Detecting for Smooth Structural Changes in GARCH Models

Bin Chen
Department of Economics
University of Rochester
and
Yongmiao Hong
Department of Economics and Statistical Science
Cornell University,
Wang Yanan Institute for Studies in Economics (WISE)
Xiamen University

May 2009

We would like to thank Zongwu Cai, Oliver Linton and Suhasini Subba Rao 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, China. Correspondences: Bin Chen, Department of Economics, University of Rochester, Rochester, NY 14627; email: 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), Xiamen University, Xiamen 361005, China; email: yh20@cornell.edu.
Detecting for Smooth Structural Changes in GARCH Models

Abstract

Detecting and modelling structural changes in GARCH processes have attracted a great amount of attention in time series econometrics over the past few years. 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 nonparametric time-varying GARCH model and a constant parameter GARCH model, where the nonparametric time-varying GARCH parameters are estimated by a local quasi-maximum likelihood estimator and the constant GARCH parameters are estimated by a standard QMLE, via the likelihood criterion. The test does not require any prior information about the alternatives of structural changes. The test statistic 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 parametric bootstrap procedure is employed to improve the finite sample performance of the proposed test and the simulation study highlights the merits of the test.

JEL Classifications: C1, C4, E0.

Key words: GARCH, Kernel, Model stability, Parameter constancy, QMLE, Smooth structural change
1. INTRODUCTION

Since the seminal works by Engle (1982) and Bollerslev (1986), various ARCH and GARCH models have been commonly used to capture volatility dynamics of macroeconomic and financial time series. However, underlying all these models is the key assumption of stationarity. Given the changing pace of the underlying economic mechanism and technological progress, modeling economic variables over a long time horizon under the stationarity assumption may not be suitable. It is quite 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 are “shocks” induced by institutional changes, such as changes of exchange rate systems from the fixed exchange rate mechanism to the 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 empirical studies. For example, Andreou and Ghysels (2002) examine the change-point hypothesis in volatility dynamics of international stock market indices and foreign exchange returns and find multiple breaks associated with the Asian and Russian financial crisis; Mikosch and Stårică (2004) apply their goodness-of-fit test to the S&P500 returns and detect structural changes related to shifts of unconditional variance.

Model stability is crucial for statistical inference, forecasts, and any sensible policy implications drawn from the model. In particular, ignoring structural changes in macroeconomic and financial time series can easily lead to spurious persistence in the conditional volatility parameters. Diebold (1986) and Lamoureux and Lastrapes (1990) are among the first to suggest that structural changes unaccounted for can yield Integrated GARCH or long memory effects. More recently, Mikosch and Stårică (2004) and Hillebrand (2005) provide some theoretical explanation for this phenomenon. The spurious IGARCH effects imply that shocks have a permanent impact on volatility and so current information remains relevant when forecasting the conditional variance for all horizons while for the short memory volatility process, shocks to variance do decay over time. 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.

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

Almost all existing change-point tests for GARCH models are constructed for abrupt changes. To our knowledge, the only exception is Amado and Teräsvirta’s (2008) test, which considers testing for a smooth transition type time varying structure of GARCH models. Smooth changes may be more realistic because volatility usually evolves over time in a continuous manner and volatility jumps are rare. Empirical evidences show that various economic events, such as liberalization of emerging markets, integration of world equity markets, changes in exchange rate or interest rate regimes, may lead to structural changes in volatility models. 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 takes time for the market to achieve some census.

Recently, time-varying time series 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 all parameters change smoothly over time. Parametric specifications for time-varying parameters lead to more efficient estimation if the underlying coefficient functions are indeed 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 (2006) assume that the variance of the process of interest can be decomposed into stationary and nonstationary components. 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) introduce a time-varying ARCH process for modeling the evolutionary behavior of volatility. The model is asymptotically locally stationary in the neighborhood of each point of time but is globally nonstationary. One advantage of this evolutionary nonparametric time-varying ARCH model is that little restriction is imposed on the functional forms of coefficients, except for the regularity condition that they evolve over time smoothly. Motivated by the flexibility of the nonparametric smooth time-varying ARCH model, we will first generalize it to a class of smooth time-varying GARCH models and derive the consistency and asymptotic normality of the nonparametric estimators for time-varying GARCH parameters. We then use the 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 stationary GARCH\((p, q)\) models, the time-varying GARCH model is not included as a special case in the time-varying ARCH\((\infty)\) class and the asymptotic analysis is much more involved. Thus, while our focus is on testing structural changes of GARCH parameters, our results on nonparametric estimation of time-varying GARCH parameters may have its own independent interest. Moreover, we study the asymptotic properties of the nonparametric estimators in both interior and boundary regions and find that the asymptotic biases have different convergence rates. All existing works on time-varying ARCH processes (e.g., Dahlhaus and Subba Rao 2006, Fryzlewicz et al. 2008) only focus on estimation in the interior region.

This paper proposes a class of consistent tests for smooth structural changes as well as abrupt structural breaks in GARCH parameters with known or unknown change points. The idea is to estimate the smooth time-varying parameters of GARCH model by a local quasi-maximum likelihood estimation (LQMLE) and compare them with the standard QMLE parameter estimator. Under the null hypothesis of parameter constancy, the tests compare the log likelihood of the unrestricted nonparametric time-varying GARCH model and the restricted constant-parameter GARCH model. Compared with the existing tests for structural breaks in GARCH models in the literature, the proposed tests have a number of appealing features.

First, the proposed tests are consistent against a large class of smooth time-varying parameter alternatives. And they are also consistent against multiple sudden structural breaks in GARCH models with 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 tests have a null asymptotic N(0,1) distribution. The only inputs required are the QMLE and LQMLE parameter estimators. Hence, any standard econometric software can carry out computational implementation easily.

Fourth, the nonparametric time-varying parameter estimator is sensitive to the local behavior of time-varying parameters. Because only local information is employed in estimating parameters at each time point, the proposed tests have symmetric power against structural breaks that occur either in the first or second half of the sample period. This is different from some existing tests that have different powers against structural breaks that have same sizes but occur at different time points.

Fifth, unlike some existing tests for structural breaks in GARCH models, no trimming procedure is needed for the proposed tests. Moreover, the nonparametric estimator of the conditional variance parameters can provide insight into the volatility dynamics.
In Section 2, we introduce the time-varying GARCH framework and hypotheses of interest. Section 3 proposes a LQMLE for the time-varying parameters in GARCH models, and establishes its consistency and asymptotic normality. Section 4 develops a nonparametric testing approach and the forms of test statistics. Section 5 derives their asymptotic null distribution and investigates their asymptotic power properties in both interior and boundary regions. In Section 6, a simulation study is conducted to examine the finite sample performance of the test via a bootstrap procedure. Section 7 provides concluding remarks. All mathematical proofs are collected in the appendix. A GAUSS code to implement the proposed tests is available from the authors upon request. Throughout the paper, $C$ denotes a generic bounded constant.

2. TIMY-VARYING GARCH MODEL AND HYPOTHESES OF INTEREST

Consider the following data generating process (DGP)

\[
\begin{align*}
X_t &= \sqrt{h_t^0} \varepsilon_t, \\
h_t^0 &= \alpha_0^0 + \sum_{j=1}^p \alpha_j^0 h_{t-j}^0 + \sum_{i=1}^q \beta_i^0 X_{t-j}^2, \\
\{\varepsilon_t\} &\sim \text{i.i.d.}(0,1),
\end{align*}
\]

(2.1)

where $X_t$ is a stochastic time series process, $\alpha_j$ and $\beta_j$ are possibly time-varying parameters, $t$ is the index of time, and $p$ and $q$ are the orders of the GARCH process. Let $\theta_t^0$ be the collection of parameters $\theta_t^0 = (\alpha_{0t}, \alpha_{1t}, \ldots, \alpha_{pt}, \beta_{1t}, \ldots, \beta_{qt})'$.

The above setup nests both stationary GARCH and time-varying GARCH processes. For example, if $\theta_t^0$ is not changing over time, we have a stationary GARCH($p, q$) process, whose asymptotic properties have been studied by Berkes, Horvath and Kokoszka (2003). Lee and Hansen (1994) and Lumsdaine (1996) establish the asymptotic theory of the QMLE of the GARCH(1,1) model when $\theta_t^0$ is a constant.

For GARCH processes with time-varying parameters, one example is the single break in the GARCH model

\[
\begin{align*}
X_t &= \sqrt{h_t^0} \varepsilon_t, \\
h_t^0 &= \begin{cases} \\
\alpha_0^0 + \sum_{j=1}^p \alpha_j^0 h_{t-j}^0 + \sum_{i=1}^q \beta_i^0 X_{t-j}^2, & \text{if } t \leq \tau, \\
\alpha_0^0 + \sum_{j=1}^p \alpha_j^0 h_{t-j}^0 + \sum_{i=1}^q \beta_i^0 X_{t-j}^2, & \text{otherwise},
\end{cases} \\
\{\varepsilon_t\} &\sim \text{i.i.d.}(0,1), \ 1 < \tau < T,
\end{align*}
\]

(2.2)

where $\tau$ is called the break point. Chu (1998) and Kulperger and Yu (2005) have used this model as an alternative to study the parameter constancy of GARCH models when $\tau$ is unknown.

Another example of time-varying GARCH processes is the smooth transition type time-varying GARCH models proposed by Amado and Teräsvirta (2008). They consider both addi-
tive and multiplicative models, where the time-varying components are included in the classical GARCH models in different forms; namely

$$
\begin{align*}
X_t &= \sqrt{h_t^0} \varepsilon_t, \\
h_t^0 &= \alpha_0^0 + \sum_{j=1}^p \alpha_j^0 h_{t-j}^0 + \sum_{i=1}^q \beta_i^0 X_{t-j}^2 + \left(\alpha_{01}^0 + \sum_{j=1}^p \alpha_{j1}^0 h_{t-j}^0 + \sum_{i=1}^q \beta_{j1}^0 X_{t-j}^2\right) G(t), \\
\{\varepsilon_t\} &\sim \text{i.i.d.}(0,1),
\end{align*}
$$

(2.3)

and

$$
\begin{align*}
X_t &= \sqrt{h_t^0} \varepsilon_t, \\
h_t^0 &= \left(\alpha_0^0 + \sum_{j=1}^p \alpha_j^0 h_{t-j}^0 + \sum_{i=1}^q \beta_i^0 X_{t-j}^2\right) [1 + \delta G(t)], \\
\{\varepsilon_t\} &\sim \text{i.i.d.}(0,1),
\end{align*}
$$

(2.4)

where $G(t)$ is a smooth transition function of time. A choice of $G(t)$ is a logistic function, namely

$$
G(t) = \left\{1 + \exp \left[-\gamma (t - c)\right]\right\}^{-1},
$$

where $c$ and $\gamma$ are scalar parameters governing the threshold and speed of transition.

To cover a wide range of possibilities and inspired by the flexibility of local stationary GARCH processes, we do not assume any parametric functional form for $\theta_t$. Instead, we assume that $\theta_t$ is an unknown smooth function of time in form of

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

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

$$
\begin{align*}
X_t &= \sqrt{h_t^0} \varepsilon_t, \\
h_t^0 &= \alpha_0^0 \left(\frac{t}{T}\right) + \sum_{j=1}^p \alpha_j^0 \left(\frac{t}{T}\right) h_{t-j} + \sum_{i=1}^q \beta_i^0 \left(\frac{t}{T}\right) X_{t-j}^2, \\
\{\varepsilon_t\} &\sim \text{i.i.d.}(0,1).
\end{align*}
$$

(2.5)

The DGP in (2.5) includes time-varying ARCH$(q)$ processes (Dahlhaus and Subba Rao 2006, Fryzlewicz et al. 2008) as a special case. We consider GARCH models in (2.5) because GARCH models are more flexible than ARCH models in capturing the volatility dynamics in time series and are more parsimonious than ARCH models in many applications. Parsimonious GARCH models are attractive in estimating and forecasting volatilities.

The specification that parameter $\theta(\cdot)$ is a function of ratio $t/T$ rather than time $t$ only is a common scaling scheme in the time series literature (see, e.g., Phillips and Hansen 1990, Robinson 1989, Dahlhaus and Subba Rao 2006 and Cai 2007). It might first appear a bit strange because
the time-varying parameter $\theta_t$ depends on the sample size $T$. The reason for this requirement is
that a nonparametric estimator for $\theta_t$ 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 $\theta_t$ at some fixed point $t$, even if some smoothness condition is imposed on $\theta_t$. The amount of
local information must increase suitably if the variance and bias of a nonparametric estimator
of $\theta_t$ are to decrease suitably. A convenient way to achieve this is to regard $\theta_t$ as ordinates of
smooth function $\theta_0(\cdot)$ on an equally spaced grid over $[0,1]$, which becomes finer as $T \to \infty$, and
then consider estimation of $\theta_0(\tau)$ at fixed points $\tau$.

A keen interest here is whether the parameter $\theta^0_t$ is changing over time. The null hypothesis
is
\[ H_0 : \theta^0_t = \theta \text{ for some constant vector } \theta \in \Theta \text{ and for all } t, \]
where $\Theta$ is a parameter space of $\theta_t$.

Under $H_0$, the DGP in (2.1) is a standard GARCH process. Hence the unknown constant
parameter vector $\theta$ can be consistently estimated by QMLE
\[
\hat{\theta} = \arg \max_{\theta \in \Theta} \frac{1}{T} \sum_{s=1}^{T} l_s(\theta), \quad (2.6)
\]
where $l_s(\theta)$ is the likelihood function; namely
\[
l_s(\theta) = -\frac{1}{2} \left[ \log h_s(\theta) + \frac{X^2_s}{h_s(\theta)} \right],
\]
\[
h_s(\theta) = \xi_0(\theta) + \sum_{j=1}^{\infty} \xi_j(\theta) X^2_{s-j}.
\]

Berkes et al. (2003) shows that the standard GARCH process has a unique representation and
the functions $\xi_j(u), 0 \leq j < \infty$, can be defined by recursion. If $q \geq p$, then

\[
\begin{align*}
\xi_0(\theta) &= \alpha_0 / \left[ 1 - (\beta_1 + \ldots + \beta_q) \right], \\
\xi_1(\theta) &= \alpha_1, \\
\xi_2(\theta) &= \alpha_2 + \beta_1 \xi_1(\theta), \\
\ldots \\
\xi_p(\theta) &= \alpha_p + \beta_1 \xi_{p-1}(\theta) + \ldots + \beta_{p-1} \xi_1(\theta), \\
\xi_{p+1}(\theta) &= \beta_1 \xi_p(\theta) + \ldots + \beta_p \xi_1(\theta), \\
\ldots \\
\xi_q(\theta) &= \beta_1 \xi_{q-1}(\theta) + \ldots + \beta_{q-1} \xi_1(\theta),
\end{align*}
\]
and if $q < p$, the preceding equations are replaced with

\[
\begin{align*}
\xi_0 (\theta) &= \alpha_0 / \left[ 1 - (\beta_1 + \ldots + \beta_q) \right], \\
\xi_1 (\theta) &= \alpha_1, \\
\xi_2 (\theta) &= \alpha_2 + \beta_1 \xi_1 (\theta), \\
\vdots & \quad \vdots \\
\xi_{q+1} (\theta) &= \alpha_{q+1} + \beta_1 \xi_q (\theta) + \ldots + \beta_q \xi_1 (\theta), \\
\xi_p (\theta) &= a_p + \beta_1 \xi_{p-1} (\theta) + \ldots \beta_q \xi_{p-q} (\theta).
\end{align*}
\] (2.8)

In general, if $j > \max (p, q)$, then

\[
\xi_j (\theta) = \beta_1 \xi_{j-1} (\theta) + \beta_2 \xi_{j-2} (\theta) + \ldots + \beta_q \xi_{j-q} (\theta).
\] (2.9)

In practice, we observe only $X_1, \ldots, X_T$ and the logarithm of the likelihood function in (2.6) can not be computed from the observed data, and so $\hat{\theta}$ is infeasible. Hence, we replace $l_s (\theta)$ with

\[
l_s^* (\theta) = -\frac{1}{2} \left[ \log h_s^* (\theta) + \frac{X_s^2}{h_s^* (\theta)} \right],
\] (2.10)

where

\[
h_s^* (\theta) = \xi_0 (\theta) + \sum_{j=1}^{s-1} \xi_j (\theta) X_{s-j}^2,
\]

and compute

\[
\hat{\theta}^* = \arg \max_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} l_s^* (\theta).
\] (2.11)

Among many others, Berkes et al. (2003) establish the consistency and asymptotic normality of both $\hat{\theta}$ and $\hat{\theta}^*$ under $H_0$, and Lee and Hansen (1994) and Lumsdaine (1996) derive the asymptotic properties of QMLE for the GARCH(1,1) model.

The alternative hypothesis $H_A$ is that $H_0$ is false. Under the alternative $H_A$, $\theta_s$ is time varying. Examples include the GARCH model with a single break in (2.2) or multiple breaks with known or unknown break points, Amado and Teräsvirta’s (2008) smooth transition GARCH models in (2.3) and (2.4), Dahlhaus and Subba Rao’s (2006) time-varying ARCH($q$) models, and the more general time-varying GARCH($p, q$) model in (2.5). We allow for abrupt changes, smooth changes and mixtures of them.

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 the GARCH model in (2.2). Amado
and Terösvirta (2008) use a $LM$ test against smooth transition GARCH alternatives in (2.3) and (2.4). Both tests specify certain parametric alternatives. While these tests have best power against the assumed alternative, no prior information about the true alternative is usually available for practitioners. Our main objective in this paper is to develop a consistent test for $H_0$ against $H_A$, using a new approach.

In a linear regression framework, Chen and Hong (2008) propose generalized Chow and generalized Hausman tests for smooth structural changes as well as abrupt structural breaks with known or unknown change points in regression models. The idea is to estimate the smooth time-varying parameters by local smoothing and compare them with the OLS parameter estimator via sums of squared residuals or fitted values. These tests are not applicable to test structural changes in GARCH models as GARCH models require different estimation and testing techniques and use different criterion functions. In this paper, we shall compare a constant parametric GARCH model with a nonparametric time-varying GARCH model via the likelihood criterion. Below, we first extend Dahlhaus and Subba Rao’s (2006) results on time-varying ARCH models and discuss how to estimate time-varying GARCH models by a LQMLE. Asymptotic properties of the LQMLE of time-varying GARCH$(p, q)$ models may have independent interests since no asymptotic results are available in the literature.

3. NONPARAMETRIC ESTIMATION OF TIME-VARYING GARCH COEFFICIENTS

Unrestricted nonstationarity may entail so much arbitrariness in the time dependent behavior of a 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 can not 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 the vicinity of each time point. This idea is similar to the notion 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 processes. A rigorous definition of local stationarity is introduced by Dahlhaus (1996a, b) which impose a smoothness condition in terms of the components in the spectral representation of the process. Heuristically we can say that a 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
global nonstationary behavior.

Here, the smoothness of the parameter function \( \theta_0(\cdot) \) guarantees that the time-varying GARCH
process displays a locally stationary behavior. In order to study the asymptotic properties of the
time-varying GARCH process \( X_t \) in (2.5), we introduce the stationary GARCH process \( \{ \tilde{X}_t(u) \} \)
associated with time-varying GARCH process at the fixed point \( u \in [0, 1] \) as follows:

\[
\begin{align*}
\tilde{X}_t(u) &= \sqrt{\tilde{h}_t^0(u)} \varepsilon_t, \\
\tilde{h}_t^0(u) &= \alpha_0(u) + \sum_{j=1}^{p} \alpha_j^0(u) \tilde{h}_{t-j}(u) + \sum_{i=1}^{q} \beta_i^0(u) \tilde{X}_{t-j}^2(u), \\
\{\varepsilon_t\} &\sim \text{i.i.d.}(0,1), \ t = 1, \ldots, T,
\end{align*}
\]

(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 approximated by the stationary process \( \tilde{X}_t^2(u) \) well (Subba Rao 2006). The
degree of the approximation depends on the rescaling factor \( T \) and the deviation \( \frac{t}{T} - u \). This
is formally stated in Lemma A.1 in the appendix.

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)

\[
\tilde{h}_t(u) = \xi_0(u) + \sum_{j=1}^{\infty} \xi_j(u) \tilde{X}_{t-j}^2(u)
\]

(3.2)

for all \( t \) with probability one under certain regularity conditions. The functions \( \xi_j \) are given in
(2.7)-(2.9), but here they are associated with a given point \( u \).

Let \( \Theta \) be the compact set

\[
\Theta = \left\{ \theta = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)' \in \mathbb{R}^{p+q+1} : \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1, \right. \\
0 < \rho \leq \min(\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q) \leq \max(\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q) \leq \bar{\rho} < 1 \}.
\]

For each \( u \in [0, 1] \), we assume that \( \theta_u \) is an interior point in \( \Theta \), where \( \theta_u = (\alpha_0(u), \alpha_1(u), \ldots, \alpha_p(u), \beta_1(u), \ldots, \beta_q(u))' \). Under \( \mathbb{H}_A \), the LQMLE to estimate \( \theta_t \),

\[
\hat{\theta}_t = \arg \max_{\theta \in \Theta} L_t(\theta) = \arg \max_{\theta \in \Theta} \frac{1}{T} \sum_{s=1}^{T} k_s l_s(\theta),
\]

(3.3)
where

\[
k_{st} = \frac{1}{b} k \left( \frac{s - t}{Tb} \right),
\]

\[
l_s(\theta) = -\frac{1}{2} \left[ \log h_s(\theta) + \frac{X_s^2}{h_s(\theta)} \right],
\]

\[
h_s(\theta) = \xi_0(\theta) + \sum_{j=1}^{\infty} \xi_j(\theta) X_{s-j}^2.
\]

the kernel \( k(\cdot) : [-1, 1] \rightarrow \mathbb{R}^+ \) is a prespecified symmetric bounded probability density, and \( b \equiv b(T) \) is a bandwidth such that \( b \rightarrow 0 \) and \( Tb \rightarrow \infty \) as \( T \rightarrow \infty \). For notational simplicity, we have suppressed the dependence of \( k_{st} \) on the sample size \( T \) and \( b \). Examples of \( k(\cdot) \) include the uniform kernel

\[
k(u) = \frac{1}{2} \mathbb{1}(|u| \leq 1),
\]

the Epanechnikov kernel

\[
k(u) = \frac{3}{4}(1 - u^2) \mathbb{1}(|u| \leq 1),
\]

and the quartic kernel

\[
k(u) = \frac{15}{16}(1 - u^2)^2 \mathbb{1}(|u| \leq 1),
\]

where \( \mathbb{1}(\cdot) \) is the indicator function.

The functions \( \xi_j(\theta), 0 \leq j < \infty \), are defined in (2.7)-(2.9). The estimator \( \hat{\theta}_t \) is regarded as an estimator of \( \theta_t = (\alpha_0 \left( \frac{t}{T} \right), \alpha_1 \left( \frac{t}{T} \right), ..., \alpha_p \left( \frac{t}{T} \right), \beta_1 \left( \frac{t}{T} \right), ..., \beta_q \left( \frac{t}{T} \right) ) \) or of \( \theta_u = (\alpha_0 (u), \alpha_1 (u), ..., \alpha_p (u), \beta_1 (u), ..., \beta_q (u)) \) where \( |u - \frac{t}{T}| < \frac{1}{T} \). We emphasize that \( \varepsilon_t \) needs not to be normally distributed although we use the normal density function in (3.3). This is an pseudo-likelihood approach.

The idea of using LQMLE is based on the idea of local fitting. It has been applied by Tibshirani and Hastie (1987) to the class of generalized linear models and to the proportional hazard model of Cox (1972). Fan, Farmen and Gijbels (1998) provide a unified approach to selecting a bandwidth and constructing confidence intervals in local MLE and apply it to least-squares nonparametric regression and to nonparametric logistic regression. There is a vast literature on applying the local likelihood method to the problem of density estimation or hazard rate estimation. See, for example, Hjort (1991) and Copas (1995).

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). We define the locally weighted likelihood of \( \tilde{X}_t^2(u) \) as

\[
\tilde{L}(u, \theta) = \frac{1}{T} \sum_{s=1}^{T} k_{st} \tilde{l}_s(u, \theta), \tag{3.4}
\]
where \( |u - \frac{t}{T}| < \frac{1}{T} \) and

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

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

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

\[
L(u, \theta) = CE \left[ \tilde{l}_0(u, \theta) \right] \tag{3.5}
\]

as \( T \to \infty, b \to 0, Tb \to \infty, |u - \frac{t}{T}| < \frac{1}{T} \), where \( C = 1, \) when \( T - Tb \geq t \geq Tb \); \( C = k_{1c} \equiv \int_{-c}^{1-c} k(u)du \), when \( t = cbT \) or \( T - cbT \), where \( 1 > c > 0 \). It is easy to see that \( L(u, \theta) \) is maximized by \( \theta_u \). By applying the extreme estimator lemma (e.g., Amemyia 1985, Theorem 4.1.1), we can establish the consistency of the estimator \( \hat{\theta}_t \). If \( b \to 0 \) and \( Tb \to \infty \) as \( T \to \infty \), we have

\[
\hat{\theta}_t \to^P \theta_u.
\]

Similar to (2.6), \( L_t(\theta) \) cannot be computed with the observed sample \( \{X_t\}_{t=1}^{T} \) so we have to replace \( L_t(\theta) \) with

\[
L^*_t(\theta) = \frac{1}{T} \sum_{s=1}^{T} k_{st}\tilde{l}_s^*(\theta), \tag{3.6}
\]

where

\[
\tilde{l}_s^*(\theta) = -\frac{1}{2} \left[ \log h_s^*(\theta) + \frac{X_s^2}{h_s^*(\theta)} \right],
\]

\[
h_s^*(\theta) = \xi_0(\theta) + \sum_{j=1}^{s-1} \xi_j(\theta) X_{s-j}^2.
\]

Then we define the feasible estimator

\[
\hat{\theta}_t^* = \arg \max_{\theta \in \Theta} L^*_t(\theta). \tag{3.7}
\]

In order to derive the asymptotic properties of \( \hat{\theta}_t^* \), we impose the following regularity conditions.

**Assumption A.1:** (i) There exists a \( \kappa \in (0, 1] \) and a constant \( K < \infty \), such that \( |\alpha_j^0(u) - \alpha_j^0(v)| \leq K|u - v|^\kappa \) and \( |\beta_j^0(u) - \beta_j^0(v)| \leq K|u - v|^\kappa \), where \( u, v \in C, C \) is the set where \( \alpha_j(\cdot) \) and \( \beta_j(\cdot) \) are continuous and \( \lambda(C^c) = 0 \), where \( \lambda(\cdot) \) is the Lebesgue measure and \( C^c \) is the complement of \( C \); (ii) \( \sup_u [\sum_{j=1}^{p} \alpha_j^0(u) + \sum_{j=1}^{q} \beta_j^0(u)] < 1 - \eta \) for some small \( \eta > 0 \), where
0 ≤ u ≤ 1.

Assumption A.2: εt is an i.i.d. (0, 1) sequence satisfying $E(|ε_t|^{4(1+δ)}) < ∞$ some $δ > 0$ and $\lim_{r→0} r^{-μ} P(ε_t^2 ≤ r) = 0$ for some $μ > 0$.

Assumption A.3: The parameter space

$$\Theta = \left\{ \theta = (α_0, α_1, ..., α_p, β_1, ..., β_q) \in \mathbb{R}^{1+p+q} : \sum_{i=1}^{p} α_i + \sum_{j=1}^{q} β_j < 1, \right\}$$

where $0 < ρ ≤ \min(α_0, α_1, ..., α_p, β_1, ..., β_q) ≤ \max(α_0, α_1, ..., α_p, β_1, ..., β_q) ≤ \bar{ρ} < 1$ is a compact set. For each $u ∈ [0, 1], θ_u ∈ \text{Int}(Θ)$, where $θ_u = (α_0(u), α_1(u), ..., α_p(u), β_1(u), ..., β_q(u))$.

Assumption A.4: The polynomials $α_1(u)x + α_2(u)x^2 + ... + α_p(u)x^p$ and $1 - β_1(u)x - β_2(u)x^2 - ... - β_q(u)x^q$ are coprimes on the set of polynomials with real coefficients for some given $0 ≤ u ≤ 1$.

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

Assumption A.6: The bandwidth $b = cT^{-λ}$ for $0 < λ < 1$ and $0 < c < ∞$.

Assumption A.7: Except for a fixed number of points on $[0, 1]$, $α_j(u)$ and $β_j(u)$ are three times differentiable with $\sup_u |\frac{∂α_j(u)}{∂u}| ≤ C$ for $j = 0, ..., p$ and $\sup_u |\frac{∂β_j(u)}{∂u}| ≤ C$ for $j = 1, ..., q$, and $l = 1, 2, 3$, where $C$ is some finite constant independent of $j$ and $l$.

Assumption A.1(i) imposes the β-Lipschitz continuity of $α_j(·)$ and $β_j(·)$, but we allow for a zero-measure set where $α_j(·)$ and $β_j(·)$ are discontinuous. Assumption A.1 is a sufficient condition that the stochastic process $\{X_t\}$ admits a time-varying state space representation as

$$X_t = A_t \left( \frac{t}{T} \right) X_{t-1} + B_t \left( \frac{t}{T} \right),$$

where $X_t = (h_t, ..., h_{t-q+1}, X_{t-1}^2, ..., X_{t-p+1}^2), B_t(·) = [α^0(·), 0, ..., 0] ∈ \mathbb{R}^{p+q-1}$ and $A_t(·)$ is a $(p+q-1) × (p+q-1)$ matrix. Assumption A.2 is imposed in Berkes et al. (2003). Assumptions A.2-A.4 guarantee that the parameter $θ$ can be uniquely identified. These are standard assumptions assumed by Berkes et al. (2003, 2004) and Kulperger and Yu (2005), among many others.

Assumption A.5 implies $\int_{-1}^{1} k(u)du = 1$, $\int_{-1}^{1} u k(u)du = 0$ and $\int_{-1}^{1} u^2 k(u)du < ∞$. All examples in Section 2 satisfy this assumption. It is possible to use kernel functions with infinite support, such as the Gaussian kernel $k(u) = \frac{1}{\sqrt{2π}} \exp(-\frac{1}{2}u^2)$ for $−∞ < u < ∞$. However, we only
use kernel functions with bounded support to simplify asymptotic analysis. Assumption A.6 is only used in Theorem 2 to derive the asymptotic bias and variance of the LQMLE.

We first state the consistency of $\hat{\theta}_t^*$.

**Theorem 1:** Suppose Assumptions A.1$-$A.6 hold. If $|u - \frac{t}{T}| < \frac{1}{T}$, we have $\hat{\theta}_t^* - \hat{\theta}_t \to^P 0$ and $\hat{\theta}_t^* \to^P \theta_u$ as $T \to \infty$.

Note that this holds even if $\varepsilon_t$ is not $i.i.d.N(0,1)$. This result generalizes Dahlhaus and Subba Rao (2006), who only consider time-varying ARCH processes. We note that unlike the case of stationary GARCH($p, q$) models, the time-varying GARCH model is not included in the time-varying ARCH($\infty$) class. Furthermore, different from the time-varying ARCH process, the Volterra expansions of time-varying GARCH process are very tedious. Instead, we rely on the stochastic recurrent relation (see, e.g., Bougerol and Picard 1992 and Subba Rao 2006) to show that the locally nonstationary time-varying GARCH process can be locally approximated by a stationary GARCH process.

Next, we derive the asymptotic normality of $\hat{\theta}_t^*$.

**Theorem 2:** Suppose Assumptions A.1$-$A.7 hold and $|u - \frac{t}{T}| < \frac{1}{T}$, where $u$ is a continuity point for $\alpha_j(\cdot)$ and $\beta_j(\cdot)$.

(i) If $t$ is in the interior, namely $bT < t < T - bT$, we have

$$\sqrt{bT} \left( \hat{\theta}_t^* - \theta_u \right) \to^d N \left( T^{\frac{1}{2}} b^{\frac{1}{2}} B_{u}, k_2 \left( \frac{K}{2} + 1 \right) I(u)^{-1} \right),$$

as $T \to \infty$, where $k_2 = \int_{-1}^{1} k^2(x) dx$, $\kappa = [E(\varepsilon_t^4) - 3]$, $I(u) = E \left[ \frac{\partial^2 L(\theta,u)}{\partial \theta \partial \theta^T} \right]$, and $B_{u} = \frac{1}{2} b^2 I(u)^{-1} \frac{\partial^2 \left[ \frac{\partial L(\theta,u)}{\partial \theta} \right]}{\partial \theta^2} \int_{-1}^{1} x^2 k(x) dx$.

(ii) If $t$ is in the left boundary, namely $t = cbT$ where $0 < c < 1$, we have

$$\sqrt{bT} \left( \hat{\theta}_t^* - \theta_u \right) \to^d N \left( T^{\frac{1}{2}} b^{\frac{1}{2}} B_{ul}, k_{2c} k_{1c}^{-2} \left( \frac{K}{2} + 1 \right) I(u)^{-1} \right),$$

as $T \to \infty$, where $k_{2c} = \int_{-c}^{1} k^2(x) dx$, $k_{1c} = \int_{-c}^{1} k(x) dx$, $B_{ul} = b k_{1c}^{-1} I(u)^{-1} \frac{\partial \left[ \frac{\partial L(\theta,u)}{\partial \theta} \right]}{\partial \theta} \int_{-c}^{1} x k(x) dx + \frac{1}{2} b^2 k_{1c}^{-1} I(u)^{-1} \frac{\partial^2 \left[ \frac{\partial L(\theta,u)}{\partial \theta} \right]}{\partial \theta^2} \int_{-c}^{1} x^2 k(x) dx$, and $\kappa$ and $I(u)$ are defined in (i).
(iii) If \( t \) is in the right boundary, namely \( t = T - cbT \) where \( 0 < c < 1 \), we have
\[
\sqrt{bT} \left( \hat{\theta}_t^* - \theta_u \right) \to^d N \left( T^{1/2} b^{1/2} B_{ur}, k_{2c} k_{1c}^{-2} \left( \frac{\kappa}{2} + 1 \right) I(u)^{-1} \right),
\]
as \( T \to \infty \), where
\[
B_{ur} = b k_{1c}^{-1} I(u)^{-1} \left( \frac{\partial^2 L(\theta, u)}{\partial \theta^2} \right) \int_{-1}^c xk(x)dx + \frac{1}{2} b^2 k_{1c}^{-1} I(u)^{-1} \left( \frac{\partial L(\theta, u)}{\partial \theta} \right) \int_{-1}^c x^2 k(x)dx
\]
and \( k_{2c}, k_{1c}, \kappa \) and \( I(u) \) are defined in (i) and (ii).

We can view \( \kappa \) as the excess kurtosis of \( \varepsilon_t \), which measures the departure from the normality in the higher moment. If \( E\varepsilon_t^4 = 3 \) as in the case of normally distributed \( \varepsilon_t \), then the asymptotic variance of interior points can be simplified to \( k_2 I(u)^{-1} \). The quantity \( I(u) \) can be viewed as a local Hessian matrix, and \( B_u \) is the asymptotic bias, which is caused by the time-varying property of GARCH parameters. We can see that \( \lim_{c \to 1} \int_{-c}^c x^2 k(x)dx = 0 \) and \( \lim_{c \to 1} \int_{-c}^c xk(x)dx = 1 \), \( \lim_{c \to 1} \int_{-c}^c x^2 k(x)dx = 1 \), and \( \lim_{c \to 1} \int_{-c}^c xk(x)dx = 1 \), and these limits are exactly the constant factors appearing in the asymptotic bias and variance for an interior point. Theorem 2 shows that although the LQMLE is consistent for both interior and boundary points, the asymptotic bias has different convergence rate. The asymptotic biases for interior points and boundary points are \( O(b^2) \) and \( O(b) \) respectively. Therefore, the LQMLE suffers from boundary effects. The reason is similar to that of the kernel estimator for linear regression models. We do not have symmetric data available in the boundary regions. The boundary problem of time-varying ARCH or GARCH models has never been studied as previous works (e.g., Dahlhaus and Subba Rao 2006) only focus on interior points. The boundary problem of time-varying linear regression models has been studied by Cai (2007).

Among other things, Theorem 2 could be used to construct confidence intervals for \( \hat{\theta}_t^* \).

4. NONPARAMETRIC TESTING

We shall 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 the existence of unknown break dates. We note that the convergence rate of the asymptotic bias in the boundary regions is slower than that of an interior point and the asymptotic variance at a boundary point tends to be larger. The latter is because on a smaller interval, fewer observations contribute to the estimator. These differences would complicate the form of the proposed test statistics.

To make the behavior of the LQMLE at boundary points similar to that at interior points, we follow Hall and Wehrly (1991) to reflect the data in the boundary regions, obtaining pseudodata
$X_t = X_{-t}$ for $-\lfloor Tb \rfloor \leq t \leq -2$, where $\lfloor Tb \rfloor$ denotes the integer part of $Tb$, and $X_t = X_{2T-t}$ for $T + 1 \leq t \leq T + \lfloor Tb \rfloor$. We use the overall data (i.e., the union of the original data and the pseudodata) to estimate $\hat{\theta}_t^*$. By construction, symmetric data are 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 a nonparametric density estimation framework and by Chen and Hong (2008) in a nonparametric regression estimation. With the QMLE $\hat{\theta}^*$ and the LQMLE $\hat{\theta}_t^*$ at hand, we shall compare a constant parametric GARCH model with a nonparametric time-varying GARCH model via the likelihood criterion. We consider two cases for our nonparametric testing depending on whether the standardized innovation $\varepsilon_t$ is i.i.d. $N(0, 1)$.

Case 1: $\{\varepsilon_t\} \sim$ i.i.d. $N(0, 1)$ is correctly specified.

Recall that under $\mathbb{H}_0$, we have a standard GARCH($p, q$) model and $\theta$ can be consistently estimated by the QMLE $\hat{\theta}^*$ in (2.11). Under the alternative $\mathbb{H}_A$, $\theta_t = \theta(t/T)$ is changing over time and can be consistently estimated by the LQMLE $\hat{\theta}_t^*$ in (3.7). With $\hat{\theta}_t^*$ and $\hat{\theta}^*$, we can construct a test using likelihood functions. The idea is to compare the log likelihood of the unrestricted nonparametric time-varying 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, giving the test its power against a wide range of alternatives. Let $l_u$ denote the log likelihood of the nonparametric time-varying GARCH model, that is,

$$l_u = \frac{1}{T} \sum_{t=1}^{T} l_t^* \left( \hat{\theta}_t^* \right),$$

(4.1)

where $\hat{\theta}_t^*$ is the nonparametric time-varying parameter estimator in (3.7). Let $l_r$ denote the log likelihood of the constant parameter model, that is,

$$l_r = \frac{1}{T} \sum_{t=1}^{T} l_t^* \left( \hat{\theta}^* \right),$$

(4.2)

where $\hat{\theta}^*$ is the QMLE estimator in (2.11). It is important to note that $l_u$ and $l_r$ are averages of log likelihoods of the original data, namely, $\{X_t\}_{t=1}^{T}$. The pseudodata are only used for estimating $\hat{\theta}_t^*$ in the boundary regions. Hence it will not affect the asymptotic distribution of test statistics. Intuitively, the use of the pseudodata only has impact on the boundary regions $[1, Tb] \cup [T - Tb, T]$ and its order of magnitude will vanish as $b \to 0$. However, it improves the finite sample
performance of the proposed tests. Meanwhile, we define that the score function
\[
S_t(\theta) \equiv \frac{\partial l_t}{\partial \theta} = \frac{1}{2} (\varepsilon_t^2 - 1) \frac{\partial \ln h_t}{\partial \theta}.
\]
We note that \(S_t(\theta)\) is a martingale difference sequence (MDS) no matter whether the distribution of \(\varepsilon_t\) is correctly specified or not. However, only under the correct distributional specification of \(\varepsilon_t\), the information matrix equality holds, namely
\[
E \left[ \frac{\partial S_t(\theta)}{\partial \theta} \right] = -E \left[ S_t(\theta) S_t(\theta)\right] = I(\theta).
\]

Our LR test statistic for \(H_0\) versus \(H_A\) is based on the comparison of \(l_u\) and \(l_r\):
\[
LR_1 = \frac{2T \sqrt{b}(l_u - l_r) - \hat{A}}{\sqrt{\hat{B}}},
\]
where
\[
\hat{A} = b^{-1/2} \left\{ 2(1+p+q)k(0) + \frac{1}{Tb} \sum_{j=-[Tb]}^{[Tb]} \hat{C}_1(j) \left( 1 - \frac{|j|}{T} \right) k^2 \left( \frac{j}{Tb} \right) + b \left[ 1 + \frac{1}{Tb} \sum_{j=-[Tb]}^{[Tb]} \hat{C}_1(j) \left( 1 - \frac{|j|}{T} \right) k \left( \frac{j}{Tb} \right) \int_{-1}^{1} k \left( \frac{j}{Tb} + 2u \right) du \right] \right\}
\]
and
\[
\hat{B} = \frac{4}{Tb} \sum_{j=1}^{T-1} \left( 1 - \frac{j}{T} \right) \hat{C}_2(j) \left[ 2k \left( \frac{j}{Tb} \right) - \int_{-1}^{1} k(u)k \left( u + \frac{j}{Tb} \right) du \right]^2,
\]
where
\[
\hat{C}_1(j) = \frac{1}{T - |j|} \sum_{t=|j|+1}^{T} \hat{S}_{t-|j|} \left( \hat{\theta} \right) \hat{I} \left( \hat{\theta} \right)^{-1} \frac{\partial S_t(\theta)}{\partial \theta} \hat{I} \left( \hat{\theta} \right)^{-1} \hat{S}_{t-|j|} \left( \hat{\theta} \right),
\]
\[
\hat{C}_2(j) = \frac{1}{T - |j|} \sum_{t=|j|+1}^{T} \hat{S}_{t-|j|} \left( \hat{\theta} \right) \hat{I} \left( \hat{\theta} \right)^{-1} \hat{S}_{t} \left( \hat{\theta} \right) \hat{S}_{t} \left( \hat{\theta} \right) \hat{I} \left( \hat{\theta} \right)^{-1} \hat{S}_{t-|j|} \left( \hat{\theta} \right),
\]
and \(\hat{S}_t(\hat{\theta})\), \(\hat{I}(\hat{\theta})\) are sample analogues of \(S_t(\theta)\), \(I(\theta)\) and \(\frac{\partial S_t(\theta)}{\partial \theta}\) respectively. Intuitively, \(\hat{A}\) and \(\hat{B}\) are approximately the mean and variance of \(2T \sqrt{b}(l_u - l_r)\). The third term of \(\hat{A}\) involving the factor \(b\) arises due to the use of the pseudodata, but it is proportional to \(b\) and will vanish to zero when \(T \to \infty\). It is a finite sample correction. For each given \(j\), \(\hat{C}_1(j)\) is a consistent
estimator of \( C_1(j) \) as \( T \to \infty \), where

\[
C_1(j) \equiv E \left[ S_{t+j} (\theta)^t I (\theta)^{-1} S_t' (\theta) I (\theta)^{-1} S_{t+j} (\theta) \right]
\]

\[
= -(1 + p + q) + \text{tr} \left[ \hat{C}_1(j) \right]
\]

and

\[
\hat{C}_1(j) = E \left\{ S_{t+j} (\theta) S_{t+j} (\theta)^t I (\theta)^{-1} \left[ S_t' (\theta) - I (\theta) \right] I (\theta)^{-1} \right\}.
\]

The function \( \hat{C}_1(j) \) may be viewed as the covariance function between \( S_{t+j} (\theta) S_{t+j} (\theta)^t I (\theta)^{-1} \) and \( S_t' (\theta) I (\theta)^{-1} \), which is generally nonzero when \( \{X_t\} \) is serially correlated. However, given the \( \beta \)-mixing condition on \( \{X_t\} \) under \( \mathbb{H}_0 \) (Carrasco and Chen, 2002), we have \( \sum_{j=-\infty}^{\infty} |\hat{C}_1(j)| < \infty \).

Similarly, for each \( j \), \( \hat{C}_2(j) \) is a consistent estimator of \( C_2(j) \) as \( T \to \infty \), where

\[
C_2(j) \equiv E \left[ S_t (\theta)^t I (\theta)^{-1} S_{t-j|1} (\theta) S_{t-j|1} (\theta)^t I (\theta)^{-1} S_t (\theta) \right]
\]

\[
= (1 + p + q) + \text{tr} \left[ \hat{C}_2(j) \right]
\]

and

\[
\hat{C}_2(j) = E \left\{ S_t (\theta) S_t (\theta)^t I (\theta)^{-1} \left[ S_{t-j|1} (\theta) S_{t-j|1} (\theta)^t + \frac{\text{var} (\varepsilon_t^2)}{2} I (\theta) \right] I (\theta)^{-1} \right\}.
\]

And like \( \hat{C}_1(j) \), given the \( \beta \)-mixing conditions on \( \{X_t\} \), we have \( \sum_{j=-\infty}^{\infty} |\hat{C}_2(j)| < \infty \). Consequently, we can replace \( \hat{A} \) and \( \hat{B} \) with the following simplified centering and scaling factors:

\[
\hat{A} = b^{-1/2} (1 + p + q) \left\{ 2k(0) - \frac{1}{Tb} \sum_{|j| \leq Tb} \left( 1 - \frac{|j|}{T} \right) k^2 \left( \frac{j}{Tb} \right) \right\}
\]

\[
+ b \left[ 1 - \frac{1}{Tb} \sum_{|j| \leq Tb} \left( 1 - \frac{|j|}{T} \right) k \left( \frac{j}{Tb} \right) \int_{-1}^{1} k \left( \frac{j}{Tb} + 2u \right) du \right]
\]

\[
= b^{-1/2} (1 + p + q) \left[ 2k(0) - \int_{-1}^{1} k^2(u) du \right] [1 + o(1)],
\]

and

\[
\hat{B} = 4 (1 + p + q) \frac{1}{Tb} \sum_{j=1}^{T-1} \left( 1 - \frac{j}{T} \right) \left[ 2k \left( \frac{j}{Tb} \right) - \int_{-1}^{1} k(u) k \left( u + \frac{j}{Tb} \right) du \right]
\]

\[
= 4 (1 + p + q) \int_{0}^{1} \left[ 2k(v) - \int_{-1}^{1} k(u) k(u + v) du \right]^2 dv + o(1).
\]

We emphasize that \( \hat{A} \) and \( \hat{B} \) do not depend on DGPs, so they are convenient to compute. Both \( l_u \) and \( l_r \) are outputs of estimation. Many statistic programs provide values of \( l_u \) and \( l_r \).
automatically. Hence, it is straightforward to compute the statistic.

**Case 2: \( \{\varepsilon_t\} \) is nonnormally distributed.**

There is a growing consensus that the disturbance \( \{\varepsilon_t\} \) may not be normal for financial data. Some fat tailed distribution may fit data better. With possible distributional misspecification of \( \{\varepsilon_t\} \), QMLE remains consistent but not efficient (see Theorems 1 and 2). The score function is still a MDS, but the information matrix equality does not hold generally. In particular,

\[
E \left[ S_t (\theta) S_t (\theta)'^T \right] = -\frac{\text{var} (\varepsilon_t^2)}{2} I (\theta) \equiv -\left( \frac{\kappa}{2} + 1 \right) I (\theta),
\]

where \( \kappa \) is the excess kurtosis of \( \varepsilon_t \), defined as \( \kappa = [E(\varepsilon_t^4) - 3] \). It measures the departure from the normality in the higher moment.

We can construct a robust generalized LR test

\[
LR_2 = \frac{2T \sqrt{b(l_u - l_v)}/(\hat{\kappa} + 1) - \hat{A}}{\sqrt{\hat{B}}},
\]

(4.5)

where \( \hat{A} \) and \( \hat{B} \) are the same as in (4.3) and \( \hat{\kappa} = \frac{1}{T} \sum_{t=1}^{T} (\hat{\varepsilon}_t^4 - 3) \) is an estimator for excess kurtosis.

Similarly, we can replace \( \hat{A} \) and \( \hat{B} \) with the simplified centering and scaling factors \( \tilde{A} \) and \( \tilde{B} \).

We note that although our focus is to test whether \( \theta_t^0 = \theta \) for some constant vector \( \theta \in \Theta \) and for all \( t \), our approach can be extended to test whether a subset of the parameters is constant; namely \( \theta_t^{0_1} = \theta_1 \) for some constant vector \( \theta_1 \in \Theta_1 \) and for all \( t \), where \( \Theta_1 \) is a subset of \( \Theta \) and \( \theta_t^0 = (\theta_t^{0_1}, \theta_t^{0_2})' \). An example is that we are only interested in whether ARCH coefficients are constant. There are two cases. If prior information restricts \( \theta_t^{0_2} \) to be some constant, it would be analogous to the "partial structural change" problem for the linear regression models where part of the regression coefficients may not be subject to structural changes and the interest is to test the constancy of the other part of the regression coefficients (see, e.g., Andrews, Lee and Ploberger 1996, Bai and Perron 1998). For this case, the null hypothesis and the alternative hypothesis are

\[ H_0 : \theta_t^0 = \theta \text{ for some constant vector } \theta \in \Theta \text{ and for all } t, \]

where \( \Theta \) is a parameter space of \( \theta_t \), and

\[ H_A : \theta_t^{0_1} \text{ is time varying and } \theta_t^{0_2} = \theta_2 \text{ for some constant vector } \theta_2 \in \Theta_2 \text{ for some } t, \]

where \( \Theta_2 \subset \Theta \), respectively. The second case is that no restriction is imposed on \( \theta_t^{0_2} \).\(^1\) Hence

\(^1\)An analogue of this case has been studied by Ang and Kristensen (2009) in a linear regression framework.
the null hypothesis and the alternative hypothesis are

\[ H_0 : \theta_0^t = \theta_1 \text{ for some constant vector } \theta_1 \in \Theta_1 \text{ for all } t, \]

where \( \Theta_1 \subset \Theta \), and

\[ H_A : \theta_0^t \text{ is time varying for some } t, \]

respectively. As a subset of the parameters may be time-varying, we can adopt the local profile QMLE method. For case 1, under \( H_A \), we first treat \( \theta_2 \) as known and estimate \( \theta_{t1} \) by

\[
\hat{\theta}_{t1}^* = \arg \max_{\theta_{t1} \in \Theta_1} \frac{1}{T} \sum_{s=1}^{T} k_s l_s^* (\theta_1|\theta_2),
\]

(4.6)

where

\[
l_s^* (\theta_1|\theta_2) = -\frac{1}{2} \left[ \log h_s^* (\theta_1|\theta_2) + \frac{X_s^2}{h_s^* (\theta_1|\theta_2)} \right].
\]

Then we obtain an estimator of the constant component \( \theta_2 \) by substituting the LQMLE \( \hat{\theta}_{t1}^* \) into the likelihood function; namely

\[
\hat{\theta}_2 = \arg \max_{\theta_2 \in \Theta_2} \frac{1}{T} \sum_{t=1}^{T} l_t^* (\theta_2|\hat{\theta}_{t1}),
\]

(4.7)

where

\[
l_t^* (\theta_2|\hat{\theta}_{t1}) = -\frac{1}{2} \left[ \log h_t^* (\theta_2|\hat{\theta}_{t1}) + \frac{X_t^2}{h_t^* (\theta_2|\hat{\theta}_{t1})} \right].
\]

Iterations between these two steps have to be employed until a certain convergence criterion is met.\(^2\) With proper estimators \( \hat{\theta} \) and \( [\hat{\theta}_{t1}^*, \hat{\theta}_2^*] \) at hand, we can compare two models via the likelihood criterion. Case 2 can be handled in a similar way, except that the local profile QMLE and LQMLE have to be applied under \( H_0 \) and \( H_A \) respectively. For simplicity, we shall focus on hypotheses of interest introduced in section 2.

### 5. ASYMPTOTIC PROPERTIES

We now state the asymptotic distribution of \( LR_1 \) and \( LR_2 \) under \( H_0 \).

**Theorem 3:** Suppose Assumptions A.1(ii)–A.6 hold. (i) If \( \varepsilon_t \sim N(0,1) \), then \( LR_1 \overset{d}{\to} N(0,1) \) under \( H_0 \) as \( T \to \infty \). (ii) If \( \varepsilon_t \) is nonnormally distributed, then \( LR_2 \overset{d}{\to} N(0,1) \) under \( H_0 \) as \( T \to \infty \).

Both the \( LR_1 \) and \( LR_2 \) tests have 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 tests do not require formulation of an alternative and are applicable when one has no prior information of the alternative. Moreover, the new tests do not require trimming data.

We require that \( b \to 0 \) and \( Tb \to \infty \) as implied by Assumption A.6. This is the standard condition for bandwidth and it covers the optimal rate \( b \propto T^{-\frac{1}{2}} \). As an important feature of \( LR_1 \) and \( LR_2 \), the use of the QML estimator in place of the true parameter \( \theta \) under \( H_0 \) has no impact on the limit distribution of \( LR_1 \) and \( LR_2 \). Intuitively, the parametric estimator \( \hat{\theta}^* \) converges to \( \theta \) faster than the nonparametric estimator \( \hat{\theta}_1^* \). Consequently, the asymptotic distributions of \( LR_1 \) and \( LR_2 \) are solely determined by the nonparametric estimator and are nuisance parameter free.

The main idea in the proofs of Theorem 3 is to perform the Hoeffding’s decomposition on the test statistics and then apply the martingale limit theorem (e.g., Brown 1971). In small samples, the distribution of \( LR_1 \) and \( LR_2 \) 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 in Section 6.

To investigate the asymptotic power properties of \( LR_1 \) and \( LR_2 \) under \( H_A \), we state the following theorem.

**Theorem 4:** Suppose Assumptions A.1–A.6 hold. (i) If \( \varepsilon_t \sim N(0,1) \), then for any sequence \( \{M_T = o(T^{\sqrt{b}})\} \), we have \( P(LR_1 > M_T) \to 1 \) under \( H_A \). (ii) If \( \varepsilon_t \) is nonnormally distributed, then for any sequence \( \{M_T = o(T^{\sqrt{b}})\} \), we have \( P(LR_2 > M_T) \to 1 \) under \( H_A \).

Assumption A.1(i) allows for both smooth structural changes and abrupt structural breaks with known or unknown break points. In other words, we permit \( \theta(\cdot) \) to have finitely many discontinuities. Hence, single structural break or multiple breaks with known or unknown break points, which are often considered in this literature, are special cases of model (2.2). For example, suppose \( \theta(\cdot) \) is a jump function, namely,

\[
\theta(\tau) = \begin{cases} 
(\alpha_0^1, \alpha_1^1, \ldots, \alpha_p^1, \beta_1^1, \ldots, \beta_q^1)', & \text{if } \tau \leq \tau_0, \\
(\alpha_0^2, \alpha_1^2, \ldots, \alpha_p^2, \beta_1^2, \ldots, \beta_q^2)', & \text{otherwise}.
\end{cases}
\]

Then we obtain the single break GARCH alternative considered in Chu (1995) when \( \tau_0 \) is unknown.

Theorem 4 suggests that the LR test is consistent against all alternatives to \( H_0 \), subject to the
regularity condition implied by Assumption A.8. 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 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 in practice.

6. FINITE SAMPLE PERFORMANCE

Theorem 3 provides the null asymptotic $N(0, 1)$ distribution of the LR test. Thus, one can implement our test for $H_0$ by comparing $LR$ with a $N(0, 1)$ critical value. However, like many other nonparametric tests in the literature, the size of $LR$ in finite samples may differ significantly from the prespecified asymptotic significance level. Therefore, we shall consider a parametric bootstrap procedure. To unify Cases 1 and 2 in Section 4, we consider this general statistic

$$
\tilde{LR} = \frac{\hat{Q}/ (\hat{k} + 1) - \tilde{A}}{\sqrt{B}},
$$

where $\hat{Q} = 2T\sqrt{b}(l_u - l_\tau)$ and $\tilde{A}$ and $\tilde{B}$ are defined in (4.4) and adopt the following parametric bootstrap procedure:

- Step (i): Obtain a bootstrap sample $X^b \equiv \{X^b_i\}_{i=1}^T$ from the estimated null GARCH model;
- Step (ii): Estimate the null model using the bootstrap sample $X^b$, and compute a bootstrap statistic $\tilde{LR}^b$ in the same way as $\tilde{LR}$, with $X^b$ replacing the original sample $X = \{X_i\}_{i=1}^T$;
- Step (iii): Repeat steps (i) and (ii) $B$ times to obtain $B$ bootstrap test statistics $\{\tilde{LR}^b_i\}_{i=1}^B$;
- Step (iv): Compute the bootstrap $p$-value $p_b \equiv B^{-1} \sum_{i=1}^B 1(\tilde{LR}^b_i > \tilde{LR})$. To obtain an accurate bootstrap $p$-value, $B$ must be sufficiently large.

The excess kurtosis $\hat{k}$ estimated from the original sample $X$ will be very close to the one estimated from the bootstrap sample $X^b$ under $H_0$, therefore the $\tilde{LR}$ statistic applies to both normally and nonnormally distributed cases. 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 testing for the correct specification of parametric conditional distributions and conditional density respectively. We state the consistency of the parametric bootstrap in the following theorem.

Theorem 5: Suppose Assumptions A.1(ii)–A.6 hold. Then

$$
\sup_{z \in \mathbb{R}} \left| P\left( \tilde{LR}^b \leq z \mid X \right) - \Phi(z) \right| = o_P(1),
$$
where $\Phi(z)$ is the cumulative distribution function of $N(0,1)$.

Theorem 5 shows that conditional on $X$, $\widetilde{LR}^b \to^d N(0,1)$ in probability as $T \to \infty$. The proof is similar to that of Theorem 3 and we need to use the fact that the parametric bootstrap ensures that in the bootstrap world, $H_0$ always holds. When the null hypothesis is true, the bootstrap procedure will lead to asymptotically correct size of the test, because $\widetilde{LR}^b$ converges in distribution to the $N(0,1)$ limiting distribution; when the null hypothesis is false, because the test statistic $\widetilde{LR}^b$ will converge to infinity in probability, whereas asymptotically the bootstrap critical value is still the same as that of $N(0,1)$, the bootstrap procedure has power.

The consistency of the parametric bootstrap does not indicate the degree of improvement of the parametric bootstrap upon the asymptotic distribution. Since $LR$ is asymptotically pivotal, it is possible that $\widetilde{LR}^b$ can achieve reasonable accuracy in finite samples. We shall examine the performance of the parametric bootstrap in our simulation study. We use $B = 100$ bootstrap iterations for each simulation iteration.

To examine the size of our test under $H_0$, we consider the following DGP:

**DGP S.1** [Standard GARCH(1,1)]:

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

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 on a personal computer.

To investigate the power of our test in detecting structural changes in GARCH models, we consider two alternatives:

**DGP P1** [Sudden break in a GARCH]:

$$
\begin{cases}
X_t = h_t^{1/2} \varepsilon_t \\
h_t = \begin{cases}
0.1 + 0.2X_{t-1}^2 + 0.4h_{t-1}, & \text{if } t \leq 0.5T, \\
0.3 + 0.4X_{t-1}^2 + 0.55h_{t-1}, & \text{otherwise}.
\end{cases} \\
\end{cases}
$$

**DGP P2** [Smooth transition GARCH]:

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

where $\varepsilon_t \sim i.i.d. N(0,1)$. DGP P2 is the smooth transition multiplicative GARCH(1,1) model proposed by Amado and Teräsvirta (2008).

For the proposed LR test, we use the quartic kernel and our simulation experience suggests that the choice of $k(\cdot)$ has little impact on the performance of tests. For simplicity, we choose $b = (1/\sqrt{12})T^{-\frac{1}{2}}$, 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 $T \to \infty$.

Under DGP S1, the LR test has good size. 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 is as small as 250, and increases to 68.6% when $T = 500$. Under DGP P2, 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.

To sum up, we observe that the LR test has good sizes in finite samples when the parametric bootstrap is used, and reasonable powers against both sudden structural break and smooth structural changes.

7. CONCLUSION

Modelling and detecting structural changes in GARCH processes have attracted a great amount of attention in time series econometrics. We have contributed to this literature by establishing the asymptotic properties of a nonparametric LQMLE for the time-varying GARCH models and proposing a new test for smooth structural changes as well as abrupt structural breaks in GARCH models. All existing works focus on the estimation of time-varying ARCH models, which are special cases of our time-varying GARCH models, and asymptotic results of LQMLE are only available for interior points, even for time-varying ARCH models. Our work fills this important gap in the literature. Moreover, our tests have intuitive appeal because they can be regarded as the generalization of the likelihood ration test from a parametric context to a nonparametric context. They have a convenient null asymptotic N(0,1) distribution, do not require trimming data, do not require prior information on the possible alternative, and are consistent against all smooth structural changes as well as multiple abrupt structural breaks in GARCH models. To overcome the adverse impact of the first stage nonparametric estimation of the time-varying parameters, we use the parametric bootstrap procedure, which provides reasonable sizes and powers for the proposed test in finite samples.
<table>
<thead>
<tr>
<th></th>
<th>$T=250$</th>
<th>$T=500$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td><strong>Size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DGP S0</td>
<td>.128</td>
<td>.058</td>
</tr>
<tr>
<td><strong>Power</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DGP P1</td>
<td>.620</td>
<td>.436</td>
</tr>
<tr>
<td>DGP P2</td>
<td>.242</td>
<td>.154</td>
</tr>
</tbody>
</table>
REFERENCES


Lamoureux, C.G. and W.D. Lastrapes (1990): "Persistence in Variance Structural Change and


