Discussion Paper No. 548

ON ADAPTIVE ESTIMATION IN NONSTATIONARY ARMA MODELS WITH GARCH ERRORS

Shiqing Ling
and
Michael McAleer

July 2001

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

On Adaptive Estimation in Nonstationary ARMA Models with GARCH Errors *

Short Title: Adaptive Estimation For ARMA-GARCH

Shiqing Ling
Hong Kong University of Science and Technology

Michael McAleer
University of Western Australia

Abstract

This paper considers adaptive estimation in nonstationary autoregressive moving average models with the noise sequence satisfying a generalised 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. Adaptivity is discussed and it is found that the parameters in the model are generally not adaptively estimable if the density of the rescaled error is asymmetric. For the model with symmetric density of the rescaled error, a new efficiency criterion is established for a class of defined M_ν-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_ν-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 reparameterisation of the GARCH process.

Key Words and Phrases: Adaptive estimation, Efficient estimation, Nonstationary ARMA-GARCH models, Kernel estimators, Limiting distribution, Locally asymptotic quadratic, Log-likelihood ratio.

AMS 1980 subject classifications: Primary 62M10, 62E20; secondary 60F17.

*The authors wish to thank Professor H. Koul and two referees for their invaluable comments and suggestions, and to acknowledge the financial support of the Australian Research Council. This paper was revised while the second author was visiting the Institute of Social and Economic Research at Osaka University.

1 Introduction

Suppose that the observations $y_1, \cdots, y_n$, are generated by the autoregressive moving average (ARMA) model with errors generated by the generalized autoregressive conditional heteroscedasticity (GARCH) process:

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

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

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 $\alpha_{00} > 0$, $u_{01}, \cdots, u_{0r}$, $v_{01}, \cdots, v_{0s} \geq 0$. Models (1.1)-(1.2) is called the nonstationary 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 the remaining 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) generalised 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 (1999)). 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. non-normal 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 Gonzàlez-Rivera (1991) proposed a semiparametric estimator for models (1.1)-(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, for example, Koul and

Pflug (1990), Philips (1991), and Elliott, Rothenberg and Stock (1996)). 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 models (1.1)-(1.2), where we allow the ARMA model to have at most one unit root. We generalise 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). The adaptivity is discussed and it is found that the parameters in the model are generally not adaptively estimable if the density f is asymmetric. For the nonstationary ARMA-GARCH model, the efficient estimator defined in Fabian and Hannan (1982) is inappropriate. We define efficient estimators in a class of M_ν-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_ν-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. The full sample adaptive procedure can be seen as an extension of the method in Kreiss (1987a). However, since the ARMA model is nonstationary and the error is not i.i.d., his proof cannot easily be extended to the current situation.

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 5-6 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 $\longrightarrow_{\mathcal{L}}$ 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 R^k, and $\mathcal{B} = \{P_{\lambda,\chi} : (\lambda, \chi) \in \Theta \times \mathcal{D}\}$ be a family of probability measures, $y_1, y_2, \cdots, y_n$ 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 $\bar{Y}_t = (Y_0, y_1, \cdots, y_t)$ 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:

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

where the rescaled errors $\eta_1(\lambda)$, $\eta_2(\lambda)$, $\cdots$ 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 $c \to f_c$ be a map from a neighbourhood Δ of the origin in R^l into $\mathcal{D}$ such that $f_0 = f$. We say that $c \to f_c$ is a regular path if there exists a measurable function ζ from R to R^l such that $\int ||\zeta(x)||^2 f(x)dx < \infty$, $\int \zeta(x)\zeta'(x)f(x)dx$ is nonsingular, and*

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

Let $P^c_{\lambda,n}$ be the restriction of P_{λ,f_c} to $\mathcal{F}_n$, a $\sigma-$field generated by $\{Y_0, y_1, \cdots, y_n\}$. Denote $P^0_{\lambda,n}$ by $P_{\lambda,n}$. Define $\Lambda_n(\lambda_1, \lambda_2, c)$ as the log-LR of $P^c_{\lambda_2,n}$ 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 $s_t(\lambda) = s_{0,t}(\lambda)$. The following assumption ensures that the Fisher information is finite for both scale and location parameters.

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

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

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

Denote $g_t(\lambda) = (\varepsilon_t(\lambda), \sqrt{h_t(\lambda)})$, where $\varepsilon_t(\lambda) = y_t - Z_{t-1}(\lambda)$. Let G_n be a sequence of diagonal non-random $k \times k$ matrices depending on n but independent of λ and χ, θ_n and ϑ_n be two bounded sequences in R^k, $\lambda_n = \lambda_0 + G_n^{-1}\theta_n$ and $\tilde{\lambda}_n = \lambda_n + G_n^{-1}\vartheta_n$, and $X_t(\lambda) = h_t^{-1/2}(\lambda)U_t(\lambda)$, where $U_t(\lambda) = (u_{ijt}(\bar{Y}_{t-1}, \lambda))_{k\times 2}$ and u_{ijt} is a measurable function from $R^{p_0+t-1} \times \Theta$ to R. Furthermore, let

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

where $\xi = (\xi_1, -\xi_2)'$, $J = E[\xi(\eta_t)\xi'(\eta_t)]$, $V_{\xi\zeta} = E[\xi(\eta_t)\zeta'(\eta_t)]'$ and $V = E[\zeta(\eta_t)\zeta'(\eta_t)]$. We make the following assumptions.

Assumption 2.2. For any sequences θ_n and ϑ_n, it follows that:

(i) $$\left[\inf_{1\le t\le n} \sqrt{h_t(\lambda_n)}\right]^{-1} = O_{\lambda_n}(1),$$

(ii) $$\sum_{t=1}^{n}\left[g_t(\tilde{\lambda}_n) - g_t(\lambda_n) - (\tilde{\lambda}_n - \lambda_n)'U_t(\lambda_n)\right]^2 = o_{\lambda_n}(1),$$

(iii) $$\sup_{1\le t\le n}\left\|G_n^{-1}U_t(\lambda_n)\right\|^2 = o_{\lambda_n}(1),$$

(iv) $$\sum_{t=1}^{n}\left\|G_n^{-1}U_t(\lambda_n)\right\|^2 = O_{\lambda_n}(1).$$

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.

Now, we give the general LAQ criterion and its proof can be found in Appendix.

Theorem 2.1. *Suppose that the path $c \to f_c$ 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 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, Ch. 7, and Jeganathan 1995, p. 850). In this case, it is sufficient to verify Assumption 2.2 with $\lambda_n = \lambda_0$ and that

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

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: $\xi_*(x) = -\xi(x) + V_{\xi\zeta} V^{-1} \zeta(x)$ 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 favourable 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, under which sense the path is least favorable and the estimator is efficient need to be defined. After defined efficiency, the efficient estimator can be constructed by the split-sample method similarly as in DKW (1997) and Koul and Schick (1997). For models (1.1)-(1.2), the efficiency and adaptivity will be discussed in the next section.

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

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

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 below). As in Drost and Klaassen (1997), model (1.2) needs to be reparameterized. Thus, we assume that, under $P_{\lambda,\chi}$, $y_t, t = 1, \cdots, n$, satisfy the following structure:

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

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

with the initial (unobservable) vector $Y_0 = (y_0, \cdots, y_{1-p}, \varepsilon_0, \cdots, \varepsilon_{1-q^*}, h_0, \cdots, h_{1-s})$ and $q^* = \max\{r, q\}$, where the rescaled errors $\eta_1(\lambda), \eta_2(\lambda), \cdots$ are i.i.d. with density $\chi \in \mathcal{D}$ and independent of Y_0, $\lambda = (\gamma, m', \tilde{\delta}')'$, $m = (\phi', \psi')'$ with $\phi = (\phi_1, \cdots, \phi_{p-1})'$ and $\psi = (\psi_1, \cdots, \psi_q)'$, $\tilde{\delta} = (\alpha_0, \delta')'$ with $\delta = (\alpha_1, \cdots, \alpha_r, \beta_1, \cdots, \beta_s)'$, and the true parameter $(\lambda_0, f) \in \Theta \times \mathcal{D}$. We assume that, for simplicity, Y_0 is a constant or random vector independent of (λ, χ), and for each $\lambda \in \Theta$, it follows that:

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 $\alpha_0 \geq$ 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. $\rho\left[E_\lambda(A_t(\lambda) \otimes A_t(\lambda))\right] < 1$, 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 and Li (1997), Ling (1999), and Ling and McAleer (2000)). 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:

$$\phi^{-1}(z)\psi(z) = \sum_{i=0}^{\infty} \upsilon_{\phi\psi}(i)z^i,\ \ \psi^{-1}(z)\phi(z) = \sum_{i=0}^{\infty} \upsilon_{\psi\phi}(i)z^i,\ \ \psi^{-1}(z) = \sum_{i=0}^{\infty} \upsilon_{\psi}(i)z^i,$$
$$(1-\alpha_0\sum_{i=1}^{s}\beta_i z^i)^{-1} = \sum_{i=0}^{\infty} \upsilon_{\beta}(i)z^i,\ \ (1-\alpha_0\sum_{i=1}^{s}\beta_i z^i)^{-1}(\alpha_0\sum_{i=1}^{r}\alpha_i z^i) = \sum_{i=1}^{\infty} \upsilon_{\alpha\beta}(i)z^i,$$

where $\upsilon_{\phi_0\psi_0}(i)$ and $\upsilon_{\phi_n\psi_n}(i)$ denote $\upsilon_{\phi\psi}(i)$ with $\lambda = \lambda_0$ and λ_n, respectively. Similarly, define $\upsilon_{\psi_0\phi_0}(i)$, $\upsilon_{\psi_n\phi_n}(i)$, $\upsilon_{\psi_0}(i)$, $\upsilon_{\psi_n}(i)$, $\upsilon_{\beta_0}(i)$, $\upsilon_{\beta_n}(i)$, $\upsilon_{\alpha_0\beta_0}(i)$ and $\upsilon_{\alpha_n\beta_n}(i)$.

Let $(w_t^0, \varepsilon_t^0, h_t^0)$ be unobservable processes generated by the following equations:

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

where $t = 0, \pm 1, \pm 2, \cdots$. Then $(w_t^0, \varepsilon_t^0, h_t^0)$ is a fixed function of the $\{\eta_t\}$. We define:

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

where $\tilde{w}_t^0 = (w_t^0, \cdots, w_{t-p+2}^0, \varepsilon_t^0, \cdots, \varepsilon_{t-q+1}^0)'$ 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 $c \to f_c$ 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 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},*

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

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

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

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

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 $E(\Re_t\Re_t') > 0$ with $\Re_t = [\eta_t, \xi'(\eta_t), \zeta'(\eta_t)]'$. However, this condition excludes the normal density. If we further assume that the path satisfies: $\lim_{c\to 0}\int(1+x^4)f_c(x)dx = \int(1+x^4)f(x)dx$, 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:

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

where $(\omega_1, \omega_2)(\tau)$ is a bivariate Brownian motion with mean zero and covariance $\tau\Omega_1 = \tau\begin{pmatrix} Eh_t^0 & 1 \\ 1 & \Omega_\gamma \end{pmatrix}$; 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 $\begin{pmatrix} \Omega_{\tilde{\delta}} & V'_{\tilde{\delta}\zeta} \\ V_{\tilde{\delta}\zeta} & V \end{pmatrix}$, respectively, and independent of $(\omega_1, \omega_2)(\tau)$; and $\Omega_\gamma = E(u^0_{\gamma t} J u^{0'}_{\gamma t})$, $\Omega_m = E(u^0_{mt} J u^{0'}_{mt})$, $\Omega_{\tilde{\delta}} = E(u^0_{\tilde{\delta}t} J u^{0'}_{\tilde{\delta}t})$ with $J = \text{diag}(I_1(f), I_2(f))$, and $V'_{\tilde{\delta}\zeta} = E u^0_{\tilde{\delta}t} V'_{\xi\zeta}$. 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 models (3.1)-(3.2) belong 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.

The LABF in Theorem 3.1 can assist in understanding the adaptivity of parametric estimation for models (3.1)-(3.2). 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 estimates has the same asymptotic information matrix as the estimates 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. This motivates us to define adaptivity directly by the asymptotic distribution.

In the following definition, we suppose that Assumptions in Theorem 3.1 hold and the $\nu : \Theta \to R^{k_1}$, $k_1 \le k$, has a total differential $\dot{\nu}(\lambda)$, a $k_1 \times k$ matrix, such that there is a $k_1 \times k_1$ matrix G^*_n satisfying $G^*_n \dot{\nu}(\lambda_0) G_n^{-1} = \dot{\nu}(\lambda_0)$. $\nu(\lambda)$, $\nu(\lambda_0)$ and $\dot{\nu}(\lambda_0)$ are abbreviated as ν, ν_0 and $\dot{\nu}_0$, respectively.

Definition 3.1. *Let $\hat{\nu}_n$ be a sequence of estimates of ν_0 and $\mathcal{Q}$ be the set of all regular paths $c \to f_c$ such that $[(\hat{\nu}_n - \nu_0)' G^{*'}_n, \sqrt{n} c'_n]' = diag(\dot{\nu}_0, I_l) \tilde{S}_n^{-1}(\lambda_0) \tilde{W}_n(\lambda_0) + o_{\lambda_0}(1)$. $\hat{\nu}_n$ is called adaptive or, precisely, $(\mathcal{Q}, \mathcal{D})-$adaptive if, for every path in $\mathcal{Q}$,*

$$G^*_n(\hat{\nu}_n - \nu_0) \longrightarrow_{\mathcal{L}} \dot{\nu}_0 S^{-1} W \ \ \textit{under} \ \ P_{\lambda_0},$$

where (W, S) is the $k \times (k+1)$ upper left corner of $(\tilde{W}, \tilde{S})$.

This definition stresses only the fact that the estimator of ν_0 without the knowledge of the true density can achieve the same asymptotic distribution as its estimator when f is known. In this sense, the adaptive estimates have the same asymptotic distribution as the MLE, if the latter is available. The optimality of adaptive estimates will be discussed later. By Theorem 3.1, the necessary and sufficient condition for $\hat{\lambda}_n$, i.e. $\hat{\nu}_n$ with $\nu = \lambda$, to be adaptive is $C = 0$ for each path in $\mathcal{Q}$.

When f is asymmetric, $\hat{\lambda}_n$ is not $(\mathcal{Q}, \mathcal{D}_2)-$ adaptive, where $\mathcal{D}_2$ denotes the set of all densities that have zero means and finite variances. In fact, let $l^*(x) = \xi_2(x) + 2(x^2 - 1)/(u_4 - 1)$ with $u_4 = E\eta_t^4$, and $f_c = f\Psi(cl^*)/\int \Psi[cl^*(x)]f(x)dx$ with $\Psi(y) = 2(1 + e^{-2y})^{-1}$. It is easy to show that the map $c \to f_c$ is a regular path from $(-\varepsilon, \varepsilon)$ to $\mathcal{D}_2$, with $\zeta = l^*$ and ε sufficiently small. Since $E[(\eta_t^2 - 1)\xi_2(\eta_t)] = -2$, it can be shown that $E[\zeta_t(\eta_t)\xi_2(\eta_t)] = I_2(f) - 4/(u_4 - 1) > 0$ and $E(\Re_t\Re_t') > 0$. Along this path, $C \neq 0$ and hence our claim holds.

When $\mathcal{D}$ includes only densities that are symmetric about zero, from Remark 3.3, we can show that $\hat{\gamma}_n$ and $\hat{m}_n$ are adaptive. However, $\tilde{\delta}_0$ is not adaptively estimable in terms of $(\mathcal{Q}, \mathcal{D})$ as $V_{\tilde{\delta}\zeta} \neq 0$ along the same path as for the above case with asymmetric densities. Similarly, we can show that α_{00} is not adaptively estimable. After projecting the score functions for $\tilde{\delta}$ and c into that for α_0, we have $\sqrt{n}(\hat{\delta}_n - \delta_0) = S_{\delta n}^{-1}(\lambda_0)W_{\delta n}(\lambda_0) + o_{\lambda_0}(1) \to_{\mathcal{L}} N(0, \Omega_\delta^{-1})$ under P_{λ_0}, where

$$
\begin{aligned}
W_{\delta n}(\lambda) &= -\frac{1}{2\sqrt{n}} \sum_{t=1}^{n} \left[\dot{l}_{\delta t}(\lambda) - \hat{\mu}_\delta(\lambda)\right] \xi_2(\eta_t(\lambda)), \\
S_{\delta n}(\lambda) &= \frac{1}{4n} \Big\{ \sum_{t=1}^{n} \dot{l}_{\delta t}(\lambda)\dot{l}_{\delta t}'(\lambda) - [\sum_{t=1}^{n} \dot{l}_{\delta t}(\lambda)][\sum_{t=1}^{n} \dot{l}_{\delta t}'(\lambda)] \Big\} I_2(f),
\end{aligned}
$$

and $\Omega_\delta = [E(\dot{l}_{\delta t}^0 \dot{l}_{\delta t}^{0\prime}) - E(\dot{l}_{\delta t}^0)E(\dot{l}_{\delta t}^{0\prime})]I_2(f)/4$, 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 $\hat{\mu}_\delta(\lambda) = n^{-1}\sum_{t=1}^{n}[h_t^{-1}(\lambda)\partial h_t(\lambda)/\partial\delta]$. That is, $\hat{\delta}_n$ is adaptive. Similar findings were given in DWK (1997) and Drost and Klaassen (1997) for the ARCH (p) and GARCH (1,1) models, respectively. In addition, this indicates that the parameters in model (1.2) are not adaptively estimable if it is not reparameterised as model (3.2).

Basing on the above discussion, we are interested in the case with symmetric density and make the following assumption:

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

The optimal properties of our adaptive estimator are as yet unknown. In LAN models, Hȧjek (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 define a class of estimators, namely M_ν-estimators. Note that our focus is on the symmetric density f, so that the corresponding restrictions are made for the $\pi-$function in the following definition.

Definition 3.2. *Let $\pi(x) = [\pi_1(x), \pi_2(x)]'$ be a bivariate real function with odd $\pi_1(x)$ and even $\pi_2(x)$, such that $E[\pi(\eta_t)] = 0$, $E[\pi(\eta_t)\pi'(\eta_t)] = diag(I_{\pi_1}, I_{\pi_2}) > 0$, $c_1 \equiv E[\pi_1(\eta_t)\xi_1(\eta_t)] > 0$ and $c_2 \equiv E[\pi_2(\eta_t)\xi_2(\eta_t)] > 0$. An estimator $\bar{\nu}_n$ of ν_0 is said to be an $M_\nu-$estimator, regardless of whether f is known or unknown, if it has the asymptotic representation: $G_n^*(\bar{\nu}_n - \nu_0) = \dot{\nu}_0 S_{Mn}^{-1}(\lambda_0) W_{Mn}(\lambda_0) + o_{\lambda_0}(1)$, where $W_{Mn}(\lambda) = G_n^{-1}\sum_{t=1}^n X_t(\lambda)\pi(\eta_t(\lambda))$ and $S_{Mn}(\lambda) = \sum_{t=1}^n G_n^{-1} X_t(\lambda) J_c X_t'(\lambda) G_n^{-1}$ with $J_c = diag(c_1, c_2)$.*

M_ν-estimation is a very wide class and includes the QMLE, adaptive estimation and MLE (if available). Now, we define the optimality properties of M_ν-estimators and present an efficiency criterion for estimates in the class $\mathcal{M}_\nu$ below. Under this criterion, the adaptive estimates $\hat{\nu}_n$ in Definition 3.1 are efficient in $\mathcal{M}_\nu$. In the following definition, we suppose that Assumptions 2.1 and 3.1-3.4 hold, under which every M_ν-estimator has a limiting distribution under P_{λ_0} (see the proof of Theorem 3.2 in Section 5).

Definition 3.3. *Let $\mathcal{M}_\nu$ be the set of all M_ν -estimators. We say that $\bar{\nu}_n$ is* *efficient if $\bar{\nu}_n \in \mathcal{M}_\nu$ with limiting distribution G under P_{λ_0}, such that $E[GG']$ is the smallest covariance matrix of the limiting distributions of M_ν-estimators in $\mathcal{M}_\nu$.*

Theorem 3.2. *Suppose that Assumptions 2.1 and 3.1-3.4 hold. If a sequence of estimators $\hat{\nu}_n$ of ν_0 has the following asymptotic representation:*

$$G_n^*(\hat{\nu}_n - \nu_0) = \dot{\nu}_0 S_n^{-1}(\lambda_0) W_n(\lambda_0) + o_{\lambda_0}(1),$$

then the estimator $\hat{\nu}_n$ belongs to $\mathcal{M}_\nu$ and is efficient.

4 Efficient and Adaptive Estimates

In order to construct the efficient estimator, we need to assume that a G_n- or 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 LeCam (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 $\{\bar{\nu}_n\}$ 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 $P_{\lambda_0}(||G_n^*(\bar{\nu}_n - \nu_0)|| < \Delta) > 1 - \varepsilon$ uniformly in n and, for each n, $\bar{\nu}_n$ takes on at most K different values in $\Theta_n^* = \{\nu \in R^{k_1} : ||G_n^*(\nu - \nu_0)|| \leq \Delta\}$.*

Lemma 4.1. *Assume $\Gamma_n(\nu), n = 1, 2, \cdots,$ to be a sequence of random variables which depends on $\nu \in \Theta^*$, an open subset in R^{k_1}. If, for each sequence $\{\nu_n\} \in \Theta^*$ satisfying $G_n^*(\nu_n - \nu_0)$ is bounded by a constant $\Delta > 0$, $\Gamma_n(\nu_n) = o_{\lambda_0}(1)$, then $\Gamma_n(\bar{\nu}_n) = o_{\lambda_0}(1)$ for discretized G_n^*-consistent estimators $\bar{\nu}_n$.*

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 with $\nu(\lambda) = \lambda$.

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. As in the discussion in Section 3, only the parameters γ_0, m_0 and δ_0 are adaptively estimable. 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:

$$W_{\gamma n}(\lambda) = \frac{1}{n} \sum_{t=1}^{n} \Big[\frac{1}{\sqrt{h_t(\lambda)}} \frac{\partial \varepsilon_t(\lambda)}{\partial \gamma} \xi_1(\eta_t(\lambda)) - \frac{1}{2h_t(\lambda)} \frac{\partial h_t(\lambda)}{\partial \gamma} \xi_2(\eta_t(\lambda)) \Big],$$

$$W_{mn}(\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Big[\frac{1}{\sqrt{h_t(\lambda)}} \frac{\partial \varepsilon_t(\lambda)}{\partial m} \xi_1(\eta_t(\lambda)) - \frac{1}{2h_t(\lambda)} \frac{\partial h_t(\lambda)}{\partial m} \xi_2(\eta_t(\lambda)) \Big],$$

$$S_{\gamma n}(\lambda) = \frac{1}{n^2} \sum_{t=1}^{n} \Big[\frac{1}{h_t(\lambda)} \Big(\frac{\partial \varepsilon_t(\lambda)}{\partial \gamma} \Big)^2 I_1(f) + \frac{1}{4h_t^2(\lambda)} \Big(\frac{\partial h_t(\lambda)}{\partial \gamma} \Big)^2 I_2(f) \Big],$$

$$S_{mn}(\lambda) = \frac{1}{n} \sum_{t=1}^{n} \Big[\frac{1}{h_t(\lambda)} \frac{\partial \varepsilon_t(\lambda)}{\partial m} \frac{\partial \varepsilon_t(\lambda)}{\partial m'} I_1(f) + \frac{1}{4h_t^2(\lambda)} \frac{\partial h_t(\lambda)}{\partial m} \frac{\partial h_t(\lambda)}{\partial m'} I_2(f) \Big],$$

$$W_{1n}(\lambda) = \begin{pmatrix} W_{\gamma n}(\lambda) \\ W_{mn}(\lambda) \\ W_{\delta n}(\lambda) \end{pmatrix}, \quad S_{1n}(\lambda) = \begin{pmatrix} S_{\gamma n}(\lambda) & 0 & 0 \\ 0 & S_{mn}(\lambda) & 0 \\ 0 & 0 & S_{\delta n}(\lambda) \end{pmatrix},$$

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

Now we construct adaptive estimators for λ_0. As in Kreiss (1987a), we use the usual kernel density estimator for $\xi_1(x)$. First, define

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

where $j = 1, \cdots, n$, $g(x, a) = (2\pi a^2)^{-1/2} exp(-x^2/2a^2)$, $x \in R$,

$$(4.2) \qquad \hat{\xi}_{1n,j}(x, \lambda) = \begin{cases} \frac{\hat{f}'_{a_n,j}(x,\lambda)}{\hat{f}_{a_n,j}(x,\lambda)} & if \begin{cases} \hat{f}_{a_n,j}(x, \lambda) \geq d_n, \\ |x| \leq g_n, \\ |\hat{f}'_{a_n,j}(x, \lambda)| \leq c_n \hat{f}_{a_n,j}(x, \lambda), \end{cases} \\ 0 & otherwise, \end{cases}$$

and $\hat{\xi}_{2n,j}(x, \lambda) = x\hat{\xi}_{1n,j}(x, \lambda) + 1$, with a_n, c_n, d_n and g_n satisfying:

Assumption 4.1. a_n, $d_n \to 0$; c_n, $g_n \to \infty$; $a_n c_n \to 0$; $n^{-1} a_n^{-3} c_n^2 g_n^3 \to 0$; and $n^{-1} g_n^4 = O(1)$.

We also define $\hat{I}_{1n}(\lambda)$ and $\hat{I}_{2n}(\lambda)$, where

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

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 diag$[\hat{S}_{\gamma n}(\lambda), \hat{S}_{mn}(\lambda)]$ by 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 $k_n/n \to \tau \in (0, 1)$. Split the residual $\eta_1(\lambda)$, $\cdots$, $\eta_n(\lambda)$ into two parts, namely $(\eta_1(\lambda), \cdots, \eta_{k_n}(\lambda))$ and $(\eta_{k_n+1}(\lambda), \cdots, \eta_n(\lambda))$. Denote

$$\begin{aligned} \hat{f}^{(1)}_{a,j}(x, \lambda) &= \frac{1}{2a(k_n - 1)} \sum_{i=1, i\neq j}^{k_n} \Big[K\Big(\frac{x + \eta_i(\lambda)}{a}\Big) + K\Big(\frac{x - \eta_i(\lambda)}{a}\Big)\Big], \\ \hat{f}^{(2)}_{a,j}(x, \lambda) &= \frac{1}{2a(n - k_n - 1)} \sum_{i=k_n+1, i\neq j}^{n} \Big[K\Big(\frac{x + \eta_i(\lambda)}{a}\Big) + K\Big(\frac{x - \eta_i(\lambda)}{a}\Big)\Big], \end{aligned}$$

where $K(x) = e^{-x}/(1 + e^{-x})^2$ is the logistic kernel. Define $\hat{\xi}^{(i)}_{2n,j}(x, \lambda) = \hat{f}^{(i)\prime}_{a_{1n},j}(x, \lambda)/ [b_n + \hat{f}^{(i)}_{a_{1n},j}(x, \lambda)]$, where $i = 1, 2$ and $n^{-1} a_{1n}^{-3} b_n^{-1} = o(1)$. $W_{\delta n}(\lambda)$ is estimated by

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

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, 3.1-3.4 and 4.1 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* $G_n = diag(n, \sqrt{n}I_{k-2})$, $\hat{W}_{1n}(\lambda) = [\hat{W}_{\gamma n}(\lambda), \hat{W}'_{mn}(\lambda), \hat{W}'_{\delta n}(\lambda)]'$ *and* $\hat{S}_{1n}(\lambda) = diag[\hat{S}_{\gamma n}(\lambda), \hat{S}_{mn}(\lambda), \hat{S}_{\delta n}(\lambda)]$.

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 (1997), 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 5-6) 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.

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:

$$y_t = \gamma_0 y_{t-1} + \varepsilon_t, \ \varepsilon_t = \eta_t\sqrt{h_t}, \ h_t = \alpha_{00}(1 + \alpha_0\varepsilon_t^2 + \beta_0 h_{t-1}), \tag{4.3}$$

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 $\alpha_{00} = 1$. This density has been frequently used for investigating the finite-sample behaviours 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 set $c_n = 5.0$, $d_n = e^{-225/3}/6\pi$, and $g_n = 15$. 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 Klaasen (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 generally smaller than those of the LSE and QMLE, except when $\gamma_0 = 0.5$.

TABLE 1

The Empirical Bias and Standard Deviation of LSE, QMLE, AE and EE
n=250, 1000 Replications, and the smoothing parameter $a_n = 0.35$

	$\gamma_0 = -0.5$		γ_0 =0.5		γ_0 =0.8		γ_0 =1.0	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD
LSE	.0088	.0640	−.0042	.0625	−.0079	.0444	.0075	.0151
QMLE	.0043	.0403	−.0004	.0408	−.0036	.0289	.0046	.0100
AE	−.0014	.0180	.0022	.0182	.0009	.0118	.0001	.0028
EE	−.0005	.0175	.0018	.0184	.0001	.0117	−.0009	.0035

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 $n(\hat{\gamma}_n - 1) \longrightarrow_{\mathcal{L}} \int_0^1 w_1(\tau)dw_2(\tau)/\kappa$ $\Omega_\gamma \int_0^1 w_1^2(\tau)d\tau$. Let

$$B_1(\tau) = \frac{1}{\sigma_\varepsilon} w_1(\tau) \text{ and } B_2(\tau) = -\frac{1}{\sigma_\varepsilon^2}\sqrt{\frac{\sigma^{2\varepsilon}}{\sigma_\varepsilon^2\Omega_\gamma - 1}}w_1(\tau) + \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2\Omega_\gamma - 1}}w_2(\tau),$$

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

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

The second term in (4.4) can be simplified to $[\sqrt{\sigma_\varepsilon^2\Omega_\gamma - 1}/(\sigma_\varepsilon^2\Omega_\gamma\kappa)]\,(\int_0^1 B_1^2(\tau)\,d\tau)^{-1/2}\xi$, 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

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

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(1999)). 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, c_n, d_n and g_n, we investigate the size for $\gamma_0 = 1.0$, and local powers for γ_0=0.95, 0.97, 0.98 and 0.99. The critical values of the DF-test come from Table

TABLE 2

The Power and Size of Lower Tail Unit Root Tests for AR(1)-GARCH(1,1) Models 1000 replications, and the smoothing parameter $a_n = 0.35$

	Significance Level 5%					Significance Level 10%				
$\gamma_0 =$	.950	.970	.980	.990	1.000	.950	.970	.980	.990	1.000
DF-test	.904	.549	.319	.166	.055	.974	.775	.534	.285	.120
$\hat{\tau}_{QEn}$	.997	.894	.643	.219	.038	1.000	.973	.839	.437	.078
$\hat{\tau}_{AEn}$	1.000	.999	.993	.860	.040	1.000	.999	.997	.923	.092

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.

5 Proofs of Theorems 3.1-3.2

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 from Bai (1993) and will be used to evaluate the coefficients in various infinite expansions. Lemma 5.2 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 5.3 gives the expansion of $(w_t, h_t)(\lambda)$. These three lemmas are used often in this section and in Section 6. Lemma 5.4 is a basic result for verifying Assumption 2.2.

Lemma 5.1. *If all the roots of $\Psi_\varpi(z) = 1 + \varpi_1 z + \cdots + \varpi_q z^q = 0$ lie outside the unit circle, then there exists a neighbourhood V_ϖ of ϖ, $M > 0$, $C > 0$ and $0 < \varrho < 1$, such that: (a) for every $u \in V_\varpi$, $\psi_i(u) \leq M\varrho^i$; (b) for any $\delta > 0$, if $u, u' \in V$, and $||u - u'|| \leq \delta$, then $|\psi_i(u) - \psi_i(u')| \leq \delta C i \varrho^{i-1}$, where $i = 0, 1, \cdots$, and $\Psi_\varpi^{-1}(z) = \sum_{i=0}^\infty \psi_i(\varpi) z^i$.*

Lemma 5.2. *Under Assumptions 3.1-3.2, the process (w_t^0, h_t^0) is strictly stationary and ergodic, and almost surely has the following causal expansions:*

$$(a)\ w_t^0 = \sum_{i=0}^{\infty} \upsilon_{\phi_0\psi_0}(i)\varepsilon_{t-i}^0 \qquad \text{and} \qquad (b)\ h_t^0 = \iota'\zeta_t + \iota' \sum_{i=1}^{\infty} \prod_{i=0}^{i-1} A_{t-i}\zeta_{t-i},$$

where $\varepsilon_t^0 = \eta_t \sqrt{h_t^0}$, $A_t = A_t(\lambda_0)$, $\iota = (0, \cdots, 0, 1, 0, \cdots, 0)'_{(r+s)\times 1}$ with the $(r+1)$th element being 1, $\zeta_t = \zeta_t(\lambda_0)$ and $\zeta_t(\lambda) = (\alpha_0\eta_t^2(\lambda), 0, \cdots, 0, \alpha_0, 0, \cdots, 0)'$, with the first and $(r+1)$th 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 5.3. *If Assumptions 3.1-3.2 hold, then under P_λ, $(w_t, h_t)(\lambda)$ has the following expansions:*

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

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

where $\varepsilon_t(\lambda) = \eta_t(\lambda)\sqrt{h_t(\lambda)}$, ι and $\zeta_t(\lambda)$ are defined as in Lemma 5.2, and $\tilde{\varepsilon}_0 = (0, \cdots, 0, \omega_0, \cdots, \omega_0)'$ with the last s elements being ω_0. Furthermore, if Assumption 3.3 holds, then $E(h_t(\lambda_0) - h_t^0)^2 = O(\varrho^t)$, $E\left(\varepsilon_t(\lambda_0) - \varepsilon_t^0\right)^2 = O(\varrho^t)$ and $E(w_t(\lambda_0) - w_t^0)^2 = O(\varrho^t)$, where $O(\cdot)$ holds uniformly in all t, $t \geq 1$, and $0 < \varrho < 1$.

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

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

where $\tilde{\varepsilon}_t(\lambda) = [\varepsilon_t^2(\lambda), \cdots, \varepsilon_{t-r+1}^2(\lambda), h_t(\lambda), \cdots, h_{t-s+1}(\lambda)]'$. 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:

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

By (5.2) and expansion (a) of this lemma, the other cases can be proved. This completes the proof. 2.

Lemma 5.4. *If Assumptions 3.1-3.3 hold, then it follows that:*

(a) $\max_{1\leq t\leq n} |n^{-1/2}y_t| = O_{\lambda_0}(1)$ *and (b)* $n^{-1/2}\max_{1\leq t\leq n} w_t^2(\lambda_n) = o_{\lambda_0}(1)$.

Proof. By Lemma 5.3 (a), under P_{λ_0},

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

where $R_{1n} = \sum_{i=0}^{[n\tau]} \upsilon_{\phi_0\psi_0}(i)(\sum_{j=[n\tau]-i+1}^{[n\tau]} \varepsilon_j(\lambda_0))$ and $R_{2n} = (\sum_{i=[n\tau]+1}^{\infty} \upsilon_{\phi_0\psi_0}(i))(\sum_{j=1}^{[n\tau]} \varepsilon_j(\lambda_0))$. By Lemma 5.3 (a), we have $n^{-1/2}\max_{1\le t\le n}|\varepsilon_t(\lambda_0)| = n^{-1/2}\max_{1\le t\le n} |\varepsilon_t^0| + o_{\lambda_0}(1)$. Since ε_t^0 is strictly stationary with a finite variance, $n^{-1/2}\max_{1\le t\le n}|\varepsilon_t^0| = o_{\lambda_0}(1)$ (see Chung (1968, p.93)). Thus, by Lemma 5.1, it is easy to show that $n^{-1/2}\max_{0\le\tau\le 1}|R_{1n}(\tau)| = o_{\lambda_0}(1)$ and $n^{-1/2}\max_{0\le\tau\le 1}|R_{2n}(\tau)| = o_{\lambda_0}(1)$. Furthermore, by Lemma 5.5 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, $n^{-1/2}\max_{1\le t\le n}(\theta_{1n}y_{t-1}/n)^2 = o_{\lambda_0}(1)$. By Lemma 5.3 (a), we have $n^{-1/2}\max_{1\le t\le n} w_t^2(\lambda_0) = n^{-1/2}\max_{1\le t\le n} w_t^{02} + o_{\lambda_0}(1)$. Since w_t^{02} is strictly stationary with a finite variance (see Lemma 5.2), $n^{-1/2}\max_{1\le t\le n} w_t^{02} = o_{\lambda_0}(1)$ (see Chung (1968, p.93)). Thus, (b) holds. This completes the proof. 2

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 Lemmas 5.1 and 5.4, and some basic inequalities. Since the techniques are similar, only the proof of (2.2) is presented. We need to prove that

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

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

where, from Assumptions 3.1-3.2 and Lemma 5.1,

$$\frac{\partial\varepsilon_t(\lambda)}{\partial\gamma} = -\sum_{i=0}^{t-1}\upsilon_\psi(i)y_{t-i-1}, \quad \frac{\partial\varepsilon_t(\lambda)}{\partial m} = -\sum_{i=0}^{t-1}\upsilon_\psi(i)\tilde{w}_{t-i-1}(\lambda),$$

$$\frac{\partial h_t(\lambda)}{\partial\tilde{m}} = 2\sum_{i=1}^{t-1}\upsilon_{\alpha\beta}(i)\varepsilon_{t-i}(\lambda)\frac{\partial\varepsilon_{t-i}(\lambda)}{\partial\tilde{m}}, \quad \frac{\partial h_t(\lambda)}{\partial\tilde{\delta}} = \sum_{i=0}^{t-1}\upsilon_\beta(i)\begin{pmatrix} u_{0t-i}(\lambda) \\ \alpha_0\tilde{\varepsilon}_{t-i-1}(\lambda)\end{pmatrix},$$

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

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

$$(5.5)\qquad \Big[\Big\|G_n\frac{\partial^2 h_t(\lambda^*{}_n)}{\partial\lambda\partial\lambda'}G_n\Big\|^2+\Big\|G_n\frac{\partial h_t(\lambda^*{}_n)}{\partial\lambda}\Big\|^4\Big]O(1)\equiv(I_{1t}+I_{2t})O(1),$$

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 Lemma 5.1, it can be obtained directly that

$$(5.6)\qquad \Big|\frac{\partial\varepsilon_t(\lambda^*{}_n)}{\partial\gamma}\Big|^2\le O(1)\max_{1\le t\le n}y_t^2,\quad \Big\|\frac{\partial\varepsilon_t(\lambda^*{}_n)}{\partial m}\Big\|^2\le O(1)\max_{1\le t\le n}w_t^2(\lambda^*{}_n),$$

$$(5.7)\qquad \Big\|\frac{\partial h_t(\lambda^*{}_n)}{\partial\gamma}\Big\|^2\le O(1)[\max_{1\le t\le n}y_t^2+\max_{1\le t\le n}w_t^2(\lambda^*{}_n)],$$

$$(5.8)\qquad \Big\|\frac{\partial h_t(\lambda^*{}_n)}{\partial m}\Big\|^2\le O(1)\max_{1\le t\le n}w_t^2(\lambda^*{}_n),\quad \Big\|\frac{\partial h_t(\lambda^*{}_n)}{\partial\tilde{\delta}}\Big\|^2\le O(1)\max_{1\le t\le n}w_t^2(\lambda^*{}_n),$$

where $O(1)$ holds uniformly in all t. Denote $D_n=[(n^{-1}\max_{1\le t\le n}y_t^2)+\max_{1\le t\le n} w_t^2(\lambda^*{}_n)]/\sqrt{n}$. By (5.6)-(5.8) and Lemma 5.4, we can show that

$$(5.9)\qquad \sum_{t=1}^n I_{1t}\le D_nO(1)=o_{\lambda_0}(1)\ \text{ and }\ \sum_{t=1}^n I_{2t}\le D_n^2O(1)=o_{\lambda_0}(1).$$

By (5.5) and (5.9), we can show that (5.4) holds. This completes the proof.2

Denote $\tilde{X}_t(\lambda)=[u'_{\gamma t}(\lambda),u'_{mt}(\lambda),u'_{\tilde{\delta}t}(\lambda)]'$ with $u_{\gamma t}(\lambda)=-[h_t^{-1/2}(\lambda),h_t^{-1}(\lambda)\sum_{i=1}^{t-1} \upsilon_{\alpha\beta}(i)\varepsilon_{t-i}(\lambda)]$, $u_{mt}(\lambda)=[h_t^{-1/2}(\lambda)\partial\varepsilon_t(\lambda)/\partial m,(2h_t(\lambda))^{-1}\partial h_t(\lambda)/\partial m]$, and $u_{\tilde{\delta}t}(\lambda)= [0,(2h_t(\lambda))^{-1}\partial h_t(\lambda)/\partial\tilde{\delta}]$. The following is an invariance principle for Theorem 3.1(b).

Lemma 5.5. *Suppose the assumptions of Theorem 3.1 hold. Then,*

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

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

Proof. By Lemma 5.3, it follows that

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

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

$$(5.10)\ \frac{1}{\sqrt{n}}\sum_{t=1}^{[n\tau]}\left[\varepsilon_t,(\tilde{X}_t\xi(\eta_t))',\zeta'(\eta_t)\right]'(\lambda_0)=\frac{1}{\sqrt{n}}\sum_{t=1}^{[n\tau]}\left[\varepsilon_t^0,(X_t^0\xi(\eta_t))',\zeta'(\eta_t)\right]'+o_{\lambda_0}(1),$$

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 $E(W_t^{*0}W_t^{*0\prime})$ $= \tilde{\Omega}$. 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 $(w_1, w_2, N'_{m\tilde{\delta}}, N'_\zeta)'(\tau)$ in $D^{k+l+1}[0,1]$. Furthermore, by (5.10), we complete the proof. 2

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

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{n}\frac{\partial g_{t-1}(\lambda_0)}{\partial\gamma}\xi'(\eta_t) &= \frac{1}{n}\sum_{t=1}^{n} y_{t-1}u_{\gamma t}^0\xi'(\eta_t) + o_{\lambda_0}(1) \\ &\longrightarrow_{\mathcal{L}} \kappa\int_0^1 w_1(\tau)dw_2(\tau), \end{aligned} \tag{5.11}$$

where the last step holds by Theorem 2.2 in Kurtz and Protter (1991) and Lemma 5.5. Similarly, we have

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

Denote $\sigma_{\gamma t} = u_{\gamma t}^0 J u_{\gamma t}^{0\prime}$. Using a similar technique as in Theorem 3.4 in Ling and Li (1998), we can show that, under P_{λ_0},

$$\frac{1}{n}\sum_{t=1}^{n}|\sigma_{\gamma t} - E\sigma_{\gamma t}| = O_{\lambda_0}(1),\ \frac{1}{\sqrt{n}}\max_{1\le t\le n}|\sigma_{\gamma t} - E\sigma_{\gamma t}| = o_{\lambda_0}(1), \tag{5.12}$$

and

$$\frac{1}{n}E[\sum_{t=1}^{n}(\sigma_{\gamma t} - E\sigma_{\gamma t})]^2 \to \sigma_0^2,\ \frac{1}{\sqrt{n}}\sum_{t=1}^{[n\tau]}(\sigma_{\gamma t} - E\sigma_{\gamma t}) \longrightarrow_{\mathcal{L}} \sigma_0\omega_0(\tau) \text{ in } D, \tag{5.13}$$

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), (5.12)-(5.13), Lemma 5.5 and the continuity theorem, it follows that

$$\begin{aligned} \frac{1}{n^2}\sum_{t=1}^{n}\frac{\partial g_{t-1}(\lambda_0)}{\partial\gamma}J\frac{\partial g'_{t-1}(\lambda_0)}{\partial\gamma} &= \frac{E\sigma_{\gamma t}}{n^2}\sum_{t=1}^{n} y_{t-1}^2 + \frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2[u_{\gamma t}^0 J u_{\gamma t}^{0\prime} - E\sigma_{\gamma t}] + o_{\lambda_0}(1) \\ &= \frac{E\sigma_{\gamma t}}{n^2}\sum_{t=1}^{n} y_{t-1}^2 + o_{\lambda_0}(1) \longrightarrow_{\mathcal{L}} \kappa^2 E\sigma_{\gamma t}\int_0^1 w_1^2(\tau)d\tau. \end{aligned} \tag{5.14}$$

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

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{\partial g_{t-1}(\lambda_0)}{\partial\gamma}J\frac{\partial g_{t-1}(\lambda_0)}{\partial m'}\longrightarrow_{\mathcal{L}} E(u^0_{\gamma t}Ju^{0'}_{mt})\kappa\int_0^1 w_1(\tau)d\tau, \tag{5.15}$$

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{\partial g_{t-1}(\lambda_0)}{\partial\gamma}J\frac{\partial g_{t-1}(\lambda)}{\partial\tilde{\delta}'}\longrightarrow_{\mathcal{L}} E(u^0_{\gamma t}Ju^{0'}_{\tilde{\delta}t})\kappa\int_0^1 w_1(\tau)d\tau, \tag{5.16}$$

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{\partial g_{t-1}(\lambda_0)}{\partial\gamma}V'_{\xi\zeta}\longrightarrow_{\mathcal{L}} Eu^0_{\gamma t}V'_{\xi\zeta}\kappa\int_0^1 w_1(\tau)d\tau. \tag{5.17}$$

Denote $u_{m\tilde{\delta}} = \begin{pmatrix} u_{mt} \\ u_{\tilde{\delta}t}\end{pmatrix}(\lambda_0)$ and $u^0_{m\tilde{\delta}} = \begin{pmatrix} u^0_{mt} \\ u^0_{\tilde{\delta}t}\end{pmatrix}$. By Lemma 5.3 and the ergodic theorem, we can show that:

$$\frac{1}{n}\sum_{t=1}^{n}(u_{m\tilde{\delta}}Ju'_{m\tilde{\delta}}) = \frac{1}{n}\sum_{t=1}^{n}(u^0_{m\tilde{\delta}}Ju^{0'}_{m\tilde{\delta}}) + o_{\lambda_0}(1) = E(u^0_{m\tilde{\delta}}Ju^{0'}_{m\tilde{\delta}}) + o_{\lambda_0}(1), \tag{5.18}$$

$$\frac{1}{n}\sum_{t=1}^{n}u_{m\tilde{\delta}}V'_{\xi\zeta} = \frac{1}{n}\sum_{t=1}^{n}u^0_{m\tilde{\delta}}V'_{\xi\zeta} + o_{\lambda_0}(1) = Eu^0_{m\tilde{\delta}}V'_{\xi\zeta} + o_{\lambda_0}(1). \tag{5.19}$$

By Theorem 2.2 in Kurtz and Protter (1991) and Lemma 5.5, all the limiting distributions involved in $\tilde{W}_n(\lambda_0)$ and $\tilde{S}_n(\lambda_0)$ are jointly convergent. Finally, by Lemma 5.5, (5.11), (5.14)-(5.19), 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. 2

Proof of Theorem 3.2. It is obvious that the estimator $\hat{\nu}_n$ belongs to $\mathcal{M}_\nu$. Let $\bar{\nu}_{\pi n}$ be any M$_\nu$-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 5.5, 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 $\tau\begin{pmatrix} Eh^0_t & 1 \\ 1 & \Omega^*_{\pi\gamma}\end{pmatrix}$, and $\Omega^*_{\pi\gamma} = E(u^0_{\gamma t}J_\pi u^{0'}_{\gamma t})$ with $J_\pi = \mathrm{diag}(I_{\pi_1}, I_{\pi_2})$. Denote $\Omega^*_{\pi m} = E(u^0_{mt}J_\pi u^{0'}_{mt})$, $\Omega^*_{\pi\tilde{\delta}} = E(u^0_{\tilde{\delta}t}J_\pi u^{0'}_{\tilde{\delta}t})$, $\Omega_{\pi\gamma} = E(u^0_{\gamma t}J_c u^{0'}_{\gamma t})$, $\Omega_{\pi m} = E(u^0_{mt}J_c u^{0'}_{mt})$, and $\Omega_{\pi\tilde{\delta}} = E(u^0_{\tilde{\delta}t}J_c u^{0'}_{\tilde{\delta}t})$. 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 $(W_{Mn}, S_{Mn})(\lambda_0)$ converges weakly to (W_M, S_M) under P_{λ_0}, where

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

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, $G^*_n(\bar{\nu}_{\pi n} - \nu_0)$ converges weakly to $G_\pi = \dot{\nu}_0 S_M^{-1} W_M$ under P_{λ_0}, and

$$E(G_\pi G'_\pi) = \dot{\nu}_0 \, \mathrm{diag}(\Sigma_{\pi\gamma}, \; \Sigma_{\pi m}, \; \Sigma_{\pi\tilde{\delta}}) \, \dot{\nu}'_0,$$

where $\Sigma_{\pi\gamma} = \kappa^{-2}\Omega_{\pi\gamma}^{-2} E\left[\int_0^1 w_1(\tau)dw_{\pi 2}(\tau)/\int_0^1 w_1^2(\tau)d\tau\right]^2$, $\Sigma_{\pi m} = (c_1 P + c_2 Q)^{-1}(P I_{\pi_1} + Q I_{\pi_2})(c_1 P + c_2 Q)^{-1}$ and $\Sigma_{\pi\tilde{\delta}} = I_{\pi_2} c_2^{-2} R^{-1}$, with

$$P = E\left(\frac{1}{h_t^0}\frac{\partial \varepsilon_t^0}{\partial m}\frac{\partial \varepsilon_t^0}{\partial m'}\right), \; Q = \frac{1}{4}E\left(\frac{1}{h_t^0}\frac{\partial h_t^0}{\partial m}\frac{\partial h_t^0}{\partial m'}\right) \text{ and } R = \frac{1}{4}E\left(\frac{1}{h_t^{02}}\frac{\partial h_t^0}{\partial \tilde{\delta}}\frac{\partial h_t^0}{\partial \tilde{\delta}'}\right).$$

Denote G as the limiting distribution of $\dot{\nu}_0 S_n^{-1}(\lambda_0) W_n(\lambda_0)$. From Theorem 3.1(b) and Remark 3.3, we obtain

$$E(GG') = \dot{\nu}_0 \, \mathrm{diag}(\Sigma_\gamma, \; \Sigma_m, \; \Sigma_{\tilde{\delta}}) \, \dot{\nu}'_0,$$

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 $I_1(f) I_{\pi_1} \geq c_1^2$ and $I_2(f) I_{\pi_2} \geq c_2^2$. It is obvious that $\Sigma_{\tilde{\delta}} \leq \Sigma_{\pi\tilde{\delta}}$. After some algebra, we have

$$\begin{aligned}
\Sigma_m - \Sigma_{\pi m} &= (c_1 P + c_2 Q)^{-1}[P(P I_{\pi_1} + Q I_{\pi_2})^{-1} P (c_1 P + c_2 Q)^{-1}(c_1^2 - I_1(f) I_{\pi_1}) \\
&\quad + Q(P I_{\pi_1} + Q I_{\pi_2})^{-1} Q (c_1 P + c_2 Q)^{-1}(c_2^2 - I_2(f) I_{\pi_2}) \\
&\quad + P(P I_{\pi_1} + Q I_{\pi_2})^{-1} Q (c_1 P + c_2 Q)^{-1}(c_1 c_2 - I_1(f) I_{\pi_2}) \\
&\quad + Q(P I_{\pi_1} + Q I_{\pi_2})^{-1} P (c_1 P + c_2 Q)^{-1}(c_1 c_2 - I_2(f) I_{\pi_1})] \\
&\leq (c_1 P + c_2 Q)^{-1}[P(P I_{\pi_1} + Q I_{\pi_2})^{-1} Q (c_1 P + c_2 Q)^{-1}(c_1 c_2 - I_1(f) I_{\pi_2}) \\
&\quad + Q(P I_{\pi_1} + Q I_{\pi_2})^{-1} P (c_1 P + c_2 Q)^{-1}(c_1 c_2 - I_2(f) I_{\pi_1})] \\
&\leq (c_1 P + c_2 Q)^{-1}(Q^{-1} I_{\pi_1} + P^{-1} I_{\pi_2})^{-1}(c_1 P + c_2 Q)^{-1} \\
&\quad \cdot \left[2 c_1 c_2 - 2(I_2(f) I_{\pi_1} I_1(f) I_{\pi_2})^{1/2}\right] \leq 0. \qquad (5.20)
\end{aligned}$$

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 5.5, we can show that, in $D^2[0,1]$, $n^{-1/2}\sum_{t=1}^{[n\tau]}(\varepsilon_t^{**},\varepsilon_{\pi t}^{**})(\lambda_0)$ converges to the bivariate Brownian motion $(\omega_2^*,\omega_{\pi 2}^*)(\tau)$, which has mean zero and covariance

$$\tau\begin{pmatrix}\Omega_\gamma^{-1} & \Omega_\gamma^{-1}\\ \Omega_\gamma^{-1} & \Omega_{\pi\gamma}^*\Omega_{\pi\gamma}^{-2}\end{pmatrix}. \tag{5.21}$$

Denote $W_{\gamma n}$ and $W_{\pi\gamma n}$ as the first elements of $W_n(\lambda_0)$ and $W_{Mn}(\lambda_0)$, respectively, and $S_{\gamma n}$ and $S_{\pi\gamma n}$ as the $(1,1)th$ elements of $S_n(\lambda_0)$ and $S_{Mn}(\lambda_0)$, 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 $\left[\kappa\int_0^1\omega_1^2(\tau)d\tau\right]^{-1}\int_0^1\omega_1(\tau)d\omega_2^*(\tau)$ and $\left[\kappa\int_0^1\omega_1^2(\tau)d\tau\right]^{-1}\int_0^1\omega_1(\tau)d\omega_{\pi 2}^*(\tau)$, respectively, which are denoted by G_γ and $G_{\pi\gamma}$, respectively. Let $\Delta=\Omega_{\pi\gamma}^*\Omega_{\pi\gamma}^{-2}-\Omega_\gamma^{-1}$. As in the proof of (5.20), we can show that $\Delta\geq 0$. Using (5.21) and Lemma 3.1 of Phillips (1989), we can show that the distribution of $G_{\pi\gamma}$ is the same as that of $G_\gamma+\kappa^{-1}\Delta^{1/2}\left[\int_0^1\omega_1^2(\tau)d\tau\right]^{-1/2}\Phi$, 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. 2

6 Proof of Theorem 4.2

From Bickel (1982), Kreiss (1987a) and Linton (1993), we know that $\hat{I}_{in}(\lambda)$ is a consistent estimator of $I_i(f)$, where $i=1,2$. Furthermore, by Lemma 4.1, it is sufficient to prove the following theorem for Theorem 4.2.

Theorem 6.1. *Let λ_n be defined as ν_n in Lemma 4.1. Then, under Assumptions 2.1, 3.1-3.4 and 4.1, $\hat{W}_{1n}(\lambda_n)-W_{1n}(\lambda_n)=o_{\lambda_0}(1)$.*

Proof. 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 this theorem under P_{λ_n}. 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,

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

$$+\frac{1}{2}E_{\lambda_n}\Big\{\frac{1}{n}\sum_{t=1}^{n}\Big[\frac{1}{h_t(\lambda_n)}\frac{\partial h_t(\lambda_n)}{\partial\gamma}(\hat{\xi}_{2t}-\xi_2(\eta_t(\lambda_n)))\Big]\Big\}^2$$

$$= 2B_{1n}+\frac{1}{2}B_{2n},\ \ say. \tag{6.1}$$

Note that $\hat{\xi}_{1t}$ and $\xi_1(\eta_t(\lambda_n))$ are odd functions of $\eta_t(\lambda_n)$. As in Kreiss (1987a) and Bickel (1982), we have

$$B_{1n}=\frac{1}{n^2}\sum_{t=1}^{n}E_{\lambda_n}\Big\{\frac{1}{h_t(\lambda_n)}\left[\hat{\xi}_{1t}-\xi_1(\eta_t(\lambda_n))\right]^2\left|\frac{\partial\varepsilon_t(\lambda_n)}{\partial\gamma}\right|^2\Big\}. \tag{6.2}$$

Since $h_t^{-1}(\lambda_n)$ is bounded and $\partial\varepsilon_t(\lambda_n)/\partial\gamma=\sum_{j=0}^{t-1}\upsilon_{\psi_n}(j)y_{t-j-1}$, by Lemma 5.1, we can show that

$$B_{1n}\le O(1)\frac{1}{n^2}\sum_{t=1}^{n}\max_{1\le j\le n}E_{\lambda_n}\left\{y_j^2\left[\hat{\xi}_{1t}-\xi_1(\eta_t(\lambda_n))\right]^2\right\}. \tag{6.3}$$

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

$$B_{1n}\le O(1)\frac{1}{n}\sum_{t=1}^{n}\max_{1\le j\le n}E_{\lambda_n}\left\{\varepsilon_j^2(\lambda_n)\left[\hat{\xi}_{1t}-\xi_1(\eta_t(\lambda_n))\right]^2\right\}=o(1), \tag{6.4}$$

where the last equation holds by Proposition 6.1 (a) below.

Now we show that $B_{2n}=o(1)$. Note that $\hat{\xi}_{2t}$ and $\xi_2(\eta_t(\lambda_n))$ are symmetric functions of $\eta_t(\lambda_n)$. Here we have to use the symmetry of f and consider the cross-terms in the expansion of B_{2n}. Denote $\xi_t^*=\eta_t(\lambda_n)\hat{\xi}_{1t}-\eta_t(\lambda_n)\xi_1(\eta_t(\lambda_n))$. Using Assumption 3.4, we can show that

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

Since $\upsilon_{\alpha_n\beta_n}(i) = O(\varrho^i)$ with $\varrho \in (0,1)$ and independent of λ_n, we have

$$|H_{nti}| \le O(1)\varrho^i \max_{1\le j\le n} E_{\lambda_n}\left[\varepsilon_j^2(\lambda_n)y_{j-1}^2|\xi_{t+i}^*\xi_t^*|\right], \tag{6.5}$$

where $O(1)$ holds uniformly in all t. By Lemma 5.1, (6.5), and the inequality, $2|\xi_{t+i}^*\xi_t^*| \le \xi_{t+i}^{*2} + \xi_t^{*2}$, we have

$$\begin{aligned} B_{2n} &= \frac{1}{n^2}\sum_{t=1}^{n} E_{\lambda_n}\Big\{\frac{1}{h_t^2(\lambda_n)}[\frac{\partial h_t(\lambda_n)}{\partial\gamma}]^2\xi_t^{*2}\Big\} + \frac{2}{n^2}\sum_{t=1}^{n-1}\sum_{i=1}^{n-t} H_{nti} \\ &\le \frac{O(1)}{n^2}\Big\{\sum_{t=1}^{n}\max_{1\le j\le n} E_{\lambda_n}[\varepsilon_j^2(\lambda_n)y_{j-1}^2\xi_t^{*2}] + \sum_{t=1}^{n-1}\sum_{i=1}^{n-t}\varrho^i \max_{1\le j\le n} E_{\lambda_n}[\varepsilon_j^2(\lambda_n)y_{j-1}^2\xi_{t+i}^{*2}]\Big\}. \qquad (6.6) \end{aligned}$$

Note that $E_{\lambda_n}\left[\varepsilon_j^2(\lambda_n)\varepsilon_i(\lambda_n)\varepsilon_{i_1}(\lambda_n)\xi_t^{*2}\right] = 0$ for any $i \ne i_1$. In a similar manner to the arguments of (6.4), we can show that

$$\begin{aligned} B_{2n} &\le \frac{O(1)}{n}\Big\{\sum_{t=1}^{n}\max_{1\le j\le n} E_{\lambda_n}[\varepsilon_j^4(\lambda_n)\xi_t^{*2}] + \sum_{t=1}^{n-1}\sum_{i=1}^{n-t}\varrho^i \max_{1\le j\le n} E_{\lambda_n}[\varepsilon_j^4(\lambda_n)\xi_{t+i}^{*2}]\Big\} \\ &= \frac{O(1)}{n}\Big\{\sum_{t=1}^{n}\max_{1\le j\le n} E_{\lambda_n}[\varepsilon_j^4(\lambda_n)\xi_t^{*2}] + \sum_{i=1}^{n-1}\varrho^i \sum_{t=i+1}^{n}\max_{1\le j\le n} E_{\lambda_n}[\varepsilon_j^4(\lambda_n)\xi_t^{*2}]\Big\} \\ &\le \frac{O(1)}{n}\sum_{t=1}^{n}\max_{1\le j\le n} E_{\lambda_n}[\varepsilon_j^4(\lambda_n)\xi_t^{*2}] = o(1), \qquad (6.7) \end{aligned}$$

where the last equation holds by Proposition 6.1 (b) below. By (6.1), (6.4) and (6.7), we can obtain $E_{\lambda_n}\left[\hat{W}_{\gamma n}(\lambda_n) - W_{\gamma n}(\lambda_n)\right]^2 = o(1)$. In a similar manner, we can obtain that $E_{\lambda_n}\left\|\hat{W}_{mn}(\lambda_n) - W_{mn}(\lambda_n)\right\|^2 = o(1)$. To complete the proof, it is sufficient to show that

$$E_{\lambda_n}\left\|\hat{W}_{\delta n}(\lambda_n) - W_{\delta n}(\lambda_n)\right\|^2 = o(1). \tag{6.8}$$

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 (6.8) is similar to that of Theorem 3.1 in DWK (1997) and hence is omitted. This completes the proof. 2

Proposition 6.1. *Under the assumptions of Theorem 6.1,*

$$(a)\quad \frac{1}{n}\sum_{i=1}^{n}\max_{1\le j\le n} E_{\lambda_n}\left[\varepsilon_j^2(\lambda_n)\int\Big\{\left[\hat{\xi}_{1n,i}(x,\lambda_n) - \xi_1(x)\right]^2 f(x)\Big\}dx\right] = o(1),$$

$$(b)\quad \frac{1}{n}\sum_{i=1}^{n}\max_{1\le j\le n} E_{\lambda_n}\left[\varepsilon_j^4(\lambda_n)\int\Big\{\left[\hat{\xi}_{1n,i}(x,\lambda_n) - \xi_1(x)\right]^2 x^2 f(x)\Big\}dx\right] = o(1).$$

In the following, we introduce six lemmas. Lemma 6.1 is a basic result for Lemmas 6.2-6.3, while Lemmas 6.2-6.3 are used in Lemma 6.5. The proof of Proposition 6.1 comes directly from the following Lemmas 6.4-6.6. The routine of the proof is similar to Bickel (1982) and Kreiss (1987a), but the technique is more complicated.

Lemma 6.1. *Denote* $f_{a_n}(x) = E_{\lambda_n}\Big[g(x+\eta_i(\lambda_n), a_n)\Big]$, $g_i^* = g(x+\eta_i(\lambda_n), a_n) - 2f_{a_n}(x) + g(x-\eta_i(\lambda_n), a_n)$, *and* $G(x) = E_{\lambda_n}\left[g_i^{*2}\eta_i^4(\lambda_n)\right] + E_{\lambda_n}\left[g_i^{*2}\eta_i^2(\lambda_n)\right] + E_{\lambda_n}g_i^{*2}$. *Under the assumptions of Theorem 6.1, it follows that:*

(a) when $i+1 \le Q \le j$ *and* $0 \le i_1 \le j-i$,

$$E_{\lambda_n}[g_{i+i_1}^{*2}(\tilde{\iota}' \prod_{t_1=Q}^{j} A_{nt_1}\zeta_{Q-1})(\tilde{\iota}' \prod_{t_1=i+1}^{j} A_{nt_1}\zeta_i)] \le M\varrho^{j-i}G(x);$$

(b) when $i+1 \le Q \le j$ *and* $0 \le i_1 \le j-i$,

$$E_{\lambda_n}[g_{i+i_1}^{*2}(\tilde{\iota}' \prod_{t_1=Q}^{j} A_{nt_1}\zeta_{Q-1})(\tilde{\iota}' \prod_{t_1=i}^{j} A_{nt_1}\tilde{\varepsilon}_{i-1}(\lambda_n))] \le M\varrho^{j-i}G(x),$$

(c) when $0 \le i_1 \le j-i$,

$$E_{\lambda_n}[g_{i+i_1}^{*2}(\tilde{\iota}' \prod_{t_1=i}^{j} A_{nt_1}\tilde{\varepsilon}_{i-1}(\lambda_n))^2] \le M\varrho^{j-i}G(x),$$

where $\varrho \in (0,1)$, ϱ *and* M *are constants and independent of* i, j, x *and* λ_n, $\tilde{\iota}$ *is an* $(r+s)-$*dimensional constant vector,* $A_{ni} = A_i(\lambda_n)$ $\zeta_i = \zeta_i(\lambda_n)$, *and* $\zeta_i(\lambda)$ *and* $\tilde{\varepsilon}_{i-1}(\lambda)$ *are defined, respectively, as in Lemma 5.2 and (5.1) with* $\alpha_0 = 1$.

Proof. First, we illustrate the following facts.

(i) Denote $\varrho_\lambda = \min\left\{\rho\left[E_\lambda(A_i(\lambda)\otimes A_i(\lambda))\right], \rho\left[E_\lambda(A_i(\lambda)\otimes I_{r+s})\right]\right\}$. By Assumption 2, which is equivalent to $\rho\left[E_\lambda(A_i(\lambda))\right] < 1$ (see Ling (1999)), and Assumption 3, we have $\varrho_{\lambda_0} \in (0,1)$. Let ϵ be a positive constant so that $\varrho \equiv \varrho_{\lambda_0} + \epsilon < 1$. Since $E_{\lambda_n}\eta_t^2(\lambda_n) = E\eta_t^2$, $E_{\lambda_n}\eta_t^4(\lambda_n) = E\eta_t^4$, and $\lambda_n - \lambda_0 = O(n^{-1/2})$, there exists an integer N so that $\varrho_{\lambda_n} \le \varrho$ for all $n > N$.

(ii) It is obvious that there is a constant M_1 independent of i, x and λ_n so that

$$\left\|E_{\lambda_n}(g_i^{*2}\zeta_i\zeta_i')\right\| \le M_1G(x), \quad \left\|E_{\lambda_n}(g_i^{*2}\zeta_i)\right\| \le M_1G(x),$$

$$\left\|E_{\lambda_n}\left[g_i^{*2}(A_{ni}\otimes\zeta_i)\right]\right\| \le M_1G(x),$$

$$\left\|E_{\lambda_n}\left[g_i^{*2}(A_{ni}\otimes I_{r+s})\right]\right\| \le M_1G(x) \text{ and } \left\|E_{\lambda_n}\left[g_i^{*2}(A_{ni}\otimes A_{ni})\right]\right\| \le M_1G(x).$$

For (a), we first consider the case with $i_1 > 0$. Denote $D_{ij} = E_{\lambda_n}[g^{*2}_{i+i_1}(\tilde{\imath}' \prod_{t_1=j}^{Q} A_{nt_1}\zeta_{Q-1})(\tilde{\imath}' \prod_{t_1=j}^{i+1} A_{nt_1}\zeta_i)]$. When $i + i_1 + 2 \leq Q \leq j$,

$$\begin{aligned} D_{ij} &= (\tilde{\imath}' \otimes \tilde{\imath}') \left[E_{\lambda_n}(A_{ni} \otimes A_{ni})\right]^{j-Q+1} E_{\lambda_n}(A_{nQ-1} \otimes \zeta_{Q-1}) \\ &\quad \cdot \left[E_{\lambda_n}(A_{ni} \otimes I_{r+s})\right]^{Q-2-(i+i_1+1)+1} E_{\lambda_n}\left[g^{*2}_{i+i_1}(A_{ni+i_1} \otimes I_{r+s})\right] \\ &\quad \cdot \left[E_{\lambda_n}(A_{ni} \otimes I_{r+s})\right]^{i_1-1} E_{\lambda_n}\left[vec(\zeta_i)\right] \leq M\varrho^{j-i}G(x); \end{aligned} \tag{6.9}$$

when $Q = i + i_1 + 1$,

$$\begin{aligned} D_{ij} &= (\tilde{\imath}' \otimes \tilde{\imath}') \left[E_{\lambda_n}(A_{ni} \otimes A_{ni})\right]^{j-(i+i_1+1)+1} E_{\lambda_n}\left[g^{*2}_{i+i_1}(A_{ni+i_1} \otimes \zeta_{i+i_1})\right] \\ &\quad \cdot \left[E_{\lambda_n}(A_{ni} \otimes I_{r+s})\right]^{i_1-1} E_{\lambda_n}\left[vec(\zeta_i)\right] \leq M\varrho^{j-i}G^2(x); \end{aligned} \tag{6.10}$$

when $i + 2 \leq Q \leq i + i_1$,

$$\begin{aligned} D_{ij} &= (\tilde{\imath}' \otimes \tilde{\imath}') \left[E_{\lambda_n}(A_{ni} \otimes A_{ni})\right]^{j-(i+i_1+1)+1} E_{\lambda_n}\left[g^{*2}_{i+i_1}(A_{ni+i_1} \otimes A_{ni+i_1})\right] \\ &\quad \cdot \left[E_{\lambda_n}(A_{ni} \otimes A_{ni})\right]^{(i+i_1-1)-Q+1} E_{\lambda_n}(A_{nQ-1} \otimes \zeta_{nQ-1}) \\ &\quad \cdot \left[E_{\lambda_n}(A_{ni} \otimes I_{r+s})\right]^{Q-2-(i+1)+1} E_{\lambda_n}\left[vec(\zeta_i)\right] \leq M\varrho^{j-i}G(x); \end{aligned} \tag{6.11}$$

when $Q = i + 1$,

$$\begin{aligned} D_{ij} &= (\tilde{\imath}' \otimes \tilde{\imath}') \left[E_{\lambda_n}(A_{ni} \otimes A_{ni})\right]^{j-(i+i_1+1)+1} E_{\lambda_n}\left[g^{*2}_{i+i_1}(A_{ni+i_1} \otimes A_{ni+i_1})\right] \\ &\quad \cdot \left[E_{\lambda_n}(A_{ni} \otimes A_{ni})\right]^{i_1-1} E_{\lambda_n}\left[vec(\zeta_i\zeta_i')\right] \leq M\varrho^{j-i}G(x), \end{aligned} \tag{6.12}$$

where M is some constant and independent of i, j, x and λ_n. By (6.9)-(6.12), we know (a) holds when $i_1 > 0$. Note that, in (6.9)-(6.12), we have used the facts (i) and (ii). Similarly, we can show that (a) holds when $i_1 = 0$. Note that, by Lemmas 5.2-5.3, we can show that $\left\|E_{\lambda_n}\left[\tilde{\varepsilon}_{i-1}(\lambda_n)\tilde{\varepsilon}'_{i-1}(\lambda_n)\right]\right\|$ is bounded uniformly in i. In a similar manner, we can show that (b) and (c) hold. This completes the proof. 2

Lemma 6.2. *Under the assumptions of Theorem 6.1:*

(a) $\max_{1\leq j\leq n} E_{\lambda_n}\left\{f^{-1}_{a_n}(x)\left[\hat{f}_{a_n,t}(x,\lambda_n) - f_{a_n}(x)\right]^2 \varepsilon_j^2(\lambda_n)\right\} \leq \frac{a_n^{-1}}{n}(\kappa_1 + \kappa_2 x^2)$,

(b) $\max_{1\leq j\leq n} E_{\lambda_n}\left\{x^2 f^{-1}_{a_n}(x)\left[\hat{f}_{a_n,t}(x,\lambda_n) - f_{a_n}(x)\right]^2 \varepsilon_j^4(\lambda_n)\right\} \leq \frac{x^2 a_n^{-1}}{n}(\kappa_1 + \frac{\kappa_2 x^4}{n})$, *where $1 \le t \le n$, κ_1 and κ_2 are constants and independent of n and x, and $f_{a_n}(x)$ is defined as in Lemma 6.1.*

Proof. We prove only (b) since the proofs of (a) and (b) are similar.

$$\begin{aligned} & E_{\lambda_n}\left\{x^2 f_{a_n}^{-1}(x)[\hat{f}_{a_n,t}(x) - f_{a_n}(x)]^2 \varepsilon_j^4(\lambda_n)\right\} \\ & \le \frac{1}{2} E_{\lambda_n}\{x^2 f_{a_n}^{-1}(x)[\frac{1}{n-1}\sum_{i=j+1, i\ne t}^{n} g_i^*]^2 \varepsilon_j^4(\lambda_n)\} \\ (6.13) \qquad & + \frac{1}{2} E_{\lambda_n}\{x^2 f_{a_n}^{-1}(x)[\frac{1}{n-1}\sum_{i=1, i\ne t}^{j} g_i^*]^2 \varepsilon_j^4(\lambda_n)\} = \frac{1}{2}(B_1 + B_2), say, \end{aligned}$$

where g_i^* is defined as in Lemma 6.1. By Lemmas 5.2-5.3, we can show that $\max_{1\le j\le n}$ $E_{\lambda_n}\left[\varepsilon_j^4(\lambda_n)\right]$ is bounded. Thus, we have

$$\begin{aligned} B_1 &= O(1) E_{\lambda_n}\{x^2 f_{a_n}^{-1}(x)[\frac{1}{n}\sum_{i=j+1, i\ne t}^{n} g_i^*]^2\} \\ (6.14) \qquad &= O(1)\frac{x^2 f_{a_n}^{-1}(x)}{n^2}\sum_{i=j+1, i\ne t}^{n} E g_i^{*2} = O(1)\frac{x^2 a_n^{-1}}{n}, \end{aligned}$$

where the last equation holds by (6.7) of Bickel (1982) and $O(1)$ holds uniformly in all x. Moreover,

$$B_2 = \frac{x^2 f_{a_n}^{-1}(x)}{(n-1)^2}[\sum_{i=1, i\ne t}^{j} E_{\lambda_n}(g_i^{*2}\varepsilon_j^4(\lambda_n)) + 2\sum_{i=1, i\ne t}^{j-1}\sum_{i_1=1}^{j-i} E_{\lambda_n}(g_{i+i_1}^* g_i^* \varepsilon_j^4(\lambda_n))].$$

Note that $\eta_j^2(\lambda_n)\iota' A_{nj} = (\alpha_1, \cdots, \alpha_r, \beta_1, \cdots, \beta_s)\eta_j^2(\lambda_n) = \tilde{\iota}' A_{nj}$, where $\tilde{\iota}' = (1, 0, \cdots, 0)'_{(r+s)\times 1}$ and A_{nj} is defined as in Lemma 6.1. From Lemma 5.3, we can show that

$$\begin{aligned} \varepsilon_j^2(\lambda_n) &= [\tilde{\iota}'\zeta_j + \tilde{\iota}' A_{nj}\zeta_{j-1} + \cdots + \tilde{\iota}'\prod_{t_1=i+2}^{j} A_{nt_1}\zeta_{i+1}] \\ &\quad + [\tilde{\iota}'\prod_{t_1=i+1}^{j} A_{nt_1}\zeta_i + \tilde{\iota}'\prod_{t_1=i}^{j} A_{nt_1}\tilde{\varepsilon}_{i-1}(\lambda_n)] = R_{j,i+1} + S_j, say, \end{aligned}$$

Since $R_{j,i+1}$ is a function of $\{\eta_j(\lambda_n), \cdots, \eta_{i+1}(\lambda_n)\}$, g_i^* is independent of $R_{j,i+1}$, and hence we have

$$\begin{aligned} E_{\lambda_n}(g_i^{*2}\varepsilon_j^4(\lambda_n)) &= E_{\lambda_n} g_i^{*2}\cdot E_{\lambda_n} R_{j,i+1}^2 + E_{\lambda_n}\left[g_i^{*2}(2R_{j,i+1}S_j + S_j^2)\right] \\ &= O(1)\left[E_{\lambda_n} g_i^{*2} + (j-i)\varrho^{j-i} G(x)\right], \end{aligned}$$

where the last equation holds because $\max_{j,i} E_{\lambda_n} R_{j,i+1} \le \max_j E_{\lambda_n} \varepsilon_j^4(\lambda_n) < \infty$ and Lemma 6.1. Similarly, since $E_{\lambda_n} g_i^* = 0$, by Lemma 6.1, we have,

$$\begin{aligned}
\left|E_{\lambda_n}(g_{i+i_1}^* g_i^* \varepsilon_j^4(\lambda_n))\right| &= \left|E_{\lambda_n}\left[g_{i+i_1}^* g_i^*(2R_{j,i+1}S_j + S_j^2)\right]\right| \\
&\le \frac{1}{2} E_{\lambda_n}\left[(g_{i+i_1}^{*2} + g_i^{*2})(2R_{j,i+1}S_j + S_j^2)\right] \\
&= O(1)\left[(j-i)\varrho^{j-i} G(x)\right],
\end{aligned}$$

where $i_1 > 0$. Thus,

$$\begin{aligned}
B_2 &= O(1)\frac{x^2 f_{a_n}^{-1}(x)}{n^2} \sum_{i=1, i\ne t}^{j} E_{\lambda_n} g_i^{*2} \\
&\quad + O(1)\frac{x^2 f_{a_n}^{-1}(x)}{n^2}\Big[\sum_{i=1,i\ne t}^{j}(j-i)\varrho^{j-i} + \sum_{i=1,i\ne t}^{j-1}\sum_{i_1=1}^{j-i}(j-i)\varrho^{j-i}\Big]G(x) \\
&= O(1)\frac{x^2 a_n^{-1}}{n} + O(1)\frac{x^2 f_{a_n}^{-1}(x)}{n^2} G(x). \qquad (6.15)
\end{aligned}$$

Note also that, when $a_n < 1$, $y^4 g(x+y, a_n) \le O(1)\,(a_n^{-1}x^4 + x^2 + a_n^3)$, $y^2 g(x+y, a_n) \le O(1)\,(a_n + x^2 a_n^{-1})$, and $g(x+y, a_n) \le O(1) a_n^{-1}$. Thus, it follows that

$$\begin{aligned}
B_2 &= O(1)\frac{x^2 a_n^{-1}}{n} + O(1)\frac{x^2}{n^2}(a_n^{-1}x^4 + a_n^{-1}x^2 + a_n^{-1}) f_{a_n}^{-1}(x) E_{\lambda_n}|g_i^*| \\
&= O(1)\frac{x^2 a_n^{-1}}{n} + O(1)\frac{x^2 a_n^{-1}}{n^2}(x^4 + x^2 + 1) = \frac{x^2 a_n^{-1}}{n}\Big[O(1) + \frac{O(1)x^4}{n}\Big], \qquad (6.16)
\end{aligned}$$

where $O(1)$ holds uniformly in all x. By (6.13), (6.14) and (6.16), result (b) holds. This completes the proof. 2

Lemma 6.3. *Under the assumptions of Theorem 6.1:*

$$(a) \qquad \max_{1\le j\le n} E_{\lambda_n}\left\{f_{a_n}^{-1}(x)\left[\hat f'_{a_n,t}(x,\lambda_n) - f'_{a_n}(x)\right]^2 \varepsilon_j^2(\lambda_n)\right\} \le \frac{a_n^{-3}}{n}(\kappa_1 + \kappa_2 x^2),$$

$$(b) \qquad \max_{1\le j\le n} E_{\lambda_n}\left\{x^2 f_{a_n}^{-1}(x)\left[\hat f'_{a_n,t}(x,\lambda_n) - f'_{a_n}(x)\right]^2 \varepsilon_j^4(\lambda_n)\right\} \le \frac{x^2 a_n^{-3}}{n}(\kappa_1 + \frac{\kappa_2 x^4}{n}),$$

where $1 \le t \le n$, *and* κ_1 *and* κ_2 *are constants independent of* n *and* x.

Proof. The proof is similar to that for Lemma 6.2, and hence it is omitted. 2

Lemma 6.4. *Under the assumptions of Theorem 6.1:*

$$(a) \qquad \frac{1}{n}\sum_{i=1}^{n}\max_{1\le j\le n} E_{\lambda_n}\left\{\varepsilon_j^2(\lambda_n)\int\Big[\frac{f'_{a_n}(x)}{\sqrt{f_{a_n}(x)}} - \frac{f'(x)}{\sqrt{f(x)}}\Big]^2 dx\right\} = o(1),$$

$$(b) \qquad \frac{1}{n}\sum_{i=1}^{n}\max_{1\le j\le n} E_{\lambda_n}\left\{\varepsilon_j^4(\lambda_n)\int\Big[\frac{f'_{a_n}(x)}{\sqrt{f_{a_n}(x)}} - \frac{f'(x)}{\sqrt{f(x)}}\Big]^2 x^2 dx\right\} = o(1).$$

Proof. Since $E_{\lambda_n}\varepsilon_i^4(\lambda_n)$ is bounded by a constant which is independent of i, the results follow from Lemma 6.2 of Bickel (1982) and Lemma 5.3 of Linton (1993). This completes the proof. 2

Lemma 6.5. *Suppose that assumptions of Theorem 6.1 hold. Then:*

$$(a)\qquad \frac{1}{n}\sum_{t=1}^{n}\max_{1\le j\le n}E_{\lambda_n}\Big\{\varepsilon_j^2(\lambda_n)\int[\hat{\xi}_{1n,t}(x,\lambda_n)-\frac{f'_{a_n}(x)}{f_{a_n}(x)}]^2 f_{a_n}(x)dx\Big\}=o(1),$$

$$(b)\qquad \frac{1}{n}\sum_{t=1}^{n}\max_{1\le j\le n}E_{\lambda_n}\Big\{\varepsilon_j^4(\lambda_n)\int[\hat{\xi}_{1n,t}(x,\lambda_n)-\frac{f'_{a_n}(x)}{f_{a_n}(x)}]^2 x^2 f_{a_n}(x)dx\Big\}=o(1).$$

Proof. The proof is similar to that in Kreiss (1987a). However, as we have the additional factor x^2 in (b), we need to avoid the requirement of higher moments. We will prove only (b), while (a) can be proved in a similar manner. Denote

$$I_1^{n,t}=\varepsilon_j^4(\lambda_n)\int_{A_{n,t}B_nC_{n,t}}\Big[\hat{\xi}_{1n,t}(x,\lambda_n)-\frac{f'_{a_n}(x)}{f_{a_n}(x)}\Big]^2x^2f_{a_n}(x)dx,$$

$$I_2^{n,t}=\varepsilon_j^4(\lambda_n)\int_{(A_{n,t}B_nC_{n,t})^c}\Big[\hat{\xi}_{1n,t}(x,\lambda_n)-\frac{f'_{a_n}(x)}{f_{a_n}(x)}\Big]^2x^2f_{a_n}(x)dx,$$

where $A_{n,t}=\Big\{x|\hat{f}_{a_n,t}(x,\lambda_n)\ge d_n\Big\}$, $B_n=\{x||x|\le g_n\}$ and $C_{n,t}=\{x|\hat{f}'_{a_n,t}(x,\lambda_n)\le c_n\ \hat{f}_{a_n,t}(x,\lambda_n)\}$. From Lemmas 6.2-6.3 and Assumption 4.1, we have

$$\frac{1}{n}\sum_{t=1}^{n}\max_{1\le j\le n}E_{\lambda_n}I_1^{n,t}\le\frac{1}{n}\sum_{t=1}^{n}\max_{1\le j\le n}$$
$$\Big\{E_{\lambda_n}\Big(\varepsilon_j^4(\lambda_n)\int_{A_{n,t}B_nC_{n,t}}\Big[\frac{\hat{f}'_{a_n,t}(x,\lambda_n)}{\hat{f}_{a_n,t}(x,\lambda_n)}-\frac{\hat{f}'_{a_n,t}(x,\lambda_n)}{f_{a_n}(x)}\Big]^2x^2f_{a_n}(x)dx\Big)$$
$$(6.17)\qquad +E_{\lambda_n}\Big(\varepsilon_j^4(\lambda_n)\int_{A_{n,t}B_nC_{n,t}}\Big[\frac{\hat{f}'_{a_n,t}(x,\lambda_n)}{f_{a_n}(x)}-\frac{f'_{a_n}(x,\lambda_n)}{f_{a_n}(x)}\Big]^2x^2f_{a_n}(x)dx\Big)\Big\}=o(1).$$

To show that $n^{-1}\sum_{t=1}^n\max_{1\le j\le n}E_{\lambda_n}I_2^{n,t}=o(1)$, because of Lemma 6.4(b), it is sufficient to show that $n^{-1}\sum_{t=1}^n\max_{1\le j\le n}E_{\lambda_n}J_2^{n,t}=o(1)$, where

$$J_2^{n,t}=\varepsilon_j^4(\lambda_n)\int_{(A_{n,t}B_nC_{n,t})^c}\left[\frac{f'(x)}{f(x)}\right]^2x^2f(x)dx.$$

As in Kreiss (1987a), we choose $t_n\in\{1,\cdots,n\}$, so that $E_{\lambda_n}J_2^{n,t_n}=\max\{E_{\lambda_n}J_2^{n,t},t=1,\cdots,n\}$. First, we show that

$$(6.18)\qquad \int\left[\frac{f'(x)}{f(x)}\right]^2x^2\max_{1\le j\le n}E_{\lambda_n}\left[\varepsilon_j^4(\lambda_n)I(A^c_{n,t_n})\right]f(x)dx=o(1).$$

Note that, for any positive constant M,

$$\max_{1\le j\le n} E_{\lambda_n}\left[\varepsilon_j^4(\lambda_n)I(A^c_{n,t_n})\right] = \max_{1\le j\le n}\left\{E_{\lambda_n}\left[\varepsilon_j^4(\lambda_n)I(\varepsilon_j^4(\lambda_n)\le M)I(A^c_{n,t_n})\right]\right.$$
$$\left.+E_{\lambda_n}\left[\varepsilon_j^4(\lambda_n)I(\varepsilon_j^4(\lambda_n)\ge M)I(A^c_{n,t_n})\right]\right\}$$
$$\le ME_{\lambda_n}\left[I(A^c_{n,t_n})\right] + \max_{1\le j\le n} E_{\lambda_n}\left[\varepsilon_j^4(\lambda_n)I(\varepsilon_j^4(\lambda_n)\ge M)\right].$$

For any $\epsilon > 0$, by Lemmas 5.2-5.3, we can show that there exists a large M_0 such that $\max_{1\le j\le n} E_{\lambda_n}[\varepsilon_j^4\ (\lambda_n)I(\varepsilon_j^4(\ \lambda_n)\ \ge\ M_0)]\ <\ \epsilon/2$. For such an M_0, since $E_{\lambda_n}(I(A^c_{n,t_n})) = o(1)$ by (6.11) of Bickel (1982), there exists an N such that, when $n > N$, $M_0E_{\lambda_n}(I(A^c_{n,t_n})) < \epsilon/2$. Thus, as $n > N$, $\max_{1\le j\le n} E_{\lambda_n}\left[\varepsilon_j^4(\lambda_n)I(A^c_{n,t_n})\right] < \epsilon$, i.e. $\max_{1\le j\le n} E_{\lambda_n}[\varepsilon_j^4(\lambda_n)\ I(A^c_{n,t_n})] = o(1)$. Furthermore, by Assumption 2.1, (6.18) holds.

Similarly, using the argument of (6.12) in Bickel (1982), we can show that

$$\int\left[\frac{f'(x)}{f(x)}\right]^2 x^2 \max_{1\le i\le n} E_{\lambda_n}\left[\varepsilon_j^4(\lambda_n)I(B^c_{n,t_n})\right] f(x)dx = o(1), \tag{6.19}$$

$$\int\left[\frac{f'(x)}{f(x)}\right]^2 x^2 \max_{1\le t\le n} E_{\lambda_n}\left[\varepsilon_j^4(\lambda_n)I(C^c_{n,t_n})\right] f(x)dx = o(1). \tag{6.20}$$

By (6.18)-(6.20), we can show that $\frac{1}{n}\sum_{i=1}^n \max_{1\le j\le n} E_{\lambda_n} I_2^{n,t} = o(1)$. Furthermore, combining (6.17), we have that (b) holds. This completes the proof. 2

Lemma 6.6. *Suppose that assumptions of Theorem 6.1 hold. Then:*

$$(a)\qquad \frac{1}{n}\sum_{t=1}^n \max_{1\le j\le n} E_{\lambda_n}\left\{\varepsilon_j^2(\lambda_n)\int\left[\hat{\xi}^2_{1n,t}(x,\lambda_n)(\sqrt{f_{a_n}(x)}-\sqrt{f(x)})\right]^2 dx\right\} = o(1),$$

$$(b)\qquad \frac{1}{n}\sum_{t=1}^n \max_{1\le j\le n} E_{\lambda_n}\left\{\varepsilon_j^4(\lambda_n)\int\left[\hat{\xi}^2_{1n,t}(x,\lambda_n)(\sqrt{f_{a_n}(x)}-\sqrt{f(x)})\right]^2 x^2dx\right\} = o(1).$$

Proof. Since $\hat{\xi}^2_{1n,t}(x,\lambda_n) \le c_n^2$ and $E_{\lambda_n}\left[\varepsilon_j^4(\lambda_n)\right]$ is bounded by some constant M which is independent of j, the proof of (a) is identical to that of Lemma 6.3 in Bickel (1982). The right-hand side of (b) is bounded by

$$c_n^2E_{\lambda_n}\left[\varepsilon_j^4(\lambda_n)\int(\sqrt{f_{a_n}(x)}-\sqrt{f(x)})^2x^2dx\right]. \tag{6.21}$$

In a similar manner to the arguments of Lemma 6.3 in Bickel (1982), we can show that (6.21) is bounded by

$$\frac{Mc_n^2a_n^2}{4}\int_0^1\int_{-\infty}^{\infty}\frac{y^2z^2f'^2(y-\mu a_n z)}{f(y-\mu a_n z)}g(z)dzd\mu$$

$$\leq \frac{Mc_n^2a_n^2}{4}\int_0^1\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+\mu a_n z)^2z^2\xi_1^2(x)f(x)g(z)dxdzd\mu. \tag{6.22}$$

By Assumption 2.1, (6.22) is bounded by $O(c_n^2a_n^2)$. Furthermore, by Assumption 4.1, (b) holds. This completes the proof. 2

A Appendix: Proof of Theorem 2.1

Before giving the proof of Theorem 2.1, we introduce the following notation and lemma. Let $\tilde{\xi}=(\xi',\zeta')'$, $\tilde{G}_n=$diag $(G_n,\sqrt{n}I_l)$, $\tilde{U}_t(\lambda)=$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:

(a) $$\sum_{t=1}^n E_{\lambda_n}\left[(u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t))^2|\mathcal{F}_{t-1}\right]=O_{\lambda_n}(1),$$

(b) $$\sum_{t=1}^n E_{\lambda_n}\left[(u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t))^2I(|u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t)|>\varepsilon)|\mathcal{F}_{t-1}\right]=o_{\lambda_n}(1),$$

(c) $$\Big|\sum_{t=1}^n\{(u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t))^2-E_{\lambda_n}[(u_n'\tilde{G}_n^{-1}\tilde{U}_t\tilde{\xi}(\eta_t))^2|\mathcal{F}_{t-1}]\}\Big|=o_{\lambda_n}(1),$$

(d) $$\sum_{t=1}^n\int\Big\{\frac{\sqrt{f_{v_n/\sqrt{n}}(Y_n)}}{\sqrt[4]{h_{nt}}}-\frac{\sqrt{f(Y)}}{\sqrt[4]{h_t}}-\frac{1}{2}u_n'\tilde{G}_n^{-1}\tilde{U}_t\frac{(\tilde{\xi}\sqrt{f})(Y)}{\sqrt[4]{h_t}}\Big\}^2dy=o_{\lambda_n}(1).$$

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 similar argument as in Koul and Schick (1997, p253), we can show that (b) holds. By (3.15) of McLeish (1974) and (a)-(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

$$T_{1n} = \sum_{t=1}^n\int h_{nt}^{-\frac{1}{2}}(f_{v_n/\sqrt{n}}^{\frac{1}{2}}-f^{\frac{1}{2}}-\frac{v_n'}{2\sqrt{n}}\zeta f^{\frac{1}{2}})^2(Y_n)dy$$

$$= \quad n\int (f^{\frac{1}{2}}_{v_n/\sqrt{n}} - f^{\frac{1}{2}} - \frac{v_n'}{2\sqrt{n}}\zeta f^{\frac{1}{2}})^2(x)dx = o(1)$$

by Definition 2.1;

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

where $R_{1n} = [\max_{1\le t\le n}(|Z_{nt} - Z_t|h_{nt}^{-\frac{1}{2}})^2]^{1/2} = o_{\lambda_n}(1)$ and $R_{2n} = [\max_{1\le t\le n}(|\sqrt{h_{nt}} - \sqrt{h_t}|h_{nt}^{-\frac{1}{2}})^2]^{1/2} = o_{\lambda_n}(1)$ by Assumption 2.2 (i)-(iii), and the above equation holds by $\int ||\zeta(x)||^2 f(x)dx < \infty$ and Lemma 19 in Jeganathan (1995); and

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

In order to show that $T_{3n} = o_{\lambda_n}(1)$, denote $U_{nt}^* = g_{nt} - g_t$, $Y_n^* = [y - Z_t - u(Z_{nt} - Z_t)]h_{nt}^{*-\frac{1}{2}}$ and $h_{nt}^* = [h_t^{\frac{1}{2}} + u(h_{nt}^{\frac{1}{2}} - h_t^{\frac{1}{2}})]^2$. 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

$$\begin{aligned} &\sum_{t=1}^{n}\int_0^1\int \left[U_{nt}^* h_{nt}^{*-\frac{3}{4}}(\xi f^{\frac{1}{2}})(Y_n^*) - \vartheta_n' G_n^{-1}U_t h_t^{-\frac{3}{4}}(\xi f^{\frac{1}{2}})(Y)\right]^2 dydu \\ &\qquad \le \sum_{t=1}^{n}\int_0^1\int \left\{\vartheta_n' G_n^{-1}U_t\left[h_{nt}^{*-\frac{3}{4}}(\xi f^{\frac{1}{2}})(Y_n^*) - h_t^{-\frac{3}{4}}(\xi f^{\frac{1}{2}})(Y)\right]\right\}^2 dydu \\ &\qquad\quad + \sum_{t=1}^{n}\left\|U_{nt}^* - \vartheta_n' G_n^{-1}U_t\right\|^2 \max\{h_t^{-1}, h_{nt}^{-1}\}\int \left\|(\xi f^{\frac{1}{2}})(x)\right\|^2 dx, \end{aligned} \tag{A. 1}$$

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

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

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. 2

Now, we prove Theorem 2.1. The basic idea of the proof comes from LeCam (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:

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

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

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

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

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

Note that $\int[s^y_{v_n/\sqrt{n},t}(\tilde{\lambda}_n) - s^y_t(\lambda_n)]^2 dy = -E_{\lambda_n}(T_{nt}|\mathcal{F}_n)$, where $s^y_t(\lambda)$ is defined as $s_t(\lambda)$ with $\eta_t(\lambda)$ replaced by Y, and similarly define $s^y_{v_n/\sqrt{n},t}(\lambda)$. 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 $a, b \in R$,

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

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

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

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

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 Lemma A.1 (b) and (d), and following the steps in DKW (1997, p.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 $\Lambda_n(\lambda_0, \lambda_n, 0) = -\Lambda_n(\lambda_n, \lambda_n + G_n'^{-1}(-\theta_n), 0) = -[\tilde{u}_n'\tilde{W}_n(\lambda_n) - \tilde{u}_n'\tilde{S}_n(\lambda_n)\tilde{u}_n/2] + o_{\lambda_n}(1)$ 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. 2

REFERENCES

Bai, J. (1993). On the partial sums of residuals in autoregressive and moving average models. *J. Time Ser. Anal.*, **14**, 247-260.

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 University 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 **IV**, 2959-3038, R.F. Engle and D.C. McFadden eds. Amsterdam, Elsevier Science.

Chung, K.L. (1968). *A Course in Probability Theory.* New York, Harcourt Brace.

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 Gonzȧlez-Riversa (1991). Semiparametric ARCH models. *J. Business & Economic Statistics*, **9**, 345-359.

Fabian, V. and Hannan, J. (1982). On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrsch. verw. Gebiet*, **47**, 31-46.

Fuller, W.A. (1976). *Introduction to Statistical Time Series.* Wiley, New York.

Hȧjek, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* I, 175-194, Univ. Cal.

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. *Econometric Theory*, **11**, 1148-1171.

Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory*, **11**, 818-887.

Kallianpur, G. (1980). *Stochastic Filtering Theory.* New York; Springer.

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. *Statistics & Decisions*, **5**, 59-76.

Kurtz, T.G. and Protter, P. (1991). Weak limit theorems to stochastic integrals and stochastic differential equations. *Ann. Prob.*, **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. (1999) A survey of recent theoretical results for time series models with GARCH errors. To appear in *Journal of Economic Surveys.*

Ling, S. (1999). On probability properties of a double threshold ARMA conditional heteroskedasticity model. *J. App. Prob.*, **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.(1999) A Monte Carlo Analysis of Unit Root Testing with GARCH (1,1) Errors, in L. Oxley, F. Scrimgeour and M. McAleer (eds.), Proc. of the International Congress on Modelling and Simulation, V.2, Uni. of Waikato, Hamilton, NZ, pp. 289-294.

Ling, S. and McAleer, M. (2000). Necessary and sufficient moment conditions for the GARCH (r, s) and asymmetric power GARCH (r, s) models. To appear in *Econometric Theory.*

Linton, O. (1993). Adaptive estimation in ARCH models. *Econometric Theory*, **9**, 535-569.

McLeish, D.L. (1974). Dependent central limit theorems and invariance principles. *Ann. Prob.*, **2**, 602-628.

Nelson, D.B. (1990) Stationarity and persistence in the GARCH (1,1) model. *Econometric Theory* **6**, 318-334.

Phillips, P.C.B. (1989). Partially identified econometric models. *Econometric 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. Applied Econometrics*, **3**, 35-51.

Schick, A. (1986). On asymptotically efficient estimation in semi-parametric models. *Ann. Statist.*, **14**, 1139-1151.

Schick, A. (1987). A note on the construction of asymptotically linear estimators. *J. Statist. Plann. Inf.*, **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. Inf.*, **19**, 217-228.

Shin, D.W. and So, B.S. (1999). Unit root tests based on adaptive maximum likelihood estimation. *Econometric Theory*, **15**, 1-23.

Weiss. A.A. (1986). Asymptotic theory for ARCH models: estimation and testing. *Econometric 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