Testing for a Unit Root with Near-Integrated Volatility*

H. Peter Boswijk
Department of Quantitative Economics, Universiteit van Amsterdam†

August 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 both cases, we find a substantial asymptotic local power gain of the likelihood ratio test for parameter values that imply heavy tails in the unconditional innovation distribution.

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 asymptotic distribution theory for likelihood based estimators of the volatility parameters in near-integrated 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 dicussions with Anders Rahbek and André Lucas.

†Address for correspondence: Department of Quantitative Economics, Universiteit van 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. In Section 5 we investigate the relevance of these local power results in finite samples, and Section 6 concludes.

2 The Model

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

$$\Delta X_t = \gamma(X_{t-1} - \mu) + \varepsilon_t, \qquad t = 1, \ldots, n, \tag{1}$$

$$\varepsilon_t = \sigma_t \eta_t, \tag{2}$$

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

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

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

$$\mathcal{H}_0 : \gamma = 0, \tag{5}$$

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 (nonnegative) 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 $\omega > 0$, $(\alpha, \beta) \geq 0$):

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, such that $\varpi > 0$, $\zeta > 0$ and $\lambda < \zeta^2/2$.*

Under Assumption 1, the process X_t is near-integrated with stationary volatility. The unit root 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 $n \to \infty$,*

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

in $D[0,1]^2$, where $\sigma^2 = \omega/(1 - \alpha - \beta)$, $W(\cdot)$ is a standard Brownian motion process on $[0,1]$, and $U(\cdot)$ is an Ornstein-Uhlenbeck process on $[0,1]$:

$$dU(s) = \kappa U(s) ds + dW(s), \qquad U(0) = 0. \tag{7}$$

The proof of this lemma is given in Ling and Li (1998, Theorem 3.3) for $\kappa = 0$, in which case $U(\cdot)$ reduces to $W(\cdot)$. This is extended to the case $\kappa \neq 0$ by writing $X_{\lfloor \cdot n \rfloor}$ as a continuous functional of the partial sum of ε_t.

Lemma 2 *Under Assumption 2, and as $n \to \infty$,*

$$\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor \cdot n \rfloor} \eta_t, \frac{1}{\sqrt{2n}} \sum_{t=1}^{\lfloor \cdot n \rfloor} (\eta_t^2 - 1), \frac{1}{\sqrt{n}} X_{\lfloor \cdot n \rfloor}, \sigma^2_{\lfloor \cdot n \rfloor} \right) \xrightarrow{\mathcal{L}} \left(W_1(\cdot), W_2(\cdot), Y(\cdot), V(\cdot) \right), \tag{8}$$

in $D[0,1]^4$, where $(W_1(\cdot), W_2(\cdot))$ is a standard bivariate Brownian motion process on $[0,1]$, and $(Y(\cdot), V(\cdot))$ is the solution to the system of stochastic differential equations

$$dY(s) = \kappa Y(s)ds + V(s)^{1/2}dW_1(s), \tag{9}$$

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

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 $X_t = \sqrt{n}Y(t/n), t = 0, 1, \ldots, n$, then the actual process generating X_t may be approximated by (1)–(4) under Assumption 2; the approximation error will vanish as $n \to \infty$, 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

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

where $\varepsilon_t(\delta) = \Delta X_t - \delta' Z_t$, and where it should be noted that $\sigma_t^2(\theta)$ 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 $\hat{\theta} = \mathrm{argmax}_{\theta\in\Theta}\, \ell(\theta)$ and $\tilde{\theta} = \mathrm{argmax}_{\theta\in\Theta_0}\, \ell(\theta)$, the unrestricted and restricted maximum likelihood estimators, respectively. The likelihood ratio statistic for the null hypothesis is

$$LR = -2\left(\ell(\tilde{\theta}) - \ell(\hat{\theta})\right). \tag{12}$$

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

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

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 $S(\theta) = \partial \ell(\theta)/\partial\theta$ and the observed information matrix $J(\theta) = -\partial^2 \ell(\theta)/\partial\theta\partial\theta'$. Conventional Taylor series expansions (corresponding to a quadratic approximation of the log-likelihood function) result in

$$\begin{aligned} LR &= \hat{\theta}' E_1 \left[E_1' J(\theta_0)^{-1} E_1\right]^{-1} E_1' \hat{\theta} + \mathrm{o}_P(1) \\ &= \left(n^{-1}\nu' + S(\theta_0)' J(\theta_0)^{-1} E_1\right) \left[E_1' J(\theta_0)^{-1} E_1\right]^{-1} \left(E_1' J(\theta_0)^{-1} S(\theta_0) + n^{-1}\nu\right) + \mathrm{o}_P(1), \quad (14) \end{aligned}$$

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 $\xi = (\omega, [\alpha+\beta], \alpha)'$, the (linearly transformed) GARCH parameters, and $w_t(\theta) = (1, \sigma_{t-1}^2(\theta), \varepsilon_{t-1}^2(\delta) - \sigma_{t-1}^2(\theta))'$. The following results are useful ingredients for the score vector:

$$\begin{aligned} \frac{\partial \ell_t(\theta)}{\partial \sigma_t(\theta)^2} &= \frac{1}{2\sigma_t^2(\theta)} \left(\frac{\varepsilon_t^2(\beta)}{\sigma_t^2(\theta)} - 1 \right) = \frac{1}{2\sigma_t^2(\theta)} \left(\eta_t^2(\theta) - 1\right), & (16) \\ \frac{\partial \sigma_t^2(\theta)}{\partial \delta} &= \beta \frac{\partial \sigma_{t-1}^2(\theta)}{\partial \delta} - 2\alpha \varepsilon_{t-1}(\delta) Z_{t-1} = -2\alpha \sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i}(\delta) Z_{t-i}, & (17) \\ \frac{\partial \sigma_t^2(\theta)}{\partial \xi} &= \beta \frac{\partial \sigma_{t-1}^2(\theta)}{\partial \xi} + w_t(\theta) = \sum_{i=0}^{t-1} \beta^i w_{t-i}(\theta). & (18) \end{aligned}$$

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

$$\begin{aligned} S_\delta(\theta) &= \frac{\partial \ell(\theta)}{\partial \delta} = \sum_{t=1}^{n} \left(Z_t \frac{\varepsilon_t(\delta)}{\sigma_t^2(\theta)} - \frac{\alpha}{\sigma_t^2(\theta)} \left(\eta_t^2(\theta) - 1\right) \sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i}(\delta) Z_{t-i} \right), & (19) \\ S_\xi(\theta) &= \frac{\partial \ell(\theta)}{\partial \xi} = \sum_{t=1}^{n} \left(\frac{1}{2\sigma_t^2(\theta)} \left(\eta_t^2(\theta) - 1\right) \sum_{i=0}^{t-1} \beta^i w_{t-i}(\theta) \right). & (20) \end{aligned}$$

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* $n \to \infty$,

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

in $D[0,1]$*, jointly with Lemma 1, where*

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

and $(W(\cdot), B(\cdot))$ *is a bivariate vector Brownian motion process with* $\mathrm{var}(W(1)) = \mathrm{var}(B(1)) = 1$ *and* $\mathrm{cov}(W(1), B(1)) = \rho = \dfrac{1}{\sigma\tau}$*. Letting* $D_{1n} = \mathrm{diag}((\sigma n)^{-1}, n^{-1/2})$ *and* $F(s) = (U(s), 1)'$*,*

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

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

Furthermore,

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

where Σ *is a positive definite matrix.*

These results leads to the following theorem, the proof of which is given in the Appendix:

Theorem 1 *Under Assumption 1, and as* $n \to \infty$*,*

$$\begin{aligned} LR \quad &\xrightarrow{\mathcal{L}} \quad \left(\int_0^1 F(s)\left[dB(s) + \frac{\kappa}{\rho} U(s) ds\right]\right)' \left[\int_0^1 F(s) F(s)' ds\right]^{-1} \\ &\qquad \times \left(\int_0^1 F(s)\left[dB(s) + \frac{\kappa}{\rho} U(s) ds\right]\right), \end{aligned} \tag{26}$$

$$\begin{aligned} 2\Phi_1 \quad &\xrightarrow{\mathcal{L}} \quad \left(\int_0^1 F(s)[dW(s) + \kappa U(s) ds]\right)' \left[\int_0^1 F(s) F(s)' ds\right]^{-1} \\ &\qquad \times \left(\int_0^1 F(s)[dW(s) + \kappa U(s) ds]\right). \end{aligned} \tag{27}$$

The limiting distribution of LR under the null hypothesis ($\kappa = 0$) depends on the nuisance parameter ρ. In practice this nuisance parameter can be estimated consistently by $\hat{\rho} = 1/\sqrt{\hat{\sigma}^2 \hat{\tau}^2}$, where $\hat{\sigma}^2 = \hat{\omega}/(1 - \hat{\alpha} - \hat{\beta})$ and $\hat{\tau}^2$ is the sample analog of (22). Although we have not been able to obtain an explicit formula for ρ in terms of α and β, an approximation yields

$$\rho(\alpha, \beta) \approx \tilde{\rho}(\alpha, \beta) = \sqrt{\frac{(1 - \alpha - \beta)(1 - \beta^2)}{(1 - \alpha - \beta + \alpha^2)(1 - \beta^2 + 2\alpha^2)}}, \tag{28}$$

which is obtained by replacing $\varepsilon_{t-i}^2/\sigma_t^2$ in (22) by 1, and using $E(1/\sigma_t^2) \approx (1 - \alpha - \beta + \alpha^2)/\omega$, which corresponds to the continuous-record stationary distribution of $1/\sigma_t^2$ obtained by Nelson (1990). In order to check the accuracy of this approximation, we estimate the expectation in (22) by the average, over 1000 Monte Carlo replications[1], of the sample mean corresponding to (22) with a sample size of $10,000$. This is done for $\alpha + \beta \in \{0.1, 0.2, \ldots, 0.9\}$ and $\alpha/(\alpha + \beta) \in \{0.1, 0.2, \ldots, 1\}$. It appears

[1]All numerical results have been obtained using Ox version 2.20, see Doornik (1999).

that (28) somewhat underestimates the true correlation; from a log-linear regression of the actual ρ's on $\tilde{\rho}(\alpha,\beta)$, we obtain the following adjusted approximation:

$$\hat{\rho}(\alpha,\beta) = \tilde{\rho}(\alpha,\beta)^{0.64}, \tag{29}$$

which is quite accurate, with a regression standard error of about 1%.

Next, the estimate of ρ can be 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 (it is invariant to σ). In the next section, we compare the power functions of the two statistics for two cases.

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

Lemma 4 *Under Assumption 2, and as $n \to \infty$,*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor \cdot n \rfloor} \left(\frac{\varepsilon_t}{\sigma_t^2} - \frac{\alpha}{\sigma_t^2} \left(\eta_t^2 - 1\right) \sum_{i=1}^{t-1} \beta^{i-1} \varepsilon_{t-i} \right) \xrightarrow{\mathcal{L}} \int_0^{\cdot} V(u)^{-1/2} dW_1(u), \tag{30}$$

in $D[0,1]$, jointly with Lemma 2. Letting $D_{2n} = \operatorname{diag}(n^{-1}, n^{-1/2})$ and $G(s) = (Y(s), 1)'$,

$$D_{2n} S_\delta \xrightarrow{\mathcal{L}} \int_0^1 G(s) V(s)^{-1/2} dW_1(s), \tag{31}$$

$$D_{2n} J_{\delta\delta} D_{2n} \xrightarrow{\mathcal{L}} \int_0^1 G(s) G(s)' V(s)^{-1} ds. \tag{32}$$

Furthermore, there exist non-singular norming matrices D_{3n} such that

$$D_{3n} S_\xi = \mathrm{O}_P(1), \qquad D_{3n} J_{\xi\xi} D_{3n} = \mathrm{O}_P(1), \qquad D_{2n} J_{\delta\xi} D_{3n} \xrightarrow{P} 0. \tag{33}$$

A proof is given in the Appendix. Note that the limiting Riemann integral in (32) is the quadratic variation of the stochastic integral in (31). 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, and as $n \to \infty$,*

$$\begin{aligned} LR \xrightarrow{\mathcal{L}} & \left(\int_0^1 G(s) V(s)^{-1/2} [dW_1(s) + \kappa V(s)^{-1/2} Y(s) ds] \right)' \left[\int_0^1 G(s) G(s)' V(s)^{-1} ds \right]^{-1} \\ & \times \left(\int_0^1 G(s) V(s)^{-1/2} [dW_1(s) + \kappa V(s)^{-1/2} Y(s) ds] \right), \end{aligned} \tag{34}$$

$$\begin{aligned} 2\Phi_1 \xrightarrow{\mathcal{L}} & \left(\int_0^1 G(s) [V(s)^{1/2} dW_1(s) + \kappa Y(s) ds] \right)' \left[\int_0^1 G(s) G(s)' ds \int_0^1 V(s) ds \right]^{-1} \\ & \times \left(\int_0^1 G(s) [V(s)^{1/2} dW_1(s) + \kappa Y(s) ds] \right). \end{aligned} \tag{35}$$

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 ($\kappa = 0$). 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 (34) and (35) 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.

A possible solution to this nuisance parameter problem is to use the *conditional* asymptotic null distribution of the two test statistics, given the volatility process $V(\cdot)$. Although this process depends on the parameters (ϖ, λ, ζ), these are variation independent of the parameter of interest κ, so that conditioning on V does not entail a loss of information on κ. In other words, V is S-ancillary for κ, see Barndorff-Nielsen (1978). Clearly the asymptotic distributions in (34) and (35) for fixed V will depend on the realization of V, and hence cannot be tabulated. However, given the independence between W_1 and V the conditional distribution is quite easy to simulate. In practice this will involve replacing the volatility process by its estimate $\hat{V}_n(s) = \hat{\sigma}^2_{\lfloor sn \rfloor}$, with $\{\hat{\sigma}_t^2\}$ the filtered estimate of $\{\sigma_t^2\}$ based on the maximum likelihood estimates of the GARCH parameters. The results of Nelson and Foster (1994) suggest that $\hat{V}_n(\cdot)$ converges in probability to $V(\cdot)$ in $D[0,1]$, which in turn would imply that the estimated conditional distribution given $\{\hat{\sigma}_t^2\}_{t=1}^n$ converges to the true conditional distribution given $\{V(s), s \in [0,1]\}$.

Before we proceed, it is of interest to discuss the difference of the two types of asymptotic approximation in Theorems 1 and 2 as $\alpha + \beta$ approaches 1, so that the unconditional variance σ^2 diverges. For fixed parameter values (Assumption 1) the approximation (28) suggests that $\rho \downarrow 0$ as $\alpha + \beta \uparrow 1$, which is confirmed by the fact that $\rho = 1/(\sigma\tau)$, and σ diverges whereas τ remains finite. This implies, first, that the limiting distribution of the LR statistic will approach the $\chi^2(2)$ distribution under the null hypothesis, because B and F become independent. Secondly, it shows that the local power will increase, and in fact approach 1 for all κ, because the non-centrality parameter in (26) is essentially κ/ρ. This suggests that in such cases the likelihood ratio test has infinite power superiority over the least-squares based test. Note, however, that Theorem 1 is only valid under Assumption 1, which involves the condition $\alpha + \beta < 1$; the quality of the asymptotic approximation might deteriorate as $\alpha + \beta \uparrow 1$. More importantly, if $\alpha \downarrow 0$ at the same time as $\alpha + \beta \uparrow 1$, then the above arguments are no longer valid, since $\lim_{\alpha\downarrow 0, \alpha+\beta\uparrow 1} \rho(\alpha, \beta)$ does not exist. This implies that for parameter values with $\alpha + \beta$ close to 1 and α close to 0, which are typically encountered with daily financial returns, this asymptotic approximation will not be reliable, and we should turn to the continuous-record asymptotic approximation implied by Assumption 2 instead.

Under Assumption 2, then, it is allowed that $\alpha + \beta = 1$ and hence $\lambda = 0$; no discontinuity in the limit theory is to be expected around $\lambda = 0$, as long as $\lambda < \zeta^2/2$, which is the condition for strict stationarity of the limiting diffusion process $V(s)$. The main difference between the cases $\lambda < 0$ and $0 \leq \lambda < \zeta^2/2$ is that in the former case the disturbances ε_t have finite variance, whereas in the latter case the unconditional variance is infinite, since the limiting distribution of ε_t is Student's $t(2 - 4\lambda/\zeta^2)$, see Nelson (1990). For $\lambda = 0$ the limiting distribution of LR will not be $\chi^2(2)$ under the null hypothesis, since $V(s)^{-1/2}Y(s) = V(s)^{-1/2}\int_0^s V(u)^{1/2}dW_1(u)$ and $W_1(s)$ are not independent for $\lambda = 0$. As ζ increases however, the variation in $V(s)$ increases, and one might expect that the behaviour of $V(s)^{-1/2}Y(s)$ will be dominated by $V(s)$, such that it becomes independent of $W_1(s)$. From the expressions in (34) and (35) it is not clear that the relative power advantage of LR will increase with ζ; this will be investigated in the next section.

In summary, the results in this section indicate we may expect a power gain of the likelihood ratio test over the Dickey-Fuller test when a large value of $\alpha + \beta$ (implying persistent volatility) is combined with a large value of α (implying a large short-run variation in the volatility). In the next section we investigate whether these predictions are reflected in the asymptotic local power behaviour of the tests, and in Section 5 we turn to the finite sample behaviour of the procedures.

4 Local Power

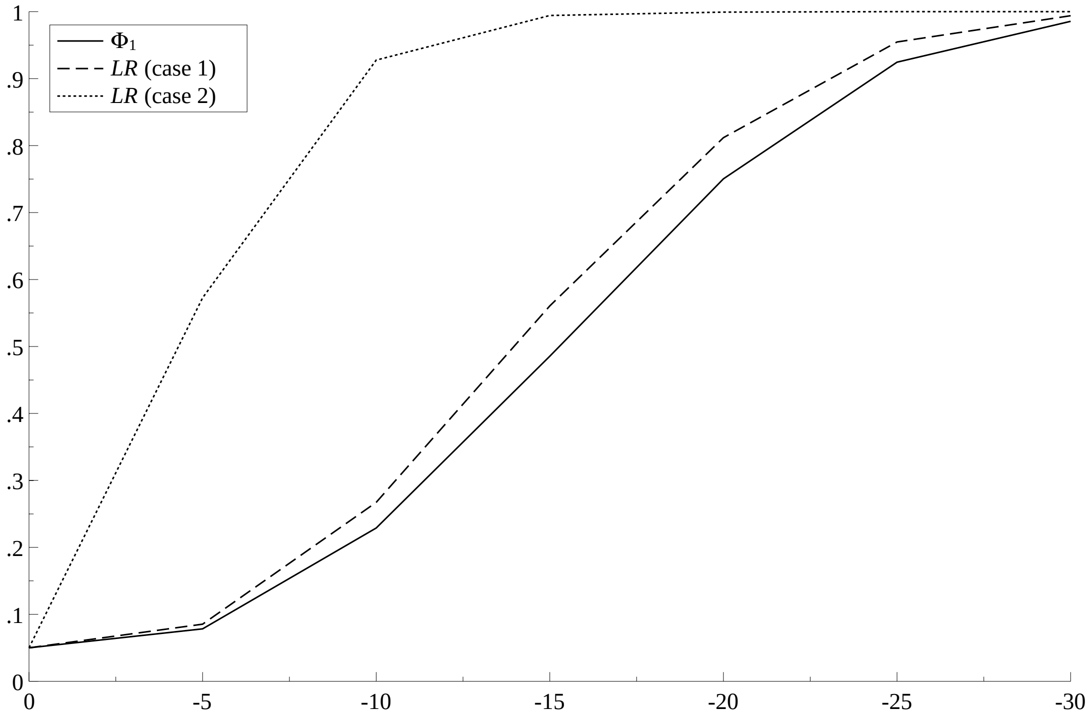
In this section we provide some numerical evidence on the local power of the two alternative test statistics. First, we consider the case of stationary volatility (Assumption 1). We consider two sets of GARCH parameters:

1. $\alpha = 0.05$, $\beta = 0.9$ and $\sigma^2 = \omega/(1 - \alpha - \beta) = 1$, which implies $\rho = 0.967$ (the value of $\rho = 1/(\sigma\tau)$ is obtained by Monte Carlo simulation, as described in Section 3). These parameter values correspond 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.

2. $\alpha = 0.35$, $\beta = 0.6$ and $\sigma^2 = 1$, which implies $\rho = 0.570$. 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.

Figure 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. Note that the local power of Φ_1 is the same as the local power of LR when $\alpha = \beta = 0$ and hence $\rho = 1$, i.e., when there are no GARCH effects. All results are obtained by Monte Carlo simulation, using a discretization (1000 equidistant points) of the processes and integrals, and with $10,000$ replications for the power calculations, and $100,000$ replications for the critical values.

Figure 1: Local power of Φ_1 and LR with stationary volatility (size = 5%).



As expected, the power gain of the LR test relative to the least-squares-based Φ_1 is very small when $(\alpha, \beta) = (0.05, 0.9)$. This suggests that for such GARCH processes, one might as well use the conventional test. For the second parameter combination, however, the power gain is much larger. Therefore, these results confirm the prediction in the previous section that only when the volatility process has itself a high volatility (corresponding to a high value of α), the likelihood ratio test yields a substantial power gain over the least-squares based Dickey-Fuller test.

Next, we consider the local power function when the volatility process is near-integrated. In this case we consider four parameter configurations:

1. $\lambda = -100$, $\zeta = \sqrt{10}$. This corresponds to the first case ($(\alpha_n, \beta_n) = (0.05, 0.9)$) considered above, with $n = 2000$.

2. $\lambda = -100$, $\zeta = 7\sqrt{10}$. For $n = 2000$, this corresponds to the second case above ($(\alpha_n, \beta_n) = (0.35, 0.6)$), which leads to more variation in the volatility process.

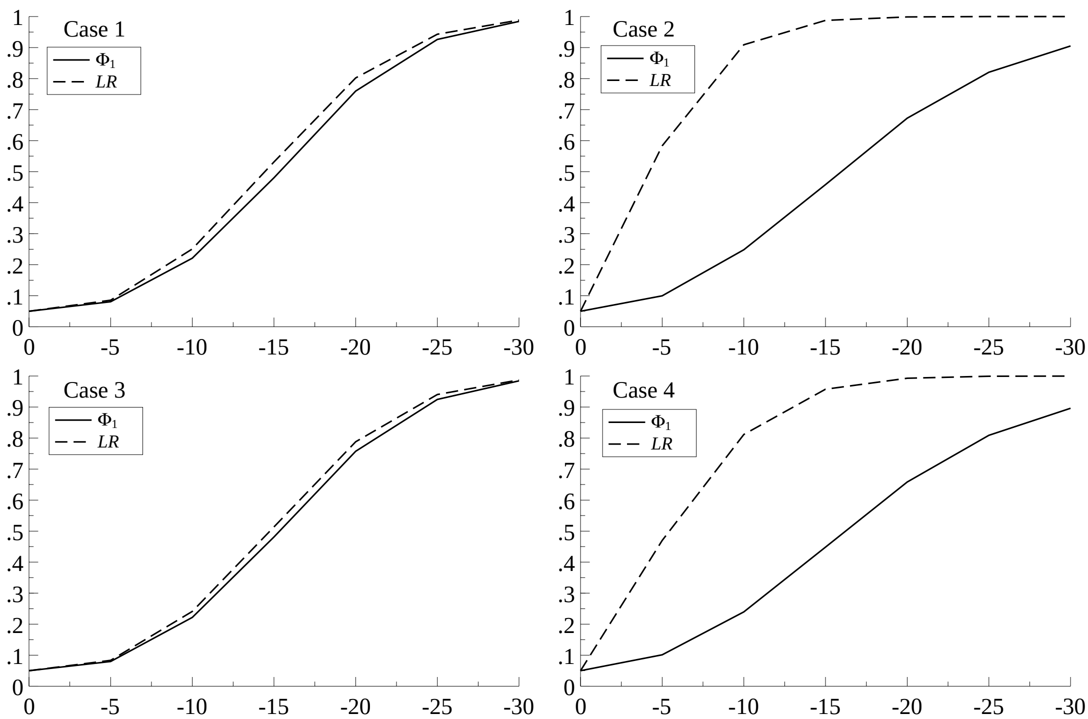
3. $\lambda = -40$, $\zeta = 2$. This is a process with less mean-reversion in the volatility than case 1, but with the same value of $-\lambda/\zeta^2 = 10$; it corresponds to $(\alpha_n, \beta_n) = (0.05, 0.9)$ for $n = 800$. Therefore, we expect roughly the same results as in case 1.

4. $\lambda = -40$, $\zeta = 14$. This is comparable to case 2 (same $-\lambda/\zeta^2 = 10/49 \approx 0.2$), but with less mean-reversion in the volatility, and corresponds to $(\alpha_n, \beta_n) = (0.35, 0.6)$ for $n = 800$.

In all cases we set $\varpi = -\lambda$, such that $V(s)$ reverts to 1, but the results are invariant to ϖ, as long as the starting value is chosen appropriately. We use a fixed start-up value $V(0) = 1$, which corresponds to the expectation of the stationary distribution of $V(s)$; alternatively, one could draw $V(0)^{-1}$ from its stationary $\Gamma(1 - 2\lambda/\zeta^2, 2\varpi/\zeta^2)$, distribution, see Nelson (1990). An important difference between cases 1 and 3 on the one hand, and cases 2 and 4 on the other hand, is the existence of higher moments of ε_t. Nelson's (1990) limiting $t(2 - 4\lambda/\zeta^2)$ distribution for the disturbances implies that ε_t has no finite integer moments beyond the variance in cases 2 and 4, whereas is has much higher moments (up to $2 - 4\lambda/\zeta^2 = 42$) in cases 1 and 3.

For each test, we perform a conditional and an unconditional version. The unconditional version involves Monte Carlo simulation of the 5% critical value for the given parameter combination (based on $100,000$ replications), and defining the local power as the rejection frequency (based on $10,000$ replications) at this critical value. In practice this is infeasible, since the volatility parameters are not known and not consistently estimable, but obtaining the critical values is of interest to investigate how sensitive they are to parameter variations. In the conditional version of the test, we simulate the p-value (based on 1000 replications), for each of the $10,000$ realizations of the test statistic, conditional on the actual volatility process for that realization, and reject when this p-value is less than 0.05. In practice the local power of the conditional and unconditional versions of the test turns out to be almost identical, so we only report the conditional versions in Figure 2.

Figure 2: Local power of Φ_1 and LR with near-integrated volatility (size = 5%).



The results clearly show the expected power gain of the LR test for cases 2 and 4, whereas the two tests are almost equivalent in cases 1 and 3. Furthermore, the power functions in cases 1 and 3 are very close to the corresponding case 1 in Figure 1 (using fixed-parameter asymptotics), and similarly the behaviour in cases 2 and 4 resembles the corresponding case 2 in Figure 1. The effect of λ is much weaker, although we see that the power is slightly lower for the small λ cases. It turns out that the critical values are also fairly close to the corresponding fixed-parameter cases, so that the two types of asymptotics are largely in agreement. As one might expect, this agreement would break down for parameter values such that $\alpha + \beta = 1$, and hence $\lambda = 0$. The fixed-parameter asymptotic analysis, although not strictly applicable anymore, would suggest that $\rho = 0$, which would imply χ^2 critical values and an infinite power gain of the LR test. However, additional simulations indicate that the behaviour of the tests under $\alpha_n + \beta_n = 1$ and $\alpha_n = \zeta/\sqrt{2n}$ depends very much on ζ, comparable to cases 1–4 in Figure 2; only when $\zeta \to \infty$ the null distribution of LR approaches the $\chi^2(2)$ distribution, and the local power becomes 1 for all κ.

5 A Monte Carlo Experiment

In this section we consider the finite-sample behaviour of the tests in a small-scale Monte Carlo experiment. We consider $n \in \{250, 2000\}$, which would correspond to approximately 1 and 8 years of daily financial data. Here $n = 250$ may be considered a small sample for GARCH estimation; usually a number of years of daily data are considered. Next, we continue to consider the two near-integrated cases $(\alpha, \beta) = \{(0.05, 0.9), (0.35, 0.6)\}$; note that these are chosen the same for both sample sizes, so that one might expect a relatively better approximation by the stationary (fixed-parameter) asymptotic distributions for larger sample sizes. Furthermore we consider $\gamma_n = \kappa/n$ with $\kappa \in \{0, -5, -20\}$, to study both the size and power properties of the tests.

For LR and Φ_1 tests[2] we compute two types of p-values, the first based on fixed-parameter asymptotics, and the second based on near-integrated asymptotics, conditional on the estimated $\{\hat{\sigma}_t^2\}$. For the fixed-parameter asymptotic p-values, we use the Gamma approximation of Boswijk and Doornik (1999), in combination with $\hat{\rho}(\alpha, \beta)$ given in (28)–(29), where α and β are replaced by their unrestricted ML estimates. Finally, we also consider QLR, the quasi-likelihood ratio test based on the assumption that $\varepsilon_t \sim \text{i.i.d.}\ t(\nu)$. This test is included to see whether the same power gain can be obtained by a test that correctly specifies the marginal distribution of the disturbances, although it misspecifies the volatility dynamics. From, e.g., Lucas (1997), it follows that when ε_t is indeed $\text{i.i.d.}\ t(\nu)$, the asymptotic distribution of QLR is the same as that of LR in (26), but with ρ the correlation between ε_t and $-\partial \log f_\nu(\varepsilon_t)/\partial \varepsilon_t$, where f_ν denotes the $t(\nu)$ density. It can be shown that for $\nu > 2$, $\rho^2 = (\nu^2 + \nu - 6)/(\nu^2 + \nu)$. Replacing ν by the unrestricted ML estimate, this yields again a p-value for this test using the Gamma approximation.

The results, based on 2000 replications, are given in Table 1, and give rise to the following conclusions. First, we see that the fixed-parameter asymptotic p-values give a better approximation for

[2] In the likelihood function, the volatility process is initialized using $\hat{\sigma}_0^2 = \hat{\varepsilon}_0^2 = n^{-1} \sum_{t=1}^n \Delta X_t^2$.

the small α generating process, but a worse approximation for the $\alpha = 0.35$ case. In the latter case, QLR has a rather large size distortion for $n = 250$, but this seems to vanish as n increases. In fact, for $n = 2000$ there hardly seem to be any substantial size distortions left, with the exception of the Φ_1 test using ordinary critical values when α is large. The power behaviour of the LR and Φ_1 tests is as predicted by the asymptotic analysis. For small α, the power of the tests is virtually identical, so that there is not much gain in using the LR test. When α is large on the other hand, the power gain is quite clear, especially for the larger sample size. Finally, we note that for this case the power curve of the Student t-based QLR test lies between that of Φ_1 and LR: although there is a clear gain in fitting the marginal tail behaviour of the disturbances by a Student's t instead of a normal distribution, we see that the misspecification of the conditional variance gives this test a power disadvantage relative to the GARCH-based LR test.

Table 1: Rejection frequencies of LR, Φ_1 and QLR

		$n = 250$			$n = 2000$		
		$\kappa = 0$	$\kappa = -5$	$\kappa = -20$	$\kappa = 0$	$\kappa = -5$	$\kappa = -20$
$\alpha = 0.05$, $\beta = 0.9$	LR, fixed	0.059	0.086	0.784	0.055	0.082	0.794
	LR, near-int.	0.067	0.094	0.809	0.061	0.091	0.813
	Φ_1, fixed	0.059	0.086	0.775	0.054	0.078	0.752
	Φ_1, near-int.	0.067	0.095	0.798	0.058	0.085	0.770
	QLR	0.062	0.085	0.760	0.054	0.079	0.755
$\alpha = 0.35$, $\beta = 0.6$	LR, fixed	0.074	0.248	0.823	0.049	0.448	0.989
	LR, near-int.	0.063	0.249	0.926	0.040	0.417	0.990
	Φ_1, fixed	0.099	0.150	0.755	0.073	0.127	0.739
	Φ_1, near-int.	0.070	0.120	0.689	0.052	0.114	0.705
	QLR	0.115	0.190	0.754	0.055	0.234	0.913

6 Conclusion

In this paper we have investigated likelihood ratio testing for a unit root when the innovations follow a near-integrated GARCH process. We have analysed the asymptotic null distribution and local power function of the likelihood ratio test and the least-squares based Dickey-Fuller test, both under fixed GARCH parameters and under near-integrated sequences. It has been found that the two types of asymptotics are largely in agreement, as long as the sum of the GARCH parameters is less than one. A considerable power gain potential for the LR test has been found to occur with GARCH processes with a large short-run variation in the volatility, corresponding to a heavy-tailed marginal distribution of the innovations. These asymptotic results have been shown to be reflected in the finite sample behaviour of the tests.

References

Barndorff-Nielsen, O.E. (1978), *Information and Exponential Families in Statistical Theory*. New York: John Wiley.

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.tinbergen.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.

Doornik, J.A. (1999), *Object-Oriented Matrix Programming Using Ox* (3rd ed). London: Timberlake Consultants Press and Oxford: `http://www.nuff.ox.ac.uk/Doornik`.

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.

Lucas, A. (1997), “Cointegration Testing Using Pseudo Likelihood Ratio Tests”, *Econometric Theory*, 13, 149–169.

Nelson, D.B. (1990), “ARCH Models as Diffusion Approximations”, *Journal of Econometrics*, 45, 7–38.

Nelson, D.B. and D.P. Foster (1994), “Asymptotic Filtering Theory for Univariate ARCH Models”, *Econometrica*, 62, 1–41.

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 Theorem 1. Consider first the limiting distribution of LR. Let $D_{1n}^* = \operatorname{diag}(D_{1n}, n^{-1/2}I_3)$, so that $E_1' D_{1n}^* = D_{1n}E_1'$. Therefore, (14) implies, with $S = S(\theta_0)$ and $J = J(\theta_0)$,

$$\begin{aligned} LR &= \left[n^{-1}\nu' D_{1n}^{-1} + S' D_{1n}^* (D_{1n}^* J D_{1n}^*)^{-1} E_1\right] \left[E_1'(D_{1n}^* J D_{1n}^*)^{-1} E_1\right]^{-1} \\ &\quad \times \left[E_1'(D_{1n}^* J D_{1n}^*)^{-1} D_{1n}^* S + D_{1n}^{-1} n^{-1}\nu\right] + \mathrm{o}_P(1). \end{aligned} \tag{A.1}$$

Lemma 3 yields

$$D_{1n}^* S \xrightarrow{\mathcal{L}} \begin{pmatrix} \tau \int_0^1 F(s) dB(s) \\ N(0,\Sigma) \end{pmatrix}, \qquad D_{1n}^* J D_{1n}^* \xrightarrow{\mathcal{L}} \begin{bmatrix} \tau^2 \int_0^1 F(s)F(s)' ds & 0 \\ 0 & \Sigma \end{bmatrix}, \tag{A.2}$$

and clearly $D_{1n}^{-1} n^{-1}\nu \to (\sigma\kappa, 0)'$. Combining these results gives

$$\begin{aligned} LR &\xrightarrow{\mathcal{L}} \left[(\sigma\kappa,0) + \tau \int_0^1 dB(s)F(s)' \left(\tau^2 \int_0^1 F(s)F(s)' ds\right)^{-1}\right] \left[\tau^2 \int_0^1 F(s)F(s)' ds\right] \\ &\quad \times \left[\left(\tau^2 \int_0^1 F(s)F(s)' ds\right)^{-1} \tau \int_0^1 F(s)dB(s) + (\sigma\kappa,0)'\right] \\ &= \left[(\sigma\tau\kappa,0) \int_0^1 F(s)F(s)' ds + \int_0^1 dB(s)F(s)'\right] \left[\int_0^1 F(s)F(s)' ds\right]^{-1} \\ &\quad \times \left[\int_0^1 F(s)F(s)' ds (\sigma\tau\kappa,0)' + \int_0^1 F(s)dB(s)\right], \end{aligned} \tag{A.3}$$

and using $\sigma\tau = 1/\rho$, this yields (26). For Φ_1, we use Lemma 1 together with the continuous mapping theorem to yield

$$D_{1n} \sum_{t=1}^n Z_t Z_t' D_{1n} \xrightarrow{\mathcal{L}} \int_0^1 F(s)F(s)' ds, \tag{A.4}$$

and

$$\begin{aligned} D_{1n} \sum_{t=1}^n Z_t \Delta X_t &= D_{1n} \sum_{t=1}^n Z_t Z_t' D_{1n} D_{1n}^{-1}\delta + D_{1n} \sum_{t=1}^n Z_t \varepsilon_t \\ &\xrightarrow{\mathcal{L}} \int_0^1 F(s)F(s)' ds (\sigma\kappa,0)' + \int_0^1 F(s)dW(s)\sigma \\ &= \sigma \int_0^1 F(s)[dW(s) + \kappa U(s)]ds. \end{aligned} \tag{A.5}$$

Furthermore,

$$\begin{aligned} \hat{\sigma}_{LS}^2 &= \frac{1}{n-2} \sum_{t=1}^n (\Delta X_t - \hat{\delta}_{LS}' Z_t)^2 \\ &= \frac{1}{n-2} \left(\sum_{t=1}^n \varepsilon_t^2 - \sum_{t=1}^n \varepsilon_t Z_t' D_{1n} \left(D_{1n} \sum_{t=1}^n Z_t Z_t' D_{1n}\right)^{-1} D_{1n} \sum_{t=1}^n Z_t \varepsilon_t\right) \\ &= \frac{1}{n-2} \sum_{t=1}^n \varepsilon_t^2 + \mathrm{o}_p(1), \end{aligned} \tag{A.6}$$

which converges in probability to σ^2. Collecting the results yields (27). $\square$

Proof of Lemma 4. Write the first term of (30) as

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor sn \rfloor} \frac{\eta_t}{\sigma_t} = \int_0^s \sigma_{\lfloor un \rfloor}^{-1} dW_{1n}(u), \tag{A.7}$$

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

Write the remainder of (30) as $n^{1/2} \sum_{i=1}^{\lfloor sn \rfloor} \upsilon_t/\sigma_t$, where υ_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 $\beta_n^2 = (1 - \zeta/\sqrt{2n} + \lambda/n)^2 = 1 - 2\zeta/\sqrt{2n} + \mathrm{o}(n^{-1/2})$, it follows that the variance of υ_t is $\mathrm{O}(n^{-1/2})$, so that $n^{-1/2} \sum_{i=1}^{\lfloor \cdot n \rfloor} \upsilon_t/\sigma_t \xrightarrow{P} 0$. This proves (30).

The results (31) and (32) follow from (30), together with the result that $(n^{1/2} D_{2n} Z_{\lfloor \cdot n \rfloor}, \sigma_{\lfloor \cdot n \rfloor}^{-1}) \xrightarrow{\mathcal{L}} (G(\cdot), V(\cdot)^{-1/2})$, and the fact that $(\eta_t + \upsilon_t)$ 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, let $e_t = (\eta_t^2 - 1)/\sqrt{2}$, $W_{2n}(s) = n^{-1/2} \sum_{t=1}^{\lfloor sn \rfloor} e_t$ and define $F_{1n}(s) = (1 - \beta_n) \sum_{i=0}^{\lfloor sn \rfloor - 1} \beta_n^i = 1 - \beta_n^{\lfloor sn \rfloor} = 1 - \exp(\lfloor sn \rfloor \log(1 - \zeta/\sqrt{2n} + \lambda/n)) \to 1$. Then the first component S_ω of S_ξ, properly normalized, satisfies

$$\sqrt{\frac{2}{n}}(1 - \beta_n) S_\omega = \int_0^1 \sigma_{\lfloor sn \rfloor}^{-2} F_{1n}(s) dW_{2n}(s) \xrightarrow{\mathcal{L}} \int_0^1 V(s)^{-1} dW_2(s). \tag{A.8}$$

For the second component $S_{\alpha+\beta}$, we use $F_{2n}(s) = (1 - \beta_n) \sum_{i=0}^{\lfloor sn \rfloor - 1} \beta_n^i \sigma_{\lfloor sn \rfloor - i - 1}^2 / \sigma_{\lfloor sn \rfloor}^2$, which converges weakly to 1, so that

$$\sqrt{\frac{2}{n}}(1 - \beta_n) S_{\alpha+\beta} = \int_0^1 F_{2n}(s) dW_{2n}(s) \xrightarrow{\mathcal{L}} \int_0^1 dW_2(s) = W_2(1). \tag{A.9}$$

Note that $\sqrt{2/n}(1 - \beta_n) = n^{-1}\zeta + \mathrm{o}(n^{-1})$, so that $n^{-1}\zeta$ (or n^{-1}) can also be used as a normalization in (A.8) and (A.9). The third part S_α of S_ω satisfies

$$\sqrt{\frac{1 - \beta_n^2}{n}} S_\alpha = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\sum_{i=0}^{t-1} \sqrt{1 - \beta_n^2} \beta_n^i \frac{\sigma_{t-i-1}^2}{\sigma_t^2} e_{t-i-1} \right) e_t = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t e_t, \tag{A.10}$$

Now u_t is a stationary process with $E(u_t^2) = q$, such that $\sqrt{(1 - \beta_n^2)/n} S_\alpha \xrightarrow{\mathcal{L}} N(0, q)$. As $1 - \beta_n^2 = 2\zeta/\sqrt{2n} + \mathrm{o}(n^{-1/2})$, the normalization in (A.10) is equivalent to $2^{1/4}\zeta^{1/2}n^{-3/4}$. In summary, letting

$$D_{3n} = \begin{bmatrix} \frac{\zeta}{n} I_2 & 0 \\ 0 & \sqrt{\frac{1 - \beta_n^2}{n}} \end{bmatrix}, \tag{A.11}$$

we have $D_{3n}S_\xi = \mathrm{O}_p(1)$. By similar methods, it can be shown that $D_{3n}J_\xi D_{3n}$ converges, and that $D_{2n}J_{\delta\xi}D_{3n}$ converges; the latter converges to zero due to the two parts of the score vector being uncorrelated because $E[\eta_t(\eta_t^2 - 1)] = 0$. □

Proof of Theorem 2. The result (34) follows from Lemma 4 and (14). Previous derivations show that

$$\begin{aligned} D_{2n}\sum_{t=1}^{n} Z_t \Delta X_t &= D_{2n}\sum_{t=1}^{n} Z_t\sigma_t\eta_t + D_{2n}\sum_{t=1}^{n} Z_t Z_t'\nu/n \\ &\xrightarrow{\mathcal{L}} \int_0^1 G(s)V(s)^{1/2}dW_1(s) + \kappa\int_0^1 G(s)Y(s)ds, \end{aligned} \tag{A.12}$$

and similarly

$$D_n\sum_{t=1}^{n} Z_t Z_t' D_n \xrightarrow{\mathcal{L}} \int_0^1 G(s)G(s)'ds. \tag{A.13}$$

Finally,

$$\begin{aligned} \frac{1}{n}\sum_{t=1}^{n}(\Delta X_t - \hat{\delta}_{LS}' Z_t)^2 &= \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2 + \mathrm{o}_P(1) \\ &= \frac{1}{n}\sum_{t=1}^{n}\sigma_t^2 + \frac{1}{n}\sum_{t=1}^{n}\sigma_t^2(\eta_t^2 - 1) + \mathrm{o}_P(1) \\ &\xrightarrow{\mathcal{L}} \int_0^1 V(s)ds. \end{aligned} \tag{A.14}$$

This proves (35). □