

Testing for a Unit Root with Near-Integrated Volatility*

H. Peter Boswijk

Department of Quantitative Economics, University of Amsterdam[†]

January 2000

Abstract

This paper considers tests for a unit root when the innovations follow a near-integrated GARCH process. We compare the asymptotic properties of the likelihood ratio statistic with that of the least-squares based Dickey-Fuller statistic. We first use asymptotics where the GARCH variance process is stationary with fixed parameters, and then consider parameter sequences such that the GARCH process converges to a diffusion process. In the fixed-parameter case, the asymptotic local power gain of the likelihood ratio test is only marginal for realistic parameter values. However, under near-integrated parameter sequences the difference in power is more pronounced.

1 Introduction

A well-known property of financial time series is that their conditional variance displays variation over time, such that persistent periods of high variation are followed by low-volatility periods. This phenomenon, known as *volatility clustering*, is modelled in the econometrics literature either by GARCH (generalized autoregressive-conditional heteroskedasticity) type models (see Bollerslev *et al.*, 1994, for an overview) or by stochastic volatility models, see e.g. Shephard (1996). When applied to daily financial returns data, both classes of models display a high degree of persistence, and hence a low degree of mean-reversion in the volatility process. Such processes are referred to as *near-integrated*, since their characteristic polynomial has a root close to but not necessarily equal to unity. Boswijk (1999) considers (quasi-) likelihood based tests for a unit root in the volatility process in exponential GARCH (EGARCH) models and stochastic volatility models.

In the present paper we study the effect of such near-integrated volatility processes on testing for an autoregressive unit root in the level of the process itself (instead of its volatility). This problem is relevant in finance, for example when models for the term structure of interest rates depend on the presence and degree of mean-reversion in the short rate. A typical model for the short rate is the one by Vasicek (1977), which is essentially a first-order autoregression with constant volatility. When applied

*This paper has benefitted from discussions with Anders Rahbek and André Lucas.

[†]Address for correspondence: Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, NL-1018 WB Amsterdam, The Netherlands. E-mail: peterb@fee.uva.nl.

to daily or weekly interest rates, the hypothesis of a unit root (i.e., no mean-reversion) often cannot be rejected, and a possible explanation of this is that least-squares-based tests are not powerful enough to discover the (weak) mean-reversion. Since interest rates clearly do not have a constant volatility, a likelihood-based testing procedure which takes this phenomenon into account might be expected to yield more efficient estimates and hence more powerful tests.

Previous work in this area is by Ling and Lee (1997, 1998) and Rahbek (1999), who consider tests for a unit autoregressive root in models with GARCH errors. They find that the maximum likelihood estimator of the mean-reversion parameter has a limiting distribution that is a weighted average of a Dickey-Fuller-type distribution and a normal distribution. They consider GARCH processes with fixed parameters in the stationarity region, whereas in this paper we study the case where the volatility parameters approach the unit root bound. Therefore, we consider parameter sequences such that the autoregressive root in the volatility process approaches unity as the sample size increases. This allows us to use the results of Nelson (1990) on continuous-time diffusion limits of GARCH processes. The present paper is also closely related to Hansen (1992b, 1995), who considers ordinary least-squares, generalized least-squares and adaptive estimation of regressions with non-stationary volatility.

The outline of the remainder of the paper is as follows. In Section 2, we define the model and hypothesis, and the parameter sequences that will be used in the asymptotic analysis. Section 3 analyses the likelihood function, the score and the information, and their asymptotic distribution under the relevant probability measures. We study the asymptotic distributions of the Dickey-Fuller test statistic, based on least-squares estimation, and the likelihood ratio test statistic, both under the null hypothesis and under local alternatives. Section 4 provides numerical evidence on the local power of these tests, and Section 5 concludes.

2 The Model

Consider a univariate first-order autoregressive process with GARCH(1,1) innovations:

$$\Delta X_t = \gamma(X_{t-1} - \mu) + \varepsilon_t, \quad t = 1, \dots, n, \quad (1)$$

$$\varepsilon_t = \sigma_t \eta_t, \quad (2)$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (3)$$

$$\eta_t \sim \text{i.i.d. } N(0, 1), \quad (4)$$

where $\Delta X_t = (X_t - X_{t-1})$, and where X_0 , ε_0 and σ_0^2 are fixed.

The parameter γ describes the degree of mean-reversion. If $-2 < \gamma < 0$, then X_t reverts back to its mean μ . The null hypothesis that we wish to test is the unit root hypothesis, or equivalently the no-mean-reversion hypothesis

$$\mathcal{H}_0 : \gamma = 0, \quad (5)$$

which is tested against the alternative $\gamma < 0$. The model (1) has a restricted constant term, such that under the null hypothesis the process does not contain a drift. Other specifications of the deterministic

component in X_t can be considered, including a restricted linear trend term (to test a random walk with drift against a trend-reverting autoregression), but this is not considered explicitly here. Similarly, the model can be extended to allow for more lags in (1).

The (non-negative) parameters ω , α and β characterize the dynamics of the volatility process. If $|\alpha + \beta| < 1$, then the variance reverts back to its mean $\sigma^2 = \omega/(1 - \alpha - \beta)$, and if $(\alpha + \beta) = 1$ then the variance follows a random walk (with drift if $\omega \neq 0$). The asymptotic distribution of the test statistics considered in the next section will depend on what we assume about the parameter of interest γ , but also on assumptions about the volatility parameters (ω, α, β) . We consider two alternative assumptions (in all cases $(\omega, \alpha, \beta) \geq 0$):

Assumption 1 For all $n \geq 1$, $\gamma_n = \kappa/n$ and $|\alpha + \beta| < 1$, with $(\kappa, \mu, \omega, \alpha, \beta)$ fixed.

Assumption 2 For all $n \geq 1$, $\gamma_n = \kappa/n$, $(\alpha_n + \beta_n) = 1 + \lambda/n$, $\omega_n = \varpi/n$ and $\alpha_n = \zeta/\sqrt{2n}$, with $(\kappa, \mu, \varpi, \lambda, \zeta)$ fixed.

Under Assumption 1, the process X_t is near-integrated with stationary volatility. The null hypothesis requires $\kappa = 0$, and values $\kappa \neq 0$ define the local alternatives. Under Assumption 2, the variance process is also near-integrated. One possible motivation for these parameter sequences is that the model (1)–(4) is viewed as a discrete-time approximation, for varying n but over a fixed time interval, of the continuous-time diffusion process defined below in Lemma 2, see Nelson (1990).

We conclude this section with two lemmas that describe the limiting behaviour of X_t under each of the two possible assumptions.

Lemma 1 Under Assumption 1, and as $n \rightarrow \infty$,

$$\left(\frac{1}{\sigma\sqrt{n}} \sum_{t=1}^{\lfloor sn \rfloor} \varepsilon_t, \frac{1}{\sqrt{n}} X_{\lfloor sn \rfloor} \right) \xrightarrow{\mathcal{L}} (W(s), U(s)), \quad (6)$$

where $W(s)$ a standard Brownian motion process on $[0, 1]$ and $U(s)$ is an Ornstein-Uhlenbeck process on $[0, 1]$:

$$dU(s) = \kappa[U(s) - \mu]ds + \sigma dW(s), \quad (7)$$

with $U(0) = 0$ and $\sigma^2 = \omega/(1 - \alpha - \beta)$.

The proof of this lemma is given in Ling and Li (1998, Theorem 3.3) for $\kappa = 0$, in which case the process $U(s)$ reduces to $\sigma W(s)$, a Brownian motion with variance σ^2 . This is extended to the case $\kappa \neq 0$ by writing $X_{\lfloor sn \rfloor}$ as a continuous functional of the partial sum of ε_t .

Lemma 2 Under Assumption 2, and as $n \rightarrow \infty$,

$$\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor sn \rfloor} \eta_t, \frac{1}{\sqrt{2n}} \sum_{t=1}^{\lfloor sn \rfloor} (\eta_t^2 - 1), \frac{1}{\sqrt{n}} X_{\lfloor sn \rfloor}, \sigma_{\lfloor sn \rfloor}^2 \right) \xrightarrow{\mathcal{L}} (W_1(s), W_2(s), Y(s), V(s)), \quad (8)$$

where $(W_1(s), W_2(s))$ is a standard bivariate Brownian motion process on $[0, 1]$, and $(Y(s), V(s))$ is the solution to the system of stochastic differential equations

$$dY(s) = \kappa[Y(s) - \mu]ds + V(s)^{1/2}dW_1(s), \quad (9)$$

$$dV(s) = [\lambda V(s) + \varpi]ds + \zeta V(s)dW_2(s), \quad (10)$$

with $Y(0) = 0$ and $V(0) = \sigma_0^2$.

The proof of this lemma follows from Nelson (1990, Theorem 2.2 and Section 2.3). The difference again is that Nelson considers the case $\kappa = 0$, but the extension of his proof to the present case is straightforward. If the process $Y(s)$ is discretely sampled at times $s = t/n$, and we define $X_t = \sqrt{n}Y(t/n)$, $t = 0, 1, \dots, n$, then the actual process generating X_t may be approximated by (1)–(4) under Assumption 2; the approximation error will vanish as $n \rightarrow \infty$, see Nelson (1990). An alternative (Euler) approximation would lead to a discrete-time stochastic volatility-type model, but we choose to work with the GARCH model because it has a closed-form expression for the likelihood function, which simplifies the construction of likelihood-based test statistics considered in the next section.

3 Likelihood Analysis

The statistical analysis of model (1)–(4) is given in Ling and Li (1997, 1998) and Rahbek (1999), but will be briefly repeated here.

It will be convenient to introduce the parameter vector $\delta = (\gamma, -\gamma\mu)'$ and $Z_t = (X_{t-1}, 1)'$, such that (1) becomes $\Delta X_t = \delta' Z_t + \varepsilon_t$, and the null hypothesis is $\mathcal{H}_0 : \delta = 0$. The full parameter vector is $\theta = (\delta', \omega, \alpha, \beta)'$, and the log-likelihood function is

$$\log L(\theta) = \sum_{t=1}^n \log L_t(\theta) = \sum_{t=1}^n -\frac{1}{2} \left(\log 2\pi + \log \sigma_t(\theta)^2 + \frac{\varepsilon_t(\delta)^2}{\sigma_t(\theta)^2} \right), \quad (11)$$

where $\varepsilon_t(\delta) = \Delta X_t - \delta' Z_t$, and where it should be noted that $\sigma_t(\theta)$ depends on the volatility parameters (ω, α, β) , but also, via ε_{t-1}^2 , on the regression parameters δ . The log-likelihood is conditional on σ_0 and ε_0 , which are not observed. In practice, they may be replaced by suitable estimates (we will assume that this has an asymptotically negligible effect).

The unrestricted parameter space for θ is $\Theta = \mathbb{R}^2 \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+$, and the restricted parameter space defined by the null hypothesis is $\Theta_0 = (0, 0) \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+$. Define $\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} \log L(\theta)$ and $\tilde{\theta} = \operatorname{argmax}_{\theta \in \Theta_0} \log L(\theta)$, the unrestricted and restricted maximum likelihood estimators, respectively. The likelihood ratio statistic for the null hypothesis is

$$LR = -2 \left(\log L(\tilde{\theta}) - \log L(\hat{\theta}) \right). \quad (12)$$

We will compare the performance of this test with that of Dickey and Fuller's (1981) F -statistic:

$$\Phi_1 = \frac{n-2}{2} \frac{\sum_{t=1}^n \Delta X_t Z_t' (\sum_{t=1}^n Z_t Z_t')^{-1} \sum_{t=1}^n Z_t \Delta X_t}{\sum_{t=1}^n (\Delta X_t - \hat{\delta}'_{LS} Z_t)^2}, \quad (13)$$

with $\hat{\delta}_{LS} = (\sum_{t=1}^n Z_t Z_t')^{-1} \sum_{t=1}^n Z_t \Delta X_t$; this is a monotonic transformation of the likelihood ratio statistic for \mathcal{H}_0 under the restriction $\alpha = \beta = 0$ (i.e., homoskedastic innovations).

Define the score vector $S(\theta) = \partial \log L(\theta) / \partial \theta$ and the observed information matrix $J(\theta) = -\partial^2 \log L(\theta) / \partial \theta \partial \theta'$. We will make use of the conventional Taylor series expansions which result in

$$LR = (n^{-1} \nu' + S(\theta_0)' J(\theta_0)^{-1} E_1) [E_1' J(\theta_0)^{-1} E_1]^{-1} (E_1' J(\theta_0)^{-1} S(\theta_0) + n^{-1} \nu) + o_P(1), \quad (14)$$

where θ_0 is the true value (which is a sequence under Assumption 1 or 2), $E_1 = [I_2 : 0]'$ is a selection matrix such that $\delta = E_1' \theta$ and ν is the normalized distance between the true and hypothesized value of δ :

$$\nu = n \left(\begin{pmatrix} \kappa/n \\ -\kappa\mu/n \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right) = \kappa \begin{pmatrix} 1 \\ -\mu \end{pmatrix}. \quad (15)$$

Therefore, we need to find an expression for $S(\cdot)$ and $J(\cdot)$, and evaluate their joint asymptotic behaviour under either Assumption 1 or 2.

Let $\xi = (\omega, \alpha, \beta)$, the GARCH parameters, and let $w_t(\theta) = (1, \varepsilon_{t-1}(\delta)^2, \sigma_{t-1}(\theta)^2)'$. It will be convenient to define:

$$\frac{\partial \log L_t(\theta)}{\partial \sigma_t(\theta)^2} = \frac{1}{2\sigma_t(\theta)^2} \left(\frac{\varepsilon_t(\beta)^2}{\sigma_t(\theta)^2} - 1 \right) = \frac{1}{2\sigma_t(\theta)^2} (\eta_t(\theta)^2 - 1), \quad (16)$$

$$\frac{\partial \sigma_t(\theta)^2}{\partial \delta} = \beta \frac{\partial \sigma_{t-1}(\theta)^2}{\partial \delta} - 2\alpha \varepsilon_{t-1}(\delta) Z_{t-1} = -2\alpha \sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i}(\delta) Z_{t-i}, \quad (17)$$

$$\frac{\partial \sigma_t(\theta)^2}{\partial \xi} = \beta \frac{\partial \sigma_{t-1}(\theta)^2}{\partial \xi} + w_t(\theta) = \sum_{i=0}^{t-1} \beta^i w_{t-i}(\theta). \quad (18)$$

where $\eta_t(\theta) = \varepsilon_t(\delta) / \sigma_t(\theta)$. Here we use the fact that a fixed startup value for σ_0^2 implies $\partial \sigma_0^2 / \partial \delta = 0$ and $\partial \sigma_0^2 / \partial \xi = 0$. Thus we find

$$S_\delta(\theta) = \frac{\partial \log L(\theta)}{\partial \delta} = \sum_{t=1}^n \left(Z_t \frac{\varepsilon_t(\delta)}{\sigma_t(\theta)^2} - \frac{\alpha}{\sigma_t(\theta)^2} (\eta_t(\theta)^2 - 1) \sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i}(\delta) Z_{t-i} \right), \quad (19)$$

$$S_\xi(\theta) = \frac{\partial \log L(\theta)}{\partial \xi} = \sum_{t=1}^n \left(\frac{1}{2\sigma_t(\theta)^2} (\eta_t(\theta)^2 - 1) \sum_{i=0}^{t-1} \beta^i w_{t-i}(\theta) \right). \quad (20)$$

Expressions for the blocks $J_{\delta\delta}$, $J_{\delta\xi}$ and $J_{\xi\xi}$ of the information matrix can be derived from this. We shall not give explicit expressions here, but only provide their limiting behaviour in the next lemma, see Ling and Li (1998).

Lemma 3 *Under Assumption 1, and as $n \rightarrow \infty$,*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor sn \rfloor} \left(\frac{\varepsilon_t}{\sigma_t^2} - \frac{\alpha}{\sigma_t^2} (\eta_t^2 - 1) \sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i} \right) \xrightarrow{\mathcal{L}} \tau B(s), \quad (21)$$

jointly with Lemma 1, where

$$\tau^2 = E \left[\frac{1}{\sigma_t^2} + 2\alpha^2 \sum_{i=1}^{\infty} \beta^{2(i-1)} \frac{\varepsilon_{t-i}^2}{\sigma_t^4} \right], \quad (22)$$

and $(W(s), B(s))$ is a bivariate vector Brownian motion process with $\text{var}(W(1)) = \text{var}(B(1)) = 1$ and $\text{cov}(W(1), B(1)) = \rho = \frac{1}{\sigma\tau}$. Letting $D_n = \text{diag}(n^{-1}, n^{-1/2})$,

$$D_n S_\delta \xrightarrow{\mathcal{L}} \tau \int_0^1 F(s) dB(s), \quad (23)$$

$$D_n J_{\delta\delta} D_n \xrightarrow{\mathcal{L}} \tau^2 \int_0^1 F(s) F(s)' ds, \quad (24)$$

where $F(s) = (U(s), 1)'$. Furthermore,

$$n^{-1/2} S_\xi \xrightarrow{\mathcal{L}} N(0, \Sigma), \quad n^{-1} J_{\xi\xi} \xrightarrow{P} \Sigma, \quad n^{-1/2} D_n J_{\delta\xi} \xrightarrow{P} 0, \quad (25)$$

where Σ is a positive definite matrix.

These results leads to the following theorem, which is obtained as a combination of the results of Ling and Li (1998) and Rahbek (1999):

Theorem 1 Under Assumption 1, and as $n \rightarrow \infty$,

$$LR \xrightarrow{\mathcal{L}} \left(\int_0^1 F(s) [dB(s) + \tau F(s)' \nu ds] \right)' \left[\int_0^1 F(s) F(s)' ds \right]^{-1} \\ \times \left(\int_0^1 F(s) [dB(s) + \tau F(s)' \nu ds] \right), \quad (26)$$

$$2\Phi_1 \xrightarrow{\mathcal{L}} \left(\int_0^1 F(s) [dW(s) + \sigma^{-1} F(s)' \nu ds] \right)' \left[\int_0^1 F(s) F(s)' ds \right]^{-1} \\ \times \left(\int_0^1 F(s) [dW(s) + \sigma^{-1} F(s)' \nu ds] \right). \quad (27)$$

The limiting distribution of LR under the null hypothesis ($\nu = 0$) depends on the nuisance parameter ρ . In practice this nuisance parameter can be estimated consistently, and used to obtain an asymptotic p -value, either by Monte Carlo simulation or by the Gamma approximation proposed by Boswijk and Doornik (1999). The power function depends, in addition to ρ , only on κ/σ (it is invariant to μ and σ). In the next section, we compare the power functions of the two statistics for various values of ρ .

Consider now the asymptotic behaviour of the score vector and information matrix under Assumption 2:

Lemma 4 Under Assumption 2, with $\lambda/\zeta^2 < 1$ and $\varpi > 0$, and as $n \rightarrow \infty$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor sn \rfloor} \left(\frac{\varepsilon_t}{\sigma_t^2} - \frac{\alpha}{\sigma_t^2} (\eta_t^2 - 1) \sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i} \right) \xrightarrow{\mathcal{L}} \int_0^s V(u)^{-1/2} dW_1(u), \quad (28)$$

jointly with Lemma 2. Letting $D_n = \text{diag}(n^{-1}, n^{-1/2})$,

$$D_n S_\delta \xrightarrow{\mathcal{L}} \int_0^1 G(s) V(s)^{-1/2} dW_1(s), \quad (29)$$

$$D_n J_{\delta\delta} D_n \xrightarrow{\mathcal{L}} \int_0^1 G(s) G(s)' V(s)^{-1} ds, \quad (30)$$

where $G(s) = (Y(s), 1)'$. Furthermore,

$$n^{-1} S_\xi = O_P(1), \quad n^{-2} J_{\xi\xi} = O_P(1), \quad n^{-1} D_n J_{\delta\xi} \xrightarrow{P} 0. \quad (31)$$

A proof is given in the Appendix. Note that the limiting Riemann integral in (30) is the quadratic variation of the stochastic integral in (29). The suitably normalized information matrix is block-diagonal in the limit, because the cross-variation between the two parts of the score vector is zero in the limit. These results imply:

Theorem 2 *Under Assumption 2, with $\lambda/\zeta^2 < 1$ and $\varpi > 0$, and as $n \rightarrow \infty$,*

$$LR \xrightarrow{\mathcal{L}} \left(\int_0^1 G(s)V(s)^{-1/2}[dW_1(s) + V(s)^{-1/2}G(s)'\nu ds] \right)' \left[\int_0^1 G(s)G(s)'V(s)^{-1} ds \right]^{-1} \\ \times \left(\int_0^1 G(s)V(s)^{-1/2}[dW_1(s) + V(s)^{-1/2}G(s)'\nu ds] \right), \quad (32)$$

$$2\Phi_1 \xrightarrow{\mathcal{L}} \left(\int_0^1 G(s)[V(s)^{1/2}dW_1(s) + G(s)'\nu ds] \right)' \left[\int_0^1 G(s)G(s)' ds \int_0^1 V(s) ds \right]^{-1} \\ \times \left(\int_0^1 G(s)[V(s)^{1/2}dW_1(s) + G(s)'\nu ds] \right). \quad (33)$$

The theorem is proved in the Appendix. The results are closely related to those obtained by Hansen (1992b, 1995), who considers ordinary least-squares, generalized least-squares and adaptive estimation of regressions with non-stationary volatility. Note that the likelihood ratio statistic is asymptotically equivalent to a Wald statistic based on weighted least-squares with known $\{\sigma_t^2\}$. Hansen shows that when the process generating the non-stationary volatility is unknown, it may be estimated non-parametrically, without loss of efficiency relative to a parametric likelihood analysis.

Both distributions in Theorem 2 depend on nuisance parameters, even under the null hypothesis ($\nu = 0$). In principle they are affected by all volatility parameters (ϖ, λ, ζ) , although parameter variations that only affect the scale of $V(s)$ will leave the distributions in (32) and (33) unaffected. From Nelson (1990), it appears that the function λ/ζ^2 is most relevant, since it determines the stationary distribution of the volatility process. Unfortunately these parameters are not consistently estimable. Further research will have to indicate the degree of dependence on the volatility parameters, and the possibility of finding a bounding distribution that would enable one to control the asymptotic size of the tests. In the next section we evaluate the power of the two tests, taking the volatility parameters as known.

4 Local Power

In this section we provide some numerical evidence on the local power of the two alternative test statistics. First, we consider the case of stationary volatility (Assumption 1). We consider two sets of GARCH parameters:

- $\alpha = 0.05$, $\beta = 0.9$ and $\sigma^2 = \omega/(1 - \alpha - \beta) = 1$, which implies $\rho = 0.93$ (the value of ρ is obtained by Monte Carlo simulation). This corresponds to a relatively smooth GARCH process with strong persistence, as typically found in empirical data sets of daily returns. The high value of the correlation coefficient suggests that the power difference between the LR and Φ_1 test will be relatively small in this case.

- $\alpha = 0.35$, $\beta = 0.6$ and $\sigma^2 = 1$, which implies $\rho = 0.32$. Again this leads to a rather slowly mean-reverting GARCH process, but now the higher value of α leads to more short-run variation in the volatility. The low value of ρ leads us to expect more power gains for the LR test in this case.

Table 1 displays the local power function of the Φ_1 test, which is the same for both parameter combinations, and that of the LR statistic for each data-generating process. All results are obtained by Monte Carlo simulation, using a discretization of the processes and integrals. As expected, the power gain of the LR test relative to the least-squares-based Φ_1 is very small when $\rho = 0.93$. This suggests that for GARCH processes, one might as well use the conventional test. For the second parameter combination, the power gain is substantial for low values of $-\kappa$. As the mean-reversion increases, we see that the power functions seem to converge to each other; it even seems that for high values of $-\kappa$, the power of Φ_1 exceeds that of LR .

Table 1: Local power of Φ_1 and LR with stationary volatility.

κ	-5	-10	-15	-20	-25	-30
Φ_1	0.079	0.214	0.451	0.716	0.893	0.973
LR ($\rho = 0.93$)	0.082	0.227	0.474	0.698	0.873	0.954
LR ($\rho = 0.32$)	0.212	0.437	0.624	0.769	0.865	0.923

Next, we consider the local power function when the volatility process is near-integrated. Again we consider two parameter configurations:

- $\lambda = -100$, $\zeta = 3.162$ and $\sigma^2 = \omega_n / (1 - \alpha_n - \beta_n) = 0.01$. This corresponds to the first case considered above ($(\alpha, \beta) = (0.05, 0.9)$), with $n = 2000$; the expected variance has been chosen smaller here, but the results in Table 1 are invariant to σ^2 , so it is fully comparable.
- $\lambda = -40$, $\zeta = 5.060$ and $\sigma^2 = 0.01$. For $n = 2000$, this corresponds to $(\alpha, \beta) = (0.08, 0.9)$, which leads to a smoother and more persistent volatility process.

Critical values and local power results for the two test statistics for each parameter combination have been obtained by Monte Carlo simulation, and are given in Table 2.

Table 2: Local power of Φ_1 and LR with near-integrated volatility.

κ	-5	-10	-15	-20	-25	-30
Φ_1 ($\lambda = -100, \zeta = 3.16$)	0.076	0.224	0.469	0.754	0.919	0.984
LR ($\lambda = -100, \zeta = 3.16$)	0.087	0.258	0.528	0.805	0.942	0.989
Φ_1 ($\lambda = -40, \zeta = 5.06$)	0.084	0.222	0.466	0.736	0.900	0.974
LR ($\lambda = -40, \zeta = 5.06$)	0.112	0.340	0.657	0.876	0.965	0.994

We observe that the power differences are more pronounced now, especially with the second parameter combination. Note that the local power function of LR in the $\lambda = -100$ case is substantially higher than the corresponding case in Table 1. Further Monte Carlo analysis would have to indicate which of the two asymptotic approximations are more accurate in finite samples.

5 Conclusion

In this paper we have investigated likelihood ratio testing for a unit root when the innovations follow a near-integrated GARCH process. With fixed GARCH parameters, the asymptotic local power function indicates that for parameter combinations that appear to be relevant in empirical finance, the possible power gain of the likelihood ratio statistic over the conventional Dickey-Fuller test is only marginal. However, when we consider near-integrated parameter sequences for the GARCH process, then the power advantages of the likelihood ratio test become more substantial. This suggests that the likelihood ratio statistic would be an attractive alternative to the usual Dickey-Fuller test in series with smooth and persistent volatility. At this point however, the test is not yet operational since its asymptotic null distribution depends on nuisance parameters that are not consistently estimable (note, however, that the same applies to the least-squares based test). Further research will have to indicate how severe this problem is, and whether replacing the unknown diffusion parameters by their maximum likelihood estimates would yield a possible practical solution.

References

- Bollerslev, T., R.F. Engle and D.B. Nelson (1994), “ARCH Models”, in R.F. Engle and D.C. McFadden (Eds.), *Handbook of Econometrics*, Vol. 4. Amsterdam: North Holland.
- Boswijk, H.P. (1999), “Some Distribution Theory for Stochastic Volatility Models”, Working paper, University of Amsterdam.
- Boswijk, H.P. and J.A. Doornik (1999), “Distribution Approximations for Cointegration Tests with Stationary Exogenous Regressors”, Tinbergen Institute Discussion Paper TI 99-013/4, <http://www.tinbinst.nl/papers/TI99013.pdf>.
- Dickey, D.A. and W.A. Fuller (1981), “Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root”, *Econometrica*, 49, 1057–1072.
- Hansen, B.E. (1992a), “Convergence to Stochastic Integrals for Dependent Heterogeneous Processes”, *Econometric Theory*, 8, 489–500.
- Hansen, B.E. (1992b), “Heteroskedastic Cointegration”, *Journal of Econometrics*, 54, 139–158.
- Hansen, B.E. (1995), “Regression with Nonstationary Volatility”, *Econometrica*, 63, 1113–1132.
- Ling, S. and W.K. Li (1997), “Estimating and Testing for Unit Root Processes with GARCH(1,1) Errors”, Technical report, Department of Statistics, Hong Kong University.
- Ling, S. and W.K. Li (1998), “Limiting Distributions of Maximum Likelihood Estimators for Unstable Autoregressive Moving-Average Time Series with General Autoregressive Heteroskedastic Errors”, *Annals of Statistics*, 26, 84–125.
- Nelson, D.B. (1990), “ARCH Models as Diffusion Approximations”, *Journal of Econometrics*, 45, 7–38.
- Rahbek, A.C. (1999), “Likelihood Ratio Tests for a Unit Root in AR-ARCH Models with and without Deterministic Terms”, Working paper, University of Copenhagen.
- Shephard, N. (1996), “Statistical Aspects of ARCH and Stochastic Volatility”, in D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen (Eds.), *Time Series Models*. London: Chapman and Hall.
- Vasicek, O.A. (1977), “An Equilibrium Characterization of the Term Structure”, *Journal of Finance*, 5, 177–188.

Appendix

Proof of Lemma 4

Write the first term of (28) as

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor sn \rfloor} \frac{\eta_t}{\sigma_t} = \int_0^s \sigma_{\lfloor un \rfloor}^{-1} dW_{1n}(u), \quad (\text{A.1})$$

where $W_{1n}(s) = n^{-1/2} \sum_{t=1}^{\lfloor sn \rfloor} \eta_t$. From Lemma 2, $(W_{1n}(s), \sigma_{\lfloor sn \rfloor}^2)$ converges weakly to $(W_1(s), V(s))$. Nelson (1990, Theorem 2.3) shows that $V(s)$ is stationary if $\lambda/\zeta^2 < 1$ and $\varpi > 0$, and that under those conditions $\sigma_{\lfloor sn \rfloor}^{-2}$ converges weakly to $V(s)^{-2}$ (and hence $\sigma_{\lfloor sn \rfloor}^{-1} \xrightarrow{\mathcal{L}} V(s)^{-1/2}$). Since $\{\eta_t\}$ are i.i.d. $N(0, 1)$, the conditions of Hansen (1992a) apply, and $\int_0^s \sigma_{\lfloor un \rfloor}^{-1} dW_{1n}(u) \xrightarrow{\mathcal{L}} \int_0^s V(u)^{-1/2} dW_1(u)$.

Write the remainder of (28) as $n^{1/2} \sum_{i=1}^{\lfloor sn \rfloor} v_t/\sigma_t$, where v_t is a martingale difference sequence with variance $2\alpha_n^2 \sum_{i=1}^{\infty} \beta_n^{2(i-1)} E(\varepsilon_{t-i}^2/\sigma_t^2)$. Using $\varepsilon_{t-i}^2/\sigma_t^2 = \eta_{t-i}^2(\sigma_{t-i}^2/\sigma_t^2)$, and substitution of $2\alpha_n^2 = \zeta^2/n$ and $\beta_n^2 = (1 - \zeta/\sqrt{2n} + \lambda/n)^2 = 1 - 2\zeta/\sqrt{2n} + o(n^{-1/2})$, it follows that the variance of v_t is $O(n^{-1/2})$, so that $n^{1/2} \sum_{i=1}^{\lfloor sn \rfloor} v_t/\sigma_t \xrightarrow{P} 0$. This proves (28).

The results (29) and (30) follow from (28), together with $(n^{1/2} D_n Z_{\lfloor sn \rfloor}, \sigma_{\lfloor sn \rfloor}^{-1}) \xrightarrow{\mathcal{L}} (G(s), V(s)^{-1/2})$, and the fact that $(\eta_t + v_t)$ has bounded variance, so that again the conditions of Hansen (1992a) for weak convergence to a stochastic integral apply.

For the results on the score and information for ξ , we note that $w_t = O_P(1)$, which implies that $\sum_{i=0}^{t-1} \beta_n^i w_{t-i} = O_P(n^{-1/2})$, because $\beta_n = 1 - \zeta/\sqrt{2n} + o(n^{-1/2})$. This explains why S_ξ and $J_{\xi\xi}$ are $O_P(n)$ and $O_P(n^2)$, respectively. The block-diagonality follows from this, together with the fact that the two parts of the score vector are uncorrelated because $E[\eta_t(\eta_t^2 - 1)] = 0$.

Proof of Theorem 2

The result (32) follows from Lemma 4 and (14). Previous derivations show that

$$\begin{aligned} D_n \sum_{t=1}^n Z_t \Delta X_t &= D_n \sum_{t=1}^n Z_t \sigma_t \eta_t + D_n \sum_{t=1}^n Z_t Z_t' \nu / n \\ &\xrightarrow{\mathcal{L}} \int_0^1 G(s) V(s)^{1/2} dW_1(s) + \int_0^1 G(s) G(s)' \nu ds, \end{aligned} \quad (\text{A.2})$$

and similarly

$$D_n \sum_{t=1}^n Z_t Z_t' D_n \xrightarrow{\mathcal{L}} \int_0^1 G(s) G(s)' ds. \quad (\text{A.3})$$

Finally,

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n (\Delta X_t - \hat{\delta}'_{LS} Z_t)^2 &= \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2 + o_P(1) \\ &= \frac{1}{n} \sum_{t=1}^n \sigma_t^2 + \frac{1}{n} \sum_{t=1}^n \sigma_t^2 (\eta_t^2 - 1) + o_P(1) \\ &\xrightarrow{\mathcal{L}} \int_0^1 V(s) ds. \end{aligned} \quad (\text{A.4})$$