

**ON ADAPTIVE ESTIMATION IN NONSTATIONARY
ARMA MODELS WITH GARCH ERRORS¹**

BY SHIQING LING AND MICHAEL MCALEER

*Hong Kong University of Science and Technology
and University of Western Australia*

This paper considers adaptive estimation in nonstationary autoregressive moving average models with the noise sequence satisfying a generalized autoregressive conditional heteroscedastic process. The locally asymptotic quadratic form of the log-likelihood ratio for the model is obtained. It is shown that the limit experiment is neither LAN nor LAMN, but is instead LABF. For the model with symmetric density of the rescaled error, a new efficiency criterion is established for a class of defined M_V -estimators. It is shown that such efficient estimators can be constructed when the density is known. Using the kernel estimator for the score function, adaptive estimators are constructed when the density of the rescaled error is symmetric, and it is shown that the adaptive procedure for the parameters in the conditional mean part uses the full sample without splitting. These estimators are demonstrated to be asymptotically efficient in the class of M_V -estimators. The paper includes the results that the stationary ARMA–GARCH model is LAN, and that the parameters in the model with symmetric density of the rescaled error are adaptively estimable after a reparameterization of the GARCH process. This paper also establishes the locally asymptotic quadratic form of the log-likelihood ratio for nonlinear time series models with ARCH-type errors.

1. Introduction. Suppose that the observations y_1, \dots, y_n , are generated by the autoregressive moving average (ARMA) model with errors generated by the generalized autoregressive conditional heteroscedasticity (GARCH) process,

$$(1.1) \quad y_t = \sum_{i=1}^p \varphi_{0i} y_{t-i} + \sum_{i=1}^q \psi_{0i} \varepsilon_{t-i} + \varepsilon_t,$$

$$(1.2) \quad \varepsilon_t = \eta_t \sqrt{h_t}, \quad h_t = \alpha_{00} + \sum_{i=1}^r u_{0i} \varepsilon_{t-i}^2 + \sum_{i=1}^s v_{0i} h_{t-i},$$

where η_t is a sequence of independent and identically distributed (i.i.d.) random variables, with mean zero, variance one and a common density f ; and $\alpha_{00} > 0$, u_{01}, \dots, u_{0r} , $v_{01}, \dots, v_{0s} \geq 0$. Model (1.1), (1.2) is called the non-

Received January 1999; revised March 2002.

¹Supported in part by the Australian Research Council.

AMS 2000 subject classifications. Primary 62M10, 62E20; secondary 60F17.

Key words and phrases. Adaptive estimation, efficient estimation, nonstationary ARMA–GARCH models, kernel estimators, limiting distribution, locally asymptotic quadratic, log-likelihood ratio.

stationary ARMA–GARCH model if the characteristic polynomial $\varphi_0(z) = 1 - \sum_{i=1}^p \varphi_{0i} z^i$ has one unit root taking the value $+1$, with other roots lying outside the unit circle.

In the traditional ARMA model, the errors ε_t are assumed to be i.i.d. Common time series practice has provided substantial evidence that these assumptions are usually inadequate. For example, the conditional variance of the errors may contain much useful information. Engle (1982) proposed the autoregressive conditional heteroscedasticity (ARCH) model; that is, model (1.2) with $s = 0$, which can capture such information. Subsequently, Bollerslev (1986) generalized the ARCH model to the popular GARCH model (1.2). This is a very important class of time series models and has been widely investigated and applied in the finance and econometric literature [see the surveys by Bollerslev, Engle and Nelson (1994), and Li, Ling and McAleer (2002)]. For ARCH-type time series, there are already some theoretical results for the quasi-maximum likelihood estimator (QMLE) in Weiss (1986) and Ling and Li (1997, 1998). However, when η_t is not normal, the QMLE is not efficient.

For various models with i.i.d. nonnormal errors, much effort has been expended in obtaining efficient estimators. Such efficiency can usually be achieved by adaptive estimation. A comprehensive account of the theory and method can be found in Bickel (1982) and Bickel, Klaassen, Ritov and Wellner (1993) (henceforth BKRW), with a valuable survey in Robinson (1988). In the time series context, Kreiss (1987a) investigated the stationary ARMA model, and proved the locally asymptotic normality (LAN) of the model and constructed adaptive estimators. Unlike Bickel (1982), Kreiss' adaptive procedure uses full samples without splitting and hence is quite useful in practical applications [see also Kreiss (1987b)]. Koul and Schick (1997) developed a general theoretical framework for nonlinear AR models with i.i.d. errors, clearly discussed the efficiency and adaptivity, and especially showed that Stein's necessary condition can be satisfied in some models with asymmetric errors. They also investigated several methods of constructing efficient estimators.

Recently, several authors have examined efficient estimation for ARCH-type time series. Engle and Gonzalez-Rivera (1991) proposed a semiparametric estimator for models (1.1) and (1.2) without a unit root and argued, through simulation, that the semiparametric approach does not seem to capture the total potential gain in efficiency. Linton (1993) considered adaptive estimation for the fixed design regression with ARCH errors. Koul and Schick (1996) investigated adaptive estimation for a random coefficient AR model, which is an ARCH-type time series model. Jeganathan (1995) and Drost, Klaassen and Werker (1997) (henceforth DKW) developed general frameworks suitable for stationary ARCH-type times series. However, apart from the simple ARCH model in DKW (1997) and the GARCH(1,1) model in Drost and Klaassen (1997), these conditions have not been established for the general-order GARCH model or the stationary ARMA–GARCH model. As Drost and Klaassen (1997) argued, greater technical

details may be required for more general cases. These general stationary GARCH and ARMA–GARCH models are included in this paper as special cases.

The above authors considered only stationary time series. There is a growing interest in efficient estimation for nonstationary time series [see, e.g., Koul and Pflug (1990), Philips (1991), Elliott, Rothenberg and Stock (1996) and Jeganathan (1997)]. Jeganathan (1995) developed a general framework for nonstationary time series models, specifically, a complete optimal inference procedure for nonstationary time series with i.i.d. errors.

In this paper, we discuss adaptive estimation for the nonstationary ARMA–GARCH model [(1.1), (1.2)], where we allow the ARMA model to have at most one unit root. We generalize the frameworks in Jeganathan (1995), DKW (1997) and Koul and Schick (1997). Under this framework, the locally asymptotic quadratic (LAQ) form of the log-likelihood ratio for the model is obtained. It is shown that the limit experiment is neither LAN nor locally asymptotic mixed normal (LAMN), but is instead the locally asymptotically Brownian functional (LABF) defined in Jeganathan (1995). For the nonstationary ARMA–GARCH model, the definition on efficient estimators given in Fabian and Hannan (1982) is inappropriate. We define efficient estimators in a class of M_v -estimators and present a new efficiency criterion for the model with symmetric density f . It is shown that such efficient estimators can be constructed when f is known. Using the kernel estimator for the score function, adaptive estimators are constructed for the model with unknown symmetric density f . It is shown that these estimators are asymptotically efficient in the class of M_v -estimators. In DKW (1997), the split sample method proposed by Schick (1986) is used for all the adaptively estimable parameters. In contrast, our adaptive estimation of the parameters in the ARMA part uses the full sample without splitting and hence may be more useful in practice.

Our adaptive estimation for the ARMA part depends heavily on the symmetry assumption. Without this assumption, some different methods of constructing adaptive estimates were given in Kreiss (1987b), DKW (1997) and Koul and Schick (1997) for the stationary ARMA model with i.i.d. errors. The research in this paper can be considered as a first step in exploring optimal inference problems in nonstationary time series with ARCH errors. Along this route, similar theories and methods can be developed for the nonstationary ARMA model with alternative ARCH-type errors, such as E-GARCH and threshold ARCH, among many others. Another important extension is towards cointegrating time series with multivariate ARCH-type errors.

This paper proceeds as follows. Section 2 presents a general framework for the LAQ. Section 3 obtains the LABF form of the log-likelihood ratio, and discusses adaptivity and efficiency for the nonstationary ARMA–GARCH model. Section 4 develops the efficient and adaptive estimators. Sections 5 and 6 provide the proofs of the main theorems. Throughout this paper, we will use the following notation: B' denotes the transpose of the vector B ; $o(1)$ [$O(1)$] denotes a series of numbers

converging to zero (being bounded); $o_\lambda(1)$ [$O_\lambda(1)$] denotes a series of random numbers converging to zero (being bounded) in $P_{\lambda,f}$ -probability; $P_{\lambda,f}$ and E_{λ_0} are abbreviated as P_λ and E , respectively; $\|\cdot\|$ denotes the Euclidean norm; and $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution.

2. A general LAQ criterion. In this section, we present a general LAQ criterion which is a generalization of the criteria in Jeganathan (1995), DKW (1997) and Koul and Schick (1997). Our discussion follows the fashion of Koul and Schick (1997).

Let \mathcal{D} be a class of Lebesgue densities, Θ be an open subset of the k -dimensional real space \mathbb{R}^k , and $\mathcal{B} = \{P_{\lambda,\chi} : (\lambda, \chi) \in \Theta \times \mathcal{D}\}$ be a family of probability measures, y_1, y_2, \dots, y_n be observable random variables, Y_0 be a $p_0 \times 1$ initial (unobservable) vector, and $Z_{t-1}(\lambda) = Z_{t-1}(\bar{Y}_{t-1}, \lambda)$ and $h_t(\lambda) = h_t(\bar{Y}_{t-1}, \lambda)$ be measurable functions of the variables \bar{Y}_{t-1} and λ , where $\bar{Y}_t = (Y_0, y_1, \dots, y_t)$ and $\lambda \in \Theta$. Suppose that, under $P_{\lambda,\chi}$, Y_0 has a Lebesgue density $q_{\lambda,\chi}$ and the time series y_t have the following structure:

$$(2.1) \quad \eta_t(\lambda) = [y_t - Z_{t-1}(\lambda)]/\sqrt{h_t(\lambda)}, \quad t = 1, 2, \dots,$$

where the rescaled errors $\eta_1(\lambda), \eta_2(\lambda), \dots$ are i.i.d. with density $\chi \in \mathcal{D}$ and independent of Y_0 , and the true parameter is (λ_0, f) .

For the nonstationary AR model with i.i.d. errors, the LAQ form of the log-likelihood ratio (LR) was given in Jeganathan (1995). However, he did not accommodate the perturbation of the unknown density and whether or not the parameters in the nonstationary AR model are adaptively estimable. By parameterizing the density, Koul and Schick (1996, 1997) gave some clear explanations as to the adaptivity of the parameters in the random AR and nonlinear AR models. This technique requiring the parameterization of densities is discussed carefully in BKRW (1993). As in Koul and Schick (1996, 1997), we introduce the following definition.

DEFINITION 2.1. Let $c \rightarrow f_c$ be a map from a neighborhood Δ of the origin in \mathbb{R}^l into \mathcal{D} such that $f_0 = f$. We say that $c \rightarrow f_c$ is a *regular* path if there exists a measurable function ζ from \mathbb{R} to \mathbb{R}^l such that $\int \|\zeta(x)\|^2 f(x) dx < \infty$, $\int \zeta(x)\zeta'(x)f(x) dx$ is nonsingular, and

$$\int [\sqrt{f_c(x)} - \sqrt{f(x)} - \frac{1}{2}c'\zeta(x)\sqrt{f(x)}]^2 dx = o(\|c\|^2).$$

Let $P_{\lambda,n}^c$ be the restriction of P_{λ,f_c} to \mathcal{F}_n , the σ -field generated by $\{Y_0, y_1, \dots, y_n\}$. Denote $P_{\lambda,n}^0$ by $P_{\lambda,n}$. Define $\Lambda_n(\lambda_1, \lambda_2, c)$ as the log-LR of $P_{\lambda_2,n}^c$ to $P_{\lambda_1,n}$:

$$\Lambda_n(\lambda_1, \lambda_2, c) = 2 \sum_{t=1}^n \left[\log \frac{s_{c,t}(\lambda_2)}{s_t(\lambda_1)} \right] + \log \frac{q_{\lambda_2,f_c}(Y_0)}{q_{\lambda_1,f}(Y_0)},$$

where $s_{c,t}(\lambda) = \sqrt{f_c(\eta_t(\lambda))}/\sqrt[4]{h_t(\lambda)}$ and $s_t(\lambda) = s_{0,t}(\lambda)$. Let $g_t(\lambda) = (\varepsilon_t(\lambda), \sqrt{h_t(\lambda)})$ with $\varepsilon_t(\lambda) = y_t - Z_{t-1}(\lambda)$. We make the following three assumptions. The first one ensures that the Fisher information is finite for both scale and location parameters. The second deals with the behavior of the functions g_t . The last deals with the initial densities.

ASSUMPTION 2.1. The density f is absolutely continuous with a.e.-derivative f' and

$$I_1(f) = \int \xi_1^2(x) f(x) dx < \infty \quad \text{and} \quad I_2(f) = \int \xi_2^2(x) f(x) dx < \infty,$$

where $\xi_1(x) = f'(x)/f(x)$ and $\xi_2(x) = 1 + x\xi_1(x)$.

ASSUMPTION 2.2. There is a sequence G_n of invertible diagonal $k \times k$ matrices with $G_n^{-1} \rightarrow 0$ and $k \times 2$ random matrices $U_t(\lambda)$ such that, for all bounded sequence θ_n and ϑ_n , the following statements are true with $\lambda_n = \lambda_0 + G_n^{-1}\theta_n$ and $\tilde{\lambda}_n = \lambda_n + G_n^{-1}\vartheta_n$:

- (i) $\left[\inf_{1 \leq t \leq n} \sqrt{h_t(\lambda_n)} \right]^{-1} = O_{\lambda_n}(1),$
- (ii) $\sum_{t=1}^n [g_t(\tilde{\lambda}_n) - g_t(\lambda_n) - (\tilde{\lambda}_n - \lambda_n)' U_t(\lambda_n)]^2 = o_{\lambda_n}(1),$
- (iii) $\sup_{1 \leq t \leq n} \|G_n^{-1} U_t(\lambda_n)\|^2 = o_{\lambda_n}(1),$
- (iv) $\sum_{t=1}^n \|G_n^{-1} U_t(\lambda_n)\|^2 = O_{\lambda_n}(1).$

ASSUMPTION 2.3. $\int |q_{\lambda, f_c}(x) - q_{\lambda_0, f}(x)| dx = o(1)$ as $\|\lambda - \lambda_0\| = o(1)$ and $\|c\| = o(1)$, where $f_c(x)$ is defined as in Definition 2.1.

We further introduce the following notation:

$$\begin{aligned} W_n(\lambda) &= G_n^{-1} \sum_{t=1}^n X_t(\lambda) \xi(\eta_t(\lambda)), & W_{\zeta n}(\lambda) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \zeta(\eta_t(\lambda)), \\ S_n(\lambda) &= G_n^{-1} \sum_{t=1}^n X_t(\lambda) V_{\xi, \xi} X_t'(\lambda) G_n^{-1}, & S_{\zeta n}(\lambda) &= \frac{G_n^{-1}}{\sqrt{n}} \sum_{t=1}^n X_t(\lambda), \\ \tilde{W}_n(\lambda) &= \begin{pmatrix} W_n(\lambda) \\ W_{\zeta n}(\lambda) \end{pmatrix}, & \tilde{S}_n(\lambda) &= \begin{pmatrix} S_n(\lambda) & S_{\zeta n}(\lambda) V_{\xi, \xi}' \\ V_{\zeta, \xi} S_{\zeta n}'(\lambda) & V_{\zeta, \zeta} \end{pmatrix}, \end{aligned}$$

where $X_t(\lambda) = U_t(\lambda)/\sqrt{h_t(\lambda)}$, $\xi = (\xi_1, -\xi_2)'$, $V_{\xi, \xi} = E[\xi(\eta_t)\xi'(\eta_t)]$, $V_{\zeta, \xi} = E[\xi(\eta_t)\zeta'(\eta_t)]'$ and $V_{\zeta, \zeta} = E[\zeta(\eta_t)\zeta'(\eta_t)]$. Now, we give the general LAQ criterion; its proof can be found in the Appendix.

THEOREM 2.1. *Suppose that the path $c \rightarrow f_c$ is regular and that Assumptions 2.1–2.3 hold. Let $u_n = (\vartheta'_n, v'_n)'$ and v_n be a bounded sequence in \mathbb{R}^l . Then*

- (a) $\Lambda_n(\lambda_n, \tilde{\lambda}_n, v_n/\sqrt{n}) = u'_n \tilde{W}_n(\lambda_n) - u'_n \tilde{S}_n(\lambda_n) u_n/2 + o_{\lambda_n}(1)$,
- (b) $P_{\lambda_0, n}$ and $P_{\lambda_n, n}$ are contiguous,
- (c) $\tilde{S}_n(\lambda_n) = \tilde{S}_n(\lambda_0) + o_{\lambda_0}(1)$ and $\tilde{W}_n(\lambda_n) = \tilde{W}_n(\lambda_0) - \tilde{S}_n(\lambda_0) \begin{pmatrix} \theta_n \\ 0 \end{pmatrix} + o_{\lambda_0}(1)$.

REMARK 2.1. If the LAQ of $\Lambda_n(\lambda_n, \tilde{\lambda}_n, v_n/\sqrt{n})$ is LAN, LAMN or LABF, then (b) automatically holds [see Kallianpur (1980), Chapter 7 and Jegannathan (1995), page 850]. In this case, it is sufficient to verify Assumption 2.2 with $\lambda_n = \lambda_0$ and that

$$(2.2) \quad \sum_{t=1}^n \|G_n^{-1}[U_t(\lambda_n) - U_t(\lambda_0)]\|^2 = o_{\lambda_0}(1).$$

Assumption 2.3 means that the starting conditions have a negligible effect. If Y_0 is assumed to be independent of (λ, χ) , as in the next section, then this assumption holds. Koul and Schick (1997) discussed this assumption carefully for some stationary nonlinear AR models.

REMARK 2.2. When the LAQ is LAN or LAMN, the error model \mathcal{D} has a two-dimensional least favorable path: $\xi_*(x) = -\xi(x) + V'_{\zeta, \xi} V^{-1} \zeta(x)$ with $\zeta(x) = (x, x^2 - 1)'$. Along this path, one can obtain the optimal estimates and discuss the efficiency and adaptivity. For the stationary nonlinear AR model, Koul and Schick (1997) showed that the LAQ is LAN, and especially, they found a one-dimensional least favorable path and generalized the criterion of efficiency in Fabian and Hannan (1982) and Schick (1988). When the LAQ is LABF, as in the next section, the notion of efficiency needs to be defined. We shall do so in the next section by defining efficiency to mean best in a class of estimates. This requires an efficient estimator to have an expansion similar to the LAMN case and efficient estimators can be constructed along the lines of DKW (1997) and Koul and Schick (1997).

3. The LABF, adaptivity and efficiency for nonstationary ARMA–GARCH model. First, it is necessary to isolate the unit root in model (1.1). Note that $\varphi_0(z)$ can be decomposed as $(1 - z)\phi_0(z)$, where $\phi_0(z) = 1 - \sum_{i=1}^{p-1} \phi_{0i} z^i$. Let $w_t = (1 - B)y_t$, where B is the backshift operator. Model (1.1) can be rewritten as

$$y_t = \gamma_0 y_{t-1} + w_t, \quad w_t = \sum_{i=1}^{p-1} \phi_{0i} w_{t-i} + \sum_{i=1}^q \psi_{0i} \varepsilon_{t-i} + \varepsilon_t,$$

where $\gamma_0 = 1$. In (1.2), we assume that the variance of η_t is one. In this case, all the parameters in (1.2) can be estimated by the QMLE method, as in Ling and

Li (1998). However, the parameters in (1.2) are not adaptively estimable; see the discussion for the GARCH(1,1) model in Drost and Klaassen (1997). Model (1.2) needs to be reparameterized. Thus, we assume that, under $P_{\lambda, \chi}$, y_t , $t = 1, \dots, n$, satisfy the following structure:

$$(3.1) \quad w_t(\lambda) = y_t - \gamma y_{t-1}, \quad \varepsilon_t(\lambda) = w_t(\lambda) - \sum_{i=1}^{p-1} \phi_i w_{t-i}(\lambda) - \sum_{i=1}^q \psi_i \varepsilon_{t-i}(\lambda),$$

$$(3.2) \quad \eta_t(\lambda) = \varepsilon_t(\lambda) / \sqrt{h_t(\lambda)}, \quad h_t(\lambda) = \alpha_0 \left[1 + \sum_{i=1}^r \alpha_i \varepsilon_{t-i}^2(\lambda) + \sum_{i=1}^s \beta_i h_{t-i}(\lambda) \right],$$

with the initial (unobservable) vector $Y_0 = (y_0, \dots, y_{1-p}, \varepsilon_0, \dots, \varepsilon_{1-q^*}, h_0, \dots, h_{1-s})$ and $q^* = \max\{r, q\}$, where the rescaled errors $\eta_1(\lambda), \eta_2(\lambda), \dots$ are i.i.d. with density $\chi \in \mathcal{D}$ and independent of Y_0 , $\lambda = (\gamma, m', \tilde{\delta}')'$, $m = (\phi', \psi')'$ with $\phi = (\phi_1, \dots, \phi_{p-1})'$ and $\psi = (\psi_1, \dots, \psi_q)'$, $\tilde{\delta} = (\alpha_0, \delta')'$ with $\delta = (\alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_s)'$, and the true parameter $(\lambda_0, f) \in \Theta \times \mathcal{D}$. We assume that, for simplicity, the density of Y_0 does not depend on the parameters (λ, χ) and, for each $\lambda \in \Theta$, the following requirements hold.

ASSUMPTION 3.1. All the roots of $\phi(z) = 1 - \sum_{i=1}^{p-1} \phi_i z^i$ and $\psi(z) = 1 + \sum_{i=1}^q \psi_i z^i$ are outside the unit circle, with $\phi_{p-1} \neq 0$ and $\psi_q \neq 0$, and $\phi(z)$ and $\psi(z)$ having no common root.

ASSUMPTION 3.2. $\alpha_0(\sum_{i=1}^r \alpha_i + \sum_{i=1}^s \beta_i) < 1$ with α_0 greater than a positive constant, $\alpha_i > 0$ and $\beta_i > 0$, and $\alpha_0 \sum_{i=1}^r \alpha_i z^i$ and $1 - \alpha_0 \sum_{i=1}^s \beta_i z^i$ having no common root.

ASSUMPTION 3.3. $\rho[E_\lambda(A_t(\lambda) \otimes A_t(\lambda))] < 1$, where \otimes denotes the Kronecker product, $\rho(B) = \max\{|x| : x \text{ is an eigenvalue of } B\}$ for some matrix B , E_λ denotes the expectation under $P_{\lambda, f}$, and

$$A_t(\lambda) = \left(\begin{array}{ccc|ccc} \alpha_0 \alpha_1 \eta_t^2(\lambda) & \cdots & \alpha_0 \alpha_r \eta_t^2(\lambda) & \alpha_0 \beta_1 \eta_t^2(\lambda) & \cdots & \alpha_0 \beta_s \eta_t^2(\lambda) \\ & I_{r-1} & O_{(r-1) \times 1} & & O_{(r-1) \times s} & \\ \hline \alpha_0 \alpha_1 & \cdots & \alpha_0 \alpha_r & \alpha_0 \beta_1 & \cdots & \alpha_0 \beta_s \\ & O_{(s-1) \times r} & & & I_{s-1} & O_{(s-1) \times 1} \end{array} \right),$$

in which I_i is the $i \times i$ identity matrix and $O_{i \times j}$ denotes the $i \times j$ zero matrix.

REMARK 3.1. Assumption 3.1 is the usual second-order stationary condition of the process $\{w_t\}$ in model (3.1). Assumptions 3.2 and 3.3 are the necessary and sufficient conditions, respectively, for the finite second- and fourth-order moments of model (3.2) [see Ling (1999) and Ling and McAleer (2002)]. Assumption 3.2 is not a necessary condition for strict stationarity of model (3.2), see Nelson (1990).

To state our main result in this section, we need the following notation:

$$\begin{aligned}\phi^{-1}(z)\psi(z) &= \sum_{i=0}^{\infty} v_{\phi\psi}(i)z^i, \\ \psi^{-1}(z)\phi(z) &= \sum_{i=0}^{\infty} v_{\psi\phi}(i)z^i, \\ \psi^{-1}(z) &= \sum_{i=0}^{\infty} v_{\psi}(i)z^i, \\ \left(1 - \alpha_0 \sum_{i=1}^s \beta_i z^i\right)^{-1} &= \sum_{i=0}^{\infty} v_{\beta}(i)z^i, \\ \left(1 - \alpha_0 \sum_{i=1}^s \beta_i z^i\right)^{-1} \left(\alpha_0 \sum_{i=1}^r \alpha_i z^i\right) &= \sum_{i=1}^{\infty} v_{\alpha\beta}(i)z^i,\end{aligned}$$

where $v_{\phi_0\psi_0}(i)$ and $v_{\phi_n\psi_n}(i)$ denote $v_{\phi\psi}(i)$ with $\lambda = \lambda_0$ and λ_n , respectively. Similarly, define $v_{\psi_0\phi_0}(i)$, $v_{\psi_n\phi_n}(i)$, $v_{\psi_0}(i)$, $v_{\psi_n}(i)$, $v_{\beta_0}(i)$, $v_{\beta_n}(i)$, $v_{\alpha_0\beta_0}(i)$ and $v_{\alpha_n\beta_n}(i)$. Furthermore, we introduce the unobservable processes $\{(w_t^0, \varepsilon_t^0, h_t^0) : t = 0, \pm 1, \pm 2, \dots\}$ generated by the following equations:

$$\begin{aligned}w_t^0 &= \sum_{i=1}^{p-1} \phi_{0i} w_{t-i}^0 + \sum_{i=1}^q \psi_{0i} \varepsilon_{t-i}^0 + \varepsilon_t^0, \\ \varepsilon_t^0 &= \eta_t \sqrt{h_t^0}, \quad h_t^0 = \alpha_{00} \left(1 + \sum_{i=1}^r \alpha_{0i} \varepsilon_{t-i}^{02} + \sum_{i=1}^s \beta_{0i} h_{t-i}^0\right).\end{aligned}$$

By Lemma 5.1 in Section 5, $(w_t^0, \varepsilon_t^0, h_t^0)$ is a fixed function of the $\{\eta_t\}$. Define

$$\begin{aligned}\frac{\partial \varepsilon_t^0}{\partial m} &= - \sum_{i=0}^{\infty} v_{\psi_0}(i) \tilde{w}_{t-i-1}^0, & \frac{\partial h_t^0}{\partial m} &= 2 \sum_{i=1}^{\infty} v_{\alpha_0\beta_0}(i) \varepsilon_{t-i}^0 \frac{\partial \varepsilon_{t-i}^0}{\partial m}, \\ \frac{\partial h_t^0}{\partial \delta} &= \sum_{i=0}^{\infty} v_{\beta_0}(i) (u_{0t-i}^0, \tilde{\varepsilon}_{t-i-1}^{0'})', & \tilde{\varepsilon}_t^0 &= \alpha_{00} (\varepsilon_t^{02}, \dots, \varepsilon_{t-r+1}^{02}, h_t^0, \dots, h_{t-s+1}^0)',\end{aligned}$$

where $\tilde{w}_t^0 = (w_t^0, \dots, w_{t-p+2}^0, \varepsilon_t^0, \dots, \varepsilon_{t-q+1}^0)'$ and $u_{0t}^0 = 1 + \sum_{i=1}^r \alpha_{0i} \varepsilon_{t-i}^{02} + \sum_{i=1}^s \beta_{0i} h_{t-i}^0$.

Using the same notation as those in Section 2 with $U_t = [\partial \varepsilon_t(\lambda)/\partial \lambda, (\partial h_t(\lambda)/\partial \lambda)/2\sqrt{h_t(\lambda)}]$, $k = p + q + r + s + 1$ and $G_n = \text{diag}(n, \sqrt{n}I_{k-1})$, our theorem is as follows.

THEOREM 3.1. *Suppose that Assumptions 2.1 and 3.1–3.3 hold and the map $c \rightarrow f_c$ is a regular path such that $\tilde{\Omega}$ below is positive definite. Let $u_n = (\vartheta'_n, v'_n)'$ and v_n be a bounded sequence in \mathbb{R}^l . Then (a) the conclusions of Theorem 2.1 hold, and (b) the matrix \tilde{S} below is almost surely positive definite and, under P_{λ_0} ,*

$$(\tilde{W}_n, \tilde{S}_n)(\lambda_0) \xrightarrow{\mathcal{L}} (\tilde{W}, \tilde{S}) = \left[\int_0^1 M(\tau) dB(\tau), \int_0^1 M(\tau) \Sigma M'(\tau) d\tau \right],$$

where $\kappa = [1 - \phi_0(1)]^{-1} \psi_0(1)$, $M(\tau) = \text{diag}(\kappa \omega_1(\tau), I_{k-1}, I_l)$, $B(\tau) = (\kappa \omega_2, N'_{m\tilde{\delta}}, N'_{\xi})'(\tau)$, $(\omega_1, \omega_2, N'_{m\tilde{\delta}}, N'_{\xi})'(\tau)$ is a $(k + l + 1)$ -dimensional Brownian motion with mean zero and covariance $\tau \tilde{\Omega} = \tau \begin{pmatrix} \Omega & C \\ C' & V_{\zeta, \zeta} \end{pmatrix}$,

$$\Omega = E \begin{pmatrix} h_t^{02} & \xi'(\eta_t) X_t^{0'} \varepsilon_t^0 \\ X_t^0 \xi(\eta_t) \varepsilon_t^0 & X_t^0 V_{\xi, \xi} X_t^{0'} \end{pmatrix}, \quad \Sigma = E \begin{pmatrix} X_t^0 V_{\xi, \xi} X_t^{0'} & X_t^0 V'_{\zeta, \xi} \\ V_{\zeta, \xi} X_t^{0'} & V_{\zeta, \zeta} \end{pmatrix},$$

$C = E[\zeta(\eta_t) \varepsilon_t^0, V_{\zeta, \xi} X_t^{0'}]'$, $X_t^0 = (u_{\gamma t}^{0'}, u_{mt}^{0'}, u_{\delta t}^{0'})'$, $u_{\gamma t}^0 = -[(h_t^0)^{-1/2}, \sum_{i=1}^{\infty} v_{\alpha_0 \beta_0}(i) \times \varepsilon_{t-i}^0 / h_t^0]$, $u_{mt}^0 = [(h_t^0)^{-1/2} \partial \varepsilon_t^0 / \partial m, (2h_t^0)^{-1} \partial h_t^0 / \partial m]$, and $u_{\delta t}^0 = [0, (2h_t^0)^{-1} \times \partial h_t^0 / \partial \tilde{\delta}]$.

REMARK 3.2. From the above theorem, we see that the LAQ form of the log-LR $\Lambda_n(\lambda_n, \tilde{\lambda}_n, v/\sqrt{n})$ is neither LAN nor LAMN, but is instead LABF. The score function and information matrix of the unit root may be correlated with those of the other parameters in the stationary mean part and the GARCH part. This phenomenon is new in the literature and results in the complicated limiting distribution (\tilde{W}, \tilde{S}) . Using Assumptions 3.1–3.3, we can show that $\Omega > 0$, as in Weiss (1986) and Ling and Li (1997). Furthermore, for $\tilde{\Omega} > 0$, one of the sufficient conditions is $E(\mathfrak{R}_t \mathfrak{R}_t') > 0$ with $\mathfrak{R}_t = [\eta_t, \xi'(\eta_t), \zeta'(\eta_t)]'$. However, this condition excludes the normal density. If we further assume that the path satisfies $\lim_{c \rightarrow 0} \int (1 + x^4) f_c(x) dx = \int (1 + x^4) f(x) dx$, then some two-dimensional regular paths such that $\tilde{\Omega} > 0$, can be constructed. Since the argument becomes more involved, we refer to Koul and Schick (1996, 1997) for the one-dimensional regular paths.

REMARK 3.3. When \mathcal{D} includes only densities that are symmetric about zero, the limiting distribution in Theorem 3.1(b) can be simplified as follows:

$$(\tilde{W}, \tilde{S}) = \begin{pmatrix} \kappa \int_0^1 \omega_1(\tau) d\omega_2(\tau) & \kappa^2 \Omega_{\gamma} \int_0^1 \omega_1^2(\tau) d\tau & 0 & 0 & 0 \\ N_1 & 0 & \Omega_m & 0 & 0 \\ N_2 & 0 & 0 & \Omega_{\tilde{\delta}} & V'_{\tilde{\delta} \zeta} \\ N_{\zeta} & 0 & 0 & V_{\tilde{\delta} \zeta} & V_{\zeta, \zeta} \end{pmatrix},$$

where $(\omega_1, \omega_2)(\tau)$ is a bivariate Brownian motion with mean zero and covariance $\tau \Omega_1 = \tau \begin{pmatrix} E h_t^0 & 1 \\ 1 & \Omega_{\gamma} \end{pmatrix}$; N_1 and $\begin{pmatrix} N_2 \\ N_{\zeta} \end{pmatrix}$ are $(p + q - 1)$ - and $(r + s + 1 + l)$ -normal

vectors with mean zero and covariances Ω_m and $\begin{pmatrix} \Omega_{\delta} & V'_{\delta\zeta} \\ V_{\delta\zeta} & V'_{\zeta\zeta} \end{pmatrix}$, respectively, and independent of $(\omega_1, \omega_2)(\tau)$; and $\Omega_\gamma = E(u_{\gamma t}^0 V_{\xi,\xi} u_{\gamma t}^{0'})$, $\Omega_m = E(u_{mt}^0 V_{\xi,\xi} u_{mt}^{0'})$, $\Omega_{\tilde{\delta}} = E(u_{\delta t}^0 V_{\xi,\xi} u_{\delta t}^{0'})$ with $V_{\xi,\xi} = \text{diag}(I_1(f), I_2(f))$, and $V'_{\delta\zeta} = E u_{\delta t}^0 V'_{\zeta,\xi}$. In this case, the LR is the product of a LABF and a LAN (a special LABF). If we assume that the unit root in (1.1) is known and not estimated, then from Theorem 3.1, we see that model (3.1), (3.2) belongs to the LAN family. Using slightly stronger conditions, this result is a generalization of Drost and Klaassen (1997) and DKW (1997) for stationary ARCH-type time series.

In LAN models, Hájek (1972), Fabian and Hannan (1982) and Koul and Schick (1997) established the precise notion of efficiency. Jeganathan [(1995), Section 3] discussed the efficiency of the estimators in LAMN models. However, the definition and discussion they gave are inappropriate for the current case. As in Jeganathan (1995), in order to obtain some useful optimality properties, we need to restrict the competing class of estimators. We first make the following assumption.

ASSUMPTION 3.4. The density f is symmetric and \mathcal{D} includes only densities that are symmetric about zero.

Furthermore, let \hat{v}_n be a k_1 -dimensional subvector of $\hat{\lambda}_n$. Similarly define v and v_0 . Denote $\dot{v}_0 = \partial v / \partial \lambda' |_{\lambda=\lambda_0}$. By the definition of G_n , it is obvious that there is a $k_1 \times k_1$ matrix G_n^* satisfying $G_n^* \dot{v}_0 G_n^{-1} = \dot{v}_0$. We define a class of estimators, namely M_v -estimators. Let Π be the set of all measurable functions $\pi = (\pi_1, \pi_2)'$ from $R \rightarrow R^2$ with π_1 odd and π_2 even such that $E[\pi(\eta_t)] = 0$ and $E[\|\pi(\eta_t)\|^2] < \infty$, and the (diagonal 2×2) matrices $V_{\pi,\pi}$ and $V_{\pi,\xi}$ are positive definite. For each $\pi \in \Pi$, let $W_n(\lambda, \pi) = G_n^{-1} \sum_{t=1}^n X_t(\lambda) \pi(\eta_t(\lambda))$ and $S_n(\lambda, \pi) = \sum_{t=1}^n G_n^{-1} X_t(\lambda) V_{\pi,\xi} X_t'(\lambda) G_n^{-1}$.

DEFINITION 3.1. An estimator \bar{v}_n of v_0 is said to be an M_v -estimator at (λ_0, f) if $G_n^*(\bar{v}_n - v_0) = \dot{v}_0 S_n^{-1}(\lambda_0, \pi) W_n(\lambda_0, \pi) + o_{\lambda_0}(1)$ for some $\pi \in \Pi$.

The class of M_v -estimators is very wide and includes the QMLE and MLE (if available). Now, we define the optimality properties of M_v -estimators and present an efficiency criterion for estimators in the class \mathcal{M}_v below. In the following definition, we suppose that Assumptions 2.1 and 3.1–3.4 hold, under which every M_v -estimator has a limiting distribution under P_{λ_0} (see Proof of Theorem 3.2 in Section 5).

DEFINITION 3.2. Let \mathcal{M}_v be the set of all M_v -estimators. We say that \bar{v}_n is efficient if $\bar{v}_n \in \mathcal{M}_v$ and $\bar{v}_n \xrightarrow{\mathcal{L}} G$ (a random vector) under P_{λ_0} , such that $E(GG') \leq E(G_\pi G'_\pi)$ for any $\bar{v}_{\pi n} \in \mathcal{M}_v$ with $\bar{v}_{\pi n} \xrightarrow{\mathcal{L}} G_\pi$ under P_{λ_0} .

THEOREM 3.2. *Suppose that Assumptions 2.1 and 3.1–3.4 hold. If a sequence of estimators \hat{v}_n of v_0 has the following asymptotic representation:*

$$G_n^*(\hat{v}_n - v_0) = \dot{v}_0 S_n^{-1}(\lambda_0) W_n(\lambda_0) + o_{\lambda_0}(1),$$

then the estimator \hat{v}_n belongs to \mathcal{M}_v and is efficient, where $(W_n, S_n)(\lambda_0)$ is the $k \times (k+1)$ upper left corner of $(\tilde{W}_n, \tilde{S}_n)(\lambda_0)$.

In LAN models, various definitions based on the locally asymptotic minimax risk for adaptivity were given in Bickel (1982), Fabian and Hannan (1982), and Koul and Schick (1997), among others. Roughly speaking, these definitions are equivalent to saying that a sequence of adaptive estimators have the same asymptotic information matrix as the estimators in the case with known density. The information matrix can completely explain the perturbation of the unknown density to the score function in LAN and LAMN models. However, in the LABF model, the information matrix does not have this advantage and it is difficult to extend their definitions to our case. The following definition stresses only the fact that the estimator of v_0 without the knowledge of the true density can achieve the same asymptotic distribution as its estimator when f is known.

DEFINITION 3.3. When the density function f is unknown, if an estimator \hat{v}_n of v_0 is efficient, then it is called adaptive.

In this sense, the adaptive estimator has the same asymptotic distribution as the MLE, if the latter is available. In Section 4, we will show how to construct an adaptive estimator $(\hat{\gamma}_n, \hat{m}_n)$ for (γ_0, m_0) . After projecting the score function and the information matrix of (δ, c) into those corresponding to α_0 , we can obtain that the score function and the information matrix of δ_0 are, respectively,

$$W_{\delta n}(\lambda) = -\frac{1}{2\sqrt{n}} \sum_{t=1}^n [\dot{l}_{\delta t}(\lambda) - \hat{\mu}_{\delta}(\lambda)] \xi_2(\eta_t(\lambda)),$$

$$S_{\delta n}(\lambda) = \frac{1}{4n} \left\{ \sum_{t=1}^n \dot{l}_{\delta t}(\lambda) \dot{l}'_{\delta t}(\lambda) - \left[\sum_{t=1}^n \dot{l}_{\delta t}(\lambda) \right] \left[\sum_{t=1}^n \dot{l}'_{\delta t}(\lambda) \right] \right\} I_2(f),$$

and $\Omega_{\delta} = [E(\dot{l}_{\delta t}^0 \dot{l}_{\delta t}^{0'}) - E(\dot{l}_{\delta t}^0) E(\dot{l}_{\delta t}^{0'})] I_2(f)/4$, with $\dot{l}_{\delta t}(\lambda) = h_t^{-1}(\lambda) \partial h_t(\lambda) / \partial \delta$, $\dot{l}_{\delta t}^0 = h_t^{0-1} \partial h_t^0 / \partial \delta$, and $\hat{\mu}_{\delta}(\lambda) = n^{-1} \sum_{t=1}^n [h_t^{-1}(\lambda) \partial h_t(\lambda) / \partial \delta]$. Section 4 will show that we can construct an adaptive estimator for δ_0 .

For the ARCH(p) and GARCH(1, 1) models, DWK (1997) and Drost and Klaassen (1997) have shown that α_0 is not adaptively estimable. It is not easy to explain that α_0 is not adaptively estimable under our framework. When the density function f is asymmetric, a similar definition on the M_v -estimators can be given, but in this case, the forms of the corresponding score function and information matrix are different from those in Definition 3.1. Having specified the class of M_v -estimators, efficiency can be defined similarly. Since our emphasis is on the symmetric case, it is not discussed here.

4. Efficient and adaptive estimators. In order to construct the efficient estimator, we need to assume that a G_n -consistent initial estimator is available. In fact, the QMLE in Ling and Li (1998) can be taken as such an initial estimator. For technical reasons, we also need to restrict the initial estimator to be discrete. The idea of discretization was first proposed by Le Cam (1960), and has become an important technical tool in the construction of efficient estimators. Some further applications of the technique can be found in Bickel (1982), Kreiss (1987a), Jeganathan (1995) and Koul and Schick (1997), among others. We now provide the following definition and lemma.

DEFINITION 4.1. A sequence of estimators $\{\bar{\lambda}_n\}$ measurable in terms of \mathcal{F}_n is called discretized G_n -consistent if, for any small $\varepsilon > 0$, there exists a constant $\Delta > 0$ and an integer $K > 0$ such that $P_{\lambda_0}(\|G_n(\bar{\lambda}_n - \lambda_0)\| < \Delta) > 1 - \varepsilon$ uniformly in n and, for each n , $\bar{\lambda}_n$ takes on at most K different values in $\Theta_n = \{\lambda \in \mathbb{R}^k : \|G_n(\lambda - \lambda_0)\| \leq \Delta\}$.

LEMMA 4.1. Assume $\Gamma_n(\lambda), n = 1, 2, \dots$, to be a sequence of random variables which depends on $\lambda \in \Theta$, an open subset in \mathbb{R}^k . If, for each sequence $\{\lambda_n\} \in \Theta$ satisfying $G_n(\lambda_n - \lambda_0)$ is bounded by a constant $\Delta > 0$, $\Gamma_n(\lambda_n) = o_{\lambda_0}(1)$, then $\Gamma_n(\bar{\lambda}_n) = o_{\lambda_0}(1)$ for discretized G_n -consistent estimators $\bar{\lambda}_n$.

The proof of this lemma is similar to Lemma 4.4 in Kreiss (1987a), and hence is omitted. Based on the initial estimator, the efficient estimator can be obtained by a one-step Newton–Raphson iteration if the density f is known. This gives the following theorem, which comes directly from Theorem 2.1(c), Theorem 3.1(a) and Lemma 4.1.

THEOREM 4.1. Suppose that $\bar{\lambda}_n$ is a discretized G_n -consistent estimator, and Assumptions 2.1 and 3.1–3.4 hold. Let

$$\tilde{\lambda}_n = \bar{\lambda}_n + G_n^{-1} S_n^{-1}(\bar{\lambda}_n) W_n(\bar{\lambda}_n).$$

Then $G_n(\tilde{\lambda}_n - \lambda_0) = S_n^{-1}(\lambda_0) W_n(\lambda_0) + o_{\lambda_0}(1)$, and hence $\tilde{\lambda}_n$ is an efficient estimator.

In practice, the density is usually unknown. In the following, we will construct an adaptive estimator which does not depend on the density but has the same efficiency as when the density is known. We merge α_{00} into f , which is equivalent to assuming that η_t has a finite variance α_{00} and that the true parameter α_{00} in model (3.2) is equal to 1. In the remainder of this section and Section 6, denote $(\gamma, m', \delta')'$ by λ . Similarly, define λ_0 and $\hat{\lambda}_n$. We introduce the

notation:

$$\begin{aligned}
W_{\gamma n}(\lambda) &= \frac{1}{n} \sum_{t=1}^n \left[\frac{1}{\sqrt{h_t(\lambda)}} \frac{\partial \varepsilon_t(\lambda)}{\partial \gamma} \xi_1(\eta_t(\lambda)) - \frac{1}{2h_t(\lambda)} \frac{\partial h_t(\lambda)}{\partial \gamma} \xi_2(\eta_t(\lambda)) \right], \\
W_{mn}(\lambda) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \left[\frac{1}{\sqrt{h_t(\lambda)}} \frac{\partial \varepsilon_t(\lambda)}{\partial m} \xi_1(\eta_t(\lambda)) - \frac{1}{2h_t(\lambda)} \frac{\partial h_t(\lambda)}{\partial m} \xi_2(\eta_t(\lambda)) \right], \\
S_{\gamma n}(\lambda) &= \frac{1}{n^2} \sum_{t=1}^n \left[\frac{1}{h_t(\lambda)} \left(\frac{\partial \varepsilon_t(\lambda)}{\partial \gamma} \right)^2 I_1(f) + \frac{1}{4h_t^2(\lambda)} \left(\frac{\partial h_t(\lambda)}{\partial \gamma} \right)^2 I_2(f) \right], \\
S_{mn}(\lambda) &= \frac{1}{n} \sum_{t=1}^n \left[\frac{1}{h_t(\lambda)} \frac{\partial \varepsilon_t(\lambda)}{\partial m} \frac{\partial \varepsilon_t(\lambda)}{\partial m'} I_1(f) + \frac{1}{4h_t^2(\lambda)} \frac{\partial h_t(\lambda)}{\partial m} \frac{\partial h_t(\lambda)}{\partial m'} I_2(f) \right], \\
W_{1n}(\lambda) &= \begin{pmatrix} W_{\gamma n}(\lambda) \\ W_{mn}(\lambda) \\ W_{\delta n}(\lambda) \end{pmatrix}, \quad S_{1n}(\lambda) = \begin{pmatrix} S_{\gamma n}(\lambda) & 0 & 0 \\ 0 & S_{mn}(\lambda) & 0 \\ 0 & 0 & S_{\delta n}(\lambda) \end{pmatrix},
\end{aligned}$$

where $W_{\delta n}(\lambda)$ and $S_{\delta n}(\lambda)$ are defined as in Section 3.

Now we construct adaptive estimators for λ_0 . Using the usual kernel density estimator for $\xi_1(x)$, we define

$$(4.1) \quad \hat{f}_{a,j}(x, \lambda) = \frac{1}{2a(n-1)} \sum_{i=1, i \neq j}^n [K(x + \eta_i(\lambda), a) + K(x - \eta_i(\lambda), a)],$$

where $j = 1, \dots, n$, and $K(x) = e^{-x}/(1 + e^{-x})^2$ is the logistic kernel,

$$(4.2) \quad \hat{\xi}_{1n,j}(x, \lambda) = \frac{\hat{f}'_{a_n,j}(x, \lambda)}{b_n + \hat{f}_{a_n,j}(x, \lambda)}$$

and $\hat{\xi}_{2n,j}(x, \lambda) = x \hat{\xi}_{1n,j}(x, \lambda) + 1$, with a_n and b_n satisfying $na_n^3 b_n \rightarrow \infty$. Define $\hat{I}_{1n}(\lambda)$ and $\hat{I}_{2n}(\lambda)$ as follows:

$$\hat{I}_{1n}(\lambda) = \frac{1}{n} \sum_{t=1}^n \hat{\xi}_{1n,t}^2(\eta_t(\lambda), \lambda) \quad \text{and} \quad \hat{I}_{2n}(\lambda) = \frac{1}{n} \sum_{t=1}^n [\eta_t(\lambda) \hat{\xi}_{1n,t}(\eta_t(\lambda), \lambda) + 1]^2.$$

Denote $[\hat{W}_{\gamma n}(\lambda), \hat{W}'_{mn}(\lambda)]$ by $[W_{\gamma n}(\lambda), W'_{mn}(\lambda)]$ with $\xi_i(\eta_t(\lambda))$ replaced by $\hat{\xi}_{in,t}(\eta_t(\lambda), \lambda)$, and $\text{diag}[\hat{S}_{\gamma n}(\lambda), \hat{S}_{mn}(\lambda)]$ by $\text{diag}[S_{\gamma n}(\lambda), S_{mn}(\lambda)]$ with $I_i(f)$ replaced by $\hat{I}_{in}(\lambda)$, where $i = 1, 2$. $W_{\gamma n}(\lambda)$, $W_{mn}(\lambda)$, $S_{\gamma n}(\lambda)$ and $S_{mn}(\lambda)$ are estimated by $\hat{W}_{\gamma n}(\lambda)$, $\hat{W}_{mn}(\lambda)$, $\hat{S}_{\gamma n}(\lambda)$ and $\hat{S}_{mn}(\lambda)$, respectively.

To estimate the score function of δ , we need the split sample technique. This technique was proposed by Schick (1986) and was also used by DKW (1997). Let k_n be an integer such that $k_n/n \rightarrow \tau \in (0, 1)$. Split the residual $\eta_1(\lambda), \dots, \eta_n(\lambda)$

into two parts, namely $(\eta_1(\lambda), \dots, \eta_{k_n}(\lambda))$ and $(\eta_{k_n+1}(\lambda), \dots, \eta_n(\lambda))$. Denote

$$\begin{aligned}\hat{f}_a^{(1)}(x, \lambda) &= \frac{1}{2ak_n} \sum_{i=1}^{k_n} \left[K\left(\frac{x + \eta_i(\lambda)}{a}\right) + K\left(\frac{x - \eta_i(\lambda)}{a}\right) \right], \\ \hat{f}_a^{(2)}(x, \lambda) &= \frac{1}{2a(n - k_n)} \sum_{i=k_n+1}^n \left[K\left(\frac{x + \eta_i(\lambda)}{a}\right) + K\left(\frac{x - \eta_i(\lambda)}{a}\right) \right].\end{aligned}$$

Define $\hat{\xi}_{2n}^{(i)}(x, \lambda) = \hat{f}_{a_n}^{(i)'}(x, \lambda) / [b_n + \hat{f}_{a_n}^{(i)}(x, \lambda)]$. $W_{\delta n}(\lambda)$ is estimated by

$$\begin{aligned}\hat{W}_{\delta n}(\lambda) &= -\frac{1}{2\sqrt{n}} \sum_{t=1}^{k_n} \left[\frac{1}{h_t(\lambda)} \frac{\partial h_t(\lambda)}{\partial \delta} - \hat{\mu}_{\delta}(\lambda) \right] \hat{\xi}_{2n,t}^{(2)}(\eta_t(\lambda), \lambda) \\ &\quad - \frac{1}{2\sqrt{n}} \sum_{t=k_n+1}^n \left[\frac{1}{h_t(\lambda)} \frac{\partial h_t(\lambda)}{\partial \delta} - \hat{\mu}_{\delta}(\lambda) \right] \hat{\xi}_{2n,t}^{(1)}(\eta_t(\lambda), \lambda),\end{aligned}$$

where $\hat{\mu}_{\delta}(\lambda)$ is defined as in Section 3.

The main result in this section is the following theorem, which indicates that the parameter λ_0 is adaptively estimable.

THEOREM 4.2. *Suppose that $\bar{\lambda}_n$ is a discretized G_n -consistent estimator, and that Assumptions 2.1 and 3.1–3.4 hold. Let*

$$\hat{\lambda}_n = \bar{\lambda}_n + G_n^{-1} \hat{S}_{1n}^{-1}(\bar{\lambda}_n) \hat{W}_{1n}(\bar{\lambda}_n).$$

Then $G_n(\hat{\lambda}_n - \lambda_0) = S_{1n}^{-1}(\lambda_0)W_{1n}(\lambda_0) + o_{\lambda_0}(1)$, and hence $\hat{\lambda}_n$ is an adaptive estimator, where $G_n = \text{diag}(n, \sqrt{n}I_{k-2})$, $\hat{W}_{1n}(\lambda) = [\hat{W}_{\gamma n}(\lambda), \hat{W}'_{mn}(\lambda), \hat{W}'_{\delta n}(\lambda)]'$ and $\hat{S}_{1n}(\lambda) = \text{diag}[\hat{S}_{\gamma n}(\lambda), \hat{S}_{mn}(\lambda), \hat{S}_{\delta n}(\lambda)]$.

REMARK 4.1. In Theorem 4.2, we use the full sample without splitting for $(\gamma_0, m'_0)'$. This method is different from that used in DKW (1997) and may be more useful in practical applications, as in the simulation evidence in Koul and Schick (1997). This method is also different from that in Koul and Schick where they need to truncate the variable \dot{H}_j . The adaptive estimate of δ_0 is constructed by the split sample method, because no symmetry can be used in the score function of δ . If we make a suitable truncation to $h_t^{-1}(\lambda)\partial h_t(\lambda)/\partial \delta - \hat{\mu}_{\delta}$, as in Koul and Schick [(1997), Sections 5 and 6] and use the results in Schick (1987) and Schick and Susarla (1988), it is possible to avoid splitting the sample.

REMARK 4.2. Theorem 4.2 includes the new results that, by deleting the corresponding component for the unit root, the adaptive procedure above can be used for the stationary ARMA–GARCH model, and that the adaptive estimators achieve the smallest asymptotic covariance matrix in LAN models.

ADAPTIVE ESTIMATION FOR ARMA–GARCH

15

TABLE 1
The empirical bias and standard deviation of LSE, QMLE, AE and EE;
 $n = 250$, and 1000 replications

	$\gamma_0 = -0.5$		$\gamma_0 = 0.5$		$\gamma_0 = 0.8$		$\gamma_0 = 1.0$	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD
LSE	0.0088	0.0640	-0.0036	0.0634	-0.0079	0.0444	-0.0075	0.0151
QMLE	0.0043	0.0403	-0.0002	0.0407	-0.0036	0.0289	-0.0046	0.0100
AE	0.0004	0.0175	0.0001	0.0176	0.0005	0.0118	-0.0005	0.0028
EE	0.0004	0.0162	0.0002	0.0160	0.0005	0.0107	0.0005	0.0026

To see how well the adaptive estimator (AE) performs in finite samples compared with the QMLE and LSE for both the nonstationary and stationary cases, we simulate the following simple AR–GARCH model:

$$(4.3) \quad y_t = \gamma_0 y_{t-1} + \varepsilon_t, \quad \varepsilon_t = \eta_t \sqrt{h_t}, \quad h_t = \alpha_0 (1 + \alpha_0 \varepsilon_t^2 + \beta_0 h_{t-1}),$$

where η_t is i.i.d. with density $f(x) = [0.5e^{-(x-3)^2/2}/\sqrt{2\pi} + 0.5e^{-(x+3)^2/2}/\sqrt{2\pi}]/\sqrt{10}$ and $\alpha_0 = 1$. This density has been frequently used for investigating the finite-sample behaviors of adaptive estimates, as in Kreiss (1987a) and Shin and So (1999). In the simulation, $\gamma_0 = -0.5, 0.5$ and 0.8 for the stationary case, and 1.0 for the nonstationary case, and $(\alpha_0, \beta_0) = (0.57, 0.02)$. We use the optimal bandwidth a_n given in Silverman [(1986), page 40] which is automatically searched from the data and $b_n = 0.0001a_n^{1/3}$. The reason for the choice of a_n and b_n was given by Shin and So (1999). The sample size is $n = 250$, and 1000 replications are used. Since the performance of the AE for (α_0, β_0) in finite samples has been investigated in Drost and Klaassen (1997), we report here only the results for γ_0 in Table 1. In this table, the efficient estimator (EE) is constructed as in Theorem 4.1 and the QMLE is described as in Ling and Li (1998). From these results, we can see that the AE and EE are much more efficient than the LSE and QMLE, while the AE and EE are very similar. Meanwhile, the biases of the AE and EE are smaller than those of the LSE and QMLE.

REMARK 4.3. The adaptive estimator $\hat{\gamma}_n$ of γ_0 can be used to construct a unit root test. From Theorems 3.1(b) and 4.2, we have $n(\hat{\gamma}_n - 1) \xrightarrow{\mathcal{L}} \int_0^1 w_1(\tau) dw_2(\tau) / \kappa \Omega_\gamma \int_0^1 w_1^2(\tau) d\tau$. Let

$$B_1(\tau) = \frac{1}{\sigma_\varepsilon} w_1(\tau)$$

and

$$B_2(\tau) = -\frac{1}{\sigma_\varepsilon^2} \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 \Omega_\gamma - 1}} w_1(\tau) + \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 \Omega_\gamma - 1}} w_2(\tau),$$

where $\sigma_\varepsilon^2 = E\varepsilon_t^{02}$. Then $B_1(\tau)$ and $B_2(\tau)$ are two independent standard Brownian motions. As shown in Ling and Li (1998), we can show that

$$(4.4) \quad n(\hat{\gamma}_n - 1) \xrightarrow{\mathcal{L}} \frac{\int_0^1 B_1(\tau) dB_1(\tau)}{\sigma_\varepsilon^2 \Omega_\gamma \kappa \int_0^1 B_1^2(\tau) d\tau} + \frac{\sqrt{\sigma_\varepsilon^2 \Omega_\gamma - 1}}{\sigma_\varepsilon^2 \Omega_\gamma \kappa} \frac{\int_0^1 B_1(\tau) dB_2(\tau)}{\int_0^1 B_1^2(\tau) d\tau}.$$

The second term in (4.4) can be simplified to $[\sqrt{\sigma_\varepsilon^2 \Omega_\gamma - 1}/(\sigma_\varepsilon^2 \Omega_\gamma \kappa)] \times (\int_0^1 B_1^2(\tau) d\tau)^{-1/2} \xi$, where ξ is a standard normal random variable independent of $\int_0^1 B_1^2(\tau) d\tau$ [see Phillips (1989)]. Let $\hat{\tau}_{AEn} = S_{\gamma_n}^{-1/2}[n(\hat{\gamma}_n - 1)]$. Then we have

$$(4.5) \quad \hat{\tau}_{AEn} \xrightarrow{\mathcal{L}} \rho \frac{\int_0^1 B_1(\tau) dB_1(\tau)}{\sqrt{\int_0^1 B_1^2(\tau) d\tau}} + \sqrt{1 - \rho^2} \xi,$$

where $\rho = 1/\sqrt{\sigma_\varepsilon^2 \Omega_\gamma} \in (0, 1)$. The asymptotic distribution of $\hat{\tau}_{AEn}$ depends on a nuisance parameter ρ . Its critical values can be obtained through the simulation method, with the estimated $\hat{\rho}$ as given in Hansen (1995) and Shin and So (1999).

Testing for unit roots has been a mainstream topic in econometrics for quite some time, so it is important to find more powerful tests for both theory and application. For the AR model with i.i.d. errors, the popular Dickey–Fuller (henceforth DF) test based on LSE has been widely used. For the AR–GARCH model, the DF test still is valid for the hypothesis $H_0: \gamma_0 = 1$ [see Ling, Li and McAleer (2002)]. The QMLE in Ling and Li (1998) may be used to construct the unit root test: $\hat{\tau}_{QEn} = (\sigma_\varepsilon^2 K_2 \rho)(\sum_{i=2}^n y_{t-1}^2)^{1/2}(\hat{\gamma}_{QEn} - 1)$, which has the same asymptotic distribution as (4.5) with $\rho = (\sigma_\varepsilon^2 K_c)^{-1/2}$, where $\hat{\gamma}_{QEn}$ denotes the QMLE of γ_0 , $K_u = E(1/h_t^0) + u\alpha_0^2 \sum_{k=1}^\infty \beta_0^{2(k-1)} E(\varepsilon_{t-k}^2/h_t^{02})$, and $c = E\eta_t^4 - 1$. Since QMLE is more efficient than LSE, $\hat{\tau}_{QEn}$ should be more powerful than the DF test. Note that the AE is more efficient than both the QMLE and LSE. It is expected that the $\hat{\tau}_{AEn}$ test is more powerful than both the DF test and $\hat{\tau}_{QEn}$.

To confirm our conjecture, we present a small simulation experiment for these unit root tests. Using the same model as in (4.3) with the same sample size, replications, and the same method for the choice of a_n and b_n , we investigate the size for $\gamma_0 = 1.0$ and local powers for $\gamma_0 = 0.95, 0.97, 0.98$ and 0.99 . The critical values of the DF test come from Table 8.5.3 in Fuller (1976). The critical values of $\hat{\tau}_{QEn}$ and $\hat{\tau}_{AEn}$ are generated through 20000 replications of an i.i.d. bivariate $N(0, I_2)$ process. From Table 2, it is clear that the sizes of the three tests are very close to the nominal 5% and 10% levels, and that their powers are consistent with our expectations.

TABLE 2
The power and size of lower tail unit root tests for AR(1)–GARCH(1, 1) models;
 $n = 250$, and 1000 replications

γ_0	Significance level 5%					Significance level 10%				
	0.950	0.970	0.980	0.990	1.000	0.950	0.970	0.980	0.990	1.000
DF test	0.904	0.549	0.319	0.166	0.055	0.974	0.775	0.534	0.285	0.120
$\hat{\tau}_{QEn}$	0.997	0.894	0.643	0.219	0.038	1.000	0.973	0.839	0.437	0.078
$\hat{\tau}_{AEn}$	1.000	0.999	0.995	0.895	0.060	1.000	0.999	0.997	0.954	0.094

5. Proofs of Theorems 3.1 and 3.2. For simplicity, we assume that the initial values are $y_i = 0$; $\varepsilon_i = 0$ and $h_i = \omega_0$ for $i \leq 0$, which does not make any essential difference to the proof. We first introduce some lemmas. Lemma 5.1 comes directly from Theorem 2.1 in Ling and Li (1997) and Theorem 6.2 in Ling (1999), which gives the basic properties of the process (w_t^0, h_t^0) . Lemma 5.2 gives the expansion of $(w_t, h_t)(\lambda)$. Lemma 5.3 is a basic result for verifying Assumption 2.2.

LEMMA 5.1. Under Assumptions 3.1 and 3.2, the process (w_t^0, h_t^0) is strictly stationary and ergodic and almost surely has the following causal expansions:

$$(a) \quad w_t^0 = \sum_{i=0}^{\infty} v_{\phi_0 \psi_0}(i) \varepsilon_{t-i}^0$$

$$(b) \quad h_t^0 = \iota' \zeta_t + \iota' \sum_{i=1}^{\infty} \prod_{j=0}^{i-1} A_{t-j} \zeta_{t-j},$$

where $\varepsilon_t^0 = \eta_t \sqrt{h_t^0}$, $A_t = A_t(\lambda_0)$, $\iota = (0, \dots, 0, 1, 0, \dots, 0)'_{(r+s) \times 1}$ with the $(r+1)$ st element being 1, $\zeta_t = \zeta_t(\lambda_0)$ and $\zeta_t(\lambda) = (\alpha_0 \eta_t^2(\lambda), 0, \dots, 0, \alpha_0, 0, \dots, 0)'$, with the first and $(r+1)$ st elements being $\alpha_0 \eta_t^2$ and α_0 , respectively. Furthermore, if Assumption 3.3 holds, then ε_t^0 and w_t^0 have finite fourth moments.

LEMMA 5.2. If Assumptions 3.1 and 3.2 hold, then under P_λ , $(w_t, h_t)(\lambda)$ has the following expansions:

$$(a) \quad w_t(\lambda) = \sum_{i=0}^{t-1} v_{\phi \psi}(i) \varepsilon_{t-i}(\lambda),$$

$$(b) \quad h_t(\lambda) = \iota' \zeta_t(\lambda) + \iota' \sum_{j=1}^{t-1} \prod_{i=0}^{j-1} A_{t-i}(\lambda) \zeta_{t-j}(\lambda) + \iota' \prod_{i=0}^{t-1} A_{t-i}(\lambda) \tilde{\varepsilon}_0,$$

where $\varepsilon_t(\lambda) = \eta_t(\lambda) \sqrt{h_t(\lambda)}$, ι and $\zeta_t(\lambda)$ are defined as in Lemma 5.1, and $\tilde{\varepsilon}_0 = (0, \dots, 0, \omega_0, \dots, \omega_0)'$ with the last s elements being ω_0 . Furthermore, if Assumption 3.3 holds, then $E(h_t(\lambda_0) - h_t^0)^2 = O(\rho^t)$, $E(\varepsilon_t(\lambda_0) - \varepsilon_t^0)^2 = O(\rho^t)$

and $E(w_t(\lambda_0) - w_t^0)^2 = O(\rho^t)$, where $O(\cdot)$ holds uniformly in all t , $t \geq 1$, and $0 < \rho < 1$.

PROOF. Under P_λ , model (3.2) can be rewritten as

$$(5.1) \quad \tilde{\varepsilon}_t(\lambda) = \zeta_t(\lambda) + A_t(\lambda)\tilde{\varepsilon}_{t-1}(\lambda),$$

where $\tilde{\varepsilon}_t(\lambda) = [\varepsilon_t^2(\lambda), \dots, \varepsilon_{t-r+1}^2(\lambda), h_t(\lambda), \dots, h_{t-s+1}(\lambda)]'$. After iterating (5.1) t -steps, we show that (b) holds. Similarly, it can be shown that (a) holds. By expansion (b) of this lemma and Assumption 3.3, we can show that

$$(5.2) \quad E(h_t(\lambda_0) - h_t^0)^2 = E \left[t' \prod_{i=0}^{t-1} A_{t-i} \tilde{\varepsilon}_0 - t' \sum_{j=t}^{\infty} \prod_{i=0}^{j-1} A_{t-i} \zeta_{t-j} \right]^2 = O(\rho^t).$$

By (5.2) and expansion (a) of this lemma, the other cases can be proved. This completes the proof. \square

LEMMA 5.3. *If Assumptions 3.1–3.3 hold, then it follows that:*

$$(a) \quad \max_{1 \leq t \leq n} |n^{-1/2} y_t| = O_{\lambda_0}(1),$$

$$(b) \quad n^{-1/2} \max_{1 \leq t \leq n} w_t^2(\lambda_n) = o_{\lambda_0}(1).$$

PROOF. By Lemma 5.2(a), under P_{λ_0} ,

$$\begin{aligned} \frac{1}{\sqrt{n}} y_{[n\tau]} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{[n\tau]} w_i(\lambda_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[n\tau]} \sum_{j=0}^{i-1} v_{\phi_0 \psi_0}(j) \varepsilon_{i-j}(\lambda_0) \\ &= \frac{1}{\sqrt{n}} \left(\sum_{i=0}^{[n\tau]} v_{\phi_0 \psi_0}(i) \right) \left(\sum_{j=1}^{[n\tau]} \varepsilon_j(\lambda_0) \right) + \frac{1}{\sqrt{n}} R_{1n}(\tau) \\ &= [1 - \phi_0(1)]^{-1} \psi_0(1) \frac{1}{\sqrt{n}} \sum_{j=1}^{[n\tau]} \varepsilon_j(\lambda_0) + \frac{1}{\sqrt{n}} R_{1n}(\tau) + \frac{1}{\sqrt{n}} R_{2n}(\tau), \end{aligned}$$

where $R_{1n} = \sum_{i=0}^{[n\tau]} v_{\phi_0 \psi_0}(i) (\sum_{j=[n\tau]-i+1}^{[n\tau]} \varepsilon_j(\lambda_0))$ and $R_{2n} = (\sum_{i=[n\tau]+1}^{\infty} v_{\phi_0 \psi_0}(i)) \times (\sum_{j=1}^{[n\tau]} \varepsilon_j(\lambda_0))$. By Lemma 5.2(a), we have $n^{-1/2} \max_{1 \leq t \leq n} |\varepsilon_t(\lambda_0)| = n^{-1/2} \times \max_{1 \leq t \leq n} |\varepsilon_t^0| + o_{\lambda_0}(1)$. Since ε_t^0 is strictly stationary with a finite variance, $n^{-1/2} \max_{1 \leq t \leq n} |\varepsilon_t^0| = o_{\lambda_0}(1)$ [see Chung (1968), page 93]. Thus, by Assumption 3.1, it is easy to show that $n^{-1/2} \max_{0 \leq \tau \leq 1} |R_{1n}(\tau)| = o_{\lambda_0}(1)$ and $n^{-1/2} \max_{0 \leq \tau \leq 1} |R_{2n}(\tau)| = o_{\lambda_0}(1)$. Furthermore, by Lemma 5.4 below and the continuity theorem, (a) holds.

Now we show that (b) holds. Under P_{λ_0} , we have $w_t(\lambda_n) = w_t(\lambda_0) - \theta_{1n} y_{t-1}/n$, where θ_{1n} is the first component of θ_n . By (a) of this lemma,

$n^{-1/2} \max_{1 \leq t \leq n} (\theta_{1n} y_{t-1}/n)^2 = o_{\lambda_0}(1)$. By Lemma 5.2(a), we have $n^{-1/2} \times \max_{1 \leq t \leq n} w_t^2(\lambda_0) = n^{-1/2} \max_{1 \leq t \leq n} w_t^{02} + o_{\lambda_0}(1)$. Since w_t^{02} is strictly stationary with a finite variance (see Lemma 5.1), $n^{-1/2} \max_{1 \leq t \leq n} w_t^{02} = o_{\lambda_0}(1)$ [see Chung (1968), page 93)]. Thus (b) holds. This completes the proof. \square

PROOF OF THEOREM 3.1(a). Since it is assumed that Y_0 is independent of (λ, χ) , Assumption 2.3 is obviously satisfied. By Theorem 2.1 and Remark 2.1, it is sufficient to verify Assumption 2.2 with $\lambda_n = \lambda_0$ and (2.2). First, (i) obviously holds. The proofs of (ii)–(iv) and (2.2) mainly use Assumption 3.1 and Lemma 5.3, and some basic inequalities. Since the techniques are similar, only the proof of (2.2) is presented. We need to prove that

$$(5.3) \quad \sum_{t=1}^n \left\| G_n^{-1} \left[\frac{\partial \varepsilon_t(\lambda_n)}{\partial \lambda} - \frac{\partial \varepsilon_t(\lambda_0)}{\partial \lambda} \right] \right\|^2 = o_{\lambda_0}(1),$$

$$(5.4) \quad \sum_{t=1}^n \left\| \frac{G_n^{-1}}{\sqrt{h_t(\lambda_n)}} \frac{\partial h_t(\lambda_n)}{\partial \lambda} - \frac{G_n^{-1}}{\sqrt{h_t(\lambda_0)}} \frac{\partial h_t(\lambda_0)}{\partial \lambda} \right\|^2 = o_{\lambda_0}(1),$$

where, by Assumptions 3.1 and 3.2, we can show that

$$\begin{aligned} \frac{\partial \varepsilon_t(\lambda)}{\partial \gamma} &= - \sum_{i=0}^{t-1} v_\psi(i) y_{t-i-1}, & \frac{\partial \varepsilon_t(\lambda)}{\partial m} &= - \sum_{i=0}^{t-1} v_\psi(i) \tilde{w}_{t-i-1}(\lambda), \\ \frac{\partial h_t(\lambda)}{\partial \tilde{m}} &= 2 \sum_{i=1}^{t-1} v_{\alpha\beta}(i) \varepsilon_{t-i}(\lambda) \frac{\partial \varepsilon_{t-i}(\lambda)}{\partial \tilde{m}}, & \frac{\partial h_t(\lambda)}{\partial \tilde{\delta}} &= \sum_{i=0}^{t-1} v_\beta(i) \begin{pmatrix} u_{0t-i}(\lambda) \\ \alpha_0 \tilde{\varepsilon}_{t-i-1}(\lambda) \end{pmatrix}, \end{aligned}$$

where $\tilde{m} = (\gamma, m')'$, $\tilde{w}_t(\lambda) = [w_t(\lambda), \dots, w_{t-p+2}(\lambda), \varepsilon_t(\lambda), \dots, \varepsilon_{t-q+1}(\lambda)]'$, $u_{0t}(\lambda) = 1 + \sum_{i=1}^r \alpha_i \varepsilon_{t-i}^2(\lambda) + \sum_{i=1}^s \beta_i h_{t-i}(\lambda)$, and $\tilde{\varepsilon}_t(\lambda)$ is defined as in (5.1). Again, since these proofs are similar, we prove only (5.4).

Note that $h_t(\lambda)$ has a lower bound uniformly in all t and in a neighborhood of λ_0 . By Taylor's expansion, it can be shown that (5.4) is bounded by

$$(5.5) \quad \left[\left\| G_n \frac{\partial^2 h_t(\lambda_n^*)}{\partial \lambda \partial \lambda'} G_n \right\|^2 + \left\| G_n \frac{\partial h_t(\lambda_n^*)}{\partial \lambda} \right\|^4 \right] O(1),$$

where $\lambda_n^* = \lambda_0 + \tilde{\kappa}_n G_n \theta_n$ with $|\tilde{\kappa}_n| < 1$, and $O(1)$ holds uniformly in all t . By Assumptions 3.1 and 3.2, it is direct to show that (5.5) is bounded by

$$(5.6) \quad \begin{aligned} O(1) & \left[\frac{1}{n^2} \max_{1 \leq t \leq n} y_t^4 + \frac{1}{n^{3/2}} \max_{1 \leq t \leq n} y_t^2 \right. \\ & \left. + \frac{1}{n} \max_{1 \leq t \leq n} w_t^4(\lambda_n^*) + \frac{1}{\sqrt{n}} \max_{1 \leq t \leq n} w_t^2(\lambda_n^*) \right], \end{aligned}$$

where $O(1)$ holds uniformly in all t . By (5.5), (5.6) and Lemma 5.3, we can know that (5.4) holds. This completes the proof. \square

Now, we introduce an invariance principle. For this, we need the following notation: $\tilde{X}_t(\lambda) = [u'_{\gamma t}(\lambda), u'_{mt}(\lambda), u'_{\delta t}(\lambda)]'$ with $u_{\gamma t}(\lambda) = -[h_t^{-1/2}(\lambda), h_t^{-1}(\lambda) \times \sum_{i=1}^{t-1} \nu_{\alpha\beta}(i)\varepsilon_{t-i}(\lambda)]$, $u_{mt}(\lambda) = [h_t^{-1/2}(\lambda)\partial\varepsilon_t(\lambda)/\partial m, (2h_t(\lambda))^{-1}\partial h_t(\lambda)/\partial m]$, and $u_{\delta t}(\lambda) = [0, (2h_t(\lambda))^{-1}\partial h_t(\lambda)/\partial \delta]$.

LEMMA 5.4. *Suppose the assumptions of Theorem 3.1 hold. Then,*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[n\tau]} [\varepsilon_t, (\tilde{X}_t \xi(\eta_t))', \zeta'(\eta_t)]'(\lambda_0) \xrightarrow{\mathcal{L}} (w_1, w_2, N'_{m\delta}, N'_\zeta)'(\tau) \\ \text{in } D^{k+l+1}[0, 1],$$

under P_{λ_0} , where $(\omega_1, \omega_2, N'_{m\delta}, N'_\zeta)'(\tau)$ are defined as in Theorem 3.1, and $D[0, 1]$ denotes the Skorokhod space.

PROOF. By Lemma 5.2, it follows that

$$\varepsilon_t(\lambda_0) - \varepsilon_t^0 = O_{\lambda_0}(\rho^t) \quad \text{and} \quad \tilde{X}_t(\lambda_0) - X_t^0 = O_{\lambda_0}(\rho^t),$$

where $0 < \rho < 1$, X_t^0 is defined as in Theorem 3.1, and $O_{\lambda_0}(\cdot)$ holds uniformly in all t . Thus,

$$(5.7) \quad \begin{aligned} & \frac{1}{\sqrt{n}} \sum_{t=1}^{[n\tau]} [\varepsilon_t, (\tilde{X}_t \xi(\eta_t))', \zeta'(\eta_t)]'(\lambda_0) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^{[n\tau]} [\varepsilon_t^0, (X_t^0 \xi(\eta_t))', \zeta'(\eta_t)]' + o_{\lambda_0}(1), \end{aligned}$$

where $o_{\lambda_0}(1)$ holds uniformly in all $\tau \in [0, 1]$. Denote $W_t^{*0} = [\varepsilon_t^0, (X_t^0 \xi(\eta_t))', \zeta'(\eta_t)]'$. Then W_t^{*0} is a strictly stationary and ergodic martingale difference with $E(W_t^{*0} W_t^{*0'}) = \tilde{\Omega}$. It is easy to verify that $n^{-1/2} \sum_{t=1}^{[n\tau]} W_t^{*0}$ satisfies the conditions of Theorem 4.1 in Hall and Heyde (1980), and hence $n^{-1/2} \sum_{t=1}^{[n\tau]} W_t^{*0}$ converges to $(w_1, w_2, N'_{m\delta}, N'_\zeta)'(\tau)$ in $D^{k+l+1}[0, 1]$. Furthermore, by (5.7), we complete the proof. \square

PROOF OF THEOREM 3.1(b). Since $\tilde{\Omega} > 0$, it is obvious that $\tilde{S} > 0$ a.s.. As in the proof of Theorem 4.1 in Ling and Li (1998), we can show that

$$(5.8) \quad \begin{aligned} & \frac{1}{n} \sum_{t=1}^n \frac{\partial g_{t-1}(\lambda_0)}{\partial \gamma} \xi'(\eta_t) = \frac{1}{n} \sum_{t=1}^n y_{t-1} u_{\gamma t}^0 \xi'(\eta_t) + o_{\lambda_0}(1) \\ & \xrightarrow{\mathcal{L}} \kappa \int_0^1 w_1(\tau) dw_2(\tau), \end{aligned}$$

where the last step holds by Theorem 2.2 in Kurtz and Protter (1991) and Lemma 5.4. Similarly, we have

$$\frac{1}{n^2} \sum_{t=1}^n \frac{\partial g_{t-1}(\lambda_0)}{\partial \gamma} V_{\xi, \xi} \frac{\partial g'_{t-1}(\lambda_0)}{\partial \gamma} = \frac{1}{n^2} \sum_{t=1}^n y_{t-1}^2 u_{\gamma t}^0 V_{\xi, \xi} u_{\gamma t}^{0'} + o_{\lambda_0}(1).$$

Denote $\sigma_{\gamma t} = u_{\gamma t}^0 V_{\xi, \xi} u_{\gamma t}^{0'}$. Using a similar technique as in Theorem 3.4 in Ling and Li (1998), we can show that, under P_{λ_0} ,

$$(5.9) \quad \frac{1}{n} \sum_{t=1}^n |\sigma_{\gamma t} - E\sigma_{\gamma t}| = O_{\lambda_0}(1), \quad \frac{1}{\sqrt{n}} \max_{1 \leq t \leq n} |\sigma_{\gamma t} - E\sigma_{\gamma t}| = o_{\lambda_0}(1),$$

and

$$(5.10) \quad \frac{1}{n} E \left[\sum_{t=1}^n (\sigma_{\gamma t} - E\sigma_{\gamma t}) \right]^2 \rightarrow \sigma_0^2,$$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor n\tau \rfloor} (\sigma_{\gamma t} - E\sigma_{\gamma t}) \xrightarrow{\mathcal{L}} \sigma_0 \omega_0(\tau) \quad \text{in } D[0, 1],$$

where σ_0 is a nonnegative constant and $\omega_0(\tau)$ is a standard Brownian motion. By Theorem 3.1 in Ling and Li (1998), (5.9), (5.10), Lemma 5.4 and the continuity theorem, it follows that

$$(5.11) \quad \begin{aligned} & \frac{1}{n^2} \sum_{t=1}^n \frac{\partial g_{t-1}(\lambda_0)}{\partial \gamma} V_{\xi, \xi} \frac{\partial g'_{t-1}(\lambda_0)}{\partial \gamma} \\ &= \frac{E\sigma_{\gamma t}}{n^2} \sum_{t=1}^n y_{t-1}^2 + \frac{1}{n^2} \sum_{t=1}^n y_{t-1}^2 [u_{\gamma t}^0 V_{\xi, \xi} u_{\gamma t}^{0'} - E\sigma_{\gamma t}] + o_{\lambda_0}(1) \\ &= \frac{E\sigma_{\gamma t}}{n^2} \sum_{t=1}^n y_{t-1}^2 + o_{\lambda_0}(1) \xrightarrow{\mathcal{L}} \kappa^2 E\sigma_{\gamma t} \int_0^1 w_1^2(\tau) d\tau. \end{aligned}$$

Similarly, we can show that, under P_{λ_0} ,

$$(5.12) \quad \frac{1}{n^{3/2}} \sum_{t=1}^n \frac{\partial g_{t-1}(\lambda_0)}{\partial \gamma} V_{\xi, \xi} \frac{\partial g_{t-1}(\lambda_0)}{\partial m'} \xrightarrow{\mathcal{L}} E(u_{\gamma t}^0 V_{\xi, \xi} u_{m t}^{0'}) \kappa \int_0^1 w_1(\tau) d\tau,$$

$$(5.13) \quad \frac{1}{n^{3/2}} \sum_{t=1}^n \frac{\partial g_{t-1}(\lambda_0)}{\partial \gamma} V_{\xi, \xi} \frac{\partial g_{t-1}(\lambda)}{\partial \tilde{\delta}'} \xrightarrow{\mathcal{L}} E(u_{\gamma t}^0 V_{\xi, \xi} u_{\tilde{\delta} t}^{0'}) \kappa \int_0^1 w_1(\tau) d\tau,$$

$$(5.14) \quad \frac{1}{n^{3/2}} \sum_{t=1}^n \frac{\partial g_{t-1}(\lambda_0)}{\partial \gamma} V'_{\zeta, \xi} \xrightarrow{\mathcal{L}} E u_{\gamma t}^0 V'_{\zeta, \xi} \kappa \int_0^1 w_1(\tau) d\tau.$$

Denote $u_{m\tilde{\delta}} = \begin{pmatrix} u_{m\tilde{\delta}} \\ u_{\tilde{\delta}t} \end{pmatrix}(\lambda_0)$ and $u_{m\tilde{\delta}}^0 = \begin{pmatrix} u_{m\tilde{\delta}}^0 \\ u_{\tilde{\delta}t}^0 \end{pmatrix}$. By Lemma 5.2 and the ergodic theorem, we can show that

$$(5.15) \quad \begin{aligned} \frac{1}{n} \sum_{t=1}^n (u_{m\tilde{\delta}} V_{\xi,\xi} u'_{m\tilde{\delta}}) &= \frac{1}{n} \sum_{t=1}^n (u_{m\tilde{\delta}}^0 V_{\xi,\xi} u_{m\tilde{\delta}}^{0'}) + o_{\lambda_0}(1) \\ &= E(u_{m\tilde{\delta}}^0 V_{\xi,\xi} u_{m\tilde{\delta}}^{0'}) + o_{\lambda_0}(1), \end{aligned}$$

$$(5.16) \quad \frac{1}{n} \sum_{t=1}^n u_{m\tilde{\delta}} V'_{\xi,\xi} = \frac{1}{n} \sum_{t=1}^n u_{m\tilde{\delta}}^0 V'_{\xi,\xi} + o_{\lambda_0}(1) = E u_{m\tilde{\delta}}^0 V'_{\xi,\xi} + o_{\lambda_0}(1).$$

By Theorem 2.2 in Kurtz and Protter (1991) and Lemma 5.4, all the limiting distributions involved in $\tilde{W}_n(\lambda_0)$ and $\tilde{S}_n(\lambda_0)$ are jointly convergent. Finally, by Lemma 5.4, (5.8) and (5.11)–(5.16), we can show that $(\tilde{W}_n, \tilde{S}_n)(\lambda_0)$ converges weakly to $(\tilde{W}, \tilde{S}) = [\int_0^1 M(\tau) dB(\tau), \int_0^1 M(\tau) \Sigma M'(\tau) d\tau]$ under P_{λ_0} . This completes the proof. \square

PROOF OF THEOREM 3.2. It is obvious that the estimator \hat{v}_n belongs to \mathcal{M}_v . Let $\bar{v}_{\pi n}$ be any M_v -estimator corresponding to the functional $\pi(x)$. Denote $\varepsilon_{\pi t}^*(\lambda_0) = u_{\gamma t}(\lambda_0)[\pi_1(\eta_t), \pi_2(\eta_t)]'$. As in the proof of Lemma 5.4, we can show that, under P_{λ_0} , $n^{-1/2} \sum_{t=1}^{[n\tau]} (\varepsilon_t, \varepsilon_{\pi t}^*)(\lambda_0)$ converges to $(\omega_1, \omega_{\pi 2})(\tau)$ in $D^2[0, 1]$, where $(\omega_1, \omega_{\pi 2})(\tau)$ is a bivariate Brownian motion with mean zero and covariance $\tau \begin{pmatrix} E h_t^0 & 1 \\ 1 & \Omega_{\pi\gamma}^* \end{pmatrix}$, and $\Omega_{\pi\gamma}^* = E(u_{\gamma t}^0 V_{\pi,\pi} u_{\gamma t}^{0'})$ with $V_{\pi,\pi} \equiv \text{diag}\{J_{\pi 1}, J_{\pi 2}\}$. Denote $\Omega_{\pi m}^* = E(u_{m\tilde{\delta}}^0 V_{\pi,\pi} u_{m\tilde{\delta}}^{0'})$, $\Omega_{\pi\tilde{\delta}}^* = E(u_{\tilde{\delta}t}^0 V_{\pi,\pi} u_{\tilde{\delta}t}^{0'})$, $\Omega_{\pi\gamma} = E(u_{\gamma t}^0 V_{\pi,\xi} u_{\gamma t}^{0'})$, $\Omega_{\pi m} = E(u_{m\tilde{\delta}}^0 V_{\pi,\xi} u_{m\tilde{\delta}}^{0'})$ and $\Omega_{\pi\tilde{\delta}} = E(u_{\tilde{\delta}t}^0 V_{\pi,\xi} u_{\tilde{\delta}t}^{0'})$ with $V_{\pi,\xi} \equiv \text{diag}\{c_1, c_2\}$. Under Assumptions 2.1 and 3.1–3.4, as in Weiss (1986) and Ling and Li (1997), we can show that the matrices $\Omega_{\pi m}^*$, $\Omega_{\pi\tilde{\delta}}^*$, $\Omega_{\pi m}$, and $\Omega_{\pi\tilde{\delta}}$ are positive definite, and $(W_n, S_n)(\lambda_0, \pi)$ converges weakly to

$$(W, S)(\pi) = \begin{pmatrix} \kappa \int_0^1 \omega_1(\tau) d\omega_{\pi 2}(\tau) & \kappa^2 \Omega_{\pi\gamma} \int_0^1 \omega_1^2(\tau) d\tau & 0 & 0 \\ N_{\pi 1} & 0 & \Omega_{\pi m} & 0 \\ N_{\pi 2} & 0 & 0 & \Omega_{\pi\tilde{\delta}} \end{pmatrix}$$

under P_{λ_0} , with $N_{\pi 1}$ and $N_{\pi 2}$ being two independent normal vectors with mean zero and covariances $\Omega_{\pi m}^*$ and $\Omega_{\pi\tilde{\delta}}^*$, respectively, independent of $(\omega_1, \omega_{\pi 2})(\tau)$. Thus, $G_n^*(\bar{v}_{\pi n} - v_0)$ converges weakly to $G_\pi = \dot{v}_0 S^{-1}(\pi) W(\pi)$ under P_{λ_0} , and

$$E(G_\pi G_\pi') = \dot{v}_0 \text{diag}(\Sigma_{\pi\gamma}, \Sigma_{\pi m}, \Sigma_{\pi\tilde{\delta}}) \dot{v}_0',$$

where $\Sigma_{\pi\gamma} = \kappa^{-2} \Omega_{\pi\gamma}^{-2} E[\int_0^1 w_1(\tau) d\omega_{\pi 2}(\tau) / \int_0^1 w_1^2(\tau) d\tau]^2$, $\Sigma_{\pi m} = (c_1 P + c_2 Q)^{-1} (P J_{\pi 1} + Q J_{\pi 2}) (c_1 P + c_2 Q)^{-1}$ and $\Sigma_{\pi\tilde{\delta}} = J_{\pi 2} c_2^{-2} R^{-1}$, with

$$P = E\left(\frac{1}{h_t^0} \frac{\partial \varepsilon_t^0}{\partial m} \frac{\partial \varepsilon_t^0}{\partial m'}\right), \quad Q = \frac{1}{4} E\left(\frac{1}{h_t^0} \frac{\partial h_t^0}{\partial m} \frac{\partial h_t^0}{\partial m'}\right)$$

and

$$R = \frac{1}{4} E \left(\frac{1}{h_t^{02}} \frac{\partial h_t^0}{\partial \tilde{\delta}} \frac{\partial h_t^0}{\partial \tilde{\delta}'} \right).$$

Let $\dot{v}_0 S_n^{-1}(\lambda_0) W_n(\lambda_0) \xrightarrow{\mathcal{L}} G$. From Theorem 3.1(b) and Remark 3.3, we obtain

$$E(GG') = \dot{v}_0 \text{diag}(\Sigma_\gamma, \Sigma_m, \Sigma_{\tilde{\delta}}) \dot{v}_0',$$

where $\Sigma_\gamma = \kappa^{-2} \Omega_\gamma^{-2} E[\int_0^1 w_1(\tau) dw_2(\tau) / \int_0^1 w_1^2(\tau) d\tau]^2$, $\Sigma_m = (P I_1(f) + Q I_2(f))^{-1}$ and $\Sigma_{\tilde{\delta}} = I_2^{-1}(f) R^{-1}$. By the Cauchy inequality and the definition of c_1 and c_2 , we know that $I_1(f) J_{\pi_1} \geq c_1^2$ and $I_2(f) J_{\pi_2} \geq c_2^2$. It is obvious that $\Sigma_{\tilde{\delta}} \leq \Sigma_{\pi \tilde{\delta}}$. After some algebra, we have

$$\begin{aligned} \Sigma_m - \Sigma_{\pi m} &= (c_1 P + c_2 Q)^{-1} [P(P J_{\pi_1} + Q J_{\pi_2})^{-1} P(c_1 P + c_2 Q)^{-1} (c_1^2 - I_1(f) J_{\pi_1}) \\ &\quad + Q(P J_{\pi_1} + Q J_{\pi_2})^{-1} Q(c_1 P + c_2 Q)^{-1} (c_2^2 - I_2(f) J_{\pi_2}) \\ &\quad + P(P J_{\pi_1} + Q J_{\pi_2})^{-1} Q(c_1 P + c_2 Q)^{-1} (c_1 c_2 - I_1(f) J_{\pi_2}) \\ &\quad + Q(P J_{\pi_1} + Q J_{\pi_2})^{-1} P(c_1 P + c_2 Q)^{-1} (c_1 c_2 - I_2(f) J_{\pi_1})] \\ (5.17) \quad &\leq (c_1 P + c_2 Q)^{-1} [P(P J_{\pi_1} + Q J_{\pi_2})^{-1} Q(c_1 P + c_2 Q)^{-1} (c_1 c_2 - I_1(f) J_{\pi_2}) \\ &\quad + Q(P J_{\pi_1} + Q J_{\pi_2})^{-1} P(c_1 P + c_2 Q)^{-1} (c_1 c_2 - I_2(f) J_{\pi_1})] \\ &\leq (c_1 P + c_2 Q)^{-1} (Q^{-1} J_{\pi_1} + P^{-1} J_{\pi_2})^{-1} (c_1 P + c_2 Q)^{-1} \\ &\quad \times [2c_1 c_2 - 2(I_2(f) J_{\pi_1} I_1(f) J_{\pi_2})^{1/2}] \leq 0. \end{aligned}$$

Now, we show that $\Sigma_\gamma \leq \Sigma_{\pi \gamma}$. Let $\varepsilon_t^{**}(\lambda_0) = \varepsilon_t^*(\lambda_0) / \Omega_\gamma$ and $\varepsilon_{\pi t}^{**}(\lambda_0) = \varepsilon_{\pi t}^*(\lambda_0) / \Omega_{\pi \gamma}$, where $\varepsilon_t^*(\lambda_0) = u_{\gamma t}(\lambda_0) \xi(\eta_t)$. As in the proof of Lemma 5.4, we can show that, in $D^2[0, 1]$, $n^{-1/2} \sum_{t=1}^{\lfloor n\tau \rfloor} (\varepsilon_t^{**}, \varepsilon_{\pi t}^{**})(\lambda_0)$ converges to the bivariate Brownian motion $(\omega_2^*, \omega_{\pi 2}^*)(\tau)$, which has mean zero and covariance

$$(5.18) \quad \tau \begin{pmatrix} \Omega_\gamma^{-1} & \Omega_\gamma^{-1} \\ \Omega_\gamma^{-1} & \Omega_{\pi \gamma}^* \Omega_{\pi \gamma}^{-2} \end{pmatrix}.$$

Denote $W_{\gamma n}$ and $W_{\pi \gamma n}$ as the first elements of $W_n(\lambda_0)$ and $W_n(\lambda_0, \pi)$, respectively, and $S_{\gamma n}$ and $S_{\pi \gamma n}$ as the $(1, 1)$ th elements of $S_n(\lambda_0)$ and $S_n(\lambda_0, \pi)$, respectively. From the proof of Theorem 3.1(b), we see that the asymptotic distributions of $S_{\gamma n}^{-1} W_{\gamma n}$ and $S_{\pi \gamma n}^{-1} W_{\pi \gamma n}$ are the same as the distributions of $G_\gamma \equiv [\kappa \int_0^1 \omega_1^2(\tau) d\tau]^{-1} \int_0^1 \omega_1(\tau) d\omega_2^*(\tau)$ and $G_{\pi \gamma} \equiv [\kappa \int_0^1 \omega_1^2(\tau) d\tau]^{-1} \times \int_0^1 \omega_1(\tau) d\omega_{\pi 2}^*(\tau)$, respectively. Let $\Delta = \Omega_{\pi \gamma}^* \Omega_{\pi \gamma}^{-2} - \Omega_\gamma^{-1}$. As in the proof of (5.17), we can show that $\Delta \geq 0$. Using (5.18) and Lemma 3.1 of Phillips (1989), we can show that the distribution of $G_{\pi \gamma}$ is the same as that of $G_\gamma +$

$\kappa^{-1} \Delta^{1/2} [\int_0^1 \omega_1^2(\tau) d\tau]^{-1/2} \Phi$, where Φ is standard normal and independent of G_γ and $\int_0^1 \omega_1^2(\tau) d\tau$. Thus, $\Sigma_{\pi\gamma} = E[G_{\pi\gamma}^2] = E[G_\gamma^2] + \kappa^{-2} \Delta E[\int_0^1 \omega_1^2(\tau) d\tau]^{-1} \geq E[G_\gamma^2] = \Sigma_\gamma$. Finally, we have that $E(GG') \leq E(G_\pi G'_\pi)$. This completes the proof. \square

6. Proof of Theorem 4.2. By Lemma 4.1, it is sufficient to prove that

$$(6.1) \quad \hat{I}_{in}(\lambda_n) = I_i(f) + o_{\lambda_0}(1), \quad i = 1, 2,$$

$$(6.2) \quad \hat{W}_{1n}(\lambda_n) = W_{1n}(\lambda_n) + o_{\lambda_0}(1),$$

where λ_n is defined as in Lemma 4.1. By Theorem 2.1(b) and Theorem 3.1(a), $P_{\lambda_0, n}$ and $P_{\lambda_n, n}$ are contiguous. Note that $\hat{W}_{1n}(\lambda_n)$ and $W_{1n}(\lambda_n)$ are measurable in terms of \mathcal{F}_n . Thus, it is sufficient to prove (6.1), (6.2) under $P_{\lambda_n, n}$.

Let $\xi_n = f'_n/(b_n + f_n)$ with $f_n(x) = \int f(x - a_n z) K(z) dz$, and $\hat{\xi}_{1n}(x, \lambda) = \hat{f}'_{a_n}(x, \lambda)/[b_n + \hat{f}_{a_n}(x, \lambda)]$ with $\hat{f}_{a_n}(x, \lambda) = \sum_{i=1}^n [K(x + \eta_i(\lambda), a) + K(x - \eta_i(\lambda), a)]/2n$. It follows from work in Schick [(1987), pages 99 and 100] and Schick and Susarla (1988) that

$$(6.3) \quad \sup_{x \in R} E_{\lambda_n} [\hat{\xi}_{1n}(x, \lambda_n) - \xi_n(x)]^2 = O(n^{-1} a_n^{-3} b_n^{-1}),$$

$$(6.4) \quad \int (1+x^2) [\xi_n(x) - \xi_1(x)]^2 f(x) dx \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

$$(6.5) \quad \sup_{x \in R} |\hat{\xi}_{1n}(x, \lambda) - \hat{\xi}_{1n,t}(x, \lambda)| \leq \frac{c}{n a_n^2 b_n},$$

for some constant c and all $t = 1, \dots, n$. From (6.3) and (6.5), it follows that

$$(6.6) \quad \frac{1}{n} \sum_{t=1}^n E_{\lambda_n} \{ \eta_t^{2i}(\lambda_n) [\hat{\xi}_{1n,t}(\eta_t(\lambda_n)) - \xi_n(\eta_t(\lambda_n))]^2 \} = O(n^{-1} a_n^{-3} b_n^{-1}),$$

where $i = 0, 1$. By (6.4) and (6.6), we know that (6.1) holds.

In the following, we show that (6.2) holds. By (6.3)–(6.5) and $n a_n^3 b_n \rightarrow \infty$, it follows that

$$(6.7) \quad \Xi_{1,n} \equiv E_{\lambda_n} \int (1+x^2) [\hat{\xi}_{1n}(x, \lambda_n) - \xi_1(x)]^2 f(x) dx = o(1)$$

and

$$(6.8) \quad \Xi_{2,n} \equiv \max_{1 \leq t \leq n} \int (1+x^2) |\hat{\xi}_{1n}(x, \lambda_n) - \hat{\xi}_{1n,t}(x, \lambda_n)|^2 f(x) dx \leq \frac{c}{n^2 a_n^4 b_n^2}$$

for some constant c . Let $c_n = \max\{5cn^{-2}a_n^{-4}b_n^{-2}, \Xi_{1,n}^{1/4}\}$,

$$(6.9) \quad A_{n,t} = \left\{ \int (1+x^2) [\hat{\xi}_{1n,t}(x, \lambda) - \xi_1(x)]^2 f(x) dx \leq c_n \right\}$$

and $A_n = \bigcap_{t=1}^n A_{n,t}$. By the Markov inequality,

$$\begin{aligned}
 P_{\lambda_n, n}(A_n^c) &\leq P_{\lambda_n, n}\left(\int (1+x^2)[\hat{\xi}_{1n}(x, \lambda) - \xi_1(x)]^2 f(x) dx \geq c_n/4\right) \\
 (6.10) \quad &+ P_{\lambda_n, n}(\Xi_{2,n} \geq c_n/4) \\
 &\leq \frac{16\Xi_{1,n}}{c_n^2} \leq 16\Xi_{1,n}^{1/2} \rightarrow 0,
 \end{aligned}$$

as $n \rightarrow \infty$. For any $\epsilon > 0$,

$$\begin{aligned}
 (6.11) \quad &P_{\lambda_n, n}(\|\hat{W}_{1n}(\lambda_n) - W_{1n}(\lambda_n)\| \geq \epsilon) \\
 &\leq P_{\lambda_n, n}(A_n^c) + \epsilon^{-2} E_{\lambda_n}[\|\hat{W}_{1n}(\lambda_n) - W_{1n}(\lambda_n)\|^2 I_{A_n}].
 \end{aligned}$$

By (6.10) and (6.11), it is sufficient for (6.2) to prove that

$$(6.12) \quad E_{\lambda_n}[\|\hat{W}_{1n}(\lambda_n) - W_{1n}(\lambda_n)\|^2 I_{A_n}] = o(1).$$

For simplicity, we denote $\hat{\xi}_{in,t}(\eta_t(\lambda_n), \lambda_n)$ as $\hat{\xi}_{it}$, $i = 1, 2$. By the triangle inequality,

$$\begin{aligned}
 (6.13) \quad &E_{\lambda_n}\{[\hat{W}_{\gamma n}(\lambda_n) - W_{\gamma n}(\lambda_n)]I_{A_n}\}^2 \\
 &\leq 2E_{\lambda_n}\left\{\frac{1}{n} \sum_{t=1}^n \left[\frac{1}{\sqrt{h_t(\lambda_n)}} \frac{\partial \varepsilon_t(\lambda_n)}{\partial \gamma} (\hat{\xi}_{1t} - \xi_1(\eta_t(\lambda_n)))\right] I_{A_n}\right\}^2 \\
 &+ \frac{1}{2} E_{\lambda_n}\left\{\frac{1}{n} \sum_{t=1}^n \left[\frac{1}{h_t(\lambda_n)} \frac{\partial h_t(\lambda_n)}{\partial \gamma} (\hat{\xi}_{2t} - \xi_2(\eta_t(\lambda_n)))\right] I_{A_n}\right\}^2 \\
 &\equiv 2B_{1n} + \frac{1}{2} B_{2n}.
 \end{aligned}$$

Note that $\hat{\xi}_{1t}$ and $\xi_1(\eta_t(\lambda_n))$ are odd functions of $\eta_t(\lambda_n)$. We have that

$$\begin{aligned}
 (6.14) \quad &B_{1n} = \frac{1}{n^2} \sum_{t=1}^n E_{\lambda_n} \left\{ \frac{1}{h_t(\lambda_n)} [\hat{\xi}_{1t} - \xi_1(\eta_t(\lambda_n))]^2 I_{A_n} \left| \frac{\partial \varepsilon_t(\lambda_n)}{\partial \gamma} \right|^2 \right\} \\
 &\leq \frac{c_n}{n^2} \sum_{t=1}^n E_{\lambda_n} \left\{ \frac{1}{h_t(\lambda_n)} \left| \frac{\partial \varepsilon_t(\lambda_n)}{\partial \gamma} \right|^2 \right\}.
 \end{aligned}$$

Since $h_t^{-1}(\lambda_n)$ is bounded and $\partial \varepsilon_t(\lambda_n)/\partial \gamma = \sum_{j=0}^{t-1} \nu_{\psi_n}(j) y_{t-j-1}$, by Assumption 3.1, we can show that

$$(6.15) \quad B_{1n} \leq O(1) \frac{c_n}{n^2} \sum_{t=1}^n \max_{1 \leq j \leq n} E_{\lambda_n} y_j^2.$$

Under $P_{\lambda_n, n}$, $y_j = \sum_{i=1}^j (1 - \theta_{1n}/n)^{j-i} w_i(\lambda_n)$ and $w_i(\lambda_n) = \sum_{j=0}^{i-1} v_{\phi_n \psi_n}(j) \times \varepsilon_{i-j}(\lambda_n)$. Thus, we can show that, as $n \rightarrow \infty$,

$$(6.16) \quad B_{1n} \leq O(1) \frac{c_n}{n} \sum_{t=1}^n \max_{1 \leq j \leq n} E_{\lambda_n} \varepsilon_j^2(\lambda_n) = O(1) c_n E_{\lambda_n} \varepsilon_j^2(\lambda_n) = o(1).$$

To study B_{2n} , let $\mathcal{G}_{n,t}$ be the σ -field generated by $\{|\eta_1(\lambda_n)|, \dots, |\eta_n(\lambda_n)|\}$ and $\{Y_0, y_1, \dots, y_t\}$. Since f is symmetric under $P_{\lambda_n, n}$, we have that

$$(6.17) \quad E_{\lambda_n}[\eta_t(\lambda_n) | \mathcal{G}_{n,t}] = 0.$$

We now consider the cross-terms in the expansion of B_{2n} . Denote $\xi_t^* = [\eta_t(\lambda_n) \hat{\xi}_{1t} - \eta_t(\lambda_n) \xi_1(\eta_t(\lambda_n))] I_{A_n}$. We have that

$$\begin{aligned} H_{nti} &\equiv E_{\lambda_n} \left[\frac{1}{h_{t+i}(\lambda_n) h_t(\lambda_n)} \frac{\partial h_{t+i}(\lambda_n)}{\partial \gamma} \frac{\partial h_t(\lambda_n)}{\partial \gamma} \xi_{t+i}^* \xi_t^* \right] \\ &= E_{\lambda_n} \left\{ \frac{2}{h_{t+i}(\lambda_n) h_t(\lambda_n)} \left[\sum_{j=1}^i v_{\alpha_n \beta_n}(j) \varepsilon_{t+i-j}(\lambda_n) y_{t+i-j-1} \right] \frac{\partial h_t(\lambda_n)}{\partial \gamma} \xi_{t+i}^* \xi_t^* \right. \\ &\quad \left. + \frac{4}{h_{t+i}(\lambda_n) h_t(\lambda_n)} \left[\sum_{j=i+1}^{t+i-1} v_{\alpha_n \beta_n}(j) \varepsilon_{t+i-j}(\lambda_n) y_{t+i-j-1} \right] \right. \\ &\quad \left. \times \left[\sum_{j=1}^{t-1} v_{\alpha_n \beta_n}(j) \varepsilon_{t-j}(\lambda_n) y_{t-j-1} \right] \xi_{t+i}^* \xi_t^* \right\}. \end{aligned}$$

Since $\hat{\xi}_{2t}$ and $\xi_2(\eta_t(\lambda_n))$ are symmetric functions of $\eta_t(\lambda_n)$, it follows that $\sqrt{h_{t+i-j}(\lambda_n)} y_{t+i-j-1} [\sum_{j_1=1}^{t-1} v_{\alpha_n \beta_n}(j_1) \varepsilon_{t-j_1}(\lambda_n) y_{t-j_1-1}] \xi_{t+i}^* \xi_t^* / h_{t+i}(\lambda_n) h_t(\lambda_n)$ is the even function of $\eta_{t+i-j}(\lambda_n)$ when $i+1 \leq j \leq t+i-1$. Since A_n is $\mathcal{G}_{n,t}$ -measurable for any $t = 1, \dots, n$, and $\varepsilon_{t+i-j}(\lambda_n) = \eta_{t+i-j}(\lambda_n) \sqrt{h_{t+i-j}(\lambda_n)}$, by (6.17), we have that

$$H_{nti} = E_{\lambda_n} \left\{ \frac{4}{h_{t+i}(\lambda_n) h_t(\lambda_n)} \left[\sum_{j=1}^{t-1} v_{\alpha_n \beta_n}(j+i) v_{\alpha_n \beta_n}(j) \varepsilon_{t-j}^2(\lambda_n) y_{t-j-1}^2 \right] \xi_{t+i}^* \xi_t^* \right\}.$$

Furthermore, since $v_{\alpha_n \beta_n}(i) = O(\rho^i)$ with $\rho \in (0, 1)$ and independent of λ_n , we have

$$(6.18) \quad |H_{nti}| \leq O(1) \rho^i \max_{1 \leq j < t} E_{\lambda_n} [\varepsilon_j^2(\lambda_n) y_{j-1}^2 | \xi_{t+i}^* \xi_t^*],$$

where $O(1)$ holds uniformly in all t . By Assumption 3.1, (6.18) and the inequality $2|\xi_{t+i}^* \xi_t^*| \leq \xi_{t+i}^{*2} + \xi_t^{*2}$, we have

$$\begin{aligned}
 B_{2n} &= \frac{1}{n^2} \sum_{t=1}^n E_{\lambda_n} \left\{ \frac{1}{h_t^2(\lambda_n)} \left[\frac{\partial h_t(\lambda_n)}{\partial \gamma} \right]^2 \xi_t^{*2} \right\} + \frac{2}{n^2} \sum_{t=1}^{n-1} \sum_{i=1}^{n-t} H_{nti} \\
 &\leq \frac{O(1)}{n^2} \left\{ \sum_{t=1}^n \max_{1 \leq j < t} E_{\lambda_n} [\varepsilon_j^2(\lambda_n) y_{j-1}^2 \xi_t^{*2}] \right. \\
 &\quad \left. + \sum_{t=1}^{n-1} \sum_{i=1}^{n-t} \rho^i \max_{1 \leq j < t} E_{\lambda_n} [\varepsilon_j^2(\lambda_n) y_{j-1}^2 \xi_{t+i}^{*2}] \right\} \\
 (6.19) \quad &\leq \frac{O(1)c_n}{n^2} \left\{ \sum_{t=1}^n \max_{1 \leq j \leq n} E_{\lambda_n} [\varepsilon_j^2(\lambda_n) y_{j-1}^2] \right. \\
 &\quad \left. + \sum_{t=1}^{n-1} \sum_{i=1}^{n-t} \rho^i \max_{1 \leq j \leq n} E_{\lambda_n} [\varepsilon_j^2(\lambda_n) y_{j-1}^2] \right\} \\
 &= O(1) \frac{c_n}{n} \max_{1 \leq j \leq n} E_{\lambda_n} [\varepsilon_j^2(\lambda_n) y_{j-1}^2].
 \end{aligned}$$

Note that $E_{\lambda_n} [\varepsilon_j^2(\lambda_n) \varepsilon_i(\lambda_n) \varepsilon_{i_1}(\lambda_n)] = 0$ for any $i \neq i_1$. We can show that

$$(6.20) \quad B_{2n} = O(1)c_n E_{\lambda_n} [\varepsilon_j^4(\lambda_n)] = o(1).$$

By (6.13), (6.16) and (6.20), we can obtain $E_{\lambda_n} [|\hat{W}_{\gamma n}(\lambda_n) - W_{\gamma n}(\lambda_n)|^2 I_{A_n}] = o(1)$. In a similar manner, we can obtain that $E_{\lambda_n} [\|\hat{W}_{mn}(\lambda_n) - W_{mn}(\lambda_n)\|^2 I_{A_n}] = o(1)$. To complete the proof, it is sufficient to show that

$$(6.21) \quad E_{\lambda_n} \|\hat{W}_{\delta n}(\lambda_n) - W_{\delta n}(\lambda_n)\|^2 = o(1).$$

Since the logistic kernel $K(x)$ for $\hat{W}_{\delta n}(\lambda_n)$ satisfies the conditions in Theorem 4.1 in Koul and Schick (1997), the proof of (6.21) is similar to that of Theorem 3.1 in DWK (1997) and hence is omitted. This completes the proof.

APPENDIX: PROOF OF THEOREM 2.1.

Before giving the proof of Theorem 2.1, we introduce the following notation and lemma. Let $\tilde{\xi} = (\xi', \zeta')'$, $\tilde{G}_n = \text{diag}(G_n, \sqrt{n}I_l)$, $\tilde{U}_t(\lambda) = \text{diag}(X_t(\lambda), I_l)$ and $Y(\lambda) = [y - Z_t(\lambda)]/\sqrt{h_t(\lambda)}$. For simplicity, we denote $Y(\lambda_n)$ by Y and $Y(\tilde{\lambda}_n)$ by Y_n . Similarly, denote h_t , h_{nt} , Z_t , Z_{nt} , g_t , g_{nt} , η_t , η_{nt} , \tilde{U}_t and \tilde{U}_{nt} .

LEMMA A.1. *Under the assumptions of Theorem 2.1, it follows that:*

$$\begin{aligned}
 (a) \quad & \sum_{t=1}^n E_{\lambda_n} [(u'_n \tilde{G}_n^{-1} \tilde{U}_t \tilde{\xi}(\eta_t))^2 | \mathcal{F}_{t-1}] = O_{\lambda_n}(1), \\
 (b) \quad & \sum_{t=1}^n E_{\lambda_n} [(u'_n \tilde{G}_n^{-1} \tilde{U}_t \tilde{\xi}(\eta_t))^2 I_{\{|u'_n \tilde{G}_n^{-1} \tilde{U}_t \tilde{\xi}(\eta_t)| > \varepsilon\}} | \mathcal{F}_{t-1}] = o_{\lambda_n}(1), \\
 (c) \quad & \left| \sum_{t=1}^n \{ (u'_n \tilde{G}_n^{-1} \tilde{U}_t \tilde{\xi}(\eta_t))^2 - E_{\lambda_n} [(u'_n \tilde{G}_n^{-1} \tilde{U}_t \tilde{\xi}(\eta_t))^2 | \mathcal{F}_{t-1}] \} \right| = o_{\lambda_n}(1), \\
 (d) \quad & \sum_{t=1}^n \int \left\{ \frac{\sqrt{f_{v_n/\sqrt{n}}(Y_n)}}{\sqrt[4]{h_{nt}}} - \frac{\sqrt{f(Y)}}{\sqrt[4]{h_t}} \right. \\
 & \quad \left. - \frac{1}{2} u'_n \tilde{G}_n^{-1} \tilde{U}_t \frac{(\tilde{\xi} \sqrt{f})(Y)}{\sqrt[4]{h_t}} \right\}^2 dy = o_{\lambda_n}(1).
 \end{aligned}$$

PROOF. By Assumptions 2.1 and 2.2(i) and (iv), and the finiteness of $\int \|\zeta(x)\|^2 f(x) dx$, (a) holds. By Assumptions 2.1 and 2.2, and using a similar argument as in Koul and Schick [(1997), page 253], we can show that (b) holds. By (3.15) of McLeish (1974) and (a) and (b) of this lemma, (c) holds.

The proof of (d) is similar to that for (2.15) in Koul and Schick (1997). The right-hand side of (d) is bounded by $3(T_{1n} + T_{2n} + T_{3n})$, where

$$\begin{aligned}
 T_{1n} &= \sum_{t=1}^n \int h_{nt}^{-1/2} \left(f_{v_n/\sqrt{n}}^{1/2} - f^{1/2} - \frac{v'_n}{2\sqrt{n}} \zeta f^{1/2} \right)^2 (Y_n) dy \\
 &= n \int \left(f_{v_n/\sqrt{n}}^{1/2} - f^{1/2} - \frac{v'_n}{2\sqrt{n}} \zeta f^{1/2} \right)^2 (x) dx = o(1)
 \end{aligned}$$

by Definition 2.1;

$$\begin{aligned}
 T_{2n} &= \frac{\|v_n\|^2}{n} \sum_{t=1}^n \int \|h_{nt}^{-1/4} (\zeta f^{1/2})(Y_n) - h_t^{-1/4} (\zeta f^{1/2})(Y)\|^2 dy \\
 &\leq O(1) R_{2n}^2 \int \|(\zeta f^{1/2})(x)\|^2 dx \\
 &\quad + O(1) \sup_{|s_1| \leq R_{1n}, |s_2| \leq R_{2n}} \int \|(\zeta f^{1/2})(x(1+s_2) + s_1) - (\zeta f^{1/2})(x)\|^2 dx \\
 &= o_{\lambda_n}(1),
 \end{aligned}$$

where

$$R_{1n} = \left[\max_{1 \leq t \leq n} (|Z_{nt} - Z_t| h_{nt}^{-1/2})^2 \right]^{1/2} = o_{\lambda_n}(1)$$

and

$$R_{2n} = \left[\max_{1 \leq t \leq n} (|\sqrt{h_{nt}} - \sqrt{h_t}| h_{nt}^{-1/2})^2 \right]^{1/2} = o_{\lambda_n}(1)$$

by Assumption 2.2(i)–(iii), and the above equation holds by $\int \|\zeta(x)\|^2 \times f(x) dx < \infty$ and Lemma 19 in Jeganathan (1995), and

$$T_{3n} = \sum_{t=1}^n \int [h_{nt}^{-1/4} f^{1/2}(Y_n) - h_t^{-1/4} f^{1/2}(Y) - \frac{1}{2} \vartheta'_n G_n^{-1} U_t h_t^{-3/4} (\xi f^{1/2})(Y)]^2 dy.$$

In order to show that $T_{3n} = o_{\lambda_n}(1)$, denote $U_{nt}^* = g_{nt} - g_t$, $Y_n^* = [y - Z_t - u(Z_{nt} - Z_t)] h_{nt}^{*-1/2}$ and $h_{nt}^* = [h_t^{1/2} + u(h_{nt}^{1/2} - h_t^{1/2})]^2$. By Assumption 2.1 and using Cauchy's form of Taylor's theorem to the function $f^*(u) = h_{nt}^{*-1/4} f^{1/2}(Y_n^*)$, T_{3n} is bounded by

$$\begin{aligned} & \sum_{t=1}^n \int_0^1 \int [U_{nt}^* h_{nt}^{*-3/4} (\xi f^{1/2})(Y_n^*) - \vartheta'_n G_n^{-1} U_t h_t^{-3/4} (\xi f^{1/2})(Y)]^2 dy du \\ (A.1) \quad & \leq \sum_{t=1}^n \int_0^1 \int \{ \vartheta'_n G_n^{-1} U_t [h_{nt}^{*-3/4} (\xi f^{1/2})(Y_n^*) - h_t^{-3/4} (\xi f^{1/2})(Y)] \}^2 dy du \\ & \quad + \sum_{t=1}^n \|U_{nt}^* - \vartheta'_n G_n^{-1} U_t\|^2 \max\{h_t^{-1}, h_{nt}^{-1}\} \int \|(\xi f^{1/2})(x)\|^2 dx, \end{aligned}$$

where the second term is $o_{\lambda_n}(1)$ by Assumptions 2.1 and 2.2(i) and (ii), and the first term is bounded by

$$\begin{aligned} & \left[O(1) \sup_{|s_1| \leq R_{1n}, |s_2| \leq R_{2n}} \int \|(\xi f^{1/2})(x(1+s_2) + s_1) - (\xi f^{1/2})(x)\|^2 dx \right. \\ & \quad \left. + O(1) R_{2n}^2 \int \|(\xi f^{1/2})(x)\|^2 dx \right] \sum_{t=1}^n \|G_n^{-1} U_{nt}\|^2 = o_{\lambda_n}(1) \end{aligned}$$

by Assumptions 2.1 and 2.2(iv) and Lemma 19 in Jeganathan (1995). Thus, $T_{3n} = o_{\lambda_n}(1)$ and hence (d) holds. This complete the proof. \square

Now, we prove Theorem 2.1. The basic idea of the proof comes from Le Cam (1970), Fabian and Hannan (1982), BKRW (1993) and DKW (1997).

PROOF OF THEOREM 2.1. Let $T_{nt} = 2[s_{v_n/\sqrt{n}, t}(\tilde{\lambda}_n)/s_t(\lambda_n) - 1]$ and $B_n = \{\max_{1 \leq t \leq n} |T_{nt}| < \varepsilon\}$ for some enough small $\varepsilon > 0$. Then, on the event B_n , the

log-LR has the Taylor expansion,

$$\begin{aligned}\Lambda_n\left(\lambda_n, \tilde{\lambda}_n, \frac{v_n}{\sqrt{n}}\right) &= 2 \sum_{t=1}^n \log\left(1 + \frac{1}{2}T_{nt}\right) + \Lambda_{n0} \\ &= \sum_{t=1}^n T_{nt} - \frac{1}{4} \sum_{t=1}^n T_{nt}^2 + \frac{1}{6} \sum_{t=1}^n \alpha_{nt} T_{nt}^3 + \Lambda_{n0},\end{aligned}$$

where $|\alpha_{nt}| < 1$ and $\Lambda_{n0} = \log[q_{\tilde{\lambda}_n, f_{v_n/\sqrt{n}}}(Y_0)/q_{\lambda_n, f}(Y_0)] = o_{\lambda_n}(1)$ by Assumption 2.3.

To prove (a), it is sufficient to show that

$$(A.2) \quad \sum_{t=1}^n \{T_{nt} - u'_n \tilde{G}_n^{-1} \tilde{U}_t \tilde{\xi}(\eta_t) + \frac{1}{4} E_{\lambda_n}[(u'_n \tilde{G}_n^{-1} \tilde{U}_t \tilde{\xi}(\eta_t))^2 | \mathcal{F}_{t-1}]\} = o_{\lambda_n}(1),$$

$$(A.3) \quad \sum_{t=1}^n \{T_{nt}^2 - E_{\lambda_n}[(u'_n \tilde{G}_n^{-1} \tilde{U}_t \tilde{\xi}(\eta_t))^2 | \mathcal{F}_{t-1}]\} = o_{\lambda_n}(1),$$

$$(A.4) \quad \max_{1 \leq t \leq n} |T_{nt}| = o_{\lambda_n}(1) \quad \text{and} \quad \sum_{t=1}^n T_{nt}^3 = o_{\lambda_n}(1).$$

Note that $\int [s_{v_n/\sqrt{n}, t}^y(\tilde{\lambda}_n) - s_t^y(\lambda_n)]^2 dy = -E_{\lambda_n}(T_{nt} | \mathcal{F}_n)$, where $s_t^y(\lambda)$ is defined as $s_t(\lambda)$ with $\eta_t(\lambda)$ replaced by Y , and similarly define $s_{v_n/\sqrt{n}, t}^y(\lambda)$. By Lemma A.1(a) and (d), and the inequality $|a^2 - b^2| \leq (1 + \alpha)(a - b)^2 + b^2/\alpha$ with $\alpha > 0$ and $a, b \in R$,

$$\begin{aligned}(A.5) \quad & \left| \sum_{t=1}^n \left\{ E_{\lambda_n}(T_{nt} | \mathcal{F}_{t-1}) + \frac{1}{4} E_{\lambda_n}[(u'_n \tilde{G}_n^{-1} \tilde{U}_t \tilde{\xi}(\eta_t))^2 | \mathcal{F}_{t-1}] \right\} \right| \\ & \leq (1 + \alpha) \sum_{t=1}^n \int \left[s_{v_n/\sqrt{n}, t}^y(\tilde{\lambda}_n) - s_t^y(\lambda_n) - \frac{1}{2} u'_n \tilde{G}_n^{-1} \tilde{U}_t \tilde{\xi}(Y_t) s_t^y(\lambda_n) \right]^2 dy \\ & \quad + \frac{1}{4\alpha} \sum_{t=1}^n E_{\lambda_n}[(u'_n \tilde{G}_n^{-1} \tilde{U}_t \tilde{\xi}(\eta_t))^2 | \mathcal{F}_{t-1}] \\ & = o_{\lambda_n}(1 + \alpha) + O_{\lambda_n}\left(\frac{1}{\alpha}\right) = o_{\lambda_n}(1),\end{aligned}$$

where the last equation holds by first letting $n \rightarrow \infty$ and then letting $\alpha \rightarrow \infty$.

Let $D_{nt} = T_{nt} - u'_n \tilde{G}_n^{-1} \tilde{U}_t \tilde{\xi}(\eta_t)$. $\sum_{t=1}^n E_{\lambda_n}\{[D_{nt} - E_{\lambda_n}(D_{nt} | \mathcal{F}_n)]^2 | \mathcal{F}_{t-1}\} \leq \sum_{t=1}^n E_{\lambda_n}(D_{nt}^2 | \mathcal{F}_{t-1}) = o_{\lambda_n}(1)$ by Lemma A.1(d), and hence $\sum_{t=1}^n [D_{nt} - E_{\lambda_n}(D_{nt} | \mathcal{F}_n)] = o_{\lambda_n}(1)$ by Remark 3.7(iii) in Fabian and Hannan (1982). Note that $E_{\lambda_n} \tilde{\xi}(\eta_t) = 0$. We have $\sum_{t=1}^n [D_{nt} - E_{\lambda_n}(T_{nt} | \mathcal{F}_n)] = o_{\lambda_n}(1)$. Furthermore,

by (A.5), we know that (A.2) holds. To (A.3), by Lemma A.1(c), it is sufficient to show that

$$(A.6) \quad \left| \sum_{t=1}^n \{T_{nt}^2 - [u_n' \tilde{G}_n^{-1} \tilde{U}_t \tilde{\xi}(\eta_t)]^2\} \right| = o_{\lambda_n}(1).$$

Note that $\sum_{t=1}^n E_{\lambda_n}[D_{nt}^2 I_{\{|D_{nt}| > \varepsilon\}} | \mathcal{F}_{t-1}] \leq \sum_{t=1}^n E_{\lambda_n}(D_{nt}^2 | \mathcal{F}_{t-1}) = o_{\lambda_n}(1)$ by Lemma A.1(d). By (3.15) of McLeish (1974), $\sum_{t=1}^n D_{nt}^2 = o_{\lambda_n}(1)$. Now, by Lemma A.1 and using a similar argument as for (A.5), we can show that (A.6) holds. By Lemmas A.1(b) and (d), and following the steps in DKW [(1997), page 794], we can show that $\max_{1 \leq t \leq n} |T_{nt}| = o_{\lambda_n}(1)$. By (A.3) and Lemma A.1(a), we have $\sum_{t=1}^n T_{nt}^2 = O_{\lambda_n}(1)$, and hence $\sum_{t=1}^n T_{nt}^3 = o_{\lambda_n}(1)$. Thus, (A.4) holds.

By (a) of this theorem, $\Lambda_n(\lambda_0, \lambda_n, 0) = u_n' \tilde{W}_n(\lambda_0) - u_n' \tilde{S}_n(\lambda_0) u_n / 2 + o_{\lambda_0}(1)$, and $\Lambda_n(\lambda_0, \lambda_n, 0) = -\Lambda_n(\lambda_n, \lambda_n + G_n'^{-1}(-\theta_n), 0) = -[\tilde{u}_n' \tilde{W}_n(\lambda_n) - \tilde{u}_n' \tilde{S}_n(\lambda_n) \times \tilde{u}_n / 2] + o_{\lambda_n}(1)$ with $\tilde{u}_n = (-\theta_n', 0)'$. By Assumptions 2.1 and 2.2, we can show that $\tilde{W}_n(\lambda_n) = O_{\lambda_n}(1)$ and $\tilde{S}_n(\lambda_n) = O_{\lambda_n}(1)$. Note that $\tilde{W}_n(\lambda_n)$ and $\tilde{S}_n(\lambda_n)$ are measurable in terms of \mathcal{F}_n and hence they are bounded under $P_{\lambda_n, n}$. Thus, $\Lambda_n(\lambda_0, \lambda_n, 0)$ is bounded under both $P_{\lambda_0, n}$ and $P_{\lambda_n, n}$, which implies (b). The first part of (c) holds by Assumption 2.2 and the second part holds by exploring the equation: $\Lambda_n(\lambda_0, \lambda_n, 0) + \Lambda_n(\lambda_n, \lambda_n + G_n'^{-1} \vartheta_n, v_n / \sqrt{n}) - \Lambda_n(\lambda_0, \lambda_n + G_n'^{-1} \vartheta_n, v_n / \sqrt{n}) = 0$. This completes the proof. \square

Acknowledgments. The authors thank Professor H. Kowl and two referees for their invaluable comments and suggestions.

REFERENCES

- BICKEL, P. J. (1982). On adaptive estimation. *Ann. Statist.* **10** 647–671.
- BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press, Baltimore.
- BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307–327.
- BOLLERSLEV, T., ENGLE, R. F. and NELSON, D. B. (1994). ARCH models. In *Handbook of Econometrics* **4** (R. F. Engle and D. C. McFadden, eds.) 2959–3038. Elsevier, Amsterdam.
- CHUNG, K. L. (1968). *A Course in Probability Theory*. Harcourt Brace, New York.
- DROST, F. C. and KLAASSEN, C. A. J. (1997). Efficient estimation in semiparametric GARCH models. *J. Econometrics* **81** 193–221.
- DROST, F. C., KLAASSEN, C. A. J. and WERKER, B. J. M. (1997). Adaptive estimation in time series models. *Ann. Statist.* **25** 786–817.
- ELLIOTT, G., ROTHENBERG, T. J. and STOCK, J. H. (1996). Efficient tests for an autoregression unit root. *Econometrica* **64** 813–836.
- ENGLE, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007.
- ENGLE, R. F. and GONZALEZ-RIVERA (1991). Semiparametric ARCH models. *J. Business Econom. Statist.* **9** 345–359.

- FABIAN, V. and HANNAN, J. (1982). On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrsch. Verw. Gebiete* **47** 31–46.
- FULLER, W. A. (1976). *Introduction to Statistical Time Series*. Wiley, New York.
- HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 175–194. Univ. California Press, Berkeley.
- HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press, San Diego.
- HANSEN, B. E. (1995). Rethinking the univariate approach to unit root testing: Using covariates to increase power. *Econom. Theory* **11** 1148–1171.
- JEGANATHAN, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econom. Theory* **11** 818–887.
- JEGANATHAN, P. (1997). On asymptotic inference in linear cointegrated time series systems. *Econom. Theory* **13** 692–745.
- KALLIANPUR, G. (1980). *Stochastic Filtering Theory*. Springer, New York.
- KOUL, H. L. and PFLUG, G. C. (1990). Weakly adaptive estimators in explosive autoregression. *Ann. Statist.* **18** 939–960.
- KOUL, H. L. and SCHICK, A. (1996). Adaptive estimation in a random coefficient autoregressive model. *Ann. Statist.* **24** 1025–1052.
- KOUL, H. L. and SCHICK, A. (1997). Efficient estimation in nonlinear autoregressive time-series models. *Bernoulli* **3** 247–277.
- KREISS, J.-P. (1987a). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15** 112–133.
- KREISS, J.-P. (1987b). On adaptive estimation in autoregressive models when there are nuisance functions. *Statist. Decisions* **5** 59–76.
- KURTZ, T. G. and PROTTER, P. (1991). Weak limit theorems to stochastic integrals and stochastic differential equations. *Ann. Probab.* **19** 1035–1070.
- LE CAM L. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.* **3** 27–98.
- LE CAM L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Statist.* **41** 802–828.
- LI, W. K., LING, S. and MCALEER, M. (2002). A survey of recent theoretical results for time series models with GARCH errors. *J. Econom. Surveys*. To appear.
- LING, S. (1999). On probability properties of a double threshold ARMA conditional heteroskedasticity model. *J. Appl. Probab.* **36** 688–705.
- LING, S. and LI, W. K. (1997). Fractional ARIMA-GARCH time series models. *J. Amer. Statist. Assoc.* **92** 1184–1194.
- LING, S. and LI, W. K. (1998). Limiting distributions of maximum likelihood estimators for unstable ARMA models with GARCH errors. *Ann. Statist.* **26** 84–125.
- LING, S., LI, W. K. and MCALEER, M. (2002). Estimation and testing for unit root processes with GARCH(1, 1) errors: Theory and Monte Carlo study. *Econometric Rev.* To appear.
- LING, S. and MCALEER, M. (2002). Necessary and sufficient moment conditions for the GARCH(r, s) and asymmetric power GARCH(r, s) models. *Econom. Theory* **18** 722–729.
- LINTON, O. (1993). Adaptive estimation in ARCH models. *Econom. Theory* **9** 535–569.
- MCLEISH, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **2** 602–628.
- NELSON, D. B. (1990). Stationarity and persistence in the GARCH(1, 1) model. *Econom. Theory* **6** 318–334.
- PHILLIPS, P. C. B. (1989). Partially identified econometric models. *Econom. Theory* **5** 181–240.
- PHILLIPS, P. C. B. (1991). Optimal inference in cointegrated systems. *Econometrica* **59** 283–306.
- ROBINSON, P. M. (1988). Semiparametric econometrics: A survey. *J. Appl. Econometrics* **3** 35–51.

- SCHICK, A. (1986). On asymptotically efficient estimation in semiparametric models. *Ann. Statist.* **14** 1139–1151.
- SCHICK, A. (1987). A note on the construction of asymptotically linear estimators. *J. Statist. Plann. Inference* **16** 89–105. [Correction (1989) **22** 269–270.]
- SCHICK, A. (1988). On estimation in LAMN families when there are nuisance parameters present. *Sankhyā Ser. A* **50** 249–268.
- SCHICK, A. and SUSARLA, V. (1988). Efficient estimation in some missing data problems. *J. Statist. Plann. Inference* **19** 217–228.
- SHIN, D. W. and SO, B. S. (1999). Unit root tests based on adaptive maximum likelihood estimation. *Econom. Theory* **15** 1–23.
- SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
- WEISS, A. A. (1986). Asymptotic theory for ARCH models: Estimation and testing. *Econom. Theory* **2** 107–131.

DEPARTMENT OF MATHEMATICS
HONG KONG UNIVERSITY
OF SCIENCE AND TECHNOLOGY
CLEAR WATER BAY
HONG KONG

DEPARTMENT OF ECONOMICS
UNIVERSITY OF WESTERN AUSTRALIA
PERTH
AUSTRALIA