Graduate School of Economics, Hitotsubashi University

Discussion Paper Series No. 2022-02

Stochastic Local and Moderate Departures from a Unit Root and Its Application to Unit Root Testing

Mikihito Nishi and Eiji Kurozumi

August 2022

Stochastic Local and Moderate Departures from a Unit Root and Its Application to Unit Root Testing

Mikihito Nishi*[1] and Eiji Kurozumi†[1]

[1]Graduate School of Economics, Hitotsubashi University

August 26, 2022

Abstract

Local-to-unity and moderate-deviations specifications have been popular alternatives to unit root modelling. This paper considers another kind of departures from a unit root, of the form cv_t/T^{β}, where v_t is random and β determines the distance from a unit root. We classify the stochastic departures into two types: local and moderate. This classification task is completed by investigating the asymptotic behavior of unit root tests that assume the stochastic unit root (STUR) processes as the alternative hypothesis. The stochastic local-to-unity model arises when $\beta = 3/4$; in this case, the test statistics have limiting distributions different from those under the unit root null, and their asymptotic powers are greater than size. Moderate deviations emerge when $1/2 \leq \beta < 3/4$, in which case the test statistics diverge. We also propose new tests for a unit root against a STUR, whose construction is based on the limit theory developed in this paper. To evaluate the performance of these new tests, we derive the limiting Gaussian power envelope under the local alternative from an approximate model.

Keywords: random coefficient model, local to unity, moderate deviation, LBI test, power envelope

JEL Codes: C12, C22

*We are grateful to Yukitoshi Matsushita and Yohei Yamamoto for helpful comments. All errors are our responsibility. Address correspondence to Mikihito Nishi, Graduate School of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan; e-mail: ed225007@g.hit-u.ac.jp

†Eiji Kurozumi's research was supported by JSPS KAKENHI Grant Number 19K01585 and 22K01422.

1 Introduction

Consider the following AR(1) model

$$y_t = \rho y_{t-1} + \varepsilon_t, \qquad t = 1, 2, \ldots, T, \tag{1}$$

with $\varepsilon_t \sim \text{i.i.d}(0, \sigma_\varepsilon^2)$. In the unit root literature, much attention has been paid to the specification given by

$$\rho = \rho_T := 1 + \frac{a}{T^\alpha}, \tag{2}$$

where $\alpha \in (0, 1]$ and $a \neq 0$ (e.g., Phillips, 1987; Elliott, Rothenberg, and Stock, 1996; Phillips and Magdalinos, 2007). Under this specification, although the autoregressive coefficient ρ strictly differs from unity by a/T^α, it approaches one as the sample size T increases. When $\alpha = 1$, the coefficient is said to have local departures from a unit root. When $\alpha \in (0, 1)$, ρ is said to have moderate deviations.

The classification of departures (local or moderate) could be better understood in terms of the asymptotic behavior of conventional unit root tests such as those proposed by Dickey and Fuller (1979) (DF, hereafter). When $\alpha = 1$, the DF test statistics have asymptotic distributions different from those under the null of a unit root, but the DF tests are not consistent; that is, the probability that the tests reject the null does not converge to one as T increases. In this sense, the alternative that the coefficient has local departures is near the null of a unit root. In contrast, when $\alpha \in (0, 1)$, the DF test statistics diverge at a rate depending on α, and hence they are consistent. This means moderate deviations are sufficient for the DF tests to detect departures from a unit root with probability approaching one.

In this paper, we consider another specification of the form

$$\rho = \rho_{T,t} := 1 + \frac{cv_t}{T^\beta}, \tag{3}$$

where $v_t \sim \text{i.i.d}(0, 1)$ and β takes some positive value. This specification is similar to (2) in that ρ approaches one as T grows. However, ρ defined by (3) is random because of v_t, and its value at time t depends on the realization of v_t. Hence, equation (3) formulates stochastic departures from a unit root. Another explanation could be given for the distinction between these two types (stochastic and nonstochastic) of departures. The nonstochastic-departures model approaches a unit root process "in mean", i.e., $\mathbb{E}[\rho_T] = 1 + aT^{-\alpha} \to 1$. By contrast, the coefficient having stochastic-departures, keeping its mean one, approaches a unit root "in variance", that is, $\mathbb{V}[\rho_{T,t}] = c^2 T^{-2\beta} \to 0$.

The usefulness of stochastic departures has been recognized in the literature. Lieberman and Phillips (2014, 2017, 2018) considered the AR(1) model with $\rho = \exp(cv_t T^{-1/2})$ and developed estimation and inference theory. They demonstrated this formulation can be successfully applied in economics and finance; for instance, they showed the model leads to a generalization of the well-known Black-Scholes model. Lieberman and Phillips (2020) extended their preceding work by jointly considering the stochastic and nonstochastic departures, i.e., $\rho = \exp(aT^{-1} + cv_t T^{-1/2})$. Tao, Phillips, and Yu (2019) considered the continuous time version of this model and demonstrated that their model can well describe several extreme behaviors of time series such as exuberance followed by collapse. Banerjee, Chevillon, and Kratz (2020) also considered the model with $\rho = \exp(aT^{-\alpha} + cv_t T^{-\alpha/2}),\ 0 < \alpha < 1$, as a model for time series data containing bubbles and flash crashes in its sample path.

One of the purposes of this paper is to classify the stochastic departures (3) into two types: the local and moderate deviations. This task can be effectively completed on the basis of the asymptotic behavior of certain test statistics, as in the case of nonstochastic departures; i.e., the model is local if the test statistics are $O_p(1)$ but their limiting distributions are different from those under the null, and the model is moderate if they diverge (at a rate depending on β). However, we cannot solve this classification problem if we employ the DF tests or other conventional unit root tests. This is because while these tests assume as the alternative a nonstochastic coefficient that differs from one (e.g., (2)), the alternative hypothesis implied by (3) is a stochastic coefficient.

As we will demonstrate later, this task can be successfully completed by utilizing unit root tests that assume as the alternative hypothesis stochastic unit root (STUR) processes, a class of random coefficient autoregressive processes that was originally proposed by Granger and Swanson (1997). Such tests were provided by several authors, including McCabe and Tremayne (1995), Lee (1998) and Nagakura (2009). By investigating the asymptotic behavior of these tests for different values of β, we can determine which values of β correspond to (stochastic) local or moderate departures. McCabe and Smith (1998) derived, under $\beta = 3/4$, the asymptotic distribution of the McCabe and Tremayne's (1995) test statistic. The asymptotic distribution differs from that under the null, which means the $\beta = 3/4$ case corresponds to the local alternative. Our results show that $\beta \in [1/2, 3/4)$ correspond to the moderate departures from a unit root.[1]

[1]Strictly speaking, the moderate-deviations region of β would cover the interval $(0, 1/2)$, but we focus only on the case $\beta \in [1/2, 3/4]$.

There have been several studies addressing similar issues. Phillips and Magdalinos (2007) considered AR(1) models whose coefficient is of the form (2) with $0 < \alpha < 1$ and bridged the gap between the (nonstochastic) local and moderate departures asymptotics. Following their work, Aue (2008) developed asymptotics for the case where

$$\rho = 1 + \frac{a}{T^\alpha} + \frac{cv_t}{T^\beta},$$

with $1/2 < \alpha < 1$ and $\beta > 1/2$, and derived the asymptotic behavior of the OLS estimator of the AR coefficient. McCabe and Smith (1998) dealt with the stochastic coefficient (3) with $\beta = 3/4$ and $1/2$. For the former case, they derived the asymptotic distribution of the McCabe-Tremayne test statistic, and for the latter case they analyzed the behavior of the DF tests. Our work differs from these earlier studies in that we develop asymptotics for the cases $\beta = 3/4$, $\beta \in (1/2, 3/4)$ and $\beta = 1/2$, derive for each case the limiting distributions of several unit root test statistics, and classify the stochastic departures into local and moderate ones.

As an application of the stochastic-departures models, we also propose several new unit root tests whose alternative is that the process has a STUR. The construction of these tests and the analysis of their asymptotic behavior are based on the limit theory developed in this paper.

The remainder of this paper is organized as follows. In Section 2, we review the concept of STUR processes and some existing tests of a unit root against a STUR. In Section 3, we localize the STUR processes by making the stochastic part of the coefficient approach zero at the rate $T^{-\beta}$, $\beta \in [1/2, 3/4]$. We establish some limit theory for this localized model, investigate the asymptotic behavior of some unit root tests, and classify the stochastic departures into the local and moderate deviations. Section 4 proposes several new tests of a unit root against a STUR, which are based on the limit theory developed in this article. In Section 5, we evaluate the asymptotic power of these tests along with that of extant tests. To this end, we derive the limiting Gaussian power envelope from an approximate model and compare it with the extant and new tests' power functions. In Section 6, Monte Carlo experiments are conducted to examine the finite-sample performance of the tests, and we also apply these tests to real data. Section 7 concludes the paper.

2 STUR Processes and Some Related Unit Root Tests

In this section, we briefly review the concept of STUR processes and existing unit root tests that assume them as the alternative.

The STUR processes, which was coined by Granger and Swanson (1997), is a class of random coefficient autoregressive processes given by

$$y_t = (1 + \omega v_t) y_{t-1} + \varepsilon_t, \qquad t = 1, 2, \ldots, T, \tag{4}$$

where $v_t \sim (0, 1)$ and $\varepsilon_t \sim (0, \sigma_\varepsilon^2)$.[2] The autoregressive coefficient of the process (4) randomly changes over time. However, its expectation is assumed to be unity, so that the value of the coefficient is one on average. The STUR model, therefore, includes unit root processes as a special case obtained by letting $\omega^2 = 0$. Granger and Swanson (1997) argued that some economic theory such as the permanent income hypothesis could be well formulated by the STUR model. Since their work, empirical studies applying the STUR model have been conducted. Among such studies are Sollis, Newbold, and Leybourne (2000), Yoon (2005, 2010a,b), and Yau and Hueng (2007).

In model (4), the process becomes an AR(1) process with a unit root when the variance ω^2 of the coefficient is zero; if ω^2 is positive, the process is a STUR process. Hence, one might want to test whether $\omega^2 = 0$, thereby determining whether the coefficient of the time series of interest is constant and one or changes randomly around unity over time. McCabe and Tremayne (1995, hereafter MT) derived a locally best invariant (LBI) test of a unit root against a STUR, of the form

$$\mathrm{MT}_T = T^{-3/2} \hat{\kappa}_T^{-1} \hat{\sigma}_T^{-2} \sum_{t=1}^{T} y_{t-1}^2 \{(\Delta y_t)^2 - \hat{\sigma}_T^2\}, \tag{5}$$

where $\hat{\sigma}_T^2 \coloneqq T^{-1} \sum_{t=1}^{T} (\Delta y_t)^2$ and $\hat{\kappa}_T^2 \coloneqq T^{-1} \sum_{t=1}^{T} \{(\Delta y_t)^2 - \hat{\sigma}_T^2\}^2$. Lee (1998) and Nagakura (2009) derived an LBI test for a more general model

$$y_t = (d + \omega v_t) y_{t-1} + \varepsilon_t, \; -1 < d \leq 1.$$

When $d = 1$, their test statistic is given by

$$\mathrm{LN}_T = \frac{T^{-3/2} \sum_{t=1}^{T} y_{t-1}^2 \{(\Delta y_t)^2 - \hat{\sigma}_T^2\}}{\hat{\kappa}_T \{T^{-3} \sum_{t=1}^{T} y_{t-1}^4 - (T^{-2} \sum_{t=1}^{T} y_{t-1}^2)^2\}^{1/2}}. \tag{6}$$

The MT and Lee-Nagakura (LN, hereafter) tests are both right-tailed. To derive the limiting null distributions of these test statistics, we impose the following assumption.

[2] In fact, Granger and Swanson considered a different but very similar model whose coefficient is given by $\exp(\omega v_t)$.

Assumption 1. $\{\varepsilon_t\}$ *is a sequence of i.i.d random variables with* $\mathbb{E}[\varepsilon_t] = 0, \mathbb{E}[\varepsilon_t^2] = \sigma_\varepsilon^2 > 0$, *and* $\mathbb{E}[\varepsilon_t^4] < \infty$. *Moreover,* $y_0 = o_p(T^{1/2})$.

Define $\eta_t := \varepsilon_t^2 - \sigma_\varepsilon^2$ and $\kappa_\varepsilon^2 := \mathbb{E}[\eta_t^2]$. Define also the partial sum process $(W_{T,\varepsilon}, W_{T,\eta})'$ on $[0,1]$ by $W_{T,\varepsilon}(r) := T^{-1/2}\sigma_\varepsilon^{-1}\sum_{t=1}^{\lfloor Tr \rfloor}\varepsilon_t$ and $W_{T,\eta}(r) := T^{-1/2}\kappa_\varepsilon^{-1}\sum_{t=1}^{\lfloor Tr \rfloor}\eta_t$. Then, it follows from the functional central limit theorem (FCLT) that under Assumption 1, as $T \to \infty$

$$\begin{pmatrix} W_{T,\varepsilon} \\ W_{T,\eta} \end{pmatrix} \Rightarrow \begin{pmatrix} W_\varepsilon \\ W_\eta \end{pmatrix},$$

where W_ε and W_η are standard Brownian motions and $\Rightarrow$ signifies weak convergence. Note that W_ε and W_η are not necessarily independent because of the covariance between ε_t and η_t, i.e., $\psi := \mathbb{E}[\varepsilon_t^3]$. A sufficient condition for them to be independent is that ε_t is symmetric.

Under Assumption 1, the test statistics MT_T and LN_T have the following asymptotic distributions under the null $\omega^2 = 0$:

$$\mathrm{MT}_T \Rightarrow \int_0^1 \left[W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\right] dW_\eta(r), \tag{7}$$

$$\mathrm{LN}_T \Rightarrow \frac{\int_0^1 \left[W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\right] dW_\eta(r)}{\left[\int_0^1 \{W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\}^2 dr\right]^{1/2}}. \tag{8}$$

These convergence results were obtained by MT and Nagakura (2009).

These tests are designed to test the unit root hypothesis $\omega^2 = 0$ against the STUR hypothesis $\omega^2 > 0$. On the other hand, the stochastic-departures model specified by (3) approaches a unit root process along the path $\omega^2 = \omega_T^2 := c^2/T^{2\beta} \searrow 0$. Hence, it is expected that for some local alternative hypothesis $\omega_T^2 = c^2/T^{2\beta} > 0$, or for some β, the MT and LN test statistics have limit distributions different from those under the null, and that for other values of β, they diverge. This observation leads us to utilize these tests as tools by which to determine what size of stochastic departures should be interpreted as local or moderate.

3 Local and Moderate Deviations from Unity

3.1 Assumptions

We turn to local- and moderate-departures models. Our model is obtained by letting $\omega = c/T^\beta$ in model (4), that is,

$$y_t = \left(1 + \frac{c}{T^\beta} v_t\right) y_{t-1} + \varepsilon_t, \qquad t = 1, 2, \ldots, T, \tag{9}$$

where $c^2 \geq 0$. As mentioned in the introduction, the local-departures model arises when $\beta = 3/4$, and the moderate-departures model when $1/2 \leq \beta < 3/4$. For model (9), we impose the following assumption in place of Assumption 1.

Assumption 2. $(\varepsilon_t, v_t)' \sim \text{i.i.d}(0, \Omega)$, *where* $\Omega := \text{diag}(\sigma_\varepsilon^2, 1)$. *Also,* $\mathbb{E}[\varepsilon_t^4] < \infty$ *and* $\mathbb{E}[v_t^8] < \infty$. *Moreover,* $y_0 = o_p(T^{1/2})$.

Assumption 2 could be modified to allow for dependence, that is, serial correlation in the innovations and contemporaneous correlation between ε_t and v_t. However, we employ the simple assumption so as not to let a complicated analysis obscure our main points. For the development of the theory under more general assumptions, interested readers are referred to Lieberman and Phillips (2017, 2018).[3]

Under Assumption 2, we have, from the FCLT,

$$\begin{pmatrix} W_{T,\varepsilon} \\ W_{T,\eta} \\ W_{T,v} \end{pmatrix} \Rightarrow \begin{pmatrix} W_\varepsilon \\ W_\eta \\ W_v \end{pmatrix},$$

where W_v is a standard Brownian motion and $W_{T,v}$ is defined by $W_{T,v}(r) := T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} v_t$.

3.2 Preliminary asymptotic theory

Before analyzing the asymptotic behavior of the MT and LN test statistics under the alternatives, we need to establish some limit theory. It is mainly concerned with the asymptotic behavior of the standardized process $T^{-1/2} y_{\lfloor Tr \rfloor}$, $\hat{\sigma}_T^2$ and $\hat{\kappa}_T^2$, which in turn determine the behavior of the test statistics. The limit theory will also play a key role in constructing new unit root tests in Section 4. Proofs of all the subsequent results are given in the appendix.

Lemma 1. *For model (9) with* $1/2 < \beta \leq 3/4$, *we have, under Assumption 2,*

(a) $Y_T \Rightarrow \sigma_\varepsilon W_\varepsilon$ *in the Skorokhod space* $D[0,1]$, *where* Y_T *is defined by* $Y_T(r) := T^{-1/2} y_{\lfloor Tr \rfloor}$ *for* $r \in [0,1]$,[4]

(b) $\hat{\sigma}_T^2 \xrightarrow{p} \sigma_\varepsilon^2$,

[3] Lieberman and Phillips (2017, 2018) considered a model in which the random component of the coefficient is observable, while we do not assume v_t's are observed.

[4] Aue (2008) showed that the finite-dimensional distributions of Y_T weakly converge to those of a Brownian motion (his Theorem 2.4). Lemma 1(a) extends this result to the weak convergence of Y_T as a stochastic process in $D[0,1]$.

(c) $\hat{\kappa}_T^2 \xrightarrow{p} \kappa_\varepsilon^2$.

This lemma states that as long as $\beta \in (1/2, 3/4]$, the standardized process Y_T weakly converges to the Brownian motion $\sigma_\varepsilon W_\varepsilon$ and estimators $\hat{\sigma}_T^2$ and $\hat{\kappa}_T^2$ are consistent, as in the pure unit root case. For the special case $\beta = 1/2$, however, we need to establish limit theory other than that for the case $\beta \in (1/2, 3/4]$. First, the large-sample behavior of the standardized process $Y_T(\cdot) = T^{-1/2} y_{\lfloor T \cdot \rfloor}$ is no longer approximated by the Brownian motion $\sigma_\varepsilon W_\varepsilon(\cdot)$, and estimators $\hat{\sigma}_T^2$ and $\hat{\kappa}_T^2$ have nondegenerate asymptotic distributions, as the following result indicates.[5]

Lemma 2. *For model (9) with $\beta = 1/2$, under Assumption 2,*

(a) $Y_T \Rightarrow \sigma_\varepsilon Y_c$, *where*

$$Y_c(r) := \exp\left(cW_v(r) - \frac{c^2}{2} r\right) \int_0^r \exp\left(-cW_v(s) + \frac{c^2}{2} s\right) dW_\varepsilon(s), \tag{10}$$

for $r \in [0, 1]$,

(b)

$$\hat{\sigma}_T^2 \Rightarrow c^2 \sigma_\varepsilon^2 \int_0^1 Y_c^2(r) dr + \sigma_\varepsilon^2,$$

(c) $\hat{\kappa}_T^2$ *has a nondegenerate limiting distribution.*

Although the standardized process Y_T still weakly converges, the limiting process $\sigma_\varepsilon Y_c$ differs from the Brownian motion $\sigma_\varepsilon W_\varepsilon$ unless $c = 0$ (i.e., the unit root case). The limiting process Y_c belongs to the class of continuous-time processes considered by Tao et al. (2019). They considered a continuous-time process that satisfies the following stochastic differential equation:

$$dy(t) = y(t)[\tilde{\mu} dt + c dW_v(t)] + \sigma_\varepsilon dW_\varepsilon(t). \tag{11}$$

The solution of this equation takes the form

$$y(t) = \exp\left[cW_v(t) + \big(\tilde{\mu} - \frac{c^2}{2}\big) t\right] y(0) + \sigma_\varepsilon \int_0^t \exp\left[c\big(W_v(t) - W_v(s)\big) + \big(\tilde{\mu} - \frac{c^2}{2}\big)(t - s)\right] dW_\varepsilon(s), \tag{12}$$

[5] McCabe and Smith (1998) derived the first-order approximation (in c) of the limit process of Y_T, which is different from that of Y_c given in Lemma 2(a). In fact, Lemma 2(a) corrects their result.

for which letting $y(0) = 0$ and $\tilde{\mu} = 0$ yields $\sigma_\varepsilon Y_c(t)$ defined by (10). This relation between (10) and (12) stems from the fact that the process defined in (11) is a modified version of the Ornstein-Uhlenbeck (OU) process

$$dy(t) = y(t)\tilde{\mu}dt + \sigma_\varepsilon dW_\varepsilon(t),$$

which is the limit of the standardized process constructed from the AR(1) process $y_t = (1 + \tilde{\mu}/T)y_{t-1} + \varepsilon_t$. The continuous-time model (11) extends the OU process by introducing $cdW_v(t)$ to the drift component. This is comparable with introducing stochastic moderate deviations $cv_t/\sqrt{T}$ into the coefficient $1 + \tilde{\mu}/T$ in the (discrete-time) AR model. Therefore, $\tilde{\mu}$ and c in (11) correspond to the nonstochastic and stochastic localizing parameters in the AR(1) model, respectively. Indeed, one can show that the standardized process constructed from $y_t = (1 + \tilde{\mu}/T + cv_t/\sqrt{T})y_{t-1} + \varepsilon_t$ weakly converges to (12) (see also Föllmer and Schweizer, 1993).

Remark 3. As shown in the appendix, a different result emerges if we allow for endogeneity, that is, a nonzero correlation between the coefficient and disturbance. When $\sigma_{\varepsilon v} := \mathbb{E}[\varepsilon_t v_t] \neq 0$, the limit process of Y_T is expressed as

$$\exp\left(cW_v(r) - \frac{c^2}{2}r\right)\left\{\sigma_\varepsilon \int_0^r \exp\left(-cW_v(s) + \frac{c^2}{2}s\right)dW_\varepsilon(s) - c\sigma_{\varepsilon v}\int_0^r \exp\left(-cW_v(s) + \frac{c^2}{2}s\right)ds\right\}, \tag{13}$$

which reduces to (10) if $\sigma_{\varepsilon v} = 0$. This is, too, included in the class of processes considered by Tao et al. (2019) for which endogeneity is taken into account. Although analyzing models with endogeneity will lead to more general results, we will focus on the model without endogeneity in the subsequent analysis to keep the main points of this paper clear.

Remark 4. Lemma 2(a) is related to results of Lieberman and Phillips (2014, 2017, 2018). The following process is a simplified version of their original model:

$$y_t = \exp\left(\frac{c}{\sqrt{T}}v_t\right)y_{t-1} + \varepsilon_t. \tag{14}$$

Model (14) may seem to be asymptotically equivalent to model (9) with $\beta = 1/2$, because the autoregressive coefficient satisfies

$$\exp\left(\frac{c}{\sqrt{T}}v_t\right) = 1 + \frac{c}{\sqrt{T}}v_t + O_p(T^{-1}). \tag{15}$$

However, our model (9) with $\beta = 1/2$ is not equivalent to (14), even asymptotically. One can show that, when $\sigma_{\varepsilon v} = 0$, model (14) can be written as

$$T^{-1/2} y_{\lfloor Tr \rfloor} = \exp\Bigg(cT^{-1/2} \sum_{k=1}^{\lfloor Tr \rfloor} v_k\Bigg) \sum_{s=1}^{\lfloor Tr \rfloor} \exp\Bigg(-cT^{-1/2} \sum_{k=1}^{s-1} v_k\Bigg)\Bigg(T^{-1/2}\varepsilon_s\Bigg) + o_p(1), \qquad (16)$$
$$\Rightarrow \exp\Big(cW_v(r)\Big) \int_0^r \exp\Big(-cW_v(s)\Big) dW_\varepsilon(s),$$

which differs from Lemma 2(a) by terms in the exponents, namely, $-c^2 r/2$ and $c^2 s/2$ in the integral (see, for example, Section 5 of Lieberman and Phillips (2014)). On the other hand, as shown in the appendix, our model can be written as (when $\sigma_{\varepsilon v} = 0$)

$$T^{-1/2} y_{\lfloor Tr \rfloor} = \exp\Bigg(cT^{-1/2} \sum_{k=1}^{\lfloor Tr \rfloor} v_k - \frac{c^2}{2} T^{-1} \sum_{k=1}^{\lfloor Tr \rfloor} v_k^2\Bigg)$$
$$\times \sum_{s=1}^{\lfloor Tr \rfloor} \exp\Bigg(-cT^{-1/2} \sum_{k=1}^{s-1} v_k + \frac{c^2}{2} T^{-1} \sum_{k=1}^{s-1} v_k^2\Bigg)\Bigg(T^{-1/2}\varepsilon_s\Bigg) + o_p(1), \qquad (17)$$

where the terms $-\frac{c^2}{2} T^{-1} \sum_{k=1}^{\lfloor Tr \rfloor} v_k^2$ and $\frac{c^2}{2} T^{-1} \sum_{k=1}^{s-1} v_k^2$ asymptotically play the roles of $-c^2 r/2$ and $c^2 s/2$ in the integral in Lemma 2(a), respectively. The difference between the two models is due to the fact that the effect of the $O_p(T^{-1})$ term in (15) (namely, $c^2 v_t^2/2T$) in fact does not vanish as $T \to \infty$. This amounts to model (14) being asymptotically equivalent to

$$y_t = \Bigg(1 + \frac{c}{\sqrt{T}} v_t + \frac{c^2}{2T} v_t^2\Bigg) y_{t-1} + \varepsilon_t.$$

Another point to be noted in Lemma 2 is that $\hat{\sigma}_T^2$ and $\hat{\kappa}_T^2$ are not consistent estimators any longer. In particular, $\hat{\sigma}_T^2$ overestimates σ_ε^2 in large samples due to the positive term $c^2 \sigma_\varepsilon^2 \int_0^1 Y_c^2(r) dr$. However, one can construct an estimator of σ_ε^2 that is consistent under the moderate-deviations case (and also under the pure unit root case). We will return to this problem in Section 4.

3.3 Asymptotic behavior of the test statistics

Using two lemmas developed in the previous subsection, we can analyze the asymptotic behavior of the MT and LN test statistics. The next result gives the asymptotic expression for these statistics under $1/2 < \beta \leq 3/4$:

Theorem 5. *For model (9) with $1/2 < \beta \leq 3/4$, under Assumption 2,*

(a)

$$\mathrm{MT}_T \overset{a}{\sim} \int_0^1 \left(W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds \right) dW_\eta(r) + T^{3/2-2\beta} \frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon} \int_0^1 \left[W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds \right]^2 dr, \tag{18}$$

(b)

$$\mathrm{LN}_T \overset{a}{\sim} \frac{\int_0^1 \left[W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\right] dW_\eta(r)}{\left[\int_0^1 \{W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\}^2 dr\right]^{1/2}} + T^{3/2-2\beta} \frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon} \left[\int_0^1 \left\{ W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds \right\}^2 dr \right]^{1/2}, \tag{19}$$

(c)

$$\mathrm{DF}_T = T(\hat{\rho}_T - 1) \Rightarrow \frac{\int_0^1 W_\varepsilon(r)dW_\varepsilon(r)}{\int_0^1 W_\varepsilon^2(r)dr},$$

where $\hat{\rho}_T := \sum_{t=1}^T y_{t-1}y_t / \sum_{t=1}^T y_{t-1}^2$.

Note that McCabe and Smith (1998) derived the limiting distribution (18) under $\beta = 3/4$, and the behavior of DF_T under $\beta = 3/4$ could be expected from Theorem 3 of McCabe and Smith (1998) by letting $\tau^2 \to 0$ in their notation.

There are three points worth mentioning. First, when $\beta = 3/4$, both the MT and LN test statistics have the local asymptotic distributions of the form

$$\mathrm{MT}_T \Rightarrow \int_0^1 \left(W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds \right) dW_\eta(r) + \frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon} \int_0^1 \left[W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds \right]^2 dr,$$

and

$$\mathrm{LN}_T \Rightarrow \frac{\int_0^1 \left[W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\right] dW_\eta(r)}{\left[\int_0^1 \{W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\}^2 dr\right]^{1/2}} + \frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon} \left[\int_0^1 \left\{ W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds \right\}^2 dr \right]^{1/2},$$

which differ from the null distributions (7) and (8) when $c^2 > 0$. For each test, the first term of the local distribution is identical to the asymptotic null distribution. The second terms, which are positive, shift the null distributions to the right and hence contribute to the increase in power. The power of each test becomes greater if the localizing parameter c^2 gets larger. Observe that the power also increases when σ_ε^2 increases or κ_ε^2 decreases. For the special case where ε_t is normally distributed, $\sigma_\varepsilon^2/\kappa_\varepsilon = 2^{-1/2}$ and thus the effects of σ_ε^2 and κ_ε^2 are constant.

The second point is that when $1/2 < \beta < 3/4$, the second terms of the asymptotic expressions in (18) and (19) dominate at the rate $O_p(T^{3/2-2\beta})$. Hence, the MT and LN test statistics diverge to positive infinity as the sample size increases. Because the MT and LN tests are right-tailed, we conclude from this theorem that the interval $1/2 < \beta < 3/4$ is included in the stochastic-moderate-deviations region.

The other point is that the DF (coefficient) test statistic converges to the null distribution either when $\beta = 3/4$ or when $1/2 < \beta < 3/4$, which means the test has power equal to size against these stochastic alternatives. In other words, the DF test cannot detect the stochastic local and moderate departures from unity. An intuitive explanation for this result is that because the standardized process $T^{-1/2}y_{\lfloor T\cdot \rfloor}$ behaves like $\sigma_\varepsilon W_\varepsilon(\cdot)$ as it does under the unit root null, the DF test takes y_t as a (pure) unit root process. It follows from this observation that conventional unit root tests that assume as the alternative hypothesis a nonstochastic coefficient differing from one are inappropriate when the true process has stochastic local or moderate departures from a unit root.

Next, using Lemma 2, we can derive the limiting behavior of the MT and LN statistics under $\beta = 1/2$.

Theorem 6. *For model (9) with $\beta = 1/2$, under Assumption 2,*

(a)

$$\mathrm{MT}_T \overset{a}{\sim} \frac{\kappa_\varepsilon \sigma_\varepsilon^2}{\kappa \sigma^2} \int_0^1 \left(Y_c^2(r) - \int_0^1 Y_c^2(s)ds \right) dW_\eta(r) + T^{1/2} \frac{c^2 \sigma_\varepsilon^4}{\kappa \sigma^2} \int_0^1 \left[Y_c^2(r) - \int_0^1 Y_c^2(s)ds \right]^2 dr,$$

where σ^2 and κ^2 are random variables distributed according to the limiting distributions of $\hat{\sigma}_T^2$ and $\hat{\kappa}_T^2$, respectively (see Lemma 2 and its proof),

(b)

$$\mathrm{LN}_T \overset{a}{\sim} \frac{\kappa_\varepsilon}{\kappa} \frac{\int_0^1 \left[Y_c^2(r) - \int_0^1 Y_c^2(s)ds \right] dW_\eta(r)}{\left[\int_0^1 \{ Y_c^2(r) - \int_0^1 Y_c^2(s)ds \}^2 dr \right]^{1/2}} + T^{1/2} \frac{c^2 \sigma_\varepsilon^2}{\kappa} \left[\int_0^1 \left\{ Y_c^2(r) - \int_0^1 Y_c^2(s)ds \right\}^2 dr \right]^{1/2}.$$

Given this result and Theorem 5, we are led to the conclusion that the stochastic-moderate-deviations interval of values of β in (9) includes $[1/2, 3/4)$.

4 New Tests of a Unit Root Against a STUR

In this section, as an application of the asymptotic theory developed thus far, we propose several new unit root tests.

Consider model (9) and suppose one wants to test the unit root null $c^2 = 0$ against the alternative $c^2 > 0$. The tests we propose here are motivated under the specification of $\beta = 1/2$. The reason why we consider the moderate deviations given by $\beta = 1/2$ rather than local departures is that $\hat{\sigma}_T^2$, an estimator of σ_ε^2, is consistent under the null but inconsistent under the alternative, when $\beta = 1/2$. This observation leads us to consider exploiting the Hausman principle: If an estimator $\tilde{\sigma}_T^2$ of σ_ε^2 is available that is consistent under both the null and alternative, then one could use the normalized difference between $\hat{\sigma}_T^2$ and $\tilde{\sigma}_T^2$ to test the hypothesis of $c^2 = 0$. Indeed, such an estimator can be constructed under Assumption 2, as the next result shows.

Lemma 7. *For model (9) with $\beta = 1/2$, under Assumption 2, $\tilde{\sigma}_T^2 \xrightarrow{p} \sigma_\varepsilon^2$, where*

$$\tilde{\sigma}_T^2 := \frac{\sum_{t=1}^T (\Delta y_t)^2 \sum_{t=1}^T y_{t-1}^4 - \sum_{t=1}^T (\Delta y_t)^2 y_{t-1}^2 \sum_{t=1}^T y_{t-1}^2}{T \sum_{t=1}^T y_{t-1}^4 - \left(\sum_{t=1}^T y_{t-1}^2\right)^2}. \tag{20}$$

Although $\tilde{\sigma}_T^2$ is a consistent estimator of σ_ε^2, finite-sample justifications for this estimator are rather weak. For one thing, $\tilde{\sigma}_T^2$ is not necessarily positive for finite T, because the denominator is always positive but the numerator is not. Another problem is that $\tilde{\sigma}_T^2$ converges very slowly to σ_ε^2. These two shortcomings of $\tilde{\sigma}_T^2$, what is worse, are aggravated when the localizing parameter c^2 is large. In Table 1, several quantiles of simulated finite-sample distributions of $\tilde{\sigma}_T^2$ and $\hat{\sigma}_T^2$ are displayed.

For the null case ($c^2 = 0$), $\hat{\sigma}_T^2$ estimates the true variance $\sigma_\varepsilon^2 = 1$ with reasonable precision, particularly when T is large. On the other hand, $\tilde{\sigma}_T^2$ can take a negative value if T is not so large. Moreover, the convergence of $\tilde{\sigma}_T^2$ needs sample size to be so large as to be unrealistic in practice. When the localizing coefficient c^2 is nonzero, the problem becomes more serious. When $c^2 = 0.8$, we see the overestimation by $\hat{\sigma}_T^2$, as is predicted by the asymptotic theory. As for $\tilde{\sigma}_T^2$, the probability that it takes an extreme value is greater than under the null, especially for small T. Although it gets concentrated around the true value as the sample size grows, the convergence speed is slow. Given these shortcomings, it will be sensible to regard $\tilde{\sigma}_T^2$ just as a building block of test statistics, rather than as an estimator of practical use.

To obtain a Hausman test statistic that has an asymptotic null distribution, the following normalization suffices:

$$\mathrm{H}_T := \frac{\sqrt{T}(\hat{\sigma}_T^2 - \tilde{\sigma}_T^2)}{\hat{\kappa}_T}. \tag{21}$$

Hereafter, we shall refer to H_T as Hausman type test or simply H test.

We propose another test statistic. This test is based on the observation that

$$\hat{c}_T^2 := \frac{\hat{\sigma}_T^2 - \tilde{\sigma}_T^2}{T^{-2}\sum_{t=1}^T y_{t-1}^2} \xrightarrow{p} c^2,$$

which is immediately obtained from Lemma 2(b), Lemma 7 and the continuous mapping theorem. When the null is true, $\hat{c}_T^2$ converges in probability to zero. Hence, it is expected that after proper normalization, we can derive the asymptotic null distribution of $\hat{c}_T^2$. Indeed, it suffices to normalize it by

$$\begin{aligned} \mathrm{C}_T &:= \sqrt{T}\hat{c}_T^2 \times \frac{\hat{\sigma}_T^2}{\hat{\kappa}_T} \\ &= \mathrm{H}_T \times \frac{\hat{\sigma}_T^2}{T^{-2}\sum_{t=1}^T y_{t-1}^2}. \end{aligned} \tag{22}$$

We shall call C_T coefficient test or C test.

Remark 8. There is a close relation among the MT, LN, H and C tests, as the following equations indicate (they are proven in the appendix):

$$\mathrm{LN}_T = \mathrm{MT}_T \times \frac{\hat{\sigma}_T^2}{\{T^{-3}\sum_{t=1}^T y_{t-1}^4 - (T^{-2}\sum_{t=1}^T y_{t-1}^2)^2\}^{1/2}}, \tag{23}$$

$$\mathrm{H}_T = \mathrm{MT}_T \times \frac{\hat{\sigma}_T^2 T^{-2}\sum_{t=1}^T y_{t-1}^2}{T^{-3}\sum_{t=1}^T y_{t-1}^4 - (T^{-2}\sum_{t=1}^T y_{t-1}^2)^2}, \tag{24}$$

$$\mathrm{C}_T = \mathrm{MT}_T \times \frac{\hat{\sigma}_T^4}{T^{-3}\sum_{t=1}^T y_{t-1}^4 - (T^{-2}\sum_{t=1}^T y_{t-1}^2)^2}. \tag{25}$$

These expressions shall be exploited to derive the asymptotic behavior of the H and C test statistics.

Theorem 9. *For model (9) with $1/2 < \beta \le 3/4$, under Assumption 2,*

(a)

$$\mathrm{H}_T \overset{a}{\sim} \int_0^1 W_\varepsilon^2(r)dr \frac{\int_0^1 [W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds] dW_\eta(r)}{\int_0^1 \{W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\}^2 dr} + T^{3/2-2\beta}\frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon}\int_0^1 W_\varepsilon^2(r)dr,$$

(b)

$$\mathrm{C}_T \overset{a}{\sim} \frac{\int_0^1 [W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds] dW_\eta(r)}{\int_0^1 \{W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\}^2 dr} + T^{3/2-2\beta}\frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon};$$

and with $\beta = 1/2$,

(c)

$$\mathrm{H}_T \overset{a}{\sim} \frac{\kappa_\varepsilon}{\kappa} \int_0^1 Y_c^2(r)dr \frac{\int_0^1 \left[Y_c^2(r) - \int_0^1 Y_c^2(s)ds\right] dW_\eta(r)}{\int_0^1 Y_c^4(r)dr - \left(\int_0^1 Y_c^2(r)dr\right)^2} + T^{1/2} \frac{c^2\sigma_\varepsilon^2}{\kappa} \int_0^1 Y_c^2(r)dr,$$

where σ^2 and κ^2 are random variables distributed according to the limiting distributions of $\hat{\sigma}_T^2$ and $\hat{\kappa}_T^2$, respectively (see Lemma 2 and its proof),

(d)

$$\mathrm{C}_T \overset{a}{\sim} \frac{\kappa_\varepsilon \sigma^2}{\kappa \sigma_\varepsilon^2} \frac{\int_0^1 \left[Y_c^2(r) - \int_0^1 Y_c^2(s)ds\right] dW_\eta(r)}{\int_0^1 Y_c^4(r)dr - \left(\int_0^1 Y_c^2(r)dr\right)^2} + T^{1/2} \frac{c^2\sigma^2}{\kappa}.$$

The null limiting distributions of H_T and C_T are given by the first terms of (a) and (b), respectively. Although H and C tests are constructed under the specification of $\beta = 1/2$, we see from Theorem 9 that they have nonnegligible asymptotic power when $\beta = 3/4$. Moreover, they are consistent not only when $\beta = 1/2$ but also when $1/2 < \beta < 3/4$. It should be emphasized here that the case $\beta = 1/2$ was just used to motivate these tests, and such tests in fact work well for other values of β.

Remark 10. The limiting distributions displayed in Theorems 5, 6 and 9 all depend on the third moment of ε_t, a nuisance parameter, through the covariance coefficient $\psi/(\kappa_\varepsilon \sigma_\varepsilon)$ between W_ε and W_η. In the subsequent analysis, we shall assume that $\psi = \mathbb{E}[\varepsilon_t^3] = 0$. Given that $\psi = 0$, the limiting null distribution of LN_T reduces to the standard normal distribution (see Nagakura, 2009). Nagakura (2009) proposed modified LN tests that are independent of ψ and verified they perform well either when $\psi = 0$ or when $\psi \neq 0$. The modification by Nagakura (2009) will be applicable to our tests, although we do not consider it here.

Table 2 contains critical values of the MT, H and C tests (when $\psi = 0$). Since critical values of the LN test are based on the standard normal distribution, we omit them from the table. They are all right-tailed tests. The asymptotic null distributions were simulated through 100,000 Monte Carlo replications where $(W_\varepsilon, W_\eta)'$ were approximated by normalized cumulative sums with 1,000 steps.

5 Evaluation of the Asymptotic Power

In Section 4, we proposed two new tests, the Hausman and coefficient tests, hoping they perform better than extant ones such as the MT and LN tests. When evaluating power properties of several tests, it is useful to compare their power functions with the power envelope, which is obtained under restrictive distributional assumptions. Because there is no uniformly most powerful test in our case, we derive the power envelope by computing the powers of a sequence of the most powerful (or point optimal) tests, which can be derived using the Neyman-Pearson lemma.[6]

For our model (9) with $\beta = 3/4$, the likelihood function l_T with a given y_0, under the Gaussian assumption on (ε_t, v_t), is given by

$$l_T(c^2, \sigma_\varepsilon^2) = (2\pi)^{-T/2} \prod_{t=1}^{T} (c^2 T^{-3/2} y_{t-1}^2 + \sigma_\varepsilon^2)^{-1/2} \times \exp\left[-\frac{1}{2} \sum_{t=1}^{T} \frac{(y_t - y_{t-1})^2}{c^2 T^{-3/2} y_{t-1}^2 + \sigma_\varepsilon^2} \right].$$

Then, by the Neyman-Pearson lemma, the most powerful (MP) test for $\mathrm{H}_0 : c^2 = 0$ vs $\mathrm{H}_1 : c^2 = \bar{c}^2$ for a given $\bar{c}^2$ rejects the null when the following statistic takes a large value:

$$\begin{aligned} q_T(\bar{c}^2, \sigma_\varepsilon^2) &:= 2\{\log l_T(\bar{c}^2, \sigma_\varepsilon^2) - \log l_T(0, \sigma_\varepsilon^2)\} \\ &= -\sum_{t=1}^{T} \log(\bar{c}^2 T^{-3/2} \tilde{y}_{t-1}^2 + 1) - \sum_{t=1}^{T} \frac{(\tilde{y}_t - \tilde{y}_{t-1})^2}{\bar{c}^2 T^{-3/2} y_{t-1}^2 + 1} \\ &\quad + \sum_{t=1}^{T} (\tilde{y}_t - \tilde{y}_{t-1})^2, \end{aligned}$$

where $\tilde{y}_t := y_t / \sigma_\varepsilon$. Note that $q_T(\bar{c}^2, \sigma_\varepsilon^2)$ is infeasible since σ_ε^2 is an unknown nuisance parameter. One possible way to obtain a feasible MP test would be to replace σ_ε^2 with the consistent estimator $\hat{\sigma}_T^2$, following Elliott et al. (1996). If the limiting distribution of $q_T(\bar{c}^2, \hat{\sigma}_\varepsilon^2)$ coincides with that of $q_T(\bar{c}^2, \sigma_\varepsilon^2)$, we can compute the asymptotic power envelope based on the common limiting distribution. Unfortunately, this is not the case as the following proposition shows.

Proposition 11. *For model (9) with $\beta = 3/4$, under Assumption 2 with (ε_t, v_t) being Gaussian, we have*

$$q_T(\bar{c}^2, \sigma_\varepsilon^2) \Rightarrow \bar{c}^2 \left[\sqrt{2} \int_0^1 W_\varepsilon^2 dW_\eta(r) + \left(c^2 - \frac{\bar{c}^2}{2} \right) \int_0^1 W_\varepsilon^4(r) dr \right] =: k(c^2, \bar{c}^2),$$

[6] Elliott et al. (1996) used this strategy to obtain the local asymptotic Gaussian power envelope for unit root tests.

and

$$q_T(\bar{c}^2, \hat{\sigma}_T^2) \Rightarrow \bar{c}^2\Bigg[\sqrt{2}\int_0^1\left(W_\varepsilon^2 - \int_0^1 W_\varepsilon^2(s)ds\right)dW_\eta(r) + c^2\int_0^1\left\{W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\right\}^2 dr - \frac{\bar{c}^2}{2}\int_0^1 W_\varepsilon^4(r)dr\Bigg] =: \hat{k}(c^2, \bar{c}^2),$$

where W_ε and W_η are independent.

We note that W_ε and W_η may be dependent and $\sqrt{2}$ in the limiting distributions is replaced by $\kappa_\varepsilon/\sigma_\varepsilon^2$ without the assumption of normality (see the proof), although we do not pursue such a general case when considering the optimality.

This proposition tells us that $q_T(\bar{c}^2, \sigma_\varepsilon^2)$ and $q_T(\bar{c}^2, \hat{\sigma}_T^2)$ have different asymptotic distributions, which implies that the asymptotic power functions based on $\hat{k}(c^2, \bar{c}^2)$ may not be relevant references in our case. Therefore, we need to seek another way to derive the MP tests without the knowledge of σ_ε^2.

Ideally, one solution would be to derive the point optimal test in a class of invariant tests with respect to σ_ε^2. In fact, McCabe and Tremayne (1995) derived the locally best ($\bar{c}^2 \to 0$) invariant test by considering the transformation of scale invariance, where $\{y_t/y_1\}_{t=2}^T$ was used as the maximal invariant. However, it is difficult to derive the point optimal invariant test for a general value of $\bar{c}^2$ based on the maximal invariant $\{y_t/y_1\}_{t=2}^T$. This is because the joint density f of the maximal invariant and y_1 depends on the latter variable in such a complicated way that it is difficult to integrate out y_1 from f to get the density only of the maximal invariant (see McCabe and Tremayne (1995) for details).

Because of such a difficulty of proceeding with $l_T(c^2, \sigma_\varepsilon^2)$, we now consider using the power envelope derived from a quasi likelihood $l_T^*(c^2, \sigma_\varepsilon^2)$ that is "near" and more tractable than the exact likelihood $l_T(c^2, \sigma_\varepsilon^2)$. Of course, the power envelope derived from l_T^* is not necessarily identical to the exact counterpart based on l_T. We emphasize here that the motivation for introducing l_T^* is to obtain benchmark test statistics; once we obtain the MP test statistics from l_T^*, we eventually derive their limiting distributions under (9) and use the derived power envelope as a benchmark. To find such a likelihood, we approximate the exact model in the following way. First, from model (9), a simple calculation yields

$$(\Delta y_t)^2 = \sigma_\varepsilon^2 + \omega_T^2 y_{t-1}^2 + \xi_t, \tag{26}$$

where $\xi_t := \omega_T^2 y_{t-1}^2(v_t^2 - 1) + 2\omega_T y_{t-1}\varepsilon_t v_t + (\varepsilon_t^2 - \sigma_\varepsilon^2)$ and $\omega_T = cT^{-3/4}$. Note that $\mathbb{E}[\xi_t] = 0$ and $\mathbb{E}[y_{t-1}^2\xi_t] = 0$ since (ε_t, v_t) is i.i.d and $\mathbb{E}[\varepsilon_t v_t] = 0$. Hence, model (26) could be viewed as the linear regression model with ξ_t playing the role of the disturbance.[7]

Our approximation strategy comes from the observation that, under the stronger assumption that $\mathbb{E}[|\varepsilon_t|^{8+\delta}] < \infty$ and $\mathbb{E}[|v_t|^{8+\delta}] < \infty$ for some $\delta > 0$, the disturbance ξ_t in (26) can be expressed as $\xi_t = \varepsilon_t^2 - \sigma_\varepsilon^2 + o_p(1)$. In view of this observation, we approximate (26) by the following Gaussian model:

$$(\Delta y_t)^2 = \sigma_\varepsilon^2 + \omega_T^2 y_{t-1}^2 + \xi_t^*, \tag{27}$$

where $\xi_t^* \sim$ i.i.d $N(0, \kappa_\varepsilon^2)$. Again, note that model (27) is an approximate one. The best we can hope for is that model (27) approximates (26), or (9) so well that the power envelope based on (27) serves as a good benchmark. Fortunately, the point optimal tests derived from (27) coincide with those based on (9), when σ_ε^2 is known. To state this precisely, let $l_T^*(c^2, \sigma_\varepsilon^2, \kappa_\varepsilon^2)$ be the likelihood for model (27):

$$l_T^*(c^2, \sigma_\varepsilon^2, \kappa_\varepsilon^2) = (2\pi\kappa_\varepsilon^2)^{-T/2} \exp\left[-\frac{1}{2\kappa_\varepsilon^2}\sum_{t=1}^{T}\left\{(\Delta y_t)^2 - \sigma_\varepsilon^2 - c^2 T^{-3/2} y_{t-1}^2\right\}^2\right].$$

Let $q_T^*(\bar{c}^2, \sigma_\varepsilon^2, \kappa_\varepsilon^2)$ be the point optimal test based on model (27) under the assumption that σ_ε^2 and κ_ε^2 are known, which is given by

$$\begin{aligned} q_T^*(\bar{c}^2, \sigma_\varepsilon^2, \kappa_\varepsilon^2) &= \log l_T^*(\bar{c}^2, \sigma_\varepsilon^2, \kappa_\varepsilon^2) - \log l_T^*(0, \sigma_\varepsilon^2, \kappa_\varepsilon^2) \\ &= \frac{\bar{c}^2}{\kappa_\varepsilon^2}\left[T^{-3/2}\sum_{t=1}^{T} y_{t-1}^2\left\{(\Delta y_t)^2 - \sigma_\varepsilon^2\right\} - \frac{\bar{c}^2}{2}T^{-3}\sum_{t=1}^{T} y_{t-1}^4\right]. \end{aligned}$$

Proposition 12. *For model (9) with $\beta = 3/4$ under Assumption 2, we have*

$$\frac{\kappa_\varepsilon^2}{\sigma_\varepsilon^4} q_T^*(\bar{c}^2, \sigma_\varepsilon^2, \kappa_\varepsilon^2) - q_T(\bar{c}^2, \sigma_\varepsilon^2) \xrightarrow{p} 0 \qquad \textit{and} \qquad q_T^*(\bar{c}^2, \sigma_\varepsilon^2, \kappa_\varepsilon^2) - q_T^*(\bar{c}^2, \sigma_\varepsilon^2, \hat{\kappa}_T^2) \xrightarrow{p} 0.$$

From Proposition 12, we can see that model (27) well approximates the exact one, (9), in that these two models induce the asymptotically equivalent point optimal tests (when σ_ε^2 is known). Furthermore, we can see that the LN test is obtained as the t-test for $\mathrm{H}_0 : \omega_T^2 = 0$ in model (27). Based on these observations, it will be reasonable to use the power envelope obtained from (27) as a benchmark for the power functions of the LN and other tests.

[7] The idea of "linearizing" model (9) into (26) is not entirely new in itself. For example, Horváth and Trapani (2019) proposed a test statistic for coefficient randomness in autoregressions using this linearization.

To deal with unknown σ_ε^2 in (27), it is natural to investigate a class of location invariant tests because it appears as a constant term in (27). To derive the location invariant test with respect to a constant term, let us define $z_t := (\Delta y_t)^2$ and $x_t := y_{t-1}^2$. Define also $Z := (z_1, z_2, \ldots, z_T)'$, $\iota := (1, 1, \ldots, 1)'$, $X := (x_1, x_2, \ldots x_T)'$ and $\Xi := (\xi_1, \xi_2, \ldots, \xi_T)'$. Then, model (27) can be expressed in matrix notation as

$$Z = \iota\sigma_\varepsilon^2 + X\omega_T^2 + \Xi, \tag{28}$$

where $\Xi \sim N(0, \kappa_\varepsilon^2 I_T)$. Define $M := I_T - \iota(\iota'\iota)^{-1}\iota'$. Then, according to Tanaka (2017), there exists a $T \times (T-1)$ matrix H such that $H'H = I_{T-1}$ and $HH' = M$. Note that $H'\iota = 0$ since $M\iota = HH'\iota = 0$. Thus, premultiplying (28) by H, we get

$$H'Z = H'X\omega_T^2 + H'\Xi. \tag{29}$$

Now, letting $F(c^2) := Z - Xc^2T^{-3/2}$ and $\mathcal{L}_T^*(c^2, \kappa_\varepsilon^2)$ be the likelihood of (29), and noting that $H'F(c^2) = H'\Xi \sim N(0, \kappa_\varepsilon^2 I_{T-1})$, we have

$$\begin{aligned}
\mathcal{L}_T^*(c^2, \kappa_\varepsilon^2) &= (2\pi\kappa_\varepsilon^2)^{\frac{-(T-1)}{2}} \exp\left[-\frac{1}{2\kappa_\varepsilon^2} F(c^2)' M F(c^2)\right] \\
&= (2\pi\kappa_\varepsilon^2)^{\frac{-(T-1)}{2}} \exp\left[-\frac{1}{2\kappa_\varepsilon^2} \sum_{t=1}^T \left\{(\Delta y_t)^2 - c^2 T^{-3/2} y_{t-1}^2 - T^{-1} \sum_{t=1}^T \big((\Delta y_t)^2 - c^2 T^{-3/2} y_{t-1}^2\big)\right\}^2\right],
\end{aligned}$$

which is independent of σ_ε^2. The point optimal test statistic for $\mathrm{H}_0 : c^2 = 0$ vs $\mathrm{H}_1 : c^2 = \bar{c}^2$ based on $\mathcal{L}_T^*$ is given by

$$\begin{aligned}
Q_T^*(\bar{c}^2, \kappa_\varepsilon^2) &:= \log \mathcal{L}_T^*(\bar{c}^2, \kappa_\varepsilon^2) - \log \mathcal{L}_T^*(0, \kappa_\varepsilon^2) \\
&= \frac{\bar{c}^2}{\kappa_\varepsilon^2}\left[T^{-3/2} \sum_{t=1}^T y_{t-1}^2 \{(\Delta y_t)^2 - \hat{\sigma}_T^2\} - \frac{\bar{c}^2}{2} T^{-3} \sum_{t=1}^T \left(y_{t-1}^2 - T^{-1} \sum_{t=1}^T y_{t-1}^2\right)^2\right].
\end{aligned}$$

Theorem 13. *For model (9) with $\beta = 3/4$, under Assumption 2 with (ε_t, v_t) being Gaussian, we have*

$$\begin{aligned}
&Q_T^*(\bar{c}^2, \kappa_\varepsilon^2) \\
&\quad \Rightarrow \frac{\bar{c}^2}{\sqrt{2}}\left[\int_0^1 \left(W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\right) dW_\eta(r) + \frac{(c^2 - \bar{c}^2/2)}{\sqrt{2}} \int_0^1 \left[W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\right]^2 dr\right] \\
&\quad =: k^*(c^2, \bar{c}^2),
\end{aligned}$$

and

$$Q_T^*(\bar{c}^2, \hat{\kappa}_T^2) - Q_T^*(\bar{c}^2, \kappa_\varepsilon^2) \xrightarrow{p} 0.$$

Theorem 13 shows that the feasible test statistic $Q_T^*(\bar{c}^2, \hat{\kappa}_T^2)$ for a given $\bar{c}^2$ has the same limiting distribution $k^*(c^2, \bar{c}^2)$ as the point optimal test statistic $Q_T^*(\bar{c}^2, \kappa_\varepsilon^2)$, and thus we can obtain the limiting power envelope as a function of c^2 at the significance level α,

$$\pi(c^2) \coloneqq P\left(k^*(c^2, c^2) > cv(c^2)\right),$$

where $cv(c^2)$ satisfies $P(k^*(0, c^2) > cv(c^2)) = \alpha$. Although there is no verifying whether the quasi power envelope $\pi(c^2)$ is identical to the exact counterpart derived from l_T with σ_ε^2 unknown, $\pi(c^2)$ seems to serve as a good benchmark according to the result of the following simulation study.

Figure 1 displays the local asymptotic power functions of the tests considered in this article, along with the limiting power envelope $\pi(c^2)$. The local asymptotic distributions were simulated by 100,000 replications. The replications are based on $\varepsilon_t \sim$ i.i.d $N(0, 1)$ and hence $\sigma_\varepsilon^2 = 1$ and $\kappa_\varepsilon^2 = 2$. We give results for the case where the significance level is 0.05.

As for tests proposed by earlier work, the LN test performs reasonably well for each value of c^2 (but its power function stays a little below the power envelope), whereas the power of the MT is relatively low except at small c^2. The low power exhibited by the MT test is in line with the results of some previous work such as Nagakura (2009) and Su and Roca (2012). The H test is more powerful than the MT test but is less powerful than the LN test, for almost all c^2. While one can see a similar pattern in these three tests (i.e., the power gradually rises as c^2 gets large), the shape of the power function of the C test looks differently. For small c^2 up to about 10, the power rises only slightly. However, for c^2 over 10, it dramatically rises as the localizing coefficient increases, excelling the powers of the other tests for $c^2 \geq 20$ and being tangent to the power envelope for $c^2 \geq 25$.

From the above results, we find that the LN test is preferred when c^2 is small, and that the C test is preferred when c^2 is large. However, in general, the true value of the variance of the coefficient is unknown, and therefore we cannot decide which test should be used based on the (true) value of c^2. To deal with this uncertainty about the data generating process, following Harvey, Leybourne, and Taylor (2009) and Harvey, Leybourne, and Sollis (2015), we consider an alternative test with the rule "reject if at least one test rejects." To fix ideas, assume two test statistics are available for some hypothesis. According to the rule, one rejects the null when either test rejects the null. Harvey et al. (2009) applied this strategy to unit root testing and Harvey et al. (2015) to right-tailed DF-type testing for a bubble. Both the studies verified that the tests based on this strategy have higher power than when only either one test is employed, across different data generating processes considered in their studies.

If we adopt this procedure with the LN and C tests employed, high power across all c^2 is expected, because while the LN test performs well particularly when c^2 is small, the C test has the highest power for c^2 large. We refer to the test constructed according to this strategy as union-of-rejections test or UR test. The UR test with significance level α can be formulated by the rule

$$\text{Reject if} \quad \mathrm{LN}_T > \tau_\alpha cv_\alpha^{\mathrm{LN}} \quad \text{or} \quad \mathrm{C}_T > \tau_\alpha cv_\alpha^{\mathrm{C}},$$

where cv_α^{LN} and cv_α^{C} denote the critical values of the LN and C tests at significance level α, respectively. Here τ_α is a scaling factor introduced to keep type one error equal to the significance level α. The values of τ_α for $\alpha = 0.01, 0.05$, and 0.1 are found through simulation to be 1.153, 1.234 and 1.277, respectively (for the procedure to calculate τ_α for given α, see Harvey et al. (2015)).

The power function of the UR test is displayed in Figure 1. Although it is slightly lower than the power of the LN test for small c^2, it attains higher power for c^2 over about 15. Since each of these two tests do not dominate the other across c^2, we evaluate their performance in terms of the mean of power over c^2. This can be directly calculated as the area under power function. The mean powers of the UR and LN tests are 0.790 and 0.781, respectively. Given this small difference in mean power, the UR test can be used as an alternative to the LN test if one thinks it is important to reject the null when c^2 is relatively large.

6 Finite Sample Performance and Empirical Application

To evaluate the finite-sample performance of the tests considered so far, we conducted Monte Carlo experiments. The experiments are based on 50,000 replications. The significance level is set to 0.05. In these experiments, data were generated from

$$y_t = (1 + \omega v_t) y_{t-1} + \varepsilon_t,$$

where $(\varepsilon_t, v_t)' \sim$ i.i.d $N(0, \Omega)$ with $\Omega = \mathrm{diag}(1, 1)$ and $y_0 = 0$. The sample size is $T \in \{100, 200\}$. For the case of $T = 100$, the values of ω^2 were chosen between 0 and 0.05, so that they give c^2 between 0 and 50 for the local model with $T = 100$. For the $T = 200$ experiment, the values of ω^2 were selected in the similar way. The results are shown in Figures 2 and 3. The power function of each test is of similar shape in Figures 2 and 3, and thus we comment only on Figure 2. As for size, all the tests considered have size around the nominal 5% level (for LN, UR, Hausman, coefficient and MT tests, the sizes are 5.90%, 6.14%, 5.06%, 6.10% and 4.15%, respectively, when $T = 100$). Overall, we conclude that these tests have a reasonable size in finite samples. Power properties in finite samples are similar to those under local asymptotics, but in finite samples, the dominance of the C test for large c^2 almost disappears.

We also apply the tests considered in preceding sections to monthly real effective exchange rates of six countries. All the series begin in January 1994 and end in September 2021, giving 333 observations. Data were taken from Federal Reserve Economic Data. Results are given in Table 3.

For Canada, Germany and Japan, the MT, LN, H and UR statistics reject the null, whereas the C test does not reject the null for these countries. This would be caused by the low power of the C test for the small variance of the coefficient. By contrast, all the tests do not reject the null for the UK, which seems to be an evidence for the hypothesis that the real effective exchange rate of the UK is a unit root process. Although the LN test does not reject the null for US and Chinese real effective exchange rates, the MT and H tests reject the null. Thus we could reject the null of difference stationarity in favor of the STUR alternative for the US and China.

7 Conclusion

In this study, we have considered departures from a unit root other than conventional ones, i.e., stochastic departures characterized by cv_t/T^{β}. The stochastic departures can be classified into two types, local and moderate deviations, by checking the asymptotic behavior of tests of a unit root against the STUR alternative, such as the McCabe-Tremayne and Lee-Nagakura tests. We have developed the asymptotic theory for $\beta \in [1/2, 3/4]$ and it turns out that the $O_p(T^{-3/4})$ neighborhood of unity corresponds to local departures, and $O_p(T^{-\beta})$ neighborhoods with $1/2 \leq \beta < 3/4$ correspond to moderate deviations. For the former case, while the Dickey-Fuller test statistic converges to its null limiting distribution, the MT and LN test statistics have asymptotic distributions different from those under the null; for the latter case, these test statistics diverge at a rate depending on β, a parameter determining the distance from a unit root. The case $\beta \in (0, 1/2)$ would be also included in the moderate-deviations model, but in this case we need to develop the asymptotic theory different from that established in this article. This is our future research.

Relying on the asymptotic theory for the moderate-deviations case developed here, we have also proposed three new unit root tests, one being based on the localizing coefficient estimator, another the Hausman principle, and the other employing the union of rejections by the Lee-Nagakura and coefficient tests. By comparing the power functions of these tests with the Gaussian power envelope derived from an approximate model, it has been verified that the union-of-rejections test can be used as an alternative test for a unit root against a STUR.

References

Aue, A. (2008): "Near-Integrated Random Coefficient Autoregressive Time Series," *Econometric Theory*, 24, 1343–1372.

Banerjee, A., G. Chevillon, and M. Kratz (2020): "Probabilistic Forecasting of Bubbles and Flash Crashes," *The Econometrics Journal*, 23, 297–315.

Dickey, D. A. and W. A. Fuller (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.

Föllmer, H. and M. Schweizer (1993): "A Microeconomic Approach to Diffusion Models for Stock Prices," *Mathematical Finance*, 3, 1–23.

Granger, C. W. J. and N. R. Swanson (1997): "An Introduction to Stochastic Unit-Root Processes," *Journal of Econometrics*, 80, 35–62.

Hansen, B. E. (1992): "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, 8, 489–500.

Harvey, D. I., S. J. Leybourne, and R. Sollis (2015): "Recursive Right-Tailed Unit Root Tests for an Explosive Asset Price Bubble," *Journal of Financial Econometrics*, 13, 166–187.

Harvey, D. I., S. J. Leybourne, and A. M. R. Taylor (2009): "Unit Root Testing in Practice: Dealing with Uncertainty over the Trend and Initial Condition," *Econometric Theory*, 25, 587–636.

Horváth, L. and L. Trapani (2019): "Testing for Randomness in a Random Coefficient Autoregression Model," *Journal of Econometrics*, 209, 338–352.

Lee, S. (1998): "Coefficient Constancy Test in a Random Coefficient Autoregressive Model," *Journal of Statistical Planning and Inference*, 74, 93–101.

Lieberman, O. and P. C. B. Phillips (2014): "Norming Rates and Limit Theory for Some Time-Varying Coefficient Autoregressions," *Journal of Time Series Analysis*, 35, 592–623.

——— (2017): "A Multivariate Stochastic Unit Root Model with an Application to Derivative Pricing," *Journal of Econometrics*, 196, 99–110.

——— (2018): "IV and GMM Inference in Endogenous Stochastic Unit Root Models," *Econometric Theory*, 34, 1065–1100.

——— (2020): "Hybrid Stochastic Local Unit Roots," *Journal of Econometrics*, 215, 257–285.

McCabe, B. P. M. and R. J. Smith (1998): "The Power of Some Tests for Difference Stationarity under Local Heteroscedastic Integration," *Journal of the American Statistical Association*, 93, 751–761.

McCabe, B. P. M. and A. R. Tremayne (1995): "Testing a Time Series for Difference Stationarity," *The Annals of Statistics*, 23, 1015–1028.

Nagakura, D. (2009): "Testing for Coefficient Stability of AR(1) Model When the Null Is an Integrated or a Stationary Process," *Journal of Statistical Planning and Inference*, 139, 2731–2745.

Phillips, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

Phillips, P. C. B. and T. Magdalinos (2007): "Limit Theory for Moderate Deviations from a Unit Root," *Journal of econometrics*, 136, 115–130.

Sollis, R., P. Newbold, and S. J. Leybourne (2000): "Stochastic Unit Roots Modelling of Stock Price Indices," *Applied Financial Economics*, 10, 311–315.

Su, J.-J. and E. Roca (2012): "Examining the Power of Stochastic Unit Root Tests without Assuming Independence in the Error Processes of the Underlying Time Series," *Applied Economics Letters*, 19, 373–377.

Tanaka, K. (2017): *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*, vol. 4, John Wiley & Sons.

Tao, Y., P. C. B. Phillips, and J. Yu (2019): "Random Coefficient Continuous Systems: Testing for Extreme Sample Path Behavior," *Journal of Econometrics*, 209, 208–237.

Yau, R. and C. J. Hueng (2007): "Output Convergence Revisited: New Time Series Results on Industrialized Countries," *Applied Economics Letters*, 14, 75–77.

Yoon, G. (2005): "Has the US Economy Really Become Less Correlated with That of the Rest of the World?" *Economic Modelling*, 22, 147–158.

——— (2010a): "Nonlinear Mean Reversion in Real Exchange Rates: Threshold Autoregressive Models and Stochastic Unit Root Processes," *Applied Economics Letters*, 17, 797–804.

——— (2010b): "Nonlinear Mean-Reversion to Purchasing Power Parity: Exponential Smooth Transition Autoregressive Models and Stochastic Unit Root Processes," *Applied Economics*, 42, 489–496.

Appendix

Inspired by the proof of Theorem 1 of McCabe and Smith (1998), we give the following lemmas as preliminary results.

Lemma A.1. *For $\beta \in [1/2, 3/4]$, under Assumption 2, with probability approaching one (w.p.a 1), $(1 + cT^{-\beta} v_t)$, $t = 1, \dots, T$, are all positive.*

Proof. This can be shown by noting that

$$\begin{aligned} P\left(\bigcup_{t=1}^{T} \{1 + cT^{-\beta} v_t \leq 0\}\right) &\leq \sum_{t=1}^{T} P(1 + cT^{-\beta} v_t \leq 0) \\ &\leq \sum_{t=1}^{T} P(|cT^{-\beta} v_t| \geq 1) \\ &\leq \sum_{t=1}^{T} |c|^3 T^{-3\beta} \mathbb{E}[|v_t|^3] \\ &= |c|^3 T^{-3\beta+1} \mathbb{E}[|v_t|^3] \to 0 \ (T \to \infty), \end{aligned}$$

where the last inequality holds by Markov's inequality and the convergence by the fact $-3\beta+1<0$ for $\beta\in[1/2,3/4]$ and $\mathbb{E}[|v_t|^3]$ is finite. □

Lemma A.2. *When $1/2<\beta\le 3/4$, we have, w.p.a one,*

$$\prod_{k=1}^{\lfloor Tr\rfloor}\big(1+cT^{-\beta}v_k\big)=1+cT^{-\beta}\sum_{k=1}^{\lfloor Tr\rfloor}v_k+O_p(T^{1-2\beta}),$$

where the O_p notation is used in the uniform sense, that is,

$$\sup_{0\le r\le 1}\left|\prod_{k=1}^{\lfloor Tr\rfloor}\big(1+cT^{-\beta}v_k\big)-\big(1+cT^{-\beta}\sum_{k=1}^{\lfloor Tr\rfloor}v_k\big)\right|=O_p(T^{1-2\beta}).$$

In what follows, the O_p and o_p notations are used in the uniform sense unless otherwise stated.

Proof. By Lemma A.1, $\log\prod_{k=1}^{\lfloor Tr\rfloor}(1+cT^{-\beta}v_k),\ 0\le r\le 1$, exist w.p.a 1. Thus, we have, on an event whose probability approaches one as $T\to\infty$,

$$\begin{aligned}\prod_{k=1}^{\lfloor Tr\rfloor}(1+cT^{-\beta}v_k)&=\exp\bigg(\log\prod_{k=1}^{\lfloor Tr\rfloor}(1+cT^{-\beta}v_k)\bigg)\\&=\exp\bigg(\sum_{k=1}^{\lfloor Tr\rfloor}\log(1+cT^{-\beta}v_k)\bigg)\\&=\exp\bigg(cT^{-\beta}\sum_{k=1}^{\lfloor Tr\rfloor}v_k-\frac{c^2}{2}T^{-2\beta}\sum_{k=1}^{\lfloor Tr\rfloor}\frac{v_k^2}{(1+\zeta_k)^2}\bigg),\end{aligned}\tag{A.1}$$

where we used a Taylor expansion for the third equality and $|\zeta_k|<|cT^{-\beta}v_k|$. As for the remainder term, we have for any $\epsilon>0$,

$$\begin{aligned}P\big(\max_{1\le k\le T}|\zeta_k|\ge\epsilon\big)&\le P\big(\max_{1\le k\le T}|cT^{-\beta}v_k|\ge\epsilon\big)\\&\le\sum_{k=1}^{T}P(|cT^{-\beta}v_k|\ge\epsilon)\\&\le c^2\epsilon^{-2}T^{-2\beta}\sum_{k=1}^{T}\mathbb{E}[v_k^2]\to 0\ \text{ as } T\to\infty,\end{aligned}$$

since $\mathbb{E}[v_k^2] = 1$ and $1 - 2\beta < 0$. Hence, ζ_k is $o_p(1)$. It follows that

$$\sup_{0\le r\le 1}\left|T^{-2\beta}\sum_{k=1}^{\lfloor Tr\rfloor}\frac{v_k^2}{(1+\zeta_k)^2}\right| = T^{-2\beta}\sum_{k=1}^{T}\frac{v_k^2}{(1+\zeta_k)^2}$$
$$\le \frac{1}{\min_{1\le k\le T}(1+\zeta_k)^2}\cdot T^{1-2\beta}\cdot T^{-1}\sum_{k=1}^{T} v_k^2 = O_p(T^{1-2\beta}),$$

for which we used $\min_{1\le k\le T}(1+\zeta_k)^2 = 1 + \min_{1\le k\le T}(2\zeta_k + \zeta_k^2) \overset{p}{\to} 1$ and $T^{-1}\sum_{k=1}^{T} v_k^2 \overset{p}{\to} \mathbb{E}[v_k^2] = 1$ by the law of large numbers (LLN). Applying this result to (A.1) yields

$$\begin{aligned}\prod_{k=1}^{\lfloor Tr\rfloor}(1+cT^{-\beta}v_k) &= \exp\Big(cT^{-\beta}\sum_{k=1}^{\lfloor Tr\rfloor} v_k\Big)\exp\Big(-\frac{c^2}{2}T^{-2\beta}\sum_{k=1}^{\lfloor Tr\rfloor}\frac{v_k^2}{(1+\zeta_k)^2}\Big)\\ &= \Big(1 + cT^{-\beta}\sum_{k=1}^{\lfloor Tr\rfloor} v_k + O_p(T^{1-2\beta})\Big)\Big(1 + O_p(T^{1-2\beta})\Big)\\ &= 1 + cT^{-\beta}\sum_{k=1}^{\lfloor Tr\rfloor} v_k + O_p(T^{1-2\beta}),\end{aligned}$$

since

$$\sup_{0\le r\le 1}\left|T^{-\beta}\sum_{k=1}^{\lfloor Tr\rfloor} v_k\right| = T^{1/2-\beta}\sup_{0\le r\le 1}\left|T^{-1/2}\sum_{k=1}^{\lfloor Tr\rfloor} v_k\right| = O_p(T^{1/2-\beta})$$

by the FCLT. □

Lemma A.3. *The stochastic process* $\prod_{k=1}^{\lfloor Tr\rfloor}(1+cT^{-1/2}v_k)$ *weakly converges to the process* $\exp\big(cW_v(r) - c^2 r/2\big)$ *in* $D[0,1]$ *under Assumption 2.*

Proof. As in the proof of Lemma A.2,

$$\begin{aligned}\prod_{k=1}^{\lfloor Tr\rfloor}(1+cT^{-1/2}v_k) &= \exp\Big(\sum_{k=1}^{\lfloor Tr\rfloor}\log\big(1+cT^{-1/2}v_k\big)\Big)\\ &= \exp\Big(cT^{-1/2}\sum_{k=1}^{\lfloor Tr\rfloor} v_k - \frac{c^2}{2}T^{-1}\sum_{k=1}^{\lfloor Tr\rfloor} v_k^2 + \frac{c^3}{3}T^{-3/2}\sum_{k=1}^{\lfloor Tr\rfloor}\frac{v_k^3}{(1+\zeta_k)^3}\Big) \qquad \text{(A.2)}\\ &= \exp\Big(cT^{-1/2}\sum_{k=1}^{\lfloor Tr\rfloor} v_k - \frac{c^2}{2}T^{-1}\sum_{k=1}^{\lfloor Tr\rfloor} v_k^2\Big)\Big(1 + R_T(r)\Big),\end{aligned}$$

w.p.a. one, where $R_T(r)$ is the remainder term associated with a Taylor expansion for $\exp(\frac{c^3}{3}T^{-3/2}\sum_{k=1}^{\lfloor Tr\rfloor}\frac{v_k^3}{(1+\zeta_k)^3})$. Notice that for any $\epsilon > 0$,

$$\begin{aligned}
P\big(\max_{1\le k\le T}|\zeta_k|\ge\epsilon\big) &\le P\big(\max_{1\le k\le T}|cT^{-1/2}v_k|\ge\epsilon\big)\\
&\le \sum_{k=1}^{T}P(|cT^{-1/2}v_k|\ge\epsilon)\\
&\le |c|^3\epsilon^{-3}T^{-3/2}\sum_{k=1}^{T}\mathbb{E}[|v_k|^3]\to 0 \ \text{ as } T\to\infty.
\end{aligned}$$

Therefore the remainder term R_T satisfies

$$\sup_{0\le r\le 1}\big|R_T(r)\big| \le \frac{|c|^3}{3}\cdot\frac{1}{\min_{1\le k\le T}|1+\zeta_k|^3}\cdot T^{-1/2}\cdot T^{-1}\sum_{k=1}^{T}|v_k|^3 = O_p(T^{-1/2}),$$

from which we deduce, by the continuous mapping theorem (CMT),

$$\begin{aligned}
\prod_{k=1}^{\lfloor Tr\rfloor}(1+cT^{-1/2}v_k) &= \exp\Bigg(cT^{-1/2}\sum_{k=1}^{\lfloor Tr\rfloor}v_k - \frac{c^2}{2}T^{-1}\sum_{k=1}^{\lfloor Tr\rfloor}v_k^2\Bigg)\big(1+R_T(r)\big)\\
&\Rightarrow \exp\big(cW_v(r)-\frac{c^2}{2}r\big),
\end{aligned}$$

in $D[0,1]$. □

Proof of Lemma 1. Consider model (9). Let $\beta\in(1/2,3/4]$. To prove part (a), by backward substitution, y_t can be written as

$$y_t = \prod_{k=1}^{t}\big(1+cT^{-\beta}v_k\big)\sum_{s=1}^{t}\Bigg\{\prod_{k=1}^{s}\big(1+cT^{-\beta}v_k\big)\Bigg\}^{-1}\varepsilon_s + y_0\prod_{k=1}^{t}\big(1+cT^{-\beta}v_k\big).$$

By Lemma A.2, we obtain, w.p.a. one,

$$\begin{aligned}
T^{-1/2}y_{\lfloor Tr\rfloor} &= \Bigg(1+cT^{-\beta}\sum_{k=1}^{\lfloor Tr\rfloor}v_k + O_p(T^{1-2\beta})\Bigg)\sum_{s=1}^{\lfloor Tr\rfloor}\Bigg(1-cT^{-\beta}\sum_{k=1}^{s}v_k+O_p(T^{1-2\beta})\Bigg)\big(T^{-1/2}\varepsilon_s\big)\\
&\quad + T^{-1/2}y_0\Bigg(1+cT^{-\beta}\sum_{k=1}^{\lfloor Tr\rfloor}v_k+O_p(T^{1-2\beta})\Bigg)\\
&= T^{-1/2}\sum_{s=1}^{\lfloor Tr\rfloor}\varepsilon_s + A_{1,T}(r)+A_{2,T}(r)+A_{3,T}(r)+o_p(1), \qquad (A.3)
\end{aligned}$$

where

$$A_{1,T}(r) \coloneqq -cT^{1/2-\beta}\sum_{s=1}^{\lfloor Tr\rfloor}\Big(T^{-1/2}\sum_{k=1}^{s}v_k + O_p(T^{1/2-\beta})\Big)(T^{-1/2}\varepsilon_s),$$
$$A_{2,T}(r) \coloneqq cT^{1/2-\beta}\Big(T^{-1/2}\sum_{k=1}^{\lfloor Tr\rfloor}v_k + O_p(T^{1/2-\beta})\Big)(T^{-1/2}\sum_{s=1}^{\lfloor Tr\rfloor}\varepsilon_s),$$

and

$$A_{3,T}(r) \coloneqq -c^2T^{1-2\beta}\Big(T^{-1/2}\sum_{k=1}^{\lfloor Tr\rfloor}v_k + O_p(T^{1/2-\beta})\Big)\sum_{s=1}^{\lfloor Tr\rfloor}\Big(T^{-1/2}\sum_{k=1}^{s}v_k + O_p(T^{1/2-\beta})\Big)(T^{-1/2}\varepsilon_s).$$

We show $A_{i,T} = o_p(1),\ i = 1,2,3$, thereby obtaining, by applying the CMT to (A.3), $Y_T \Rightarrow \sigma_\varepsilon W_\varepsilon$, where $Y_T(r) = T^{-1/2}y_{\lfloor Tr\rfloor}$. Now, as for $A_{1,T}$, we have

$$\begin{aligned}
&\sum_{s=1}^{\lfloor Tr\rfloor}\Big(T^{-1/2}\sum_{k=1}^{s}v_k + O_p(T^{1/2-\beta})\Big)(T^{-1/2}\varepsilon_s)\\
&\quad= T^{-1}\sum_{s=1}^{\lfloor Tr\rfloor}v_s\varepsilon_s + \sum_{s=1}^{\lfloor Tr\rfloor}\Big(T^{-1/2}\sum_{k=1}^{s-1}v_k + O_p(T^{1/2-\beta})\Big)(T^{-1/2}\varepsilon_s)\\
&\quad\Rightarrow r\mathbb{E}[\varepsilon_t v_t] + \sigma_\varepsilon\int_0^r W_v(s)dW_\varepsilon(s),
\end{aligned}$$

by the LLN and Theorem 2.1 of Hansen (1992). This result, combined with $cT^{1/2-\beta} \to 0$, gives $A_{1,T} = o_p(1)$. $A_{2,T}$ also vanishes because $cT^{1/2-\beta} \to 0$ and

$$\Big(T^{-1/2}\sum_{k=1}^{\lfloor Tr\rfloor}v_k + O_p(T^{1/2-\beta})\Big)\Big(T^{-1/2}\sum_{s=1}^{\lfloor Tr\rfloor}\varepsilon_s\Big) \Rightarrow \sigma_\varepsilon W_v(r)W_\varepsilon(r).$$

That $A_{3,T} = o_p(1)$ can be verified by a similar argument.

Given that $Y_T \Rightarrow \sigma_\varepsilon W_\varepsilon$, it is straightforward to show the consistency of $\hat\sigma_T^2$ and $\hat\kappa_T^2$. For part (b), we have

$$\begin{aligned}
\hat\sigma_T^2 &= T^{-1}\sum_{t=1}^{T}(\Delta y_t)^2\\
&= c^2T^{-1-2\beta}\sum_{t=1}^{T}y_{t-1}^2v_t^2 + 2cT^{-1-\beta}\sum_{t=1}^{T}y_{t-1}v_t\varepsilon_t + T^{-1}\sum_{t=1}^{T}\varepsilon_t^2. \qquad \text{(A.4)}
\end{aligned}$$

The first term is

$$T^{-1-2\beta}\sum_{t=1}^{T} y_{t-1}^2 v_t^2 \le \sup_{0\le r\le 1}\left|T^{-1/2}y_{\lfloor Tr\rfloor}\right|^2 \cdot T^{1-2\beta}\cdot T^{-1}\sum_{t=1}^{T} v_t^2 = O_p(T^{1-2\beta}).$$

The second term is

$$\left|T^{-1-\beta}\sum_{t=1}^{T} y_{t-1}v_t\varepsilon_t\right| \le \sup_{0\le r\le 1}\left|T^{-1/2}y_{\lfloor Tr\rfloor}\right| T^{1/2-\beta}\cdot T^{-1}\sum_{t=1}^{T}|\varepsilon_t v_t| = O_p(T^{1/2-\beta}).$$

Substituting these results into (A.4), we have

$$\hat{\sigma}_T^2 = T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 + o_p(1) \xrightarrow{p} \sigma_\varepsilon^2,$$

since $1/2-\beta<0$ when $\beta\in(1/2,3/4]$.

For part (c), we write $\hat{\kappa}_T^2$ as

$$\hat{\kappa}_T^2 = T^{-1}\sum_{t=1}^{T}\left\{\left(\Delta y_t\right)^2 - \hat{\sigma}_T^2\right\}^2 = T^{-1}\sum_{t=1}^{T}\left(\Delta y_t\right)^4 - \hat{\sigma}_T^4. \tag{A.5}$$

The first term is

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}\left(\Delta y_t\right)^4 = {} & c^4T^{-1-4\beta}\sum_{t=1}^{T} y_{t-1}^4 v_t^4 + 4c^3T^{-1-3\beta}\sum_{t=1}^{T} y_{t-1}^3\varepsilon_t v_t^3 + 6c^2T^{-1-2\beta}\sum_{t=1}^{T} y_{t-1}^2\varepsilon_t^2 v_t^2 \\ & + 4cT^{-1-\beta}\sum_{t=1}^{T} y_{t-1}\varepsilon_t^3 v_t + T^{-1}\sum_{t=1}^{T}\varepsilon_t^4, \end{aligned}$$

for which we have

$$\begin{aligned} \left|T^{-1-4\beta}\sum_{t=1}^{T} y_{t-1}^4 v_t^4\right| &\le \sup_{0\le r\le 1}\left|T^{-1/2}y_{\lfloor Tr\rfloor}\right|^4 T^{2-4\beta}\cdot T^{-1}\sum_{t=1}^{T} v_t^4 = O_p(T^{2-4\beta}), \\ \left|T^{-1-3\beta}\sum_{t=1}^{T} y_{t-1}^3\varepsilon_t v_t^3\right| &\le \sup_{0\le r\le 1}\left|T^{-1/2}y_{\lfloor Tr\rfloor}\right|^3 T^{3/2-3\beta}\cdot T^{-1}\sum_{t=1}^{T}|\varepsilon_t v_t^3| = O_p(T^{3/2-3\beta}), \\ \left|T^{-1-2\beta}\sum_{t=1}^{T} y_{t-1}^2\varepsilon_t^2 v_t^2\right| &\le \sup_{0\le r\le 1}\left|T^{-1/2}y_{\lfloor Tr\rfloor}\right|^2 T^{1-2\beta}\cdot T^{-1}\sum_{t=1}^{T}|\varepsilon_t^2 v_t^2| = O_p(T^{1-2\beta}), \end{aligned}$$

and

$$\left|T^{-1-\beta}\sum_{t=1}^{T} y_{t-1}\varepsilon_t^3 v_t\right| \le \sup_{0\le r\le 1}\left|T^{-1/2}y_{\lfloor Tr\rfloor}\right| T^{1/2-\beta}\cdot T^{-1}\sum_{t=1}^{T}|\varepsilon_t^3 v_t| = O_p(T^{1/2-\beta}).$$

Thus, the first term of (A.5) is

$$T^{-1}\sum_{t=1}^{T}\big(\Delta y_t\big)^4 = T^{-1}\sum_{t=1}^{T}\varepsilon_t^4 + o_p(1) \xrightarrow{p} \mathbb{E}[\varepsilon_t^4].$$

Hence

$$\hat{\kappa}_T^2 \xrightarrow{p} \mathbb{E}[\varepsilon_t^4] - \sigma_\varepsilon^4 = \kappa_\varepsilon^2.$$

□

Proof of Lemma 2. We prove this lemma under a more general condition, replacing $\sigma_{\varepsilon v} = \mathbb{E}[\varepsilon_t v_t] = 0$ with $\sigma_{\varepsilon v} \neq 0$.[8] According to Theorem 3.1 of Föllmer and Schweizer (1993), part (a) could be proved if what they call goodness condition is seen to hold under Assumption 2. However, instead of relying on their result, we here take an approach similar to the proof of Lemma 1 of Lieberman and Phillips (2020), which relies on the standard continuous mapping argument. As in the proof of part (a) of Lemma 1, we write y_t as

$$y_t = \prod_{k=1}^{t}\Big(1 + cT^{-1/2}v_k\Big)\sum_{s=1}^{t}\bigg\{\prod_{k=1}^{s}\Big(1 + cT^{-1/2}v_k\Big)\bigg\}^{-1}\varepsilon_s + y_0\prod_{k=1}^{t}\Big(1 + cT^{-1/2}v_k\Big).$$

Setting $t = \lfloor Tr \rfloor$ and using (A.2) give

$$\begin{aligned} T^{-1/2}y_{\lfloor Tr \rfloor} &= \prod_{k=1}^{\lfloor Tr \rfloor}\Big(1 + cT^{-1/2}v_k\Big)\sum_{s=1}^{\lfloor Tr \rfloor}\exp\bigg(-cT^{-1/2}\sum_{k=1}^{s}v_k + \frac{c^2}{2}T^{-1}\sum_{k=1}^{s}v_k^2 \\ &\qquad - \frac{c^3}{3}T^{-3/2}\sum_{k=1}^{s}\frac{v_k^3}{(1+\zeta_k)^3}\bigg)\Big(T^{-1/2}\varepsilon_s\Big) + T^{-1/2}y_0\prod_{k=1}^{\lfloor Tr \rfloor}\Big(1 + cT^{-1/2}v_k\Big) \\ &=: D_{1,T}(r) \times D_{2,T}(r) + o_p(1), \end{aligned}$$

in view of Lemma A.3 and the fact that $y_0 = o_p(T^{1/2})$. From Lemma A.3

$$D_{1,T}(r) = \prod_{k=1}^{\lfloor Tr \rfloor}\Big(1 + cT^{-1/2}v_k\Big) \Rightarrow \exp\big(cW_v(r) - \frac{c^2}{2}r\big) \text{ in } D[0,1]. \tag{A.6}$$

[8] Note that when proving the other results, we maintain $\sigma_{\varepsilon v} = 0$, i.e., Assumption 2. We consider the case $\sigma_{\varepsilon v} \neq 0$ only for Lemma 2.

To deal with $D_{2,T}$, we note

$$\begin{aligned}
D_{2,T}(r) &= \sum_{s=1}^{\lfloor Tr \rfloor} \exp\Big(-\frac{c}{\sqrt{T}} \sum_{k=1}^{s} v_k + \frac{c^2}{2T} \sum_{k=1}^{s} v_k^2 + O_p(T^{-1/2})\Big)\big(T^{-1/2}\varepsilon_s\big) \\
&= \sum_{s=1}^{\lfloor Tr \rfloor} \exp\Big(-\frac{c}{\sqrt{T}} \sum_{k=1}^{s} v_k + \frac{c^2}{2T} \sum_{k=1}^{s} v_k^2\Big)\Big(1 + O_p(T^{-1/2})\Big)\big(T^{-1/2}\varepsilon_s\big) \\
&= \sum_{s=1}^{\lfloor Tr \rfloor} \exp\Big(-\frac{c}{\sqrt{T}} \sum_{k=1}^{s} v_k + \frac{c^2}{2T} \sum_{k=1}^{s} v_k^2\Big)\big(T^{-1/2}\varepsilon_s\big) + o_p(1).
\end{aligned}$$

The dominant term is

$$\begin{aligned}
&\sum_{s=1}^{\lfloor Tr \rfloor} \exp\Big(-\frac{c}{\sqrt{T}} \sum_{k=1}^{s} v_k + \frac{c^2}{2T} \sum_{k=1}^{s} v_k^2\Big)\big(T^{-1/2}\varepsilon_s\big) \\
&= \sum_{s=1}^{\lfloor Tr \rfloor} \exp\Big(-\frac{c}{\sqrt{T}} \sum_{k=1}^{s-1} v_k + \frac{c^2}{2T} \sum_{k=1}^{s-1} v_k^2\Big)\Big(1 - \frac{cv_s}{\sqrt{T}} + \frac{c^2 v_s^2}{2T} + O_p(T^{-1})\Big)\big(T^{-1/2}\varepsilon_s\big) \\
&= \sigma_\varepsilon \int_0^r \exp\big(-cW_{T,v}(s) + \frac{c^2}{2}s\big)dW_{T,\varepsilon}(s) - c\sigma_{\varepsilon v} \int_0^r \exp\big(-cW_{T,v}(s) + \frac{c^2}{2}s\big)ds \\
&- \frac{c}{\sqrt{T}}\mathbb{E}[(\varepsilon_s v_s - \sigma_{\varepsilon v})^2]^{1/2} \int_0^r \exp\big(-cW_{T,v}(s) + \frac{c^2}{2}s\big)dW_{T,\varepsilon v}(s) + o_p(1),
\end{aligned}$$

where

$$W_{T,\varepsilon v}(r) := \mathbb{E}[(\varepsilon_s v_s - \sigma_{\varepsilon v})^2]^{-1/2} T^{-1/2} \sum_{s=1}^{\lfloor Tr \rfloor} (\varepsilon_s v_s - \sigma_{\varepsilon v}).$$

Since $\{\varepsilon_s v_s - \sigma_{\varepsilon v}\}$ is i.i.d and has zero mean and finite variance under Assumption 2, $W_{T,\varepsilon v}$ weakly converges by the FCLT. Thus

$$D_{2,T}(r) \Rightarrow \sigma_\varepsilon \int_0^r \exp\big(-cW_v(s) + \frac{c^2}{2}s\big)dW_\varepsilon(s) - c\sigma_{\varepsilon v} \int_0^r \exp\big(-cW_v(s) + \frac{c^2}{2}s\big)ds \tag{A.7}$$

in $D[0,1]$. By (A.6), (A.7) and the CMT, we deduce

$$T^{-1/2} y_{\lfloor Tr \rfloor} \Rightarrow \exp\big(cW_v(r) - \frac{c^2}{2}r\big)\Big\{\sigma_\varepsilon \int_0^r \exp\big(-cW_v(s) + \frac{c^2}{2}s\big)dW_\varepsilon(s) - c\sigma_{\varepsilon v} \int_0^r \exp\big(-cW_v(s) + \frac{c^2}{2}s\big)ds\Big\},$$

for which letting $\sigma_{\varepsilon v} = 0$ produces $\sigma_\varepsilon Y_c$ defined by (10).

For part (b), we have

$$\begin{aligned}
\hat{\sigma}_T^2 &= c^2 T^{-2} \sum_{t=1}^{T} y_{t-1}^2 v_t^2 + 2cT^{-3/2} \sum_{t=1}^{T} y_{t-1} v_t \varepsilon_t + T^{-1} \sum_{t=1}^{T} \varepsilon_t^2 \\
&= c^2 \int_0^1 Y_T^2(r)dr + c^2 T^{-1/2} \int_0^1 Y_T^2(r) dW_{T,v^2}(r) + 2c\sigma_{\varepsilon v} \int_0^1 Y_T(r)dr \\
&+ 2c\sigma_\varepsilon T^{-1/2} \int_0^1 Y_T(r) dW_{T,\varepsilon v}(r) + \sigma_\varepsilon^2 + o_p(1) \\
&\Rightarrow c^2 \sigma_\varepsilon^2 \int_0^1 Y_c^2(r)dr + 2c\sigma_\varepsilon \sigma_{\varepsilon v} \int_0^1 Y_c(r)dr + \sigma_\varepsilon^2. \qquad \text{(A.8)}
\end{aligned}$$

When $\sigma_{\varepsilon v} = 0$, the limit reduces to $c^2\sigma_\varepsilon^2 \int_0^1 Y_c^2(r)dr + \sigma_\varepsilon^2$.

For part (c), we write $\hat{\kappa}_T^2$ as

$$\begin{aligned}
\hat{\kappa}_T^2 &= T^{-1} \sum_{t=1}^{T} (\Delta y_t)^4 - \hat{\sigma}_T^4 \\
&= c^4 T^{-3} \sum_{t=1}^{T} y_{t-1}^4 v_t^4 + 4c^3 T^{-5/2} \sum_{t=1}^{T} y_{t-1}^3 v_t^3 \varepsilon_t + 6c^2 T^{-2} \sum_{t=1}^{T} y_{t-1}^2 v_t^2 \varepsilon_t^2 \\
&+ 4cT^{-3/2} \sum_{t=1}^{T} y_{t-1} v_t \varepsilon^3 + T^{-1} \sum_{t=1}^{T} \varepsilon_t^4 - \hat{\sigma}_T^4 \\
&= c^4 \mathbb{E}[v_t^4] \int_0^1 Y_T^4(r)dr + 4c^3 \mathbb{E}[v_t^3 \varepsilon_t] \int_0^1 Y_T^3(r)dr + 6c^2 \mathbb{E}[v_t^2 \varepsilon_t^2] \int_0^1 Y_T^2(r)dr \\
&+ 4c\mathbb{E}[v_t \varepsilon_t^3] \int_0^1 Y_T(r)dr + \sigma_\varepsilon^4 - \hat{\sigma}_T^4 + o_p(1),
\end{aligned}$$

since $\mathbb{E}[|v_t|^8] < \infty$. Then, we obtain

$$\begin{aligned}
\hat{\kappa}_T^2 \Rightarrow\ & c^4 \sigma_\varepsilon^4 \mathbb{E}[v_t^4] \int_0^1 Y_c^4(r)dr + 4c^3 \sigma_\varepsilon^3 \mathbb{E}[v_t^3 \varepsilon_t] \int_0^1 Y_c^3(r)dr + 6c^2 \sigma_\varepsilon^2 \mathbb{E}[v_t^2 \varepsilon_t^2] \int_0^1 Y_c^2(r)dr \\
&+ 4c\sigma_\varepsilon \mathbb{E}[v_t \varepsilon_t^3] \int_0^1 Y_c(r)dr + \sigma_\varepsilon^4 - \left[c^2 \sigma_\varepsilon^2 \int_0^1 Y_c^2(r)dr + 2c\sigma_\varepsilon \sigma_{\varepsilon v} \int_0^1 Y_c(r)dr + \sigma_\varepsilon^2 \right]^2. \qquad \text{(A.9)}
\end{aligned}$$

□

Proof of Theorem 5. Consider model (9) and let $\beta \in (1/2, 3/4]$. For part (a), we write

MT_T as

$$\begin{aligned}\mathrm{MT}_T &= \hat{\kappa}_T^{-1}\hat{\sigma}_T^{-2}T^{-3/2}\sum_{t=1}^{T} y_{t-1}^2\Big\{\big(\Delta y_t\big)^2 - \hat{\sigma}_T^2\Big\} \\ &= \hat{\kappa}_T^{-1}\hat{\sigma}_T^{-2}T^{-3/2}\sum_{t=1}^{T}\Big(y_{t-1}^2 - T^{-1}\sum_{t=1}^{T} y_{t-1}^2\Big)\Big\{\big(\Delta y_t\big)^2 - \hat{\sigma}_T^2\Big\} \\ &= \hat{\kappa}_T^{-1}\hat{\sigma}_T^{-2}T^{-3/2}\sum_{t=1}^{T}\Big(y_{t-1}^2 - T^{-1}\sum_{t=1}^{T} y_{t-1}^2\Big)\Big\{\big(\Delta y_t\big)^2 - \sigma_\varepsilon^2\Big\}.\end{aligned}$$

In deriving the above expression, we used the property of deviations from mean, noting $\hat{\sigma}_T^2 = T^{-1}\sum_{t=1}^{T}(\Delta y_t)^2$, as in McCabe and Tremayne (1995). Since

$$\big(\Delta y_t\big)^2 = c^2T^{-2\beta}y_{t-1}^2 v_t^2 + 2cT^{-\beta}y_{t-1}\varepsilon_t v_t + \varepsilon_t^2,$$

we have

$$\begin{aligned}\mathrm{MT}_T = {}& \hat{\kappa}_T^{-1}\hat{\sigma}_T^{-2}T^{-3/2}\sum_{t=1}^{T}\Big(y_{t-1}^2 - T^{-1}\sum_{t=1}^{T} y_{t-1}^2\Big)(\varepsilon_t^2 - \sigma_\varepsilon^2) \\ &+ \hat{\kappa}_T^{-1}\hat{\sigma}_T^{-2}T^{-3/2}\sum_{t=1}^{T}\Big(y_{t-1}^2 - T^{-1}\sum_{t=1}^{T} y_{t-1}^2\Big)(c^2T^{-2\beta}y_{t-1}^2 v_t^2 + 2cT^{-\beta}y_{t-1}\varepsilon_t v_t) \\ = {}& \hat{\kappa}_T^{-1}\hat{\sigma}_T^{-2}T^{-3/2}\sum_{t=1}^{T}\Big(y_{t-1}^2 - T^{-1}\sum_{t=1}^{T} y_{t-1}^2\Big)(\varepsilon_t^2 - \sigma_\varepsilon^2) \\ &+ \hat{\kappa}_T^{-1}\hat{\sigma}_T^{-2}\Big(c^2T^{-3/2-2\beta}\sum_{t=1}^{T} y_{t-1}^4 v_t^2 + 2cT^{-3/2-\beta}\sum_{t=1}^{T} y_{t-1}^3\varepsilon_t v_t\Big) \\ &- \hat{\kappa}_T^{-1}\hat{\sigma}_T^{-2}T^{-2}\sum_{t=1}^{T} y_{t-1}^2\Big(c^2T^{-1/2-2\beta}\sum_{t=1}^{T} y_{t-1}^2 v_t^2 + 2cT^{-1/2-\beta}\sum_{t=1}^{T} y_{t-1}\varepsilon_t v_t\Big) \\ &=: B_{1,T} + B_{2,T} - B_{3,T}.\end{aligned}$$

One can easily verify that by the CMT

$$\begin{aligned}B_{1,T} &\Rightarrow \kappa_\varepsilon^{-1}\sigma_\varepsilon^{-2}\Big\{\sigma_\varepsilon^2\kappa_\varepsilon\int_0^1 W_\varepsilon^2(r)dW_\eta(r) - \sigma_\varepsilon^2\kappa_\varepsilon\int_0^1 W_\varepsilon^2(r)dr W_\eta(1)\Big\} \\ &= \int_0^1\Big\{W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\Big\}dW_\eta(r), \qquad \text{(A.10)}\end{aligned}$$

in view of Lemma 1. Note that this result holds for both the cases $\beta = 3/4$ and $1/2 < \beta < 3/4$.

The limiting behavior of $B_{2,T}$ and $B_{3,T}$ depends on the value of β. First consider the case $\beta = 3/4$. In this case, we have

$$\begin{aligned} B_{2,T} &= \hat{\kappa}_T^{-1}\hat{\sigma}_T^{-2}\bigg(c^2T^{-3}\sum_{t=1}^{T} y_{t-1}^4 + c^2T^{-3}\sum_{t=1}^{T} y_{t-1}^4(v_t^2-1) + 2cT^{-9/4}\sum_{t=1}^{T} y_{t-1}^3\varepsilon_t v_t\bigg) \\ &= \hat{\kappa}_T^{-1}\hat{\sigma}_T^{-2}\bigg(c^2T^{-1}\sum_{t=1}^{T}(T^{-1/2}y_{t-1})^4 + c^2T^{-1/2}\sum_{t=1}^{T}(T^{-1/2}y_{t-1})^4T^{-1/2}(v_t^2-1) \\ &\quad + 2cT^{-1/4}\sum_{t=1}^{T}(T^{-1/2}y_{t-1})^3T^{-1/2}\varepsilon_t v_t\bigg). \end{aligned}$$

The first term in the parentheses weakly converges to $c^2\sigma_\varepsilon^4\int_0^1 W_\varepsilon^4(r)dr$. To analyze the behavior of the second and third terms, it should be noticed that the processes $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}(v_t^2-1)$ and $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\varepsilon_t v_t$ on $[0,1]$ weakly converge under Assumption 2. It follows from Theorem 2.1 of Hansen (1992) that

$$\sum_{t=1}^{T}\big(T^{-1/2}y_{t-1}\big)^4T^{-1/2}(v_t^2-1) \Rightarrow \sigma_\varepsilon^4\mathbb{E}[(v_t^2-1)^2]^{1/2}\int_0^1 W_\varepsilon^4(r)dW_{v^2}(r)$$

and

$$\sum_{t=1}^{T}\big(T^{-1/2}y_{t-1}\big)^3T^{-1/2}\varepsilon_t v_t \Rightarrow \sigma_\varepsilon^3\mathbb{E}[\varepsilon_t^2v_t^2]^{1/2}\int_0^1 W_\varepsilon^3(r)dW_{\varepsilon v}(r),$$

say. This leads to obtaining

$$B_{2,T} = \hat{\kappa}_T^{-1}\hat{\sigma}_T^{-2}c^2T^{-1}\sum_{t=1}^{T}\big(T^{-1/2}y_{t-1}\big)^4 + O_p(T^{-1/4}) \Rightarrow \frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon}\int_0^1 W_\varepsilon^4(r)dr. \qquad \text{(A.11)}$$

As for $B_{3,T}$, note that

$$\begin{aligned} &T^{-1}\sum_{t=1}^{T}\big(T^{-1/2}y_{t-1}\big)^2 \Rightarrow \sigma_\varepsilon^2\int_0^1 W_\varepsilon^2(r)dr, \\ &T^{-2}\sum_{t=1}^{T} y_{t-1}^2v_t^2 = T^{-2}\sum_{t=1}^{T} y_{t-1}^2 + T^{-2}\sum_{t=1}^{T} y_{t-1}^2(v_t^2-1) \Rightarrow \sigma_\varepsilon^2\int_0^1 W_\varepsilon^2(r)dr, \end{aligned}$$

and

$$T^{-5/4}\sum_{t=1}^{T} y_{t-1}\varepsilon_t v_t = T^{-1/4}\sum_{t=1}^{T}\big(T^{-1/2}y_{t-1}\big)T^{-1/2}\varepsilon_t v_t = O_p(T^{-1/4}).$$

Substituting these results into $B_{3,T}$ gives

$$B_{3,T} \Rightarrow \frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon}\left(\int_0^1 W_\varepsilon^2(r)dr\right)^2. \tag{A.12}$$

Combining (A.10) through (A.12), we have

$$\mathrm{MT}_T \Rightarrow \int_0^1\left(W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\right)dW_\eta(r) + \frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon}\int_0^1\left[W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\right]^2 dr.$$

This proves part (a) under $\beta = 3/4$. For part (b), since

$$\mathrm{LN}_T = \mathrm{MT}_T \times \frac{\hat{\sigma}_T^2}{\{T^{-3}\sum_{t=1}^T y_{t-1}^4 - (T^{-2}\sum_{t=1}^T y_{t-1}^2)^2\}^{1/2}}, \tag{A.13}$$

the convergence of MT_T, Lemma 1 and the CMT yield, when $\beta = 3/4$,

$$\mathrm{LN}_T \Rightarrow \frac{\int_0^1\left[W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\right]dW_\eta(r)}{\left[\int_0^1\{W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\}^2 dr\right]^{1/2}} + \frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon}\left[\int_0^1 \left\{W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\right\}^2 dr\right]^{1/2}.$$

Next, we prove parts (a) and (b) under $1/2 < \beta < 3/4$. Multiplying $B_{2,T}$ and $B_{3,T}$ by $T^{2\beta-3/2}$ leads to

$$\begin{aligned} T^{2\beta-3/2}B_{2,T} &= \hat{\kappa}_T^{-1}\hat{\sigma}_T^{-2}\left(c^2T^{-3}\sum_{t=1}^T y_{t-1}^4 v_t^2 + 2cT^{-3+\beta}\sum_{t=1}^T y_{t-1}^3\varepsilon_t v_t\right) \\ &\Rightarrow \frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon}\int_0^1 W_\varepsilon^4(r)dr, \end{aligned}$$

and

$$\begin{aligned} T^{2\beta-3/2}B_{3,T} &= \hat{\kappa}_T^{-1}\hat{\sigma}_T^{-2}T^{-2}\sum_{t=1}^T y_{t-1}^2\left(c^2T^{-2}\sum_{t=1}^T y_{t-1}^2 v_t^2 + 2cT^{-2+\beta}\sum_{t=1}^T y_{t-1}\varepsilon_t v_t\right) \\ &\Rightarrow \frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon}\left(\int_0^1 W_\varepsilon^2(r)dr\right)^2, \end{aligned}$$

because $-3+\beta < -2$ and $-2+\beta < -1$. Therefore, we deduce

$$\begin{aligned} \mathrm{MT}_T &= B_{1,T} + T^{3/2-2\beta} \times (T^{2\beta-3/2}B_{2,T} + T^{2\beta-3/2}B_{3,T}) \\ &\overset{a}{\sim} \int_0^1\left(W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\right)dW_\eta(r) + T^{3/2-2\beta}\frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon}\int_0^1\left[W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\right]^2 dr. \end{aligned}$$

Using the relation (A.13) and the CMT, one can derive the asymptotic expression of LN_T under $1/2 < \beta < 3/4$.

To prove part (c), note that the DF test statistic is

$$\mathrm{DF}_T = \frac{T^{-1}\sum_{t=1}^T y_{t-1}\Delta y_t}{T^{-2}\sum_{t=1}^T y_{t-1}^2} = \frac{cT^{-1-\beta}\sum_{t=1}^T y_{t-1}^2 v_t}{T^{-2}\sum_{t=1}^T y_{t-1}^2} + \frac{T^{-1}\sum_{t=1}^T y_{t-1}\varepsilon_t}{T^{-2}\sum_{t=1}^T y_{t-1}^2}.$$

The second term converges to the so-called Dickey-Fuller distribution when $1/2 < \beta \leq 3/4$, in view of part (a) of Lemma 1. As for the first term, the numerator is

$$cT^{-1-\beta}\sum_{t=1}^T y_{t-1}^2 v_t = cT^{1/2-\beta}\sum_{t=1}^T \big(T^{-1/2}y_{t-1}\big)^2 T^{-1/2}v_t = o_p(1),$$

since $1/2 - \beta < 0$ when $1/2 < \beta \leq 3/4$. Therefore the first term is negligible, and it follows that

$$\mathrm{DF}_T \Rightarrow \frac{\int_0^1 W_\varepsilon(r)dW_\varepsilon(r)}{\int_0^1 W_\varepsilon^2(r)dr},$$

which completes the proof. □

Proof of Theorem 6. We imitate the proof of Theorem 5. To prove part (a), let us express MT_T as

$$\mathrm{MT}_T = B_{1,T}^* + B_{2,T}^* - B_{3,T}^*,$$

where

$$B_{1,T}^* := \hat\kappa_T^{-1}\hat\sigma_T^{-2}T^{-3/2}\sum_{t=1}^T\left(y_{t-1}^2 - T^{-1}\sum_{t=1}^T y_{t-1}^2\right)\big(\varepsilon_t^2 - \sigma_\varepsilon^2\big),$$

$$B_{2,T}^* := \hat\kappa_T^{-1}\hat\sigma_T^{-2}\left(c^2T^{-5/2}\sum_{t=1}^T y_{t-1}^4 v_t^2 + 2cT^{-2}\sum_{t=1}^T y_{t-1}^3\varepsilon_t v_t\right),$$

and

$$B_{3,T}^* := \hat\kappa_T^{-1}\hat\sigma_T^{-2}T^{-2}\sum_{t=1}^T y_{t-1}^2\left(c^2T^{-3/2}\sum_{t=1}^T y_{t-1}^2 v_t^2 + 2cT^{-1}\sum_{t=1}^T y_{t-1}\varepsilon_t v_t\right).$$

By the same argument as in the proof of Theorem 5, we have

$$B_{1,T}^* \Rightarrow \frac{\kappa_\varepsilon\sigma_\varepsilon^2}{\kappa\sigma^2}\int_0^1\left(Y_c^2(r) - \int_0^1 Y_c^2(s)ds\right)dW_\eta(r),$$

$$T^{-1/2}B_{2,T}^* \Rightarrow \frac{c^2\sigma_\varepsilon^4}{\kappa\sigma^2}\int_0^1 Y_c^4(r)dr,$$

and

$$T^{-1/2}B_{3,T}^{*} \Rightarrow \frac{c^2\sigma_\varepsilon^4}{\kappa\sigma^2}\left(\int_0^1 Y_c^2(r)dr\right)^2,$$

where σ^2 and κ^2 are random variables that are distributed according to (A.8) and (A.9) with $\sigma_{\varepsilon v}=0$, respectively. Therefore, we get

$$\mathrm{MT}_T \overset{a}{\sim} \frac{\kappa_\varepsilon\sigma_\varepsilon^2}{\kappa\sigma^2}\int_0^1\left(Y_c^2(r)-\int_0^1 Y_c^2(s)ds\right)dW_\eta(r)+T^{1/2}\frac{c^2\sigma_\varepsilon^4}{\kappa\sigma^2}\int_0^1\left[Y_c^2(r)-\int_0^1 Y_c^2(s)ds\right]^2dr.$$

The result for LN_T follows immediately in view of (A.13), which proves part (b). □

Proof of Lemma 7. Divide both the numerator and denominator of (20) by T^4. Then, the first term of the numerator is

$$T^{-1}\sum_{t=1}^{T}(\Delta y_t)^2\cdot T^{-3}\sum_{t=1}^{T}y_{t-1}^4 \Rightarrow \sigma_\varepsilon^6\left(c^2\int_0^1 Y_c^2(r)dr+1\right)\int_0^1 Y_c^4(r)dr, \tag{A.14}$$

by Lemma 2 and the CMT. The second term of the numerator is

$$\begin{aligned}
&T^{-2}\sum_{t=1}^{T}(\Delta y_t)^2y_{t-1}^2\cdot T^{-2}\sum_{t=1}^{T}y_{t-1}^2\\
&=\left(c^2T^{-3}\sum_{t=1}^{T}y_{t-1}^4v_t^2+2cT^{-5/2}\sum_{t=1}^{T}y_{t-1}^3\varepsilon_tv_t+T^{-2}\sum_{t=1}^{T}y_{t-1}^2\varepsilon_t^2\right)T^{-2}\sum_{t=1}^{T}y_{t-1}^2\\
&\Rightarrow \sigma_\varepsilon^6\left(c^2\int_0^1 Y_c^4(r)dr+\int_0^1 Y_c^2(r)dr\right)\int_0^1 Y_c^2(r)dr.
\end{aligned} \tag{A.15}$$

The denominator becomes

$$T^{-3}\sum_{t=1}^{T}y_{t-1}^4-\left(T^{-2}\sum_{t=1}^{T}y_{t-1}^2\right)^2 \Rightarrow \sigma_\varepsilon^4\left\{\int_0^1 Y_c^4(r)dr-\left(\int_0^1 Y_c^2(r)dr\right)^2\right\}. \tag{A.16}$$

Using (A.14) through (A.16), we obtain

$$\begin{aligned}
\tilde{\sigma}_T^2 \Rightarrow &\left\{\sigma_\varepsilon^6\left(c^2\int_0^1 Y_c^2(r)dr+1\right)\int_0^1 Y_c^4(r)dr-\sigma_\varepsilon^6\left(c^2\int_0^1 Y_c^4(r)dr+\int_0^1 Y_c^2(r)dr\right)\int_0^1 Y_c^2(r)dr\right\}\\
&\div\left\{\sigma_\varepsilon^4\int_0^1 Y_c^4(r)dr-\sigma_\varepsilon^4\left(\int_0^1 Y_c^2(r)dr\right)^2\right\}\\
=\ &\sigma_\varepsilon^2.
\end{aligned}$$

□

Proof of equations (23) to (25). (23) is immediate from the definitions (5) and (6). (24) can be derived by a direct calculation:

$$
\begin{aligned}
\mathrm{H}_T &= \frac{\sqrt{T}(\hat{\sigma}_T^2 - \tilde{\sigma}_T^2)}{\hat{\kappa}_T} \\
&= \frac{\sqrt{T}}{\hat{\kappa}_T}\Big\{T^{-1}\sum_{t=1}^{T}(\Delta y_t)^2 - \frac{\sum_{t=1}^{T}(\Delta y_t)^2\sum_{t=1}^{T}y_{t-1}^4 - \sum_{t=1}^{T}(\Delta y_t)^2 y_{t-1}^2\sum_{t=1}^{T}y_{t-1}^2}{T\sum_{t=1}^{T}y_{t-1}^4 - \big(\sum_{t=1}^{T}y_{t-1}^2\big)^2}\Big\} \\
&= \frac{\sqrt{T}}{\hat{\kappa}_T}\Big\{\frac{\sum_{t=1}^{T}\big(\Delta y_t\big)^2 y_{t-1}^2\sum_{t=1}^{T}y_{t-1}^2 - T^{-1}\sum_{t=1}^{T}\big(\Delta y_t\big)^2\big(\sum_{t=1}^{T}y_{t-1}^2\big)^2}{T\sum_{t=1}^{T}y_{t-1}^4 - \big(\sum_{t=1}^{T}y_{t-1}^2\big)^2}\Big\} \\
&= \sqrt{T}\sum_{t=1}^{T}y_{t-1}^2\frac{\sum_{t=1}^{T}y_{t-1}^2\big\{(\Delta y_t)^2 - T^{-1}\sum_{t=1}^{T}(\Delta y_t)^2\big\}}{\hat{\kappa}_T\big\{T\sum_{t=1}^{T}y_{t-1}^4 - \big(\sum_{t=1}^{T}y_{t-1}\big)^2\big\}} \\
&= T^{-3/2}\hat{\kappa}_T^{-1}\hat{\sigma}_T^{-2}\sum_{t=1}^{T}y_{t-1}^2\{(\Delta y_t)^2 - \hat{\sigma}_T^2\} \times \frac{\hat{\sigma}_T^2 T^{-2}\sum_{t=1}^{T}y_{t-1}^2}{T^{-3}\sum_{t=1}^{T}y_{t-1}^4 - \big(T^{-2}\sum_{t=1}^{T}y_{t-1}^2\big)^2},
\end{aligned}
$$

which is identical to (24). Equation (25) follows immediately from (24) and the fact that

$$
\mathrm{C}_T = \mathrm{H}_T \times \frac{\hat{\sigma}_T^2}{T^{-2}\sum_{t=1}^{T}y_{t-1}^2}.
$$

□

Proof of Theorem 9.

(a) In view of (24), (18) and Lemma 1, an application of the CMT gives

$$
\begin{aligned}
\mathrm{H}_T &\overset{a}{\sim} \Big\{\int_0^1\Big(W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\Big)dW_\eta(r) + T^{3/2-2\beta}\frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon}\int_0^1\Big[W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\Big]^2 dr\Big\} \\
&\quad\times \frac{\sigma_\varepsilon^4\int_0^1 W_\varepsilon^2(r)dr}{\sigma_\varepsilon^4\big\{\int_0^1 W_\varepsilon^4(r)dr - \big(\int_0^1 W_\varepsilon^2(r)dr\big)^2\big\}} \\
&= \int_0^1 W_\varepsilon^2(r)dr\frac{\int_0^1\big[W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\big]dW_\eta(r)}{\int_0^1 W_\varepsilon^4(r)dr - \big(\int_0^1 W_\varepsilon^2(r)dr\big)^2} + T^{3/2-2\beta}\frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon}\int_0^1 W_\varepsilon^2(r)dr.
\end{aligned}
$$

(b) In view of (25), (18) and Lemma 1, an application of the CMT gives

$$
\begin{aligned}
\mathrm{C}_T &\overset{a}{\sim} \Big\{\int_0^1\Big(W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\Big)dW_\eta(r) + T^{3/2-2\beta}\frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon}\int_0^1\Big[W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\Big]^2 dr\Big\} \\
&\quad\times \frac{\sigma_\varepsilon^4}{\sigma_\varepsilon^4\big\{\int_0^1 W_\varepsilon^4(r)dr - \big(\int_0^1 W_\varepsilon^2(r)dr\big)^2\big\}} \\
&= \frac{\int_0^1\big[W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s)ds\big]dW_\eta(r)}{\int_0^1 W_\varepsilon^4(r)dr - \big(\int_0^1 W_\varepsilon^2(r)dr\big)^2} + T^{3/2-2\beta}\frac{c^2\sigma_\varepsilon^2}{\kappa_\varepsilon}.
\end{aligned}
$$

(c) The proof is similar to that of part (a) and hence is omitted.

(d) The proof is similar to that of part (b) and hence is omitted.

□

Proof of Proposition 11. First, note that $T^{-1/2}\tilde{y}_{t-1} \Rightarrow W_\varepsilon(r)$. Using the Taylor expansion for $\log(1+x)$ and $(1+x)^{-1}$, $q_T(\bar{c}^2, \sigma_\varepsilon^2)$ is expanded as

$$\begin{aligned}
q_T(\bar{c}^2, \sigma_\varepsilon^2) &= -\sum_{t=1}^{T}\left(\bar{c}^2 T^{-3/2}\tilde{y}_{t-1}^2 - \frac{\bar{c}^4}{2}T^{-3}\tilde{y}_{t-1}^4 + O_p(T^{-3/2})\right) \\
&\quad - \sum_{t=1}^{T}(\Delta\tilde{y}_t)^2\left(1 - \bar{c}^2 T^{-3/2}\tilde{y}_{t-1}^2 + \bar{c}^4 T^{-3}\tilde{y}_{t-1}^4 + O_p(T^{-3/2})\right) + \sum_{t=1}^{T}(\Delta\tilde{y}_t)^2 \\
&= \bar{c}^2\left[T^{-3/2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2\{(\Delta\tilde{y}_t)^2 - 1\} - \frac{\bar{c}^2}{2}T^{-3}\sum_{t=1}^{T}\tilde{y}_{t-1}^4\right] - \bar{c}^4 T^{-3}\sum_{t=1}^{T}\tilde{y}_{t-1}^4\{(\Delta\tilde{y}_t)^2 - 1\} + O_p(T^{-1/2}) \\
&=: \bar{c}^2\left[G_{1,T} - G_{2,T}\right] - G_{3,T} + O_p(T^{-1/2}). \qquad \text{(A.17)}
\end{aligned}$$

The first term satisfies

$$\begin{aligned}
G_{1,T} &= T^{-3/2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2\{c^2T^{-3/2}\tilde{y}_{t-1}^2 + c^2T^{-3/2}\tilde{y}_{t-1}^2(v_t^2 - 1) + 2cT^{-3/4}\tilde{y}_{t-1}\tilde{\varepsilon}_t v_t + (\tilde{\varepsilon}^2 - 1)\} \\
&= c^2T^{-3}\sum_{t=1}^{T}\tilde{y}_{t-1}^4 + \frac{1}{\sigma_\varepsilon^2}T^{-3/2}\sum_{t=1}^{T}\tilde{y}_{t-1}^2(\varepsilon_t^2 - \sigma_\varepsilon^2) + O_p(T^{-1/4}) \\
&\Rightarrow c^2\int_0^1 W_\varepsilon^4(r)dr + \frac{\kappa_\varepsilon}{\sigma_\varepsilon^2}\int_0^1 W_\varepsilon^2(r)dW_\eta(r).
\end{aligned}$$

For the second term, it is easy to show $G_{2,T} \Rightarrow \frac{\bar{c}^2}{2}\int_0^1 W_\varepsilon^4(r)dr$. The third term becomes

$$G_{3,T} = \bar{c}^4T^{-3}\sum_{t=1}^{T}\tilde{y}^4\left\{c^2T^{-3/2}\tilde{y}_{t-1}^2 + c^2T^{-3/2}\tilde{y}_{t-1}^2(v_t^2 - 1) + 2cT^{-3/4}\tilde{y}_{t-1}\tilde{\varepsilon}_t v_t + (\tilde{\varepsilon}_t^2 - 1)\right\} = O_p(T^{-1/2}).$$

Thus, we arrive at

$$q_T(\bar{c}^2, \sigma_\varepsilon^2) \Rightarrow \bar{c}^2\left[\frac{\kappa_\varepsilon}{\sigma_\varepsilon^2}\int_0^1 W_\varepsilon^2 dW_\eta(r) + \left(c^2 - \frac{\bar{c}^2}{2}\right)\int_0^1 W_\varepsilon^4(r)dr\right].$$

If (ε_t, v_t) is Gaussian, then $\kappa_\varepsilon/\sigma_\varepsilon^2 = \sqrt{2}$, and W_ε and W_η are independent.

Next, let $\hat{\tilde{y}}_t := y_t/\hat{\sigma}_T$. Since $T^{-1/2}\hat{\tilde{y}}_{\lfloor Tr\rfloor} \Rightarrow W_\varepsilon(r)$, which implies $\hat{\tilde{y}}_t = O_p(T^{1/2})$, from the calculation leading to (A.17), we obtain

$$q_T(\bar{c}^2, \hat{\sigma}_T^2) = \bar{c}^2\left[\hat{G}_{1,T} - \hat{G}_{2,T}\right] - \hat{G}_{3,T} + O_p(T^{-1/2}),$$

where

$$\hat{G}_{1,T} \coloneqq \frac{1}{\hat{\sigma}_T^4} T^{-3/2} \sum_{t=1}^{T} y_{t-1}^2 \big\{ (\Delta y_t)^2 - \hat{\sigma}_T^2 \big\},$$
$$\hat{G}_{2,T} \coloneqq \frac{\bar{c}^2}{2} T^{-3} \sum_{t=1}^{T} \hat{\tilde{y}}_{t-1}^4,$$

and

$$\hat{G}_{3,T} \coloneqq \frac{\bar{c}^4}{\hat{\sigma}_T^6} T^{-3} \sum_{t=1}^{T} y_{t-1}^4 \big\{ (\Delta y_t)^2 - \hat{\sigma}_T^2 \big\}.$$

Since $\hat{G}_{1,T} = \hat{\kappa}_T \hat{\sigma}_T^{-2} \mathrm{MT}_T$, we have

$$\hat{G}_{1,T} \Rightarrow \frac{\kappa_\varepsilon}{\sigma_\varepsilon^2} \int_0^1 \left(W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s) ds \right) dW_\eta(r) + c^2 \int_0^1 \left[W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s) ds \right]^2 dr.$$

We also have $\hat{G}_{2,T} \Rightarrow \frac{\bar{c}^2}{2} \int_0^1 W_\varepsilon^4(r) dr$. Finally, for the third term $\hat{G}_{3,T}$, we get

$$\begin{aligned}
\hat{G}_{3,T} &= \frac{\bar{c}^4}{\hat{\sigma}_T^6} T^{-3} \sum_{t=1}^{T} \left(y_{t-1}^4 - T^{-1} \sum_{t=1}^{T} y_{t-1}^4 \right) \left\{ (\Delta y_t)^2 - \sigma_\varepsilon^2 \right\} \\
&= \frac{\bar{c}^4}{\hat{\sigma}_T^6} T^{-3} \sum_{t=1}^{T} \left(y_{t-1}^4 - T^{-1} \sum_{t=1}^{T} y_{t-1}^4 \right) \Big\{ c^2 T^{-3/2} y_{t-1}^2 + c^2 T^{-3/2} y_{t-1}^2 (v_t^2 - 1) \\
&\qquad 2cT^{-3/4} y_{t-1} \varepsilon_t v_t + (\varepsilon_t^2 - \sigma_\varepsilon^2) \Big\} = O_p(T^{-1/2}).
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
q_T(\bar{c}^2, \hat{\sigma}_T^2) \Rightarrow \bar{c}^2 \Bigg[& \frac{\kappa_\varepsilon}{\sigma_\varepsilon^2} \int_0^1 \left(W_\varepsilon^2 - \int_0^1 W_\varepsilon^2(s) ds \right) dW_\eta(r) \\
& + c^2 \int_0^1 \left\{ W_\varepsilon^2(r) - \int_0^1 W_\varepsilon^2(s) ds \right\}^2 dr - \frac{\bar{c}^2}{2} \int_0^1 W_\varepsilon^4(r) dr \Bigg].
\end{aligned}$$

□

Proof of Proposition 12. Note that

$$\begin{aligned}
\frac{\kappa_\varepsilon^2}{\sigma_\varepsilon^4} q_T^*(\bar{c}^2, \sigma_\varepsilon^2, \kappa_\varepsilon^2) &= \frac{\bar{c}^2}{\sigma_\varepsilon^4} \left[T^{-3/2} \sum_{t=1}^{T} y_{t-1}^2 \big\{ (\Delta y_t)^2 - \sigma_\varepsilon^2 \big\} - \frac{\bar{c}^2}{2} T^{-3} \sum_{t=1}^{T} y_{t-1}^4 \right] \\
&= \bar{c}^2 \big[G_{1,T} - G_{2,T} \big],
\end{aligned}$$

where $G_{1,T}$ and $G_{2,T}$ are defined in (A.17). Thus

$$\frac{\kappa_\varepsilon^2}{\sigma_\varepsilon^4} q_T^*(\bar{c}^2, \sigma_\varepsilon^2, \kappa_\varepsilon^2) - q_T(\bar{c}^2, \sigma_\varepsilon^2) = G_{3,T} + O_p(T^{-1/2}) \xrightarrow{p} 0.$$

The second statement is obtained by noting that $\hat{\kappa}_T^2 \xrightarrow{p} \kappa_\varepsilon^2$. □

Proof of Theorem 13. Noting that

$$Q_T^*(\bar{c}^2, \kappa_\varepsilon^2) = \frac{\bar{c}^2}{\kappa_\varepsilon^2}\left[\hat{\kappa}_T \hat{\sigma}_T^2 \mathrm{MT}_T - \frac{\bar{c}^2}{2}\int_0^1 \left(Y_T^2(r) - \int_0^1 Y_T^2(s)ds\right)^2 dr\right],$$

the desired results follow from the consistency of $\hat{\kappa}_T^2$, Theorem 5(a) and the CMT. □

Table 1: Selected quantiles of $\hat{\sigma}_T^2$ and $\tilde{\sigma}_T^2$

	0%	1%	5%	25%	50%	75%	95%	99%	100%
					$\mathbf{c^2 = 0}$				
	$T = 10^2$								
$\hat{\sigma}_T^2$	0.53	0.70	0.77	0.89	0.98	1.08	1.23	1.35	1.67
$\tilde{\sigma}_T^2$	−0.10	0.55	0.68	0.85	0.98	1.12	1.36	1.56	2.44
	$T = 10^4$								
$\hat{\sigma}_T^2$	0.95	0.97	0.98	0.99	1	1.01	1.02	1.03	1.06
$\tilde{\sigma}_T^2$	0.89	0.95	0.97	0.99	1	1.01	1.03	1.05	1.11
	$T = 10^5$								
$\hat{\sigma}_T^2$	0.98	0.99	0.99	1	1	1	1.01	1.01	1.02
$\tilde{\sigma}_T^2$	0.97	0.98	0.99	1	1	1	1.01	1.02	1.03
					$\mathbf{c^2 = 0.8}$				
	$T = 10^2$								
$\hat{\sigma}_T^2$	0.58	0.78	0.88	1.06	1.24	1.56	2.97	5.88	51.43
$\tilde{\sigma}_T^2$	−21.53	0.05	0.54	0.82	0.98	1.14	1.46	1.89	12.52
	$T = 10^4$								
$\hat{\sigma}_T^2$	0.98	1.02	1.04	1.11	1.23	1.53	2.87	5.56	48.89
$\tilde{\sigma}_T^2$	0.38	0.91	0.96	0.98	1.00	1.02	1.05	1.09	1.94
	$T = 10^5$								
$\hat{\sigma}_T^2$	1.01	1.03	1.04	1.10	1.22	1.53	2.88	5.46	47.00
$\tilde{\sigma}_T^2$	0.64	0.97	0.99	0.99	1.00	1.01	1.01	1.03	1.22

Entries are based on 50,000 Monte Carlo replications where data are generated from $y_t = (1 + cv_t/\sqrt{T})y_{t-1} + \varepsilon_t$ with $y_0 = 0$ and $(\varepsilon_t, v_t)' \sim$ i.i.d $\mathrm{N}(0, \Omega)$ for $\Omega = \mathrm{diag}(1, 1)$

Table 2: Asymptotic critical values of the tests

	10%	5%	1%
C	6.08	10.21	23.21
H	1.27	1.69	2.62
MT	0.49	0.84	1.93

Entries are based on 100,000 Monte Carlo replications.

Table 3: Results of tests for real effective exchange rates

	MT	LN	C	H	UR
CA	6.745***	2.611***	0.985	12.14***	**
CN	2.501***	0.480	0.114	1.836**	-
GM	14.29***	2.686***	0.471	18.28***	**
JP	9.421***	4.957***	2.595	12.61***	***
UK	-1.921	-0.618	-0.146	-2.094	-
US	3.495***	1.047	0.334	6.853***	-

[a] CA, CN, GM, JP, UK, and US signifies Canada, China, Germany, Japan, the United Kingdom and the United States, respectively.

[b] *, **, *** denote significance at the 10%,5% and 1% levels, respectively. - for UR means no significance at any level.

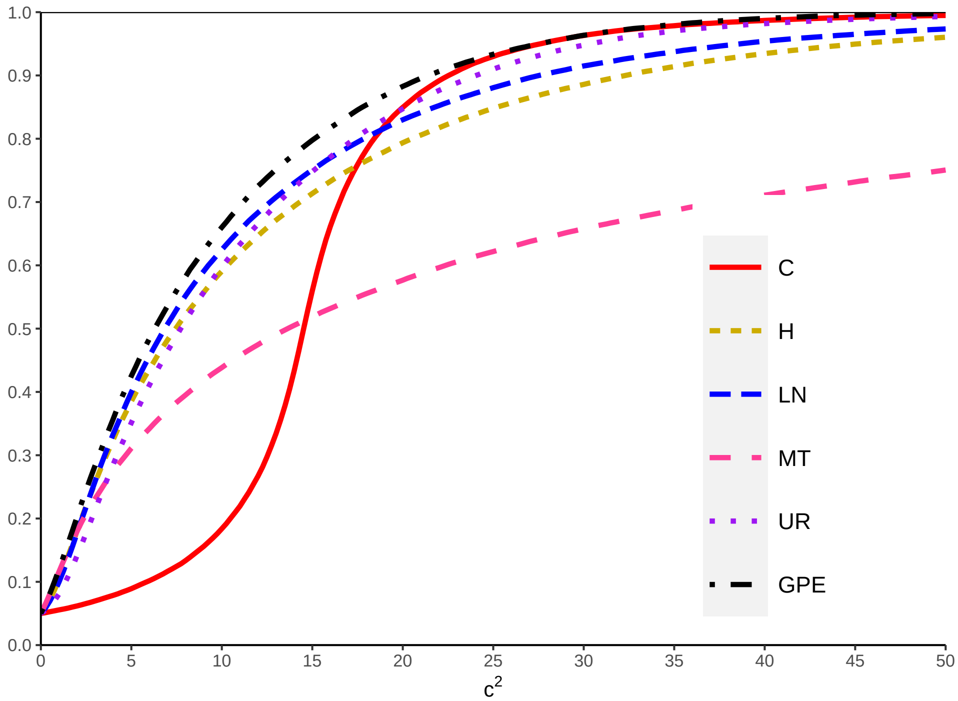

Figure 1: Local asymptotic power functions and the Gaussian power envelope (GPE)

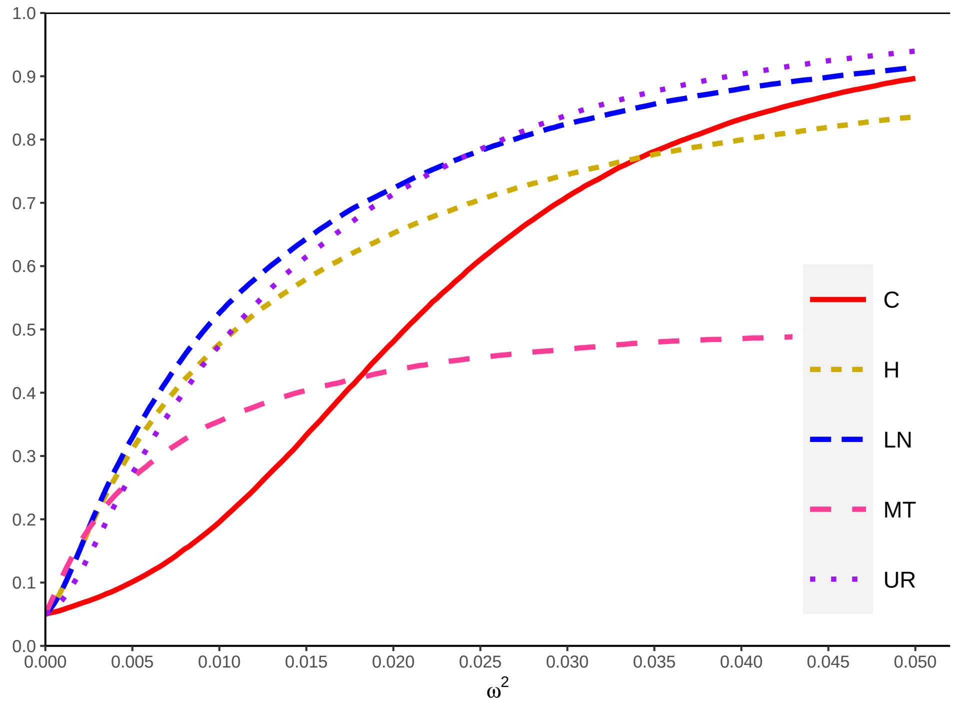

Figure 2: Finite-sample size-adjusted power functions, $T = 100$

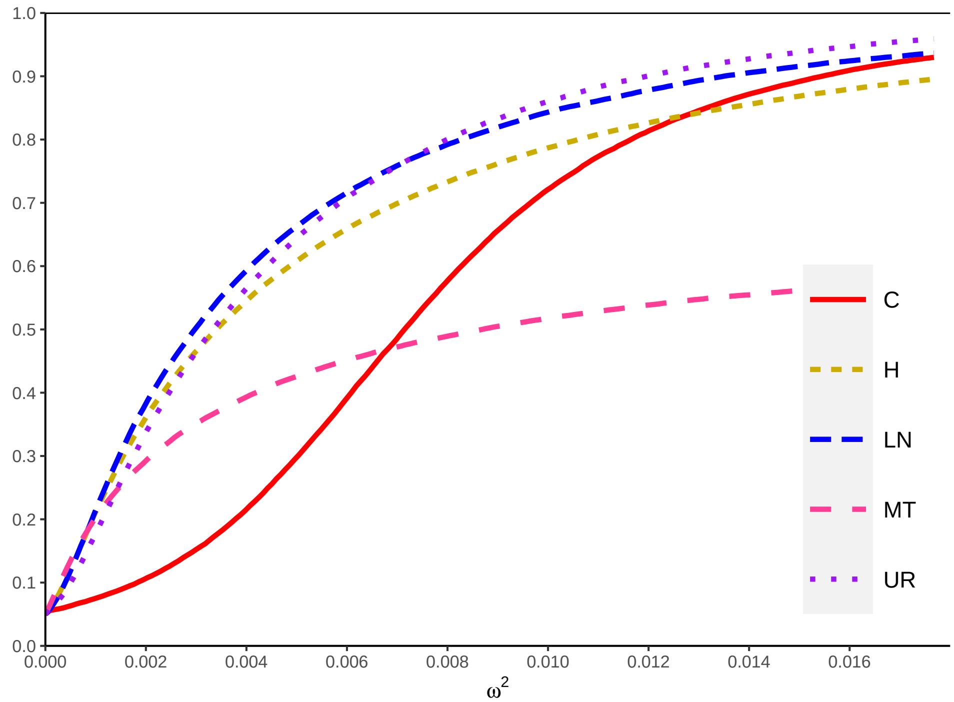

Figure 3: Finite-sample size-adjusted power functions, $T = 200$