_{m(s-1)} & O_{m(s-1) \times m}\end{array}\right)$, and here "the matrix $A \geqslant$ the matrix $B$ " means that each component of $A$ is larger than or equal to the corresponding component of $B$. Thus, we have
$\tilde{A}^{i} \geqslant\left(\begin{array}{cc}O & O \\ O & \tilde{B}^{i}\end{array}\right)$.
By Assumption 3, $\rho(\tilde{A})<1$, and hence $\sum_{i=0}^{k} \tilde{A}^{i}$ converges to a finite limit as $k \rightarrow \infty$. By (B.1), $\sum_{i=0}^{k} \tilde{B}^{i}$ also converges to a finite limit as $k \rightarrow \infty$, and hence $\rho(\tilde{B})<1$, which is equivalent to all the roots of $\left|I_{m}-\sum_{i=1}^{s} B_{i} L^{i}\right|=0$ lying outside the unit circle. This completes the proof.

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

Proof of Lemma 4.2. First, by (3.2),
$\varepsilon_{t}^{\epsilon}=\Psi(L)^{-1} \Phi(L)\left(Y_{t}-\mu\right), \quad \frac{\partial \varepsilon_{t}^{\epsilon}}{\partial \varphi^{\prime}}=\Psi^{-1}(L)\left[-\Phi(1), X_{t-1} \otimes I_{m}\right]$,
where $X_{t-1}=\left(Y_{t-1}^{\prime}-\mu^{\prime}, \ldots, Y_{t-p+1}^{\prime}-\mu^{\prime}, \varepsilon_{t-1}^{\epsilon^{\prime}}, \ldots, \varepsilon_{t-q+1}^{\epsilon^{\prime}}\right)$ and the preceding vector differentiation follows rules in Lütkepohl (1993, Appendix A). Denote $U_{t}=\partial \varepsilon_{t}^{\epsilon} / \partial \varphi^{\prime}$ and $V_{t}=\left[-\Phi(1), X_{t-1} \otimes I_{m}\right]$. Then
$U_{t}+\Psi_{1} U_{t-1}+\cdots+\Psi_{q} U_{t-q}=V_{t}$.
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 $c_{\varphi}$, whereas $c_{2}$ is the vector consisting of the remaining elements of $c_{\varphi}$. Then $-\Phi(1) c_{1}+\left(X_{t-1} \otimes I_{m}\right) c_{2}=0$. Because $X_{t-1}$ is not degenerate, $\left(X_{t-1} \otimes I_{m}\right) c_{2}=0$ 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),

$$
\begin{align*}
H_{t}^{\epsilon} & =\left(I_{m}-\sum_{i=1}^{s} B_{i} L^{i}\right)^{-1}\left[W+\left(\sum_{i=1}^{r} A_{i} L^{i}\right) \vec{\varepsilon}_{t}^{\epsilon}\right],  \tag{B.4}\\
\frac{\partial H_{t}^{\epsilon}}{\partial \delta^{\prime}} & =\left(I_{m}-\sum_{i=1}^{s} B_{i} L^{i}\right)^{-1}\left(I_{m}, \widetilde{H}_{t-1}^{\epsilon} \otimes I_{m}\right), \tag{B.5}
\end{align*}
$$

where $\tilde{H}_{t-1}^{\epsilon}=\left(\vec{\varepsilon}_{t-1}^{\epsilon^{\prime}}, \ldots, \vec{\varepsilon}_{t-r}^{\epsilon^{\prime}}, H_{t-1}^{\epsilon^{\prime}}, \ldots, H_{t-s}^{\epsilon^{\prime}}\right)$. Denoting $U_{1 t}=\partial H_{t}^{\epsilon} / \partial \delta^{\prime}$ and $V_{1 t}=$ ( $I_{m}, \widetilde{H}_{t-1}^{\epsilon} \otimes I_{m}$ ), we have the following recursive equation:
$U_{1 t}=B_{1} U_{1 t-1}+\cdots+B_{s} U_{1 t-s}+V_{1 t}$.

If $U_{1 t} c=0$ a.s., then $V_{1 t} c=0$ a.s. By Assumptions 3 and 4, in a similar manner to $V_{t} c_{\varphi}=0$, we can conclude $c=0$ (also refer to Jeantheau, 1998, the proof of Proposition 3.4). 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 $0<\varrho<1$, independent of all $\lambda \in \Theta$, such that
$\left\|\varepsilon_{t}^{\epsilon}\right\| \leq c_{0}+c_{1} \sum_{i=0}^{\infty} \varrho^{i}\left\|Y_{t-i}\right\| \equiv \varepsilon_{t}^{*}$.
Thus, $E \sup _{\lambda \in \Theta}\left\|\varepsilon_{t}^{\epsilon}\right\|^{2}<\infty$ by Theorem 2.1. Note that, by Assumption 3, $\left|D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right|$ has a lower bound uniformly over $\Theta$. We have $E \sup _{\lambda \in \Theta}\left[\varepsilon_{t}^{\epsilon^{\prime}}\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1} \varepsilon_{t}^{\epsilon}\right]<\infty$. By Assumption 3 and Lemma 4.1, we can show that
$\left\|H_{t}^{\epsilon}\right\| \leq c_{2}+c_{3} \sum_{i=1}^{\infty} \varrho_{1}^{i}\left\|Y_{t-i}\right\|^{2} \equiv \varepsilon_{h t}^{*}$,
where $c_{2}, c_{3}>0$ and $0<\varrho_{1}<1$ are constants independent of all $\lambda \in \Theta$. Thus, $E \sup _{\lambda \in \Theta}\left\|H_{t}^{\epsilon}\right\|<\infty$, and hence $E \sup _{\lambda \in \Theta}\left|D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right|<\infty$. By Jensen's inequality, $E \sup _{\lambda \in \Theta}|\ln | D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}| |<\infty$. Thus, $E\left|l_{t}^{\epsilon}(\lambda)\right|<\infty$ for all $\lambda \in \Theta$. Let $g\left(\underline{Y}_{t}, \lambda\right)=l_{t}^{\epsilon}-E l_{t}^{\epsilon}$, where $\underline{Y}_{t}=\left(Y_{t}, Y_{t-1}, \ldots\right)$. Then $E \sup _{\lambda \in \Theta}\left|g\left(\underline{Y}_{t}, \lambda\right)\right|<\infty$. Furthermore, because $g\left(\underline{Y}_{t}, \lambda\right)$ is strictly stationary with $E g\left(\underline{Y}_{t}, \lambda\right)=0$, by Theorem 3.1, $\sup _{\lambda \in \Theta}\left|n^{-1} \sum_{t=1}^{n} g\left(\underline{Y}_{t}, \lambda\right)\right|=$ $o_{p}(1)$. This completes the proof.

## Proof of Lemma 4.4. First,

$$
\begin{align*}
-E \ln \mid & D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon} \mid-E\left[\varepsilon_{t}^{\epsilon^{\prime}}\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1} \varepsilon_{t}^{\epsilon}\right] \\
= & -E \ln \left|D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right|-E\left[\left(\varepsilon_{t}^{\epsilon}-\varepsilon_{0 t}+\varepsilon_{0 t}\right)^{\prime}\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1}\left(\varepsilon_{t}^{\epsilon}-\varepsilon_{0 t}+\varepsilon_{0 t}\right)\right] \\
= & \left\{-E \ln \left|D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right|-E\left[\varepsilon_{0 t}^{\prime}\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1} \varepsilon_{0 t}\right]\right\} \\
& -E\left[\left(\varepsilon_{t}^{\epsilon}-\varepsilon_{0 t}\right)^{\prime}\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1}\left(\varepsilon_{t}^{\epsilon}-\varepsilon_{0 t}\right)\right] \equiv L_{1}(\lambda)+L_{2}(\lambda) . \tag{B.9}
\end{align*}
$$

The term $L_{2}(\lambda)$ obtains its maximum at zero, and this occurs if and only if $\varepsilon_{t}^{\epsilon}=\varepsilon_{0 t}$. Thus,
$\varepsilon_{t}^{\epsilon}-\varepsilon_{0 t}=\left.\frac{\partial \varepsilon_{t}^{\epsilon}}{\partial \varphi^{\prime}}\right|_{\varphi^{*}}\left(\varphi-\varphi_{0}\right)=0$.
By Lemma 4.2, we know that equation (B.10) holds if and only if $\varphi=\varphi_{0}$.

$$
\begin{align*}
L_{1}(\lambda) & =-E \ln \left|D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right|-E \operatorname{tr}\left(M_{t}\right) \\
& =-\left[-E \ln \left|M_{t}\right|+\operatorname{Etr}\left(M_{t}\right)\right]-E \ln \left|D_{0 t} \Gamma_{0} D_{0 t}\right|, \tag{B.11}
\end{align*}
$$

where $M_{t}=\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1 / 2}\left(D_{0 t} \Gamma_{0} D_{0 t}\right)\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1 / 2}$. Note that, for any positive definite matrix $M,-f(M) \equiv-\ln |M|+t r M \geq m$ (see Johansen, 1995, Lemma A.6), and hence
$-E \ln \left|M_{t}\right|+E t r\left(M_{t}\right) \geq m$.
(B.12)

When $M_{t}=I_{m}$, we have $f\left(M_{t}\right)=f\left(I_{m}\right)=-m$. If $M_{t} \neq I_{m}$, then $f\left(M_{t}\right)<f\left(I_{m}\right)$, so that $E f\left(M_{t}\right) \leq E f\left(I_{m}\right)$ with equality only if $M_{t}=I_{m}$ with probability one. Thus, $L_{1}(\lambda)$ reaches its maximum $-m-E \ln \left(D_{0 t} \Gamma_{0} D_{0 t}\right)$, and this occurs if and only if $D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}=$ $D_{0 t} \Gamma_{0} D_{0 t}$. From the definition of $\Gamma$, we have $h_{i t}=h_{0 i t}$, and hence $\Gamma=\Gamma_{0}$. Note that
$\max _{\lambda \in \Theta} L(\lambda) \leq \max _{\lambda \in \Theta} L_{1}(\lambda)+\max _{\lambda \in \Theta} L_{2}(\lambda)$.
The expression $\max _{\lambda \in \Theta} L(\lambda)=-m-E \ln \left(D_{0 t} \Gamma_{0} D_{0 t}\right)$ if and only if $\max _{\lambda \in \Theta} L_{2}(\lambda)=0$ and $\max _{\lambda \in \Theta} L_{1}(\lambda)=-m-E \ln \left(D_{0 t} \Gamma_{0} D_{0 t}\right)$, which occurs if and only if $\varphi=\varphi_{0}, \Gamma=$ $\Gamma_{0}$, and $h_{i t}=h_{0 i t}$. From $\varphi=\varphi_{0}$ and $h_{i t}=h_{0 i t}$, we have
$\left.\left(H_{t}^{\epsilon}-H_{0 t}\right)\right|_{\varphi=\varphi_{0}}=\left.\frac{\partial H_{t}^{\epsilon}}{\partial \delta^{\prime}}\right|_{\left(\varphi_{0}, \delta^{*}\right)}\left(\delta-\delta_{0}\right)=0$
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 maximized at $\lambda_{0}$. This completes the prod

Proof of Lemma 4.5. First, for any positive constant $M$,
$\left|\frac{1}{n} \sum_{t=1}^{n} X_{t} \xi_{t} I\left(\left|X_{t}\right|>M\right)\right| \leq \frac{C}{n} \sum_{t=1}^{n}\left|X_{t}\right| I\left(\left|X_{t}\right|>M\right)$,
where $I(\cdot)$ is the indicator function. For any small $\epsilon, \kappa>0$, because $E\left|X_{t}\right|<\infty$, there exists a constant $M_{0}$ such that

$$
\begin{align*}
& P\left(\left|\frac{1}{n} \sum_{t=1}^{n} X_{t} \xi_{t} I\left(\left|X_{t}\right|>M_{0}\right)\right|>\kappa\right) \\
& \leq \frac{1}{\kappa} E\left(\left|\frac{1}{n} \sum_{t=1}^{n} X_{t} \xi_{t} I\left(\left|X_{t}\right|>M_{0}\right)\right|\right) \\
& \leq \frac{C}{\kappa} E\left(\frac{1}{n} \sum_{t=1}^{n}\left|X_{t}\right| I\left(\left|X_{t}\right|>M_{0}\right)\right) \\
& \quad \leq \frac{C}{\kappa} \int_{|x|>M_{0}}|x| d F(x)<\frac{\epsilon}{2} \tag{B.15}
\end{align*}
$$

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 $n>N$,
$\left.P\left(\left|\frac{1}{n} \sum_{t=1}^{n} X_{t} \xi_{t} I\left(\left|X_{t}\right| \leq M_{0}\right)\right|>\kappa\right) \leq P\left(\frac{1}{n} \sum_{t=1}^{n}\left|\xi_{t}\right|>\kappa / M_{0}\right)\right)<\frac{\epsilon}{2}$.
By (B.15) and (B.16), as $n>N, P\left(\left|n^{-1} \sum_{t=1}^{n} X_{t} \xi_{t}\right|>2 \kappa\right)<\epsilon$, 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 $1, \varepsilon_{t}^{\epsilon}$ and $\varepsilon_{t}$ have the expansions
$\varepsilon_{t}^{\epsilon}=\sum_{k=0}^{\infty} \mathrm{Y}_{k}\left(Y_{t-k}-\mu\right), \quad \varepsilon_{t}=\sum_{k=0}^{t-1} \mathrm{Y}_{k}\left(Y_{t-k}-\mu\right)$,
where $\Phi^{-1}(L) \Psi(L)=\sum_{k=0}^{\infty} \Upsilon_{k} L^{k}$. By (B.17),
$\left\|\varepsilon_{t}^{\epsilon}-\varepsilon_{t}\right\| \leq c_{1} \sum_{k=t}^{\infty} \varrho_{1}^{k}\left\|Y_{t-k}-\mu\right\|$,
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
$H_{t}^{\epsilon}=\sum_{k=0}^{\infty} \Gamma_{k}\left[W+\left(\sum_{i=1}^{r} A_{i} L^{i}\right) \vec{\varepsilon}_{t-k}^{\epsilon}\right], \quad H_{t}=\sum_{k=0}^{t-1} \Gamma_{k}\left[W+\left(\sum_{i=1}^{r} A_{i} L^{i}\right) \vec{\varepsilon}_{t-k}\right]$,
where $\left(I_{m}-\sum_{i=1}^{s} B_{i} L^{i}\right)^{-1}=\sum_{k=0}^{\infty} \Gamma_{k} L^{k}$. By (B.19)
$\left\|H_{t}^{\epsilon}-H_{t}\right\| \leq \sum_{k=t}^{\infty} \varrho_{2}^{k}\left(c_{2}+c_{3}\left\|\vec{\varepsilon}_{t-k}^{\epsilon}-\vec{\varepsilon}_{t-k}\right\|\right)$,
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
$E \sup _{\lambda \in \Theta}\left(\varepsilon_{i t}^{\epsilon}-\varepsilon_{i t}\right)^{2}=O\left(\varrho^{t}\right) \quad$ and $\quad E \sup _{\lambda \in \Theta}\left|h_{i t}^{\epsilon}-h_{i t}\right|=O\left(\varrho^{t}\right)$,
where $i=1, \ldots, m, 0<\varrho<1$, and $O(\cdot)$ holds uniformly in all $t$. Because $h_{i t}$ has a lower bound, by (B.21), it follows that

$$
\begin{align*}
& \frac{1}{n} \sum_{t=1}^{n} E \sup _{\lambda \in \Theta}|\ln | D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}|-\ln | D_{t} \Gamma D_{t}| | \\
& \quad=\sum_{i=1}^{m}\left[\frac{1}{n} \sum_{t=1}^{n} E \sup _{\lambda \in \Theta}\left|\ln \left(\frac{h_{i t}^{\epsilon}}{h_{i t}}\right)\right|\right] \\
& \quad \leq \sum_{i=1}^{m}\left[\frac{1}{n} \sum_{t=1}^{n} E \sup _{\lambda \in \Theta}\left|\frac{h_{i t}^{\epsilon}-h_{i t}}{h_{i t}}\right|\right] \\
& \quad=O(1) \sum_{i=1}^{m}\left[\frac{1}{n} \sum_{t=1}^{n} E \sup _{\lambda \in \Theta}\left|h_{i t}^{\epsilon}-h_{i t}\right|\right] \\
& \quad=O(1) \sum_{i=1}^{m} \frac{1}{n} \sum_{t=1}^{n} O\left(\varrho^{t}\right)=o(1) . \tag{B.22}
\end{align*}
$$

Again, because $h_{i t}^{\varepsilon}$ and $h_{i t}$ have a lower bound uniformly in all $t, i$, and $\lambda$,

$$
\begin{equation*}
\sum_{i=1}^{m}\left|\frac{\varepsilon_{i t}^{\epsilon}}{\sqrt{h_{i t}^{\epsilon}}}-\frac{\varepsilon_{i t}}{\sqrt{h_{i t}}}\right|^{2} \leq \sum_{i=1}^{m}\left[\varepsilon_{i t}^{\epsilon \in}\left|\frac{1}{\sqrt{h_{i t}^{\epsilon}}}-\frac{1}{\sqrt{h_{i t}}}\right|^{2}+\left(\varepsilon_{i t}^{\epsilon}-\varepsilon_{i t}\right)^{2}\right] O(1), \tag{B.23}
\end{equation*}
$$

where $O(1)$ holds uniformly in all $t$. We have

$$
\begin{align*}
& \left|\varepsilon_{t}^{\epsilon^{\prime}}\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1} \varepsilon_{t}^{\epsilon}-\varepsilon_{t}^{\prime}\left(D_{t} \Gamma D_{t}\right)^{-1} \varepsilon_{t}\right| \\
& =\left|2 \varepsilon_{t}^{\epsilon^{\prime}} D_{t}^{\epsilon-1} \Gamma^{-1}\left(D_{t}^{\epsilon-1} \varepsilon_{t}^{\epsilon}-D_{t}^{-1} \varepsilon_{t}\right)-\left(\varepsilon_{t}^{\epsilon^{\prime}} D_{t}^{\epsilon-1}-\varepsilon_{t}^{\prime} D_{t}^{-1}\right) \Gamma^{-1}\left(D_{t}^{\epsilon-1} \varepsilon_{t}^{\epsilon}-D_{t}^{-1} \varepsilon_{t}\right)\right| \\
& \leq\left(\sum_{i=1}^{m}\left|\frac{\varepsilon_{i t}^{\epsilon}}{\sqrt{h_{i t}^{\epsilon}}}-\frac{\varepsilon_{i t}}{\sqrt{h_{i t}}}\right|^{2}\right)^{1 / 2}\left\|\varepsilon_{t}^{\epsilon}\right\| O(1)+\left(\sum_{i=1}^{m}\left|\frac{\varepsilon_{i t}^{\epsilon}}{\sqrt{h_{i t}^{\epsilon}}}-\frac{\varepsilon_{i t}}{\sqrt{h_{i t}}}\right|^{2}\right) O(1) \\
& \leq \sum_{i=1}^{m}\left[\left\|\varepsilon_{t}^{\epsilon}\right\|\left|\varepsilon_{i t}^{\epsilon}\right|\left|\frac{1}{\sqrt{h_{i t}^{\epsilon}}}-\frac{1}{\sqrt{h_{i t}}}\right|+\varepsilon_{i t}^{\epsilon 2}\left|\frac{1}{\sqrt{h_{i t}^{\epsilon}}}-\frac{1}{\sqrt{h_{i t}}}\right|^{2}\right] O(1) \\
& +\sum_{i=1}^{m}\left[\left\|\varepsilon_{t}^{\epsilon}\right\|\left|\varepsilon_{i t}^{\epsilon}-\varepsilon_{i t}\right|+\left(\varepsilon_{i t}^{\epsilon}-\varepsilon_{i t}\right)^{2}\right] O(1) \\
& =O(1) R_{1 t}+O(1) R_{2 t}, \tag{B.24}
\end{align*}
$$

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 $n^{-1} \sum_{t=1}^{n} \sup _{\lambda \in \Theta} R_{2 t}=o_{p}(1)$. Thus, it is sufficient to show that $n^{-1} \sum_{t=1}^{n} \sup _{\lambda \in \Theta} R_{1 t}=o_{p}(1)$. Let $X_{t}=\varepsilon_{t}^{* 2}$ and $\xi_{t}=$ $\sup _{\lambda \in \Theta}\left|h_{i t}^{\epsilon-1 / 2}-h_{i t}^{-1 / 2}\right|^{2}$, where $\varepsilon_{t}^{*}$ is defined by (B.7). Then, $X_{t}$ is a strictly stationary and ergodic time series, with $E X_{t}<\infty$ and $\left|\xi_{t}\right| \leq C$, a constant. Furthermore, by (B.21),

$$
\begin{aligned}
\frac{1}{n} \sum_{t=1}^{n} \xi_{t} & =\frac{1}{n} \sum_{t=1}^{n} \sup _{\lambda \in \Theta}\left|\frac{h_{i t}^{\epsilon}-h_{i t}}{\sqrt{h_{i t}^{\epsilon} h_{i t}}\left(\sqrt{h_{i t}^{\epsilon}}+\sqrt{h_{i t}}\right)}\right|^{2} \\
& \leq \frac{1}{n} \sum_{t=1}^{n} \sup _{\lambda \in \Theta} \frac{\left|h_{i t}^{\epsilon}-h_{i t}\right|\left(h_{i t}^{\epsilon}+h_{i t}\right)}{h_{i t}^{\epsilon} h_{i t}\left(\sqrt{h_{i t}^{\epsilon}}+\sqrt{h_{i t}}\right)^{2}} \\
& \leq O(1) \frac{1}{n} \sum_{t=1}^{n} \sup _{\lambda \in \Theta}\left|h_{i t}^{\epsilon}-h_{i t}\right| \\
& =O(1) \frac{1}{n} \sum_{t=1}^{n} O_{p}\left(\varrho^{t}\right)=o_{p}(1) .
\end{aligned}
$$

By Lemma 4.5, $n^{-1} \sum_{t=1}^{n} X_{t} \sup _{\lambda \in \Theta}\left|h_{i t}^{\epsilon-1 / 2}-h_{i t}^{-1 / 2}\right|^{2}=o_{p}(1)$. Similarly, we can show that $n^{-1} \sum_{t=1}^{n} X_{t} \sup _{\lambda \in \Theta}\left|h_{i t}^{\epsilon-1 / 2}-h_{i t}^{-1 / 2}\right|=o_{p}(1)$. Thus,

$$
\begin{aligned}
\frac{1}{n} \sum_{t=1}^{n} \sup _{\lambda \in \Theta} R_{1 t} \leq \sum_{i=1}^{m}\left\{\frac{1}{n} \sum_{t=1}^{n}[ \right. & {\left[X_{t} \sup _{\lambda \in \Theta}\left|h_{i t}^{\epsilon-1 / 2}-h_{i t}^{-1 / 2}\right|^{2}\right.} \\
& \left.\left.+X_{t} \sup _{\lambda \in \Theta}\left|h_{i t}^{\epsilon-1 / 2}-h_{i t}^{-1 / 2}\right|\right]\right\}=o_{p}(1) .
\end{aligned}
$$

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 $Y_{t}, t=1, \ldots, n$ for all $\lambda \in \Theta$. Third, by Lemmas 4.3 and 4.4, $L_{n}^{\epsilon}(\lambda) \rightarrow_{p} L(\lambda)$ uniformly in $\Theta$. From Lemma 4.6, we have

$$
\sup _{\lambda \in \Theta}\left|L_{n}(\lambda)-L(\lambda)\right| \leq \sup _{\lambda \in \Theta}\left|L_{n}^{\epsilon}(\lambda)-L(\lambda)\right|+\sup _{\lambda \in \Theta}\left|L_{n}^{\epsilon}(\lambda)-L_{n}(\lambda)\right|=o_{p}(1) .
$$

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 $E \sup _{\lambda \in \Theta}\left\|\varepsilon_{t}^{\epsilon}\right\|^{2}<\infty$. With the same argument, it can be shown that $E \sup _{\lambda \in \Theta} \|\left(\partial \varepsilon_{t}^{\epsilon^{\prime}} /\right.$ $\partial \varphi) \|^{2}<\infty$. Because $D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}$ has a lower bound uniformly for all $\lambda \in \Theta$, we have $E \sup _{\lambda \in \Theta}\left\|\left(\partial \varepsilon_{t}^{\epsilon^{\prime}} / \partial \varphi\right)\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1}\left(\partial \varepsilon_{t}^{\epsilon} / \partial \varphi^{\prime}\right)\right\|<\infty$. Let $c$ be any constant vector with the same dimension as $\varphi$. If $c^{\prime} E\left[\left(\partial \varepsilon_{t}^{\epsilon^{\prime}} / \partial \varphi\right)\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1}\left(\partial \varepsilon_{t}^{\epsilon} / \partial \varphi^{\prime}\right)\right] c=0$, then $c^{\prime}\left(\partial \varepsilon_{t}^{\epsilon^{\prime}} / \partial \varphi\right)\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1 / 2}=0$ a.s., and hence $c^{\prime} \partial \varepsilon_{t}^{\epsilon^{\prime}} / \partial \varphi=0$ a.s. By Lemma 4.2, $c=0$. Thus $E\left[\left(\partial \varepsilon_{t}^{\epsilon^{\prime}} / \partial \varphi\right)\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1}\left(\partial \varepsilon_{t}^{\epsilon} / \partial \varphi^{\prime}\right)\right]>0$. This completes the proof.

Proof of Lemma 5.2. First,

$$
\begin{align*}
\frac{\partial l_{t}^{\epsilon}}{\partial \varphi} & =-\frac{\partial \varepsilon_{t}^{\epsilon^{\prime}}}{\partial \varphi}\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1} \varepsilon_{t}^{\epsilon}-\frac{1}{2} \frac{\partial H_{t}^{\epsilon^{\prime}}}{\partial \varphi} D_{t}^{\epsilon-2} \zeta_{t}  \tag{B.25}\\
\frac{\partial H_{t}^{\epsilon}}{\partial \varphi^{\prime}} & =\left(I_{m}-\sum_{i=1}^{s} B_{i} L^{i}\right)^{-1}\left(\sum_{i=1}^{r} A_{i} L^{i}\right)\left(2 \vec{\varepsilon}_{t}^{*} \frac{\partial \varepsilon_{t}^{\epsilon}}{\partial \varphi^{\prime}}\right)  \tag{B.26}\\
\frac{\partial l_{t}^{\epsilon}}{\partial \delta} & =-\frac{1}{2} \frac{\partial H_{t}^{\epsilon^{\prime}}}{\partial \delta} D_{t}^{\epsilon-2} \zeta_{t}  \tag{B.27}\\
\frac{\partial l_{t}^{\epsilon^{\prime}}}{\partial \sigma} & =-\frac{1}{2} \frac{\partial \operatorname{vec}^{\prime}(\Gamma)}{\partial \sigma} \operatorname{vec}\left(\Gamma^{-1}-\Gamma^{-1} D_{t}^{\epsilon-1} \varepsilon_{t}^{\epsilon} \varepsilon_{t}^{\epsilon^{\prime}} D_{t}^{\epsilon-1} \Gamma^{-1}\right) \tag{B.28}
\end{align*}
$$

where $\vec{\varepsilon}_{t}^{*}=\operatorname{diag}\left(\varepsilon_{1 t}^{\epsilon}, \ldots, \varepsilon_{m t}^{\epsilon}\right), \zeta_{t}=\Pi-\tilde{\eta}_{t}^{\epsilon} \Gamma^{-1} \eta_{t}^{\epsilon}, \Pi=(1, \ldots, 1)_{m \times 1}^{\prime}$, and $\eta_{t}^{\epsilon}$ and $\tilde{\eta}_{t}^{\epsilon}$ are defined as in Lemma 5.3. When $\lambda=\lambda_{0}, \eta_{t}^{\epsilon}=\eta_{0 t}$ and, in this case, we denote $\zeta_{t}$ and $\tilde{\eta}_{t}^{\epsilon}$ by $\zeta_{0 t}$ and $\tilde{\eta}_{0 t}^{\epsilon}$, respectively.

For models (2.10) and (2.11),
$\frac{\partial H_{t}^{\epsilon}}{\partial \mu^{\prime}}=-2 \sum_{i=1}^{r} A_{i} \vec{\varepsilon}_{t-i}^{*}$.
Because $\left|\varepsilon_{j t-i}^{\epsilon 2}\right| \leq h_{j t}^{\epsilon} / \alpha_{i j}$ and $\alpha_{i j} \geq a_{i i j}^{l}>0, j=1, \ldots, m$ and $i=1, \ldots, r$, we have
$\left\|\frac{\partial H_{t}^{\epsilon^{\prime}}}{\partial \mu} D_{t}^{\epsilon-2}\right\| \leq \kappa_{1} \sum_{j=1}^{m} \sum_{i=1}^{r} \frac{\left|\varepsilon_{j t-i}^{\epsilon}\right|}{h_{j t}^{\epsilon}}<\kappa_{2}$,
where $\kappa_{1}$ and $\kappa_{2}$ are some constants independent of $\lambda$. Furthermore, because all the terms in $\partial h_{i t} / \partial \delta$ appear in $h_{i t}^{\epsilon},\left\|\left(\partial H_{t}^{\epsilon^{\prime}} / \partial \delta\right) D_{t}^{\epsilon-2}\right\|<M$, a constant independent of $\lambda$. Because $E \eta_{0 i t}^{4}<\infty$ and $E\left\|\zeta_{0 t}\right\|^{2}<\infty$, it follows that $\Omega_{0}<\infty$.

For models (2.4) and (2.5), because (B.25) and (B.26), $E\left\|\zeta_{0 t}\right\|^{2}<\infty, E\left\|Y_{t}\right\|^{4}<\infty$, Q11 and $D_{0 t}$ has a lower bound, we have
$E\left\|\frac{\partial l_{0 t}^{\epsilon}}{\partial \varphi}\right\|^{2} \leq 2 E\left\|\frac{\partial \varepsilon_{0 t}^{\epsilon^{\prime}}}{\partial \varphi}\right\|^{2}+\frac{1}{2} E\left\|\frac{\partial H_{0 t}^{\epsilon^{\prime}}}{\partial \varphi} D_{0 t}^{-2}\right\|^{2} E\left\|\zeta_{0 t}\right\|^{2}<\infty$.
Similarly, we can show that $E\left\|\partial l_{0 t}^{\epsilon} / \partial \delta\right\|^{2}$ is finite. It is obvious that $E\left\|\partial l_{0 t}^{\epsilon} / \partial \sigma\right\|^{2}<\infty$. Thus, we also have $\Omega_{0}<\infty$. In a similar manner, it can be shown that $\Omega_{0}<\infty$ for models (2.8) and (2.9).

Let $S_{t}=\sum_{t=1}^{n} c^{\prime} \partial l_{0 t}^{\epsilon} / \partial \lambda$, 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, $E S_{n} / n=c^{\prime} E\left[\partial l_{0 t}^{\epsilon} / \partial \lambda \partial l_{0 t}^{\epsilon} / \partial \lambda^{\prime}\right] c>0$. Using the central limit theorem of Stout (1974), $n^{-1 / 2} S_{n}$ converges to $N\left(0, c^{\prime} \Omega_{0} c\right)$ in distribution. Finally, by the Cramér-Wold device, $n^{-1 / 2} \sum_{t=1}^{n} \partial l_{0 t}^{\epsilon} / \partial \lambda$ converges to $N\left(0, \Omega_{0}\right)$ in distribution.

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

Proof of Lemma 5.3. For models (2.10) and (2.11), from the proof of Lemma 5.2, we have shown that
$\sup _{\lambda \in \Theta}\left\|\left(\partial H_{t}^{\epsilon^{\prime}} / \partial \tilde{\lambda}\right) D_{t}^{\epsilon-2}\right\|<C<\infty \quad$ with probability one,
where $C$ is a nonrandom constant. Furthermore,

$$
\sup _{\lambda \in \Theta}\left\|\Delta_{t}^{\epsilon}\right\| \leq \kappa_{1}\left\|\eta_{t}^{\epsilon}\right\|^{2} \leq \kappa_{1}\left\|\varepsilon_{t}^{\epsilon}\right\|^{2} \leq \kappa_{3} \varepsilon_{t}^{* 2}
$$

where $\varepsilon_{t}^{*}$ is defined as in (B.7). Thus, $E \sup _{\lambda \in \Theta} \|\left(\partial H_{t}^{\epsilon^{\prime}} / \partial \tilde{\lambda}\right) D_{t}^{\epsilon-2} \Delta_{t}^{\epsilon} D_{t}^{\epsilon-2}\left(\partial H_{t}^{\epsilon} /\right.$ $\left.\partial \tilde{\lambda}^{\prime}\right) \|<\infty$.

For models (2.8) and (2.9),
$\frac{\partial H_{t}^{\epsilon}}{\partial \varphi^{\prime}}=2 \sum_{i=1}^{r} A_{i} \vec{\varepsilon}_{t-i}^{*} \frac{\partial \varepsilon_{t}^{\epsilon}}{\partial \varphi^{\prime}}$,
where $\vec{\varepsilon}_{t}^{*}$ is defined as in (B.26). Thus, with probability one,
$\left\|\frac{\partial H_{t}^{\epsilon^{\prime}}}{\partial \varphi} D_{t}^{\epsilon-2}\right\| \leq \kappa_{1} \sum_{j=1}^{m} \sum_{i=1}^{r} \frac{\left|\varepsilon_{j t-i}^{\epsilon}\right|}{h_{j t}^{\epsilon}}\left\|\frac{\partial \varepsilon_{t}^{\epsilon^{\prime}}}{\partial \varphi}\right\| \leq \kappa_{2} \sum_{j=1}^{m} \sum_{i=1}^{r}\left\|\frac{\partial \varepsilon_{t-i}^{\epsilon^{\prime}}}{\partial \varphi}\right\|$,
where $\kappa_{1}$ and $\kappa_{2}$ are constants independent of $\lambda$. Because all the components in $\partial H_{t}^{\varepsilon^{\prime}} / \partial \delta$ also appear in $D_{t}^{\epsilon 2}$, we have
$\sup _{\lambda \in \Theta}\left\|\frac{\partial H_{t}^{\epsilon^{\prime}}}{\partial \delta} D_{t}^{\epsilon-2}\right\|<C<\infty$,
where $C$ is a nonrandom constant independent of $\lambda$. By (B.31) and (B.32), it is easy to show that, if $E\left\|Y_{t}\right\|^{4}<\infty, E \sup _{\lambda \in \Theta}\left\|\left(\partial H_{t}^{\epsilon^{\prime}} / \partial \tilde{\lambda}\right) D_{t}^{\epsilon-2} \Delta_{t}^{\epsilon} D_{t}^{\epsilon-2}\left(\partial H_{t}^{\epsilon} / \partial \tilde{\lambda}^{\prime}\right)\right\|<\infty$.
For models (2.4) and (2.5), because $E\left\|Y_{t}\right\|^{6}<\infty$,
$E \sup _{\lambda \in \Theta}\left\|\frac{\partial H_{t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}} D_{t}^{\epsilon-2} \Delta_{t}^{\epsilon} D_{t}^{\epsilon-2} \frac{\partial H_{t}^{\epsilon}}{\partial \tilde{\lambda}^{\prime}}\right\| \leq C E \sup _{\lambda \in \Theta}\left\|\frac{\partial H_{t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}} \Delta_{t}^{\epsilon^{\epsilon}} \frac{\partial H_{t}^{\epsilon}}{\partial \tilde{\lambda}^{\prime}}\right\|<\infty$,
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) and (B.28), we have
$\frac{\partial^{2} l_{t}^{\epsilon}}{\partial \tilde{\lambda} \partial \tilde{\lambda}^{\prime}}=-R_{t}^{(1)}-R_{t}^{(2)}-R_{t}^{(3)}$,
where

$$
\begin{aligned}
R_{t}^{(1)}= & \frac{\partial \varepsilon_{t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}}\left(D_{t}^{\epsilon} \Gamma D_{t}^{\epsilon}\right)^{-1} \frac{\partial \varepsilon_{t}^{\epsilon}}{\partial \tilde{\lambda}^{\prime}}, \quad R_{t}^{(2)}=\frac{1}{4} \frac{\partial H_{\epsilon}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}} D_{t}^{\epsilon-2} \Delta_{t}^{\epsilon} D_{t}^{\epsilon-2} \frac{\partial H_{t}^{\epsilon}}{\partial \tilde{\Lambda}^{\prime}}, \\
R_{t}^{(3)}= & \left(\varepsilon_{t}^{\epsilon^{\prime}} \otimes I_{m}\right) \frac{\partial}{\partial \tilde{\Lambda}^{\prime}} \operatorname{vec}\left[\frac{\partial \varepsilon_{t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}}\left(D_{t}^{\epsilon}\left[D_{t}^{\epsilon}\right)^{-1}\right]\right. \\
& +\left(\zeta_{t}^{\prime} \otimes I_{m}\right) \frac{\partial}{\partial \tilde{\lambda}^{\prime}} \operatorname{vec}\left[\frac{1}{2} \frac{\partial H_{t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}} D_{t}^{\epsilon-2}\right] \\
& -\frac{1}{2} \frac{\partial H_{t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}} D_{t}^{\epsilon-2}\left[\tilde{\eta}_{t}^{s} \Gamma^{-1} D_{t}^{\epsilon-1}+\tilde{\Delta}_{t}^{\epsilon} D_{t}^{\epsilon-1}\right] \frac{\partial \varepsilon_{t}^{\epsilon}}{\partial \tilde{\lambda}^{\prime}},
\end{aligned}
$$

and $\Delta_{t}^{\epsilon}, \tilde{\Delta}_{t}^{\epsilon}$, and $\tilde{\eta}_{t}^{\epsilon}$ 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), $E \sup _{\lambda \in \Theta}\left\|\partial^{2} l_{t}^{\epsilon} / \partial \tilde{\lambda} \partial \tilde{\lambda}^{\prime}\right\|<\infty$. Furthermore,

$$
\begin{aligned}
\frac{\partial^{2} l_{t}^{\epsilon}}{\partial \varphi \partial \sigma^{\prime}} & =\frac{\partial \varepsilon_{t}^{\epsilon^{\prime}}}{\partial \varphi}\left(\varepsilon_{t}^{\epsilon^{\prime}} D_{t}^{\epsilon-1} \Gamma^{-1} \otimes D_{t}^{\epsilon-1} \Gamma^{-1}\right) \mathcal{K}-\frac{1}{2} \frac{\partial H_{t}^{\epsilon^{\prime}}}{\partial \varphi} D_{t}^{\epsilon-2} \frac{\partial \zeta_{t}}{\partial \sigma^{\prime}} \\
\frac{\partial^{2} l_{t}^{\epsilon}}{\partial \delta \partial \sigma^{\prime}}= & -\frac{1}{2} \frac{\partial H_{t}^{\epsilon^{\prime}}}{\partial \delta} D_{t}^{\epsilon-2} \frac{\partial \zeta_{t}}{\partial \sigma^{\prime}}, \\
\frac{\partial \zeta_{t}}{\partial \sigma^{\prime}}= & \left(\eta_{t}^{\epsilon^{\prime}} \Gamma^{-1} \otimes \tilde{\eta}_{t}^{\epsilon}\right)\left(I_{m} \otimes \Gamma^{-1}\right) \mathcal{K} \\
\frac{\partial^{2} l_{t}^{\epsilon}}{\partial \sigma \partial \sigma^{\prime}}= & \frac{1}{2} \mathcal{K}^{\prime}\left(\Gamma^{-1} \otimes I_{m}\right)\left[I_{m}-\left(\Gamma^{-1} D_{t}^{\epsilon-1} \varepsilon_{t}^{\epsilon} \varepsilon_{t}^{\epsilon^{\prime}} D_{t}^{\epsilon-1} \otimes I_{m}\right)\right. \\
& \left.\quad\left(I_{m} \otimes \Gamma^{-1} D_{t}^{\epsilon-1} \varepsilon_{t}^{\epsilon} \varepsilon_{t}^{\epsilon^{\prime}} D_{t}^{\epsilon-1}\right)\right]\left(I_{m} \otimes \Gamma^{-1}\right) \mathcal{K} .
\end{aligned}
$$

In a similar manner, it is straightforward to show that $E \sup _{\lambda \in \Theta}\left\|\partial^{2} l_{t}^{\epsilon} / \partial \varphi \partial \sigma^{\prime}\right\|<\infty$, $E \sup _{\lambda \in \Theta}\left\|\partial^{2} l_{t}^{\epsilon} / \partial \delta \partial \sigma^{\prime}\right\|<\infty$, and $E \sup _{\lambda \in \Theta}\left\|\partial^{2} l_{t}^{\epsilon} / \partial \sigma \partial \sigma^{\prime}\right\|<\infty$. Finally, by the triangle inequality, we can show that $E \sup _{\lambda \in \Theta}\left\|\partial^{2} l_{t}^{\epsilon} / \partial \lambda \partial \lambda^{\prime}\right\|<\infty$. 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 $\left\|\Sigma_{0}\right\|<\infty$. By Lemma 5.4, we have $\Sigma_{n}=\Sigma_{0}+o_{p}(1)$.

Let $c$ be a constant vector with the same dimension as $\delta$. If $c^{\prime} E\left[\partial H_{0 t}^{\epsilon^{\prime}} / \partial \delta D_{0 t}^{-4} \partial\right.$ $\left.H_{0 t}^{\epsilon} / \partial \delta^{\prime}\right] c=0$, then $c^{\prime}\left(\partial H_{0 t}^{\epsilon^{\prime}} / \partial \delta\right) D_{0 t}^{-2}=0$ and hence $c^{\prime} \partial H_{0 t}^{\epsilon^{\prime}} / \partial \delta=0$. By Lemma 4.2, $c=0$. Thus, $E\left[\partial H_{0 t}^{\epsilon^{\prime}} / \partial \delta D_{0 t}^{-4} \partial H_{0 t}^{\epsilon} / \partial \delta^{\prime}\right]>0$.

Denote
$\Sigma_{\delta 0}=E\left[\left(\begin{array}{cc}\frac{1}{2} \frac{\partial H_{0 t}^{\epsilon_{0 t}^{\prime}}}{\partial \delta} D_{0 t}^{-2} & 0 \\ 0 & P^{\prime}\end{array}\right)\left(\begin{array}{cc}C & C_{1} \\ C_{1}^{\prime} & I_{m^{2}} / 2\end{array}\right)\left(\begin{array}{cc}\frac{1}{2} D_{0 t}^{-2} \frac{\partial H_{0 t}^{\epsilon}}{\partial \delta^{\prime}} & 0 \\ 0 & P\end{array}\right)\right]$.
By the condition given, $C \geq 2 I_{m}$. Thus, it is easy to show that $\left(\begin{array}{cc}C & C_{1} \\ C_{\mathrm{i}}^{\prime} & I_{m^{2}} / 2\end{array}\right)$ is positive by Theorem 14.8.5 in Harville (1997). Because $P^{\prime} P=\mathcal{K}^{\prime}\left(\Gamma_{0}^{-1} \otimes \Gamma_{0}^{-1}\right) \mathcal{K}$ and $E\left[\partial H_{0 t}^{\epsilon^{\prime}} / \partial \delta D_{0 t}^{-4} \partial H_{0 t}^{\epsilon} / \partial \delta^{\prime}\right]$ are positive, we know that $\Sigma_{\delta 0}$ is positive.
$-\Sigma_{0}=\left(\begin{array}{cc}E\left[\frac{\partial \varepsilon_{0 t}^{\prime}}{\partial \varphi}\left(D_{0 t} \Gamma_{0} D_{0 t}\right)^{-1} \frac{\partial \varepsilon_{0 t}}{\partial \varphi^{\prime}}\right] & 0 \\ 0 & 0\end{array}\right)+\left(\begin{array}{cc}\Sigma_{\varphi 0} & \Sigma_{\varphi \delta \sigma 0} \\ \Sigma_{\varphi \delta \sigma 0}^{\prime} & \Sigma_{\delta 0}\end{array}\right)$,
where $\Sigma_{\varphi 0}=E\left[\left(\partial H_{0 t}^{\epsilon^{\prime}} / \partial \varphi\right) D_{0 t}^{-2} C D_{0 t}^{-2}\left(\partial H_{0 t}^{\epsilon} / \partial \varphi^{\prime}\right)\right] / 4, \Sigma_{\varphi \delta \sigma 0}=\left(\Sigma_{\varphi \delta 0}, \Sigma_{\varphi \sigma 0}\right), \Sigma_{\varphi \delta 0}=$ $E\left[\left(\partial H_{0 t}^{\epsilon^{\prime}} / \partial \varphi\right) D_{0 t}^{-2} C D_{0 t}^{-2}\left(\partial H_{0 t}^{\epsilon} / \partial \delta^{\prime}\right)\right] / 4$, and $\Sigma_{\varphi \sigma 0}=E\left[\left(\partial H_{0 t}^{\epsilon^{\prime}} / \partial \varphi\right) D_{0 t}^{-2}\right] C_{1} P / 2$. Let $c=$ $\left(c_{1}^{\prime}, c_{2}^{\prime}\right)^{\prime}$ be any constant vector with the same dimension as $\lambda$ and let $c_{1}$ have the same dimension as $\varphi$, that is, $m+(p+q) m^{2}$ for models (2.4) and (2.5) and (2.8) and (2.9) and $m$ for models (2.10) and (2.11). If $-c^{\prime} \Sigma_{0} c=0$, then $c_{1}^{\prime} E\left[\left(\partial \varepsilon_{0 t}^{\prime} / \partial \varphi\right)\left(D_{0 t} \Gamma_{0} D_{0 t}\right)^{-1}\right.$ $\left.\left(\partial \varepsilon_{0 t} / \partial \varphi^{\prime}\right)\right] c_{1}=0$. By Lemma 5.1, $c_{1}=0$. Thus, $c_{2}^{\prime} \Sigma_{\delta 0} c_{2}=0$. As we have shown that $\Sigma_{\delta 0}$ 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) and (2.5). The proofs for models (2.8) and (2.9) and models (2.10) and (2.11) are similar, except that (B.29) and (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

$$
\begin{equation*}
\left\|\frac{\partial \varepsilon_{t}^{\epsilon}}{\partial \varphi}\right\| \leq c_{2}+c_{3} \sum_{i=1}^{\infty} \rho_{1}^{i}\left\|Y_{t-i}\right\| \equiv X_{1 t} . \tag{B.34}
\end{equation*}
$$

Because $X_{1 t}$ is a strictly stationary time series with $E X_{1 t}^{2}<\infty$, we have (see Chung, 1968, p. 93)
$\frac{1}{\sqrt{n}} \max _{1 \leq t \leq n} \sup _{\lambda \in \Theta}\left\|\frac{\partial \varepsilon_{t}^{\epsilon}}{\partial \varphi}\right\|=o_{p}(1)$.

By (B.5), (B.7), (B.8), and (B.26), it follows that
$\sup _{\lambda \in \Theta}\left\|\frac{\partial H_{t}^{\epsilon}}{\partial \tilde{\lambda}}\right\| \leq c_{4}+c_{5} \sum_{i=1}^{\infty} \rho_{2}^{i}\left\|Y_{t-i}\right\|^{2} \equiv X_{2 t}$.
Because $X_{2 t}$ is a strictly stationary time series with $E X_{2 t}^{2}<\infty$, we have

$$
\begin{equation*}
\frac{1}{\sqrt{n}} \max _{1 \leq t \leq n} \sup _{\lambda \in \Theta}\left\|\frac{\partial H_{t}^{\epsilon}}{\partial \tilde{\lambda}}\right\|=o_{p}(1) . \tag{B.37}
\end{equation*}
$$

In the following discussion, $\zeta_{t}$ 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 $\eta_{n t}^{\epsilon}, \tilde{\eta}_{n t}^{\epsilon}, \zeta_{n t}$, and $D_{n t}^{\epsilon}$, respectively, when $\lambda=$ $\lambda_{n}$. By (B.35) and (B.37),

$$
\begin{align*}
\left|\eta_{n i t}^{\epsilon}-\eta_{0 i t}\right| \leq & \left|\varepsilon_{n i t}^{\epsilon}-\varepsilon_{0 i t}\right| \frac{1}{h_{n i t}^{\epsilon 1 / 2}}+\left|h_{n i t}^{\epsilon \in / 2}-h_{0 i t}^{1 / 2}\right| \frac{\left|\varepsilon_{0 i t}\right|}{h_{0 i t}^{1 / 2} h_{n i t}^{\epsilon 1 / 2}} \\
\leq & \left\|\sqrt{n}\left(\tilde{\lambda}_{n}-\tilde{\lambda}_{0}\right)\right\|\left[\frac{1}{h_{n i t}^{\epsilon 1 / 2} \sqrt{n}} \max _{1 \leq t \leq n}\left(\left.\left\|\frac{\partial \varepsilon_{t}^{\epsilon}}{\partial \tilde{\lambda}}\right\|\right|_{\lambda_{1 n}^{*}}\right)\right. \\
& +\frac{\left|\varepsilon_{0 i t}\right|}{h_{0 i t}^{1 / 2} h_{n i t}^{\epsilon!/ 2}} \left\lvert\, \frac{1}{\sqrt{n}} \max _{1 \leq t \leq n}\left(\left.\left\|\frac{1}{h_{i t}^{\epsilon 1 / 2}} \frac{\partial h_{i t}^{\epsilon}}{\partial \tilde{\lambda}}\right\|\right|_{\lambda_{2 n}^{*}}\right)\right. \\
= & o_{p}(1)+o_{p}(1)\left|\eta_{0 i t}\right|, \tag{B.38}
\end{align*}
$$

where $o_{p}(1)$ holds uniformly in all $t, i=1, \ldots, m$, and $\lambda_{1 n}^{*}$ and $\lambda_{2 n}^{*}$ lie between $\lambda_{0}$ and $\lambda_{n}$. From (B.38), we have

$$
\begin{align*}
\left\|\zeta_{n t}-\zeta_{0 t}\right\|= & \left\|\tilde{\eta}_{n t}^{\epsilon} \Gamma_{n}^{-1} \eta_{n t}^{\epsilon}-\tilde{\eta}_{0 t} \Gamma_{0}^{-1} \eta_{0 t}\right\| \\
\leq & \left\|\tilde{\eta}_{n t}^{\epsilon}\right\|\left\|\tilde{\eta}_{0 t}\right\|\left\|\Gamma_{n}^{-1}-\Gamma_{0}^{-1}\right\| \\
& +2\left\|\tilde{\eta}_{n t}^{\epsilon}-\tilde{\eta}_{0 t}\right\|\left\|\Gamma_{0}^{-1} \eta_{0 t}\right\|+\left\|\tilde{\eta}_{n t}^{\epsilon}-\tilde{\eta}_{0 t}\right\|^{2}\left\|\Gamma_{0}^{-1}\right\| \\
= & o_{p}(1)+o_{p}(1)\left\|\eta_{0 t}\right\|^{2}, \tag{B.39}
\end{align*}
$$

where $o_{p}(1)$ holds uniformly in all $t$. By (B.37),

$$
\begin{align*}
\max _{1 \leq t \leq n}\left|h_{n i t}^{\epsilon-1}-h_{0 i t}^{-1}\right| & =\left\|\sqrt{n}\left(\tilde{\lambda}_{n}-\tilde{\lambda}_{0}\right)\right\| \frac{1}{\sqrt{n}} \max _{1 \leq t \leq n}\left(\left\|\frac{1}{h_{i t}^{\epsilon 2}} \frac{\partial h_{i t}^{\epsilon}}{\partial \tilde{\lambda}}\right\|_{\lambda_{3 n}^{*}}\right) \\
& =o_{p}(1) \tag{B.40}
\end{align*}
$$

where $\lambda_{3 n}^{*}$ lies in between $\lambda_{0}$ and $\lambda_{n}$. By (B.39) and (B.40),

$$
\begin{align*}
\left\|D_{n t}^{\epsilon-2} \zeta_{n t}-D_{0 t}^{-2} \zeta_{0 t}\right\| & \leq\left\|D_{n t}^{\epsilon-2}-D_{0 t}^{-2}\right\|\left\|\zeta_{0 t}\right\|+\left\|D_{n t}^{\epsilon-2}\right\|\left\|\zeta_{n t}-\xi_{0 t}\right\| \\
& =o_{p}(1)+o_{p}(1)\left\|\eta_{0 t}\right\|^{2} . \tag{B.41}
\end{align*}
$$

By (B.41),

$$
\begin{align*}
& \left\|D_{n t}^{\epsilon-2} \zeta_{n t} \zeta_{n t}^{\prime} D_{n t}^{\epsilon-2}-D_{0 t}^{-2} \zeta_{0 t} \zeta_{0 t}^{\prime} D_{0 t}^{-2}\right\| \\
& \quad \leq 2\left\|D_{n t}^{\epsilon-2} \zeta_{n t}-D_{0 t}^{-2} \zeta_{t}\right\|\left\|D_{0 t}^{-2} \zeta_{0 t}\right\|+\left\|D_{n t}^{\epsilon-2} \zeta_{n t}-D_{0 t}^{-2} \zeta_{t}\right\|^{2} \\
& \quad=o_{p}(1)+o_{p}(1)\left\|\eta_{0 t}^{4}\right\| . \tag{B.42}
\end{align*}
$$

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

$$
\begin{equation*}
\sup _{\lambda \in \Theta}\left\|\frac{\partial^{2} h_{i t}^{\epsilon}}{\partial \tilde{\lambda} \partial \tilde{\lambda}}\right\| \leq c_{6}+c_{7} \sum_{j=1}^{\infty} \rho_{3}^{j}\left\|Y_{t-j}\right\|^{2} \equiv X_{3 i t}, \tag{B.43}
\end{equation*}
$$

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

$$
\begin{align*}
&\left\|\frac{\partial H_{n t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}} D_{n t}^{\epsilon-2} \zeta_{n t} \zeta_{n t}^{\prime} D_{n t}^{\epsilon-2} \frac{\partial H_{n t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}}-\frac{\partial H_{0 t}^{\prime}}{\partial \tilde{\lambda}} D_{0 t}^{-2} \zeta_{0 t} \zeta_{0 t}^{\prime} D_{0 t}^{-2} \frac{\partial H_{0 t}^{\prime}}{\partial \tilde{\lambda}}\right\| \\
& \leq\left\|\frac{\partial H_{n t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}} D_{0 t}^{-2} \zeta_{0 t} \zeta_{0 t}^{\prime} D_{0 t}^{-2} \frac{\partial H_{n t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}}-\frac{\partial H_{0 t}^{\prime}}{\partial \tilde{\lambda}} D_{0 t}^{-2} \zeta_{0 t} \zeta_{0 t}^{\prime} D_{0 t}^{-2} \frac{\partial H_{0 t}^{\prime}}{\partial \tilde{\lambda}}\right\| \\
&+\left\|\frac{\partial H_{n t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}}\right\|^{2}\left[o_{p}(1)+o_{p}(1)\left\|\eta_{0 t}\right\|^{4}\right] \\
& \leq 2\left\|\frac{\partial H_{n t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}}-\frac{\partial H_{0 t}^{\prime}}{\partial \tilde{\lambda}}\right\|\left\|\frac{\partial H_{0 t}^{\prime} \|}{\partial \tilde{\lambda}}\right\| D_{0 t}^{-2} \zeta_{0 t} \zeta_{0 t}^{\prime} D_{0 t}^{-2} \| \\
&+\left\|\frac{\partial H_{n t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}}-\frac{\partial H_{0 t}^{\prime}}{\partial \tilde{\lambda}}\right\|^{2}\left\|D_{0 t}^{-2} \zeta_{0 t} \zeta_{0 t}^{\prime} D_{0 t}^{-2}\right\|+\left\|\frac{\partial H_{n t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}}\right\|^{2}\left[o_{p}(1)+o_{p}(1)\left\|\eta_{0 t}\right\|^{4}\right] \\
& \leq \frac{1}{\sqrt{n}} O_{p}(1)\left\|\sqrt{n}\left(\tilde{\lambda}_{n}-\tilde{\lambda}_{0}\right) \sum_{i=1}^{m} X_{3 i t}\right\| \frac{\partial H_{0 t}^{\prime}}{\partial \tilde{\lambda}}\| \| y \zeta_{0 t} \|^{2} \\
&+\frac{1}{\sqrt{n}}\left\|\sqrt{n}\left(\tilde{\lambda}_{n}-\tilde{\lambda}_{0}\right)\right\|^{2}\left(\sum_{i=1}^{m} X_{3 i t}\right)^{2}\left\|\zeta_{0 t}\right\|^{2}+X_{2 t}^{2}\left(1+\left\|\eta_{0 t}\right\|^{4}\right) o_{p}(1) \\
&= o_{p}(1)\left[\sum_{i=1}^{m} X_{3 i t}\left\|\frac{\partial H_{0 t}^{\prime} \|}{\partial \tilde{\lambda}}\right\|+\left(\sum_{i=1}^{m} X_{3 i t}\right)^{2}+X_{2 t}^{2}\right]\left(1+\left\|\eta_{0 t}\right\|^{4}\right) \\
& \equiv o_{p}(1) X_{t}^{*}\left(1+\left\|\eta_{0 t}\right\|^{4}\right), \tag{B.44}
\end{align*}
$$

where $O_{p}(1)$ and $o_{p}(1)$ hold uniformly in all $t$. Note that $X_{t}^{*}\left(1+\left\|\eta_{0 t}\right\|^{4}\right)$ is strictly stationary, with $E\left[X_{t}^{*}\left(1+\left\|\eta_{0 t}\right\|^{4}\right)\right]=E X_{t}^{*} E\left(1+\left\|\eta_{0 t}\right\|^{4}\right)<\infty$. By the ergodic theorem, we have $n^{-1} \sum_{t=1}^{n} X_{t}^{*}\left(1+\left\|\eta_{0 t}\right\|^{4}\right)=O_{p}(1)$. Thus, by (B.44), we have
$\frac{1}{n} \sum_{t=1}^{n}\left\|\frac{\partial H_{n t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}} D_{n t}^{\epsilon-2} \zeta_{n t} \zeta_{n t}^{\prime} D_{n t}^{\epsilon-2} \frac{\partial H_{n t}^{\epsilon^{\prime}}}{\partial \tilde{\lambda}}-\frac{\partial H_{0 t}^{\prime}}{\partial \tilde{\lambda}} D_{0 t}^{-2} \zeta_{0 t} \zeta_{0 t}^{\prime} D_{0 t}^{-2} \frac{\partial H_{0 t}}{\partial \tilde{\lambda}^{\prime}}\right\|=o_{p}(1)$.
(B.45)

Similarly, we can show that

$$
\begin{align*}
\frac{1}{n} \sum_{t=1}^{n} \| & \frac{\partial \varepsilon_{n t}^{\epsilon^{\prime}}}{\partial \varphi}\left(D_{n t}^{\epsilon} \Gamma D_{n t}^{\epsilon}\right)^{-1} \varepsilon_{n t}^{\epsilon} \varepsilon_{n t}^{\epsilon^{\prime}}\left(D_{n t}^{\epsilon} \Gamma D_{n t}^{\epsilon}\right)^{-1} \frac{\partial \varepsilon_{n t}^{\epsilon^{\prime}}}{\partial \varphi} \\
& -\frac{\partial \varepsilon_{0 t}^{\prime}}{\partial \varphi}\left(D_{0 t} \Gamma_{0} D_{0 t}\right)^{-1} \varepsilon_{0 t} \varepsilon_{0 t}^{\prime}\left(D_{0 t} \Gamma D_{0 t}\right)^{-1} \frac{\partial \varepsilon_{0 t}}{\partial \varphi^{\prime}} \|=o_{p}(1) \tag{B.46}
\end{align*}
$$

and
$\frac{1}{n} \sum_{t=1}^{n}\left\|\frac{\partial l_{n t}^{\epsilon}}{\partial \sigma} \frac{\partial l_{n t}^{\epsilon}}{\partial \sigma^{\prime}}-\frac{\partial l_{0 t}^{\epsilon}}{\partial \sigma} \frac{\partial l_{0 t}^{\epsilon}}{\partial \sigma^{\prime}}\right\|=o_{p}(1)$.
Thus, by (B.45)-(B.47) and the triangle inequality, we can show that
$\frac{1}{n} \sum_{t=1}^{n}\left\|\frac{\partial l_{n t}^{\epsilon}}{\partial \lambda} \frac{\partial l_{n t}^{\epsilon}}{\partial \lambda^{\prime}}-\frac{\partial l_{0 t}^{\epsilon}}{\partial \lambda} \frac{\partial l_{0 t}^{\epsilon}}{\partial \lambda^{\prime}}\right\|=o_{p}(1)$.
Thus, (a) holds. In a similar manner to the proof of Lemma 4.6, we can show that
$\frac{1}{n} \sum_{t=1}^{n}\left\|\frac{\partial l_{n t}^{\epsilon}}{\partial \lambda} \frac{\partial l_{n t}^{\epsilon}}{\partial \lambda^{\prime}}-\frac{\partial l_{n t}}{\partial \lambda} \frac{\partial l_{n t}}{\partial \lambda^{\prime}}\right\|=o_{p}(1)$.
Note that $\left(\partial l_{0 t}^{\epsilon} / \partial \lambda\right)\left(\partial l_{0 t}^{\epsilon} / \partial \lambda^{\prime}\right)$ is strictly stationary and ergodic with $E \|\left(\partial l_{0 t}^{\epsilon} / \partial \lambda\right)$ $\left(\partial l_{0 t}^{\epsilon} / \partial \lambda^{\prime}\right) \|<\infty$. By the ergodic theorem, we have $n^{-1} \sum_{t=1}^{n}\left\|\left(\partial l_{0 t}^{\epsilon} / \partial \lambda\right)\left(\partial l_{0 t}^{\epsilon} / \partial \lambda^{\prime}\right)\right\|=$ $\Omega_{0}+o_{p}(1)$. Furthermore, by (B.48) and (B.49), (b) holds. This completes the proof.

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, $n^{-1} \sum_{t=1}^{n}\left(\partial l_{t}^{2} / \partial \lambda \partial \lambda^{\prime}\right)$ exists and is continuous in $\Theta$. Third, by Lemmas 5.4 and 5.5 , we can immediately obtain that $n^{-1} \sum_{t=1}^{n}\left(\partial l_{n t}^{2} / \partial \lambda \partial \lambda^{\prime}\right)$ converges to $\Sigma_{0}>0$ for any sequence $\lambda_{n}$, such that $\lambda_{n} \rightarrow \lambda_{0}$ in probability. Fourth, by Lemma 5.2, $n^{-1 / 2} \sum_{t=1}^{n}\left(\partial l_{0 t} / \partial \lambda\right)$ converges to $N\left(0, \Omega_{0}\right)$ in distribution. Thus, we have established all the conditions in Theorem 4.1.3 in Amemiya (1985), and hence $n^{1 / 2}\left(\hat{\lambda}_{n}-\lambda_{0}\right)$ converges to $N\left(0, \Sigma_{0}^{-1} \Omega_{0} \Sigma_{0}^{-1}\right)$. Finally, by Lemmas 5.5 and 5.6, $\Sigma_{0}$ and $\Omega_{0}$ can be estimated consistently by $\hat{\Sigma}_{n}$ and $\hat{\Omega}_{n}$, respectively. This completes the proof.


[^0]:    The authors thank the co-Editor, Bruce Hansen, and two referees for very helpful comments and suggestions and acknowledge the financial support of the Australian Research Council. Address correspondence to: Shiqing Ling, e-mail: sling@ecel.uwa.edu.au.

