*Econometric Theory*, **32**, 2016, 740–791.
doi:10.1017/S0266466614000942

# DETECTING FOR SMOOTH STRUCTURAL CHANGES IN GARCH MODELS

Bin Chen
*University of Rochester*

Yongmiao Hong
*Cornell University and Xiamen University*

Detecting and modeling structural changes in GARCH processes have attracted increasing attention in time series econometrics. In this paper, we propose a new approach to testing structural changes in GARCH models. The idea is to compare the log likelihood of a time-varying parameter GARCH model with that of a constant parameter GARCH model, where the time-varying GARCH parameters are estimated by a local quasi-maximum likelihood estimator (QMLE) and the constant GARCH parameters are estimated by a standard QMLE. The test does not require any prior information about the alternatives of structural changes. It has an asymptotic N(0,1) distribution under the null hypothesis of parameter constancy and is consistent against a vast class of smooth structural changes as well as abrupt structural breaks with possibly unknown break points. A consistent parametric bootstrap is employed to provide a reliable inference in finite samples and a simulation study highlights the merits of our test.

## 1. INTRODUCTION

Since Engle's (1982) seminal work, various ARCH and GARCH models have been used to capture volatility dynamics of macroeconomic and financial time series. Underlying all these models is the key assumption of stationarity. Given the changing pace of the underlying economic mechanism and technological progress, modeling economic processes over a long time horizon under the stationarity assumption may not be suitable. It is plausible that structural changes may occur, causing the time series to deviate from stationarity. Indeed, various economic factors may lead to structural changes in economic time series. For example, one driving force for structural changes is the "shocks" induced by

We thank the co-editor (Pentti Saikkonen), two referees, Zongwu Cai, Kajal Lahiri, Oliver Linton, Suhasini Subba Rao, and seminar participants at 2009 New York Camp Econometrics in Lake Placid for their comments. All remaining errors are solely ours. Hong acknowledges support from the National Science Foundation of China and the Cheung Kong Visiting Scholarship by the Chinese Ministry of Education and Xiamen University. Address correspondence to Bin Chen, Department of Economics, University of Rochester, Rochester, NY 14627, USA; e-mail: bchen8@mail.rochester.edu; Yongmiao Hong, Department of Economics & Department of Statistical Science, Cornell University, Ithaca, NY 14850, USA and Wang Yanan Institute for Studies in Economics (WISE) and MOE Key Laboratory in Econometrics, Xiamen University, Xiamen 361005, China; e-mail: yh20@cornell.edu.

 

institutional changes, such as the changes of exchange rate systems from a fixed exchange rate mechanism to a floating exchange rate mechanism, or the introduction of Euro. The prevalence of structural instability in macroeconomic and financial time series has been documented by numerous studies. For example, Andreou and Ghysels (2002) examine the change-point hypothesis in the volatility dynamics of international stock market indices and foreign exchange returns and find multiple breaks associated with the Asian and Russian financial crises; Mikosch and Stărică (2004) apply their goodness-of-fit test to Standard & Poor's 500 returns and detect structural changes related to shifts of the unconditional variance.

Model stability is crucial for statistical inference, forecasts, and policy recommendations drawn from the model. In particular, ignoring structural changes in macroeconomic and financial time series may easily lead to spurious persistence in volatility dynamics. Diebold (1986) and Lamoureux and Lastrapes (1990) are among the first to suggest that structural changes unaccounted for can yield spurious Integrated GARCH (IGARCH) or long memory effects. More recently, Mikosch and Stărică (2004) and Hillebrand (2005) provide some theoretical explanation for this phenomenon. The IGARCH process implies that shocks have a permanent impact on volatility and so current information remains relevant when forecasting the conditional variance at long horizons. In contrast, for a short memory volatility process, shocks to variance decay quickly over time even when structural changes exist. Moreover, model instability may affect asset allocation or lead to large errors in pricing, hedging, and managing risk. Pettenuzzo and Timmerman (2005) show that the possibility of future breaks has its largest effect at long investment horizons, but historical breaks can significantly change investment decisions even at short horizons through its effect on current parameter estimates.

Tests have been proposed to detect structural breaks in GARCH models in the literature. For example, Chu (1995) considers a supremum Lagrange multiplier (LM) test for a GARCH model. Berkes, Gombay, Horvath, and Kokoszka (2004) develop a sequential likelihood-ratio ($LR$) test for parameter constancy of a GARCH model. The test is more informative than any sequential cumulative sum (CUSUM) test performed on observed asset returns or residual transformations. It is, however, computationally intensive as it involves the calculation of quasi-likelihood scores. Kulperger and Yu (2005) derive the properties of structural break tests based on the partial sums of squared estimated standardized residuals of a GARCH model. These tests all consider one-time shift as the alternative so they may not have good power against multiple breaks.

Almost all existing change-point tests for GARCH models are constructed for abrupt changes. To our knowledge, the only exception is Amado and Teräsvirta (2008), who consider testing for a time-varying smooth transition GARCH model. Smooth changes may be more realistic because volatility usually evolves over time in a continuous manner and volatility jumps are rare. Empirical evidence shows that various economic events, such as liberalization of emerging markets, integration of world equity markets, changes in exchange rate or interest

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

rate regimes, may lead to structural changes in volatility dynamics. The changes induced by policy switch, preference changes, and technology progress usually exhibit evolutionary changes in the long term. In general, as Hansen (2001) points out, "it may seem unlikely that a structural break could be immediate and might seem more reasonable to allow a structural change to take a period of time to take effect". In particular, volatility is a measure of risk and it may take time for the market to achieve some consensus.

Recently, time-varying parameter ARCH and GARCH models have appeared as a novel tool to capture the evolutionary behavior of economic time series. For example, Amado and Teräsvirta (2008) propose both additive and multiplicative time-varying GARCH models. They introduce a smooth transition function that allows model parameters to change smoothly over time. Parametric specifications for time-varying parameters lead to more efficient estimation if the coefficient functions are correctly specified. However, economic theories usually do not suggest any concrete functional form for time-varying parameters; the choice of a functional form is somewhat arbitrary. Engle and Rangel (2008) assume that the variance of the process of interest can be decomposed into stationary and nonstationary components, where the nonstationary component is modeled using spline functions of time and the stationary component follows a GARCH process. Dahlhaus and Subba Rao (2006) and Fryzlewicz, Sapatinas, and Subba Rao (2008) study a time-varying parameter ARCH process for modeling the evolutionary behavior of volatility. The model is locally stationary in the neighborhood of each point of time but is globally nonstationary. One advantage of this evolutionary time-varying parameter ARCH model is that little restriction is imposed on the functional forms of ARCH coefficients, except for the regularity condition that they evolve over time smoothly.

Motivated by the flexibility of smooth time-varying parameter ARCH models and the popularity of GARCH models in practice, we will first generalize the smooth time-varying parameter ARCH models to a class of smooth time-varying GARCH models and derive the consistency and asymptotic normality of a local QMLE for time-varying GARCH parameters in both the interior and boundary regions of time. We then use a time-varying GARCH$(p,q)$ model as the alternative to test smooth structural changes and sudden structural breaks for a GARCH model. We emphasize that unlike the case of a stationary GARCH$(p,q)$ model, a time-varying GARCH$(p,q)$ model is not included as a special case in the time-varying ARCH$(\infty)$ class and therefore the asymptotic analysis is much more involved. Thus, while the main focus of this paper is on testing structural changes of GARCH parameters, our results on local QMLE of time-varying GARCH parameters may have its own independent interest. Moreover, we study the asymptotic properties of the local QMLE of time-varying GARCH parameters in both the interior and boundary regions of time. We find that the asymptotic biases of the local QMLE in the interior and boundary regions have different convergence rates, and a simple boundary-correction will make the bias in the boundary regions vanish to zero at the same rate as in the interior region.

As the main contribution, this paper proposes a consistent test for smooth structural changes as well as abrupt structural breaks in GARCH parameters with either known or unknown change points. The idea is to estimate the smooth time-varying GARCH parameters by a local QMLE and compare them with the standard QMLE for constant GARCH parameters. Compared with the existing tests for structural breaks in GARCH models in the literature, the proposed test has a number of appealing features:

First, the proposed test is consistent against a large class of smooth time-varying parameter alternatives. It is also consistent against multiple sudden structural breaks in GARCH models with known or unknown break points.

Second, no prior information on a structural change GARCH alternative is needed. In particular, we do not need to know whether the structural changes are smooth or abrupt, and in the cases of abrupt structural breaks, we do not need to know the dates or the number of breaks.

Third, unlike most tests for structural breaks in GARCH models in the literature, which often have nonstandard asymptotic distributions, the proposed test has a null asymptotic N(0,1) distribution. The only inputs required are the log likelihoods of QMLE and local QMLE. Any standard econometric software can carry out computational implementation easily.

Fourth, the local QMLE can capture the local behavior of time-varying GARCH parameters. Because only local information is employed in estimating parameters at each time point, the proposed test has symmetric power against structural breaks that occur in either the first or second half of the sample period. This is different from some existing tests (e.g., CUSUM tests) that have asymmetric powers against structural breaks that have same break sizes but occur at different time points.

Fifth, unlike some existing tests for structural breaks in GARCH models, no trimming procedure is required for the proposed test. Thus, the proposed test is expected to have nontrivial powers for structural changes near the boundary regions of time, provided that the sample size is large enough. Moreover, the local QMLE for the time-varying parameters can provide insight into the nature of volatility dynamics.

In Section 2, we introduce a time-varying GARCH framework and hypotheses of interest. Section 3 proposes a local QMLE for the smooth time-varying parameters in a GARCH model and establishes its consistency and asymptotic normality for both the interior and boundary regions of time. Section 4 develops a likelihood ratio test. Section 5 derives its asymptotic null distribution and investigates its asymptotic power property. In Section 6, a simulation study is conducted to examine the finite sample performance of the test via a parametric bootstrap, which is shown to be consistent. Section 7 provides concluding remarks. All mathematical proofs are collected in the Appendix. A GAUSS code to implement the proposed test is available from the authors upon request. Throughout the paper, $C$ denotes a generic bounded constant, $\|\cdot\|_d$ denotes the $l_d$-norm, and $|\cdot|_{abs}$ denotes the absolute matrix, where $(|A|_{abs})_{i,j} = \left|A_{i,j}\right|$.

## 2. TIME-VARYING GARCH MODEL AND HYPOTHESES OF INTEREST

Consider the following data generating process (DGP)

$$\begin{cases} X_t = \sqrt{h_t^0}\varepsilon_t, \\ h_t^0 = \alpha_{0t}^0 + \sum_{i=1}^p \alpha_{it}^0 h_{t-i}^0 + \sum_{j=1}^q \beta_{jt}^0 X_{t-j}^2, \\ \{\varepsilon_t\} \sim \text{i.i.d.}(0,1), \end{cases} \tag{2.1}$$

where $X_t$ is a stochastic time series process, the $\alpha_{jt}^0$ and $\beta_{jt}^0$ are possibly time-varying parameters, $t$ is the index of time, $p$ and $q$ are the orders of the GARCH process, and $\{\varepsilon_t\}$ is an $i.i.d.$ sequence of standardized innovations with mean 0 and variance 1. Let $\theta_t^0$ be the collection of parameters; namely, $\theta_t^0 = \left(\alpha_{0t}^0, \alpha_{1t}^0, \ldots, \alpha_{pt}^0, \beta_{1t}^0, \ldots, \beta_{qt}^0\right)'$, a $(p+q+1)$-dimensional vector.

The above setup nests both constant parameter GARCH and time-varying GARCH processes. For example, if $\theta_t^0$ is not changing over time, we have a constant parameter GARCH$(p,q)$ process, whose asymptotic properties have been studied by Berkes, Horvath, and Kokoszka (2003). Francq and Zakoïan (2004) derive the consistency and asymptotic normality under strict stationarity and Escanciano (2009) extends Francq and Zakoïan (2004) to GARCH models with martingale difference centered squared innovations. Lee and Hansen (1994) and Lumsdaine (1996) also establish the asymptotic theory of the QMLE for a GARCH(1,1) model when $\theta_t^0$ is a constant.

For time-varying parameter GARCH processes, one example is the single break at time $u$ in a GARCH model. Chu (1995) and Kulperger and Yu (2005) have used this model as an alternative to study parameter constancy of GARCH models. Another example of time-varying GARCH processes is the time-varying smooth transition GARCH models proposed by Amado and Teräsvirta (2008). They consider both additive and multiplicative GARCH models, where the time-varying components are included in the conventional GARCH models in different forms.

To cover a wide range of possibilities, we do not assume any parametric functional form for $\theta_t^0$. Instead, we assume that $\theta_t^0$ is an unknown smooth function of time in form of

$$\theta_t^0 = \theta_0\left(\frac{t}{T}\right),$$

where $\theta_0 : [0,1] \to \mathbb{R}^{(p+q+1)}$ is a vector-valued smooth function. The parameter $\theta_t^0$ changes over time but in an evolutionary manner. The DGP in (2.1) becomes the following time-varying GARCH process, where

$$h_t^0 = \alpha_0^0\left(\frac{t}{T}\right) + \sum_{i=1}^p \alpha_i^0\left(\frac{t}{T}\right) h_{t-i} + \sum_{j=1}^q \beta_j^0\left(\frac{t}{T}\right) X_{t-j}^2. \tag{2.2}$$

This includes time-varying ARCH$(q)$ processes (Dahlhaus and Subba Rao, 2006; Fryzlewicz et al., 2008) as a special case when $\alpha_i^0\left(\frac{t}{T}\right) = 0$ for all $t, i = 1, \ldots, p$.

In this paper, we consider GARCH models in (2.2) because GARCH models are more flexible and parsimonious than ARCH models in capturing volatility dynamics. Parsimonious GARCH models are attractive in estimating and forecasting volatilities.

The specification that $\theta_0(\cdot)$ is a function of ratio $t/T$ rather than time $t$ is a common scaling scheme in the time series literature (see, e.g., Robinson, 1989; Phillips and Hansen, 1990; Dahlhaus and Subba Rao, 2006; Cai 2007). It might first appear a bit strange because the time-varying parameter $\theta_t^0$ depends on the sample size $T$. The reason for this requirement is that a nonparametric estimator for $\theta_t^0$ will not be consistent unless the amount of data on which it depends increases, and merely increasing the sample size will not necessarily improve estimation of $\theta_t^0$ at a fixed point $t$, even if some smoothness condition is imposed on $\theta_t^0$. The amount of local information must increase suitably if the variance and bias of a nonparametric estimator of $\theta_t^0$ are to decrease suitably as the sample size $T$ increases. A convenient way to achieve this is to regard $\theta_t^0$ as the ordinates of the smooth function $\theta_0(\cdot)$ on an equally spaced grid over $[0,1]$, which becomes finer and finer as $T \to \infty$, and then consider estimation of $\theta_0(u)$ at a fixed point $u \in [0,1]$. See Robinson (1989) for more discussion in a linear regression context.

A keen interest here is whether the parameter $\theta_t^0$ is changing over time. The null hypothesis is

$$\mathbb{H}_0 : \theta_t^0 = \theta^0 \text{ for some unknown constant vector } \theta^0 \in \Theta \text{ and for all } t,$$

where $\Theta$ is a $(p+q+1)$-dimensional parameter space of $\theta_t$.

Under $\mathbb{H}_0$, the DGP in (2.1) is a standard GARCH process with constant parameter $\theta^0$. The unknown constant parameter vector $\theta^0$ could be consistently estimated by the global QMLE

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} l_t(\theta), \tag{2.3}$$

where $l_t(\theta)$ is the likelihood function; namely

$$l_t(\theta) = -\frac{1}{2}\left[\log h_t(\theta) + \frac{X_t^2}{h_t(\theta)}\right],$$

$$h_t(\theta) = \xi_0(\theta) + \sum_{j=1}^{\infty} \xi_j(\theta) X_{t-j}^2,$$

where the functions $\xi_j(\theta)$, $0 \le j < \infty$, are defined in Berkes et al. (2003).

In practice, we observe only $\{X_1, \ldots, X_T\}$ of size $T$ and the logarithm of the likelihood function in (2.3) cannot be computed from the observed data, and so

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

the estimator $\hat{\theta}$ is infeasible. Hence, we replace $l_s(\theta)$ with

$$\bar{l}_t(\theta) = -\frac{1}{2}\left[\log \bar{h}_t(\theta) + \frac{X_t^2}{\bar{h}_t(\theta)}\right], \tag{2.4}$$

where

$$\bar{h}_t(\theta) = \xi_0(\theta) + \sum_{j=1}^{s-1} \xi_j(\theta) X_{t-j}^2,$$

and the feasible global QMLE is given by

$$\bar{\theta} = \arg\max_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} \bar{l}_t(\theta). \tag{2.5}$$

Among many others, Berkes et al. (2003) establish the consistency and asymptotic normality of both $\hat{\theta}$ and $\bar{\theta}$ under $\mathbb{H}_0$, and Lee and Hansen (1994) and Lumsdaine (1996) derive the asymptotic properties of QMLE for a GARCH(1,1) model under $\mathbb{H}_0$. Heuristically, $\xi_j(\theta)$ decays exponentially so that replacing $l_s(\theta)$ with $\bar{l}_s(\theta)$ has asymptotically negligible impact and both $\hat{\theta}$ and $\bar{\theta}$ have the same asymptotic distribution.

The alternative hypothesis $\mathbb{H}_A$ is that $\mathbb{H}_0$ is false. Under $\mathbb{H}_A$, $\theta_t^0$ is time varying with an unknown functional form. Examples include GARCH models with a single break or multiple breaks with possibly unknown break points, Amado and Teräsvirta's (2008) time-varying smooth transition GARCH models, Dahlhaus and Subba Rao's (2006) time-varying ARCH($q$) models, and the more general time-varying GARCH($p, q$) models in (2.2). We allow for smooth changes and a finite number of abrupt changes under $\mathbb{H}_A$, which covers a wide range of alternatives.

All existing tests for structural changes in GARCH models in the literature consider a parametric alternative of structural changes. For example, Chu (1995) considers a supremum LM test to check parameter constancy against a single break in a GARCH model. Amado and Teräsvirta (2008) use a LM test against time-varying smooth transition GARCH alternatives. Both tests specify certain parametric alternatives, and they have best power against the assumed alternatives. However, usually no prior information about the true alternative is available for practitioners. A main objective in this paper is to develop a consistent test for $\mathbb{H}_0$ against a wide range of alternatives, using a new approach.

In a linear regression framework, Chen and Hong (2012) propose generalized Chow and generalized Hausman tests for smooth structural changes as well as abrupt structural breaks in regression models. The idea is to estimate the smooth time-varying parameters by local linear smoothing and compare it with the constant parameter OLS estimator via sums of squared residuals and fitted values respectively. These tests are not applicable to test structural changes in GARCH models, since GARCH models require different estimation methods and use

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

different criterion functions. In this paper, we shall compare a constant parameter GARCH model with a time-varying parameter GARCH model via a quasi-log-likelihood criterion. Naturally, our test can be applied to check structural changes in an ARCH model, as it is just a special case. We note that no such a test for volatility structural changes was available in the previous literature, even for a smooth time-varying parameter ARCH model, although Dahlhaus and Subba Rao (2006) point out "from a practical point of view, one could evaluate the sum of squared deviations between the kernel-QML estimator at each time point and the global QML estimator. We conjecture that the asymptotic distribution under the null hypothesis of stationarity is a chi-square". Our quasi-likelihood ratio test is rather natural and computationally simple, because log-likelihood values are the outputs of estimation. And we obtain a convenient asymptotic N(0,1) distribution after suitable centering and scaling under $\mathbb{H}_0$.

To introduce our test, below we first extend Dahlhaus and Subba Rao's (2006) results on smooth time-varying ARCH models and discuss how to estimate smooth time-varying GARCH models by a local QMLE. Asymptotic properties of the local QMLE of smooth time-varying GARCH$(p,q)$ parameters for both the interior and boundary regions of time may have independent interests since no such asymptotic results were available in the literature.

## 3. ESTIMATION OF SMOOTH TIME-VARYING GARCH PARAMETERS

Unrestricted nonstationarity may entail so much arbitrariness in the time dependent behavior of a time series process that it may be impossible to develop a meaningful asymptotic theory. When a process is changing over time smoothly, increasing the number of observations over time does not necessarily imply an increase in information. For example, one cannot expect an ensemble average to be consistently estimated by the corresponding temporal average. To avoid pathological cases arising from extreme nonstationarity, we impose some restrictions on the process to control the extent of the deviations from stationarity. A natural way of doing so is to embed a stationary structure on the process in some neighborhood of each time point. This is similar to the idea that underlies the nonparametric technique of fitting a line locally to a nonlinear curve. In this case a smoothness condition on the curve is required to validate the approach. Likewise in the present case, the imposition of local stationarity involves the use of a smoothness constraint on the evolution of the nonstationary process. A rigorous definition of local stationarity is introduced by Dahlhaus (1996a, 1996b, 1997) who imposes a smoothness condition in terms of the components in the spectral representation of the process. Heuristically, one can say that a time series process is locally stationary if the law of motion is smoothly time-varying. Thus a locally stationary process behaves like a stationary process in the neighborhood of each instant in time but has a global nonstationary behavior.

Here, the smoothness of the parameter function $\theta_0(\cdot)$ guarantees that the time-varying GARCH process in (2.2) displays a locally stationary behavior.

In order to study the asymptotic properties of $\{X_t\}$ in (2.2), we introduce a stationary GARCH process $\{\tilde{X}_t(u)\}$ that is associated with $\{X_t\}$ at the fixed point $u \in [0,1]$:

$$\begin{cases} \tilde{X}_t(u) = \sqrt{\tilde{h}_t\left(u,\theta_u^0\right)}\varepsilon_t, \\ \tilde{h}_t\left(u,\theta_u^0\right) = \alpha_0^0(u) + \sum_{i=1}^{p}\alpha_i^0(u)\tilde{h}_{t-i}(u) + \sum_{j=1}^{q}\beta_j^0(u)\tilde{X}_{t-j}^2(u), \\ \{\varepsilon_t\} \sim \text{i.i.d.}(0,1), \quad t = 1,\ldots,T, \end{cases} \tag{3.1}$$

where all coefficients depend on the fixed point $u$ but do not depend on time $t$.

It has been shown in the literature that $X_t^2$ admits a time-varying state space representation and thus can be well approximated by the stationary process $\tilde{X}_t^2(u)$ (Subba Rao, 2006). The degree of the approximation depends on the rescaling factor $T$ and the deviation $\left|\frac{t}{T}-u\right|$. This is formally stated in Lemma A.1 in the Appendix.

Let $\Theta$ be the compact set

$$\Theta = \left\{\theta = (\alpha_0,\alpha_1,\ldots,\alpha_p,\beta_1,\ldots,\beta_q)' \in \mathbb{R}^{p+q+1} : \sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j \le (1-\eta)/\mu, \textit{ for some } \eta > 0 \right.$$
$$\left. \underline{\rho} \le \min\left(\alpha_0,\alpha_1,\ldots,\alpha_p,\beta_1,\ldots,\beta_q\right) \le \max\left(\alpha_0,\alpha_1,\ldots,\alpha_p,\beta_1,\ldots,\beta_q\right) \le \bar{\rho}\right\}, \tag{3.2}$$

where $\mu = \left(E\varepsilon_t^4\right)^{1/2}$, $0 < \underline{\rho} < \bar{\rho} < \infty$, $\underline{\rho}q < \rho_0$, and $0 < \rho_0 < 1$. For each $u \in [0,1]$, we assume that $\theta_u^0$ is an interior point in $\Theta$, where $\theta_u^0 = (\alpha_0^0(u), \alpha_1^0(u), \ldots, \alpha_p^0(u), \beta_1^0(u), \ldots, \beta_q^0(u))'$. The hypothetical process in (3.1) is a stationary GARCH process at a given point $u \in [0,1]$ and thus has a unique representation (Berkes et al., 2003)

$$\tilde{h}_t\left(u,\theta_u^0\right) = \xi_0\left(\theta_u^0\right) + \sum_{j=1}^{\infty}\xi_j\left(\theta_u^0\right)\tilde{X}_{t-j}^2(u) \tag{3.3}$$

for all $t$ with probability one under certain regularity conditions. The functions $\left\{\xi_j(\theta_u^0)\right\}$ are given in Berkes et al. (2003). Under $\mathbb{H}_A$, the local QMLE to estimate $\theta_t^0$ is given by

$$\hat{\theta}_t = \arg\max_{\theta\in\Theta} L_t(\theta) = \arg\max_{\theta\in\Theta} \frac{1}{T}\sum_{s=1}^{T} k_{st} l_s(\theta), \tag{3.4}$$

where $l_s(\theta)$ and $h_s(\theta)$ are defined in (2.3), $k_{st} = \frac{1}{b}k\left(\frac{s-t}{Tb}\right)$, the kernel $k : [-1,1] \to \mathbb{R}^+$ is a prespecified symmetric bounded probability density, and $b \equiv b(T)$ is a bandwidth such that $b \to 0$ and $Tb \to \infty$ as $T \to \infty$. For notational simplicity, we have suppressed the dependence of $k_{st}$ on the sample size $T$ and the bandwidth $b$. Examples of $k(\cdot)$ include the uniform, Epanechniov, and quartic kernels. The estimator $\hat{\theta}_t$ in (3.4) is regarded as

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

an estimator of $\theta_t^0 = \left(\alpha_0^0\left(\frac{t}{T}\right), \alpha_1^0\left(\frac{t}{T}\right), \ldots, \alpha_p^0\left(\frac{t}{T}\right), \beta_1^0\left(\frac{t}{T}\right), \ldots, \beta_q^0\left(\frac{t}{T}\right)\right)'$ or $\theta_u^0 = (\alpha_0^0(u), \alpha_1^0(u), \ldots, \alpha_p^0(u), \beta_1^0(u), \ldots, \beta_q^0(u))'$ where $\left|\frac{t}{T} - u\right| < \frac{1}{T}$.

In the derivation of the asymptotic properties of $\hat{\theta}_t$, we rely on the local approximation of $X_t^2$ by the stationary process $\tilde{X}_t^2(u)$ defined in (3.1) for $t/T$ close to $u$. We define the locally weighted likelihood of $\tilde{X}_t(u)$ as

$$\tilde{L}(u,\theta) = \frac{1}{T}\sum_{s=1}^{T} k_{st}\tilde{l}_s(u,\theta), \tag{3.5}$$

where $\left|\frac{t}{T} - u\right| < \frac{1}{T}$ and

$$\tilde{l}_s(u,\theta) = -\frac{1}{2}\left[\log \tilde{h}_s(u,\theta) + \frac{\tilde{X}_s(u)^2}{\tilde{h}_s(u,\theta)}\right],$$

$$\tilde{h}_s(u,\theta) = \xi_0(\theta) + \sum_{j=1}^{\infty} \xi_j(\theta)\tilde{X}_{s-j}^2(u).$$

It is shown in the Appendix that $L_t(\theta)$ in (3.4) and $\tilde{L}(u,\theta)$ in (3.5) become arbitrarily close to each other, and both converge in probability to

$$L(u,\theta) = DE\left[\tilde{l}_0(u,\theta)\right] \tag{3.6}$$

as $T \to \infty$, $b \to 0$, $Tb \to \infty$, $\left|\frac{t}{T} - u\right| < \frac{1}{T}$, where $D = 1$ when $T - \lfloor Tb \rfloor \geq t \geq \lfloor Tb \rfloor$, where $\lfloor Tb \rfloor$ denotes the integer part of $Tb$; and $D = k_{1c} \equiv \int_{-c}^{1} k(u)du$ when $t = [cbT]$ or $T - [cbT]$, where $1 \geq c \geq 0$. It is easy to see that $L(u,\theta)$ is maximized by $\theta_u^0$. By applying the extreme estimator lemma (e.g., Amemiya, 1985, Thm. 4.1.1), we can establish the consistency of the local QMLE $\hat{\theta}_t$. Specifically, if $b \to 0$ and $Tb \to \infty$ as $T \to \infty$, we have

$$\hat{\theta}_t \to^P \theta_u^0 \quad \text{for } u \in [0,1] \quad \text{and} \quad \left|\frac{t}{T} - u\right| < \frac{1}{T}.$$

Similar to (2.3), $L_t(\theta)$ in (3.4) cannot be computed with the observed sample $\{X_t\}_{t=1}^T$, so we have to replace $L_t(\theta)$ with

$$\bar{L}_t(\theta) = \frac{1}{T}\sum_{s=1}^{T} k_{st}\bar{l}_s(\theta), \tag{3.7}$$

where $\bar{l}_s(\theta)$ and $\bar{h}_s(\theta)$ are defined in (2.4). Then we define the feasible local QMLE

$$\bar{\theta}_t = \arg\max_{\theta \in \Theta} \bar{L}_t(\theta). \tag{3.8}$$

To derive the asymptotic properties of $\bar{\theta}_t$, we impose the following regularity conditions.

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

**Assumption A.1.** *The* $\alpha_j^0(\cdot)$ *and* $\beta_j^0(\cdot)$ *are continuous on* $[0,1]$ *except for a finite number of points in* $[0,1]$. *For each continuity point* $u \in [0,1]$, *there exists some constant* $\varphi \in (0,1]$ *and some constant* $C < \infty$, *such that* $|\alpha_j^0(u) - \alpha_j^0(v)| \le C|u-v|^{\varphi}$ *and* $|\beta_j^0(u) - \beta_j^0(v)| \le C|u-v|^{\varphi}$, *where* $v \in N_{\varepsilon}(u)$, *a small neighborhood containing* $u$.

**Assumption A.2.** *Let the* $(p+q+1)$*-dimensional parameter space* $\Theta$ *be defined as (3.2). For each* $u \in [0,1]$, $\theta_u^0 \in \mathrm{Int}(\Theta)$, *where* $\theta_u^0 = (\alpha_0^0(u), \alpha_1^0(u), \ldots, \alpha_p^0(u), \beta_1^0(u), \ldots, \beta_q^0(u))'$.

**Assumption A.3.** *The polynomials* $\alpha_1^0(u)x + \alpha_2^0(u)x^2 + \cdots + \alpha_p^0(u)x^p$ and $1 - \beta_1^0(u)x - \beta_2^0(u)x^2 - \cdots - \beta_q^0(u)x^q$ *are coprime on the set of polynomials with real coefficients for some given* $u \in [0,1]$.

**Assumption A.4.** *The standardized innovation* $\{\varepsilon_t\}$ *is an i.i.d.*$(0,1)$ *sequence satisfying* $\lim_{r\to 0} r^{-\nu} P\left(\varepsilon_t^2 \le r\right) = 0$ *for some* $\nu > 0$, *and with (i)* $E|\varepsilon_t|^{4(1+\delta)} < \infty$ *for some* $\delta > 0$ *or (ii)* $E\left(\varepsilon_t^{12}\right) < \infty$.

**Assumption A.5.** *The kernel* $k : [-1,1] \to \mathbb{R}^+$ *is a symmetric bounded probability density function.*

**Assumption A.6.** *The bandwidth* $b = cT^{-\lambda}$ $(0 < c < \infty)$ *with either (i)* $0 < \lambda < 1$ *or (ii)* $1/13 < \lambda < 1$.

**Assumption A.7.** *Except for a finite number of points on* $[0,1]$, *the* $\alpha_j^0(u)$ *and* $\beta_j^0(u)$ *are three times differentiable with* $\sup_{u\in[0,1]} \left|\left(\partial^l/\partial u^l\right)\alpha_j^0(u)\right| \le C$ *for* $j = 0, \ldots, p$ *and* $\sup_{u\in[0,1]} \left|\left(\partial^l/\partial u^l\right)\beta_j^0(u)\right| \le C$ *for* $j = 1, \ldots, q$, *and* $l = 1,2,3$, *where* $C$ *is some bounded constant independent of* $j$ *and* l.

**Assumption A.8.** *Let* $\mathcal{X}_t(2,u) = [\tilde{h}_t\left(u,\theta_u^0\right), \ldots, \tilde{h}_{t-q+1}\left(u,\theta_u^0\right), \tilde{X}_{t-1}^2(u), \ldots, \tilde{X}_{t-p+1}^2(u), d\tilde{h}_t\left(u,\theta_u^0\right)/du, \ldots, d\tilde{h}_{t-q+1}\left(u,\theta_u^0\right)/du, d\tilde{X}_{t-1}^2(u)/du, \ldots, d\tilde{X}_{t-p+1}^2(u)/du]'$. *Then* $\mathcal{X}_t(2,u)$ *is* $\phi$*-irreducible.*

Assumption A.1 imposes the $\varphi$-Lipschitz continuity of parameter functions $\alpha_j^0(\cdot)$ and $\beta_j^0(\cdot)$, but we allow for a set of a finite number of points where $\alpha_j^0(\cdot)$ and $\beta_j^0(\cdot)$ are discontinuous. Assumptions A.1 and A.2 provide sufficient conditions that the stochastic process $\{X_t\}$ admits a time-varying state space representation (Subba Rao, 2006). Assumptions A.2 and A.3 guarantee that the time-varying parameter $\theta_t^0$ can be uniquely identified. These are standard assumptions imposed by Berkes et al. (2003, 2004) and Kulperger and Yu (2005), among many others, for the case of constant parameters. Assumption A.4 imposes mildly strong moment conditions, which may seem a bit restrictive for some financial applications. Assumption A.4(i) is used to prove a similar result for sums of martingale arrays as opposed to sums of martingale differences in the stationary context. Assumption A.4(ii) is imposed to derive the bias of the local QMLE.

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

Assumption A.5 implies $\int_{-1}^{1} k(u)du = 1$, $\int_{-1}^{1} uk(u)du = 0$, and $\int_{-1}^{1} u^2 k(u) du < \infty$. All examples in Section 2 satisfy this assumption. It is possible to use kernels with infinite support, such as the Gaussian kernel $k(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$ for $-\infty < u < \infty$. However, we use kernels with bounded support to simplify analysis. Assumptions A.4(ii), A.6(ii), and A.7 are used only in Theorems 2 and 3. Assumptions A.4(ii) and A.6(ii) are used to derive the closed form of the asymptotic bias, and Assumption A.7, which implies Assumption A.1, is to guarantee that the asymptotic bias and variance of the local QMLE are well-defined. Similar assumptions have been imposed in Dahlhaus and Subba Rao (2006).

As shown in Subba Rao (2006), $\mathcal{X}_t(2,u)$ can be represented as a Markov chain. The $\phi$-irreducibility in Assumption A.8 guarantees that all parts of the space can be reached by the Markov chain, no matter what the starting point is. Assumption A.8 is used to derive the strong mixing property of the stationary derivative process $\mathcal{X}_t(2,u)$, which restricts the degree of temporal dependence in $\mathcal{X}_t(2,u)$. Kristensen (2009) and Meitz and Saikkonen (2008) have derived and the primitive conditions of the $\phi$-irreducibility for a stationary GARCH(1, 1) model, but it seems nontrivial to derive the primitive conditions for a general stationary GARCH$(p,q)$ model (see Lee, 2003).

We first state the consistency of $\bar{\theta}_t$ in (3.8).

THEOREM 1. *Suppose Assumptions A.1–A.3, A.4(i), A.5, and A.6(i) hold. If* $\left|\frac{t}{T} - u\right| < \frac{1}{T}$ *where u is a continuity point in* $[0,1]$, *we have* $\bar{\theta}_t - \hat{\theta}_t \to^P 0$ *and* $\bar{\theta}_t \to^P \theta_u^0$ *as* $T \to \infty$.

Theorem 1 holds even if $\varepsilon_t$ is not $i.i.d.N(0,1)$. This generalizes Dahlhaus and Subba Rao (2006), who consider time-varying ARCH processes. We note that unlike the case of stationary GARCH$(p,q)$ models, a time-varying GARCH$(p,q)$ model is not included as a special case in the time-varying ARCH$(\infty)$ class. Furthermore, different from a time-varying ARCH process, the Volterra expansion of a time-varying GARCH process is rather tedious. Here, we rely on a stochastic recurrence relation (see, e.g., Bougerol and Picard, 1992; Subba Rao 2006) to show that a time-varying parameter GARCH process can be approximated by a local stationary GARCH process indexed by $u \in [0,1]$.

Next, we derive the asymptotic normality of $\bar{\theta}_t$.

THEOREM 2. *Suppose Assumptions A.2, A.3, A.4(ii), A.5, A.6(ii), A.7, and A.8 hold and* $\left|\frac{t}{T} - u\right| < \frac{1}{T}$, *where* $u \in [0,1]$ *is a continuity point for the coefficient functions* $\alpha_j^0(\cdot)$ *and* $\beta_j^0(\cdot)$.

(i) *If t is in the interior region of the sample period in the sense that* $\lfloor Tb \rfloor \leq t \leq T - \lfloor Tb \rfloor$, *we have*

$$\sqrt{Tb}\left(\bar{\theta}_t - \theta_u^0 - B_u\right) \to^d N\left(0, -k_2\left(\frac{\kappa}{2}+1\right) H^{-1}(u)\right),$$

*as* $T \to \infty$, *where* $k_2 = \int_{-1}^{1} k^2(x)dx$, $\kappa = E(\varepsilon_t^4) - 3$, $H(u) = E\left[\frac{\partial^2 L(\theta_u^0, u)}{\partial\theta\partial\theta'}\right]$, *and*

$$B_u = \frac{1}{2} b^2 H^{-1}(u) \frac{\partial^3 L\left(\theta_u^0, u\right)}{\partial\theta\partial u^2} \int_{-1}^{1} x^2 k(x)dx.$$

(ii) *If $t$ is in the left boundary of the sample period in the sense that $t = \lfloor cTb \rfloor$ where $0 \le c \le 1$, we have*

$$\sqrt{Tb}\left(\bar{\theta}_t - \theta_u^0 - B_{ul}\right) \to^d N\left(0, -k_{2c}k_{1c}^{-2}\left(\frac{\kappa}{2}+1\right)H^{-1}(u)\right),$$

*as* $T \to \infty$, *where* $k_{2c} = \int_{-c}^{1} k^2(x)dx$, $k_{1c} = \int_{-c}^{1} k(x)\,dx$,

$$B_{ul} = bk_{1c}^{-1} H^{-1}(u) \frac{\partial^2 L\left(\theta_u^0, u\right)}{\partial\theta\partial u} \int_{-c}^{1} xk(x)\,dx + \frac{1}{2} b^2 k_{1c}^{-1} H^{-1}(u) \frac{\partial^3 L\left(\theta_u^0, u\right)}{\partial\theta\partial u^2} \int_{-c}^{1} x^2 k(x)dx,$$

*and $\kappa$ and $H(u)$ are defined in (i).*

(iii) *If $t$ is in the right boundary of the sample period in the sense that $t = T - \lfloor cTb \rfloor$ where $0 \le c \le 1$, we have*

$$\sqrt{Tb}\left(\bar{\theta}_t - \theta_u^0 - B_{ur}\right) \to^d N\left(0, -k_{2c}k_{1c}^{-2}\left(\frac{\kappa}{2}+1\right)H^{-1}(u)\right),$$

*as* $T \to \infty$, *where*

$$B_{ur} = bk_{1c}^{-1} H^{-1}(u) \frac{\partial^2 L\left(\theta_u^0, u\right)}{\partial\theta\partial u} \int_{-1}^{c} xk(x)\,dx + \frac{1}{2} b^2 k_{1c}^{-1} H^{-1}(u) \frac{\partial^3 L\left(\theta_u^0, u\right)}{\partial\theta\partial u^2} \int_{-1}^{c} x^2 k(x)dx$$

*and $k_{1c}, k_{2c}$, $\kappa$, and $H(u)$ are defined in (i) and (ii).*

Parameter $\kappa$ is the excess kurtosis of $\varepsilon_t$, which measures the departure from the assumed normality in the fourth moment. If $E\varepsilon_t^4 = 3$ as in the case of a normally distributed $\varepsilon_t$, then the asymptotic variance of interior points can be simplified to $-k_2 H^{-1}(u)$. The quantity $H(u)$ can be viewed as the expected value of the local Hessian matrix at point $u \in [0,1]$, and $B_u$ is the asymptotic bias, which is caused by the time-varying property of GARCH parameters. From Theorem 2, we expect that for any interior continuity point $u \in [b, 1-b]$ and all $t$ with $|\frac{t}{T} - u| < \frac{1}{T}$, the asymptotic mean square error (AMSE) is given by

$$\text{AMSE}\left(\bar{\theta}_t\right) = \frac{b^4}{4}\left[\int_{-1}^{1} x^2 k(x)dx\right]^2 \left\| H^{-1}(u) \frac{\partial^3 L\left(\theta_u^0, u\right)}{\partial\theta\partial u^2}\right\|_2^2 - \frac{k_2\left(\frac{\kappa}{2}+1\right)}{Tb} trace\left[H^{-1}(u)\right].$$

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

By minimizing the AMSE($\bar{\theta}_t$), we could obtain the optimal bandwidth $b^*$, which is of the order $T^{-\frac{1}{5}}$ for all interior continuity points $u$.

We observe that $\lim_{c\to 1}\int_{-c}^{1} k^2(x)dx = k_2$, $\lim_{c\to 1}\int_{-c}^{1} k(x)dx = 1$, $\lim_{c\to 1}\int_{-c}^{1} xk(x)\,dx = 0$, and $\lim_{c\to 1}\int_{-c}^{1} x^2k(x)\,dx = \int_{-1}^{1} x^2k(x)\,dx$, and these limits are exactly the constant factors appearing in the asymptotic bias and variance for an interior continuity point $u$. Theorem 2 shows that although the local QMLE is consistent at both interior and boundary points, the asymptotic bias of the local QMLE has a slower convergence rate in the boundary regions than in the interior region. The asymptotic biases for time-varying parameters at interior and boundary points are $O\left(b^2\right)$ and $O(b)$ respectively. Therefore, the local QMLE suffers from the well-known boundary problem. The main reason is that we do not have symmetric data available in the boundary regions. On the other hand, the asymptotic variances at interior and boundary points have the same order of magnitude and the difference is only a scale factor. Previous works on time-varying ARCH models mainly focus on interior points although some bias-correction methods could be used in the boundary regions. We note that the boundary problem of time-varying linear regression models has been studied by Cai (2007) and Chen and Hong (2012).

To overcome the boundary problem of the local QMLE, we consider a reflection method, following Hall and Wehrly (1991). Specifically, we reflect the data in the boundary regions, obtaining pseudo data $X_t = X_{-t}$ for $-\lfloor Tb\rfloor \le t \le -2$ and $X_t = X_{2T-t}$ for $T+1 \le t \le T+\lfloor Tb\rfloor$. We then use the augmented data (i.e., the union of the original data and the pseudo data) to estimate $\theta_t^0$ in the boundary regions. By construction, symmetric data become available in the original boundary regions $[1, Tb]\cup[T-Tb, T]$. This method has also been described as "reflection about the boundaries" or "boundary folding" by Schuster (1985), Silverman (1986), and Cline and Hart (1991) in nonparametric density estimation and by Chen and Hong (2012) in nonparametric regression estimation. The bias reduction is achieved by considering the expectation of the bias, which is equivalent to using the boundary kernel $\left[k\left(\frac{t+s}{Tb}\right)+k\left(\frac{t-s}{Tb}\right)\right]/2$ in the boundary regions.

Our reflection method has advantages over some alternative solutions to the boundary problem. One popular solution is to simply ignore the data in the boundary regions and use only the data in the interior region. Such trimming is simple, but it may lead to the loss of a significant amount of information, even for fairly large sample sizes. For example, if $b = \left(1/\sqrt{12}\right)T^{-\frac{1}{5}}$, where $1/\sqrt{12}$ is the standard deviation of $U(0,1)$, which could be viewed as the limiting distribution of the grid points $\frac{t}{T}$, $t = 1,\ldots,T$, as $T\to\infty$, then about 23%, 19%, and 17% of the sample observations will fall into the boundary regions when $T = 100, 250$, and 500 respectively. In the context of estimating a time-varying ARCH model, Dahlhaus and Subba Rao (2007) suggest running recursive estimation concurrently and taking a linear combination of these estimators to reduce bias. However, this method is rather computationally intensive as it involves recursive estimation with two different step sizes. The reflection method we employ is simpler.

We now define the boundary-corrected local QMLE

$$\bar{\theta}_t^c = \begin{cases} \arg\max_{\theta\in\Theta} \frac{1}{T} \sum_{s=-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st}\bar{l}_s(\theta) & \text{if } t = \lfloor cTb\rfloor,\ 0 \le c \le 1, \\ \bar{\theta}_t & \text{if } \lfloor Tb\rfloor \le t \le T - \lfloor Tb\rfloor, \\ \arg\max_{\theta\in\Theta} \frac{1}{T} \sum_{s=t-\lfloor Tb\rfloor}^{T+\lfloor Tb\rfloor} k_{st}\bar{l}_s(\theta) & \text{if } t = T - \lfloor cTb\rfloor,\ 0 \le c \le 1, \end{cases} \tag{3.9}$$

where $\bar{l}_s(\theta)$ and $\bar{\theta}_t$ are defined in (3.7) and (3.8) respectively. Note that the pseudo data are only used to estimate $\theta_t^0$ in the boundary regions.

Now we derive the asymptotic distribution of $\bar{\theta}_t^c$ in the boundary regions.

THEOREM 3. *Suppose Assumptions A.2, A.3, A.4(ii), A.5, A.6(ii), A.7, and A.8 hold, and* $\left|\frac{t}{T} - u\right| < \frac{1}{T}$, *where* $u \in [0,1]$ *is a continuity point for the* $\alpha_j^0(\cdot)$ *and* $\beta_j^0(\cdot)$. *If t is in the left boundary or the right boundary in the sense that* $t = \lfloor cTb\rfloor$ *or* $T - \lfloor cTb\rfloor$, *where* $0 \le c \le 1$, *we have*

$$\sqrt{Tb}\left(\bar{\theta}_t^c - \theta_u^0 - B_u\right) \to^d N\left(0, -k_b\left(\frac{\kappa}{2}+1\right)H^{-1}(u)\right),$$

*as* $T \to \infty$, *where* $k_b = k_2 + \int_{-1}^{1} k(x)k(x+2c)\,dx$, $k_2, \kappa, H(u)$, *and* $B_u$ *are defined in Theorem 2(i).*

Theorem 3 shows that the reflection method reduces the asymptotic bias of the local QMLE at the boundary regions of time. Now, the asymptotic biases at both interior and boundary points are the same order of magnitude; namely, $O\left(b^2\right)$. That is because symmetric data are now available in the boundary regions $[1, Tb] \cup [T - Tb, T]$ and the bias term related to the first order Taylor expansion can be approximated by an integral of $uk(u)$ from $-1$ to 1, which vanishes to 0. On the other hand, the asymptotic variance of $\bar{\theta}_t^c$ at boundary points is larger than that at interior points since by construction, the pseudo data and the original data are correlated with each other. Note that the reflection method has no impact on the interior region $[Tb, T - Tb]$. Heuristically, this method can be viewed as using the boundary kernel $\left[k\left(\frac{s-t}{Tb}\right) + k\left(\frac{s+t}{Tb}\right)\right]/2$ in the boundary regions and the standard kernel $\frac{1}{b}k\left(\frac{s-t}{Tb}\right)$ in the interior region respectively.

We note that the local QMLE in the regions near any fixed discontinuity point does not suffer from the boundary problem for $T$ sufficiently large. Suppose $u_0$ is a fixed continuity point. Then for any arbitrary $t > 0$, we can always find a $\delta > 0$, such that for all $u_0 - \delta < u < u_0 + \delta$, $|\alpha_i(u) - \alpha_i(u_0)| < t$ and $|\beta_j(u) - \beta_j(u_0)| < t$ for $i = 0, \ldots, p$ and $j = 1, \ldots, q$. Since $b = b(T) \to 0$ as $T \to \infty$ and $\delta$ does not depend on the sample size $T$, $b$ will eventually become smaller than $\delta$ for all $T$ sufficiently large so that $u_0$ becomes an interior point. This is different from the boundary problem in the boundary regions, which is due to the asymmetric coverage of observations in the boundary regions.

## 4. NONPARAMETRIC TESTING

We now propose a consistent test for smooth structural changes in GARCH models, which will complement the existing tests for sudden structural breaks and avoid the difficulty associated with the possibility of multiple breaks and/or unknown break dates. We note that the facts that the convergence rate of the asymptotic bias of the local QMLE in the boundary regions is slower than that in the interior region and the asymptotic variance of the local QMLE in the boundary regions tends to be larger would complicate the form of test statistic to be constructed. The finite sample performance of the test may be affected as well. Hence, we apply the reflection method. We use the augmented data (i.e., the union of the original data and the pseudo data) to construct the boundary-corrected local QMLE $\bar{\theta}_t^c$ in (3.8). We note that under $\mathbb{H}_0$, $\theta_t^0$ is constant, so no bias exists even in the boundary regions. However, under $\mathbb{H}_A$, $\theta_t^0$ is time-varying and correcting the boundary problem is expected to help improve power. With the global QMLE $\bar{\theta}$ and the boundary-corrected local QMLE $\bar{\theta}_t^c$ at hand, we can construct a likelihood ratio test. The idea is to compare the log likelihood of the unrestricted time-varying parameter GARCH model with that of the restricted constant parameter GARCH model. Intuitively, under $\mathbb{H}_0$, two likelihoods are close to each other. Under $\mathbb{H}_A$, the nonparametric likelihood is larger than the parametric likelihood when the sample size $T$ is sufficiently large, giving the test its power against a wide range of alternatives. Let $l_U$ denote the averaged log likelihood of the (unrestricted) time-varying parameter GARCH model, that is,

$$l_U = \frac{1}{T}\sum_{t=1}^{T} \bar{l}_t\left(\bar{\theta}_t^c\right), \tag{4.1}$$

where $\bar{\theta}_t^c$ is the boundary-corrected local QMLE in (3.8). Let $l_R$ denote the averaged log likelihood of the (restricted) constant parameter GARCH model, that is,

$$l_R = \frac{1}{T}\sum_{t=1}^{T} \bar{l}_t\left(\bar{\theta}\right), \tag{4.2}$$

where $\bar{\theta}$ is the global QMLE in (2.5). It is important to note that $l_U$ and $l_R$ are averages of log likelihoods of the observed sample $\{X_t\}_{t=1}^T$, rather than the augmented sample. The pseudo data augmented by the reflection method are only used in estimating $\theta_t^0$ via $\bar{\theta}_t^c$ in the boundary regions. Hence it will not affect the asymptotic distribution of our test statistic. Intuitively, the use of the pseudo data only has impact on the boundary regions $[1, Tb] \cup [T - Tb, T]$ and its cumulative effect over the boundary regions is asymptotically negligible as $T \to \infty$. However, it improves the finite sample performance of the proposed test. Meanwhile, we define the score function

$$S_t(\theta) \equiv \frac{\partial l_t}{\partial \theta} = \frac{1}{2}\left(\varepsilon_t^2 - 1\right)\frac{\partial \ln h_t(\theta)}{\partial \theta}. \tag{4.3}$$

We note that under $\mathbb{H}_0$, $S_t(\theta^0)$ is a martingale difference sequence (MDS) no matter whether the distribution of $\varepsilon_t$ is correctly specified or not. However, with possible distributional misspecification of $\{\varepsilon_t\}$, the local QMLE remains consistent but not efficient (see Theorems 1–3). And the information matrix equality does not hold generally even at $\theta = \theta^0$. In particular,

$$E\left[S_t\left(\theta^0\right)S_t'\left(\theta^0\right)\right] = -\frac{var\left(\varepsilon_t^2\right)}{2}H_0 \equiv -\left(\frac{\kappa}{2}+1\right)H_0, \tag{4.4}$$

where $H_0 = E\left[\frac{\partial^2 l_t(\theta^0)}{\partial\theta\partial\theta'}\right]$ is the expected value of the Hessian matrix, and $\kappa = E\left(\varepsilon_t^4\right) - 3$ is the excess kurtosis of $\varepsilon_t$. It measures the departure from normality.

Our robust $LR$ test statistic for $\mathbb{H}_0$ versus $\mathbb{H}_A$ is based on the comparison of $l_U$ and $l_R$ :

$$LR = \frac{2T\sqrt{b}(l_U - l_R)/\left(\frac{\hat{\kappa}}{2}+1\right) - \hat{A}}{\sqrt{\hat{B}}}, \tag{4.5}$$

where the centering factor

$$\begin{aligned}\hat{A} = b^{-1/2}(p+q+1)&\left\{2k(0) - \frac{1}{Tb}\sum_{j=-\lfloor Tb\rfloor}^{\lfloor Tb\rfloor}\left(1-\frac{|j|}{T}\right)k^2\left(\frac{j}{Tb}\right)\right.\\ &\left. + b\left[1 - \frac{1}{Tb}\sum_{j=-\lfloor Tb\rfloor}^{\lfloor Tb\rfloor}\left(1-\frac{|j|}{T}\right)k\left(\frac{j}{Tb}\right)\int_{-1}^{1}k\left(\frac{j}{Tb}+2u\right)du\right]\right\}\\ = b^{-1/2}(p+q+1)&\left[2k(0) - \int_{-1}^{1}k^2(u)du\right][1+o(1)],\end{aligned}$$

the scaling factor

$$\begin{aligned}\hat{B} &= 4(p+q+1)\frac{1}{Tb}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\left[2k\left(\frac{j}{Tb}\right) - \int_{-1}^{1}k(u)k\left(u+\frac{j}{Tb}\right)du\right]\\ &= 4(p+q+1)\int_0^1\left[2k(v) - \int_{-1}^{1}k(u)k(u+v)du\right]^2 dv + o(1),\end{aligned}$$

and $\hat{\kappa} = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_t^4 - 3\right)$ is a consistent estimator for excess kurtosis $\kappa$. Both $\hat{A}$ and $\hat{B}$ do not depend on the DGP and are nonstochastic, so they are convenient to compute. The log likelihoods $l_U$ and $l_R$ are outputs of estimation. Many statistical programs provide values of $l_U$ and $l_R$ automatically. Hence, it is straightforward to compute the $LR$ test statistic. In fact, as will be seen below, a consistent bootstrap is even simpler: one only needs to compare the value of log-likelihood ratio $l_U - l_R$ based on the observed data with those based on bootstrap samples. There is no need to compute $\hat{A}$ and $\hat{B}$.

We note that although our focus is to test whether $\theta_t^0 = \theta^0$ for some unknown constant vector $\theta^0 \in \Theta$ and for all $t$, our approach could be extended to test whether a subset of parameters is constant; namely $\theta_{1t}^0 = \theta_1^0$ for some unknown constant vector $\theta_1^0 \in \Theta_1$ and for all $t$, where $\Theta = \Theta_1 \times \Theta_2$ and $\theta_t^0 = \left(\theta_{1t}^{0\prime}, \theta_{2t}^{0\prime}\right)'$. An example is that we are only interested in whether ARCH coefficients are constant. There are two possibilities. We now discuss the first case. If prior information restricts $\theta_{2t}^0$ to be some constant, it would be analogous to the "partial structural change" problem for the linear regression models where part of regression coefficients may not be subject to structural changes and the interest is to test the constancy of the other part of regression coefficients (see, e.g., Andrews, Lee, and Ploberger, 1996; Bai and Perron 1998). In this case, the null hypothesis is

$$\mathbb{H}_0 : \theta_t^0 = \theta^0 \text{ for some unknown constant vector } \theta^0 \in \Theta \text{ and for all } t,$$

where $\Theta$ is a parameter space of $\theta_t$, and the alternative hypothesis is

$$\mathbb{H}_A : \theta_{1t}^0 \text{ is time varying and } \theta_{2t}^0 = \theta_2^0 \text{ for some unknown constant vector } \theta_2^0 \in \Theta_2 \text{ for some } t,$$

where $\Theta = \Theta_1 \times \Theta_2$. As a subset of parameters may be time-varying, we can adopt the local profile QMLE method. Under $\mathbb{H}_A$, we first fix $\theta_{2t} = \theta_2$ and estimate $\theta_{1t}$ by

$$\bar{\theta}_{1t} = \bar{\theta}_{1t}(\theta_2) = \arg\max_{\theta_1 \in \Theta_1} \frac{1}{T} \sum_{s=1}^{T} k_{st} \bar{l}_s(\theta_1 | \theta_2), \tag{4.6}$$

where $\bar{l}_s(\theta_1|\theta_2)$ has the same functional form as $\bar{l}_s(\theta)$ defined in (2.4), but with $\theta_2$ fixed and $\theta = (\theta_1', \theta_2')'$.

Next, we obtain an estimator $\hat{\theta}_2$ of the constant component $\theta_2^0$ by substituting the local QMLE $\bar{\theta}_{1t}$ into the likelihood function; namely

$$\hat{\theta}_2 = \arg\max_{\theta_2 \in \Theta_2} \frac{1}{T} \sum_{t=1}^{T} \bar{l}_t\left(\theta_2 | \bar{\theta}_{1t}\right), \tag{4.7}$$

where $\bar{l}_s\left(\theta_2|\bar{\theta}_{1t}\right)$ has the same functional form as $\bar{l}_s(\theta)$ defined in (2.4), but with $\theta$ replaced by $\left(\bar{\theta}_{1t}', \theta_2'\right)'$ and $\bar{\theta}_{1t}$ is defined in (4.6).

Iterations between these two steps have to be employed until a certain convergence criterion is met (see, e.g., Speckman, 1988; Carroll, Fan, Gijbels, and Wand 1997; Fan and Huang 2005 for details on the profile likelihood or profile least-squares method). With proper estimators $\bar{\theta}$ and $\left(\hat{\theta}_{1t}, \hat{\theta}_2'\right)'$ at hand, where $\hat{\theta}_{1t} = \bar{\theta}_{1t}\left(\hat{\theta}_2\right)$, we can compare two models via likelihood ratio.

The second case of the partial structural change is that no restriction is imposed on $\theta_{2t}^0$. Hence the null hypothesis is

$$\mathbb{H}_0 : \theta_{1t}^0 = \theta_1^0 \text{ for some unknown constant vector } \theta_1^0 \in \Theta_1 \text{ for all } t,$$

where $\Theta = \Theta_1 \times \Theta_2$, and the alternative hypothesis is

$\mathbb{H}_A : \theta_t^0$ is time varying for some $t$.

An example of this case has been studied by Ang and Kristensen (2012) in a linear regression framework. They test whether the conditional alphas of CAPM models are constant over time while allowing the conditional betas to be time-varying. This second case of the partial structural change can be dealt with in a similar way to the first case, except that the local profile QMLE and local QMLE have to be applied under $\mathbb{H}_0$ and $\mathbb{H}_A$ respectively. In subsequent sections, we shall focus on hypotheses of interest introduced in Section 2.

## 5. ASYMPTOTIC PROPERTIES

We now state the asymptotic distribution of $LR$ in (4.5) under $\mathbb{H}_0$.

THEOREM 4. *Suppose Assumptions A.2, A.3, A.4(i), A.5, A.6(i), and A.8 hold. (i) Under* $\mathbb{H}_0$, $LR \xrightarrow{d} N(0,1)$ *as* $T \to \infty$. *(ii) If in addition* $\varepsilon_t \sim N(0,1)$, *then the* $LR$ *test statistic can be simplified as* $LR = [2T\sqrt{b}(l_U - l_R) - \hat{A}]/\sqrt{\hat{B}}$.

The $LR$ test has a convenient null asymptotic N(0,1) distribution. This is quite appealing in light of the facts that most existing tests for structural breaks in GARCH models have nonstandard distributions, which may depend on the DGP. The proposed test does not require formulation of an alternative and is applicable when one has no prior information of the alternative. Moreover, the new test does not require trimming data (i.e., we test all points $u \in [0,1]$ rather than restrict $u$ to be a strict subset of $[0,1]$, as usually done in existing tests). When the standardized innovation $\varepsilon_t \sim N(0,1)$, we have $\kappa = E\left(\varepsilon_t^4\right) - 3 = 0$ and hence the well-known information matrix equality $E\left[S_t\left(\theta^0\right) S_t'\left(\theta^0\right)\right] + H_0 = 0$ holds. Consequently, we can simplify the robust $LR$ test statistic to $LR = [2T\sqrt{b}(l_U - l_R) - \hat{A}]/\sqrt{\hat{B}}$.

We require that $b \to 0$ and $Tb \to \infty$ as implied by Assumption A.6(i). This is a standard condition for bandwidth $b$ and it covers the optimal rate $b^* \propto T^{-\frac{1}{5}}$ for estimation. As an important feature of $LR$, the use of the global QMLE in place of the true parameter $\theta^0$ under $\mathbb{H}_0$ has no impact on the limit distribution of $LR$. Intuitively, the global QMLE $\bar{\theta}$ converges to $\theta^0$ at a $\sqrt{T}$-rate, which is faster than the nonparametric local QMLE $\bar{\theta}_t^c$. Consequently, the asymptotic distribution of $LR$ is solely determined by $\bar{\theta}_t^c$ and thus is nuisance parameter free.

In small samples, the distribution of $LR$ may not be well approximated by the asymptotic N(0,1) distribution. Accurate finite sample critical values can be obtained by using a bootstrap procedure, which we shall discuss and justify in Section 6.

To investigate the asymptotic power property of $LR$ under $\mathbb{H}_A$, we impose the following assumption.

**Assumption A.9.** *$\theta^* = \arg\max_{\theta\in\Theta} \int_0^1 E[\tilde{l}_0(u,\theta)]du$ is the unique maximizer over $\Theta$.*

Assumption A.9 assumes that $\theta^*$ maximizes the time integrated probability limit of the expected log-likelihood function, which is a highly nonlinear function of $\theta$ in the present context. Therefore, it is not obvious how to find an explicit closed form expression for $\theta^*$ (if any) and we conjecture that $\theta^*$ may not coincide with the integrated parameter $\int_0^1 \theta_0(u)du$.

THEOREM 5. *Suppose Assumptions A.1–A.3, A.4(i), A.5, A.6(i), A.8, and A.9 hold. Then for any nonstochastic sequence $\{M_T = o(T\sqrt{b})\}$, we have $\Pr(LR > M_T) \to 1$ under $\mathbb{H}_A$ as $T \to \infty$.*

Assumption A.1 allows for both smooth structural changes and abrupt structural breaks with known or unknown break points. We permit $\theta_0(\cdot)$ to have a finite number of discontinuities. Hence, single structural break and multiple breaks with known or unknown break points, which are often considered in this literature, are included as special cases of (2.2). For example, suppose $\theta_0(\cdot)$ is a jump function, namely,

$$\theta_0(u) = \begin{cases} \left(\alpha_0^1, \alpha_1^1, \ldots, \alpha_p^1, \beta_1^1, \ldots, \beta_q^1\right)', & \text{if } u \le u_0, \\ \left(\alpha_0^2, \alpha_1^2, \ldots, \alpha_p^2, \beta_1^2, \ldots, \beta_q^2\right)', & \text{if } u > u_0. \end{cases}$$

Then we obtain the single break GARCH alternative considered in Chu (1995).

Theorem 5 suggests that the $LR$ test is consistent against all alternatives to $\mathbb{H}_0$, subject to a set of regularity conditions (*i.e.*, Assumption A.1). Thus, the proposed test will be able to detect any structural changes in GARCH models as long as the sample size $T$ is sufficiently large. This is appealing in light of the fact that usually no prior information about the alternative of structural changes is available in practice. It avoids the blindness of searching for possible alternatives of structural changes. Moreover, as we do not use any trimming procedure, that is, we test all points $u$ in the interval $[0, 1]$ rather than a strict subset of it, our test can detect structural changes that occur near the boundary of $[0, 1]$, provided that $T$ is sufficiently large and the bandwidth $b$ is sufficiently small. And, unlike some existing tests in the literature (e.g., the CUSUM test), our test has symmetric asymptotic power against structural breaks that occur either in the first or second half of the sample period.

## 6. FINITE SAMPLE PERFORMANCE

### 6.1. Parametric Bootstrap

Theorem 4 provides the null asymptotic $N(0, 1)$ distribution of the $LR$ test. Thus, one can implement our test for $\mathbb{H}_0$ by comparing $LR$ with a $N(0, 1)$ critical value, which is rather convenient in practice. However, like many other nonparametric tests in the literature, the sizes of $LR$ in finite samples based on asymptotic approximation may differ significantly from the prespecified significance level. Therefore, we shall consider the following bootstrap procedure:

- Step (i): Estimate the null GARCH model and compute the estimated standardized residual sample $\{\hat{\varepsilon}_t\}_{t=1}^T$;
- Step (ii): Obtain a bootstrap standardized residual $\varepsilon_t^*$ from the centered empirical distribution of $\{\hat{\varepsilon}_t\}_{t=1}^T$ and then a bootstrap sample $\mathcal{X}^* \equiv \{X_t^*\}_{t=1}^T$ from the estimated null GARCH model and $\{\varepsilon_t^*\}_{t=1}^T$;
- Step (iii): Estimate the null GARCH model using the bootstrap sample $\mathcal{X}^*$, and compute the bootstrap test statistic $LR^*$ in the same way as we compute $LR$, with $\mathcal{X}^*$ replacing the original sample $\mathcal{X} = \{X_t\}_{t=1}^T$;
- Step (iv): Repeat steps (ii) and (iii) $B$ times to obtain $B$ bootstrap test statistics $\{LR_l^*\}_{l=1}^B$, where $B$ is sufficiently large;
- Step (v): Compute the bootstrap $p$-value $p^* \equiv B^{-1}\sum_{l=1}^B \mathbf{1}(LR_l^* > LR)$.

The excess kurtosis $\hat{\kappa}$ estimated from the original sample $\mathcal{X}$ will be very close to the one estimated from the bootstrap sample $\mathcal{X}^*$ under $\mathbb{H}_0$, so the $LR$ statistic applies to both normally and nonnormally distributed innovations. The parametric bootstrap has been widely used to improve the finite sample performance of nonparametric tests. For example, Fan, Li, and Min (2006) and Li and Tkacz (2006) apply it to test the correct specification of parametric conditional distribution and conditional density in different contexts respectively.

We first show the consistency of the bootstrap in the following theorem.

THEOREM 6. *Suppose Assumptions A.1–A.3, A.4(i), A.5, A.6(i), and A.8 hold. Then conditional on* $\mathcal{X}$, $LR^* \to^d N(0,1)$ *as* $T \to \infty$.

The proof is similar to that of Theorem 4 and we need to use the fact that the parametric bootstrap ensures that in the bootstrap world, $\mathbb{H}_0$ always holds. When the null hypothesis $\mathbb{H}_0$ is true, the bootstrap procedure will lead to asymptotically correct size of the test, because $LR^*$ converges in distribution to $N(0,1)$. On the other hand, when the null hypothesis is false, the bootstrap procedure has asymptotic unit power. This follows because the test statistic $LR$ will converge to infinity in probability, whereas the bootstrap test statistic $LR^*$ still converges in distribution to $N(0,1)$.

In fact, when the same kernel $k(\cdot)$ and the same bandwidth $b$ are used for both $LR$ and $LR^*$, the parametric bootstrap described above can be greatly simplified by replacing $LR$ and $LR^*$ with $l_U - l_R$ and $l_U^* - l_R^*$ respectively, where $l_U^*$ and $l_R^*$ are the average log-likelihood values of the local QMLE and global QMLE based on the bootstrap sample $\mathcal{X}^*$. This procedure is rather convenient because there is no need to compute factors $\hat{A}$, $\hat{B}$ and to estimate $\kappa$. It is also applicable no matter whether $\varepsilon_t$ is normal or nonnormal.

The consistency of the parametric bootstrap does not indicate the degree of improvement of the parametric bootstrap upon the asymptotic approximation. Since $LR$ is asymptotically pivotal, it is possible that $LR^*$ can achieve reasonable

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

accuracy in finite samples. We shall examine the performance of the parametric bootstrap via simulation study.

### 6.2. Simulation Study

To examine the size of our test under $\mathbb{H}_0$, we consider the following DGP:
DGP S1 [*Standard GARCH(1,1)*]:

$$\begin{cases} X_t = \sqrt{h_t^0}\varepsilon_t \\ h_t^0 = 0.1 + 0.2X_{t-1}^2 + 0.7h_{t-1}^0 \\ \varepsilon_t \sim i.i.d.N\,(0,1)\,. \end{cases} \tag{6.1}$$

The standard GARCH(1,1) model is the most popular GARCH model and has been widely used in modeling volatilities in financial econometrics. We generate 500 data sets of a random sample $\{X_t\}_{t=1}^T$ for $T = 250$ and 500 respectively, using the Matlab Windows Version 7 Random Number Generator. We generate an initial value $X_0$ from its unconditional density $N(0,1)$ and discard the first 5000 realizations to eliminate the impact of the initial value.

To investigate the power of our test in detecting structural changes in GARCH models, we consider the following four alternatives:
DGP P1 [*Single Break in a GARCH*]:

$$\begin{cases} X_t = \sqrt{h_t^0}\varepsilon_t \\ h_t^0 = \begin{cases} 0.1 + 0.2X_{t-1}^2 + 0.4h_{t-1}^0, & \text{if } t \le 0.5T, \\ 0.3 + 0.4X_{t-1}^2 + 0.55h_{t-1}^0, & \text{otherwise;} \end{cases} \end{cases} \tag{6.2}$$

DGP P2 [*Multiple Breaks in a GARCH*]:

$$\begin{cases} X_t = \sqrt{h_t^0}\varepsilon_t \\ h_t^0 = \begin{cases} 0.1 + 0.2X_{t-1}^2 + 0.3h_{t-1}^0, & \text{if } t \le 0.3T, \\ 0.3 + 0.3X_{t-1}^2 + 0.4h_{t-1}^0, & \text{if } 0.3T < t \le 0.6T, \\ 0.5 + 0.4X_{t-1}^2 + 0.55h_{t-1}^0, & \text{otherwise;} \end{cases} \end{cases} \tag{6.3}$$

DGP P3 [*Smooth Transition GARCH*]:

$$\begin{cases} X_t = \sqrt{h_t^0}\varepsilon_t \\ h_t^0 = \left(0.1 + 0.2X_{t-1}^2 + 0.4h_{t-1}^0\right)[1 + 0.5G\,(t)] \\ G\,(t) = \{1 + \exp[-5\,(t/T - 0.5)]\}^{-1}; \end{cases} \tag{6.4}$$

where $\varepsilon_t \sim i.i.d.N\,(0,1)$.

The single break has been a structural change with classical importance. Under DGP P1, an abrupt break occurs in a GARCH(1) model at some unknown time $t$.

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

This alternative has been considered by Chu (1995) and Berkes et al. (2004). DGP P2 admits monotonic multiple breaks. DGP P3 is the time-varying smooth transition multiplicative GARCH(1,1) model proposed by Amado and Teräsvirta (2008).

For the proposed $LR$ test, we use the quartic kernel $k(u) = \frac{15}{16}(1-u^2)^2\mathbf{1}(|u| \leq 1)$, where $\mathbf{1}(\cdot)$ is the indicator function taking value 1 if $|u| \leq 1$ and 0 otherwise. In fact, our simulation evidence (not reported here) suggests that the choice of $k(\cdot)$ has little impact on the performance of the test. For simplicity, we choose the bandwidth $b = \left(1/\sqrt{12}\right)T^{-\frac{1}{5}}$, where $1/\sqrt{12}$ is the standard deviation of $U(0,1)$, which could be viewed as the limiting distribution of the grid points $\frac{t}{T}$, $t = 1, \ldots, T$, as $T \to \infty$. We use the parametric bootstrap described above with the number of bootstrap iterations $B = 100$. Both 10% and 5% significance levels are considered.

The simulated results are summarized in Table 1. Under DGP S1, the $LR$ test has good size. For example, the rejection rate is 5.8% at the 5% level when $T = 250$ and decreases to 5.4% when $T = 500$. DGP P1 has a single sudden break with unknown break date. The $LR$ test has reasonable power. The rejection rate is 43.6% at the 5% level even when the sample size $T$ is as small as 250, and increases to 68.6% when $T = 500$. DGP P2 has multiple breaks. As expected, the rejection rate is higher than that under DGP P1. Under DGP P3, the coefficients of the GARCH model are changing over time smoothly. The rejection rate is a bit low when $T = 250$, but increases with the sample size $T$.

To sum up, the $LR$ test has good sizes in finite samples when the parametric bootstrap is used. It has reasonable powers against both sudden structural breaks and smooth structural changes in GARCH models.

**TABLE 1.** Size and power of $LR$

| | T = 250 | | T = 500 | |
|---|---|---|---|---|
| | 10% | 5% | 10% | 5% |
| | | Size | | |
| DGP S0 | .128 | .058 | .112 | .054 |
| | | Power | | |
| DGP P1 | .620 | .436 | .814 | .686 |
| DGP P2 | .844 | .722 | .956 | .912 |
| DGP P3 | .242 | .154 | .480 | .346 |

*Notes:* (1) DGP S0 is a classical GARCH(1,1) model, given in (6.2); DGP P1 is a GARCH model with a single break, given in (6.3); DGP P2 is a GARCH model with multiple breaks, given in (6.4); DGP P3 is a smooth transition GARCH model, given in (6.5); (2) The parametric bootstrap procedure is used and the bootstrap iteration number $B = 100$; (3) The empirical rejection rates are based on the results of 500 iterations.

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

## 7. CONCLUSION

Modeling and detecting structural changes in GARCH processes have attracted increasing attention in time series econometrics. We have contributed to this literature by establishing the asymptotic properties of a local QMLE for a class of smooth time-varying parameter GARCH models in both the interior and boundary regions of the sample period, and more importantly, proposing a new consistent test against smooth structural changes as well as abrupt structural breaks in GARCH models. Existing works mainly focus on the estimation of time-varying ARCH models, which are special cases of our time-varying GARCH models, and asymptotic properties of the local QMLE were only available for interior points in the previous literature. Moreover, no test on parameter constancy was available for time-varying ARCH models. On the other hand, our $LR$ test is intuitively appealing and straightforward to compute. It has a convenient null asymptotic N(0,1) distribution, does not require trimming data, does not require prior information on the alternative, and is consistent against all smooth structural changes as well as multiple abrupt structural breaks in GARCH or ARCH models. To improve the size, we use a parametric bootstrap procedure, which provides reasonable size and power for the proposed test in small and finite samples.

## *REFERENCES*


Amado, C. & T. Teräsvirta (2008) Modelling Conditional and Unconditional Heteroskedasticity with Smoothly Time-Varying Structure. Working paper, Stockholm School of Economics.

Amemiya, T. (1985) *Advanced Econometrics*. Harvard University Press.

Andreou, E. & E. Ghysels (2002) Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics* 17, 579–600.

Andrews, D.W.K., I. Lee, & W. Ploberger (1996) Optimal changepoint tests for normal linear regression. *Journal of Econometrics* 70, 9–38.

Ang, A. & D. Kristensen (2012) Testing conditional factor models. *Journal of Financial Economics* 106, 132–156.

Bai, J. & P. Perron (1998) Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47–78.

Berkes, I., E. Gombay, L. Horvath, & P. Kokoszka (2004) Sequential change-point detection in GARCH (p,q) models. *Econometric Theory* 20, 1140–1167.

Berkes, I., L. Horvath, & P. Kokoszka (2003) GARCH processes: Structure and estimation. *Bernoulli* 9, 201–227.

Bougerol, P. & N. Picard (1992) Strict stationarity of generalized autoregressive processes. *Annals of Probability* 20, 1714–1730.

Brown, B.M. (1971) Martingale limit theorems. *Annals of Mathematical Statistics* 42, 59–66.

Cai, Z. (2007) Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics* 136, 163–188.

Carroll, R., J. Fan, I. Gijbels, & M.P. Wand (1997) Generalized partially linear single-index models. *Journal of the American Statistical Association* 92, 477–489.

Chen, B. & Y. Hong (2012) Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica* 80, 1157–1183.

Chu, C.-S. (1995) Detecting parameter shift in GARCH models. *Econometric Reviews* 14, 241–266.


https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

Cline, D.B.H. & J.D. Hart (1991) Kernel estimation of densities with discontinuities or discontinuous derivative. *Statistics* 22, 69–84.

Dahlhaus, R. (1996a) On the Kullback Leibler information divergence of locally stationary processes. *Stochastic Processes and Their Applications* 62, 139–168.

Dahlhaus, R. (1996b) Maximum likelihood estimation and model selection for locally stationary processes. *Journal of Nonparametric Statistics* 6, 171–191.

Dahlhaus, R. (1997) Fitting time series models to nonstationary processes. *Annals of Statistics* 25, 1–37.

Dahlhaus, R. & S. Subba Rao (2006) Statistical inference for time-varying ARCH process. *Annals of Statistics* 34, 1075–1114.

Dahlhaus, R. & S. Subba Rao (2007) A recursive online algorithm for the estimation of time-varying ARCH parameters. *Bernoulli* 13, 389–422.

Diebold, F.X. (1986) Modeling the persistence of conditional variances: A comment. *Econometric Reviews* 5, 51–56.

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

Engle, R. & J. Gonzalo Rangel (2008) The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies* 21, 1187–1222.

Escanciano, J.C. (2009) Quasi-maximum likelihood estimation of semi-strong GARCH models. *Econometric Theory* 25, 561–570.

Fan, J. & T. Huang (2005) Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* 11, 1031–1057.

Fan, Y., Q. Li, & I. Min (2006) A nonparametric bootstrap test of conditional distributions. *Econometric Theory* 22, 587–612.

Francq, C. & J.M. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH. *Bernoulli* 10, 605–637.

Fryzlewicz, P., T. Sapatinas, & S. Subba Rao (2008) Normalized least-squares estimation in time-varying ARCH models. *Annals of Statistics* 36, 742–786.

Hall, P. & T.E. Wehrly (1991) A geometrical method for removing edge effects from kernel-type nonparametric regression estimators. *Journal of American Statistical Association* 86, 665–672.

Hansen, B. (2001) The new econometrics of structural change: Dating breaks in U.S. labor productivity. *Journal of Economic Perspectives* 15, 117–128.

Hillebrand, E. (2005) Neglecting parameter changes in GARCH models. *Journal of Econometrics* 129, 121–138.

Hjellvik, V., Q. Yao, & D. Tjøstheim (1998) Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference* 68, 295–321.

Kristensen, D. (2009) On stationarity and ergodicity of the bilinear model with applications to GARCH models. *Journal of Time Series Analysis* 30, 125–144.

Kulperger, R. & H. Yu (2005) High moment partial sum processes of residuals in GARCH models and their applications. *Annals of Statistics* 33, 2395–2422.

Lamoureux, C.G. & W.D. Lastrapes (1990) Persistence in variance structural change and the GARCH model. *Journal of Business and Economic Statistics* 8, 225–234.

Lee, O. (2003) A study on GARCH(p, q) process. *Communications of the Korean Mathematical Society* 18, 541–550.

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

Li, F. & G. Tkacz (2006) A consistent test for conditional density functions with time dependent data. *Journal of Econometrics* 133, 863–886.

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.

Meitz, M. & P. Saikkonen (2008) Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. *Econometric Theory* 24, 1291–1320.

Mikosch, T. & C. Stărică (2004) Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics* 86, 378–390.

Pettenuzzo, D. & A. Timmerman (2005) Predictability of Stock Returns and Asset Allocation under Structural Breaks. Working paper, UCSD.

Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99–125.

Robinson, P.M. (1989) Nonparametric estimation of time-varying parameters. In P. Hackl (ed.), *Statistical Analysis and Forecasting of Economic Structural Change*, pp. 253–264. Springer.

Schuster, E.F. (1985) Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics: Theory and Methods* 14, 1123–1136.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman & Hall.

Speckman, P. (1988) Kernel smoothing in partial linear models. *Journal of Royal Statistical Society Series B* 50, 413–436.

Stout, W. (1974) *Almost Sure Convergence*. Academic Press.

Subba Rao, S. (2006) On some nonstationary, nonlinear random processes and their stationary approximations. *Advances in Applied Probability* 38, 1155–1172.

Tibshirani, R. & T. Hastie (1987) Local likelihood estimation. *Journal of American Statistical Association* 82, 559–567.

# MATHEMATICAL APPENDIX

Throughout the appendix, $C$, $C_1$, and $C_2$ denote generic bounded constants that may differ in different places. We will use $\|\cdot\|_d$ for the $l_d$-norm and $|\cdot|_{abs}$ for the absolute matrix, where $(|A|_{abs})_{i,j} = \left|A_{i,j}\right|$. We say $A \le B$ if $A_{i,j} \le B_{i,j}$ for all $i$ and $j$, where $A$ and $B$ are two matrices with the same dimension. All convergences are taken as $T \to \infty$.

**Proof of Theorem 1.** Our proof follows Dahlhaus and Subba Rao (2006) and Berkes, Horvath, and Kokoszka (2003, BHK) closely. First we state two lemmas.

Following Subba Rao (2006), we introduce some notations. Define $\mathcal{X}_t^{\mathsf{T}} = \left(h_t^0, \ldots, h_{t-q+1}^0, X_{t-1}^2, \ldots, X_{t-p+1}^2\right)$ and $\tilde{\mathcal{X}}_t^{\mathsf{T}}(u) = \left(\tilde{h}_t\left(u, \theta_u^0\right), \ldots, \tilde{h}_{t-q+1}\left(u, \theta_u^0\right), \tilde{X}_{t-1}^2(u), \ldots, \tilde{X}_{t-p+1}^2(u)\right)$. By Assumption A.1, there exists a constant $C$ such that for all continuity points $u$, $|\alpha_j^0(u) - \alpha_j^0(v)| \le C|u-v|^{\varphi}$ and $|\beta_j^0(u) - \beta_j^0(v)| \le C|u-v|^{\varphi}$, where $v \in N_{\varepsilon}(u)$. Define an integer $M$ such that $M \ge \{\eta/[2C(p+q)]\}^{-1/\varphi}$. For each $r = 1, \ldots, M$, define

$$\mathcal{A}_t(r) = \begin{bmatrix} \tilde{\tau}_t(r) & b_q(r) & \mathbf{a}(r) & a_p(r) \\ \mathbf{I}_{q-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \varepsilon_{t-1}^2 & 0 & \mathbf{0} & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{p-2} & \mathbf{0} \end{bmatrix},$$

where $\tilde{\tau}_t(r) = [b_1(r) + a_1(r)\varepsilon_{t-1}^2, b_2(r), \ldots, b_{q-1}(r)]$, $\mathbf{a}(r) = [a_2(r), \ldots, a_{p-1}(r)]$, $\varepsilon_{t-1}^2 = [\varepsilon_{t-1}^2, 0, \ldots, 0]$, $a_i(r) = \alpha_i^0((r-1)/M) + CM^{-\varphi}$ for $i = 1, \ldots, p$, $b_j(r) = \beta_j^0((r-1)/M) + CM^{-\varphi}$ for $j = 1, \ldots, q$, and $\mathbf{I}_d$ is a $d \times d$ identity matrix. Let $\tilde{b}_t^{'} = (\sup_u \alpha_0^0(u), 0, \ldots, 0) \in \mathbb{R}^{p+q-2}$, $Y_t = \sum_{\substack{k=1 \\ k \ne t-\eta_t}}^{\infty} \sum_{r=1}^{M} \prod_{j=0}^{k-1} \mathcal{A}_{t-j}(r)\tilde{b}_{t-k}$, where $\eta_t$ is the location of discontinuity points less than $t$, and $\mathcal{A}_t = (1+\varepsilon_{t-1}^2)\mathcal{I}_{(p+q+1)}$, where $\mathcal{I}_d$ is an $d \times d$ matrix with all entries being 1.

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

LEMMA A.1. *Under the conditions of Theorem 1, we have*

$$\left|\mathcal{X}_t - \tilde{\mathcal{X}}_t(\tfrac{t}{T})\right|_{abs} \le \tfrac{1}{T^{\varphi}} V_t + U_t, \; \left|\tilde{\mathcal{X}}_t(u) - \tilde{\mathcal{X}}_t(v)\right|_{abs} \le |u-v|^{\varphi} W_t + R_t,$$

*where*

$$V_t = C \sum_{\substack{k=1 \\ k \neq t-\eta_t}}^{\infty} k \prod_{j=0}^{k-1} \mathcal{A}_{t-j}(i_1) \left[\mathcal{A}_{t-k} \left|\mathcal{X}_{t-k-1}\right|_{abs} + \tilde{b}_{t-k}\right],$$

$$W_t = C \sum_{\substack{k=1 \\ k \neq t-\eta_t}}^{\infty} \sum_{r=1}^{M} \prod_{j=0}^{k-1} \mathcal{A}_{t-j}(r) \left(\mathcal{A}_{t-k} Y_{t-k-1} + \tilde{b}_{t-k}\right),$$

$$U_t = C \sum_{k=t-\eta_t} \prod_{j=0}^{k-1} \mathcal{A}_{t-j}(i_1) \left[\left|\mathcal{X}_{t-k-1}\right|_{abs} + C\right],$$

$$R_t = C \sum_{k=t-\eta_t} \sum_{r=1}^{M} \prod_{j=0}^{k-1} \mathcal{A}_{t-j}(r)(Y_{t-k-1} + C),$$

*where $\eta_t$ is the location of discontinuity points less than $t$, and $i_1$ is such that $(i_1 - 1)/M \le t/T < i_1/M$. Moreover, for some $n \in [1, \infty)$,*

$$\begin{aligned} &\sup_{t,T} E \left\|\mathcal{X}_t\right\|_n^n < \infty, && \sup_u E \left\|\tilde{\mathcal{X}}_t(u)\right\|_n^n < \infty, \\ &\sup_{t,T} E \left\|V_t\right\|_n^n < \infty, && E \left\|W_t\right\|_n^n < \infty, \\ &E \left\|U_t\right\|_n^n \to 0, && E \left\|R_t\right\|_n^n \to 0. \end{aligned}$$

**Proof of Lemma A.1.** The proof of Lemma A.1 is rather similar to the proof of Theorem 2.1 of Subba Rao (2006) with a finite number of discontinuities. Therefore, we follow her proof with some modification. Note that $\mathcal{X}_t = A_t(\frac{t}{T})\mathcal{X}_{t-1} + b_t(\frac{t}{T})$ and $\tilde{\mathcal{X}}_t(u) = A_t(u)\tilde{\mathcal{X}}_{t-1}(u) + b_t(u)$, where $b_t(u)' = (\alpha_0^0(u), 0, \ldots, 0) \in \mathbb{R}^{p+q-2}$,

$$A_t(u) = \begin{bmatrix} \tau_t(u) & \beta_q^0(u) & \alpha^0(u) & \alpha_p^0(u) \\ \mathbf{I}_{q-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \varepsilon_{t-1}^2 & 0 & \mathbf{0} & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{p-2} & \mathbf{0} \end{bmatrix},$$

$\tau_t(u) = [\beta_1^0(u) + \alpha_1^0(u)\varepsilon_{t-1}^2, \beta_2^0(u), \ldots, \beta_{q-1}^0(u)]$, $\alpha^0(u) = [\alpha_2^0(u), \ldots, \alpha_{p-1}^0(u)]$.

By iteratively expanding $\mathcal{X}_t$ and $\tilde{\mathcal{X}}_t(\frac{t}{T})$, we have

$$\begin{aligned} &\left|\mathcal{X}_t - \tilde{\mathcal{X}}_t(\frac{t}{T})\right|_{abs} \\ &\quad = \left|A_t\left(\frac{t}{T}\right)\left[A_{t-1}\left(\frac{t-1}{T}\right) - A_{t-1}\left(\frac{t}{T}\right)\right]\mathcal{X}_{t-2} + A_{t-1}\left(\frac{t}{T}\right)\left[\mathcal{X}_{t-2} - \tilde{\mathcal{X}}_{t-2}\left(\frac{t}{T}\right)\right]\right. \\ &\qquad \left. + A_t\left(\frac{t}{T}\right)\left[b_{t-1}\left(\frac{t-1}{T}\right) - b_{t-1}\left(\frac{t}{T}\right)\right]\right|_{abs} \\ &\quad \le \frac{1}{T^{\varphi}} \mathcal{A}_t(i_1)\mathcal{A}_{t-1}\left|\mathcal{X}_{t-2}\right|_{abs} + \mathcal{A}_{t-1}(i_1)\left|\mathcal{X}_{t-2} - \tilde{\mathcal{X}}_{t-2}\left(\frac{t}{T}\right)\right|_{abs} + \frac{1}{T^{\varphi}} \mathcal{A}_t(i_1)\tilde{b}_{t-1} \\ &\quad \le \frac{1}{T^{\varphi}} V_t + U_t, \end{aligned}$$

where we obtain the first inequality by using the facts that $\left|A_{t-1}\left(\frac{t-1}{T}\right)-A_{t-1}\left(\frac{t}{T}\right)\right|_{abs} \le \frac{C}{T^{\varphi}}\mathcal{A}_{t-1}$ and $\left|b_{t-1}\left(\frac{t-1}{T}\right)-b_{t-1}\left(\frac{t}{T}\right)\right|_{abs} \le \frac{C}{T^{\varphi}}\tilde{b}_{t-1}$ as shown in Subba Rao (2006) and obtain the second inequality by continuing the iteration and separating continuous and discontinuous points. We note that $E\|U_t\|_n^n \le \left(\sum_{k=t-\eta_t}\rho^k\right)^n \le C\rho^{Tbn} = o(1)$, where $0<\rho<1$ and $\sup_{t,T} E\|V_t\|_n^n < \infty$, as shown in Subba Rao (2006). Using a similar argument, we can show that $\left|\tilde{\mathcal{X}}_t(u)-\tilde{\mathcal{X}}_t(v)\right|_{abs} \le |u-v|^{\varphi} W_t + R_t$, where $E\|W_t\|_n^n<\infty$ and $E\|R_t\|_n^n \to 0$. ■

LEMMA A.2. *Under the conditions of Theorem 1, we have*

$$\tilde{h}_t\left(u,\theta_u^0\right) = \xi_0\left(\theta_u^0\right) + \sum_{j=1}^{\infty}\xi_j\left(\theta_u^0\right)\tilde{X}_{t-j}^2(u)$$

*for all $t$ with probability one, where the functions $\{\xi_j(\theta)\}$ are given in BHK (2003). Moreover the representation is unique.*

**Proof of Lemma A.2.** This is shown in BHK (2003). ■

In the derivation of the asymptotic properties of $\bar{\theta}_t$, we make use of the local approximation of $X_t^2$ by the stationary process $\tilde{X}_t^2(u)$ for all $t$ with $\left|\frac{t}{T}-u\right| < \frac{1}{T}$, where $u$ is a continuity point in $[0,1]$. As the parameter space $\Theta$ is a compact set, the proof of Theorem 1 consists of the proofs of Theorems A.1–A.4. Then from the compactness of $\Theta$ and the continuity of $L(u,\cdot)$ in (3.6), we conclude $\bar{\theta}_t \xrightarrow{P} \theta_u^0$ by Theorem 4.1.1 of Amemiya (1985). ■

THEOREM A.1. *Under the conditions of Theorem 1,* $\sup_{\theta\in\Theta}\left|\tilde{L}(u,\theta)-L(u,\theta)\right| \xrightarrow{P} 0$, *where $\tilde{L}(u,\theta)$ and $L(u,\theta)$ are defined in (3.5) and (3.6) respectively.*

**Proof of Theorem A.1.** To prove uniform convergence, it is sufficient to show pointwise convergence and stochastic equicontinuity of $\tilde{L}(u,\cdot)$. Theorem A.1 follows from the two propositions below.

PROPOSITION A.1. *Under the conditions of Theorem 1, for any $\theta\in\Theta$,*

$$\frac{1}{T}\sum_{s=1}^{T}k_{st}\ln\tilde{h}_s(u,\theta) \xrightarrow{P} E\left[\ln\tilde{h}_0(u,\theta)\right] \tag{A.1}$$

*and*

$$\frac{1}{T}\sum_{s=1}^{T}k_{st}\frac{\tilde{X}_s(u)^2}{\tilde{h}_s(u,\theta)} \xrightarrow{P} E\left[\frac{\tilde{X}_0(u)^2}{\tilde{h}_0(u,\theta)}\right]. \tag{A.2}$$

**Proof of Proposition A.1.** Note that $\tilde{X}_t^2(u)$ is a stationary process indexed by $u$. By Lemmas 3.1 and 5.1 of BHK (2003), we have

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

$$0 < C_1 \le \tilde{h}_s(u,\theta) \le C_2 \left(1 + \sum_{j=1}^{\infty} \rho_0^{j/q} \tilde{X}_{t-j}^2(u)\right),$$

$$E\left|\ln \tilde{h}_0(u,\theta)\right| \le E\left[C_2\left(1 + \sum_{j=1}^{\infty} \rho_0^{j/q} \tilde{X}_{t-j}^2(u)\right)\right] < \infty,$$

$$E\left[\frac{\tilde{X}_0^2(u)}{\tilde{h}_0(u,\theta)}\right] = E\left(\varepsilon_0^2\right) E\left[\frac{\tilde{h}_0\left(u,\theta_u^0\right)}{\tilde{h}_0(u,\theta)}\right] < \infty,$$

where $\rho_0 = q\underline{\rho}$. $\tilde{X}_t^2(u)$ is a measurable function of $\{\varepsilon_t\}$ for each $u$, so it is stationary and ergodic at a given $u$ by Theorem 3.5.8 of Stout (1974). Both (A.1) and (A.2) follow from Lemma A.2 of Dahlhaus and Subba Rao (2006). ■

PROPOSITION A.2. *Under the conditions of Theorem 1, $\tilde{L}(u,\cdot)$ is equicontinuous over $\Theta$ in probability.*

**Proof of Proposition A.2.** We write

$$\sup_{\theta_1,\theta_2\in\Theta,\ \theta_1\neq\theta_2} \frac{1}{\|\theta_1-\theta_2\|_2^2}\left|\tilde{L}(u,\theta_1)-\tilde{L}(u,\theta_2)\right|^2 \le \frac{C}{2T}\sum_{s=1}^{T} k_{st}\tau_s(u),$$

where

$$\tau_s(u) = \sup_{\theta_1,\theta_2\in\Theta,\ \theta_1\neq\theta_2} \frac{1}{\|\theta_1-\theta_2\|_2^2} \times \left\{\left|\ln\tilde{h}_s(u,\theta_1) - \ln\tilde{h}_s(u,\theta_2)\right|^2 + \left|\frac{\tilde{X}_s^2(u)}{\tilde{h}_s(u,\theta_1)} - \frac{\tilde{X}_s^2(u)}{\tilde{h}_s(u,\theta_2)}\right|^2\right\}.$$

By Theorem 3.5.8 of Stout (1974), $\tau_s(u)$ is a stationary and ergodic for a given $u$. By Lemma A.3, we have $E\left[\tau_0(u)\right] < \infty$. Hence

$$\frac{1}{2T}\sum_{s=1}^{T} k_{st}\tau_s(u) = O_P(1)$$

and

$$\sup_{\theta_1,\theta_2\in\Theta,\ \theta_1\neq\theta_2} \frac{1}{\|\theta_1-\theta_2\|_2^2}\left|\tilde{L}(u,\theta_1)-\tilde{L}(u,\theta_2)\right|^2 = O_P(1)$$

by Lemma A.2 of Dahlhaus and Subba Rao (2006). Therefore, $\tilde{L}(u,\cdot)$ is equicontinuous over $\Theta$ in probability. ■

LEMMA A.3. *Under the conditions of Theorem 1,*

$$E\sup_{\theta_1,\theta_2\in\Theta,\ \theta_1\neq\theta_2} \frac{1}{\|\theta_1-\theta_2\|_2^2}\left|\frac{\tilde{X}_s^2(u)}{\tilde{h}_s(u,\theta_1)} - \frac{\tilde{X}_s^2(u)}{\tilde{h}_s(u,\theta_2)}\right|^2 < \infty$$

*and*

$$E\sup_{\theta_1,\theta_2\in\Theta,\ \theta_1\neq\theta_2} \frac{1}{\|\theta_1-\theta_2\|_2^2}\left|\ln\tilde{h}_s(u,\theta_1) - \ln\tilde{h}_s(u,\theta_2)\right|^2 < \infty.$$

**Proof of Lemma A.3.** Similar to the proof of Lemma 5.3 of BHK (2003). ■

THEOREM A.2. *Under the conditions of Theorem 1,* $\sup_{\theta\in\Theta}|\mathcal{B}_t(u,\theta)| \xrightarrow{P} 0$, *where* $\mathcal{B}_t(u,\theta) \equiv L_t(\theta) - \tilde{L}(u,\theta)$, *and* $L_t(\theta)$ *is defined in (3.4).*

**Proof of Theorem A.2.** We decompose

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left|L_t(\theta)-\tilde{L}(u,\theta)\right| \\
&\le \sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}k_{st}\left|\frac{X_s^2}{h_s(\theta)}-\frac{\tilde{X}_s^2(u)}{\tilde{h}_s(u,\theta)}\right| + \sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}k_{st}\left|\ln h_s(\theta)-\ln\tilde{h}_s(u,\theta)\right| \\
&\le \sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}\frac{k_{st}}{h_s(\theta)}\left|X_s^2-\tilde{X}_s^2(u)\right| + \sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}\frac{k_{st}\tilde{X}_s^2(u)}{h_s(\theta)\tilde{h}_s(u,\theta)}\left|h_s(\theta)-\tilde{h}_s(u,\theta)\right| \\
&\quad + \sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}k_{st}\left|\ln h_s(\theta)-\ln\tilde{h}_s(u,\theta)\right|. \qquad \textbf{(A.3)}
\end{aligned}$$

Theorem A.2 follows from three lemmas below under the conditions of Theorem 1.

LEMMA A.4. $\sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}\frac{k_{st}}{h_s(\theta)}\left|X_s^2-\tilde{X}_s^2(u)\right| = o_P(1)$.

**Proof of Lemma A.4.**

$$\begin{aligned}
&\sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}\frac{k_{st}}{h_s(\theta)}\left|X_s^2-\tilde{X}_s^2(u)\right| \\
&\le \sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}\frac{k_{st}}{h_s(\theta)}\left(\left|\frac{s}{T}-u\right|^{\varphi}W_{s+1,q+1}+\frac{1}{T^{\varphi}}V_{s+1,q+1}\right) \\
&\le \sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}\frac{k_{st}}{h_s(\theta)}\left(\left|\frac{s-t}{T}\right|^{\varphi}W_{s+1,q+1}+\left|\frac{t}{T}-u\right|^{\varphi}W_{s+1,q+1}+\frac{1}{T^{\varphi}}V_{s+1,q+1}\right) \\
&= O_P\left(b^{\varphi}\right),
\end{aligned}$$

where $W_{s,d}$ and $V_{s,d}$ are the $dth$ element of $W_s$ and $V_s$, which are defined in Lemma A.1 respectively. We have used the fact that $\left|\frac{t}{T}-u\right| < \frac{1}{T}$ and Lemma A.1. ■

LEMMA A.5. $\sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}\frac{k_{st}\tilde{X}_s^2(u)}{h_s(\theta)\tilde{h}_s(u,\theta)}\left|h_s(\theta)-\tilde{h}_s(u,\theta)\right| = o_P(1)$.

**Proof of Lemma A.5.**

$$\begin{aligned}
&\sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}\frac{k_{st}\tilde{X}_s^2(u)}{h_s(\theta)\tilde{h}_s(u,\theta)}\left|h_s(\theta)-\tilde{h}_s(u,\theta)\right| \\
&\le C\sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}\sum_{j=1}^{\infty}k_{st}\xi_j(\theta)\frac{\tilde{h}_s\left(u,\theta_u^0\right)\varepsilon_s^2}{\tilde{h}_s(u,\theta)}\left(\left|\frac{s-j}{T}-u\right|^{\varphi}W_{s-j+1,q+1}+\frac{1}{T^{\varphi}}V_{s-j+1,q+1}\right)
\end{aligned}$$

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

$$\le C \sup_{\theta\in\Theta} \frac{1}{T}\sum_{s=1}^{T}\sum_{j=1}^{\infty} k_{st}\rho_0^{j/q} \frac{\tilde{h}_s\left(u,\theta_u^0\right)\varepsilon_s^2}{\tilde{h}_s\left(u,\theta\right)}\left[\left|\frac{s-t}{T}\right|^{\varphi} W_{s-j+1,q+1}\right.$$
$$\left. + \left|\frac{t}{T}-u\right|^{\varphi} W_{s-j+1,q+1} + \left|\frac{j}{T}\right|^{\varphi} W_{s-j+1,q+1} + \frac{1}{T^{\varphi}} V_{s-j+1,q+1}\right]$$
$$= O_P\left(b^{\varphi}\right). \tag{A.4}$$

We have used the monotone convergence theorem, Lemma A.1, and the fact that $\left|\frac{t}{T}-u\right| < \frac{1}{T}$. ■

LEMMA A.6. $\sup_{\theta\in\Theta} \frac{1}{T}\sum_{s=1}^{T} k_{st}\left|\ln h_s(\theta) - \ln\tilde{h}_s(u,\theta)\right| = o_P(1).$

**Proof of Lemma A.6.**

$$\sup_{\theta\in\Theta} \frac{1}{T}\sum_{s=1}^{T} k_{st}\left|\ln h_s(\theta) - \ln\tilde{h}_s(u,\theta)\right|$$
$$\le \sup_{\theta\in\Theta} \frac{1}{T}\sum_{s=1}^{T} k_{st}\left|\frac{h_s(\theta)-\tilde{h}_s(u,\theta)}{h_s^*(u,\theta)}\right|$$
$$\le \sup_{\theta\in\Theta} \frac{1}{T}\sum_{s=1}^{T}\sum_{j=1}^{\infty} C k_{st}\xi_j(\theta)\left[\left|\frac{s-j}{T}-u\right|^{\varphi} W_{s-j+1,q+1} + \frac{1}{T^{\varphi}} V_{s-j+1,q+1}\right]$$
$$\le \sup_{\theta\in\Theta} \frac{1}{T}\sum_{s=1}^{T}\sum_{j=1}^{\infty} C k_{st}\rho_0^{j/q}\left[\left|\frac{s-t}{T}\right|^{\varphi} W_{s-j+1,q+1} + \left|\frac{t}{T}-u\right|^{\varphi} W_{s-j+1,q+1}\right.$$
$$\left. + \left|\frac{j}{T}\right|^{\varphi} W_{s-j+1,q+1} + \frac{1}{T^{\varphi}} V_{s-j+1,q+1}\right]$$
$$= O_P\left(b^{\varphi}\right), \tag{A.5}$$

where $h^*(u,\theta)$ lies between $h_s(\theta)$ and $\tilde{h}_s(u,\theta)$. We have used the mean value theorem, the monotone convergence theorem, Lemma A.1, and the fact that $\left|\frac{t}{T}-u\right| < \frac{1}{T}$. ■

THEOREM A.3. *Under the conditions of Theorem 1, $L(u,\theta)$ has a unique maximum at $\theta_u^0$.*

**Proof of Theorem A.3.** It follows from Lemma 5.5 of BHK (2003) for the stationary case. ■

From the compactness of $\Theta$ and the continuity of $L(u,\cdot)$, we can conclude $\hat{\theta}_t \xrightarrow{P} \theta_u^0$ by Theorem 4.1.1 of Amemiya (1985). ■

THEOREM A.4. *Under the conditions of Theorem 1,* $\sup_{\theta\in\Theta}\left|L_t(\theta) - \bar{L}_t(\theta)\right| \xrightarrow{P} 0$, *where $\bar{L}_t(\theta)$ is defined in (3.7).*

**Proof of Theorem A.4.** We rewrite

$$\sup_{\theta\in\Theta}\left|L_t(\theta) - \bar{L}_t(\theta)\right| \le \sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T} k_{st}\left|\frac{X_s^2}{h_s(\theta)} - \frac{X_s^2}{\bar{h}_s(\theta)}\right| + \sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T} k_{st}\left|\ln h_s(\theta) - \ln\bar{h}_s(\theta)\right|.$$

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

It suffices to show the following two lemmas.

LEMMA A.7. $\sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}k_{st}\left|\frac{X_s^2}{h_s(\theta)}-\frac{X_s^2}{\bar{h}_s(\theta)}\right| = o_P(1)$, *where* $\bar{h}_s(\theta)$ *is defined in (3.7).*

**Proof of Lemma A.7.** We have

$$\sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}k_{st}\left|\frac{X_s^2}{h_s(\theta)}-\frac{X_s^2}{\bar{h}_s(\theta)}\right| \le \sup_{\theta\in\Theta}\frac{C}{T}\sum_{s=1}^{T}\sum_{j=0}^{\infty}k_{st}\rho_0^{s/q}\frac{X_s^2}{h_s(\theta)}\rho_0^{j/q}X_{-j}^2 = o_P(1),$$

where we have used Lemma A.1 and the monotone convergence theorem. ■

LEMMA A.8. $\sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}k_{st}\left|\ln h_s(\theta)-\ln\bar{h}_s(\theta)\right| = o_P(1)$.

**Proof of Lemma A.8.** We have

$$\begin{aligned}\sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}k_{st}\left|\ln h_s(\theta)-\ln\bar{h}_s(\theta)\right| &\le \sup_{\theta\in\Theta}\frac{1}{T}\sum_{s=1}^{T}k_{st}\frac{\left|h_s(\theta)-\bar{h}_s(\theta)\right|}{h_s^*(\theta)}\\ &\le \sup_{\theta\in\Theta}\frac{C}{T}\sum_{s=1}^{T}\sum_{j=0}^{\infty}k_{st}\rho_0^{s/q}\rho_0^{j/q}X_{-j}^2 = o_P(1),\end{aligned}$$

where we have used Lemma A.1, the mean value theorem, and the monotone convergence theorem. ■

Therefore, Theorem A.4 follows. ■

**Proof of Theorem 2.** By a Taylor expansion, we have

$$\frac{\partial L_t\left(\hat{\theta}_t\right)}{\partial\theta}-\frac{\partial L_t\left(\theta_u^0\right)}{\partial\theta} = \frac{\partial^2 L_t\left(\theta_t^*\right)}{\partial\theta\partial\theta'}\left(\hat{\theta}_t-\theta_u^0\right),$$

where $\theta_t^*$ lies between $\hat{\theta}_t$ and $\theta_u^0$. As $\theta_t^*\xrightarrow{P}\theta_u^0$ and $\sup_{\theta\in\Theta}\left|\frac{\partial^2 L_t(\theta)}{\partial\theta\partial\theta'}-\frac{\partial^2 L(u,\theta)}{\partial\theta\partial\theta'}\right|\xrightarrow{P}0$ by Theorems A.5 and A.6, we have $\frac{\partial^2 L_t(\theta_t^*)}{\partial\theta\partial\theta'}\xrightarrow{P}H(u)$ and

$$\sqrt{bT}\left[\left(\hat{\theta}_t-\theta_u^0\right)+H^{-1}(u)\frac{\partial\mathcal{B}_t\left(u,\theta_u^0\right)}{\partial\theta}\right] = -\sqrt{bT}H^{-1}(u)\frac{\partial\tilde{L}\left(u,\theta_u^0\right)}{\partial\theta}+o_P(1),$$

where $\mathcal{B}_t(u,\theta)$ is defined in Theorem A.2. The proof of Theorem 2(i) consists of the proofs of Theorems A.5–A.8.

THEOREM A.5. *Under the conditions of Theorem 2(i),* $\sup_{\theta\in\Theta}\left|\frac{\partial^2\tilde{L}(u,\theta)}{\partial\theta\partial\theta'}-\frac{\partial^2 L(u,\theta)}{\partial\theta\partial\theta'}\right|\xrightarrow{P}0.$

THEOREM A.6. *Under the conditions of Theorem 2(i),* $\sup_{\theta\in\Theta}\left|\frac{\partial^2\mathcal{B}_t(u,\theta)}{\partial\theta\partial\theta'}\right|\xrightarrow{P}0.$

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

The proofs of Theorems A.5 and A.6 are rather similar to the proofs of Theorems A.1 and A.2 respectively. We omit the details here.

THEOREM A.7. *Under the conditions of Theorem 2(i),*

$$\sqrt{bT}\frac{\partial \tilde{L}\left(u,\theta_u^0\right)}{\partial \theta} \to^d N\left(0, -k_2\left(\frac{\kappa}{2}+1\right)H(u)\right).$$

**Proof of Theorem A.7.** By definition, we have

$$\sqrt{bT}\frac{\partial \tilde{L}\left(u,\theta_u^0\right)}{\partial \theta} = \frac{1}{\sqrt{bT}}\sum_{s=1}^{T} k\left(\frac{s-t}{Tb}\right)\frac{\partial \tilde{l}_s\left(u,\theta_u^0\right)}{\partial \theta},$$

where $\left[\frac{\partial \tilde{l}_s(u,\theta_u^0)}{\partial \theta}\right]$ is a martingale difference sequence. It is straightforward to verify the conditional Linderberg and variance conditions. The result follows from the martingale central limit theorem and the Cramér–Wold device. ∎

THEOREM A.8. *Under the conditions of Theorem 2(i),*

$$E\left[\frac{\partial \mathcal{B}_t\left(u,\theta_u^0\right)}{\partial \theta}\right] = B_u + O\left(b^3 + \frac{1}{T}\right)$$

*and*

$$var\left[\frac{\partial \mathcal{B}_t\left(u,\theta_u^0\right)}{\partial \theta}\right] = O\left(b^6 + \frac{1}{T}\right).$$

**Proof of Theorem A.8.** We define

$$\tilde{l}_s^\tau\left(u,\theta_u^0\right) = -\frac{1}{2}\left[\ln \tilde{h}_s^\tau\left(u,\theta_u^0\right) + \frac{\tilde{X}_s^2(u)}{\tilde{h}_s^\tau\left(u,\theta_u^0\right)}\right],$$

$$\tilde{h}_s^\tau\left(u,\theta_u^0\right) = \xi_0\left(\theta_u^0\right) + \sum_{j=1}^{\tau}\xi_j\left(\theta_u^0\right)\tilde{X}_{s-j}^2(u),$$

where $\tau$ is some deterministic function of $T$ such that $T\rho_0^{\tau+1/q} = O(1)$ and $b\tau = O(1)$ as $T \to \infty$.

$$\frac{\partial \mathcal{B}_t\left(u,\theta_u^0\right)}{\partial \theta} = \frac{1}{T}\sum_{s=1}^{T} k_{st}\left[\frac{\partial \tilde{l}_s^\tau\left(\frac{s}{T},\theta_u^0\right)}{\partial \theta} - \frac{\partial \tilde{l}_s^\tau\left(u,\theta_u^0\right)}{\partial \theta}\right]$$
$$+ \frac{1}{T}\sum_{s=1}^{T} k_{st}\left[\frac{\partial \tilde{l}_s^\tau\left(u,\theta_u^0\right)}{\partial \theta} - \frac{\partial \tilde{l}_s\left(u,\theta_u^0\right)}{\partial \theta}\right]$$

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

$$+\frac{1}{T}\sum_{s=1}^{T}k_{st}\left[\frac{\partial\tilde{l}_s\left(\frac{s}{T},\theta_u^0\right)}{\partial\theta}-\frac{\partial\tilde{l}_s^{\tau}\left(\frac{s}{T},\theta_u^0\right)}{\partial\theta}\right]$$
$$+\frac{1}{T}\sum_{s=1}^{T}k_{st}\left[\frac{\partial l_s\left(\theta_u^0\right)}{\partial\theta}-\frac{\partial\tilde{l}_s\left(\frac{s}{T},\theta_u^0\right)}{\partial\theta}\right]$$
$$=B_0+R_1+R_2+R_3. \tag{A.6}$$

For the second term of (A.6), we have

$$E\left|R_1\right|\leq\frac{1}{T}\sum_{s=1}^{T}k_{st}E\left|\frac{\frac{\partial\tilde{h}_s(u,\theta_u^0)}{\partial\theta}\left[\tilde{h}_s^{\tau}\left(u,\theta_u^0\right)-\tilde{h}_s\left(u,\theta_u^0\right)\right]}{\tilde{h}_s\left(u,\theta_u^0\right)\tilde{h}_s^{\tau}\left(u,\theta_u^0\right)}\right|$$
$$+\frac{1}{T}\sum_{s=1}^{T}k_{st}E\left|\frac{\frac{\partial\tilde{h}_s(u,\theta_u^0)}{\partial\theta}-\frac{\partial\tilde{h}_s^{\tau}(u,\theta_u^0)}{\partial\theta}}{\tilde{h}_s^{\tau}\left(u,\theta_u^0\right)}\right|$$
$$+\frac{1}{T}\sum_{s=1}^{T}k_{st}E\left|\frac{\tilde{X}_s^2(u)\frac{\partial\tilde{h}_s(u,\theta_u^0)}{\partial\theta}\left[\tilde{h}_s^{\tau 2}\left(u,\theta_u^0\right)-\tilde{h}_s^2\left(u,\theta_u^0\right)\right]}{\tilde{h}_s^2\left(u,\theta_u^0\right)\tilde{h}_s^{\tau 2}\left(u,\theta_u^0\right)}\right|$$
$$+\frac{1}{T}\sum_{s=1}^{T}k_{st}E\left|\frac{\tilde{X}_s^2(u)\left[\frac{\partial\tilde{h}_s(u,\theta_u^0)}{\partial\theta}-\frac{\partial\tilde{h}_s^{\tau}(u,\theta_u^0)}{\partial\theta}\right]}{\tilde{h}_s^{\tau 2}\left(u,\theta_u^0\right)}\right|$$
$$\equiv\sum_{j=1}^{4}R_{1j},\ \text{say,}$$

where

$$R_{11}\leq\frac{1}{T}\sum_{s=1}^{T}k_{st}\left\{C+\sum_{j=1}^{\infty}Cj\rho_0^{j/q}\left[E\tilde{X}_{s-j}^4(u)\right]^{1/2}\right\}\left\{\sum_{j=\tau+1}^{\infty}\rho_0^{j/q}\left[E\tilde{X}_{s-j}^4(u)\right]^{1/2}\right\}$$
$$=O\left(\rho_0^{\tau+1/q}\right)=O\left(\frac{1}{T}\right),$$

and we have used the facts that $\xi_j\left(\theta_u^0\right)\leq C\rho_0^{j/q}$ and $\left|\partial\xi_j\left(\theta_u^0\right)/\partial\theta\right|\leq Cj\rho_0^{j/q}$ by Lemmas 3.1 and 3.2 of BHK (2003). And

$$R_{12}\leq\frac{1}{T}\sum_{s=1}^{T}k_{st}E\left|\frac{\sum_{j=\tau+1}^{\infty}Cj\rho_0^{j/q}\tilde{X}_{s-j}^2(u)}{\tilde{h}_s^{\tau}\left(u,\theta_u^0\right)}\right|$$
$$\leq\frac{1}{T}\sum_{s=1}^{T}k_{st}\sum_{j=\tau+1}^{\infty}C\tilde{\rho}^{j}E\frac{\tilde{X}_{s-j}^2(u)}{\tilde{h}_s^{\tau}\left(u,\theta_u^0\right)}=O\left(\tilde{\rho}^{\tau+1}\right)=O\left(\frac{1}{T}\right),$$

where $\tilde{\rho}=e^{\frac{1}{e}}\rho_0$. The proofs of $R_{13}$ and $R_{14}$ are rather similar to those of $R_{11}$ and $R_{12}$ respectively. The third term of (A.6) can be shown in a similar way to the second term.

And for the last term, we have

$$E\left|R_3\right| \le \frac{1}{T}\sum_{s=1}^{T} k_{st} E\left|\frac{\frac{\partial h_s(\theta_u^0)}{\partial\theta}\left[\tilde{h}_s\left(\frac{s}{T},\theta_u^0\right)-h_s\left(\theta_u^0\right)\right]}{h_s\left(\theta_u^0\right)\tilde{h}\left(\frac{s}{T},\theta_u^0\right)}\right|\left[1+\frac{\tilde{X}_s^2\left(\frac{s}{T}\right)}{\tilde{h}\left(\frac{s}{T},\theta_u^0\right)}\right]$$

$$+k_{st}E\left|\frac{\frac{\partial h_s(\theta_u^0)}{\partial\theta}-\frac{\partial \tilde{h}_s(\frac{s}{T},\theta_u^0)}{\partial\theta}}{\tilde{h}_s\left(\frac{s}{T},\theta_u^0\right)}\right|\left[1+\frac{\tilde{X}_s^2\left(\frac{s}{T}\right)}{\tilde{h}\left(\frac{s}{T},\theta_u^0\right)}\right]$$

$$+k_{st}E\left|\frac{\frac{\partial h_s(\theta_u^0)}{\partial\theta}}{h_s^2\left(\theta_u^0\right)}\right|\left|X_s^2-\tilde{X}_s^2\left(\frac{s}{T}\right)\right|$$

$$+k_{st}E\left|\frac{\tilde{X}_s^2\left(\frac{s}{T}\right)}{h_s\left(\theta_u^0\right)\tilde{h}\left(\frac{s}{T},\theta_u^0\right)}\right|\left|h_s\left(\theta_u^0\right)-\tilde{h}_s\left(\frac{s}{T},\theta_u^0\right)\right|$$

$$=\frac{1}{T^{\varphi}},$$

where we have repeatedly used Lemma A.1 four times.

Next, we consider the first term of (A.6). A Taylor expansion yields

$$B_0=\frac{1}{T}\sum_{s=1}^{T}k_{st}\left(\frac{s}{T}-u\right)\frac{\partial^2\tilde{l}_s^{\tau}\left(u,\theta_u^0\right)}{\partial\theta\partial u}+\frac{1}{2T}\sum_{s=1}^{T}k_{st}\left(\frac{s}{T}-u\right)^2\frac{\partial^3\tilde{l}_s^{\tau}\left(u,\theta_u^0\right)}{\partial\theta\partial u^2}$$

$$+\frac{1}{3!T}\sum_{s=1}^{T}k_{st}\left(\frac{s}{T}-u\right)^3\frac{\partial^4\tilde{l}_s^{\tau}\left(\bar{u}_s,\theta_u^0\right)}{\partial\theta\partial u^3}=B_1+B_2+B_3,\text{ say,}$$

where $\bar{u}_s$ lies between $\frac{s}{T}$ and $u$.

For the first term, we have

$$E\left(B_1\right)=\frac{1}{T}\sum_{s=1}^{T}k_{st}\left(\frac{s-t}{T}\right)E\left[\frac{\partial^2\tilde{l}_s^{\tau}\left(u,\theta_u^0\right)}{\partial\theta\partial u}\right]+O\left(\frac{1}{T}\right)$$

$$=b\int_{-1}^{1}k(u)udu+O\left(\frac{1}{T}\right)=O\left(\frac{1}{T}\right)$$

and

$$var\left(B_1\right)=\frac{1}{T^2}\sum_{s=1}^{T}\sum_{r=1}^{T}k_{st}k_{rt}\left(\frac{s}{T}-u\right)\left(\frac{r}{T}-u\right)cov\left(\frac{\partial^2\tilde{l}_s^{\tau}\left(u,\theta_u^0\right)}{\partial\theta\partial u},\frac{\partial^2\tilde{l}_r^{\tau}\left(u,\theta_u^0\right)}{\partial\theta\partial u}\right)$$

$$\le\frac{b^2}{b^2T^2}\sum_s k\left(\frac{s-t}{Tb}\right)\sum_r k\left(\frac{r+s-t}{Tb}\right)\left|cov\left(\frac{\partial^2\tilde{l}_s^{\tau}\left(u,\theta_u^0\right)}{\partial\theta\partial u},\frac{\partial^2\tilde{l}_{s+r}^{\tau}\left(u,\theta_u^0\right)}{\partial\theta\partial u}\right)\right|$$

$$\le\frac{Cb}{T}\sum_r\left|cov\left(\frac{\partial^2\tilde{l}_s^{\tau}\left(u,\theta_u^0\right)}{\partial\theta\partial u},\frac{\partial^2\tilde{l}_{s+r}^{\tau}\left(u,\theta_u^0\right)}{\partial\theta\partial u}\right)\right|=O\left(\frac{1}{T}\right),$$

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

where we have used the result that $b\sum_r \left|cov\left(\frac{\partial^2 \tilde{l}_s^\tau(u,\theta_u^0)}{\partial\theta\partial u}, \frac{\partial^2 \tilde{l}_{s+r}^\tau(u,\theta_u^0)}{\partial\theta\partial u}\right)\right| = O(1)$. We now verify it below. First note that

$$\begin{aligned}
\frac{\partial^2 \tilde{l}_s^\tau(u,\theta_u^0)}{\partial\theta\partial u} &= -\frac{1}{2}\left[-\frac{1}{\tilde{h}_s^{\tau 2}(u,\theta_u^0)}\frac{\partial\tilde{h}_s^\tau(u,\theta_u^0)}{\partial\theta}\frac{\partial\tilde{h}_s^\tau(u,\theta_u^0)}{\partial u} + \frac{1}{\tilde{h}_s^{\tau 2}(u,\theta_u^0)}\frac{\partial^2\tilde{h}_s^\tau(u,\theta_u^0)}{\partial\theta\partial u}\right. \\
&\quad - \frac{1}{\tilde{h}_s^{\tau 2}(u,\theta_u^0)}\frac{\partial\tilde{h}_s^\tau(u,\theta_u^0)}{\partial\theta}\frac{\partial\tilde{X}_s^2(u)}{\partial u} - \frac{\tilde{X}_s^2(u)}{\tilde{h}_s^{\tau 2}(u,\theta_u^0)}\frac{\partial^2\tilde{h}_s^\tau(u,\theta_u^0)}{\partial\theta\partial u} \\
&\quad \left. + \frac{2\tilde{X}_s^2(u)}{\tilde{h}_s^{\tau 3}(u,\theta_u^0)}\frac{\partial\tilde{h}_s^\tau(u,\theta_u^0)}{\partial\theta}\frac{\partial\tilde{h}_s^\tau(u,\theta_u^0)}{\partial u}\right] \\
&= \sum_{j=1}^{5} B_{sj}, \text{ say.}
\end{aligned}$$

It would be sufficient to show that

$$b\sum_r \left|cov\left(B_{si}, B_{(s+r)j}\right)\right| = O(1)$$

for $i, j = 1, \ldots, 5$. We first consider $i = j = 1$. In this case,

$$\begin{aligned}
b\sum_r \left|cov\left(B_{s1}, B_{(s+r)1}\right)\right| &= b\sum_{r\le\tau-1}\left|cov\left(B_{s1}, B_{(s+r)1}\right)\right| \\
&\quad + b\sum_{r>\tau-1}\left|cov\left(B_{s1}, B_{(s+r)1}\right)\right| = B_{11}^1 + B_{11}^2,
\end{aligned}$$

where

$$\begin{aligned}
B_{11}^1 &\le Cb\sum_{r\le\tau-1} E\left|\sum_{j,i=1}^{\tau}\left[\frac{\partial\xi_0(\theta_u^0)}{\partial\theta}\right]\left[\frac{\partial\xi_0(\theta_u^0)}{\partial\theta}\right]'\xi_i\left(\theta_u^0\right)\xi_j\left(\theta_u^0\right)\frac{d\tilde{X}_{s-i}^2(u)}{du}\frac{d\tilde{X}_{s+r-j}^2(u)}{du}\right| \\
&\quad + E\left|\sum_{k,j,i=1}^{\tau}\left[\frac{\partial\xi_0(\theta_u^0)}{\partial\theta}\right]\left[\frac{\partial\xi_i(\theta_u^0)}{\partial\theta}\right]'\xi_j\left(\theta_u^0\right)\xi_k\left(\theta_u^0\right)\tilde{X}_{s+r-i}^2(u)\frac{d\tilde{X}_{s-j}^2(u)}{du}\frac{d\tilde{X}_{s+r-k}^2(u)}{du}\right| \\
&\quad + E\left|\sum_{k,j,i=1}^{\tau}\left[\frac{\partial\xi_0(\theta_u^0)}{\partial\theta}\right]\left[\frac{\partial\xi_j(\theta_u^0)}{\partial\theta}\right]'\xi_i\left(\theta_u^0\right)\xi_k\left(\theta_u^0\right)\tilde{X}_{s-j}^2(u)\frac{d\tilde{X}_{s-i}^2(u)}{du}\frac{d\tilde{X}_{s+r-k}^2(u)}{du}\right| \\
&\quad + E\left|\sum_{l,k,j,i=1}^{\tau}\left[\frac{\partial\xi_j(\theta_u^0)}{\partial\theta}\right]\left[\frac{\partial\xi_k(\theta_u^0)}{\partial\theta}\right]'\xi_i\left(\theta_u^0\right)\xi_l\left(\theta_u^0\right)\right. \\
&\qquad \left.\times\,\tilde{X}_{s-j}^2(u)\tilde{X}_{s+r-k}^2(u)\frac{d\tilde{X}_{s-i}^2(u)}{du}\frac{d\tilde{X}_{s+r-l}^2(u)}{du}\right| \\
&\le Cb\sum_{r\le\tau-1}\left\{\sum_{j,i=1}^{\tau}\rho_0^{i/q}\rho_0^{j/q}\left[E\left|\frac{d\tilde{X}_{s-i}^2(u)}{du}\right|^2\right]^{1/2}\left[E\left|\frac{d\tilde{X}_{s+r-j}^2(u)}{du}\right|^2\right]^{1/2}\right. \\
&\quad + \sum_{k,j,i=1}^{\tau} i\rho_0^{i/q}\rho_0^{j/q}\rho_0^{k/q}\left[E\left|\tilde{X}_{s+r-i}^2(u)\right|^2\right]^{1/2}\left[E\left|\frac{d\tilde{X}_{s-j}^2(u)}{du}\right|^4\right]^{1/4}\left[E\left|\frac{d\tilde{X}_{s+r-k}^2(u)}{du}\right|^4\right]^{1/4}
\end{aligned}$$

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

$$+\sum_{k,j,i=1}^{\tau} j\rho_0^{j/q}\rho_0^{i/q}\rho_0^{k/q}\left[E\left|\tilde{X}_{s-j}^2(u)\right|^2\right]^{1/2}\left[E\left|\frac{d\tilde{X}_{s-i}^2(u)}{du}\right|^4\right]^{1/4}\left[E\left|\frac{d\tilde{X}_{s+r-k}^2(u)}{du}\right|^4\right]^{1/4}$$

$$+\sum_{l,k,j,i=1}^{\tau} j\rho_0^{j/q}k\rho_0^{k/q}\rho_0^{i/q}\rho_0^{l/q}\left[E\left|\tilde{X}_{s-j}^2(u)\right|^4\right]^{1/4}\left[E\left|\tilde{X}_{s+r-k}^2(u)\right|^4\right]^{1/4}$$

$$\left.\times\left[E\left|\frac{d\tilde{X}_{s-i}^2(u)}{du}\right|^4\right]^{1/4}\left[E\left|\frac{d\tilde{X}_{s+r-l}^2(u)}{du}\right|^4\right]^{1/4}\right\}$$

$$= O(b\tau) = O(1).$$

$$B_{11}^2 \le \frac{1}{4}\sum_{r>\tau-1}\beta^{\delta/\delta+1}(r-\tau)\left[E\left|\frac{1}{\tilde{h}_s^{\tau 2}(u,\theta_u^0)}\frac{\partial\tilde{h}_s^{\tau}(u,\theta_u^0)}{\partial\theta}\frac{\partial\tilde{h}_s^{\tau}(u,\theta_u^0)}{\partial u}\right|^{2(1+\delta)}\right]^{\frac{1}{2(1+\delta)}}$$

$$\times\left[E\left|\frac{1}{\tilde{h}_{s+r}^{\tau 2}(u,\theta_u^0)}\frac{\partial\tilde{h}_{s+r}^{\tau}(u,\theta_u^0)}{\partial\theta}\frac{\partial\tilde{h}_{s+r}^{\tau}(u,\theta_u^0)}{\partial u}\right|^{2(1+\delta)}\right]^{\frac{1}{2(1+\delta)}}$$

$$\le C\sum_{r>\tau-1}\alpha^{\delta/\delta+1}(r-\tau)\left[E\left|\sum_{j=1}^{\tau}\frac{\partial\xi_j(\theta_u^0)}{\partial\theta}\tilde{X}_{s-j}^2(u)\right|^{4(1+\delta)}\right]^{1/4(1+\delta)}$$

$$\times\left[E\left|\sum_{j=1}^{\tau}\xi_j\left(\theta_u^0\right)\frac{d\tilde{X}_{s-j}^2(u)}{du}\right|^{4(1+\delta)}\right]^{1/4(1+\delta)}\left[E\left|\sum_{j=1}^{\tau}\xi_j\left(\theta_u^0\right)\frac{d\tilde{X}_{s+r-j}^2(u)}{du}\right|^{4(1+\delta)}\right]^{1/4(1+\delta)}$$

$$\times\left[E\left|\sum_{j=1}^{\tau}\frac{\partial\xi_j(\theta_u^0)}{\partial\theta}\tilde{X}_{s+r-j}^2(u)\right|^{4(1+\delta)}\right]^{1/4(1+\delta)}$$

$$= O(1),$$

where $\alpha(r)$ is the mixing coefficient for the process $\mathcal{X}_t(2,u)$ and $\mathcal{X}_t(2,u)$, which is defined in Assumption A.8. We note that Subba Rao (2006, Sect. 5) has shown that $\mathcal{X}_t(2,u)$ satisfies Assumption 3.1 of Subba Rao (2006). Therefore the strong mixing property of $\mathcal{X}_t(2,u)$ with a geometric rate follows by Theorem 4.1 of Subba Rao (2006). The proofs for other $i$ and $j$ are very similar and hence we omit the details here. Similarly, we obtain $E(B_2) = B_u + O(\frac{1}{T})$, $var(B_2) = O(\frac{1}{T})$ and $E(B_3^2) = O(b^6)$, leading to the desired result of Theorem 2(i). The proofs of Theorem 2(ii) and (iii) are very similar. ■

**Proof of Theorem 3.** It follows from the proof of Theorem 2(i) with the augmented random sample and we omit the details of the proof. The major difference is the calculation of variance and bias. That is, with the augmented sample, we have

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

$$\begin{aligned}
&\operatorname{var}\left(\sqrt{bT}\frac{\partial \tilde{L}\left(u,\theta_u^0\right)}{\partial\theta}\right)\\
&\quad=\frac{1}{bT}\sum_{s=1}^{t+\lfloor Tb\rfloor}\left[k^2\left(\frac{s-t}{bT}\right)+k^2\left(\frac{s+t}{bT}\right)+k\left(\frac{s-t}{bT}\right)k\left(\frac{s-t}{bT}+\frac{2t}{bT}\right)\right.\\
&\qquad\left.+k\left(\frac{s+t}{bT}\right)k\left(\frac{s+t}{bT}-\frac{2t}{bT}\right)\right]E\left[\frac{\partial \tilde{l}_s\left(u,\theta_u^0\right)}{\partial\theta}\frac{\partial \tilde{l}_s\left(u,\theta_u^0\right)}{\partial\theta'}\right]\\
&\quad=-\left[\int_{-1}^{1}k^2(x)\,dx+\int_{-1}^{1}k(x)k(x+2c)\,dx\right]\left(\frac{E\varepsilon_t^4-1}{2}\right)H(u)+o(1),
\end{aligned}$$

and

$$\begin{aligned}
E(B_1)&=\frac{1}{T}\sum_{s=-\lfloor Tb\rfloor}^{T}\left\{k_{st}\left(\frac{s-t}{T}\right)E\left[\frac{\partial^2\tilde{l}_s(u,\theta_u^0)}{\partial\theta\partial u}\right]+k_{st}\left(\frac{t}{T}-u\right)E\left[\frac{\partial^2\tilde{l}_s(u,\theta_u^0)}{\partial\theta\partial u}\right]\right\}\\
&=b\int_{-1}^{1}xk(x)dx\,E\left[\frac{\partial^2\tilde{l}_s(u,\theta_u^0)}{\partial\theta\partial u}\right]+\int_{-1}^{1}k(x)dx\,E\left[\frac{\partial^2\tilde{l}_s\left(u,\theta_u^0\right)}{\partial\theta\partial u}\right]\left(\frac{1}{T}\right)+o\left(\frac{1}{T}\right)\\
&=O\left(\frac{1}{T}\right),
\end{aligned}$$

where, in calculating the variance, we have made use of the identity that

$$E\left[\frac{\partial \tilde{l}_s\left(u,\theta_u^0\right)}{\partial\theta}\frac{\partial \tilde{l}_s\left(u,\theta_u^0\right)}{\partial\theta'}\right]=-\left(\frac{E\varepsilon_t^4-1}{2}\right)H(u)$$

as can be easily verified. ■

**Proof of Theorem 4.** Under $\mathbb{H}_0$, we can decompose

$$\begin{aligned}
2Tb^{1/2}(l_U-l_R)&=b^{1/2}\sum_{t=1}^{T}\left[2S_t(\theta^0)'+\left(\bar{\theta}_t^c-\theta^0\right)'\frac{\partial S_t(\theta_t^1)}{\partial\theta}\right]\left(\bar{\theta}_t^c-\theta^0\right)\\
&\quad-2b^{1/2}\sum_{t=1}^{T}S_t(\theta^0)'\left(\bar{\theta}-\theta^0\right)-b^{1/2}\sum_{t=1}^{T}\left(\bar{\theta}-\theta^0\right)'\frac{\partial S_t(\theta^1)}{\partial\theta}\left(\bar{\theta}-\theta^0\right)\\
&=Q_1+Q_2+Q_3,\ \text{say,}
\end{aligned}$$

where the score function is defined in (4.3), $\theta_t^1$ lies between $\bar{\theta}_t^c$ and $\theta^0$, and $\theta^1$ lies between $\bar{\theta}$ and $\theta^0$.

The proof of Theorem 4 consists of the proofs of Theorems A.9–A.11. ■

THEOREM A.9. *Under the conditions of Theorem 4,* $Q\equiv(Q_1-\tilde{A})/\sqrt{\tilde{B}}\xrightarrow{d}N(0,1)$, *where*

$$\tilde{A}=b^{-1/2}\frac{var\left(\varepsilon_t^2\right)}{2}(1+p+q)\left[2k(0)-\int_{-1}^{1}k^2(u)du\right]$$

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

*and*

$$\tilde{B} = \left[var\left(\varepsilon_t^2\right)\right]^2 (1+p+q)\int_0^1 \left[2k(v) - \int_{-1}^1 k(u)k(u+v)du\right]^2 dv.$$

**Proof of Theorem A.9.** To show $Q \xrightarrow{d} N(0,1)$, it suffices to show three propositions below. ■

PROPOSITION A.3. *Under the conditions of Theorem 4,*

$$\bar{\theta}_t^c - \theta^0 = -H_0^{-1}\left\{\frac{1}{T}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st} S_s(\theta^0)\right\} + O_P\left(T^{-\frac{1}{2}}b^{-\frac{1}{2}}\right),$$

*where $H_0$ is defined in (4.4).*

**Proof of Proposition A.3.** By the Taylor expansion of the first order condition, we have

$$\frac{1}{T}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st} S_s(\theta^0) + \frac{1}{T}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st}\frac{\partial S_s\left(\theta_t^1\right)}{\partial\theta}\left(\bar{\theta}_t^c - \theta^0\right) = 0,$$

where $\theta_t^1$ lies between $\bar{\theta}_t^c$ and $\theta^0$. Because $\sqrt{\frac{b}{T}}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st} S_s(\theta^0) \to^d N\left(0, -\frac{k_2}{2} var(\varepsilon_t^2) H_0\right)$, it suffices to show the following lemma. ■

LEMMA A.9. *Under the conditions of Theorem 4,*

$$\frac{1}{T}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st}\frac{\partial S_s\left(\theta_t^1\right)}{\partial\theta} = H_0 + o_P(1).$$

**Proof of Lemma A.9.** We decompose

$$\begin{aligned}\frac{1}{T}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st}\frac{\partial S_s(\theta_t^1)}{\partial\theta} - H_0 &= \left[\frac{1}{T}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st}\frac{\partial S_s(\theta_t^1)}{\partial\theta} - \frac{1}{T}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st}\frac{\partial S_s(\theta^0)}{\partial\theta}\right] \\ &\quad + \left[\frac{1}{T}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st}\frac{\partial S_s(\theta^0)}{\partial\theta} - H_0\right].\end{aligned}$$

For the first term,

$$\begin{aligned}&\left\|\frac{1}{T}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st}\frac{\partial S_{s,j}(\theta_t^1)}{\partial\theta} - \frac{1}{T}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st}\frac{\partial S_{s,j}(\theta^0)}{\partial\theta}\right\|_2^2 \\ &\quad \le \frac{1}{T}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st}\sup_\theta\left\|\frac{\partial^2}{\partial\theta\partial\theta'} S_{s,j}(\theta)\right\|_2^2 \left\|\theta_t^1 - \theta^0\right\|_2^2 \xrightarrow{P} 0,\end{aligned}$$

where we have used the mean value theorem and Theorem 1.

For the second term, since $\left\{\frac{\partial S_s(\theta^0)}{\partial\theta}\right\}$ is an ergodic process with $E\left|\frac{\partial S_s(\theta^0)}{\partial\theta}\right| < \infty$, we have

$$\frac{1}{T}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st}\frac{\partial S_s(\theta^0)}{\partial\theta} \xrightarrow{P} H_0$$

by Lemma A.2 of Dahlhaus and Subba Rao (2006). ■

PROPOSITION A.4. *Under the conditions of Theorem 4,*

$$Q_1 = \tilde{A} + 2\tilde{U} + o_P(1),$$

*where*

$$\tilde{U} = T^{-2}b^{1/2}\sum_{s=2}^{T}\sum_{t=1}^{s-1} S_s(\theta^0)' H_0^{-1} S_t(\theta^0)\left(2Tk_{st} - \sum_{r=1}^{T} k_{rs}k_{rt}\right). \tag{A.7}$$

**Proof of Proposition A.4.** We first decompose

$$\begin{aligned}
Q_1 &= -T^{-1}b^{-1/2}\sum_{t=1}^{T} S_t(\theta^0)' H_0^{-1}\left[2k(0)\mathbf{I}_d - T^{-1}b\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st}^2\frac{\partial S_s(\theta^0)}{\partial\theta}H_0^{-1}\right]S_t(\theta^0)\\
&\quad - 2T^{-1}b^{-1/2}\sum_{t=1}^{Tb} S_t(\theta^0)' H_0^{-1}\left[2k\left(\frac{2t}{Tb}\right)\mathbf{I}_d - T^{-1}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st}k\left(\frac{s+t}{Tb}\right)\frac{\partial S_s(\theta^0)}{\partial\theta}H_0^{-1}\right]S_t(\theta^0)\\
&\quad - 2T^{-1}b^{-1/2}\sum_{t>s=1}^{T} S_t(\theta^0)' H_0^{-1}\left[2bk_{ts}\mathbf{I}_d - T^{-1}b\sum_{r} k_{rt}k_{rs}\frac{\partial S_r(\theta^0)}{\partial\theta}H_0^{-1}\right]S_s(\theta^0)\\
&\quad + T^{-2}b^{1/2}\sum_{t=1}^{T}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} S_s(\theta^0)' H_0^{-1}\left\{k_{st}^2\left[\frac{\partial S_t(\theta_t^1)}{\partial\theta} - \frac{\partial S_t(\theta^0)}{\partial\theta}\right]\right\}H_0^{-1}S_s(\theta^0)\\
&\quad + 2T^{-2}b^{-1/2}\sum_{t=1}^{Tb}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} S_s(\theta^0)' H_0^{-1}\left\{k_{st}k\left(\frac{s+t}{Tb}\right)\left[\frac{\partial S_t(\theta_t^1)}{\partial\theta} - \frac{\partial S_t(\theta^0)}{\partial\theta}\right]\right\}H_0^{-1}S_s(\theta^0)\\
&\quad + 2T^{-2}b^{1/2}\sum_{t=1}^{Tb}\sum_{s>r} S_s(\theta^0)' H_0^{-1}\left\{k_{st}k_{rt}\left[\frac{\partial S_t(\theta_t^1)}{\partial\theta} - \frac{\partial S_t(\theta^0)}{\partial\theta}\right]\right\}H_0^{-1}S_r(\theta^0) + o_P(1)\\
&= M_1 + M_2 + U_1 + L_1 + L_2 + L_3 + o_P(1),\ \text{say.}
\end{aligned} \tag{A.8}$$

We will show that the first two terms jointly determine the asymptotic mean, the third term determines the asymptotic variance, and the remainders are higher order terms. This is established by the following four lemmas.

LEMMA A.10. *Let $M_1$ be defined as in (A.8). Then*

$$M_1 - b^{-1/2}(1+p+q)\frac{var(\varepsilon_t^2)}{2}\left[2k(0) - \frac{1}{Tb}\sum_{j=-\lfloor Tb\rfloor}^{\lfloor Tb\rfloor}\left(1-\frac{|j|}{T}\right)k^2\left(\frac{j}{Tb}\right)\right] = o_P(1).$$

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

LEMMA A.11. *Let $M_2$ be defined as in (A.8). Then*

$$M_2 - b^{1/2}(1+p+q)\frac{var(\varepsilon_t^2)}{2}\left[1-\frac{1}{Tb}\sum_{j=-\lfloor Tb\rfloor}^{\lfloor Tb\rfloor}\left(1-\frac{|j|}{T}\right)k\left(\frac{j}{Tb}\right)\int_{-1}^{1}k\left(\frac{j}{Tb}+2u\right)du\right] = o_P(1).$$

LEMMA A.12. *Let $U_1$ and $\tilde{U}$ be defined as in (A.8) and (A.7) respectively. Then*

$U_1 = \tilde{U} + o_P(1)$.

LEMMA A.13. *Let $L_i$ be defined as in (A.8), where $i = 1,2,3$. Then $L_i = o_P(1)$ for $i = 1,2,3$.*

The proofs of Lemmas A.10–A.13 are tedious but straightforward. Therefore we omit them here. Detailed derivations are available upon request. ■

PROPOSITION A.5. *Under the conditions of Theorem 4, $2\tilde{U}/\sqrt{\tilde{B}} \xrightarrow{d} N(0,1)$.*

**Proof of Proposition A.5.** Let

$$R_s = -T^{-2}b^{1/2}\sum_{t=1}^{s-1}S_s(\theta^0)'H_0^{-1}S_t(\theta^0)\left(2Tk_{st}-\sum_{r=1}^{T}k_{rs}k_{rt}\right).$$

We apply Brown's (1971) martingale limit theorem, which states $\text{var}(2\tilde{U})^{-\frac{1}{2}}2\tilde{U} \xrightarrow{d} \text{N}(0,1)$ if

$$\text{var}(2\tilde{U})^{-1}\sum_{s=1}^{T}(2R_s)^2\mathbf{1}\left[|2R_s| > \eta\cdot\text{var}(2\tilde{U})^{\frac{1}{2}}\right]\to 0\ \ \forall\,\eta>0, \tag{A.9}$$

$$\text{var}(2\tilde{U})^{-1}\sum_{s=1}^{T}E[(2R_s)^2|\mathcal{F}_{s-1}]\xrightarrow{P} 1. \tag{A.10}$$

First,

$$\begin{aligned}
\text{var}\left(2\tilde{U}\right) &= 4T^{-4}b\sum_{s=2}^{T}\sum_{t=1}^{s-1}E\left[S_s(\theta^0)'H_0^{-1}S_t\left(\theta^0\right)S_t(\theta^0)'H_0^{-1}S_s\left(\theta^0\right)\right]\\
&\quad\times\left(2Tk_{st}-\sum_{r=1}^{T}k_{rs}k_{rt}\right)^2\\
&\quad+4T^{-4}b\sum_{s=1}^{T}\sum_{t_1=1}^{s-1}\sum_{t_2=1,t_1\neq t_2}^{s-1}E\left[S_s(\theta^0)'H_0^{-1}S_{t_1}(\theta^0)S_{t_2}(\theta^0)'H_0^{-1}S_s(\theta^0)\right]\\
&\quad\times\left(2Tk_{st_1}-\sum_r k_{rt_1}k_{rs}\right)\left(2Tk_{st_2}-\sum_r k_{rt_2}k_{rs}\right)\\
&= V_1+V_2,\ \text{say}.
\end{aligned} \tag{A.11}$$

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

For the first term, we have

$$V_1 = 4T^{-1}b^{-1}\sum_{j=1}^{T-1} C_2(j)\left(1-\frac{j}{T}\right)\left[2k\left(\frac{j}{Tb}\right)-\int_{-1}^{1}k(u)k\left(u+\frac{j}{Tb}\right)du\right]^2 + o(1)$$

$$= \left[var\left(\varepsilon_t^2\right)\right]^2(1+p+q)\int_0^1\left[2k(v)-\int_{-1}^{1}k(u)k(u+v)du\right]^2 dv$$

$$+4T^{-1}b^{-1}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)tr\left[\tilde{C}_2(j)\right]+o(1),$$

where

$$C_2(j) \equiv E\left[S_t(\theta^0)' H_0^{-1} S_{t-|j|}(\theta^0) S_{t-|j|}(\theta^0)' H_0^{-1} S_t(\theta^0)\right]$$

$$= \frac{\left[var\left(\varepsilon_t^2\right)\right]^2}{4}(1+p+q)+tr\left[\tilde{C}_2(j)\right]$$

and

$$\tilde{C}_2(j) = E\left\{S_t(\theta^0)S_t(\theta^0)' H_0^{-1}\left[S_{t-|j|}(\theta^0)S_{t-|j|}\left(\theta^0\right)' + \frac{var\left(\varepsilon_t^2\right)}{2}H_0\right]H_0^{-1}\right\}.$$

Given the fact that $\sum_{j=-\infty}^{\infty}\left|\tilde{C}_2(j)\right| \le C$, we have

$$V_1 = \left[var\left(\varepsilon_t^2\right)\right]^2(1+p+q)\int_0^1\left[2k(v)-\int_{-1}^{1}k(u)k(u+v)du\right]^2 dv + o(1). \qquad \textbf{(A.12)}$$

For the second term in (A.11), we have

$$V_2 = 4T^{-4}b\sum_{s=1}^{T}\sum_{t_1=1}^{s-1}\sum_{t_2=1,t_1\neq t_2}^{s-1} tr\left\{E\left[S_s(\theta^0)S_s\left(\theta^0\right)' H_0^{-1}S_{t_1}(\theta^0)S_{t_2}(\theta^0)' H_0^{-1}\right]\right\}$$

$$\times\left(2Tk_{st_1}-\sum_r k_{rt_1}k_{rs}\right)\left(2Tk_{st_2}-\sum_r k_{rt_2}k_{rs}\right)$$

$$= 4T^{-2}b^{-1}\sum_{s=1}^{T}\sum_{t_1=1}^{s-1}\sum_{t_2=1,t_1\neq t_2}^{s-1} tr\left[C_{22}(s-t_1,s-t_2)\right]$$

$$\times\left[2bk_{st_1}-\int_{-1}^{1}k(u)k\left(u+\frac{s-t_1}{Tb}\right)du\right]\left[2bk_{st_2}-\int_{-1}^{1}k(u)k\left(u+\frac{s-t_2}{Tb}\right)du\right]+o(1)$$

$$= o(1), \qquad \textbf{(A.13)}$$

where the fourth order cumulant function

$$C_{22}(j,l) = E\left\{\left[S_s(\theta^0)S_s(\theta^0)' + \frac{var\left(\varepsilon_s^2\right)}{2}H_0\right]H_0^{-1}S_{s-j}(\theta^0)S_{s-l}(\theta^0)' H_0^{-1}\right\}.$$

We have used the fact that $\sum_j\sum_l |C_{22}(j,l)| < \infty$, which can be obtained by using the mixing inequality. It follows that

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

$$var\left(2\tilde{U}\right)=\left[var\left(\varepsilon_t^2\right)\right]^2(1+p+q)\int_0^1\left[2k(v)-\int_{-1}^1 k(u)k(u+v)du\right]^2 dv+o(1)$$

from (A.12) and (A.13).

We now verify condition (A.9). Let $W_{st}=2bk_{st}-T^{-1}b\sum_{r=1}^T k_{rs}k_{rt}$. Then we have

$$\begin{aligned}\sum_{s=1}^T E(R_s^4) &= T^{-4}b^{-2}\Bigg\{\sum_{s=1}^T\sum_{t=1}^{s-1}E\left[S_s(\theta^0)'H_0^{-1}S_t(\theta^0)\right]^4 W_{st}^4\\
&\quad+\sum_{s=1}^T\sum_{t_1\neq t_2}E\left[S_s(\theta^0)'H_0^{-1}S_{t_1}(\theta^0)\right]^2\left[S_s(\theta^0)'H_0^{-1}S_{t_2}(\theta^0)\right]^2 W_{st_1}^2W_{st_2}^2\\
&\quad+\sum_{s=1}^T\sum_{t_1\neq t_2\neq t_3}E\left[S_s(\theta^0)'H_0^{-1}S_{t_1}(\theta^0)\right]^2\left[S_s(\theta^0)'H_0^{-1}S_{t_2}(\theta^0)S_{t_3}(\theta^0)'H_0^{-1}S_s(\theta^0)\right]\\
&\quad\times W_{st_1}^2W_{st_2}W_{st_3}\\
&\quad+\sum_{s=1}^T\sum_{t_1\neq t_2\neq t_3\neq t_4}E\Big[S_s(\theta^0)'H_0^{-1}S_{t_1}(\theta^0)S_{t_2}(\theta^0)'H_0^{-1}S_s(\theta^0)S_s(\theta^0)'H_0^{-1}S_{t_3}(\theta^0)\\
&\quad\times S_{t_4}(\theta^0)'H_0^{-1}S_s(\theta^0)\Big]W_{st_1}W_{st_2}W_{st_3}W_{st_4}\\
&=O(T^{-2}b^{-1})+O(T^{-1})+O(T^{-1}b^{-1})+O(b)\\
&=o(1).\end{aligned}$$

So that $[\text{var}(2\tilde{U})]^{-2}\sum_{s=1}^T E(R_s^4)\to 0$ and Condition (A.9) holds.

Next we verify Condition (A.10). Let $Q_s=\sum_{t=1}^{s-1}S_t(\theta^0)'\left(2Tb^2k_{sr}-\sum_t b^2k_{tr}k_{ts}\right)$. We have

$$\begin{aligned}E\left(R_s^2|\mathcal{F}_{s-1}\right)&=T^{-4}b^{-3}Q_sH_0^{-1}E\left[S_s(\theta^0)S_s(\theta^0)'|\mathcal{F}_{s-1}\right]H_0^{-1}Q_s'\\
&=T^{-4}b^{-3}Q_s\left\{H_0^{-1}E\left[S_s(\theta^0)S_s(\theta^0)'|\mathcal{F}_{s-1}\right]+\frac{var\left(\varepsilon_s^2\right)}{2}\mathbf{I}_{p+q+1}\right\}H_0^{-1}Q_s'\\
&\quad-\frac{var\left(\varepsilon_s^2\right)}{2}T^{-4}b^{-3}Q_sH_0^{-1}Q_s'\\
&=V_{1s}+R_{1s},\ \text{say,}\end{aligned}\tag{A.14}$$

where $\mathbf{I}_{p+q+1}$ is a $(p+q+1)\times(p+q+1)$ identity matrix. We further decompose

$$\begin{aligned}R_{1s}&=-\frac{var\left(\varepsilon_s^2\right)}{2}T^{-4}b^{-3}\left\{Q_sH_0^{-1}Q_s'-E\left[Q_sH_0^{-1}Q_s'\right]\right\}\\
&\quad-\frac{var\left(\varepsilon_s^2\right)}{2}T^{-4}b^{-3}E\left[Q_sH_0^{-1}Q_s'\right]\\
&=R_{2s}+\left[\frac{var\left(\varepsilon_s^2\right)}{2}\right]^2T^{-2}b^{-1}\sum_{r=1}^T(1+p+q)W_{sr}^2.\end{aligned}\tag{A.15}$$

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

Then we write

$$R_{2s} = -\frac{var\left(\varepsilon_s^2\right)}{2} T^{-2} b^{-1} \sum_{r=1}^{s-1} \left[ S_r(\theta^0)' H_0^{-1} S_r(\theta^0) - E\left( S_r\left(\theta^0\right)' H_0^{-1} S_r(\theta^0)\right)\right]$$

$$\times W_{sr}^2 - var\left(\varepsilon_s^2\right) T^{-2} b^{-1} \sum_{r_1=1}^{s-1} \sum_{r_2=1}^{r_1} S_{r_1}\left(\theta^0\right)' H_0^{-1} S_{r_2} W_{sr_1} W_{sr_2}$$

$$= V_{2s} + V_{3s}, \text{ say.} \tag{A.16}$$

It follows from A.11–A.16 that $\sum_{s=1}^T \{E[(2R_s)^2|\mathcal{F}_{s-1}] - E[(2R_s)^2]\} = \sum_{i=1}^3 \sum_{s=1}^T 4V_{is} - V_2 + o(1)$. It suffices to show Lemmas A.14–A.16, which imply $E|\sum_{s=1}^T E[(2R_s)^2|\mathcal{F}_{s-1}] - E[(2R_s)^2]|^2 = o(1)$. Thus, Condition (A.10) holds, and so $2\tilde{U}/\sqrt{\tilde{B}} \xrightarrow{d} N(0,1)$ by Brown's (1971) theorem. ■

LEMMA A.14. *Let $V_{1s}$ be defined as in (A.14). Then* $E\left(\sum_{s=1}^T V_{1s}\right)^2 = o(1)$.

**Proof of Lemma A.14.** Let

$$\Omega(S_{r_1}, S_s, S_{r_2})$$

$$= S_{r_1}(\theta^0)' \left[ H_0^{-1} E\left[ S_s(\theta^0) S_s(\theta^0)' | \mathcal{F}_{s-1}\right] + \frac{var(\varepsilon_s^2)}{2} \mathbf{I}_{p+q+1}\right] H_0^{-1} S_{r_2}(\theta^0) W_{sr_1} W_{sr_2}$$

$$+ S_{r_2}(\theta^0)' \left[ H_0^{-1} E\left[ S_{r_1}(\theta^0) S_{r_1}(\theta^0)' | \mathcal{F}_{r_1-1}\right] + \frac{var(\varepsilon_{r_1}^2)}{2} \mathbf{I}_{p+q+1}\right] H_0^{-1} S_s(\theta^0) W_{r_1 r_2} W_{r_1 s}$$

$$+ S_s(\theta^0)' \left[ H_0^{-1} E\left[ S_{r_2}(\theta^0) S_{r_2}(\theta^0)' | \mathcal{F}_{r_2-1}\right] + \frac{var(\varepsilon_{r_2}^2)}{2} \mathbf{I}_{p+q+1}\right] H_0^{-1} S_{r_1}(\theta^0) W_{r_2 s} W_{r_2 r_1},$$

where $\mathbf{I}_{p+q+1}$ is a $(p+q+1)\times(p+q+1)$ identity matrix. Then we have

$$E\left(\sum_{s=1}^T V_{1s}\right)^2 = CT^{-4} b^{-2} E\left[\sum_{s\neq r_1 \neq r_2} \Omega\left(S_{r_1}, \zeta_s, \zeta_{r_2}\right)\right]^2$$

$$= O\left(T^{-1} b^{-1}\right) = o(1),$$

where we have used Lemma A(i) of Hjellvik, Yao, and Tjøstheim (1998). ■

LEMMA A.15. *Let $V_{2s}$ be defined as in (A.16). Then* $E\left(\sum_{s=1}^T V_{2s}\right)^2 = o(1)$.

**Proof of Lemma A.15.**

$$E\left(\sum_{s=1}^T V_{2s}\right)^2$$

$$= T^{-4} b^{-2} \left[\frac{var(\varepsilon_s^2)}{2}\right]^2 \sum_{s=1}^T \sum_{r=1}^{s-1} E\left\{ S_r(\theta^0)' H_0^{-1} S_r(\theta^0) - E\left[ S_r(\theta^0)' H_0^{-1} S_r(\theta^0)\right]\right\}^2 W_{sr}^4$$

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

$$
\begin{aligned}
&+T^{-4}b^{-2}\left[\frac{var(\varepsilon_s^2)}{2}\right]^2 \\
&\times \sum_{s_1=1}^{T}\sum_{r_1=1}^{s_1-1}\sum_{s_2=1}^{T}\sum_{r_2=1}^{s_2-1} E\left\{\left\{S_{r_1}(\theta^0)' H_0^{-1} S_{r_1}(\theta^0) - E\left[S_{r_1}(\theta^0)' H_0^{-1} S_{r_1}(\theta^0)\right]\right\}\right. \\
&\times \left.\left\{S_{r_2}(\theta^0)' H_0^{-1} S_{r_2}(\theta^0) - E\left[S_{r_2}(\theta^0)' H_0^{-1} S_{r_2}(\theta^0)\right]\right\}\right\} W_{s_1 r_1}^2 W_{s_2 r_2}^2 \\
&= O\left(T^{-2}b^{-1}\right) + O\left(T^{-1}b^{-1}\right) \\
&= o(1).
\end{aligned}
$$

■

LEMMA A.16. *Let $V_{3s}$ be defined as in (A.16). Then* $E\left(\sum_{s=1}^{T} V_{3s}\right)^2 = o(1)$.

**Proof of Lemma A.16.**

$$
\begin{aligned}
E\left(\sum_{s=1}^{T} V_{3s}\right)^2 &= T^{-4}b^{-2}\left[\frac{var\left(\varepsilon_s^2\right)}{2}\right]^2 \sum_{s=1}^{T} E\left[\sum_{r_1=1}^{s-1}\sum_{r_2=1}^{r_1-1} S_{r_1}\left(\theta^0\right)' H_0^{-1} S_{r_2}\left(\theta^0\right)' W_{sr_1} W_{sr_2}\right]^2 \\
&\quad + T^{-4}b^{-2}\left[\frac{var(\varepsilon_s^2)}{2}\right]^2 \sum_{s_1=1}^{T}\sum_{s_2=1}^{s_1-1} E\left[\sum_{r_1=1}^{s_1-1}\sum_{r_2=1}^{r_1-1} S_{r_1}\left(\theta^0\right)' H_0^{-1} S_{r_2}\left(\theta^0\right)' W_{s_1 r_1} W_{s_1 r_2}\right] \\
&\quad \times \left[\sum_{r_3=1}^{s_2-1}\sum_{r_4=1}^{r_3-1} S_{r_3}\left(\theta^0\right)' H_0^{-1} S_{r_4}\left(\theta^0\right)' W_{s_2 r_3} W_{s_2 r_4}\right] \\
&= O\left(T^{-1}b^{-1}\right) = o(1).
\end{aligned}
$$

■

THEOREM A.10. *Under the conditions of Theorem 4,* $Q_2 \xrightarrow{P} 0$.

**Proof of Theorem A.10.**

$$
Q_2 = -2b^{1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} S_t\left(\theta^0\right)' \sqrt{T}\left(\bar{\theta} - \theta^0\right) = o_P(1),
$$

where we have used the fact $\sqrt{T}\left(\bar{\theta} - \theta^0\right) = O_P(1)$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} S_t\left(\theta^0\right) = O_p(1)$. ■

THEOREM A.11. *Under the conditions of Theorem 4,* $Q_3 \xrightarrow{P} 0$.

**Proof of Theorem A.11.**

$$
Q_3 = -b^{1/2}\sqrt{T}\left(\bar{\theta} - \theta^0\right)' \left[\frac{1}{T}\sum_{t=1}^{T} \frac{\partial S_t\left(\theta^1\right)}{\partial \theta}\right] \sqrt{T}\left(\bar{\theta} - \theta^0\right) = o_P(1),
$$

where $\theta^1$ lies between $\bar{\theta}$ and $\theta^0$, and we have used the facts that $\sqrt{T}\left(\bar{\theta} - \theta^0\right) = O_P(1)$ and $\theta^1 \xrightarrow{P} \theta^0$. ■

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

**Proof of Theorem 5.** Under the alternative hypothesis,

$$2(l_U - l_R) = 2T^{-1}\sum_{t=1}^{T}[\bar{l}_t\left(\theta_t^0\right) - \bar{l}_t(\theta^*)] + T^{-1}\sum_{t=1}^{T}\left[2S_t\left(\theta_t^0\right)' + \left(\bar{\theta}_t^c - \theta_t^0\right)'\frac{\partial S_t\left(\theta_t^1\right)}{\partial\theta}\right]$$
$$\times\left(\bar{\theta}_t^c - \theta_t^0\right) + T^{-1}\sum_{t=1}^{T}\left[-2S_t(\theta^*)' - (\bar{\theta}-\theta^*)'\frac{\partial S_t\left(\tilde{\theta}\right)}{\partial\theta}\right](\bar{\theta}-\theta^*)$$
$$= Q_0 + Q_4 + Q_5, \text{ say,}$$

where $\theta_t^1$ lies between $\bar{\theta}_t^c$ and $\theta_t^0$, and $\tilde{\theta}$ lie between $\bar{\theta}$ and $\theta^*$.

It is straightforward to show that $Q_4 = o_P(1)$ and $Q_5 = o_P(1)$ by Theorem A.12. Moreover, $T^{-1}b^{-1/2}\hat{A} = o(1)$. Hence, $T^{-1}b^{-1/2}LR = \{4(p+q+1)\int_0^1\left[2k(v) - \int_{-1}^1 k(u)k(u+v)du\right]^2 dv\}^{-1/2}Q_0 + o_P(1)$, where $Q_0$ converges to a strictly positive constant $2\left\{\int_0^1 E[\tilde{l}_0(u,\theta_u^0)]du - \int_0^1 E[\tilde{l}_0(u,\theta^*)]du\right\}$ by Theorem A.13. It follows that for any nonstochastic sequence $\{M_T = o(T\sqrt{b})\}$, we have $P(LR_2 > M_T) \to 1$. ∎

THEOREM A.12. *Under the conditions of Theorem 5,* $\bar{\theta} \xrightarrow{P} \theta^*$.

**Proof of Theorem A.12.** It is assumed that $\theta^* = \arg\max_{\theta\in\Theta} L(\theta) \equiv \arg\max_{\theta\in\Theta} \int_0^1 E[\tilde{l}_0(u,\theta)]du$ is the unique maximizer over $\Theta$. Then we need to show

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{s=1}^{T} l_s(\theta) - L(\theta)\right| \to 0. \quad \textbf{(A.17)}$$

To show (A.17), we need to show (a) for any $\theta \in \Theta$, $\frac{1}{T}\sum_{s=1}^{T} l_s(\theta) - L(\theta) \to 0$, and (b) $\frac{1}{T}\sum_{s=1}^{T} l_s(\theta)$ is equicontinuous in probability. For the pointwise law of large numbers in (a), we decompose

$$\frac{1}{T}\sum_{s=1}^{T} l_s(\theta) - L(\theta) = \left[\frac{1}{T}\sum_{s=1}^{T} l_s(\theta) - \frac{1}{T}\sum_{s=1}^{T}\tilde{l}_s\left(\frac{s}{T},\theta\right)\right] + \left[\frac{1}{T}\sum_{s=1}^{T}\left[\tilde{l}_s\left(\frac{s}{T},\theta\right) - E\tilde{l}_s\left(\frac{s}{T},\theta\right)\right]\right]$$
$$+\left[\frac{1}{T}\sum_{s=1}^{T} E\tilde{l}_s\left(\frac{s}{T},\theta\right) - \int_0^1 E[\tilde{l}_0(u,\theta)]du\right]$$
$$= A_1 + A_2 + A_3,$$

where

$$|A_1| \le \frac{1}{2T}\sum_s \frac{1}{h_s(\theta)}\left|X_s^2 - \tilde{X}_s^2\left(\frac{s}{T}\right)\right| + \frac{1}{2T}\sum_s \frac{\tilde{X}_s^2\left(\frac{s}{T}\right)}{h_s(\theta)\tilde{h}_s\left(\frac{s}{T},\theta\right)}\left|h_s(\theta) - \tilde{h}_s\left(\frac{s}{T},\theta\right)\right|$$
$$+\frac{1}{2T}\sum_s\left|\frac{h_s(\theta) - \tilde{h}_s\left(\frac{s}{T},\theta\right)}{\bar{h}_s(\theta)}\right|$$
$$\le \frac{1}{2T}\sum_s h_s^{-1}(\theta)\left(\frac{V_{s+1,q+1}}{T^{\varphi}} + U_{t+1,q+1}\right) + \frac{1}{2T}\sum_s \frac{\tilde{X}_s^2\left(\frac{s}{T}\right)}{h_s(\theta)\tilde{h}_s\left(\frac{s}{T},\theta\right)}$$

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

$$
\begin{aligned}
&\times \sum_{j=1}^{\infty} \xi_j(\theta)\left(\frac{V_{s-j+1,q+1}}{T^{\varphi}} + U_{s-j+1,q+1}\right) \\
&+ \frac{1}{2T}\sum_s \left| \frac{\sum_{j=1}^{\infty} \xi_j(\theta)\left(\frac{V_{s-j+1,q+1}}{T^{\varphi}} + U_{s-j+1,q+1}\right)}{\bar{h}_s(\theta)} \right| \\
\leq\; & \frac{1}{2T}\sum_s h_s^{-1}(\theta)\left(\frac{V_{s+1,q+1}}{T^{\varphi}} + U_{t+1,q+1}\right) + \frac{C}{2T}\sum_s \frac{\tilde{X}_s^2\left(\frac{s}{T}\right)}{h_s(\theta)\tilde{h}_s\left(\frac{s}{T},\theta\right)} \\
&\times \sum_{j=1}^{\infty} \rho_0^{j/q}\left(\frac{V_{s-j+1,q+1}}{T^{\varphi}} + U_{s-j+1,q+1}\right) \\
&+ \frac{C}{2T}\sum_s \left| \frac{\sum_{j=1}^{\infty} \rho_0^{j/q}\left(\frac{V_{s-j+1,q+1}}{T^{\varphi}} + U_{s-j+1,q+1}\right)}{\bar{h}_s(\theta)} \right| \\
=\; & o_P(1),
\end{aligned}
$$

where $V_{s,d}$ and $U_{s,d}$ are the $dth$ element of $V_s$ and $U_s$ defined in Lemma A.1.

To show $A_2 = o_P(1)$, we define

$$
\tilde{l}_s^{\tau}(u,\theta) = -\frac{1}{2}\left[\ln \tilde{h}_s^{\tau}(u,\theta) + \frac{\tilde{X}_s^2(u)}{\tilde{h}_s^{\tau}(u,\theta)}\right],
$$

$$
\tilde{h}_s^{\tau}(u,\theta) = \xi_0(\theta) + \sum_{j=1}^{\tau} \xi_j(\theta)\tilde{X}_{s-j}^2(u),
$$

where $\tau$ is some deterministic function of $T$. We decompose

$$
\begin{aligned}
A_2 =\; & \left\{\frac{1}{T}\sum_{s=1}^{T}\left[\tilde{l}_s^{\tau}\left(\frac{s}{T},\theta\right) - E\tilde{l}_s^{\tau}\left(\frac{s}{T},\theta\right)\right]\right\} + \left\{\frac{1}{T}\sum_{s=1}^{T}\left[\tilde{l}_s\left(\frac{s}{T},\theta\right) - \tilde{l}_s^{\tau}\left(\frac{s}{T},\theta\right)\right]\right\} \\
& - \left\{\frac{1}{T}\sum_{s=1}^{T}\left[E\tilde{l}_s\left(\frac{s}{T},\theta\right) - E\tilde{l}_s^{\tau}\left(\frac{s}{T},\theta\right)\right]\right\} \\
=\; & A_{21} + A_{22} + A_{23}.
\end{aligned}
$$

For the first term, we have

$$
\begin{aligned}
var(A_{21}) =\; & \frac{1}{T^2}\sum_{s,r} cov\left[\tilde{l}_s^{\tau}\left(\frac{s}{T},\theta\right), \tilde{l}_r^{\tau}\left(\frac{r}{T},\theta\right)\right] \\
=\; & \frac{1}{T^2}\sum_{|s-r|<2\tau} cov\left[\tilde{l}_s^{\tau}\left(\frac{s}{T},\theta\right), \tilde{l}_r^{\tau}\left(\frac{r}{T},\theta\right)\right] \\
& + \frac{1}{T^2}\sum_{|s-r|\geq 2\tau} cov\left[\tilde{l}_s^{\tau}\left(\frac{s}{T},\theta\right), \tilde{l}_r^{\tau}\left(\frac{r}{T},\theta\right)\right]
\end{aligned}
$$

$$\leq \frac{1}{T^2} \sum_{|s-r|<2\tau} \left[ E\tilde{l}_s^\tau \left(\frac{s}{T},\theta\right)^2 \right]^{1/2} \left[ E\tilde{l}_r^\tau \left(\frac{r}{T},\theta\right)^2 \right]^{1/2}$$
$$+\alpha^{1-\frac{1}{q}}(\tau) \left[ E\tilde{l}_s^\tau \left(\frac{s}{T},\theta\right)^{2q} \right]^{1/2q} \left[ E\tilde{l}_r^\tau \left(\frac{r}{T},\theta\right)^{2q} \right]^{1/2q}$$
$$= O\left(\frac{\tau}{T}\right) + O(\rho^{-\tau}) = o(1)$$

as $\tau/T \to 0$ and $\tau \to \infty$.

For the second term,

$$E|A_{22}| \leq \frac{1}{T}\sum_{s=1}^{T} E\left|\ln\tilde{h}_s\left(\frac{s}{T},\theta\right) - \ln\tilde{h}_s^\tau\left(\frac{s}{T},\theta\right)\right| + E\left|\frac{\tilde{X}_s^2\left(\frac{s}{T}\right)}{\tilde{h}_s\left(\frac{s}{T},\theta\right)} - \frac{\tilde{X}_s^2\left(\frac{s}{T}\right)}{\tilde{h}_s^\tau\left(\frac{s}{T},\theta\right)}\right|$$
$$\leq \frac{C}{T}\sum_{s=1}^{T} E\left|\tilde{h}_s\left(\frac{s}{T},\theta\right) - \tilde{h}_s^\tau\left(\frac{s}{T},\theta\right)\right|$$
$$+\frac{C}{T}\sum_{s=1}^{T}\sup_u [E\tilde{X}_s^4(u)]^{1/2} \sum_{j=\tau+1}^{\infty} \xi_j(\theta)\left[E\tilde{X}_s^4\left(\frac{s}{T}\right)\right]^{1/2}$$
$$= O\left(\rho^{(\tau+1)/q}\right) = o(1).$$

The last term can be shown in a similar way to the second term. Therefore, $A_2 = o_P(1)$.

Finally,

$$|A_3| \leq \frac{1}{T}\sum_{s=1}^{T}\left|E\tilde{l}_s\left(\frac{s}{T},\theta\right) - E\tilde{l}_{s-1}\left(\frac{s-1}{T},\theta\right)\right|$$
$$\leq \frac{1}{T}\sum_{s=1}^{T}\left|-\frac{1}{2}E\ln\left[\frac{\tilde{h}_s\left(\frac{s}{T},\theta\right)}{\tilde{h}_s\left(\frac{s-1}{T},\theta\right)}\right]\right| + \left|E\left[\frac{\tilde{X}_s^2\left(\frac{s}{T}\right)}{\tilde{h}_s\left(\frac{s}{T},\theta\right)} - \frac{\tilde{X}_s^2\left(\frac{s-1}{T}\right)}{\tilde{h}_s\left(\frac{s-1}{T},\theta\right)}\right]\right|$$
$$\leq \frac{C}{T}\sum_{s=1}^{T} E\sum_{j=1}^{\infty}\xi_j(\theta)\left|\tilde{X}_{s-j}^2\left(\frac{s}{T}\right) - \tilde{X}_{s-j}^2\left(\frac{s-1}{T}\right)\right|$$
$$+\frac{C}{T}\sum_{s=1}^{T} E\left|\frac{\tilde{X}_{s-j}^2\left(\frac{s}{T}\right) - \tilde{X}_s^2\left(\frac{s-1}{T}\right)}{\tilde{h}_s\left(\frac{s}{T},\theta\right)}\right| + \frac{C}{T}\sum_{s=1}^{T} E\left|\frac{\tilde{h}_s^2\left(\frac{s}{T},\theta\right) - \tilde{h}_s^2\left(\frac{s-1}{T},\theta\right)}{\tilde{h}_s\left(\frac{s}{T},\theta\right)}\right|$$
$$= O\left(\frac{1}{T}\right),$$

where we have used the fact that $E\left|\tilde{X}_s^2\left(\frac{s}{T}\right) - \tilde{X}_s^2\left(\frac{s-1}{T}\right)\right| = O\left(\frac{1}{T}\right)$.

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

Now we show the stochastic equicontinuity in (b). For any pair of $\theta_1$ and $\theta_2 \in \Theta$ with $\theta_1 \neq \theta_2$, there exist a $\tilde{\theta} \in \Theta$ that lies between $\theta_1$ and $\theta_2$, such that

$$
\begin{aligned}
\frac{|L(\theta_1)-L(\theta_2)|^2}{\|\theta_1-\theta_2\|^2} &\leq \frac{1}{2T}\sum_{s=1}^{T}\left\| \frac{\frac{\partial h_s(\tilde{\theta})}{\partial \theta}}{h_s(\tilde{\theta})} - \frac{X_s^2 \frac{\partial h_s(\tilde{\theta})}{\partial \theta}}{h_s^2(\tilde{\theta})} \right\|^2 \\
&\leq \frac{1}{2T}\sum_{s=1}^{T}\sum_{j=1}^{\infty} j\rho^{j-1}\left[\left|X_{s-j}^2 - \tilde{X}_{s-j}^2\left(\frac{s-j}{T}\right)\right| + \tilde{X}_{s-j}^2\left(\frac{s-j}{T}\right)\right] \\
&\quad + \frac{1}{2T}\sum_{s=1}^{T}\sum_{j=1}^{\infty} j\rho^{j-2}\left\{\left[X_s^2 - \tilde{X}_s^2\left(\frac{s}{T}\right)\right]\left[X_{s-j}^2 - \tilde{X}_{s-j}^2\left(\frac{s-j}{T}\right)\right]\right. \\
&\quad + \tilde{X}_s^2\left(\frac{s}{T}\right)\left[X_{s-j}^2 - \tilde{X}_{s-j}^2\left(\frac{s-j}{T}\right)\right] + \left[X_s^2 - \tilde{X}_s^2\left(\frac{s}{T}\right)\right]\tilde{X}_{s-j}^2\left(\frac{s-j}{T}\right) \\
&\quad \left. + \tilde{X}_s^2\left(\frac{s}{T}\right)\tilde{X}_{s-j}^2\left(\frac{s-j}{T}\right)\right\} \\
&= O_P(1).
\end{aligned}
$$

■

THEOREM A.13. *Under the conditions of Theorem 5,* $T^{-1}\sum_{t=1}^{T}\bar{l}_t\left(\theta_t^0\right) = \int_0^1 E[\tilde{l}_0(u,\theta_u^0)]du + o_P(1)$ *and* $T^{-1}\sum_{t=1}^{T}\bar{l}_t\left(\theta^*\right) = \int_0^1 E[\tilde{l}_0(u,\theta^*)]du + o_P(1)$.

**Proof of Theorem A.13.** First we note that replacing $\bar{l}_t(\cdot)$ with $l_t(\cdot)$ has asymptotically negligible impact by a similar proof to that of Theorem A.4. Therefore, $T^{-1}\sum_{t=1}^{T}\bar{l}_t\left(\theta^*\right) = T^{-1}\sum_{t=1}^{T} l_t\left(\theta^*\right) + o_P(1) = \int_0^1 E[\tilde{l}_0(u,\theta^*)]du + o_P(1)$ by (A.17).

Next, we decompose

$$
T^{-1}\sum_{t=1}^{T} l_t\left(\theta_t^0\right) = -\frac{1}{2T}\sum_{t=1}^{T}\ln h_t\left(\theta_t^0\right) - \frac{1}{2T}\sum_{t=1}^{T}\varepsilon_t^2 = -A_4 - A_5,
$$

where $A_5 \xrightarrow{P} \frac{1}{2}$ by the law of large numbers. We further decompose

$$
\begin{aligned}
A_4 &= \frac{1}{2}\int_0^1 E[\ln \tilde{h}_0(u,\theta_u^0)]du + \frac{1}{2T}\sum_{t=1}^{T}\left[\ln h_t\left(\theta_t^0\right) - \ln \tilde{h}_t\left(\frac{t}{T},\theta_t^0\right)\right] \\
&\quad + \frac{1}{2T}\sum_{t=1}^{T}\left[\ln \tilde{h}_t\left(\frac{t}{T},\theta_t^0\right) - E\ln \tilde{h}_t\left(\frac{t}{T},\theta_t^0\right)\right] \\
&\quad + \left[\frac{1}{2T}\sum_{t=1}^{T} E\ln \tilde{h}_t\left(\frac{t}{T},\theta_t^0\right) - \frac{1}{2}\int_0^1 E[\ln \tilde{h}_0\left(u,\theta_u^0\right)]du\right] \\
&= \frac{1}{2}\int_0^1 E[\ln \tilde{h}_0(u,\theta_u^0)]du + A_{41} + A_{42} + A_{43}
\end{aligned}
$$

and we shall show that $A_{4j} = o_P(1),\ j = 1,2,3.$

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press

For the first term, we have

$$|A_{41}| \leq \frac{1}{2T}\sum_{t=1}^{T}\left|\frac{h_t\left(\theta_t^0\right)-\tilde{h}_t\left(\frac{t}{T},\theta_t^0\right)}{h_t^*\left(\frac{t}{T},\theta_t^0\right)}\right| = O_P\left(\frac{1}{T^{\varphi}}\right),$$

where $h_t^*\left(\frac{t}{T},\theta_t^0\right)$ lies between $h_t\left(\theta_t^0\right)$ and $\tilde{h}_t\left(\frac{t}{T},\theta_t^0\right)$. We have used the mean value theorem and Lemma A.1.

To show $A_{42} = o_P(1)$, we define

$$\tilde{h}_t^{\tau}\left(\frac{t}{T},\theta_t^0\right) = \xi_0\left(\theta_t^0\right) + \sum_{j=1}^{\tau}\xi_j\left(\theta_t^0\right)\tilde{X}_{t-j}^2\left(\frac{u}{T}\right),$$

where $\tau$ is some deterministic function of $T$. We decompose

$$\begin{aligned}
A_{42} = &\left\{\frac{1}{2T}\sum_{t=1}^{T}\left[\ln\tilde{h}_t^{\tau}\left(\frac{t}{T},\theta_t^0\right) - E\ln\tilde{h}_t^{\tau}\left(\frac{t}{T},\theta_t^0\right)\right]\right\} \\
&+\left\{\frac{1}{2T}\sum_{t=1}^{T}\left[\ln\tilde{h}_t\left(\frac{t}{T},\theta_t^0\right) - \ln\tilde{h}_t^{\tau}\left(\frac{t}{T},\theta_t^0\right)\right]\right\} \\
&-\left\{\frac{1}{2T}\sum_{t=1}^{T}\left[E\ln\tilde{h}_t\left(\frac{t}{T},\theta_t^0\right) - E\ln\tilde{h}_t^{\tau}\left(\frac{t}{T},\theta_t^0\right)\right]\right\}.
\end{aligned}$$

The proof of $A_{42} = o_P(1)$ is rather similar to the above proof of $A_2 = o_P(1)$ and hence we omit the details here.

For the last term, we have

$$\begin{aligned}
|A_{43}| &\leq \frac{1}{2T}\sum_{t=1}^{T}\left|E\ln\tilde{h}_t\left(\frac{t}{T},\theta_t^0\right) - E\ln\tilde{h}_{t-1}\left(\frac{t-1}{T},\theta_{t-1}^0\right)\right| \\
&\leq \frac{C}{T}\sum_{t=1}^{T}E\sum_{j=1}^{\infty}\xi_j\left(\theta_t^0\right)\left|\tilde{X}_{t-j}^2\left(\frac{t}{T}\right) - \tilde{X}_{t-j}^2\left(\frac{t-1}{T}\right)\right| \\
&\quad + \frac{C}{T}\sum_{t=1}^{T}E\sum_{j=1}^{\infty}\left|\xi_j\left(\theta_t^0\right) - \xi_j\left(\theta_{t-1}^0\right)\right|\tilde{X}_{t-j}^2\left(\frac{t-1}{T}\right) \\
&= O\left(\frac{1}{T}\right),
\end{aligned}$$

where we have used the fact that $E\left|\tilde{X}_s^2\left(\frac{s}{T}\right) - \tilde{X}_s^2\left(\frac{s-1}{T}\right)\right| = O\left(\frac{1}{T}\right)$, $\left|\theta_t^0 - \theta_{t-1}^0\right| = O\left(\frac{1}{T}\right)$ when $t$ and $t-1$ are continuity points, $\left|\theta_t^0 - \theta_{t-1}^0\right| = O(1)$ when $t$ or $t-1$ is a discontinuity point, and the number of discontinuity points is finite and fixed. It follows that $T^{-1}\sum_{t=1}^{T}\bar{l}_t\left(\theta_t^0\right) = \int_0^1 E[\tilde{l}_0(u,\theta_u^0)]du + o_P(1)$. ■

**Proof of Theorem 6.** First we note that the parametric bootstrap ensures that in the bootstrap world, $\mathbb{H}_0$ always holds with $\theta = \bar{\theta}$, the global QMLE, and conditional on the random sample $\mathcal{X} = \{X_t\}_{t=1}^T$, the bootstrap standardized residual $\{\varepsilon_t^*\}$ is an i.i.d. sequence.

Define $l_U^*$ and $l_R^*$ in a similar way to $l_U$ and $l_R$ in (4.1) and (4.2) respectively with proper substitutions, namely, replacing $X_t$, $\bar{\theta}_t^c$, and $\bar{\theta}$ with $X_t^*$, $\bar{\theta}_t^{c*}$ and $\bar{\theta}^*$ respectively. We decompose

$$\begin{aligned}2Tb^{1/2}\left(l_U^* - l_R^*\right) &= b^{1/2}\sum_{t=1}^T \left[2S_t^*\left(\bar{\theta}\right)' + \left(\bar{\theta}_t^{c*} - \bar{\theta}\right)' \frac{\partial S_t^*\left(\theta_t^{1*}\right)}{\partial \theta}\right]\left(\bar{\theta}_t^{c*} - \bar{\theta}\right) \\ &\quad -2b^{1/2}\sum_{t=1}^T S_t^*\left(\bar{\theta}\right)'\left(\bar{\theta}^* - \bar{\theta}\right) - b^{1/2}\sum_{t=1}^T \left(\bar{\theta}^* - \bar{\theta}\right)' \frac{\partial S_t^*\left(\theta^{1*}\right)}{\partial \theta}\left(\bar{\theta}^* - \bar{\theta}\right) \\ &\equiv Q_1^* + Q_2^* + Q_3^*,\end{aligned}$$

where $\theta_t^{1*}$ lies between $\bar{\theta}_t^{c*}$ and $\bar{\theta}$, and $\theta^{1*}$ lies between $\bar{\theta}^*$ and $\bar{\theta}$. The first term $Q_1^*$ comes from nonparametric estimation based on the bootstrap sample $\mathcal{X}^*$, which determines the asymptotic distribution of $LR^*$ conditional on the observed random sample $\mathcal{X}$. The second and third terms come from parametric estimation, whose impact is negligible asymptotically. The proof of Theorem 6 consists of following three steps:

(1) Noting that conditional on $\mathcal{X}$, $\bar{\theta}^*$ is a $\sqrt{T}$ consistent estimator for $\bar{\theta}$, so we have

$$Q_2^* = -2b^{1/2}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^T S_t^*\left(\bar{\theta}\right)'\right]\sqrt{T}\left(\bar{\theta}^* - \bar{\theta}\right) = o_P(1) \text{ and}$$

$$Q_3^* = -b^{1/2}\sqrt{T}\left(\bar{\theta}^* - \bar{\theta}\right)'\left[\frac{1}{T}\sum_{t=1}^T \frac{\partial S_t^*\left(\theta^{1*}\right)}{\partial \theta}\right]\sqrt{T}\left(\bar{\theta}^* - \bar{\theta}\right) = o_P(1).$$

(2) Let $\bar{H}_* = E^*\left[\frac{\partial^2 l_t^*(\bar{\theta})}{\partial \theta \partial \theta'}|\mathcal{X}\right]$. Using a similar decomposition to (A.8), we have

$$\begin{aligned}Q_1^* &= -T^{-1}b^{-1/2}\sum_{t=1}^T S_t^*\left(\bar{\theta}\right)'\bar{H}_*^{-1} \\ &\quad \times \left[2k(0)\mathbf{I}_d - T^{-1}b\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st}^2 \frac{\partial S_s^*\left(\bar{\theta}\right)}{\partial \theta}\bar{H}_*^{-1}\right] S_t^*\left(\bar{\theta}\right) - 2T^{-1}b^{-1/2} \\ &\quad \times \sum_{t=1}^{Tb} S_t^*\left(\bar{\theta}\right)'\bar{H}_*^{-1}\left[2k\left(\frac{2t}{Tb}\right)\mathbf{I}_d - T^{-1}\sum_{s=t-\lfloor Tb\rfloor}^{t+\lfloor Tb\rfloor} k_{st}k\left(\frac{s+t}{Tb}\right)\frac{\partial S_s^*\left(\bar{\theta}\right)}{\partial \theta}\bar{H}_*^{-1}\right] S_t^*\left(\bar{\theta}\right) \\ &\quad -2T^{-1}b^{-1/2}\sum_{t>s=1}^T S_t^*\left(\bar{\theta}\right)'\bar{H}_*^{-1}\left[2bk_{ts}\mathbf{I}_d - T^{-1}b\sum_r k_{rt}k_{rs}\frac{\partial S_r^*\left(\bar{\theta}\right)}{\partial \theta}\bar{H}_*^{-1}\right] S_s^*\left(\bar{\theta}\right) \\ &\quad + o_P(1) \\ &= \frac{var^*\left(\varepsilon_t^{*2}|\mathcal{X}\right)\hat{A}}{2} + \tilde{U}^* + o_P(1),\end{aligned}$$

where $\hat{A}$ is defined in (4.3) and $\tilde{U}^* = -T^{-2}b^{1/2}\Sigma_{s=2}^{T}\Sigma_{t=1}^{s-1}S_s^*\left(\bar{\theta}\right)'\bar{H}_*^{-1}S_t^*\left(\bar{\theta}\right)\left(2Tk_{st} - \Sigma_{r=1}^{T}k_{rs}k_{rt}\right)$.

(3) Note that $E^*\left(\varepsilon_t^*|\mathcal{X}\right) = 0$, $E^*\left(\varepsilon_t^{*2}|\mathcal{X}\right) = 1$ and

$$
\begin{aligned}
&var^*\left(2\tilde{U}^*|\mathcal{X}\right)\\
&= 4T^{-4}b\sum_{s=2}^{T}\sum_{t=1}^{s-1}E^*\left[S_s^*\left(\bar{\theta}\right)'\bar{H}_*^{-1}S_t^*\left(\bar{\theta}\right)S_t^*\left(\bar{\theta}\right)'\bar{H}_*^{-1}S_s^*\left(\bar{\theta}\right)|\mathcal{X}\right]\left(2Tk_{st} - \Sigma_{r=1}^{T}k_{rs}k_{rt}\right)^2\\
&\quad + 4T^{-4}b\sum_{s=1}^{T}\sum_{t_1=1}^{s-1}\sum_{t_2=1,t_1\neq t_2}^{s-1}E^*\left[S_s^*\left(\bar{\theta}\right)'\bar{H}_*^{-1}S_{t_1}^*\left(\bar{\theta}\right)S_{t_2}^*\left(\bar{\theta}\right)'\bar{H}_*^{-1}S_s^*\left(\bar{\theta}\right)|\mathcal{X}\right]\\
&\quad \times\left(2Tk_{st_1} - \Sigma_{r=1}^{T}k_{rt_1}k_{rs}\right)\left(2Tk_{st_2} - \Sigma_{r=1}^{T}k_{rt_2}k_{rs}\right)\\
&= \left[var^*\left(\varepsilon_t^{*2}|\mathcal{X}\right)\right]^2(1+p+q)\int_0^1\left[2k(v) - \int_{-1}^{1}k(u)k(u+v)du\right]^2 dv + o(1).
\end{aligned}
$$

Then by verifying the conditions of Brown's (1971) CLT theorem, we can obtain that conditional on $\mathcal{X}$, the bootstrap test statistic $LR^* \to^d N(0,1)$. Combining Steps 1–3 yields the desired result. ■

https://doi.org/10.1017/S0266466614000942 Published online by Cambridge University Press