Journal of Econometrics 167 (2012) 95–112

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

# ARCH/GARCH with persistent covariate: Asymptotic theory of MLE☆

Heejoon Han[a], Joon Y. Park[b,*]

[a] Department of Economics and Risk Management Institute, National University of Singapore, Singapore 117570, Singapore

[b] Department of Economics, Indiana University and Sungkyunkwan University, Bloomington, IN 47405-7104, USA

ARTICLE INFO

*Article history:*
Received 18 August 2009
Received in revised form
16 August 2011
Accepted 18 October 2011
Available online 28 October 2011

*JEL classification:*
C22
C50
G12

*Keywords:*
ARCH
GARCH
Persistent covariate
Maximum likelihood estimator
Asymptotic distribution theory

ABSTRACT

The paper considers a volatility model which introduces a persistent, integrated or near-integrated, covariate to the standard GARCH(1, 1) model. For such a model, we derive the asymptotic theory of the quasi-maximum likelihood estimator. In particular, we establish consistency and obtain limit distribution. The limit distribution is generally non-Gaussian and represented as a functional of Brownian motions. However, it becomes Gaussian if the covariate has innovation uncorrelated with the squared innovation of the model or the volatility function is linear in parameter. We provide a simulation study to demonstrate the relevance and usefulness of our asymptotic theory.



## 1. Introduction

The ARCH model introduced by Engle (1982) and its variants such as the GARCH and EGARCH models developed later by Bollerslev (1986) and Nelson (1991) are extremely popular. However, they are purely statistical, specifying the conditional variance as being determined only by its past or the lagged values of the process itself. They do not allow, at least in their prototypical forms, for other covariates to affect the volatility of the underlying process. Needless to say, this is quite undesirable. We often find it necessary, for some theoretical reasons or for better forecastability of the model, to include some relevant covariates in modeling volatilities of many economic and financial time series. Indeed, it seems fair to say that empirical researchers and practitioners routinely try potentially relevant covariates, when they fit ARCH type models.

Not surprisingly, many previous works considered ARCH type models, predominantly the GARCH(1, 1) model, with covariates. Glosten et al. (1993) and Engle and Patton (2001) used three month US Treasury bill rates for modeling stock return volatility, and Gray (1996) added the level of interest rates to explain the conditional variance in his generalized regime-switching model of short-term interest rates. Likewise, interest rate differentials between countries were used as covariates by Hagiwara and Herce (1999) to model exchange rate return volatility. There are many other examples, to which the reader is referred to Fleming et al. (2008). Moreover, many ARCH type models used in practice include covariates. For instance, the US congressional budget office uses inflation as a covariate in their ARCH(1) model for the volatilities of interest rate spreads, as can be seen in the memorandum by Congressional Budget Office (2000).

We have an extensive literature on the statistical analysis of ARCH type models, and their statistical theories have now been well developed. See Weiss (1986) for the ARCH($p$) model, and

☆ We are very grateful to a Co-Editor and three anonymous referees for many useful comments. This paper combines two working papers, "ARCH with persistent covariate" by Han and Park and "GARCH process with persistent covariates" by Han, which have been circulated in 2006 and 2007 respectively. We would like to thank Yoosoon Chang, Yongok Choi, Bradley Paye and seminar participants of the 15th annual meeting of the Midwest Econometric Group (Carbondale), the Rice University, the Indiana University Bloomington, the National University of Singapore, Bates White, the LLC, the Texas A&M University, the Conference in Honor of the 25th Anniversary of Beveridge–Nelson Decomposition at Atlanta Fed, the Korean Econometrics Study Group Meeting, and the International Conference on Econometrics at the Shanghai University of Finance and Economics for helpful comments and suggestions. This work was supported by the Korea Research Foundation Grant (KRF-2004-041-B00060). Han acknowledges research support from the NUS Risk Management Institute.

* Corresponding author.
*E-mail addresses:* ecshhj@nus.edu.sg (H. Han), joon@indiana.edu (J.Y. Park).

0304-4076/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2011.10.004

Lumsdaine (1996) and Lee and Hansen (1994) for the stationary GARCH(1, 1) and IGARCH(1, 1) models. Recently, Berkes et al. (2003) established the limit distribution theory of the maximum likelihood estimator (MLE) for the GARCH($p$, $q$) model.[1] Robinson and Zaffaroni (2006) developed the asymptotic distribution theory of the MLE in the class of ARCH($\infty$) models that allow for more general ARCH type models. Moreover, Jensen and Rahbek (2004a,b) obtained the asymptotic theory of the MLE for the ARCH(1) and GARCH(1, 1) models in the nonstationary case.

Nevertheless, there is no existing literature which investigates the statistical properties of ARCH models with additional covariates. This paper attempts to fill this gap. More specifically, we provide the asymptotics of the MLE[2] for the ARCH(1) and GARCH(1, 1) models with a persistent covariate, which has either a unit root or a near unit root. The persistence of covariate is an essential feature of our model. This feature is adopted in our model for two reasons. First, all of the aforementioned ARCH type models with covariates include variables that are widely believed to be persistent. Therefore, it is practically much more relevant to study ARCH type models with persistent covariates. Second, as shown in Park (2002) and Han and Park (2008), the presence of persistent covariates may very well explain some of the common characteristics such as the long memory property in volatility and leptokurtosis that we observe in many economic and financial time series including various speculative return series. The ARCH type models with stationary covariates cannot generate such characteristics.

In the paper, we establish the consistency of the MLE of the parameter in our model. We also derive its limit distribution, which is in general non-Gaussian and represented as a functional of Brownian motions as in other models involving unit root or near unit root processes. The standard inference is therefore generally invalid. However, the limit distribution reduces to be Gaussian in some special cases. In particular, it becomes mixed normal when the covariate has innovation uncorrelated with the squared innovation of the model. Furthermore, it becomes normal, if the volatility function is specified as linear in parameter. In these cases, the standard inference is valid and applicable even in the presence of a persistent covariate. Of the two special cases, the latter has a particularly important implication in practical applications. Indeed, empirical researchers often use the simple linear volatility function, in which case they may rely on the standard method of inference.

The rest of the paper is organized as follows. Section 2 introduces the GARCH(1, 1) model with a persistent covariate and assumptions. They allow for the ARCH(1) model with a persistent covariate as a special case. The asymptotics of the MLE for our model is presented in Section 3. In particular, we establish the consistency of the MLE and obtain its limit distribution. The simulation results are reported in Section 4. Section 5 concludes the paper, and Appendices A and B contain mathematical proofs for the technical results in the paper. Finally a word on notation. We denote by $\mathbb{R}_+$ and $\mathbb{R}_{++}$ the sets of real numbers that are nonnegative and positive, respectively, and $\|\cdot\|$ signifies the usual Euclidean norm for $\mathbb{R}^n$. Standard terminologies and notations in probability and measure theory are used throughout the paper. In particular, notations for various convergence modes such as $\to_{a.s.}$, $\to_p$ and $\to_d$ will frequently appear. Throughout the paper, we denote the sample size by $n$, and all limits are taken as $n \to \infty$, except where otherwise indicated.

1 Giraitis and Robinson (2001) considered the Whittle estimator for the general GARCH($p$, $q$) model. The Whittle estimation is asymptotically equivalent to the least squares regression of $(y_t^2)$ on its past values and thus is not fully efficient.

2 We do not assume normality, and therefore, it is the quasi-maximum likelihood estimator (QMLE). However, it will be referred simply to as the MLE throughout the paper.

## 2. The model and assumptions

We consider the volatility model specified as

$$y_t = \sigma_t \varepsilon_t,$$

and let $(\mathcal{F}_t)$ be the filtration representing the information available at time $t$.

**Assumption 1.** Assume that

(a) $(\varepsilon_t)$ is iid (0, 1), and adapted to $(\mathcal{F}_t)$,
(b) $(\sigma_t)$ is given by

$$\sigma_t^2 = \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 + f(x_t, \pi)$$

for parameters $\alpha, \beta \in \mathcal{A} \subset \mathbb{R}_+$ such that $\alpha + \beta < 1$ and $\pi \in \Pi \subset \mathbb{R}^d$ with $\mathcal{A}$ and $\Pi$ compact and a volatility function $f : \mathbb{R} \times \Pi \to \mathbb{R}_+$, and let

$$x_t = \left(1 - \frac{c}{n}\right) x_{t-1} + v_t,$$

where $(x_t)$ is adapted to $(\mathcal{F}_{t-1})$ and $c \geq 0$, and
(c) $(f(x_t, \pi))$ is strictly positive a.s. for all $t$ and $\pi \in \Pi \subset \mathbb{R}^d$.

In our model, the covariate $(x_t)$ is defined as a near-integrated process with the dominant AR coefficient given as a function of the sample size $n$. Strictly speaking, $(x_t)$ should be denoted by $(x_{n,t})$ for being triangular arrays, and we should subsequently write $(\sigma_t)$ and $(y_t)$ also as $(\sigma_{n,t})$ and $(y_{n,t})$ accordingly. In the paper, however, we follow the usual convention and use the simple notation for expositional brevity. This should cause no confusion. The process $(x_t)$ with the exact unit root will be considered as a special case with $c = 0$ in the subsequent development of our asymptotic theory. It is also possible to accommodate for the models with stationary covariate $(x_t)$, if we let $c \to \infty$ in our asymptotics. However, it will not be pursued in the paper, since the asymptotics for the models with stationary covariates are largely identical to those of stationary ARCH type models.

Under Assumption 1, $(y_t)$ becomes the GARCH(1, 1) process with a persistent, integrated or near-integrated, covariate. We have

$$y_t^2 = (\alpha + \beta) y_{t-1}^2 + f(x_t, \pi) + (u_t - \beta u_{t-1}),$$

where $(u_t)$ is a martingale difference sequence defined by $u_t = y_t^2 - \sigma_t^2$. Now it can be clearly seen that $(y_t^2)$ becomes a nonlinear ARMAX(1, 1) process under our volatility model, whereas, as is well known, it reduces to the ARMA(1, 1) process under the GARCH(1, 1) model without any covariate. Throughout the paper, we assume that $\alpha + \beta < 1$, under which $(y_t^2)$ is expected to behave like an integrated process of order one. We do not consider the possibility of $\alpha + \beta = 1$ in the paper, since in this case $(y_t^2)$ becomes as persistent as an integrated process of order two. This is not very likely for most economic and financial time series. The time series properties of the process we consider here are investigated in Han and Park (2008).[3] In particular, they find that the process exhibits persistence in volatility and leptokurtosis, the characteristics that are commonly observed in many economic and financial time series. If squared, the process indeed yields the sample autocorrelation function that does not vanish at all lags or decays very slowly at a polynomial rate. The sample kurtosis of the process also either diverges or remains to be random over the support with a lower bound given by the kurtosis of innovation $(\varepsilon_t)$ in the model.

**Assumption 2.** Assume that

(a) $(v_t)$ is generated by

$$v_t = \phi(L)\eta_t = \sum_{k=0}^{\infty} \phi_k \eta_{t-k},$$

3 Though they only consider the ARCH(1) process, all the time series properties they derive are shared at least qualitatively by the more general GARCH(1, 1) model.

where $\phi_0 = 1$, $\phi(1) \neq 0$ with $\sum_{k=0}^{\infty} k|\phi_k| < \infty$, and $(\eta_{t+1})$ is iid jointly with $(\varepsilon_t)$, adapted to $(\mathcal{F}_t)$ with mean zero and $\mathbb{E}|\eta_t|^p < \infty$ for some $p > 2$,

(b) $\mathbb{E}|\varepsilon_t|^q < \infty$ and $\mathbb{E}\left(\beta + \alpha\varepsilon_t^2\right)^{q/2} < 1$ for some $q > 4$, and

(c) $1/p + 2/q < 1/2$.

Assumption 2 more precisely defines the covariate $(x_t)$ as an integrated or near-integrated process driven by a general linear process, and introduces moment conditions for the innovation sequences $(\eta_{t+1})$ and $(\varepsilon_t)$.

The conditions in part (a) are standard. For the conditions in part (b), note that

$$\mathbb{E}\left(\ln\left(\beta + \alpha\varepsilon_t^2\right)\right) \leq \ln\left[\mathbb{E}\left(\beta + \alpha\varepsilon_t^2\right)\right] \leq \frac{2}{q}\ln\left[\mathbb{E}\left(\beta + \alpha\varepsilon_t^2\right)^{q/2}\right]$$

for any $q > 2$, by the successive applications of Jensen's inequality. As a result, it follows from part (b) that $\mathbb{E}\left(\ln\left(\beta + \alpha\varepsilon_t^2\right)\right) < 0$, which as shown in Nelson (1990) is the necessary and sufficient condition for the GARCH(1, 1) process (without covariates) to be strictly stationary and ergodic.[4] Moreover, note that it also follows from part (b) that

$$\mathbb{E}\left(\beta + \alpha\varepsilon_t^2\right)^2 \leq \left(\mathbb{E}\left(\beta + \alpha\varepsilon_t^2\right)^{q/2}\right)^{4/q} < 1 \tag{1}$$

by Jensen's inequality. The condition, $\mathbb{E}\left(\beta + \alpha\varepsilon_t^2\right)^2 < 1$, is necessary in the investigation of the statistical properties of the GARCH(1, 1) process. If $\mathbb{E}\left(\beta + \alpha\varepsilon_t^2\right)^2 < 1$, the sample autocorrelation of squared process and the sample kurtosis have probability limits in the GARCH(1, 1) model. Due to part (c), the conditions in part (b) should hold for large $q$ if $p$ is small in part (a). As $p \to \infty$, we may choose $q$ arbitrarily close to 4. Under the Gaussianity of innovation sequences, we have $p = \infty$ and $\mathbb{E}\varepsilon_t^4 = 3$. Consequently, we may allow for any $\alpha$ and $\beta$ in the range of $0 \leq \left(\beta^2 + 2\alpha\beta + 3\alpha^2\right) < 1$ by taking $q$ sufficiently close to 4.

Now we introduce the conditions for volatility function $f$. Throughout the paper, we assume that $f$ is asymptotically homogeneous in the sense of Park and Phillips (2001). Roughly, asymptotically homogeneous functions are the functions that behave like homogeneous functions in the limit. More precisely, $f$ is said to be asymptotically homogeneous if it can be represented as

$$f(\lambda x, \pi) = \kappa(\lambda, \pi)\bar{f}(x, \pi) + o(\kappa(\lambda, \pi)) \tag{2}$$

for all large $\lambda$, uniformly in $x$ over any compact interval on $\mathbb{R}$ and uniformly in $\pi$ over the parameter set $\Pi$, with some non-vanishing function $\bar{f}$ on $\mathbb{R} \times \Pi$. We call $\kappa$ and $\bar{f}$, respectively, the asymptotic order and limit homogeneous function of $f$,[5] and say that $f$ is asymptotically homogeneous with asymptotic order $\kappa$ and limit homogeneous function $\bar{f}$. The reader is referred to Park and Phillips (2001) for more discussions on asymptotically homogeneous functions. They show that many functions that are commonly used in nonlinear econometric models are asymptotically homogeneous, including the logarithmic function, logistic function (and other distribution function-like functions) and power function. In what follows, we assume that the limit homogeneous function $\bar{f}$ of $f$ satisfies the mild technical regularity conditions introduced in Park and Phillips (2001).

Simple volatility functions such as $f(x, \pi) = \pi|x|$ and $f(x, \pi) = \pi x^2$, which are linear in parameter, are often used in practical applications. Another volatility function that is also of primary interest is given by

$$f(x, \pi) = \pi_1|x|^{\pi_2} \tag{3}$$

with $\pi = (\pi_1, \pi_2)' \in \Pi \subset \mathbb{R}^2_{++}$. The function $f$ in (3) is often called the constant elasticity of variance (CEV) volatility, since it implies that the elasticity of variance is constant. It is easy to see that the CEV volatility function is asymptotically homogeneous with asymptotic order $\kappa(\lambda, \pi) = \lambda^{\pi_2}$ and limit homogeneous function $\bar{f}(x, \pi) = \pi_1|x|^{\pi_2}$.[6] In what follows, we say that the two asymptotically homogeneous functions are 'equivalent' if they are different only over a compact interval. Clearly, the equivalent asymptotically homogeneous functions have the same asymptotic order and limit homogeneous function.

**Assumption 3.** The volatility function $f$ satisfies the following conditions:

(a) $f$ is asymptotically homogeneous with asymptotic order $\kappa$ such that $\inf_{\pi\in\Pi} \kappa(\lambda, \pi) \to \infty$ as $\lambda \to \infty$, and

(b) $f$ has an equivalent asymptotically homogeneous function $\tilde{f}$, say, such that

$$|\tilde{f}(x, \pi) - \tilde{f}(y, \pi)| \leq \nabla f(z, \pi)|x - y|$$

for any $x, y$ in a compact subset of $\mathbb{R}$, where $x \leq z \leq y$ and $\nabla f$ is asymptotically homogeneous with asymptotic order $\nabla\kappa$ such that

$$\nabla\kappa(\lambda, \pi) = O(\kappa(\lambda, \pi)/\lambda) \quad \text{for all } \pi \in \Pi, \text{ as } \lambda \to \infty.$$

It is not difficult to see that the conditions in Assumption 3 hold for the CEV volatility function introduced in (3). Regarding the conditions in Assumption 3, the role of parameter $\pi_1$ is unimportant. To avoid unnecessary complications and focus on the role of parameter $\pi_2$ in discussing these conditions, we therefore look at a simpler version $f(x, \pi) = |x|^{\pi}$ with $\pi \in \Pi \subset \mathbb{R}_{++}$. Obviously, the function is asymptotically homogeneous and its asymptotic order is given by $\kappa(\lambda, \pi) = \lambda^{\pi}$. Therefore, the condition in part (a) is clearly satisfied. For $\pi > 1$, it is also easy to see that the condition in part (b) holds with $\tilde{f} = f$, $\nabla f(x, \pi) = (\partial/\partial x)f(x, \pi)$ and $\nabla\kappa(\lambda, \pi) = \lambda^{\pi-1}$. For $0 < \pi \leq 1$, we may just define $\tilde{f}$ to be $f$ smoothed locally around the origin to see that the condition is met with $\nabla f(x, \pi) = (\partial/\partial x)\tilde{f}(x, \pi)$ and $\nabla\kappa(\lambda, \pi) = \lambda^{\pi-1}$. As can be readily seen from this example, the introduction of an equivalent asymptotically homogeneous function in part (b) allows us to consider volatility functions, which are not smooth in every point of the domain. For the sake of brevity in exposition, we assume in what follows that the local smoothing has already been done and $f$ itself satisfies the condition in part (b).

We need additional technical conditions to fully develop the asymptotic theory of the MLE. Before we state the conditions, it will be convenient to introduce some new notations and conventions. Define

$$\dot{f} = \left(\frac{\partial f}{\partial \pi_i}\right), \qquad \ddot{f} = \left(\frac{\partial^2 f}{\partial \pi_i \partial \pi_j}\right), \qquad \dddot{f} = \left(\frac{\partial^3 f}{\partial \pi_i \partial \pi_j \partial \pi_k}\right)$$

to be all vectors, arranged by the lexicographic ordering of their indices, if they exist. Moreover, we will denote the asymptotic order and limit homogeneous function of asymptotically homogeneous

[4] Most works pertaining to the asymptotic distribution theory of the MLE in GARCH models depend on this finding. See Lee and Hansen (1994), Lumsdaine (1996) and Berkes et al. (2003). Recently, Jensen and Rahbek (2004b) established the asymptotic distribution theory of the MLE in the nonstationary GARCH(1, 1) model under $\mathbb{E}\left[\ln\left(\beta + \alpha\varepsilon_t^2\right)\right] \geq 0$. See also Berkes et al. (2005) for the asymptotic analysis of the near-integrated GARCH(1, 1) processes.

[5] In general, the asymptotic order and limit homogeneous function are not uniquely determined. We may define any $\kappa$ and $\bar{f}$ to be the asymptotic order and limit homogeneous function of $f$ as long as they yield the representation in (2).

[6] As discussed, they are not uniquely defined. For instance, we may set $\kappa(\lambda, \pi) = \pi_1\lambda^{\pi_2}$ and $\bar{f}(x, \pi) = |x|^{\pi_2}$. Of course, the choice of $\kappa$ and $\bar{f}$ does not affect our asymptotic theory.

$\dot{f}$ by $\dot{\kappa}$ and $\bar{f}^{\cdot}$, respectively, and the asymptotic order of asymptotically homogeneous $\ddot{f}$ by $\ddot{\kappa}$. Throughout the paper, we write $\kappa_0(\lambda) = \kappa(\lambda, \pi_0)$, $\dot{\kappa}_0(\lambda) = \dot{\kappa}(\lambda, \pi_0)$ and $\ddot{\kappa}_0(\lambda) = \ddot{\kappa}(\lambda, \pi_0)$ for notational simplicity. Finally, we define a neighborhood of $\pi_0$ by

$$N = \left\{\pi \in \Pi : \left\|(\dot{\kappa}_0/\kappa_0)(\lambda)'(\pi - \pi_0)\right\| \leq \lambda^{-1+\varepsilon}\right\}$$

for $\varepsilon > 0$ given.

**Assumption 4.** Assume

(a) $\dot{f}, \ddot{f}$ and $\dddot{f}$ exist, and $\dot{f}$ and $\ddot{f}$ are asymptotically homogeneous,
(b) $\dot{f}/f$ is asymptotically homogeneous with asymptotic order $\dot{\kappa}/\kappa$ and limit homogeneous function $\bar{f}^{\cdot}/\bar{f}$, and
(c) for any $\bar{s} > 0$ given, there exists $\varepsilon > 0$ such that

$$\lambda^{-1}\left\|\kappa_0(\lambda)\left(\dot{\kappa}_0 \otimes \dot{\kappa}_0\right)(\lambda)^{-1}\ddot{\kappa}_0(\lambda)\right\| \to 0, \tag{4}$$

$$\lambda^{-1+\varepsilon}\left\|\kappa_0^2(\lambda)\dot{\kappa}_0(\lambda)^{-1}\sup_{|s|\leq\bar{s}} \times\left(\sup_{\pi\in N}\left|f(\lambda s, \pi)^{-2}\right|\sup_{\pi\in N}\left|\dot{f}(\lambda s, \pi)\right|\right)\right\| \to 0, \tag{5}$$

$$\lambda^{-1+\varepsilon}\left\|\kappa_0(\lambda)\dot{\kappa}_0(\lambda)^{-1}\sup_{|s|\leq\bar{s}}\sup_{\pi\in N}\left|\dot{f}(\lambda s, \pi)\right|\right\| \times\left\|\kappa_0^2(\lambda)\left(\dot{\kappa}_0 \otimes \dot{\kappa}_0\right)(\lambda)^{-1}\sup_{|s|\leq\bar{s}} \times\left(\sup_{\pi\in N}\left|f(\lambda s, \pi)^{-2}\right|\sup_{\pi\in N}\left|\ddot{f}(\lambda s, \pi)\right|\right)\right\| \to 0, \tag{6}$$

$$\lambda^{-1+\varepsilon}\left\|\kappa_0^3(\lambda)\left(\dot{\kappa}_0 \otimes \dot{\kappa}_0 \otimes \dot{\kappa}_0\right)(\lambda)^{-1}\sup_{|s|\leq\bar{s}} \times\left(\sup_{\pi\in N}\left|f(\lambda s, \pi)^{-1}\right|\sup_{\pi\in N}\left|\dddot{f}(\lambda s, \pi)\right|\right)\right\| \to 0 \tag{7}$$

as $\lambda \to \infty$.

It is easy to see that all the conditions in Assumption 4 hold for simple volatility functions such as $f(x, \pi) = \pi|x|$ or $f(x, \pi) = \pi x^2$, which are linear in parameter. For the CEV volatility function $f(x, \pi) = \pi_1|x|^{\pi_2}$, we have in particular

$$\dot{\kappa}(\lambda, \pi) = \begin{pmatrix} \lambda^{\pi_2} & 0 \\ \pi_1\lambda^{\pi_2}\log\lambda & \pi_1\lambda^{\pi_2} \end{pmatrix} \quad \text{and}$$

$$\bar{f}^{\cdot}(x, \pi) = \begin{pmatrix} x^{\pi_2} \\ x^{\pi_2}\log x \end{pmatrix}.$$

It is tedious, but straightforward to show that the CEV volatility function satisfies all the conditions in Assumption 4.[7]

## 3. Asymptotic theory

In this section, we derive the asymptotic theory of the MLE for our model. Let $\theta = (\alpha, \beta, \pi')'$ with the true value $\theta_0 = (\alpha_0, \beta_0, \pi_0')'$ and the parameter set $\Theta = \mathcal{A} \times \Pi$, and let $\sigma_t^2(\theta) = \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2 + f(x_t, \pi)$. Also, we denote by $\ell_t(\theta)$ the conditional log-likelihood for $y_t$ given $\mathcal{F}_{t-1}$ for $t = 1, \ldots, n$. Then the log-likelihood function of the entire sample $(y_1, \ldots, y_n)$ is given by

$$\sum_{t=1}^n \ell_t(\theta) = -\frac{1}{2}\sum_{t=1}^n\left(\log\sigma_t^2(\theta) + \frac{y_t^2}{\sigma_t^2(\theta)}\right)$$

ignoring the unimportant constant term, and the MLE $\hat{\theta}_n$ is defined as

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\operatorname{argmax}}\sum_{t=1}^n \ell_t(\theta).$$

We do not assume Gaussianity in our asymptotic theory, and hence $\hat{\theta}_n$ is the QMLE to be more precise. However, we will simply refer to $\hat{\theta}_n$ as the MLE in the paper.

Let $\sigma_t^2 = \sigma_t^2(\theta)$. Then the score vector $s_n(\theta)$ and Hessian matrix $H_n(\theta)$ are given by

$$s_n(\theta) = \sum_{t=1}^n \frac{\partial\ell_t(\theta)}{\partial\theta} = \frac{1}{2}\sum_{t=1}^n\left(\frac{y_t^2}{\sigma_t^2} - 1\right)\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta},$$

$$H_n(\theta) = \sum_{t=1}^n \frac{\partial^2\ell_t(\theta)}{\partial\theta\partial\theta'} = \frac{1}{2}\sum_{t=1}^n\left[\left(1 - \frac{2y_t^2}{\sigma_t^2}\right)\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'} + \left(\frac{y_t^2}{\sigma_t^2} - 1\right)\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta'}\right].$$

The asymptotics of $\hat{\theta}_n$ in our model can be obtained from the first order Taylor expansion of the score vector, i.e.,

$$s_n(\hat{\theta}_n) = s_n(\theta_0) + H_n(\theta_n)(\hat{\theta}_n - \theta_0), \tag{8}$$

where $\theta_n$ lies in the line segment connecting $\hat{\theta}_n$ and $\theta_0$. If $\hat{\theta}_n$ is an interior solution, we have $s_n(\hat{\theta}_n) = 0$. Therefore, we may write from (8)

$$\nu_n'(\hat{\theta}_n - \theta_0) = -\left[\nu_n^{-1}H_n(\theta_n)\nu_n^{-1'}\right]^{-1}\left[\nu_n^{-1}s_n(\theta_0)\right] \tag{9}$$

for an appropriately defined sequence $(\nu_n)$ of normalization matrix.

The following conditions ML1–ML3 are sufficient to derive the asymptotics for $\hat{\theta}_n$ upon appropriately choosing the sequence $(\nu_n)$ of normalization matrix.

ML1: $\nu_n^{-1}s_n(\theta_0) \to_d P$ for some $P$,
ML2: $-\nu_n^{-1}H_n(\theta_0)\nu_n^{-1'} \to_d Q$ for some $Q > 0$ a.s., and
ML3: there exists a sequence $(\mu_n)$ of invertible normalization matrices such that $\mu_n\nu_n^{-1} \to 0$, and such that

$$\sup_{\theta\in N_n}\left\|\mu_n^{-1}\left(H_n(\theta) - H_n(\theta_0)\right)\mu_n^{-1'}\right\| \to_p 0,$$

where $N_n = \left\{\theta \in \Theta : \left\|\mu_n'(\theta - \theta_0)\right\| \leq 1\right\}$ is a sequence of shrinking neighborhoods of $\theta_0$.

As shown in Wooldridge (1994), we may indeed deduce from ML1–ML3 that

$$\nu_n'(\hat{\theta}_n - \theta_0) = -\left[\nu_n^{-1}H_n(\theta_0)\nu_n^{-1'}\right]^{-1}\left[\nu_n^{-1}s_n(\theta_0)\right] + o_p(1) \to_d Q^{-1}P.$$

This is as expected from (9). In particular, ML3 ensures that $s_n(\hat{\theta}_n) = 0$ with probability approaching to one and

$$\nu_n^{-1}\left(H_n(\theta_n) - H_n(\theta_0)\right)\nu_n^{-1'} \to_p 0.$$

Now we let

$$z_t = 1 + \sum_{k=1}^\infty\prod_{i=1}^k\left(\beta_0 + \alpha_0\varepsilon_{t-i}^2\right) \tag{10}$$

for $t \geq 1$.[8] Moreover, we define $u_t = (u_{1t}, u_{2t}, u_{3t})'$ to be given by

[7] The detailed proof is not given to save the space. The proof will be available from the authors upon request.

[8] Note that $(z_t)$ also appears in the analysis of usual GARCH models. In fact, if $(y_t)$ is generated as the GARCH(1, 1) process with its conditional variance given by $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2$, then we may easily deduce by the recursive substitution that $\sigma_t^2 = \omega z_t$ for $t \geq 1$.

$$u_{1t} = \sum_{i=1}^{\infty} \beta_0^{i-1} \frac{z_{t-i}\varepsilon_{t-i}^2}{z_t}(\varepsilon_t^2 - 1),$$

$$u_{2t} = \sum_{i=1}^{\infty} \beta_0^{i-1} \frac{z_{t-i}}{z_t}(\varepsilon_t^2 - 1),$$

$$u_{3t} = \frac{1}{(1-\beta_0)\, z_t}(\varepsilon_t^2 - 1),$$

and let $w_t = (u_t', v_t)'$. Then we have

**Lemma 1.** *Under Assumptions 1 and 2, we have*

$$W_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} w_t \to_d W(r),$$

*where $W$ is a vector Brownian motion with covariance matrix given by the long-run variance of $(w_t)$.*

For the subsequent development of our asymptotics, we decompose $W = (U', V)'$, $U = (U_1, U_2, U_3)'$, conformably with $(w_t)$, and note that the variance of $U$ is given by $(K^4 - 1)\Omega$, where (see the equation in Box I)

The two limit Brownian motions $U$ and $V$ become independent if $\mathbb{E}(\eta_{t+1}\varepsilon_t^2) = 0$. To develop our asymptotics, we also need to introduce an Ornstein–Uhlenbeck process $V_c$, which is defined as

$$V_c(r) = \int_0^r \exp(-c(r-s))\, dV(s)$$

from the limit Brownian motion $V$ introduced in Lemma 1, where we let $W = (U', V)'$. As is well known, $V_c$ is the solution of the stochastic differential equation

$$dV_c(r) = -cV_c(r)\, dr + dV(r)$$

with the initial condition $V_c(0) = 0$.

It follows that

**Lemma 2.** *Let Assumptions 1–4 hold. Then all the conditions in* ML1–ML3 *are satisfied, and*

$$\nu_n^{-1} s_n(\theta_0) \to_d \frac{1}{2} \int_0^1 M(r)\, dU(r),$$

$$-\nu_n^{-1} H_n(\theta_0) \nu_n^{-1'} \to_d \frac{1}{2} \int_0^1 M(r)\Omega M(r)' dr,$$

*where*

$$\nu_n = \sqrt{n}\, \mathrm{diag}\left(1, 1, \frac{\dot{\kappa}_0(\sqrt{n})}{\kappa_0(\sqrt{n})}\right),$$

$$M(r) = \mathrm{diag}\left(1, 1, \frac{\dot{\bar{f}}_0(V_c(r))}{\bar{f}_0(V_c(r))}\right)$$

*and other notations are defined earlier.*

We may readily obtain the limit distribution of the MLE $\hat{\theta}_n$ from Lemma 2, which is presented in the following theorem.

**Theorem 3.** *Let Assumptions 1–4 hold. Then*

$$\nu_n'(\hat{\theta}_n - \theta_0) \to_d \left(\int_0^1 M(r)\Omega M(r)' dr\right)^{-1} \int_0^1 M(r)\, dU(r)$$

*in notations introduced in Lemma 2.*

The ARCH(1) model with a persistent covariate may simply be regarded as a special case of our model.[9] If $\beta = 0$, the asymptotic distribution of the MLE $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\pi}_n')'$ is the same as the distribution in Theorem 3 with

$$\nu_n = \sqrt{n}\, \mathrm{diag}\left(1, \frac{\dot{\kappa}_0(\sqrt{n})}{\kappa_0(\sqrt{n})}\right), \qquad M(r) = \mathrm{diag}\left(1, \frac{\dot{\bar{f}}_0(V_c(r))}{\bar{f}_0(V_c(r))}\right)$$

and

$$\Omega = \mathbb{E}\begin{pmatrix} (z_{t-1}^2/z_t^2)\varepsilon_{t-1}^4 & (z_{t-1}/z_t^2)\varepsilon_{t-1}^2 \\ (z_{t-1}/z_t^2)\varepsilon_{t-1}^2 & 1/z_t^2 \end{pmatrix},$$

where $z_t = 1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} \alpha_0 \varepsilon_{t-i}^2$.

The MLE $\hat{\theta}_n$ is generally consistent. In particular, the GARCH parameters $\hat{\alpha}_n$ and $\hat{\beta}_n$ have the standard $\sqrt{n}$ rate of convergence. In contrast, the covariate parameter $\hat{\pi}_n$ has a rate $\sqrt{n}(\dot{\kappa}_0/\kappa_0)(\sqrt{n})$, which is dependent upon the asymptotic orders $\kappa$ and $\dot{\kappa}$ of the volatility function $f$ and its derivative $\dot{f}$ with respect to $\pi$. For instance, for the CEV volatility function in (3), we have

$$\sqrt{n}\frac{\dot{\kappa}_0(\sqrt{n})}{\kappa_0(\sqrt{n})} = \begin{pmatrix} (1/\pi_{10})\sqrt{n} & 0 \\ \sqrt{n}\ln\sqrt{n} & \sqrt{n} \end{pmatrix}$$

as the convergence rate for $\hat{\pi}_n$.

In general, the limit distribution of the MLE $\hat{\theta}_n$ is non-Gaussian. This is true for both the GARCH parameters $\hat{\alpha}_n$ and $\hat{\beta}_n$ and the covariate parameter $\hat{\pi}_n$. However, there are two cases, for which the limit distribution of $\hat{\theta}_n$ reduces to mixed normal or even normal. First, consider the special case where the limit Brownian motion $V$ is independent of $U$, in which case we will simply say that the covariate $(x_t)$ is *exogenous*. As discussed, the covariate $(x_t)$ becomes exogenous when its innovation $(\eta_{t+1})$ is uncorrelated with the squared model innovation $(\varepsilon_t^2)$, i.e., $\mathbb{E}(\eta_{t+1}\varepsilon_t^2) = 0$.[10] In this case, it follows immediately from Theorem 3 that

$$\nu_n'(\hat{\theta}_n - \theta_0) \to_d \mathbb{MN}\left(0, (K^4 - 1)\left(\int_0^1 M(r)\Omega M(r)' dr\right)^{-1}\right),$$

where $\mathbb{MN}$ stands for mixed normal distribution. Second, if the volatility function is *linear in parameter* and given by $f(x, \pi) = \pi g(x)$ for some asymptotically homogeneous function $g$, then we have $\nu_n = \sqrt{n}\,\mathrm{diag}(1, (1/\pi_0))$ and $M(r)$ becomes an identity matrix for all $r \in [0, 1]$. Consequently, we have

$$\nu_n'(\hat{\theta}_n - \theta_0) \to_d \mathbb{N}\left(0, (K^4 - 1)\Omega^{-1}\right),$$

where $\mathbb{N}$ signifies normal distribution. Note that the variance of the limit Brownian motion $U$ is given by $(K^4 - 1)\Omega$, which reduces to $2\Omega$ if the model innovation $(\varepsilon_t)$ is normal.

The normality and mixed normality in these two cases ensure the standard inference to be generally valid for our model. In fact, we may easily show that the usual asymptotic tests such as Likelihood Ratio, Lagrange Multiplier and Wald tests based on the MLE are all valid in these cases, if the model innovation $(\varepsilon_t)$ is normal. To see this, we consider the null hypothesis given by

$$\mathrm{H}_0 : \varphi(\theta_0) = 0, \tag{11}$$

where $\varphi : \mathbb{R}^a \to \mathbb{R}^b$ is continuously differentiable with $a$ and $b$ signifying the numbers of parameters and restrictions, respectively. The null hypothesis (11) can be tested using the statistic

$$\tau_n = \varphi(\hat{\theta}_n)' \left[-\Phi(\hat{\theta}_n) H_n(\hat{\theta}_n)^{-1} \Phi(\hat{\theta}_n)'\right]^{-1} \varphi(\hat{\theta}_n), \tag{12}$$

[9] For the asymptotic analysis of GARCH processes with some zero true coefficients, the reader is referred to Francq and Zakoian (2007, 2010).

[10] Note that we allow the innovation $(\eta_{t+1})$ of covariate $(x_t)$ to be correlated with the model innovation $(\varepsilon_t)$ itself. In fact, if we set $\eta_{t+1} = \varepsilon_t$, then $(\eta_{t+1})$ becomes uncorrelated with $(\varepsilon_t^2)$ whenever $\mathbb{E}\varepsilon_t^3 = 0$.

$$\Omega = \mathbb{E}\left[\frac{1}{z_t^2}\begin{pmatrix} \left(\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\varepsilon_{t-i}^2\right)^2 & \sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\varepsilon_{t-i}^2\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i} & \frac{1}{1-\beta_0}\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\varepsilon_{t-i}^2 \\ \sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\varepsilon_{t-i}^2\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i} & \left(\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\right)^2 & \frac{1}{1-\beta_0}\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i} \\ \frac{1}{1-\beta_0}\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\varepsilon_{t-i}^2 & \frac{1}{1-\beta_0}\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i} & \frac{1}{(1-\beta_0)^2}\end{pmatrix}\right],$$

and $K^4 = \mathbb{E}\varepsilon_t^4$. Note that $(u_t)$ is a martingale difference sequence. As we show in the proof of Lemma 1, $\Omega$ is well defined. The variance of $V$ is $[\phi(1)]^2\mathbb{E}(\eta_t^2)$, and the covariance between $U$ and $V$ is given by

$$\phi(1)\left[\mathbb{E}(\eta_{t+1}\varepsilon_t^2)\right]\mathbb{E}\left[\frac{1}{z_t}\left(\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\varepsilon_{t-i}^2, \sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}, \frac{1}{1-\beta_0}\right)\right]'.$$

**Box I.**

where $\Phi(\theta) = (\partial/\partial\theta')\varphi(\theta)$. In what follows, we assume that there exists a matrix $\Psi$ of full row rank such that

$$\omega_n\Phi(\theta_0) = \left[\Psi + o(1)\right]\nu_n'$$

for some sequence of nonsingular matrices $(\omega_n)$.

Under the null hypothesis (11), we have

$$\tau_n = [\nu_n'(\hat\theta_n - \theta_0)]'\Psi'\left[-\Psi H_n(\theta_0)^{-1}\Psi'\right]^{-1} \times \Psi[\nu_n'(\hat\theta_n - \theta_0)] + o_p(1), \tag{13}$$

and we may readily obtain the limit distribution of the statistic $\tau_n$ in (12) from Lemma 2 and Theorem 3. In particular, in case that the covariate $(x_t)$ is exogenous and the two limit Brownian motions $U$ and $V$ are independent, we have

$$\tau_n \to_d \frac{K^4 - 1}{2}\chi_b^2.$$

Consequently, $\tau_n$ is asymptotically distributed as chi-square with degrees of freedom given by the number of restrictions in this case, if the innovation $(\varepsilon_t)$ of our volatility model is normal and $K^4 = 3$. The standard inference based on $\tau_n$ is therefore valid.[11] It is straightforward to see that all our results here are also applicable for other standard asymptotic tests. In the paper, we mainly consider the two-sided version of $\tau_n$ for each of individual parameters and the one-sided version of $\tau_n$ jointly for all model parameters, which we simply call the $t$ and $F$-statistics, respectively.

Now we provide some simple examples that will be helpful in understanding our asymptotic results.

**Example 1.** For $f(x,\pi) = \pi x^2$, we have $\kappa(\sqrt{n}) = \sqrt{n}$ and $\bar f(x,\pi) = \pi x^2$. Moreover, it follows that $\dot f(x,\pi) = x^2$ with $\dot\kappa\left(\sqrt{n}\right) = \sqrt{n}$ and $\bar f^{\cdot}(x,\pi) = x^2$. Therefore, we have

$$\nu_n = \sqrt{n}\,\mathrm{diag}(1,1,1) \quad\text{and}\quad M(r) = \mathrm{diag}(1,1,1/\pi_0),$$

and we may easily see from Theorem 3 that the MLE $\hat\theta_n$ is consistent and has convergence rate $\sqrt{n}$. This is the simplest example, where the volatility function is *linear in parameter*. Since $M$ is an identity matrix, we may indeed readily see from our result in Theorem 3 that the asymptotic distribution of the MLE $\hat\theta_n$ is normal, and the $t$-statistics for all individual parameters have standard normal limit distribution. In particular, we do not require the exogeneity of the covariate $(x_t)$ for their asymptotic normality.

**Example 2.** For $f(x,\pi) = \pi_1|x|^{\pi_2}$, we have $\kappa\left(\sqrt{n}\right) = n^{\pi_2/2}$ and $\bar f(x,\pi) = \pi_1|x|^{\pi_2}$. Moreover, it follows that $\dot f(x,\pi) = (|x|^{\pi_2}, \pi_1|x|^{\pi_2}\ln|x|)'$ with

$$\dot\kappa\left(\sqrt{n}\right) = \begin{pmatrix} n^{\pi_2/2} & 0 \\ n^{\pi_2/2}\ln\sqrt{n} & n^{\pi_2/2}\end{pmatrix} \quad\text{and}$$

$$\bar f^{\cdot}(x,\pi) = \begin{pmatrix} |x|^{\pi_2} \\ \pi_1|x|^{\pi_2}\ln|x|\end{pmatrix}.$$

Therefore, we have

$$\nu_n = \sqrt{n}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \ln\sqrt{n} & 1\end{pmatrix} \quad\text{and}$$

$$M(r) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/\pi_{10} \\ 0 & 0 & \ln|V_c(r)|\end{pmatrix}.$$

Again, the MLE $\hat\theta_n$ is consistent. We have the standard $\sqrt{n}$ convergence rate for $\hat\alpha_n$, $\hat\beta_n$ and $\hat\pi_{2n}$. The convergence rate for $\hat\pi_{1n}$ is exceptional and given by a reduced rate $\sqrt{n}/\ln\sqrt{n}$, which follows directly from

$$\frac{\sqrt{n}}{\ln\sqrt{n}}(\hat\pi_{1n} - \pi_{10}) = -\sqrt{n}(\hat\pi_{2n} - \pi_{20}) + o_p(1).$$

The limit distribution of the MLE $\hat\theta_n$ is generally non-Gaussian and the standard inference is invalid. The limit distributions of $t$-statistics for $\alpha$, $\beta$ and $\pi_2$ are given by (13), together with Lemma 2 and Theorem 3, respectively with $\Psi = (1,0,0,0), (0,1,0,0)$ and $(0,0,0,1)$. The limit distribution of the $t$-statistic for $\pi_1$ is given by the negative of the limit distribution of the $t$-statistic for $\pi_2$, since we have

$$\frac{1}{\ln\sqrt{n}}(0,0,1,0) = \left(0,0,\frac{1}{\ln\sqrt{n}},-1\right)\nu_n' = \left[(0,0,0,-1) + o(1)\right]\nu_n'.$$

The limit distribution of the $F$-statistic can also be easily obtained from (13) with $\Psi = I$, due to Lemma 2 and Theorem 3. Clearly, the limit distributions of all the $t$-statistics and the $F$-statistic are non-Gaussian unless the covariate is exogenous and $V_c$ becomes independent of $U$.

In the presence of a persistent covariate, the constant term in the ARCH/GARCH models becomes unidentified. This is exactly

[11] If the model innovation $(\varepsilon_t)$ is nonnormal, $\tau_n$ and other standard asymptotic tests have scaled $\chi^2$ limit distribution, with the scale given by the fourth moment of the innovation.

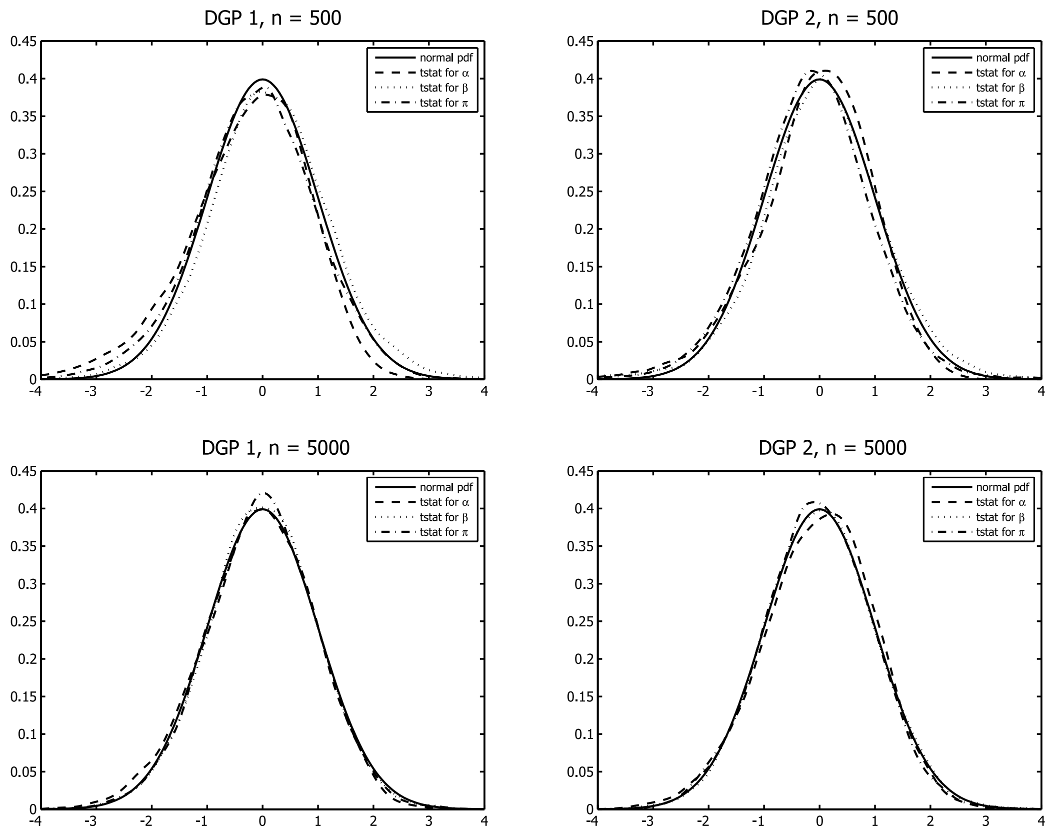


**Fig. 1.** The simulated densities of $t$-statistics for Model 1.

as in the nonstationary ARCH/GARCH models studied by Jensen and Rahbek (2004a,b). In these models, the constant term is dominated by the persistent covariate or nonstationary volatility, and becomes unimportant asymptotically.

## 4. Simulation

To investigate the relevance and usefulness of our asymptotic theory, we perform a simulation study. Our simulation study is based on two volatility functions considered respectively in Examples 1 and 2 in the previous section. More explicitly, we consider

Model 1: $f(x,\pi) = \pi x^2$
Model 2: $f(x,\pi) = \pi_1 |x|^{\pi_2}$

for the volatility model $y_t = \sigma_t \varepsilon_t$ with

$$\sigma_t^2 = \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 + f(x_t, \pi)$$

and

$$x_t = \left(1 - \frac{c}{n}\right) x_{t-1} + v_t$$

for some $c \geq 0$. The parameter values are set to be $\alpha_0 = 0.1$, $\beta_0 = 0.4$ and $c = 1$, and respectively for Models 1 and 2, $\pi = 0.1$ and $\pi = (\pi_{10}, \pi_{20})' = (0.1, 2)'$. The innovation process $(\varepsilon_t)$ is generated as iid standard normal, and we let

DGP 1: $\eta_{t+1} = 0.1\varepsilon_t$
DGP 2: $\eta_{t+1} = \frac{0.1}{\sqrt{2}}(\varepsilon_t^2 - 1)$

with $v_t = \eta_t$. The initial values are set $x_1 = 0$ and $\sigma_0^2 = 0.01$.

The covariate $(x_t)$ becomes exogenous under DGP 1, whereas it is endogenous under DGP 2. Model 1 is linear in parameter, and therefore, the MLE of its parameter has asymptotic normal distribution under both DGP's 1 and 2. In contrast, the MLE of the model parameter in Model 2 is mixed normal only under DGP 1. Under DGP 2, the MLE in Model 2 has non-Gaussian limit distribution. We thus expect that the conventional asymptotic tests are applicable in Model 1 for both DGP's 1 and 2, while in Model 2 they are valid only for DGP 1. The critical parameter in our simulation is $c$. As $c$ gets large, the covariate tends to be stationary and the standard Gaussian limit theory becomes more relevant for all cases. Our simulation results are robust with respect to all other model parameter values. In our simulation, the null distributions of the $t$-statistics for individual model parameters and the joint $F$-statistic for all model parameters are simulated for $n = 500$ and 5,000 with 10,000 iterations. The simulation results are reported in Figs. 1–3.

Figs. 1 and 2 present the simulated distributions of the $t$-statistics respectively for Models 1 and 2. Our simulation results are largely consistent with the asymptotic theory developed in the paper. The Gaussian limit distribution theory for Model 1 is effectively demonstrated and confirmed in Fig. 1. The empirical distributions of the $t$-statistics are close to normal, and become more so as the sample size increases. This is true for both DGP's 1 and 2. Our asymptotic theory is also well reflected in the simulation results for Model 2 reported in Fig. 2. Under DGP 1, it can be clearly seen that the empirical distributions of the $t$-statistics approach to standard normal as the sample size increases. On the other hand, under DGP 2, they remain to be away from standard normal distribution even for large samples. In particular, the empirical distribution of the $t$-statistic for $\pi_2$ is quite distinctive of standard normal distribution.[12]

12 The distributions of the $t$-statistics for $\pi_1$ and $\pi_2$ have the leading terms, one of which is the negative of the other. However, they differ by a term of order $1/\ln n$ for samples of size $n$, and it appears that their empirical distributions are quite different unless the sample size is very large.

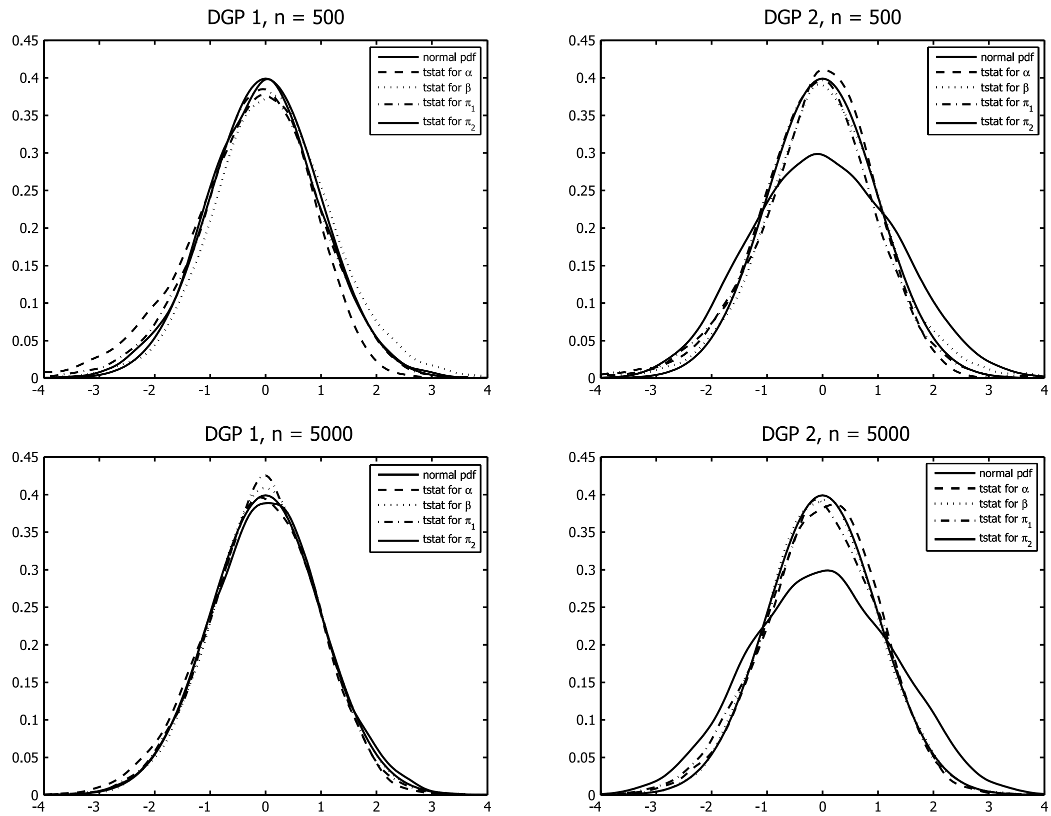


**Fig. 2.** The simulated densities of $t$-statistics for Model 2.

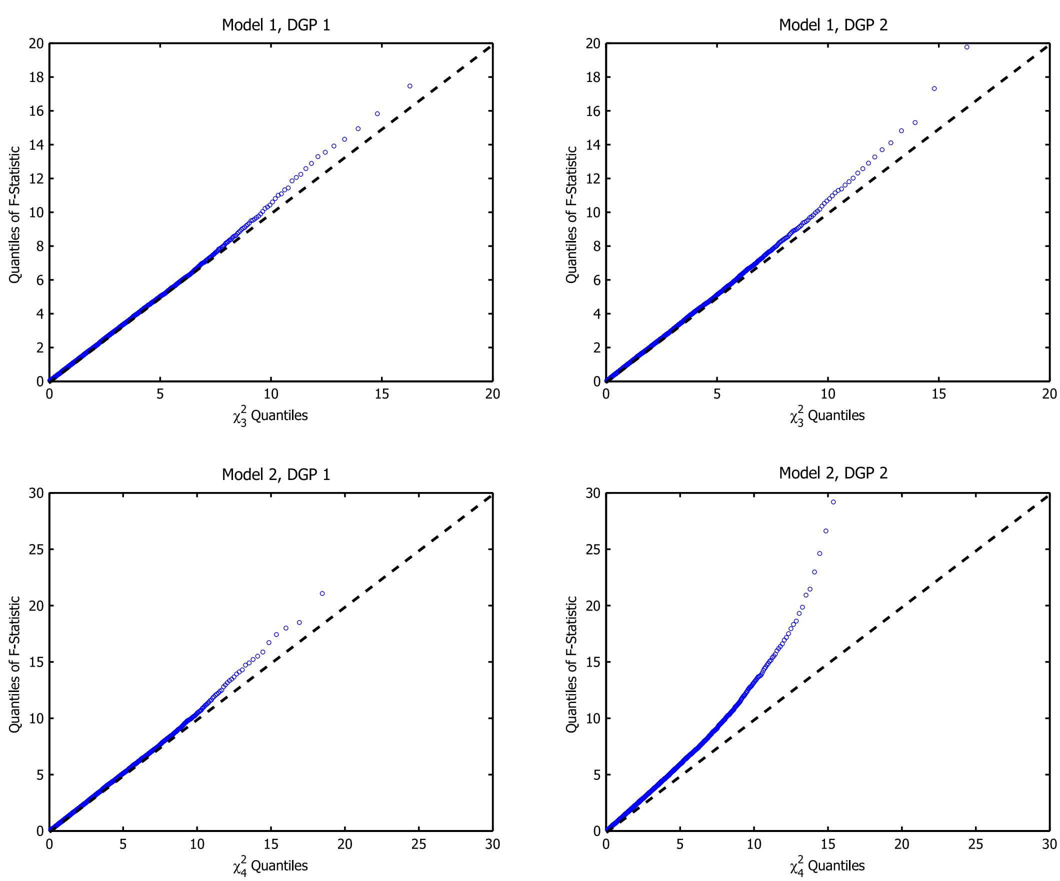


**Fig. 3.** The QQ-plots of $F$-statistics.

To contrast two cases yielding different types of limit distributions more effectively, we simulate and compare the $F$-statistics for Models 1 and 2 under DGP's 1 and 2. Our results are presented in Fig. 3. Fig. 3 contrasts the empirical quantiles of the $F$-statistics with the quantiles of chi-square distributions for samples of size $n = 5,000$. For Model 1, the limit distribution of the $F$-statistic is chi-square with three degrees of freedom under both DGP's 1 and 2. For Model 2, the limit distribution of the $F$-statistic is chi-square with four degrees of freedom under only DGP 1, and it becomes nonstandard and depends upon nuisance parameters under DGP 2. Our simulation results in Fig. 3 are consistent with the asymptotic theory. The $F$-statistics have distributions close to chi-square under both DGP's 1 and 2 for Model 1, whereas their distributions are noticeably better approximated by chi-square under DGP 1 than DGP 2 for Model 2.

## 5. Conclusion

In this paper, we developed the asymptotic theory for the MLE of the GARCH(1, 1) model with an integrated or near-integrated covariate. For a wide class of volatility functions, the MLE's of the parameters in the model are consistent. The MLE's of the GARCH parameters have the usual $\sqrt{n}$ convergence rate, whereas the MLE's of the covariate parameters have the convergence rates depending upon the asymptotic orders of the volatility function. For both parameters, the limit distributions are generally non-Gaussian and the standard testing procedures are not valid. However, there are cases in which they have Gaussian limit distributions. If the covariate has innovation uncorrelated with the squared innovation of the model, then the limit distributions for the MLE's of all parameters in the model reduce to mixed normal. Moreover, their limit distributions become normal even in the presence of nonzero correlation between the covariate innovation and squared model innovation, if the volatility function is linear in parameter. Consequently, the standard tests based on the MLE are generally applicable in these cases.

## Appendix A. Useful lemmas and their proofs

The proofs of the theorems in the paper rely on the results from the following lemmas. Here we write $f_0(x) = f(x, \pi_0)$ for notational simplicity.

**Lemma A.** *Under Assumptions 1–3, we have*

$$\kappa_0\left(\sqrt{n}\right)^{-1}\max_{1\le t\le n}\left|\sigma_t^2(\theta_0) - z_t f_0(x_t)\right| = o_p(1)$$

*for all large n.*

**Proof of Lemma A.** Throughout the proof, we let $(\tau_n)$ be a sequence of numbers such that

$$\tau_n = n^r$$

with $0 < r < 1/4 - 1/2p - 1/q$, where $p$ and $q$ are introduced in Assumption 2. Note in particular that

$$\tau_n \to \infty \quad \text{and} \quad \tau_n^2 n^{-1/2+1/p+2/q} = n^{2r-1/2+1/p+2/q} \to 0. \tag{14}$$

Moreover, $K$ denotes a generic constant whose precise definition varies from a line to another.

By the recursive substitution, it can be easily deduced that

$$\sigma_t^2(\theta_0) = f_0(x_t) + \sum_{i=1}^{\infty} f_0(x_{t-i}) \prod_{j=1}^{i}\left(\beta_0 + \alpha_0\varepsilon_{t-j}^2\right).$$

Therefore, we may write

$$\sigma_t^2(\theta_0) = z_t f_0(x_t) + e_t$$

with

$$e_t = e_t(A) + e_t(B) + e_t(C),$$

where

$$e_t(A) = \sum_{i=1}^{\tau_n}[f_0(x_{t-i}) - f_0(x_t)] \prod_{j=1}^{i}\left(\beta_0 + \alpha_0\varepsilon_{t-j}^2\right), \tag{15}$$

$$e_t(B) = \sum_{i=\tau_n+1}^{\infty} f_0(x_{t-i}) \prod_{j=1}^{i}\left(\beta_0 + \alpha_0\varepsilon_{t-j}^2\right), \tag{16}$$

$$e_t(C) = -\sum_{i=\tau_n+1}^{\infty} f_0(x_t) \prod_{j=1}^{i}\left(\beta_0 + \alpha_0\varepsilon_{t-j}^2\right). \tag{17}$$

First, we analyze $e_t(A)$ defined in (15). Note that

$$\max_{1\le t\le n}|v_t| = O_p(n^{1/p}), \tag{18}$$

since for any constant $K > 0$ we have

$$\begin{aligned}\mathbb{P}\left\{\max_{1\le t\le n} n^{-1/p}|v_t| > K\right\} &\le \sum_{t=1}^{n}\mathbb{P}\left\{n^{-1/p}|v_t| > K\right\} \\ &= n\mathbb{P}\left\{n^{-1/p}|v_t| > K\right\} \\ &\le K^{-p}\mathbb{E}|v_t|^p.\end{aligned}$$

It follows from part (b) of Assumption 3 that

$$\begin{aligned}\kappa_0\left(\sqrt{n}\right)^{-1}|f_0(x_t) - f_0(x_{t-i})| \le &\left[\left(\frac{\kappa_0\left(\sqrt{n}\right)}{\sqrt{n}}\right)^{-1}|\nabla f_0(x_t^*)|\right] \\ &\times\left[n^{-1/2}|v_t + \cdots + v_{t-i+1}|\right],\end{aligned} \tag{19}$$

where

$$\left|x_t^* - x_t\right| \le |v_t + \cdots + v_{t-i+1}|.$$

However, we have

$$\left(\frac{\kappa_0\left(\sqrt{n}\right)}{\sqrt{n}}\right)^{-1}\max_{1\le t\le n}|\nabla f_0(x_t)| = O_p(1)$$

and, for all $i \le \tau_n$, we have

$$\begin{aligned}\max_{1\le t\le n} n^{-1/2}|v_t + \cdots + v_{t-i+1}| &\le \tau_n n^{-1/2}\max_{1\le t\le n}|v_t| \\ &= O_p(\tau_n n^{-1/2+1/p}),\end{aligned}$$

due to (18). Consequently, it follows from (19) that

$$\kappa_0\left(\sqrt{n}\right)^{-1}\max_{1\le t\le n}|f_0(x_t) - f_0(x_{t-i})| = O_p(\tau_n n^{-1/2+1/p}). \tag{20}$$

We may further deduce that

$$\max_{1\le t\le n}\sum_{i=1}^{\tau_n}\prod_{j=1}^{i}\left(\beta_0 + \alpha_0\varepsilon_{t-j}^2\right) = O_p(\tau_n n^{2/q}). \tag{21}$$

For this, we note that

$$\begin{aligned}&\mathbb{P}\left\{\max_{1\le t\le n}\tau_n^{-1}n^{-2/q}\sum_{i=1}^{\tau_n}\prod_{j=1}^{i}\left(\beta_0 + \alpha_0\varepsilon_{t-j}^2\right) > K\right\} \\ &\quad\le \sum_{t=1}^{n}\mathbb{P}\left\{\sum_{i=1}^{\tau_n}\prod_{j=1}^{i}\left(\beta_0 + \alpha_0\varepsilon_{t-j}^2\right) > K\tau_n n^{2/q}\right\},\end{aligned}$$

and that

$$\mathbb{P}\left\{\sum_{i=1}^{\tau_n}\prod_{j=1}^{i}\left(\beta_0+\alpha_0\varepsilon_{t-j}^2\right) > K\tau_n n^{2/q}\right\}$$
$$\leq \sum_{i=1}^{\tau_n}\mathbb{P}\left\{\prod_{j=1}^{i}\left(\beta_0+\alpha_0\varepsilon_{t-j}^2\right) > Kn^{2/q}\right\}$$
$$\leq \frac{K^{-q/2}}{n}\sum_{i=1}^{\tau_n}\left(\mathbb{E}\left(\beta_0+\alpha_0\varepsilon_t^2\right)^{q/2}\right)^i$$
$$\leq \frac{K^{-q/2}}{n\left(1-\mathbb{E}\left(\beta_0+\alpha_0\varepsilon_t^2\right)^{q/2}\right)},$$

which holds for all $1 \leq t \leq n$. Therefore, we have

$$\mathbb{P}\left\{\max_{1\leq t\leq n}\tau_n^{-1}n^{-2/q}\sum_{i=1}^{\tau_n}\prod_{j=1}^{i}\left(\beta_0+\alpha_0\varepsilon_{t-j}^2\right) > K\right\}$$
$$\leq \frac{K^{-q/2}}{1-\mathbb{E}\left(\beta_0+\alpha_0\varepsilon_t^2\right)^{q/2}},$$

as required to establish (21). Now it follows from (20) and (21) that

$$\kappa_0\left(\sqrt{n}\right)^{-1}\max_{1\leq t\leq n}|e_t(A)|$$
$$\leq \left(\kappa_0\left(\sqrt{n}\right)^{-1}\max_{1\leq t\leq n}\max_{1\leq i\leq\tau_n}|f_0(x_t)-f_0(x_{t-i})|\right)$$
$$\times\left(\max_{1\leq t\leq n}\sum_{i=1}^{\tau_n}\prod_{j=1}^{i}\left(\beta_0+\alpha_0\varepsilon_{t-j}^2\right)\right)$$
$$= O_p(\tau_n^2 n^{-1/2+1/p+2/q}) = o_p(1), \tag{22}$$

due in particular to (14), and we may readily conclude that $e_t(A)$ is negligible.

Second, we look at $e_t(B)$ and $e_t(C)$ introduced in (16) and (17) respectively. To analyze the terms, we establish

$$\max_{1\leq t\leq n}\sum_{i=\tau_n+1}^{\infty}\prod_{j=1}^{i}\left(\beta_0+\alpha_0\varepsilon_{t-j}^2\right) = o_p(1). \tag{23}$$

For this, let $\delta > 0$ be given arbitrarily and note that

$$\mathbb{P}\left\{\max_{1\leq t\leq n}\sum_{i=\tau_n+1}^{\infty}\prod_{j=1}^{i}\left(\beta_0+\alpha_0\varepsilon_{t-j}^2\right) > \delta\right\}$$
$$\leq \sum_{t=1}^{n}\mathbb{P}\left\{\sum_{i=\tau_n+1}^{\infty}\prod_{j=1}^{i}\left(\beta_0+\alpha_0\varepsilon_{t-j}^2\right) > \delta\right\},$$

and that

$$\mathbb{P}\left\{\sum_{i=\tau_n+1}^{\infty}\prod_{j=1}^{i}\left(\beta_0+\alpha_0\varepsilon_{t-j}^2\right) > \delta\right\}$$
$$\leq \sum_{i=\tau_n+1}^{\infty}\mathbb{P}\left\{\prod_{j=1}^{i}\left(\beta_0+\alpha_0\varepsilon_{t-j}^2\right) > \delta n^{-1}\right\}$$
$$\leq \delta^{-q/2}n^{q/2}\sum_{i=\tau_n+1}^{\infty}\left(\mathbb{E}\left(\beta_0+\alpha_0\varepsilon_t^2\right)^{q/2}\right)^i$$
$$\leq \delta^{-q/2}n^{q/2}\frac{\left(\mathbb{E}\left(\beta_0+\alpha_0\varepsilon_t^2\right)^{q/2}\right)^{\tau_n+1}}{1-\mathbb{E}\left(\beta_0+\alpha_0\varepsilon_t^2\right)^{q/2}},$$

which holds for all $1 \leq t \leq n$. Therefore, we have

$$\mathbb{P}\left\{\max_{1\leq t\leq n}\sum_{i=\tau_n+1}^{\infty}\prod_{j=1}^{i}\left(\beta_0+\alpha_0\varepsilon_{t-j}^2\right) > \delta\right\}$$
$$\leq \frac{\delta^{-q/2}n^{1+q/2}\left(\mathbb{E}\left(\beta_0+\alpha_0\varepsilon_t^2\right)^{q/2}\right)^{\tau_n+1}}{1-\mathbb{E}\left(\beta_0+\alpha_0\varepsilon_t^2\right)^{q/2}} \to 0,$$

due in particular to (14) and that $|\mathbb{E}\left(\beta_0+\alpha_0\varepsilon_t^2\right)^{q/2}| < 1$. This proves (23). We may now easily obtain

$$\kappa_0\left(\sqrt{n}\right)^{-1}\max_{1\leq t\leq n}|e_t(B)| = \kappa_0\left(\sqrt{n}\right)^{-1}\max_{1\leq t\leq n}|e_t(C)| = o_p(1), \tag{24}$$

due to (23) and the fact that

$$\kappa_0\left(\sqrt{n}\right)^{-1}\max_{1\leq t\leq n}|f_0(x_t)| = O_p(1)$$

for all large $n$. The stated result now follows immediately from (22) and (24). □

**Lemma B.** *Under Assumptions 1–3,*

$$\kappa_0\left(\sqrt{n}\right)^{-1}\max_{1\leq t\leq n}|z_tf_0(x_t)-z_tf_0(x_{t-1})| = o_p(1)$$

*for all large n.*

**Proof of Lemma B.** Similarly as (20) in the proof of Lemma A, we may deduce that

$$\kappa_0\left(\sqrt{n}\right)^{-1}\max_{1\leq t\leq n}|f_0(x_t)-f_0(x_{t-1})| = O_p(n^{-1/2+1/p}). \tag{25}$$

Moreover, if we let $(\tau_n)$ be the sequence of numbers given as (14) in the proof of Lemma A, it follows that

$$\max_{1\leq t\leq n}|z_t| \leq 1+\max_{1\leq t\leq n}\sum_{i=1}^{\tau_n}\prod_{j=1}^{i}\left(\beta_0+\alpha_0\varepsilon_{t-j}^2\right)$$
$$+\max_{1\leq t\leq n}\sum_{i=\tau_n+1}^{\infty}\prod_{j=1}^{i}\left(\beta_0+\alpha_0\varepsilon_{t-j}^2\right)$$
$$= O_p(\tau_n n^{2/q}) + o_p(1), \tag{26}$$

due to (21) and (23) in the proof of Lemma A.

Now it follows from (25) and (26) that

$$\kappa_0\left(\sqrt{n}\right)^{-1}\max_{1\leq t\leq n}|z_tf_0(x_t)-z_tf_0(x_{t-1})|$$
$$\leq \left(\max_{1\leq t\leq n}|z_t|\right)\left(\kappa_0\left(\sqrt{n}\right)^{-1}\max_{1\leq t\leq n}|f_0(x_t)-f_0(x_{t-1})|\right)$$
$$= O_p(\tau_n n^{-1/2+1/p+2/q}) + o_p(n^{-1/2+1/p}) = o_p(1)$$

as was to be shown. The proof is therefore complete. □

## Appendix B. Proofs of the main results

**Proof of Lemma 1.** We may write $(v_t)$ as

$$v_t = \phi(1)\eta_t + (\tilde{v}_{t-1}-\tilde{v}_t)$$

with

$$\tilde{v}_t = \sum_{k=0}^{\infty}\tilde{\phi}_k\eta_{t-k}, \qquad \tilde{\phi}_k = \sum_{i=k+1}^{\infty}\phi_i,$$

using the so-called Beveridge–Nelson decomposition, for the details of which we refer to Phillips and Solo (1992). Under our condition in Assumption 2(a), $(\tilde{v}_t)$ is well defined as a linear process generated by a sequence of iid random variables $(\eta_t)$ such that $\mathbb{E}|\tilde{v}_t|^p < \infty$ for some $p > 2$. In what follows, we let

$$v_t^* = \phi(1)\eta_t,$$

and let

$$w_t^* = (u_t', v_t^*)'$$

conformably with the definition of $(w_t)$.

If we define

$$W_n^*(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} w_t^*,$$

then it follows that

$$\sup_{0\le r\le 1} \left\| W_n(r) - W_n^*(r) \right\| \le \frac{1}{\sqrt{n}} \left( |\tilde{v}_0| + \max_{1\le t\le n} |\tilde{v}_t| \right) = o_p(1),$$

as shown in Phillips and Solo (1992). Therefore, it suffices to show that

$$W_n^* \to_d W, \tag{27}$$

where we use "$\to_d$" to denote the weak convergence in the product space of $C[0, 1]$, i.e., the space of continuous functions on $[0, 1]$ endowed with the usual uniform metric.

To establish the invariance principle in (27), we note that $(w_t^*)$ is a martingale difference sequence, which is strictly stationary. Furthermore, we have

$$\mathbb{E}\|w_t\|^2 < \infty, \tag{28}$$

as will be shown below. The invariance principle in (27) therefore follows immediately from Theorem 4.1 of Hall and Heyde (1980).

To complete the proof, we only need to prove (28). We have

$$\left| \frac{1}{z_t} \right| \le 1 \quad \text{a.s.} \tag{29}$$

Moreover, it follows from He and Teräsvirta (1999) that

$$\mathbb{E}z_t^2 < \infty \tag{30}$$

under our Assumption 2(b). Note that we may deduce from Jensen's inequality

$$\mathbb{E}\left(\beta_0 + \alpha_0\varepsilon_t^2\right)^2 \le \left( \mathbb{E}\left(\beta_0 + \alpha_0\varepsilon_t^2\right)^{q/2} \right)^{4/q} < 1$$

for $q \ge 4$.

Since

$$\begin{aligned} \mathbb{E}u_{1t}^2 &= (K^4 - 1)\mathbb{E}\left[ \frac{1}{z_t^2} \left( \sum_{i=1}^{\infty} \beta_0^{i-1} z_{t-i}\varepsilon_{t-i}^2 \right)^2 \right] \\ &\le (K^4 - 1)\mathbb{E}\left[ \left( \sum_{i=1}^{\infty} \beta_0^{i-1} z_{t-i}\varepsilon_{t-i}^2 \right)^2 \right], \\ \mathbb{E}u_{2t}^2 &= (K^4 - 1)\mathbb{E}\left[ \frac{1}{z_t^2} \left( \sum_{i=1}^{\infty} \beta_0^{i-1} z_{t-i} \right)^2 \right] \\ &\le (K^4 - 1)\mathbb{E}\left[ \left( \sum_{i=1}^{\infty} \beta_0^{i-1} z_{t-i} \right)^2 \right] \end{aligned}$$

and

$$\mathbb{E}u_{3t}^2 \le (K^4 - 1)\sum_{i=0}^{\infty} \beta_0^i < \infty$$

due to (29), it suffices to show $\mathbb{E}\left(\sum_{i=1}^{\infty} \beta_0^{i-1} z_{t-i}\varepsilon_{t-i}^2\right)^2 < \infty$.

Note that

$$\begin{aligned} \mathbb{E}\left( \sum_{i=1}^{\infty} \beta_0^{i-1} z_{t-i}\varepsilon_{t-i}^2 \right)^2 &= \sum_{i=1}^{\infty} \beta_0^{2(i-1)} \mathbb{E}\left( z_{t-i}^2 \varepsilon_{t-i}^4 \right) \\ &\quad + 2\sum_{j=1}^{\infty}\sum_{i=1}^{\infty} \beta_0^{2(j-1)+i} \mathbb{E}\left( z_{t-j} z_{t-j-i} \varepsilon_{t-j}^2 \varepsilon_{t-j-i}^2 \right) \end{aligned}$$

and

$$\begin{aligned} \sum_{i=1}^{\infty} \beta_0^{2(i-1)} \mathbb{E}\left( z_{t-i}^2 \varepsilon_{t-i}^4 \right) &= \sum_{i=1}^{\infty} \beta_0^{2(i-1)} K^4 \mathbb{E}\left( z_{t-i}^2 \right) \\ &= K^4 \mathbb{E}\left( z_t^2 \right) / (1 - \beta_0^2). \end{aligned}$$

Since

$$z_{t-j} = z_{t-j-i} \prod_{k=1}^{i} \left( \beta_0 + \alpha_0 \varepsilon_{t-j-k}^2 \right) + \sum_{l=0}^{i-1} \prod_{k=1}^{l} \left( \beta_0 + \alpha_0 \varepsilon_{t-j-k}^2 \right),$$

where we set $\prod_{k=1}^{0} \left( \beta_0 + \alpha_0 \varepsilon_{t-j-k}^2 \right) = 1$ by convention, we have

$$\begin{aligned} \mathbb{E}\left( z_{t-j} z_{t-j-i} \varepsilon_{t-j}^2 \varepsilon_{t-j-i}^2 \right) &= \mathbb{E}\left( z_{t-j} z_{t-j-i} \varepsilon_{t-j-i}^2 \right) \\ &= \mathbb{E}\left( z_{t-j-i}^2 \right) \mathbb{E}\left( \prod_{k=1}^{i} \left( \beta_0 + \alpha_0 \varepsilon_{t-j-k}^2 \right) \varepsilon_{t-j-i}^2 \right) \\ &\quad + \mathbb{E}\left( z_{t-j-i} \right) \mathbb{E}\left( \sum_{l=0}^{i-1} \prod_{k=1}^{l} \left( \beta_0 + \alpha_0 \varepsilon_{t-j-k}^2 \right) \varepsilon_{t-j-i}^2 \right) \\ &= \mathbb{E}\left( z_{t-j-i}^2 \right) (\alpha_0 + \beta_0)^{i-1} \left( \beta_0 + \alpha_0 K^4 \right) \\ &\quad + \mathbb{E}\left( z_{t-j-i} \right) \frac{1 - (\alpha_0 + \beta_0)^i}{1 - (\alpha_0 + \beta_0)} \\ &< \mathbb{E}\left( z_t^2 \right) \left( \beta_0 + \alpha_0 K^4 \right) + \mathbb{E}\left( z_t \right) \frac{1}{1 - (\alpha_0 + \beta_0)}. \end{aligned}$$

If we let

$$\upsilon = \mathbb{E}\left( z_t^2 \right) \left( \beta_0 + \alpha_0 K^4 \right) + \mathbb{E}\left( z_t \right) \frac{1}{1 - (\alpha_0 + \beta_0)},$$

we have $\upsilon < \infty$ due to (30). Therefore,

$$\begin{aligned} \mathbb{E}\left( \sum_{i=1}^{\infty} \beta_0^{i-1} z_{t-i}\varepsilon_{t-i}^2 \right)^2 &< K^4 \mathbb{E}\left( z_t^2 \right) \frac{1}{1 - \beta_0^2} \\ &\quad + 2\upsilon \frac{\beta_0}{(1 - \beta_0)\left(1 - \beta_0^2\right)} < \infty. \end{aligned}$$

This completes the proof. □

**Proof of Lemma 2.** The proof will be done in four steps. For notational simplicity, we write $f(x) = f(x, \pi)$ and $f_0(x) = f(x, \pi_0)$, and $\sigma_t^2 = \sigma_t^2(\theta)$ and $\sigma_{0t}^2 = \sigma_t^2(\theta_0)$. Also, we let $\kappa_n = \kappa_0\left(\sqrt{n}\right)$, $\dot{\kappa}_n = \dot{\kappa}_0\left(\sqrt{n}\right)$ and $\ddot{\kappa}_n = \ddot{\kappa}_0\left(\sqrt{n}\right)$. We denote by $\ddot{F} = \partial^2 f / \partial\pi\partial\pi'$ the second derivative of $f$ in matrix form.

*First step*. Since

$$\begin{aligned} \sigma_t^2(\theta) &= \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2 + f(x_t, \pi) \\ &= \alpha \sum_{i=1}^{\infty} \beta^{i-1} y_{t-i}^2 + \sum_{i=0}^{\infty} \beta^i f(x_{t-i}, \pi), \end{aligned}$$

we have

$$\begin{aligned} \frac{\partial\sigma_t^2}{\partial\theta} &= \left( \frac{\partial\sigma_t^2}{\partial\alpha}, \frac{\partial\sigma_t^2}{\partial\beta}, \frac{\partial\sigma_t^2}{\partial\pi'} \right)' \\ &= \left( \sum_{i=1}^{\infty} \beta^{i-1} y_{t-i}^2, \sum_{i=1}^{\infty} \beta^{i-1} \sigma_{t-i}^2, \sum_{i=0}^{\infty} \beta^i \dot{f}(x_{t-i})' \right)' \end{aligned}$$

and the score function is given by

$$s_n(\theta) = \frac{1}{2}\sum_{t=1}^{n}\left(\frac{y_t^2}{\sigma_t^2}-1\right)\frac{1}{\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta} = \frac{1}{2}\sum_{t=1}^{n}\left(\frac{y_t^2}{\sigma_t^2}-1\right)\frac{1}{\sigma_t^2} \times\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2,\sum_{i=1}^{\infty}\beta^{i-1}\sigma_{t-i}^2,\sum_{i=0}^{\infty}\beta^{i}\dot f(x_{t-i})'\right)'. \quad (31)$$

We have

$$\frac{y_t^2}{\sigma_{0t}^2} = \varepsilon_t^2. \quad (32)$$

Moreover, it follows from Lemmas A and B that

$$\kappa_n^{-1}\sigma_{0t}^2 = z_t\left(\kappa_n^{-1}f_0(x_t)\right)+o_p(1) = z_t\left(\kappa_n^{-1}f_0(x_{t-1})\right)+o_p(1)$$

uniformly in $t = 1,\ldots,n$, which in turn yields

$$\frac{y_{t-i}^2}{\sigma_{0t}^2} = \frac{\kappa_n^{-1}\sigma_{0,t-i}^2}{\kappa_n^{-1}\sigma_{0t}^2}\varepsilon_{t-i}^2 = \frac{z_{t-i}\varepsilon_{t-i}^2}{z_t}+o_p(1), \quad (33)$$

$$\left(\frac{\dot\kappa_n}{\kappa_n}\right)^{-1}\frac{\dot f_0(x_{t-i})}{\sigma_{0t}^2} = \frac{\dot\kappa_n^{-1}\dot f_0(x_{t-i})}{\kappa_n^{-1}f_0(x_{t-i})}\frac{1}{z_t}+o_p(1) \quad (34)$$

uniformly in $t = 1,\ldots,n$.

Now we let

$$s_n(\theta_0) = (s_{1n}(\theta_0), s_{2n}(\theta_0), s_{3n}'(\theta_0))'.$$

It follows from Lemma 1 and (31)–(34) that

$$\frac{1}{\sqrt n}s_{1n}(\theta_0) = \frac{1}{2\sqrt n}\sum_{t=1}^{n}\sum_{i=1}^{\infty}\beta_0^{i-1}\frac{z_{t-i}\varepsilon_{t-i}^2}{z_t}(\varepsilon_t^2-1)+o_p(1) = \frac{1}{2\sqrt n}\sum_{t=1}^{n}u_{1t}+o_p(1)\to_d\frac{1}{2}U_1(1) \quad (35)$$

and

$$\frac{1}{\sqrt n}s_{2n}(\theta_0) = \frac{1}{2\sqrt n}\sum_{t=1}^{n}\sum_{i=1}^{\infty}\beta_0^{i-1}\frac{z_{t-i}}{z_t}(\varepsilon_t^2-1)+o_p(1) = \frac{1}{2\sqrt n}\sum_{t=1}^{n}u_{2t}+o_p(1)\to_d\frac{1}{2}U_2(1), \quad (36)$$

where $U_2$ is the limit Brownian motion introduced in Lemma 1, where we let $W = (U', V)'$ with $U = (U_1, U_2, U_3)'$. Furthermore, we may also deduce from Lemma 1 and (31)–(34) that

$$\begin{aligned}\frac{1}{\sqrt n}\left(\frac{\dot\kappa_n}{\kappa_n}\right)^{-1}s_{3n}(\theta_0) &= \frac{1}{2\sqrt n}\sum_{t=1}^{n}\sum_{i=0}^{\infty}\beta_0^{i}\frac{\dot\kappa_n^{-1}\dot f_0(x_{t-i})}{\kappa_n^{-1}f_0(x_t)}\frac{1}{z_t}(\varepsilon_t^2-1)+o_p(1)\\ &= \sum_{i=0}^{\infty}\beta_0^{i}\frac{1}{2\sqrt n}\sum_{t=1}^{n}\frac{\dot\kappa_n^{-1}\dot f_0(x_{t-i})}{\kappa_n^{-1}f_0(x_{t-i})}\frac{1}{z_t}(\varepsilon_t^2-1)+o_p(1)\\ &= \sum_{i=0}^{\infty}\beta_0^{i}\frac{1}{2\sqrt n}\sum_{t=1}^{n}\frac{\dot\kappa_n^{-1}\dot f_0(x_{t-i})}{\kappa_n^{-1}f_0(x_{t-i})}(1-\beta_0)u_{3t}+o_p(1)\\ &\to_d \sum_{i=0}^{\infty}\beta_0^{i}\frac{1}{2}\int_0^1\frac{\dot{\bar f}_0(V_c(r))}{\bar f_0(V_c(r))}d(1-\beta_0)U_3(r)\\ &= \frac{1}{2}\int_0^1\frac{\dot{\bar f}_0(V_c(r))}{\bar f_0(V_c(r))}dU_3(r). \end{aligned}\quad (37)$$

Consequently, it follows from (35)–(37) that

$$\nu_n^{-1}s_n(\theta_0)\to_d\frac{1}{2}\int_0^1 M(r)dU(r),$$

which establishes ML1.

*Second step*. The Hessian function is given by (See the equation in Box II)

Dividing $H_n(\theta)$ into two parts, we let

$$H_n^a(\theta) = \frac{1}{2}\sum_{t=1}^{n}\left[\left(1-\frac{2y_t^2}{\sigma_t^2}\right)\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}\right] = \begin{pmatrix}H_{11}^n(\theta) & H_{12}^n(\theta) & H_{13}^n(\theta)\\ H_{21}^n(\theta) & H_{22}^n(\theta) & H_{23}^n(\theta)\\ H_{31}^n(\theta) & H_{32}^n(\theta) & H_{33}^n(\theta)\end{pmatrix}$$

and

$$H_n^b(\theta) = \frac{1}{2}\sum_{t=1}^{n}\left[\left(\frac{y_t^2}{\sigma_t^2}-1\right)\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta'}\right]$$

in what follows. In this step, we consider only $H_n^a(\theta_0)$ and $H_n^b(\theta_0)$ will be analyzed in the fourth step.

It follows from (32)–(34) that

$$\begin{aligned}-\frac{H_{11}^n(\theta_0)}{n} &= \frac{1}{2n}\sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right)\frac{1}{\left(\kappa_n^{-1}\sigma_{0t}^2\right)^2}\times\left(\sum_{i=1}^{\infty}\beta_0^{i-1}\kappa_n^{-1}\sigma_{0,t-i}^2\varepsilon_{t-i}^2\right)^2\\ &= \frac{1}{2n}\sum_{t=1}^{n}\frac{1}{(z_t)^2}\left(\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\varepsilon_{t-i}^2\right)^2+o_p(1)\\ &\to_p\frac{1}{2}\mathbb{E}\left[\frac{1}{(z_t)^2}\left(\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\varepsilon_{t-i}^2\right)^2\right],\end{aligned}\quad (38)$$

$$\begin{aligned}-\frac{H_{21}^n(\theta_0)}{n} &= \frac{1}{2n}\sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right)\frac{1}{\left(\kappa_n^{-1}\sigma_{0t}^2\right)^2}\times\left(\sum_{i=1}^{\infty}\beta_0^{i-1}\kappa_n^{-1}\sigma_{0,t-i}^2\varepsilon_{t-i}^2\right)\left(\sum_{i=1}^{\infty}\beta_0^{i-1}\kappa_n^{-1}\sigma_{0,t-i}^2\right)\\ &= \frac{1}{2n}\sum_{t=1}^{n}\frac{1}{(z_t)^2}\left(\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\varepsilon_{t-i}^2\right)\left(\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\right)+o_p(1)\\ &\to_p\frac{1}{2}\mathbb{E}\left[\frac{1}{(z_t)^2}\left(\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\varepsilon_{t-i}^2\right)\left(\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\right)\right]\end{aligned}\quad (39)$$

and

$$\begin{aligned}-\frac{H_{22}^n(\theta_0)}{n} &= \frac{1}{2n}\sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right)\frac{1}{\left(\kappa_n^{-1}\sigma_{0t}^2\right)^2}\times\left(\sum_{i=1}^{\infty}\beta_0^{i-1}\kappa_n^{-1}\sigma_{0,t-i}^2\right)^2\\ &= \frac{1}{2n}\sum_{t=1}^{n}\frac{1}{(z_t)^2}\left(\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\right)^2+o_p(1)\\ &\to_p\frac{1}{2}\mathbb{E}\left[\frac{1}{(z_t)^2}\left(\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\right)^2\right].\end{aligned}\quad (40)$$

$$H_n(\theta) = \frac{1}{2}\sum_{t=1}^{n}\left[\left(1-\frac{2y_t^2}{\sigma_t^2}\right)\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'} + \left(\frac{y_t^2}{\sigma_t^2}-1\right)\frac{1}{\sigma_t^2}\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta'}\right],$$

where

$$\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'} = \begin{pmatrix} \left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2\right)^2 & \sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2\sum_{i=1}^{\infty}\beta^{i-1}\sigma_{t-i}^2 & \sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2\sum_{i=0}^{\infty}\beta^{i}\dot{f}(x_{t-i})' \\ \sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2\sum_{i=1}^{\infty}\beta^{i-1}\sigma_{t-i}^2 & \left(\sum_{i=1}^{\infty}\beta^{i-1}\sigma_{t-i}^2\right)^2 & \sum_{i=1}^{\infty}\beta^{i-1}\sigma_{t-i}^2\sum_{i=0}^{\infty}\beta^{i}\dot{f}(x_{t-i})' \\ \sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2\sum_{i=0}^{\infty}\beta^{i}\dot{f}(x_{t-i}) & \sum_{i=1}^{\infty}\beta^{i-1}\sigma_{t-i}^2\sum_{i=0}^{\infty}\beta^{i}\dot{f}(x_{t-i}) & \sum_{i=0}^{\infty}\beta^{i}\dot{f}(x_{t-i})\sum_{i=0}^{\infty}\beta^{i}\dot{f}(x_{t-i})' \end{pmatrix}$$

and

$$\frac{\partial^2\sigma_t^2}{\partial\theta\partial\theta'} = \begin{pmatrix} 0 & \sum_{i=1}^{\infty}(i-1)\beta^{i-2}y_{t-i}^2 & 0 \\ \sum_{i=1}^{\infty}\beta^{i-1}\sum_{k=1}^{\infty}\beta^{k-1}y_{t-i-k}^2 & \sum_{i=1}^{\infty}\begin{pmatrix}(i-1)\beta^{i-2}\sigma_{t-i}^2 \\ +\beta^{i-1}\sum_{k=1}^{\infty}\beta^{k-1}\sigma_{t-i-k}^2\end{pmatrix} & \sum_{i=1}^{\infty}\beta^{i-1}\sum_{k=0}^{\infty}\beta^{k}\dot{f}(x_{t-i-k})' \\ 0 & \sum_{i=0}^{\infty}i\beta^{i-1}\dot{f}(x_{t-i}) & \sum_{i=0}^{\infty}\beta^{i}\ddot{F}(x_{t-i}) \end{pmatrix}.$$

**Box II.**

Moreover, we have from Lemma 1 and (32)–(34)

$$\begin{aligned}
-\frac{\dot{\kappa}_n^{-1}H_{31}^n(\theta_0)}{n\kappa_n^{-1}} &= \frac{1}{2n}\sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right) \\
&\quad\times\frac{\sum_{i=1}^{\infty}\beta_0^{i-1}\kappa_n^{-1}\sigma_{0,t-i}^2\varepsilon_{t-i}^2}{\kappa_n^{-1}\sigma_{0,t}^2}\frac{\sum_{i=0}^{\infty}\beta_0^{i}\dot{\kappa}_n^{-1}\dot{f}_0(x_{t-i})}{\kappa_n^{-1}\left(z_tf_0(x_t)\right)} + o_p(1) \\
&= \sum_{i=0}^{\infty}\beta_0^i\frac{1}{2n}\sum_{t=1}^{n}\frac{\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\varepsilon_{t-i}^2}{(z_t)^2}\frac{\dot{\kappa}_n^{-1}\dot{f}_0(x_{t-i})}{\kappa_n^{-1}f_0(x_{t-i})} + o_p(1) \\
&\to_d \frac{1}{2}\mathbb{E}\left[\frac{1}{(1-\beta_0)(z_t)^2}\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\varepsilon_{t-i}^2\right] \\
&\quad\times\int_0^1\frac{\bar{f}_0^{\cdot}(V_c(r))}{\bar{f}_0(V_c(r))}dr
\end{aligned}\tag{41}$$

and

$$\begin{aligned}
-\frac{\dot{\kappa}_n^{-1}H_{32}^n(\theta_0)}{n\kappa_n^{-1}} &= \frac{1}{2n}\sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right) \\
&\quad\times\frac{\sum_{i=1}^{\infty}\beta_0^{i-1}\kappa_n^{-1}\sigma_{0,t-i}^2}{\kappa_n^{-1}\sigma_{0,t}^2}\frac{\sum_{i=0}^{\infty}\beta_0^{i}\dot{\kappa}_n^{-1}\dot{f}_0(x_{t-i})}{\kappa_n^{-1}\left(z_tf_0(x_t)\right)} + o_p(1) \\
&= \sum_{i=0}^{\infty}\beta_0^i\frac{1}{2n}\sum_{t=1}^{n}\frac{\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}}{(z_t)^2}\frac{\dot{\kappa}_n^{-1}\dot{f}_0(x_{t-i})}{\kappa_n^{-1}f_0(x_{t-i})} + o_p(1) \\
&\to_d \frac{1}{2}\mathbb{E}\left[\frac{1}{(1-\beta_0)(z_t)^2}\sum_{i=1}^{\infty}\beta_0^{i-1}z_{t-i}\right]\int_0^1\frac{\bar{f}_0^{\cdot}(V_c(r))}{\bar{f}_0(V_c(r))}dr.
\end{aligned}\tag{42}$$

Finally, for $H_{33}^n(\theta_0)$, we have from Lemma 1 and (32)–(34)

$$\begin{aligned}
-\frac{\dot{\kappa}_n^{-1}H_{33}^n(\theta_0)\dot{\kappa}_n^{-1'}}{n\kappa_n^{-2}} &= \frac{\kappa_n^2}{2n}\sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right) \\
&\quad\times\frac{\dot{\kappa}_n^{-1}\sum_{i=0}^{\infty}\beta_0^i\dot{f}_0(x_{t-i})}{\sigma_{0t}^2}\frac{\sum_{i=0}^{\infty}\beta_0^i\dot{f}_0(x_{t-i})'\dot{\kappa}_n^{-1'}}{\sigma_{0t}^2} \\
&= \frac{1}{2n}\sum_{t=1}^{n}\frac{1}{(z_t)^2}\left(\sum_{i=0}^{\infty}\beta_0^i\frac{\dot{\kappa}_n^{-1}\dot{f}_0(x_{t-i})}{\kappa_n^{-1}f_0(x_{t-i})}\right) \\
&\quad\times\left(\sum_{i=0}^{\infty}\beta_0^i\frac{\dot{\kappa}_n^{-1}\dot{f}_0(x_{t-i})}{\kappa_n^{-1}f_0(x_{t-i})}\right)' + o_p(1) \\
&\to_d \frac{1}{2}\mathbb{E}\left[\frac{1}{(1-\beta_0)^2(z_t)^2}\right]\int_0^1\frac{\bar{f}_0^{\cdot}(V_c(r))\bar{f}_0^{\cdot}(V_c(r))'}{\bar{f}_0^2(V_c(r))}dr.
\end{aligned}\tag{43}$$

Note that we use the result of Han and Park (2008, Lemma 3(b)) for the last line.

Consequently, it follows from (38)–(43) that

$$-\nu_n^{-1}H_n^a(\theta_0)\nu_n^{-1'}\to_d\frac{1}{2}\int_0^1 M(r)\Omega M(r)'dr,$$

as was to be shown to establish ML2.

*Third step*. To establish ML3, fix $\delta$ such that $0<\delta<\varepsilon/6$, and define $\mu_n=\nu_n^{1-\delta}$ so that $\mu_n\nu_n^{-1}\to 0$ as required. Moreover, let $\bar{s}=\max(s_{\max},-s_{\min})+1$ as in Park and Phillips (2001, Proof of Theorem 5.3). It follows that

$$\mu_n = \begin{pmatrix} n^{1/2-\delta} & 0 & 0 \\ 0 & n^{1/2-\delta} & 0 \\ 0 & 0 & n^{1/2-\delta}\kappa_n^{-1}\dot{\kappa}_n \end{pmatrix},$$

and therefore, we have

$$\left\|n^{1/2-\delta}\left(\alpha-\alpha_0\right)\right\| \leq 1, \tag{44}$$

$$\left\|n^{1/2-\delta}\left(\beta-\beta_0\right)\right\| \leq 1, \tag{45}$$

$$\left\|n^{1/2-\delta}(\dot{\kappa}_n/\kappa_n)'\left(\pi-\pi_0\right)\right\| \leq 1 \tag{46}$$

for all $\theta \in N_n$. We will show in this step that

$$\left\|\mu_n^{-1}\left(H_n^a(\theta)-H_n^a(\theta_0)\right)\mu_n^{-1'}\right\| = o_p(1),$$

and $\left\|\mu_n^{-1}\left(H_n^b(\theta)-H_n^b(\theta_0)\right)\mu_n^{-1'}\right\| = o_p(1)$ will be proven in the next step.

Equation given in Box III, it suffices to show that

$$\varpi_{1n}^2(\theta) = \left\|\frac{1}{n^{1-2\delta}}\left(H_{11}^n(\theta)-H_{11}^n(\theta_0)\right)\right\| \to_p 0, \tag{47}$$

$$\varpi_{2n}^2(\theta) = \left\|\frac{1}{n^{1-2\delta}}\left(H_{21}^n(\theta)-H_{21}^n(\theta_0)\right)\right\| \to_p 0, \tag{48}$$

$$\varpi_{3n}^2(\theta) = \left\|\frac{1}{n^{1-2\delta}}\left(H_{22}^n(\theta)-H_{22}^n(\theta_0)\right)\right\| \to_p 0, \tag{49}$$

$$\varpi_{4n}^2(\theta) = \left\|\frac{1}{n^{1-2\delta}}(\dot{\kappa}_n/\kappa_n)^{-1}\left(H_{31}^n(\theta)-H_{31}^n(\theta_0)\right)\right\| \to_p 0, \tag{50}$$

$$\varpi_{5n}^2(\theta) = \left\|\frac{1}{n^{1-2\delta}}(\dot{\kappa}_n/\kappa_n)^{-1}\left(H_{32}^n(\theta)-H_{32}^n(\theta_0)\right)\right\| \to_p 0, \tag{51}$$

$$\varpi_{6n}^2(\theta) = \left\|\frac{1}{n^{1-2\delta}}(\dot{\kappa}_n/\kappa_n)^{-1}\left(H_{33}^n(\theta)-H_{33}^n(\theta_0)\right)(\dot{\kappa}_n/\kappa_n)^{-1'}\right\| \to_p 0 \tag{52}$$

uniformly for all $\theta \in N_n$.

To derive (47), note that

$$\begin{aligned}
&\frac{1}{\sigma_t^4}\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2\right)^2 - \frac{1}{\sigma_{0t}^4}\left(\sum_{i=1}^{\infty}\beta_0^{i-1}y_{t-i}^2\right)^2 \\
&= \left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2\right)^2\left(\frac{1}{\sigma_t^4}-\frac{1}{\sigma_{0t}^4}\right) \\
&\quad + \frac{1}{\sigma_{0t}^4}\left\{\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2\right)^2 - \left(\sum_{i=1}^{\infty}\beta_0^{i-1}y_{t-i}^2\right)^2\right\}.
\end{aligned} \tag{53}$$

For the first term in (53), we have

$$\begin{aligned}
&\sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right)\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2\right)^2\frac{\left(\sigma_{0t}^4-\sigma_t^4\right)}{\sigma_t^4\sigma_{0t}^4} \\
&= \sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right)\frac{\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2\right)^2}{\sigma_{0t}^2}\frac{\sigma_{0t}^2-\sigma_t^2}{\sigma_t^4} \\
&\quad + \sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right)\frac{\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2\right)^2}{\sigma_{0t}^4}\frac{\sigma_{0t}^2-\sigma_t^2}{\sigma_t^2}.
\end{aligned}$$

By the definition of $\sigma_t^2$, we have

$$\begin{aligned}
\sigma_{0t}^2-\sigma_t^2 &= \left(\alpha_0-\alpha\right)y_{t-1}^2 + \left(\beta_0-\beta\right)\sigma_{0t-1}^2 \\
&\quad + \left(f_0(x_t)-f(x_t)\right) + \beta\left(\sigma_{0t-1}^2-\sigma_{t-1}^2\right) \\
&= \left(\alpha_0-\alpha\right)\sum_{k=1}^{\infty}\beta^{k-1}y_{t-k}^2 + \left(\beta_0-\beta\right)\sum_{k=1}^{\infty}\beta^{k-1}\sigma_{0t-k}^2 \\
&\quad + \sum_{k=1}^{\infty}\beta^{k-1}\left(f_0(x_{t-k+1})-f(x_{t-k+1})\right).
\end{aligned} \tag{54}$$

Using (54), we can divide $\sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right)\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2\right)^2\left(1/\sigma_{0t}^2\right)\left(\sigma_{0t}^2-\sigma_t^2\right)/\sigma_t^4$ into three terms. For the first term, we have

$$\begin{aligned}
&\left\|\frac{1}{n^{1-2\delta}}\sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right)\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2\right)^2 \times \frac{1}{\sigma_{0t}^2}\frac{1}{\sigma_t^4}\left(\alpha_0-\alpha\right)\sum_{k=1}^{\infty}\beta^{k-1}y_{t-k}^2\right\| \\
&\leq \left\{\frac{n^{3\delta}}{\sqrt{n}}\left\|\kappa_n^2\sup_{|s|\leq\bar{s}}\sup_{\theta\in N_n}\left|\frac{1}{f^2\left(\sqrt{n}s,\theta\right)}\right|\right\|\left\|\frac{1}{n}\sum_{t=1}^{n}\right|\left(\kappa_n^2\right)^{-1}\left(2\varepsilon_t^2-1\right)\right. \\
&\quad \left.\times\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2\right)^2\frac{1}{\sigma_{0t}^2}\sum_{k=1}^{\infty}\beta^{k-1}y_{t-k}^2\right|\Bigg\},
\end{aligned} \tag{55}$$

due to (44) and

$$0 < \frac{1}{\sigma_t^2} \leq \frac{1}{f(x_t)}. \tag{56}$$

Note that

$$\frac{1}{n}\sum_{t=1}^{n}\left|\left(\kappa_n^2\right)^{-1}\left(2\varepsilon_t^2-1\right)\frac{y_{t-1}^4}{\sigma_{0t}^2}\sum_{k=1}^{\infty}\beta^{k-1}y_{t-k}^2\right| = O_p(1)$$

due to Han and Park (2008, Lemma 4). Similarly, for the second term, we have

$$\begin{aligned}
&\left\|\frac{1}{n^{1-2\delta}}\sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right)\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2\right)^2 \times \frac{1}{\sigma_{0t}^2}\frac{1}{\sigma_t^4}\left(\beta_0-\beta\right)\sum_{k=1}^{\infty}\beta^{k-1}\sigma_{0t-k}^2\right\| \\
&\leq \left\{\frac{n^{3\delta}}{\sqrt{n}}\left\|\kappa_n^2\sup_{|s|\leq\bar{s}}\sup_{\theta\in N_n}\left|\frac{1}{f^2\left(\sqrt{n}s,\theta\right)}\right|\right\|\left\|\frac{1}{n}\sum_{t=1}^{n}\right|\left(\kappa_n^2\right)^{-1}\right. \\
&\quad \left.\times\left(2\varepsilon_t^2-1\right)\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^2\right)^2\frac{1}{\sigma_{0t}^2}\sum_{k=1}^{\infty}\beta^{k-1}\sigma_{0t-k}^2\right|\Bigg\}
\end{aligned} \tag{57}$$

due to (45) and (56). Let $\bar{\pi}$ lie in the line segment connecting $\pi$ and $\pi_0$. Then, for the third term, we have

Since

$$\mu_n^{-1}H_n^a(\theta)\mu_n^{-1'} = -\begin{pmatrix} n^{-1+2\delta}H_{11}^n(\theta) & n^{-1+2\delta}H_{12}^n(\theta) & n^{-1+2\delta}H_{13}^n(\theta)\left(\frac{\dot\kappa_n}{\kappa_n}\right)^{-1'} \\ n^{-1+2\delta}H_{21}^n(\theta) & n^{-1+2\delta}H_{22}^n(\theta) & n^{-1+2\delta}H_{23}^n(\theta)\left(\frac{\dot\kappa_n}{\kappa_n}\right)^{-1'} \\ n^{-1+2\delta}\left(\frac{\dot\kappa_n}{\kappa_n}\right)^{-1}H_{31}^n(\theta) & n^{-1+2\delta}\left(\frac{\dot\kappa_n}{\kappa_n}\right)^{-1}H_{32}^n(\theta) & n^{-1+2\delta}\left(\frac{\dot\kappa_n}{\kappa_n}\right)^{-1}H_{33}^n(\theta)\left(\frac{\dot\kappa_n}{\kappa_n}\right)^{-1'} \end{pmatrix},$$

**Box III.**

$$\begin{aligned}
&\Bigg\| \frac{1}{n^{1-2\delta}}\sum_{t=1}^n (2\varepsilon_t^2-1)\left(\sum_{i=1}^\infty \beta^{i-1}y_{t-i}^2\right)^2 \frac{1}{\sigma_{0t}^2}\frac{1}{\sigma_t^4} \\
&\quad\times \sum_{k=1}^\infty \beta^{k-1}\left(f_0(x_{t-k+1})-f(x_{t-k+1})\right)\Bigg\| \\
&\le \frac{1}{n^{1-2\delta}}\sum_{t=1}^n \left|(2\varepsilon_t^2-1)\left(\sum_{i=1}^\infty \beta^{i-1}y_{t-i}^2\right)^2\frac{1}{\sigma_{0t}^2}\right| \\
&\quad\times \left\|\frac{1}{\sigma_t^4}\sum_{k=1}^\infty \beta^{k-1}\dot f(x_{t-k+1},\bar\pi)'(\pi_0-\pi)\right\| \\
&\le \Bigg\{\frac{1}{1-\beta}\frac{n^{3\delta}}{\sqrt n} \\
&\quad\times \left\|\kappa_n^2\dot\kappa_n^{-1}\sup_{|s|\le\bar s}\left(\sup_{\theta\in N_n}\left|\frac{1}{f^2(\sqrt n s,\theta)}\right|\sup_{\theta\in N_n}\left|\dot f(\sqrt n s,\theta)\right|\right)\right\| \\
&\quad\times \frac1n\sum_{t=1}^n\left|\kappa_n^{-1}(2\varepsilon_t^2-1)\left(\sum_{i=1}^\infty\beta^{i-1}y_{t-i}^2\right)^2\frac{1}{\sigma_{0t}^2}\right|\Bigg\},
\end{aligned}\tag{58}$$

due to (46) and (56). Similarly, we can obtain

$$\begin{aligned}
&\left\|\frac{1}{n^{1-2\delta}}\sum_{t=1}^n(2\varepsilon_t^2-1)\left(\sum_{i=1}^\infty\beta^{i-1}y_{t-i}^2\right)^2\frac{1}{\sigma_{0t}^4}\frac{\sigma_{0t}^2-\sigma_t^2}{\sigma_t^2}\right\| \\
&\le \frac{n^{3\delta}}{\sqrt n}\left\|\kappa_n\sup_{|s|\le\bar s}\sup_{\theta\in N_n}\left|\frac{1}{f(\sqrt n s,\theta)}\right|\right\| \\
&\quad\times\frac1n\sum_{t=1}^n\left|\kappa_n^{-1}(2\varepsilon_t^2-1)\left(\sum_{i=1}^\infty\beta^{i-1}y_{t-i}^2\right)^2\frac{1}{\sigma_{0t}^4}\sum_{k=1}^\infty\beta^{k-1}y_{t-k}^2\right| \\
&\quad+\frac{n^{3\delta}}{\sqrt n}\left\|\kappa_n\sup_{|s|\le\bar s}\sup_{\theta\in N_n}\left|\frac{1}{f(\sqrt n s,\theta)}\right|\right\| \\
&\quad\times\frac1n\sum_{t=1}^n\left|\kappa_n^{-1}(2\varepsilon_t^2-1)\left(\sum_{i=1}^\infty\beta^{i-1}y_{t-i}^2\right)^2\frac{1}{\sigma_{0t}^4}\sum_{k=1}^\infty\beta^{k-1}\sigma_{0t-k}^2\right| \\
&\quad+\Bigg\{\frac{1}{1-\beta}\frac{n^{3\delta}}{\sqrt n} \\
&\quad\times\left\|\kappa_n\dot\kappa_n^{-1}\sup_{|s|\le\bar s}\left(\sup_{\theta\in N_n}\left|\frac{1}{f(\sqrt n s,\theta)}\right|\sup_{\theta\in N_n}\left|\dot f(\sqrt n s,\theta)\right|\right)\right\| \\
&\quad\times\frac1n\sum_{t=1}^n\left|(2\varepsilon_t^2-1)\left(\sum_{i=1}^\infty\beta^{i-1}y_{t-i}^2\right)^2\frac{1}{\sigma_{0t}^4}\right|\Bigg\}.
\end{aligned}\tag{59}$$

For the second term in (53), we can easily deduce that

$$\begin{aligned}
&\Bigg\|\frac{1}{n^{1-2\delta}}\sum_{t=1}^n(2\varepsilon_t^2-1)\frac{1}{\sigma_{0t}^4} \\
&\quad\times\left\{\left(\sum_{i=1}^\infty\beta^{i-1}y_{t-i}^2\right)^2-\left(\sum_{i=1}^\infty\beta_0^{i-1}y_{t-i}^2\right)^2\right\}\Bigg\|\le\frac{n^{3\delta}}{\sqrt n}O_p(1),
\end{aligned}\tag{60}$$

because

$$\begin{aligned}
&\left\|\frac1n\sum_{t=1}^n\kappa_n^{-2}\left\{\left(\sum_{i=1}^\infty\beta^{i-1}y_{t-i}^2\right)^2-\left(\sum_{i=1}^\infty\beta_0^{i-1}y_{t-i}^2\right)^2\right\}\right\| \\
&=\Bigg\|\frac1n\sum_{t=1}^n\kappa_n^{-2}\left(\sum_{i=1}^\infty\beta^{i-1}y_{t-i}^2+\sum_{i=1}^\infty\beta_0^{i-1}y_{t-i}^2\right) \\
&\quad\times\sum_{i=1}^\infty\left(\beta^{i-1}-\beta_0^{i-1}\right)y_{t-i}^2\Bigg\| \\
&=\|\beta-\beta_0\|\,O_p(1)
\end{aligned}$$

due to Han and Park (2008, Lemma 4). From (55) and (57)–(60), (47) follows immediately due to (5). We can obtain (48) and (49) similarly as (47).

To derive (50), note that

$$\begin{aligned}
&\frac{1}{\sigma_t^4}\sum_{i=0}^\infty\beta^i\dot f(x_{t-i})\sum_{i=1}^\infty\beta^{i-1}y_{t-i}^2-\frac{1}{\sigma_{0t}^4}\sum_{i=0}^\infty\beta_0^i\dot f_0(x_{t-i})\sum_{i=1}^\infty\beta_0^{i-1}y_{t-i}^2 \\
&=\sum_{i=1}^\infty\beta^{i-1}y_{t-i}^2\left(\frac{1}{\sigma_t^4}\sum_{i=0}^\infty\beta^i\dot f(x_{t-i})-\frac{1}{\sigma_{0t}^4}\sum_{i=0}^\infty\beta_0^i\dot f_0(x_{t-i})\right) \\
&\quad+\sum_{i=1}^\infty\left(\beta^{i-1}-\beta_0^{i-1}\right)y_{t-i}^2\frac{1}{\sigma_{0t}^4}\sum_{i=0}^\infty\beta_0^i\dot f_0(x_{t-i}).
\end{aligned}\tag{61}$$

Regarding the second term in (61), we have

$$\begin{aligned}
&\Bigg\|\frac{\kappa_n}{n^{1-2\delta}}\dot\kappa_n^{-1}\sum_{t=1}^n(2\varepsilon_t^2-1)\sum_{i=1}^\infty\left(\beta^{i-1}-\beta_0^{i-1}\right)y_{t-i}^2 \\
&\quad\times\frac{1}{\sigma_{0t}^4}\sum_{i=0}^\infty\beta_0^i\dot f_0(x_{t-i})\Bigg\|\le\frac{n^{3\delta}}{\sqrt n}O_p(1),
\end{aligned}\tag{62}$$

because

$$\left\|\frac1n\sum_{t=1}^n\kappa_n^{-1}\sum_{i=0}^\infty\left(\beta^i-\beta_0^i\right)y_{t-i}^2\right\|=\|\beta-\beta_0\|\,O_p(1).$$

For the other terms in (61), note that (see the equation in Box IV).

$$\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^{2}\left(\frac{1}{\sigma_t^4}\sum_{i=0}^{\infty}\beta^{i}\dot f(x_{t-i})-\frac{1}{\sigma_{0t}^4}\sum_{i=0}^{\infty}\beta_0^{i}\dot f_0(x_{t-i})\right)$$
$$=\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^{2}\left(\begin{array}{c}\frac{1}{\sigma_t^4}\sum_{i=0}^{\infty}\beta^{i}\left(\dot f(x_{t-i})-\dot f_0(x_{t-i})\right)+\sum_{i=0}^{\infty}\beta^{i}\dot f_0(x_{t-i})\frac{\sigma_{0t}^4-\sigma_t^4}{\sigma_t^4\sigma_{0t}^4}\\ +\frac{1}{\sigma_{0t}^4}\sum_{i=0}^{\infty}\left(\beta^{i}-\beta_0^{i}\right)\dot f_0(x_{t-i})\end{array}\right). \tag{63}$$

**Box IV.**

For the first term in (63), we have

$$\begin{aligned}&\left\|\frac{\kappa_n}{n^{1-2\delta}}\dot\kappa_n^{-1}\sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right)\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^{2}\right)\frac{1}{\sigma_t^4}\sum_{i=0}^{\infty}\beta^{i}\left(\dot f(x_{t-i})-\dot f_0(x_{t-i})\right)\right\|\\
&\le\frac{1}{n^{3/2-3\delta}}\sum_{t=1}^{n}\left|\kappa_n^{-1}\left(2\varepsilon_t^2-1\right)\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^{2}\right)\right|\\
&\times\left\|\kappa_n^3\frac{1}{\sigma_t^4}\sum_{i=0}^{\infty}\beta^{i}\left(\dot\kappa_n\otimes\dot\kappa_n\right)^{-1}\ddot f(x_{t-i},\bar\pi)\right\|\\
&\le\left\{\frac{1}{1-\beta}\frac{n^{3\delta}}{\sqrt n}\right.\\
&\times\left\|\kappa_n^3\left(\dot\kappa_n\otimes\dot\kappa_n\right)^{-1}\sup_{|s|\le\bar s}\left(\sup_{\theta\in N_n}\left|\frac{1}{f^2\left(\sqrt n s,\theta\right)}\right|\sup_{\theta\in N_n}\left|\ddot f\left(\sqrt n s,\theta\right)\right|\right)\right\|\\
&\left.\times\frac{1}{n}\sum_{t=1}^{n}\left|\kappa_n^{-1}\left(2\varepsilon_t^2-1\right)\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^{2}\right)\right|\right\},\end{aligned}\tag{64}$$

due to (46). For the second term in (63), we have

$$\begin{aligned}&\left\|\frac{\kappa_n}{n^{1-2\delta}}\dot\kappa_n^{-1}\sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right)\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^{2}\right)\right.\\
&\left.\times\sum_{i=0}^{\infty}\beta^{i}\dot f_0(x_{t-i})\frac{\left(\sigma_{0t}^4-\sigma_t^4\right)}{\sigma_t^4\sigma_{0t}^4}\right\|\\
&\le\frac{1}{n^{1-2\delta}}\sum_{t=1}^{n}\left\|\kappa_n^{-1}\dot\kappa_n^{-1}\left(2\varepsilon_t^2-1\right)\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^{2}\right)\right.\\
&\left.\times\sum_{i=0}^{\infty}\beta^{i}\dot f_0(x_{t-i})\right\|\left|\kappa_n^2\frac{\left(\sigma_{0t}^4-\sigma_t^4\right)}{\sigma_t^4\sigma_{0t}^4}\right|\to_p 0\end{aligned}\tag{65}$$

similarly as (47). For the third term in (63),

$$\begin{aligned}&\left\|\frac{\kappa_n}{n^{1-2\delta}}\dot\kappa_n^{-1}\sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right)\left(\sum_{i=1}^{\infty}\beta^{i-1}y_{t-i}^{2}\right)\frac{1}{\sigma_{0t}^4}\right.\\
&\left.\times\sum_{i=0}^{\infty}\left(\beta^{i}-\beta_0^{i}\right)\dot f_0(x_{t-i})\right\|\le\frac{n^{3\delta}}{\sqrt n}O_p(1),\end{aligned}\tag{66}$$

because

$$\left\|\frac{1}{n}\sum_{t=1}^{n}\dot\kappa_n^{-1}\sum_{i=0}^{\infty}\left(\beta^{i}-\beta_0^{i}\right)\dot f_0(x_{t-i})\right\|=\|\beta-\beta_0\|\,O_p(1).$$

From (62)–(66), (50) follows immediately due to (6). We can obtain (51) similarly as (50).

To establish (52), note that

$$\begin{aligned}&\frac{1}{\sigma_t^4}\sum_{i=0}^{\infty}\beta^{i}\dot f(x_{t-i})\sum_{i=0}^{\infty}\beta^{i}\dot f(x_{t-i})'\\
&-\frac{1}{\sigma_{0t}^4}\sum_{i=0}^{\infty}\beta_0^{i}\dot f_0(x_{t-i})\sum_{i=0}^{\infty}\beta_0^{i}\dot f_0(x_{t-i})'\\
&=\frac{1}{\sigma_t^4}\sum_{i=0}^{\infty}\beta^{i}\dot f(x_{t-i})\left[\sum_{i=0}^{\infty}\beta^{i}\dot f(x_{t-i})'-\sum_{i=0}^{\infty}\beta_0^{i}\dot f_0(x_{t-i})'\right]\\
&+\frac{1}{\sigma_t^4}\left[\sum_{i=0}^{\infty}\beta^{i}\dot f(x_{t-i})-\sum_{i=0}^{\infty}\beta_0^{i}\dot f_0(x_{t-i})\right]\sum_{i=0}^{\infty}\beta_0^{i}\dot f_0(x_{t-i})'\\
&+\sum_{i=0}^{\infty}\beta_0^{i}\dot f_0(x_{t-i})\sum_{i=0}^{\infty}\beta_0^{i}\dot f_0(x_{t-i})'\frac{\left(\sigma_{0t}^4-\sigma_t^4\right)}{\sigma_t^4\sigma_{0t}^4}.\end{aligned}\tag{67}$$

For the first term in (67), we have

$$\begin{aligned}&\left\|\frac{\kappa_n^2}{n^{1-2\delta}}\dot\kappa_n^{-1}\sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right)\frac{1}{\sigma_t^4}\sum_{i=0}^{\infty}\beta^{i}\dot f(x_{t-i})\right.\\
&\left.\times\sum_{i=0}^{\infty}\beta^{i}\left(\dot f(x_{t-i})'-\dot f_0(x_{t-i})'\right)\dot\kappa_n^{-1\prime}\right\|\\
&\le\left\{\frac{1}{(1-\beta)^2}\frac{n^{3\delta}}{\sqrt n}\left\|\dot\kappa_n^{-1}\sup_{|s|\le\bar s}\sup_{\theta\in N_n}\left|\dot f\left(\sqrt n s,\theta\right)\right|\right\|\right.\\
&\times\left\|\kappa_n^3\left(\dot\kappa_n\otimes\dot\kappa_n\right)^{-1}\sup_{|s|\le\bar s}\left(\sup_{\theta\in N_n}\left|\frac{1}{f^2\left(\sqrt n s,\theta\right)}\right|\right.\right.\\
&\left.\left.\left.\times\sup_{\theta\in N_n}\left|\ddot f\left(\sqrt n s,\theta\right)\right|\right)\right\|\left\|\frac{1}{n}\sum_{t=1}^{n}\left|2\varepsilon_t^2-1\right|\right\|\right\}\end{aligned}\tag{68}$$

and

$$\begin{aligned}&\left\|\frac{\kappa_n^2}{n^{1-2\delta}}\dot\kappa_n^{-1}\sum_{t=1}^{n}\left(2\varepsilon_t^2-1\right)\frac{1}{\sigma_t^4}\sum_{i=0}^{\infty}\beta^{i}\dot f(x_{t-i})\right.\\
&\left.\times\sum_{i=0}^{\infty}\left(\beta^{i}-\beta_0^{i}\right)\dot f_0(x_{t-i})'\dot\kappa_n^{-1\prime}\right\|\\
&\le\frac{n^{3\delta}}{\sqrt n}\left\|\kappa_n^2\dot\kappa_n^{-1}\sup_{|s|\le\bar s}\left(\sup_{\theta\in N_n}\left|\frac{1}{f^2\left(\sqrt n s,\theta\right)}\right|\right.\right.\\
&\left.\left.\times\sup_{\theta\in N_n}\left|\dot f\left(\sqrt n s,\theta\right)\right|\right)\right\|O_p(1),\end{aligned}\tag{69}$$

because

$$\left\|\frac{1}{n}\sum_{t=1}^{n}\sum_{i=0}^{\infty}\left(\beta^{i}-\beta_0^{i}\right)\dot f_0(x_{t-i})'\dot\kappa_n^{-1\prime}\right\|=\|\beta-\beta_0\|\,O_p(1).$$

For the second term in (67), we have

$$\left\| \frac{\kappa_n^2}{n^{1-2\delta}} \dot{\kappa}_n^{-1} \sum_{t=1}^n \left(2\varepsilon_t^2 - 1\right) \frac{1}{\sigma_t^4} \sum_{i=0}^\infty \beta^i \times \left(\dot{f}(x_{t-i}) - \dot{f}_0(x_{t-i})\right) \sum_{i=0}^\infty \beta^i \dot{f}_0(x_{t-i})' \dot{\kappa}_n^{-1'} \right\|$$
$$\le \left\{ \frac{1}{(1-\beta)^2} \frac{n^{3\delta}}{\sqrt{n}} \left\| \kappa_n^3 \left(\dot{\kappa}_n \otimes \dot{\kappa}_n\right)^{-1} \sup_{|s|\le \bar{s}} \times \left( \sup_{\theta\in N_n} \left| \frac{1}{f^2\left(\sqrt{n}s, \theta\right)} \right| \sup_{\theta\in N_n} \left| \ddot{f}\left(\sqrt{n}s, \theta\right) \right| \right) \right\| \times \frac{1}{n} \sum_{t=1}^n \left\| \left(2\varepsilon_t^2 - 1\right) \dot{f}_0(x_t)' \dot{\kappa}_n^{-1'} \right\| \right\} \tag{70}$$

and

$$\left\| \frac{\kappa_n^2}{n^{1-2\delta}} \dot{\kappa}_n^{-1} \sum_{t=1}^n \left(2\varepsilon_t^2 - 1\right) \frac{1}{\sigma_t^4} \sum_{i=0}^\infty \left(\beta^i - \beta_0^i\right) \dot{f}_0(x_{t-i}) \times \sum_{i=0}^\infty \beta^i \dot{f}_0(x_{t-i})' \dot{\kappa}_n^{-1'} \right\|$$
$$\le \frac{n^{3\delta}}{\sqrt{n}} \left\| \kappa_n^2 \sup_{|s|\le\bar{s}} \sup_{\theta\in N_n} \left| \frac{1}{f^2\left(\sqrt{n}s, \theta\right)} \right| \right\| O_p(1). \tag{71}$$

For the third term in (67), we have

$$\left\| \frac{\kappa_n^2}{n^{1-2\delta}} \dot{\kappa}_n^{-1} \sum_{t=1}^n \left(2\varepsilon_t^2 - 1\right) \sum_{i=0}^\infty \beta^i \dot{f}_0(x_{t-i}) \times \sum_{i=0}^\infty \beta^i \dot{f}_0(x_{t-i})' \frac{\sigma_{0t}^4 - \sigma_t^4}{\sigma_t^4 \sigma_{0t}^4} \dot{\kappa}_n^{-1'} \right\|$$
$$\le \frac{1}{n^{1-2\delta}} \sum_{t=1}^n \left\| \left(2\varepsilon_t^2 - 1\right) \dot{\kappa}_n^{-1} \sum_{i=0}^\infty \beta^i \dot{f}_0(x_{t-i}) \times \sum_{i=0}^\infty \beta^i \dot{f}_0(x_{t-i})' \dot{\kappa}_n^{-1'} \right\| \left| \kappa_n^2 \frac{\sigma_{0t}^4 - \sigma_t^4}{\sigma_t^4 \sigma_{0t}^4} \right| \to_p 0 \tag{72}$$

similarly as (47). From (68)–(72), (52) follows immediately due to (6). Therefore, we may easily deduce that $\varpi_{in}^2(\theta) = o_p(1)$ for $i = 1, 2, \ldots, 6$ uniformly in $\theta \in N_n$.

*Fourth step.* As the last step to establish ML2 and ML3, we will show

$$\nu_n^{-1} H_n^b(\theta_0) \nu_n^{-1'} = o_p(1) \tag{73}$$

and

$$\left\| \mu_n^{-1} \left(H_n^b(\theta) - H_n^b(\theta_0)\right) \mu_n^{-1'} \right\| = o_p(1). \tag{74}$$

All elements of $H_n^b(\theta)$ are martingale difference sequences. Hence (73) and (74) hold due to Theorem 3.8 of Park (2003) for all elements of $H_n^b(\theta)$ except for its $3 \times 3$ element, $\sum_{t=1}^n \left(1-\varepsilon_t^2\right) \sum_{i=0}^\infty \beta^i \ddot{F}(x_{t-i})/\sigma_t^2$. Note that if $f$ is linear in parameter then $\ddot{F} = 0$ and (73) and (74) hold automatically.

Now we consider the $3 \times 3$ element of $H_n^b(\theta)$ for the case of $\ddot{F} \neq 0$. Since

$$\frac{\dot{\kappa}_n^{-1} H_{33}^n(\theta_0) \dot{\kappa}_n^{-1'}}{n\kappa_n^{-2}} = O_p(1)$$

from the second step, we can deduce that

$$\frac{\kappa_n^2}{n} \dot{\kappa}_n^{-1} \left[ \sum_{t=1}^n \left(1-\varepsilon_t^2\right) \frac{1}{\sigma_{0t}^2} \sum_{i=0}^\infty \beta^i \ddot{F}_0(x_{t-i}) \right] \dot{\kappa}_n^{-1'} = o_p(1),$$

because

$$\operatorname{vec}\left[ \frac{\kappa_n^2}{n} \dot{\kappa}_n^{-1} \sum_{t=1}^n \left(1-\varepsilon_t^2\right) \frac{1}{\sigma_{0t}^2} \sum_{i=0}^\infty \beta^i \ddot{F}_0(x_{t-i}) \dot{\kappa}_n^{-1'} \right]$$
$$= \left[ \frac{\kappa_n}{\sqrt{n}} \left(\dot{\kappa}_n^{-1} \otimes \dot{\kappa}_n^{-1}\right) \ddot{\kappa}_n \right] \sum_{i=0}^\infty \beta^i \frac{1}{\sqrt{n}} \sum_{t=1}^n (1-\varepsilon_t^2) \times \frac{\ddot{\kappa}_n^{-1} \ddot{f}_0(x_{t-i})}{\kappa_n^{-1} \left(z_t f_0(x_{t-i})\right)} = o_p(1)$$

due in particular to (4) in Assumption 4(c). This establishes (73).

To establish (74), we will show

$$\left\| \frac{\kappa_n^2}{n^{1-2\delta}} \dot{\kappa}_n^{-1} \sum_{t=1}^n \left(1-\varepsilon_t^2\right) \times \left( \frac{1}{\sigma_t^2} \sum_{i=0}^\infty \beta^i \ddot{F}(x_{t-i}) - \frac{1}{\sigma_{0t}^2} \sum_{i=0}^\infty \beta_0^i \ddot{F}_0(x_{t-i}) \right) \dot{\kappa}_n^{-1'} \right\| \to_p 0. \tag{75}$$

Note that

$$\frac{\beta^i \ddot{F}(x_{t-i})}{\sigma_t^2} - \frac{\beta_0^i \ddot{F}_0(x_{t-i})}{\sigma_{0t}^2} = \beta^i \frac{\ddot{F}(x_{t-i}) - \ddot{F}_0(x_{t-i})}{\sigma_t^2} + \beta^i \frac{\ddot{F}_0(x_{t-i}) \left(\sigma_{0t}^2 - \sigma_t^2\right)}{\sigma_t^2 \sigma_{0t}^2} + \left(\beta^i - \beta_0^i\right) \frac{\ddot{F}_0(x_{t-i})}{\sigma_{0t}^2}. \tag{76}$$

For the first term in (76), we have

$$\left\| \frac{\kappa_n^2}{n^{1-2\delta}} \dot{\kappa}_n^{-1} \sum_{t=1}^n \left(1-\varepsilon_t^2\right) \frac{1}{\sigma_t^2} \sum_{i=0}^\infty \beta^i \left(\ddot{F}(x_{t-i}) - \ddot{F}_0(x_{t-i})\right) \dot{\kappa}_n^{-1'} \right\|$$
$$\le \frac{1}{n^{1-2\delta}} \sum_{t=1}^n \left|1-\varepsilon_t^2\right| \times \left\| \kappa_n^2 \frac{1}{\sigma_t^2} \sum_{i=0}^\infty \beta^i \dot{\kappa}_n^{-1} \left(\ddot{F}(x_{t-i}) - \ddot{F}_0(x_{t-i})\right) \dot{\kappa}_n^{-1'} \right\|$$
$$\le \left\{ \frac{1}{1-\beta} \frac{n^{3\delta}}{\sqrt{n}} \left\| \kappa_n^3 \left(\dot{\kappa}_n \otimes \dot{\kappa}_n \otimes \dot{\kappa}_n\right)^{-1} \sup_{|s|\le\bar{s}} \times \left( \sup_{\theta\in N_n} \left| \frac{1}{f\left(\sqrt{n}s, \theta\right)} \right| \sup_{\theta\in N_n} \left| \dddot{f}\left(\sqrt{n}s, \theta\right) \right| \right) \right\| \times \frac{1}{n} \sum_{t=1}^n \left|1-\varepsilon_t^2\right| \right\}. \tag{77}$$

For the second term in (76), we have (see the equation in Box V)

The second line in (78) follows because of (54). For the third term in (76), we have

$$\left\| \frac{\kappa_n^2}{n^{1-2\delta}} \dot{\kappa}_n^{-1} \sum_{t=1}^n \left(1-\varepsilon_t^2\right) \sum_{i=0}^\infty \left(\beta^i - \beta_0^i\right) \frac{\ddot{F}_0(x_{t-i})}{\sigma_{0t}^2} \dot{\kappa}_n^{-1'} \right\|$$
$$\le \frac{n^{3\delta}}{\sqrt{n}} \left\| \kappa_n \left(\dot{\kappa}_n \otimes \dot{\kappa}_n\right)^{-1} \sup_{|s|\le\bar{s}} \left| \ddot{f}\left(\sqrt{n}s, \theta_0\right) \right| \right\| O_p(1). \tag{79}$$

From (77)–(79), (75) follows immediately due to (6) and (7). This completes the proof. □

$$\left\| \frac{\kappa_n^2}{n^{1-2\delta}} \dot{\kappa}_n^{-1} \sum_{t=1}^{n} \left(1-\varepsilon_t^2\right) \sum_{i=0}^{\infty} \beta^i \ddot{F}_0(x_{t-i}) \frac{\sigma_{0t}^2 - \sigma_t^2}{\sigma_t^2 \sigma_{0t}^2} \dot{\kappa}_n^{-1'} \right\| \leq \frac{1}{n^{1-2\delta}} \sum_{t=1}^{n} \left| \kappa_n \frac{\left(1-\varepsilon_t^2\right)}{\sigma_{0t}^2} \right| \left\| \sum_{i=0}^{\infty} \beta^i \kappa_n \left(\dot{\kappa}_n \otimes \dot{\kappa}_n\right)^{-1} \ddot{f}_0(x_{t-i}) \right\| \left| \frac{\sigma_{0t}^2 - \sigma_t^2}{\sigma_t^2} \right|$$
$$\leq \left\{ \frac{1}{1-\beta} \frac{n^{3\delta}}{\sqrt{n}} \left\| \kappa_n^2 \left(\dot{\kappa}_n \otimes \dot{\kappa}_n\right)^{-1} \sup_{|s|\leq \bar{s}} \left( \left( \sup_{\theta \in N_n} \left| \frac{1}{f\left(\sqrt{n}s, \theta\right)} \right| \right) \left| \ddot{f}\left(\sqrt{n}s, \theta_0\right) \right| \right) \right\| \right.$$
$$\times \left[ \frac{1}{n} \sum_{t=1}^{n} \left| \frac{\left(1-\varepsilon_t^2\right)}{\sigma_{0t}^2} \sum_{k=1}^{\infty} \beta^{k-1} y_{t-k}^2 \right| + \frac{1}{n} \sum_{t=1}^{n} \left| \frac{\left(1-\varepsilon_t^2\right)}{\sigma_{0t}^2} \sum_{k=1}^{\infty} \beta^{k-1} \sigma_{0t-k}^2 \right| \right] \right\} + \left\{ \frac{1}{1-\beta} \frac{n^{3\delta}}{\sqrt{n}} \left\| \dot{\kappa}_n^{-1} \sup_{|s|\leq \bar{s}} \sup_{\theta \in N_n} \left| \dot{f}\left(\sqrt{n}s, \theta\right) \right| \right\| \right.$$
$$\left. \times \left\| \frac{1}{1-\beta} \kappa_n^2 \left(\dot{\kappa}_n \otimes \dot{\kappa}_n\right)^{-1} \sup_{|s|\leq \bar{s}} \left( \left( \sup_{\theta \in N_n} \left| \frac{1}{f\left(\sqrt{n}s, \theta\right)} \right| \right) \left| \ddot{f}\left(\sqrt{n}s, \theta_0\right) \right| \right) \right\| \frac{1}{n} \sum_{t=1}^{n} \left| \kappa_n \frac{\left(1-\varepsilon_t^2\right)}{\sigma_{0t}^2} \right| \right\}. \tag{78}$$

**Box V.**

**Proof of Theorem 3.** Due to ML1–ML3, we have

$$\nu_n'(\hat{\theta}_n - \theta_0) = -\left[\nu_n^{-1} H_n(\theta_0) \nu_n^{-1'}\right]^{-1} \left[\nu_n^{-1} s_n(\theta_0)\right] + o_p(1),$$

and the stated result follows immediately from Lemma 2. □

## References


Berkes, L., Horváth, L., Kokoszka, P., 2003. GARCH processes: structure and estimation. Bernoulli 9, 201–207.

Berkes, L., Horváth, L., Kokoszka, P., 2005. Near integrated GARCH sequences. Annals of Applied Probability 15, 890–913.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Congressional Budget Office, 2000, A framework for projecting interest rate spreads and volatilities. CBO memorandum.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. Econometrica 50, 987–1008.

Engle, R.F., Patton, A.J., 2001. What good is a volatility model? Quantitative Finance 1, 237–245.

Fleming, J., Kirby, C., Ostdiek, B., 2008. The specification of GARCH models with stochastic covariates. Journal of Futures Markets 28, 911–934.

Francq, C., Zakoian, J.-M., 2007. Qausi-maximum likelihood estimation in GARCH processes when some of the coefficients are equal to zero. Stochastic Processes and Their Applications 117, 1265–1284.

Francq, C., Zakoian, J.-M., 2010. GARCH Models. Wiley, New York.

Giraitis, L., Robinson, P.M., 2001. Whittle estimation of ARCH models. Econometric Theory 17, 608–631.

Glosten, L.R., Jagannathan, R., Runkle, D., 1993. On the relation between the expected value and the volatility of nominal excess returns on stocks. Journal of Finance 48, 1779–1801.

Gray, S.F., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42, 27–62.

Hagiwara, M., Herce, M.A., 1999. Endogenous exchange rate volatility, trading volume and interest rate differentials in a model of portfolio selection. Review of International Economics 7, 202–218.

Hall, P., Heyde, C.C., 1980. Martingale limit theory and its application. Academic Press, New York.

Han, H., Park, J.Y., 2008. Time series properties of ARCH processes with persistent covariates. Journal of Econometrics 146, 275–292.

He, C., Teräsvirta, T., 1999. Fourth moment structure of the GARCH$(p, q)$ processes. Econometric Theory 15, 824–846.

Jensen, S.T., Rahbek, A., 2004a. Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. Econometrica 72, 641–646.

Jensen, S.T., Rahbek, A., 2004b. Asymptotic inference for nonstationary GARCH. Econometric Theory 20, 1203–1226.

Lee, S.W., Hansen, B.E., 1994. Asymptotic theory for the GARCH(1, 1) quasi-maximum likelihood estimator. Econometric Theory 10, 29–52.

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

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347–370.

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

Park, J.Y., 2002. Nonstationary nonlinear heteroskedasticity. Journal of Econometrics 110, 383–415.

Park, J.Y., 2003. Weak unit root. Mimeograph, Department of Economics, Rice University.

Park, J.Y., Phillips, P.C.B., 2001. Nonlinear regressions with integrated time series. Econometrica 69, 117–161.

Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. Annals of Statistics 20, 971–1001.

Robinson, P.M., Zaffaroni, P., 2006. Pseudo-maximum likelihood estimation of ARCH($\infty$) models. Annals of Statistics 34, 1049–1074.

Weiss, A., 1986. Asymptotic theory for ARCH models. Econometric Theory 2, 107–131.

Wooldridge, J.M., 1994. Estimation and inference for dependent processes. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, Vol. 4. Elsevier, Amsterdam, pp. 2639–2738.