Bootstrap Unit Root Tests

Joon Y. Park[1]
Department of Economics
Rice University
and
School of Economics
Seoul National University

Abstract

We consider the bootstrap unit root tests based on finite order autoregressive integrated models driven by iid innovations, with or without deterministic time trends. A general methodology is developed to approximate asymptotic distributions for the models driven by integrated time series, and used to obtain asymptotic expansions for the Dickey-Fuller unit root tests. The second-order terms in their expansions are of stochastic orders $O_p(n^{-1/4})$ and $O_p(n^{-1/2})$, and involve functionals of Brownian motions and normal random variates. The asymptotic expansions for the bootstrap tests are also derived and compared with those of the Dickey-Fuller tests. We show in particular that the bootstrap offers asymptotic refinements for the Dickey-Fuller tests, i.e., it corrects their second-order errors. More precisely, it is shown that the critical values obtained by the bootstrap resampling are correct up to the second-order terms, and the errors in rejection probabilities are of order $o(n^{-1/2})$ if the tests are based upon the bootstrap critical values. Through simulations, we investigate how effective is the bootstrap correction in small samples.

First Draft: February, 1999
This version: October, 2002

Key words and phrases: bootstrap, unit root test, asymptotic expansion.

[1] I am grateful to Coeditor, Joel Horowitz, and three anonymous referees for many constructive suggestions, which have greatly improved the expostion of the paper. I also wish to thank Bill Brown and Yoosoon Chang for helpful discussions and comments. Lars Muus suggested the research reported here a decade ago, while he was my colleague at Cornell. Earlier versions of this paper were presented at World Congress of the Econometric Society, Seattle, 2000 and Texas Econometrics Camp, Tanglewood, 1999.

1. Introduction

It is now well perceived that the bootstrap, if applied appropriately, helps to compute the critical values of asymptotic tests more accurately in finite samples, and that the tests based on the bootstrap critical values generally have actual finite sample rejection probabilities closer to their asymptotic nominal values. See, e.g., Hall (1992) and Horowitz (2001). The bootstrap unit root tests, i.e., the unit root tests relying on the bootstrap critical values, seem particularly attractive in this respect. For most of the commonly used unit root tests, the discrepancies in the actual and nominal rejection probabilities are known to be large and often too large for the tests to be any reliable. It has indeed been observed by various authors including Ferretti and Romo (1996) and Nankervis and Savin (1996) that the bootstrap tests have actual rejection probabilities that are much closer to their nominal values, compared to the asymptotic tests, in the unit root models.

The main purpose of this paper is to provide a theory for the asymptotic refinement of bootstrap unit root tests. Bootstrap theories for unit root models have previously been studied by, among others, Basawa et al. (1991a, 1991b), Datta (1996), Park (2002) and Chang and Park (2002). However, they have all been restricted to the consistency (and inconsistency) of the bootstrap estimators and statistics from unit root models. None of them considers the asymptotic refinement of bootstrap. In this paper, we develop asymptotic expansions that are applicable for a wide class of unit root tests and their bootstrap versions, and provide a framework within which we investigate the bootstrap asymptotic refinement of various unit root tests. Our asymptotic expansions are obtained by analyzing the Skorohod embedding, i.e., the embedding of the partial sum process into a Brownian motion defined on an extended probability space.

In the paper, we consider more specifically the Dickey-Fuller unit root tests for the finite order autoregressive unit root models driven by iid errors, possibly with constant and linear time trend. It can be clearly seen, however, that our methodology may also be used to analyze many other unit root tests as well. For the Dickey-Fuller unit root tests, the expansions have as the leading term the functionals of Brownian motion representing their asymptotic distributions. This is as expected. The second-order terms in the expansions are, however, quite different from the standard Edgeworth-type expansions for the stationary models. They are represented by functionals of Brownian motions and normal random variates, which are of stochastic orders $O_p(n^{-1/4})$ and $O_p(n^{-1/2})$. The second-order expansion terms involve various unknown model parameters. The expansions are obtained for the tests in models with deterministic trends, as well as for the tests in purely stochastic models. They have similar characteristics.

We show that the limiting distributions of the bootstrap statistics have expansions that are analogous to the original statistics. The bootstrap statistics have the same leading expansion terms. This is well expected, since the statistics that we consider are asymptotically pivotal. More importantly, their second-order terms are also exactly the same as the original statistics except that the unknown parameters included in the expansions of the original statistics are now replaced by their sample analogues, which strongly converge to the corresponding population parameters. Consequently, using the critical values obtained by the bootstrap is expected to reduce the order of discrepancy between the actual (finite sample) and nominal (asymptotic) rejection probabilities of the tests. The bootstrap thus provides an asymptotic refinement for the tests. Though our asymptotic expansions for the unit root models are quite different from the Edgeworth-type expansions for stationary models, the reason that the bootstrap offers a refinement of asymptotics is precisely the same.

Through simulations, we investigate how effective the bootstrap correction is in small samples. We consider both Gaussian and non-Gaussian unit root models. For the non-Gaussian models, we investigate models driven by innovations that are distributed symmetrically and asymmetrically. Our findings are generally supportive of the theory developed in the paper. Moreover, they are consistent with the simulation results obtained earlier by Nankervis and Savin (1996). Overall, the bootstrap does provide some obvious improvements over the asymptotics. The tests based on the bootstrap critical values in general have rejection probabilities that are substantially closer to their nominal values. The actual magnitudes of improvements, however, somewhat vary depending upon the distributional characteristics of innovations, the size of samples and the presence of deterministic trends in the model. It appears in particular that the benefits from the bootstrap are more noticeable for the models with trends and for the samples of small sizes.

The rest of the paper is organized as follows. Section 2 introduces the model, tests and bootstrap method. The test statistics are introduced together with the autoregressive unit root model and the moment condition, and how to obtain bootstrap samples from such a model is explained here. The asymptotic expansions are derived in Section 3. The section starts with the probabilistic embeddings that are essential for the development of our subsequent theory, and present the asymptotic expansions for the original and bootstrap tests. Some of their implications are also discussed. The asymptotic powers of the bootstrap tests against the local-to-unity model are considered in Section 4. Section 5 extends the theory to the models with deterministic trends. The asymptotic expansions for the tests in models with constant and linear time trend are presented and compared with the earlier results. The simulation results are reported in Section 6, and Section 7 concludes the paper. Mathematical proofs are given in Section 8.

2. The Model, Tests and Bootstrap Method

2.1 The Model and Test Statistics

We consider the test of the unit root hypothesis

$$\mathrm{H}_0 : \alpha = 1 \tag{1}$$

in the AR(p) unit root model

$$y_t = \alpha y_{t-1} + \sum_{i=1}^{p} \alpha_i \triangle y_{t-i} + \varepsilon_t \tag{2}$$

where $\triangle$ is the usual difference operator. We define

$$\alpha(z) = 1 - \sum_{i=1}^{p} \alpha_i z^i$$

so that under the null hypothesis of the unit root (1) we may write $\alpha(L)\triangle y_t = \varepsilon_t$ using the lag operator L. Assume

Assumption 2.1 Let (ε_t) be an iid sequence with $\mathbf{E}\varepsilon_t = 0$ and $\mathbf{E}|\varepsilon_t|^r < \infty$ for some $r > 1$. Also, we assume that $\alpha(z) \neq 0$ for all $|z| \leq 1$.

Under Assumption 2.1 (with $r \geq 2$) and the unit root hypothesis (1), the time series $(\triangle y_t)$ becomes a (second-order) stationary AR(p) process.

The unit root hypothesis is customarily tested using the t-statistic on α in regression (2). Denote by $\hat{\alpha}_n$ the OLS estimator for α in regression (2). If we let

$$x_{t-1} = (\triangle y_{t-1}, \ldots, \triangle y_{t-p})'$$

and define

$${}_p y_{t-1} = y_{t-1} - \left(\sum_{t=1}^{n} y_{t-1} x_{t-1}' \right) \left(\sum_{t=1}^{n} x_{t-1} x_{t-1}' \right)^{-1} x_{t-1},$$

then we may explicitly write the t-statistic for the null hypothesis (1) as

$$F_n = \frac{\hat{\alpha}_n - 1}{\sigma_n \left(\sum_{t=1}^{n} {}_p y_{t-1}^2 \right)^{-1/2}} \tag{3}$$

where σ_n^2 is the usual variance estimator for the regression errors. The test is first proposed and investigated by Dickey and Fuller (1979, 1981), and it is commonly referred to as the Dickey-Fuller test (if applied to the regressions with no lagged difference term) or the augmented Dickey-Fuller (ADF) test (if based on the regressions augmented with lagged difference terms).

We may also use the statistic

$$G_n = \frac{n(\hat{\alpha}_n - 1)}{\alpha_n(1)} \tag{4}$$

to test the unit root hypothesis, where

$$\alpha_n(1) = 1 - \sum_{i=1}^{p} \alpha_{ni}$$

with the least squares estimators α_{ni} of α_i for $i = 1, \ldots, p$. The statistic G_n reduces to the normalized coefficient $n(\hat{\alpha}_n - 1)$ in the simple model with no lagged difference term.

The asymptotic distributions of the statistics F_n and G_n under the presence of a unit root are well known [see, e.g., Stock (1994)], and given by

$$F_n \to_d F = \frac{\displaystyle\int_0^1 W(t)dW(t)}{\left(\displaystyle\int_0^1 W(t)^2 dt\right)^{1/2}}, \quad G_n \to_d G = \frac{\displaystyle\int_0^1 W(t)dW(t)}{\displaystyle\int_0^1 W(t)^2 dt}$$

where W is the standard Brownian motion. Since F and G do not involve any nuisance parameter, the statistics F_n and G_n are asymptotically pivotal. The distributions represented by F and G are however non-standard, and they are tabulated in Fuller (1996). See Evans and Savin (1981, 1984) for a detailed discussion on some of their distributional characteristics.

The initialization of (y_t) is important for some of our subsequent theories. In what follows, we let $(y_0, \ldots, y_{-p})$ be fixed and make all our arguments conditional on them. If we let $\alpha = 1$ and define $u_t = \triangle y_t$, then we may equivalently assume that $(y_0, (u_0, \ldots, u_{-p+1}))$ are given. This convention on the initialization of (y_t) is crucial for the theory developed in Section 3 for the model with no constant term. It will however be unimportant for the model with constant or linear time trend considered in Section 4. Under the unit root hypothesis, our statistics become invariant with respect to the initial values of (y_t) in the regression with intercept.

2.2 The Bootstrap Method

Implementation of the bootstrap method in our unit root model is quite straightforward, once we fit the regression

$$\triangle y_t = \sum_{i=1}^{p} \alpha_i \triangle y_{t-i} + \varepsilon_t \tag{5}$$

and obtain the coefficient estimates (α_{ni}) and the fitted residuals $(\hat{\varepsilon}_t)$. Since our purpose is to bootstrap the distributions of the statistics under the null hypothesis of the unit root, it seems natural to resample from the restricted regression (5) instead of the unrestricted one in (2). It is indeed well known that the bootstrap must be based on regression (5), not on regression (2), for consistency [see Basawas, et al. (1991a)].[2]

The first step is to draw bootstrap samples for the innovations (ε_t) after mean correction. As usual, we denote by (ε_t^*) their bootstrap samples, i.e., (ε_t^*) are the samples from

$$\left(\hat{\varepsilon}_t - \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i\right)_{t=1}^{n}$$

which can be viewed as iid samples from the empirical distribution given by $(\hat{\varepsilon}_t - \sum_{i=1}^n \hat{\varepsilon}_i/n)$. Note that the mean adjustment is necessary, since otherwise the mean of the bootstrap samples is nonzero.

[2] We may estimate (α_i) and (ε_t) from regression (2), as long as we set the value of α to unity (instead of its estimated value) and use regression (6) to generate bootstrap samples. The resulting differences are of order $o(n^{-1}\log n)$ a.s., and therefore, will not change any of our subsequent theory.

Once the bootstrap samples (ε_t^*) are obtained, we may construct the values for (u_t^*) recursively from (ε_t^*) as

$$u_t^* = \sum_{i=1}^{p} \alpha_{ni} u_{t-i}^* + \varepsilon_t^* \tag{6}$$

starting from $(u_0, \ldots, u_{-p+1})$. Finally, the bootstrap samples (y_t^*) for (y_t) can be obtained just by taking partial sums of (u_t^*), i.e.,

$$y_t^* = y_0 + \sum_{i=1}^{t} u_i^*$$

given y_0. For the model with no intercept term, the initializations of (u_t^*) and (y_t^*) are important and should be done as specified here to make our theory applicable. However, they become unimportant for the models with deterministic trends including constant, as in the case of the initializations of (u_t) and (y_t).

The bootstrap versions of the statistics F_n and G_n, which we denote by F_n^* and G_n^* respectively, are defined from (y_t^*) exactly in the same way that F_n and G_n in (3) and (4) are constructed from (y_t). Of course, the distributions of the bootstrap statistics F_n^* and G_n^* can now be found by repeatedly generating bootstrap samples and computing their values in each bootstrap repetition. These distributions are regarded as approximations of the null distributions of F_n and G_n. The bootstrap unit root tests use the critical values calculated from the distributions of the bootstrap statistics F_n^* and G_n^*.

3. Asymptotic Expansions of Test Statistics

3.1 Probabilistic Embeddings

Our subsequent theoretical development relies heavily on the probabilistic embedding of the partial sum process constructed from the innovation sequence (ε_i) into a Brownian motion in an expanded probability space. This will be given below. Throughout the paper, we denote by $\mathbf{E}\varepsilon_i^2 = \sigma^2, \mathbf{E}\varepsilon_i^3 = \mu^3$ and $\mathbf{E}\varepsilon_i^4 = \kappa^4$, whenever they exist.

Lemma 3.1 Let Assumption 2.1 hold with $r \geq 2$. Then there exist a standard Brownian motion $(W(t))_{t\geq 0}$ and a time change $(T_i)_{i\geq 0}$ such that $T_0 \equiv 0$ and for all $n \geq 1$,

$$W\left(T_i/n\right) =_d \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{i} \varepsilon_k \tag{7}$$

$i = 1, \ldots, n$, and if we let $\Delta_i = T_i - T_{i-1}$, then Δ_i's are iid with $\mathbf{E}\Delta_i = 1$ and $\mathbf{E}|\Delta_i|^{r/2} \leq K\mathbf{E}|\varepsilon_t|^r$ for all $r \geq 2$, where K is an absolute constant depending only upon r.

The reader is referred to Hall and Heyde (1980) for the explicit construction of the time change $(T_i)_{i\geq 0}$. The result in Lemma 3.1 is originally due to Skorohod (1965). If Assumption

2.1 holds with $r > 2$, we have as shown in Park and Phillips (1999)

$$\max_{1 \le i \le n} \left| \frac{T_i - i}{n^s} \right| \to_{a.s.} 0 \tag{8}$$

for any $s > 1/2$.

In what follows, we will assume that (ε_i) and $(W, (T_i))$ are defined on the common probability space $(\Omega, \mathcal{F}, \mathbf{P})$. This causes no loss in generality since we are concerned only with the distributional results of the test statistics defined in (3) and (4), yet it will greatly simplify and clarify our subsequent exposition. The convention will be made throughout the paper. From now on, we would thus interpret the distributional equality in (7) as the usual equality. If we define a stochastic process W_n on $[0,1]$ by $W_n(t) = n^{-1/2} \sum_{i=1}^{[nt]} \varepsilon_i/\sigma$, then it follows from the Hölder continuity of the Brownian sample path and the result in (8) that

$$\sup_{0 \le t \le 1} |W_n(t) - W(t)| \le \sup_{0 \le t \le 1} \left| T_{[nt]}/n - t \right|^{1/2-\epsilon} = o(n^{-1/4+\epsilon}) \text{ a.s.} \tag{9}$$

for any $\epsilon > 0$. Therefore, we have in particular $W_n \to_{a.s.} W$ uniformly on $[0,1]$. Throughout the paper, we let $T_{ni} = T_i/n$, $i = 1, \ldots, n$, for notational brevity.

For the development of our asymptotic expansions, it is necessary to define additional sequences defined from the Brownian motion W and the time change (T_i) introduced in Lemma 3.1. We let

$$\delta_i = \Delta_i - 1$$

for $i = 1, \ldots, n$. Moreover, we define

$$\eta_i = n \int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})] dW(t)$$

for $i = 1, \ldots, n$. Note that (δ_i) and (η_i) are iid sequences of random variables. We also need to consider the sequence (ξ_i) given by

$$\xi_i = x_{i-1} \varepsilon_i.$$

Clearly, (ξ_i) is a martingale difference sequence. Under the null hypothesis of the unit root, it has conditional covariance matrix whose expectation is given by $\sigma^4 \Gamma$, where $\Gamma = \mathbf{E} x_i x_i'/\sigma^2$. Finally, we let $\mathbf{E}\,\delta_i^2 = \tau^4/\sigma^4$, which is finite under Assumption 2.1. Note that $\delta_i \equiv 0$, when and only when (ε_i) are normal. The parameter τ can therefore be regarded as the *non-normality* parameter. Subsequently, we set $\tau = 0$ if and only if (ε_i) are normal. The parameters Γ and τ^4 defined here, in addition to σ^2, μ^3 and κ^4 introduced earlier, will appear frequently in the development of our asymptotic expansions.

Now we define

$$v_i = (\varepsilon_i/\sigma, \delta_i, \eta_i, \xi_i'/\sigma^2)'$$

and let

$$B_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} v_i. \tag{10}$$

Then invariance principle holds, and $B_n \to_d B$ for a properly defined vector Brownian motion B. We present this formally as a lemma.

Lemma 3.2 Let Assumption 2.1 hold with $r > 4$. Then $B_n \to_d B$, where B is a vector Brownian motion with covariance matrix Σ given by

$$\Sigma = \begin{pmatrix} 1 & \mu^3/3\sigma^3 & \mu^3/3\sigma^3 & 0 \\ & \tau^4/\sigma^4 & (\kappa^4 - 3\sigma^4 - 3\tau^4)/12\sigma^4 & 0 \\ & & \kappa^4/6\sigma^4 & 0 \\ & & & \Gamma \end{pmatrix}$$

where the parameters are defined earlier in this section.

Following our earlier convention, we subsequently assume that both B_n and B are defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$, and that $B_n \to_{a.s.} B$. It is well known that any weakly convergent random sequence can be represented, up to the distributional equivalence, by a random sequence which converges a.s. [see, e.g., Pollard (1984)].

Remark We make a partition of the limit Brownian motion B as

$$B = (W, V, U, Z')'$$

conformably with (v_i). Let $(W^{\cdot}, W^{\cdot\cdot})$ be a bivariate standard Brownian motion independent of W. Clearly, we may then write

$$\begin{aligned} U &= \omega W + \dot{\omega} W^{\cdot}, \\ V &= \omega W + \omega^{\cdot} W^{\cdot} + \omega^{\cdot\cdot} W^{\cdot\cdot}, \end{aligned}$$

where

$$\begin{aligned} \omega &= \frac{\mu^3}{3\sigma^3}, \\ \dot{\omega} &= \left(\frac{\kappa^4}{6\sigma^4} - \frac{\mu^6}{9\sigma^6} \right)^{1/2}, \\ \omega^{\cdot} &= \left(\frac{\kappa^4}{6\sigma^4} - \frac{\mu^6}{9\sigma^6} \right)^{-1/2} \left(\frac{\kappa^4 - 3\sigma^4}{12\sigma^4} - \frac{\tau^4}{4\sigma^4} - \frac{\mu^6}{9\sigma^6} \right), \\ \omega^{\cdot\cdot} &= \left[\frac{\tau^4}{\sigma^4} - \frac{\mu^6}{9\sigma^6} - \left(\frac{\kappa^4}{6\sigma^4} - \frac{\mu^6}{9\sigma^6} \right)^{-1} \left(\frac{\kappa^4 - 3\sigma^4}{12\sigma^4} - \frac{\tau^4}{4\sigma^4} - \frac{\mu^6}{9\sigma^6} \right)^2 \right]^{1/2}. \end{aligned}$$

The representations can be greatly simplified for the Gaussian models, for which we have $\mu^3 = 0$ and $\kappa^4 = 3\sigma^4$ as well as $\tau = 0$. Consequently, we have $\dot{\omega} = 1/\sqrt{2}$ and $\omega = \omega^{\cdot} = \omega^{\cdot\cdot} = 0$, and therefore, $V \equiv 0$ and U becomes independent of W. In addition to the representations of U and V given above, we may write $Z(1) = \Gamma^{1/2} S$ using a multivariate normal random vector S with the identity covariance matrix. Since Z is independent of (W, V, U), so is S. Finally, our subsequent expansions also involve stochastic processes M and N. We let M be an extended standard Brownian motion on $\mathbf{R}$ independent of B (and therefore all of the Brownian motions and normal random variates defined above), and let N be another extended Brownian motion on $\mathbf{R}$ defined by $N(t) = W(1+t) - W(1)$.[3] The notations defined here will be used throughout the paper without any further reference.

3.2 Asymptotic Expansions

We are now ready to obtain asymptotic expansions for the distributions of the statistics F_n and G_n introduced in (3) and (4). We define

$$P_n = \frac{1}{n}\sum_{t=1}^{n} y_{t-1}\varepsilon_t - \frac{1}{n}\left(\sum_{t=1}^{n} y_{t-1}x_{t-1}'\right)\left(\sum_{t=1}^{n} x_{t-1}x_{t-1}'\right)^{-1}\left(\sum_{t=1}^{n} x_{t-1}\varepsilon_t\right), \tag{11}$$

$$Q_n = \frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2 - \frac{1}{n^2}\left(\sum_{t=1}^{n} y_{t-1}x_{t-1}'\right)\left(\sum_{t=1}^{n} x_{t-1}x_{t-1}'\right)^{-1}\left(\sum_{t=1}^{n} x_{t-1}y_{t-1}\right). \tag{12}$$

Also, we write the error variance estimate σ_n^2 as

$$\sigma_n^2 = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t^2 - \frac{1}{n}\left(\sum_{t=1}^{n} \varepsilon_t x_{t-1}'\right)\left(\sum_{t=1}^{n} x_{t-1}x_{t-1}'\right)^{-1}\left(\sum_{t=1}^{n} x_{t-1}\varepsilon_t\right) \tag{13}$$

and write

$$\alpha_n(1) = \alpha(1) - \iota'\left(\sum_{t=1}^{n} x_{t-1}x_{t-1}'\right)^{-1}\left(\sum_{t=1}^{n} x_{t-1}\varepsilon_t\right). \tag{14}$$

Here and elsewhere in the paper, ι denotes the p-vector of ones. The statistics F_n and G_n can now be written respectively as

$$F_n = \frac{P_n}{\sigma_n\sqrt{Q_n}} \quad \text{and} \quad G_n = \frac{P_n}{\alpha_n(1)Q_n}. \tag{15}$$

Here we assume that σ^2 and $\alpha(1)$ are estimated under the unit root restriction. This assumption is made purely for the expositional purpose. All of our subsequent results also hold for the unrestricted estimators of σ^2 and $\alpha(1)$.

To derive the asymptotic expansions for the statistics F_n and G_n, we need to consider various sample product moments in (11) – (14). The asymptotics for some of them are presented in Lemma 3.3, which can be directly obtained from the probabilistic embeddings developed in the previous section. Proposition 3.4 is a direct consequence of Lemma 3.3. To simplify the subsequent exposition, we use X to denote $X(1)$, as well as the process itself, for Brownian motion X. This should cause no confusion.

[3]The definition of N, of course, requires that $W(t)$ be defined for $t < 0$ as well as for $t \geq 0$. In the subsequent development of our theory, we assume that the necessary extension is made and W is an extended Brownian motion defined on $\mathbf{R}$.

Lemma 3.3 Let Assumption 2.1 hold with $r > 4$. Then we have

(a) $\dfrac{1}{n\sigma^2}\sum_{t=1}^{n}\varepsilon_t^2 = 1 + n^{-1/2}(V + 2U) + o_p(n^{-1/2})$,

(b) $\dfrac{1}{n^{1/2}\sigma^2}\sum_{t=1}^{n} x_{t-1}\varepsilon_t = Z + o_p(1)$,

(c) $\dfrac{1}{n\sigma^2}\sum_{t=1}^{n} x_{t-1}x_{t-1}' = \Gamma + O_p(n^{-1/2})$,

for large n.

Proposition 3.4 Let Assumption 2.1 hold with $r > 4$. Then we have

(a) $\sigma_n^2 = \sigma^2\left[1 + n^{-1/2}(V + 2U)\right] + o_p(n^{-1/2})$,

(b) $\alpha_n(1) = \alpha(1) - n^{-1/2}\iota'\Gamma^{-1}Z + o_p(n^{-1/2})$,

for large n.

We now obtain the asymptotic expansions for the sample product moments $\sum y_{t-1}\varepsilon_t$, $\sum y_{t-1}^2$ and $\sum x_{t-1}y_{t-1}$. To effectively analyze these product moments, we define $w_t = \sum_{i=1}^{t}\varepsilon_i$ for $t \geq 1$ and $w_0 \equiv 0$ and first consider the asymptotic expansions for the sample product moments of (w_t) and (ε_t). We let $u_t = \triangle y_t$ as before, so that $\alpha(L)u_t = \varepsilon_t$ under the null hypothesis of the unit root. Under the unit root hypothesis, (u_t) is just a linearly filtered process of (ε_t), and (y_t) becomes an integrated process generated by such a process. Our subsequent asymptotic expansions involve various functionals of Brownian motions. To ease the exposition, we let for Brownian motions X and Y,

$$I(X) = \int_0^1 X(t)dt \quad \text{and} \quad J(X,Y) = \int_0^1 X(t)dY(t),$$

in the subsequent development of our theory. This shorthand notation, together with $X = X(1)$ introduced above, will be used repeatedly for the rest of the paper.

Lemma 3.5 Let Assumption 2.1 hold with $r \geq 8$. Then we have

(a) $\dfrac{1}{n^{1/2}\sigma}\sum_{t=1}^{n}\varepsilon_t = W + n^{-1/4}M(V) + n^{-1/2}N(V) + o_p(n^{-1/2})$,

(b) $\dfrac{1}{n^{3/2}\sigma}\sum_{t=1}^{n} w_{t-1} = I(W) + n^{-1/2}[WV - J(W,V) - \omega] + o_p(n^{-1/2})$,

(c) $\dfrac{1}{n^2\sigma^2}\sum_{t=1}^{n} w_{t-1}^2 = I(W^2) + n^{-1/2}\left[W^2V - J(W^2,V) - 2\omega I(W)\right] + o_p(n^{-1/2})$,

(d) $\dfrac{1}{n\sigma^2}\sum_{t=1}^{n} w_{t-1}\varepsilon_t = J(W,W) + n^{-1/4}WM(V)$,

$$+ n^{-1/2}\left[(1/2)M(V)^2 + WN(V) - (1/2)(V+2U)\right] + o_p(n^{-1/2}),$$

for large n.

The asymptotic expansions for $\sum y_{t-1}\varepsilon_t$, $\sum y_{t-1}^2$ and $\sum x_{t-1}y_{t-1}$ can now be obtained using the relationships between (y_t) and (w_t), and between (u_t) and (ε_t). To write down more explicitly their relationships, we need to define some new notation. We let

$$\pi = 1/\alpha(1) \quad \text{and} \quad \pi_i = \sum_{j=i}^{p} \alpha_j/\alpha(1)$$

for $i = 1, \ldots, p$, and let

$$\varpi = (\pi_1, \ldots, \pi_p)'.$$

We also define

$$\nu = (1/\pi\sigma)\left(y_0 + \sum_{i=1}^{p} \pi_i u_{1-i}\right).$$

Note that we assume $(y_0, (u_0, \ldots, u_{-p+1}))$ to be given. Therefore, we may and will regard ν as a parameter in our subsequent analysis.

With the notation introduced above, we may write after some algebra

$$u_t = \pi\varepsilon_t + \varpi'(x_{t-1} - x_t)$$

and subsequently get

$$y_t = \pi\sigma\nu + \pi w_t - \varpi' x_t. \tag{16}$$

It is now straightforward to deduce from Lemma 3.5 that

Proposition 3.6 Let Assumption 2.1 hold with $r \geq 8$. Then we have

(a) $\displaystyle \frac{1}{n^{3/2}\pi\sigma}\sum_{t=1}^{n} y_{t-1} = I(W) + n^{-1/2}\left[WV - J(W,V) + (\nu - \omega)\right] + o_p(n^{-1/2}),$

(b) $\displaystyle \frac{1}{n\pi^2\sigma^2}\sum_{t=1}^{n} x_{t-1}y_{t-1} = \iota[1 + J(W,W)] - \Gamma\varpi/\pi^2 + o_p(1),$

(c) $\displaystyle \frac{1}{n^2\pi^2\sigma^2}\sum_{t=1}^{n} y_{t-1}^2 = I(W^2),$

$$+ n^{-1/2}\left[W^2V - J(W^2,V) + 2(\nu-\omega)I(W)\right] + o_p(n^{-1/2}),$$

(d) $\displaystyle \frac{1}{n\pi\sigma^2}\sum_{t=1}^{n} y_{t-1}\varepsilon_t = J(W,W) + n^{-1/4}WM(V),$

$$+ n^{-1/2}[(1/2)M(V)^2 + WN(V) + \nu W - (1/2)(V+2U) - \varpi' Z/\pi] + o_p(n^{-1/2}),$$

for large n.

The asymptotic expansions for the statistics F_n and G_n can now be easily obtained from (15), using the results in Lemma 3.3 and Propositions 3.4 and 3.6.

Theorem 3.7 Let Assumption 2.1 hold with $r \geq 8$. Then we have for large n

$$\begin{aligned} F_n &= F + F_1/\sqrt[4]{n} + F_2/\sqrt{n} + o_p(n^{-1/2}), \\ G_n &= G + G_1/\sqrt[4]{n} + G_2/\sqrt{n} + o_p(n^{-1/2}), \end{aligned}$$

where $F_1 = WM(V)/I(W^2)^{1/2}$, $G_1 = WM(V)/I(W^2)$ and

$$\begin{aligned} F_2 &= \frac{(1/2)M(V)^2 + WN(V) + \nu W - [1 + J(W,W)][(V+2U)/2 + \pi\iota'\Gamma^{-1}Z]}{I(W^2)^{1/2}} \\ &\quad - \frac{J(W,W)[W^2V - J(W^2,V) + 2(\nu-\omega)I(W)]}{2I(W^2)^{3/2}}, \\ G_2 &= \frac{(1/2)M(V)^2 + WN(V) + \nu W - (V+2U)/2 - \pi\iota'\Gamma^{-1}Z}{I(W^2)} \\ &\quad - \frac{J(W,W)[W^2V - J(W^2,V) + 2(\nu-\omega)I(W)]}{I(W^2)^2} \end{aligned}$$

in notation introduced earlier in Lemma 3.5 and Proposition 3.6.

Naturally, the asymptotic expansions for the statistics F_n and G_n have the leading terms F and G representing their asymptotic distributions. For both F_n and G_n, the second terms $F_1/\sqrt[4]{n}$ and $G_1/\sqrt[4]{n}$ in our expansions are of stochastic order $O_p(n^{-1/4})$. Their effects are, however, distributionally of order $O(n^{-1/2})$. More precisely, we have

$$\begin{aligned} \mathbf{P}\left\{F + F_1/\sqrt[4]{n} \leq x\right\} &= \mathbf{P}\left\{F \leq x\right\} + O(n^{-1/2}), \\ \mathbf{P}\left\{G + G_1/\sqrt[4]{n} \leq x\right\} &= \mathbf{P}\left\{G \leq x\right\} + O(n^{-1/2}), \end{aligned}$$

uniformly in x. This is because the process M included in F_1 and G_1 is a Gaussian process independent of (W,V,U). Note that for any functionals $a(W)$ and $b(W)$ of W, we have

$$a(W) + (1/\sqrt[4]{n})b(W)M(V) =_d \mathbf{MN}\left(a(W), (1/\sqrt{n})b(W)^2|V|\right)$$

where **MN** stands for mixed normal distribution.[4] Therefore, we call

$$F_{nn} = F_1/\sqrt[4]{n} + F_2/\sqrt{n}, \quad G_{nn} = G_1/\sqrt[4]{n} + G_2/\sqrt{n}$$

the second-order terms in our asymptotic expansions of F_n and G_n. The remainder terms in the expansions are given to be of order $o_p(n^{-1/2})$.

The results in Theorem 3.7 suggest that our second-order asymptotic expansions of the statistics F_n and G_n provide refinements of their asymptotic distributions up to order $o(n^{-1/2})$. This can be shown rigorously, if we assume higher moments exist. More precisely, if we let

$$ {}_2F_n = F + F_{nn}, \quad {}_2G_n = G + G_{nn}, \tag{17}$$

then we have

[4]The characteristic functions of $F + F_1/\sqrt[4]{n}$ and $G + G_1/\sqrt[4]{n}$ can therefore be expanded in integral powers of $n^{-1/2}$ with the leading terms being the characteristic functions of F and G, respectively. This shows that the second terms $F_1/\sqrt[4]{n}$ and $G_1/\sqrt[4]{n}$ have distributional effects of order $O(n^{-1/2})$.

Corollary 3.8 Let Assumption 2.1 hold with $r > 12$. Then we have

$$\mathbf{P}\{F_n \leq x\} = \mathbf{P}\left\{{}_2F_n \leq x\right\} + o(n^{-1/2}),$$
$$\mathbf{P}\{G_n \leq x\} = \mathbf{P}\left\{{}_2G_n \leq x\right\} + o(n^{-1/2}),$$

uniformly in $x \in \mathbf{R}$.

It is thus expected in general that the actual finite sample rejection probabilities of the tests F_n and G_n disagree with their nominal values only by order $o(n^{-1/2})$, if the second-order corrected critical values are used, i.e., a_λ and b_λ such that $\mathbf{P}\{{}_2F_n \leq a_\lambda\} = \lambda$ and $\mathbf{P}\{{}_2G_n \leq b_\lambda\} = \lambda$ for tests with nominal rejection probability λ.

For both statistics, the second-order terms F_{nn} and G_{nn} involve various functionals of Brownian motions. The functionals are dependent upon various model parameters, not only those included explicitly, but also those given implicitly by the variances and covariances of (W, V, U, Z) in Lemma 3.2. More precisely, if we represent V, U and Z as suggested in Remark following Lemma 3.2, then F_{nn} and G_{nn} can be written explicitly as functionals of three independent Brownian motions $W, W^{\cdot}, W^{\cdot\cdot}$ and another independent normal random vector S. The functionals involve the parameter θ defined by

$$\theta = (\nu, \pi, \sigma^2, \mu^3, \kappa^4, \tau^4, \Gamma). \tag{18}$$

We denote by $F_{nn}(\theta)$ and $G_{nn}(\theta)$ the resulting functionals respectively for F_{nn} and G_{nn}. Symbolically, we write

$$F_{nn}(\theta) = F_{nn}(\theta, (W, W^{\cdot}, W^{\cdot\cdot}, S)), \quad G_{nn}(\theta) = G_{nn}(\theta, (W, W^{\cdot}, W^{\cdot\cdot}, S)) \tag{19}$$

to signify such functionals.

Our asymptotic expansions of the statistics F_n and G_n provide some important informations on their finite sample distributions. For instance, our expansions make it clear that the initial values have effects, which are distributionally of order $O(n^{-1/2})$, on their finite sample distributions. Note that they are parametrized as ν. Moreover, we may learn from the expansions that the presence of shortrun dynamics, if it is correctly modelled, has distributional effects also of order $O(n^{-1/2})$. As is well known, neither the initial values nor the shortrun dynamics affect the limiting distributions of F_n and G_n.

Though we will not discuss the details in the paper, it is rather straightforward to obtain the second-order asymptotic expansions for many other unit root tests using our results here. For the tests considered in Stock (1994, pp2772–2773), it is indeed not difficult to see that the tests classified as $\hat{\rho}$-class, $\hat{\tau}$-class, SB-class, $J(p,q)$, LMPI (no-deterministic case) and P_T all have the asymptotic expansions that are obtainable from the results in Lemmas 3.3 and 3.5 and Propositions 3.4 and 3.6. This, of course, is true only when the nuisance parameter is estimated from the AR(p) model as for F_n and G_n considered in the paper. The nonparametric estimation of the nuisance parameter would fundamentally change the nature of asymptotic expansions, and our results do not apply to the unit root tests with nuisance parameters estimated nonparametrically. Our approach developed here can also be used to analyze the models with the local-to-unity formulation of the unit root hypothesis. The asymptotics for such models are quite similar to those for the unit root models, except that they involve Ornstein-Uhlenbeck diffusion process in place of Brownian motion. Their asymptotic expansions can be obtained exactly in the same manner using the probabilistic embedding of Ornstein-Uhlenbeck process.

3.3 Bootstrap Asymptotic Expansions

To develop the asymptotic expansions for the bootstrap statistics F_n^* and G_n^* corresponding to those for F_n and G_n presented in the previous section, we first need a probabilistic embedding of the standardized partial sum of the bootstrap samples (ε_i^*) into a Brownian motion defined on an extended probability space. Once this embedding is done in an appropriately extended probability space, the rest of the procedure to obtain the asymptotic expansions for F_n^* and G_n^* is essentially identical to that for F_n and G_n. Following the usual convention in the bootstrap literature, we use superscript $*$ for the quantities and relationships that are dependent upon the realizations of (ε_i).

Let W be a standard Brownian motion independent of (ε_i),[5] and assume that they are defined on the common probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Of course, there exists a probability space rich enough to support W together with (ε_i), since we assume that they are independent. We then let $(T_i^*)_{i\geq 0}$ be a time change defined on $(\Omega, \mathcal{F}, \mathbf{P})$ such that

$$W(T_i^*/n) =_{d^*} \frac{1}{\sigma_n \sqrt{n}} \sum_{k=1}^{i} \varepsilon_k^* \quad \text{a.s.} \tag{20}$$

where $=_{d^*}$ denotes the equivalence of distribution conditional on a realization of (ε_i). Note that, for each n and for any possible realization of $(\varepsilon_i)_{i=1}^n$, we may find a time change $(T_i^*)_{i=1}^n$ for which (20) holds with the same Brownian motion W. The Brownian motion W therefore is not dependent upon the realizations of (ε_i).

Just as the convention made in Section 3.1, we identify (ε_i^*) only up to their distributional equivalences so that we may assume (ε_i^*) are also defined on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$, and interpret the equality $=_{d^*}$ in conditional distributions as the usual equality in (20). Under the convention, we construct the sequences (δ_i^*) and (η_i^*) from $(W, (T_i^*))$ for each realization of (ε_i), analogously as (δ_i) and (η_i). We also let (ξ_i^*) be given similarly as (ξ_i) for each realization of (ε_i). Clearly, we may alternatively define (δ_i^*, η_i^*) to be the iid samples from the empirical distribution of (δ_i, η_i), which are drawn together with (ε_i^*) from (ε_i). We may thus regard $(\varepsilon_i^*, \delta_i^*, \eta_i^*)$ as the iid samples from the empirical distribution of $(\varepsilon_i, \delta_i, \eta_i)$. To simplify the subsequent exposition, however, we will assume that (δ_i^*, η_i^*) are defined from the embedding (20) of (ε_i^*) given a realization of (ε_i).

Now we define

$$v_i^* = (\varepsilon_i^*/\sigma_n, \delta_i^*, \eta_i^*, \xi_i^{*\prime}/\sigma_n^2)'$$

[5] The Brownian motion W here is, of course, distinct from the one introduced in Sections 3.1 and 3.2. We just use the same notation here to make our results for bootstrap tests more directly comparable to those for asymptotic tests.

and let

$$B_n^*(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} v_i^*$$

as in (10). It may be readily deduced that

Lemma 3.9 Let Assumption 2.1 hold with $r > 4$. Then $B_n^* \to_{d^*} B^*$ a.s., where B^* is a vector Brownian motion with covariance matrix Σ_n given by the sample analogue estimator of Σ defined in Lemma 3.2.

Analogously as for B, we let

$$B^* = (W, V^*, U^*, Z^{*\prime})'$$

and further represent V^*, U^* and Z^* in terms of independent standard Brownian motions W, $W^{\cdot}$ and $W^{\cdot\cdot}$, as in Remark below Lemma 3.2, with the coefficients given by the sample analogue estimators ω_n, ${}^{\cdot}\omega_n$, $\omega_n^{\cdot}$ and $\omega_n^{\cdot\cdot}$, say, of ω, ${}^{\cdot}\omega$, $\omega^{\cdot}$ and $\omega^{\cdot\cdot}$, i.e.,

$$\begin{aligned} U^* &= \omega_n W + {}^{\cdot}\omega_n W^{\cdot}, \\ V^* &= \omega_n W + \omega_n^{\cdot} W^{\cdot} + \omega_n^{\cdot\cdot} W^{\cdot\cdot}. \end{aligned}$$

Moreover, we may write $Z^*(1) = \Gamma_n^{1/2} S$, where Γ_n is the sample analogue estimator of Γ. Note that we may use the same $W^{\cdot}, W^{\cdot\cdot}$ and S for all realizations of (ε_i) to represent V^*, U^* and Z^* as above. Therefore, we may assume that $(W^{\cdot}, W^{\cdot\cdot}, S)$ are defined on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$ as (ε_i) and $(W, (T_i^*))$, and independent of (ε_i) as well as W. We also let (M, N) be defined as earlier, which we may also regard as being independent of (ε_i). Finally, corresponding to θ in (18), we define

$$\theta_n = (\nu, \pi_n, \sigma_n^2, \mu_n^3, \kappa_n^4, \tau_n^4, \Gamma_n) \tag{21}$$

where $\pi_n = 1/\alpha_n(1)$, and $\sigma_n^2, \mu_n^3, \kappa_n^4, \tau_n^4$ and Γ_n are the sample analogue estimators of $\sigma^2, \mu^3, \kappa^4, \tau^4$ and Γ, respectively.

As usual, $\mathbf{P}^*$ and $\mathbf{E}^*$ refer respectively to the probability and expectation operators given a realization of (ε_i). They can be more formally defined as the conditional probability and expectation operators $\mathbf{P}(\,\cdot\,|(\varepsilon_i))$ and $\mathbf{E}(\,\cdot\,|(\varepsilon_i))$ on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ introduced above. For the functionals of $(W, W^{\cdot}, W^{\cdot\cdot}, S)$ and (M, N), however, $\mathbf{P}^*$ and $\mathbf{E}^*$ agree with $\mathbf{P}$ and $\mathbf{E}$ respectively, since they are independent of (ε_i) by construction.

For the subsequent development of our theory, it is convenient to introduce the bootstrap stochastic order symbols. For a sequence of random sequences (X_n) on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$, we let $X_n = o_p^*(1)$ if $\mathbf{P}^*\{|X_n| > \epsilon\} \to_{a.s.} 0$ for any $\epsilon > 0$. Likewise, we denote by $Y_n = O_p^*(1)$ for (Y_n) on $(\Omega, \mathcal{F}, \mathbf{P})$ if, for a.s. all realizations of (ε_i) and for any $\epsilon > 0$, there exists a constant K such that $\mathbf{P}^*\{|Y_n| > K\} \leq \epsilon$. The constant K may vary depending upon the realizations of (ε_i). The symbols $o_p^*(1)$ and $O_p^*(1)$ are the bootstrap versions of the stochastic order symbols $o_p(1)$ and $O_p(1)$. For the random sequences whose distributions are independent of the realizations of (ε_i), the two notions become identical. It is easy to see that $X_n = o_p^*(1)$ if $\mathbf{E}^*|X_n|^s \to_{a.s.} 0$ for some $s > 0$. Moreover, $o_p^*(1)$ and $O_p^*(1)$ satisfy the usual addition and product rules that apply to $o_p(1)$ and $O_p(1)$, as one may easily check. Needless to say, the definitions of $o_p^*(1)$ and $O_p^*(1)$ naturally extend to $o_p^*(a_n)$ and $O_p^*(b_n)$ for some numerical sequences (a_n) and (b_n).

Theorem 3.10 Let Assumption 2.1 hold with $r \geq 8$. Then we have for large n

$$\begin{aligned} F_n^* &= F + F_{nn}(\theta_n) + o_p^*(n^{-1/2}), \\ G_n^* &= G + G_{nn}(\theta_n) + o_p^*(n^{-1/2}), \end{aligned}$$

where F_{nn} and G_{nn} are introduced in (19) and θ_n is defined in (21).

Corollary 3.11 Let Assumption 2.1 hold with $r > 12$. Then we have for large n

$$\begin{aligned} \mathbf{P}^*\{F_n^* \leq x\} &= \mathbf{P}\{{}_2F_n \leq x\} + o(n^{-1/2}) \text{ a.s.}, \\ \mathbf{P}^*\{G_n^* \leq x\} &= \mathbf{P}\{{}_2G_n \leq x\} + o(n^{-1/2}) \text{ a.s.} \end{aligned}$$

uniformly in $x \in \mathbf{R}$, where ${}_2F_n$ and ${}_2G_n$ are defined in (17).

The asymptotics for the bootstrap statistics F_n^* and G_n^* are completely analogous to those for the corresponding statistics F_n and G_n. Theorem 3.10 and Corollary 3.11 are respectively the bootstrap versions of Theorem 3.7 and Corollary 3.8. In Theorem 3.10, the parameters appeared in the asymptotic expansions of the original statistics are replaced by their estimates, as in the bootstrap Edgeworth expansions for the standard stationary models. Due to the law of iterated logarithm for iid sequences, we have for any $\epsilon > 0$

$$\theta_n = \theta + o_p^*(n^{-1/2+\epsilon})$$

under the given moment condition. We may therefore rewrite the results in Theorem 3.10 as

$$\begin{aligned} F_n^* &= F + F_{nn}(\theta) + o_p^*(n^{-1/2}), \\ G_n^* &= G + G_{nn}(\theta) + o_p^*(n^{-1/2}). \end{aligned}$$

Corollary 3.11 shows that these second-order expansions of F_n^* and G_n^* actually provide the refinements of their asymptotic distributions a.s.

Corollaries 3.8 and 3.11 yield under the required moment condition

$$\begin{aligned} \mathbf{P}^*\{F_n^* \leq x\} &= \mathbf{P}\{F_n \leq x\} + o(n^{-1/2}) \text{ a.s.}, \\ \mathbf{P}^*\{G_n^* \leq x\} &= \mathbf{P}\{G_n \leq x\} + o(n^{-1/2}) \text{ a.s.}, \end{aligned}$$

uniformly in $x \in \mathbf{R}$. Now we define a_λ^* and b_λ^* as

$$\mathbf{P}^*\left\{F_n^* \leq a_\lambda^*\right\} = \mathbf{P}^*\left\{G_n^* \leq b_\lambda^*\right\} = \lambda$$

for tests with nominal rejection probability λ. The values a^*_λ and b^*_λ are the bootstrap critical values for the λ-level tests based on the statistics F_n and G_n. Then it follows that

$$\mathbf{P}\left\{F_n \leq a^*_\lambda\right\},\ \mathbf{P}\left\{G_n \leq b^*_\lambda\right\} = \lambda + o(n^{-1/2})$$

for large n. The tests using the bootstrap critical values a^*_λ and b^*_λ thus have rejection probabilities with errors of order $o(n^{-1/2})$.[6]

4. Asymptotics under Local Alternatives

We now consider local alternatives

$$\mathrm{H}_1 : \alpha = 1 - \frac{c}{n} \tag{22}$$

where $c > 0$ is a fixed constant, and let (y_t) be generated as

$$y_t = \alpha y_{t-1} + \sum_{i=1}^{p} \alpha_i \triangle_c y_{t-i} + \varepsilon_t \tag{23}$$

where $\triangle_c = 1 - (1 - c/n)L$ is the quasi-differencing operator. The model given by (22) and (23) is commonly referred to as the *local-to-unity* model, and introduced here to investigate the asymptotic powers of the bootstrap tests.

For the local-to-unity model, it is well known [see, e.g., Stock (1994)] that

$$F_n \to_d F(c) \;=\; -c\left(\int_0^1 W_c(t)^2 dt\right)^{1/2} + \frac{\displaystyle\int_0^1 W_c(t)dW(t)}{\left(\displaystyle\int_0^1 W_c(t)^2 dt\right)^{1/2}}, \tag{24}$$

$$G_n \to_d G(c) \;=\; -c + \frac{\displaystyle\int_0^1 W_c(t)dW(t)}{\displaystyle\int_0^1 W_c(t)^2 dt}, \tag{25}$$

where $W_c(t) = W(t) - c\int_0^t e^{-c(t-s)}W(s)ds$ is Ornstein-Uhlenbeck process, which may be defined as the solution to the stochastic differential equation $dW_c(t) = -cW_c(t)dt + dW(t)$. As is well known,

$$\mathbf{P}\{F(c) \leq x\} > \mathbf{P}\{F \leq x\}, \quad \mathbf{P}\{G(c) \leq x\} > \mathbf{P}\{G \leq x\}, \tag{26}$$

for all $x \in \mathbf{R}$, and we may thus expect that the unit root tests relying on F_n and G_n have some discriminatory powers against the local-to-unity model.

[6] Note that the results here hold only under the assumption that the underlying model is AR(p) with known p and iid errors. For the model driven by more general, possibly conditionally heterogeneous, martingale differences, only the first-order asymptotics are valid. If p is unknown or given as infinity, we may increase p with the sample size and apply the results for the sieve bootstrap established in Park (2002).

The limiting distributions of the bootstrap statistics F_n^* and G_n^* are, however, unaffected, i.e., their limiting distributions under local alternatives are precisely the same as their limiting null distributions. This is shown below in Theorem 4.1. We may indeed expect that the bootstrap samples asymptotically behave as the unit root processes under many other alternatives as well, since they are generated under the unit root restriction regardless of the true data generating mechanism. It is therefore not surprising that the bootstrap statistics F_n^* and G_n^* have the same limiting distributions under both the exact-unit root and local-to-unit root specifications.

Theorem 4.1 Let Assumption 2.1 hold with $r > 2$. Then we have under the local-to-unity model

$$F_n^* \to_{d^*} F \quad \text{a.s.}, \qquad G_n^* \to_{d^*} G \quad \text{a.s.}$$

as $n \to \infty$.

Under the alternative of the local-to-unity model, we have in particular

$$\mathbf{P}\{F \le a_\lambda^*\},\ \mathbf{P}\{G \le b_\lambda^*\} \to \lambda$$

as $n \to \infty$, and therefore,

$$\begin{aligned}\lim_{n\to\infty} \mathbf{P}\{F_n \le a_\lambda^*\} &= \lim_{n\to\infty} \mathbf{P}\{F(c) \le a_\lambda^*\} > \lambda, \\ \lim_{n\to\infty} \mathbf{P}\{G_n \le b_\lambda^*\} &= \lim_{n\to\infty} \mathbf{P}\{G(c) \le b_\lambda^*\} > \lambda\end{aligned}$$

due to (24), (25) and (26). The bootstrap unit root tests would thus have non-trivial powers against the local-to-unity model.

5. Tests in Models with Deterministic Trends

In this section, we investigate the unit root tests in the model

$$y_t = D_t + \alpha y_{t-1} + \sum_{i=1}^{p} \alpha_i \triangle y_{t-i} + \varepsilon_t \tag{27}$$

where D_t is deterministic trend. In what follows, we only explicitly consider D_t specified as

$$D_t = \beta_0 \quad \text{or} \quad \beta_0 + \beta_1 t \tag{28}$$

since they are most frequently used in practical applications. Our theories and methodologies here, however, apply to more general models with higher order polynomials possibly with structural changes, i.e., $D_t = \sum_{i=0}^{q} \beta_i t^i$ or $\sum_{i=0}^{q} \beta_i t^i + \sum_{i=0}^{q} \beta^i t^i \{t \ge s_i\}$, where s_i, $i = 1, \ldots q$, are known break points. We only need some obvious modifications for such models.

We need to consider model (27), instead of (2), when it is believed that the observed time series (y_t) includes deterministic trend D_t and is generated as

$$y_t = D_t + y_t^\circ \tag{29}$$

where the stochastic component (y_t°) is assumed to follow (2). As an alternative to testing for the unit root in regression (27), we may detrend (y_t) directly from the regression given by (29) with (28) to obtain the fitted residuals $(\hat{y}_t^\circ)$, and base the unit root tests on regression (2) using $(\hat{y}_t^\circ)$. It turns out that they are asymptotically equivalent not only in the first order, but also in the second order. All our subsequent results are therefore applicable for both procedures.[7]

To obtain the asymptotic expansions for the Dickey-Fuller tests in the presence of linear time trends, we need the following lemma and the subsequent proposition. We denote by $\imath$ the identity function $\imath(x) = x$ in what follows.

Lemma 5.1 Let Assumption 2.1 hold. Then we have

(a) $$\frac{1}{n^{1/2}\sigma}\sum_{t=1}^{n}\frac{t}{n}\varepsilon_t = J(\imath, W) + n^{-1/4}M(V) - n^{-1/2}\left[WV - J(W,V) - N(V) - \omega\right] + o_p(n^{-1/2}),$$

(b) $$\frac{1}{n^{3/2}\sigma}\sum_{t=1}^{n}\frac{t}{n}w_{t-1} = I(\imath W) + n^{-1/2}\left[WV - I(WV) - J(\imath W, V) - \omega/2\right] + o_p(n^{-1/2}),$$

for large n.

Proposition 5.2 Let Assumption 2.1 hold. Then we have

$$\frac{1}{n^{3/2}\pi\sigma}\sum_{t=1}^{n}\frac{t}{n}y_{t-1} = I(\imath W) + n^{-1/2}\left[WV - I(WV) - J(\imath W, V) + (\nu - \omega/2)\right] + o_p(n^{-1/2})$$

for large n

We now present the asymptotic expansions of the Dickey-Fuller tests for the models with constant, $D_t = \beta_0$, and for the models with linear time trend, $D_t = \beta_0 + \beta_1 t$. They are quite similar, and we present them together in a single framework. For both cases, we denote by $\tilde{F}_n$ and $\tilde{G}_n$ the Dickey-Fuller statistics based on regression (27), or equivalently, the ones defined as in (3) and (4) from the regression (2) run with the demeaned or detrended (y_t). We denote by $\tilde{W}$ the demeaned or detrended Brownian motion, for the case of $D_t = \beta_0$ or $D_t = \beta_0 + \beta_1 t$. Moreover, we let $\tilde{F}$ and $\tilde{G}$ respectively be the functionals of Brownian motions defined similarly as F and G with W replaced by $\tilde{W}$. It is well known that $\tilde{F}_n$ and $\tilde{G}_n$ have the limiting distributions given by $\tilde{F}$ and $\tilde{G}$ respectively. We also define ${}_2\tilde{F}_n$ and ${}_2\tilde{G}_n$ to be the second-order expansions of $\tilde{F}_n$ and $\tilde{G}_n$, similarly as ${}_2F_n$ and ${}_2G_n$ for F_n and G_n.

[7]We do not consider in the paper the GLS detrending proposed by Elliot, Rothenberg and Stock (1996) based on the local-to-unity formulation of the unit root hypothesis. Such detrending in general yields asymptotics that are different from those for the usual OLS detrending considered here.

Theorem 5.3 Let Assumption 2.1 hold with $r \geq 8$. Then we have for large n

$$\tilde{F}_n = \tilde{F} + \tilde{F}_1/\sqrt[4]{n} + \tilde{F}_2/\sqrt{n} + o_p(n^{-1/2}),$$
$$\tilde{G}_n = \tilde{G} + \tilde{G}_1/\sqrt[4]{n} + \tilde{G}_2/\sqrt{n} + o_p(n^{-1/2}),$$

where $\tilde{F}_1 = \tilde{W}M(V)/I(\tilde{W}^2)^{1/2}$, $\tilde{G}_1 = \tilde{W}M(V)/I(\tilde{W}^2)$ and

$$\begin{aligned}\tilde{F}_2 &= \frac{(1/2)M(V)^2 + \tilde{W}N(V) - [1 + J(\tilde{W},\tilde{W})][(V+2U)/2 + \pi\iota'\Gamma^{-1}Z]}{I(\tilde{W}^2)^{1/2}} \\ &\quad - \frac{J(\tilde{W},\tilde{W})[\tilde{W}^2V - J(\tilde{W}^2,V) - 2\omega I(\tilde{W})]}{2I(\tilde{W}^2)^{3/2}}, \\ \tilde{G}_2 &= \frac{(1/2)M(V)^2 + \tilde{W}N(V) - (V+2U)/2 - \pi\iota'\Gamma^{-1}Z}{I(\tilde{W}^2)} \\ &\quad - \frac{J(\tilde{W},\tilde{W})[\tilde{W}^2V - J(\tilde{W}^2,V) - 2\omega I(\tilde{W})]}{I(\tilde{W}^2)^2}.\end{aligned}$$

Moreover, if Assumption 2.1 holds with $r > 12$, then for large n

$$\begin{aligned}\mathbf{P}\{\tilde{F}_n \leq x\} &= \mathbf{P}\{{}_2\tilde{F}_n \leq x\} + o(n^{-1/2}), \\ \mathbf{P}\{\tilde{G}_n \leq x\} &= \mathbf{P}\{{}_2\tilde{G}_n \leq x\} + o(n^{-1/2}),\end{aligned}$$

uniformly in $x \in \mathbf{R}$.

The asymptotic expansions for $\tilde{F}_n$ and $\tilde{G}_n$ in Theorem 5.3 are quite similar to those for F_n and G_n in Theorem 3.7. We only have two differences. First, all of the terms in the expansions for F_n and G_n representing the dependency on the initial value ν disappear, and are not present in the expansions of $\tilde{F}_n$ and $\tilde{G}_n$. This is naturally expected, since the demeaning or detrending makes the statistics $\tilde{F}_n$ and $\tilde{G}_n$ invariant with respect to the initial values. Second, the Brownian motion W is replaced by the demeaned or detrended Brownian motion $\tilde{W}$ in all of the expansion terms. The demeaning or detrending thus affects not only the first-order asymptotics, but also the second-order asymptotics.

Now we define the second-order expansion terms

$$\tilde{F}_{nn} = \tilde{F}_1/\sqrt[4]{n} + \tilde{F}_2/\sqrt{n}, \quad \tilde{G}_{nn} = \tilde{G}_1/\sqrt[4]{n} + \tilde{G}_2/\sqrt{n}$$

for $\tilde{F}_n$ and $\tilde{G}_n$, and let

$$\tilde{F}_{nn}(\theta) = \tilde{F}_{nn}(\theta, (\tilde{W}, W^{\cdot}, W^{\cdot\cdot}, S)), \quad \tilde{G}_{nn}(\theta) = \tilde{G}_{nn}(\theta, (\tilde{W}, W^{\cdot}, W^{\cdot\cdot}, S)), \tag{30}$$

analogously as in (19). Moreover, we let

$${}_2\tilde{F}_n = \tilde{F} + \tilde{F}_{nn}, \quad {}_2\tilde{G}_n = \tilde{G} + \tilde{G}_{nn}, \tag{31}$$

be the second-order approximations of $\tilde{F}_n$ and $\tilde{G}_n$ correspondingly to (17).

Theorem 5.4 Let Assumption 2.1 hold with $r \geq 8$. Then we have for large n

$$\begin{aligned} \tilde{F}_n^* &= \tilde{F} + \tilde{F}_{nn}(\theta_n) + o_p^*(n^{-1/2}), \\ \tilde{G}_n^* &= \tilde{G} + \tilde{G}_{nn}(\theta_n) + o_p^*(n^{-1/2}), \end{aligned}$$

where $\tilde{F}_{nn}$ and $\tilde{G}_{nn}$ are introduced in (30) and θ_n is the sample moment estimator of the parameter θ defined in (18). Moreover, if Assumption 2.1 holds with $r > 12$, then for large n

$$\begin{aligned} \mathbf{P}^*\{\tilde{F}_n^* \leq x\} &= \mathbf{P}\{{}_2\tilde{F}_n \leq x\} + o(n^{-1/2}) \text{ a.s.}, \\ \mathbf{P}^*\{\tilde{G}_n^* \leq x\} &= \mathbf{P}\{{}_2\tilde{G}_n \leq x\} + o(n^{-1/2}) \text{ a.s.}, \end{aligned}$$

uniformly in $x \in \mathbf{R}$, where ${}_2\tilde{F}_n$ and ${}_2\tilde{G}_n$ are defined in (31).

The results in Theorem 5.4 make it clear that the main conclusions on the asymptotic refinements of the bootstraps in Section 3.3 continue to hold for the tests in models with constant and linear trends. Using bootstrap critical values would reduce the finite sample distortion in rejection probability to the order $o(n^{-1/2})$ also in models with such deterministic trends.

6. Monte Carlo Simulations

We perform Monte Carlo simulations to investigate the actual finite sample performances of the bootstrap tests. The model we use for the simulations is specified as

$$y_t = \alpha y_{t-1} + \beta \triangle y_{t-1} + \varepsilon_t$$

where α and β are parameters and (ε_t) are iid innovations. The parameter values are chosen to be $\alpha = 1$ and $\beta = 0.4, 0.0, -0.4$. We set $\alpha = 1$ and investigate only the finite sample sizes of the tests.[8] The innovations are generated as standard normal $\mathbf{N}(0,1)$, normal-mixture $\mathbf{N}(0,1)$ and $\mathbf{N}(0,16)$ with mixing probabilities 0.8 and 0.2, and shifted chi-square $\chi^2(8) - 8$ distributions.[9] We thus consider both normal and non-normal innovations, and for the non-normal innovations we look at skewed ones as well as those that are not skewed. The samples of sizes $n = 25, 50, 100$ are generated. The rejection probabilities for the tests with fitted mean and time trend are given respectively in Tables 1 and 2. The nominal rejection probabilities of the test are 5%.

The simulation results reported in Tables 1 and 2 are generally supportive of the theory developed in the paper. In particular, they make it clear that the bootstrap does provide asymptotic refinements for the tests of a unit root in finite samples. In all cases that we investigate here, the bootstrap tests, i.e., the tests based on the critical values computed by

[8] We also looked at the finite sample powers of the tests for various values of α. The bootstrap tests have essentially the same powers as the asymptotic tests. This confirms the findings by Nankervis and Savin (1996).

[9] To make our results more comparable to theirs, we look at the distributions considered by Nankervis and Savin (1996). However, we do not follow them in standardizing the distributions to have unit variance, since the unit root tests considered here are invariant with respect to scaling.

the bootstrap sampling, have rejection probabilities that are closer to their nominal values, if compared with the usual asymptotic tests. The actual magnitudes of the refinements, however, depend upon various factors such as the sample size, the model specification and the distribution of innovations. Overall, it appears that the bootstrap offers more significant refinements for small samples and for models with fitted time trend, respectively in terms of the sample size and the model specification. The distribution of innovations seems to have only minor effects. Our simulation results are largely comparable to those obtained earlier by Nankervis and Savin (1996).

Indeed, it can be seen clearly from Tables 1 and 2 that the bootstrap correction in finite samples is highly effective for the tests of a unit root. The rejection probabilities of the bootstrap tests are quite close to their nominal values regardless of the sample size, the model specification and the distribution of innovations. The discrepancies between the actual and nominal rejection probabilities never exceed more than 0.5% in most cases. This is in contrast with the asymptotic tests. For the asymptotic tests, the actual rejection probabilities are larger than 10% in several cases for the 5% tests. It seems clear that the use of asymptotic critical values can seriously distort the test results in finite samples, and that the bootstrap provides an effective tool to prevent such a distortion. Our simulations suggest that the bootstrap correction is needed more for the tests using smaller samples and based on models with maintained time trend. The asymptotics provide poor approximations especially when the sample size is small and the model includes a maintained time trend.

7. Conclusion

In the paper, we develop asymptotic expansions for the unit root models and show that the bootstrap provides asymptotic refinements for the unit root tests. It is demonstrated through simulations that the bootstrap indeed offers asymptotic refinements in finite samples and the bootstrap corrections are in general quite effective in eliminating finite sample biases of the test statistics. Though we consider exclusively the Dickey-Fuller tests, it is made clear that our results are applicable for other unit root tests as well. Our methodology here can also be extended to analyze the bootstrap for more general models, nonlinear as well as linear, with integrated time series and near-integrated time series. This will be reported in future work.

8. Mathematical Proofs

We first present some useful lemmas and their proofs. They will be used in the proofs of the main results in the text, which will follow subsequently. Throughout the proof, $|\cdot|$ denotes the Euclidian norm, and K signifies a generic constant depending possibly only upon r, which may vary from place to place.

8.1 Useful Lemmas and Their Proofs

We write

$$\frac{\varepsilon_i}{\sigma\sqrt{n}} = \int_{T_{n,i-1}}^{T_{ni}} dW(t) = W(T_{ni}) - W(T_{n,i-1}).$$

Then it follows from Ito's formula that

$$\left(\frac{\varepsilon_i}{\sigma\sqrt{n}}\right)^{k+2} = (k+2)\int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]^{k+1} dW(t) + \frac{(k+1)(k+2)}{2}\int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]^{k} dt \tag{32}$$

for $k \geq 0$. Consequently, we have

Lemma A1 Let Assumption 2.1 hold with $r \geq 2$. We have

(a) $\varepsilon_i^2/\sigma^2 - 1 = \delta_i + 2\eta_i$.

Moreover, it follows that

(b) $\mathbf{E}\int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]^{k} dt = \frac{2}{(k+1)(k+2)}\left(\frac{1}{\sigma\sqrt{n}}\right)^{k+2} \mathbf{E}\,\varepsilon_i^{k+2}$

for any integer $k \geq 0$ such that $k \leq r-2$.

Proof of Lemma A1 The result in part (a) may easily be deduced from

$$\left(\frac{\varepsilon_i}{\sigma\sqrt{n}}\right)^{2} = 2\int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})] dW(t) + (T_{ni} - T_{n,i-1}),$$

which follows from Ito's formula (32) with $k = 0$. To derive part (b), we rewrite the Ito's formula (32) as

$$\int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]^{k} dt = \frac{2}{(k+1)(k+2)}\left(\frac{\varepsilon_i}{\sigma\sqrt{n}}\right)^{k+2} - \frac{2}{k+1}\int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]^{k+1} dW(t).$$

The stated result follows immediately upon noticing that

$$\int_{T_{n,i-1}}^{\cdot} [W(s) - W(T_{n,i-1})]^{k+1} dW(s)$$

is a martingale, and therefore

$$\mathbf{E}\int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]^{k+1} dW(t) = 0$$

due to the optional stopping theorem. □

Let

$$u_t = \varphi(L)\varepsilon_t = \sum_{i=0}^{\infty} \varphi_i \varepsilon_{t-i}$$

and let $\upsilon^2 = \sum_{i=0}^{\infty} \varphi_i^2$. Also, we define $\Gamma_{ij} = \mathbf{E} u_{t-i} u_{t-j}$, so that we have in particular $\Gamma_0 = \sigma^2 \upsilon^2$.

Lemma A2 If Assumption 2.1 holds with $r \geq 2$, then

(a) $\mathbf{E}|u_i|^r \leq \upsilon^r \mathbf{E}|\varepsilon_i|^r$,

(b) $\mathbf{E}|\delta_i|^{r/2} \leq K(1 + \mathbf{E}|\varepsilon_i|^r)$, $\mathbf{E}|1 + \delta_i|^{r/2} \leq K \mathbf{E}|\varepsilon_i|^r$,

(c) $\mathbf{E}|\eta_i|^{r/2} \leq K(1 + \sigma^{-r})\mathbf{E}|\varepsilon_i|^r$.

If Assumption 2.1 holds with $r \geq 4$, then

(d) $\mathbf{E}\left|\sum_{k=1}^{n}(1+\delta_k)u_{k-i}\right|^{r/2} \leq n^{r/4}\left(1+\upsilon^r\right) K \left[1 + \left(\mathbf{E}|\varepsilon_i|^r\right)^2\right]$,

(e) $\mathbf{E}\left|\sum_{k=1}^{n}\delta_k u_{k-i} u_{k-j}\right|^{r/2} \leq n^{r/4}\upsilon^r K \left[1 + \left(\mathbf{E}|\varepsilon_i|^r\right)^2\right]$,

(f) $\mathbf{E}\left|\sum_{k=1}^{n}(u_{k-i}u_{k-j} - \Gamma_{ij})\right|^{r/2} \leq n^{r/4}\upsilon^r K \left(\sigma^r + \mathbf{E}|\varepsilon_i|^r\right)$,

for all $i, j = 1, \ldots, p$.

Proof of Lemma A2 Part (a) is well known. Part (b) is due to Lemma 3.1. To prove part (c), use part (a) of Lemma A1 and Minkowski's inequality to deduce

$$\mathbf{E}|\eta_i|^{r/2} \leq K\left(\mathbf{E}|1+\delta_i|^{r/2} + \sigma^{-r}\mathbf{E}|\varepsilon_i|^r\right)$$

from which and Lemma 3.1 the stated result readily follows. Given parts (a) and (b), parts (d) and (e) can easily be deduced from the successive applications of Burkholder's inequality [see, e.g., Hall and Heyde (1980, Theorem 2.10)] and Minkowski's inequality. Indeed we have

$$\mathbf{E}\left|\sum_{k=1}^{n}(1+\delta_k)u_{k-i}\right|^{r/2} \leq K\mathbf{E}\left|\sum_{k=1}^{n}(1+\delta_k)^2 u_{k-i}^2\right|^{r/4} \leq n^{r/4} K \mathbf{E}|1+\delta_i|^{r/2}\mathbf{E}|u_i|^{r/2}$$

and part (d) follows immediately, due to parts (a) and (b). Note that $\upsilon^{r/2} \leq 1 + \upsilon^r$ and $\mathbf{E}|\varepsilon_i|^{r/2}, \mathbf{E}|\varepsilon_i|^r \leq 1 + (\mathbf{E}|\varepsilon_i|^r)^2$. The proof for part (e) is entirely analogous. For part (f), we write

$$\sum_{k=1}^{n}(u_{k-i}u_{k-j} - \Gamma_{ij}) = \sum_{k=1}^{n}\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\varphi_p\varphi_q(\varepsilon_{k-i-p}\varepsilon_{k-j-q} - \sigma^2\delta_{i+p,j+q})$$

where δ_{ij} is the usual Kronecker delta, i.e., $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. The stated result can then be easily obtained as above by applying Burkholder's and Minkowski's inequalities successively. $\square$

Our asymptotic expansions rely on a strong approximation of B_n by B. It involves extending the underlying probability space to redefine B_n, without changing its distribution, on the same probability space as B, and provides an explicit rate for the convergence of B_n to B. The relevant theory will be developed in the lemma given below. Following our earlier convention, we will not distinguish B_n from its distributionally equivalent copy defined, together with B, in the newly extended probability space. We let

$$B_n = (A_n', Z_n')', \quad B = (A', Z')' \tag{33}$$

where $A = (W, V, U)'$ and $A_n = (W_n, V_n, U_n)'$ is defined conformably with A, i.e., W_n, V_n and U_n are the partial sum processes that a.s. converge respectively to W, V and U.

Lemma A3 If Assumption 2.1 holds with $r > 4$, then we may choose A_n and Z_n jointly such that

$$\mathbf{P}\left\{\sup_{0\le t\le 1} |A_n(t) - A(t)| > c\right\} \le n^{1-r/4} c^{-r/2} (1 + \sigma^{-r}) K (1 + \mathbf{E}|\varepsilon_i|^r)$$

for any $c \ge n^{-1/2+2/r}$, and

$$\mathbf{P}\left\{\sup_{0\le t\le 1} |Z_n(t) - Z(t)| > c\right\} \le n^{-r/4} c^{-r} (1 + \upsilon^r) K \left[1 + (\mathbf{E}|\varepsilon_i|^r)^2\right]$$

for any $c \ge n^{-1/4}$.

Proof of Lemma A3 The strong approximation by Courbot (2001) for general multidimensional continuous time martingales is most directly applicable here, but his result only provides the convergence rate that is far from optimal and depends also upon the dimensionality parameter. Therefore, we will develop a more direct embedding for the martingale difference sequence (ξ_i), and subsequently use the strong approximation by Einmahl (1987a, 1987b, 1989) for the iid random vectors $(\varepsilon_i, \delta_i, \eta_i)$. The first step embedding for (ξ_i) only introduces a limit process independent of W, and therefore, does not interfere with the second step embedding for $(\varepsilon_i, \delta_i, \eta_i)$, which are determined solely by W. On the other hand, the distributions of $(\varepsilon_i, \delta_i, \eta_i)$ fully specify those of $(\varepsilon_i, \delta_i, \eta_i, \xi_i)$, since the values of (ε_i) completely specify (ξ_i). The second step embedding would therefore provide the desired strong approximation for $(\varepsilon_i, \delta_i, \eta_i, \xi_i)$.

Let

$$C_n(t) = \frac{1}{\sigma} \sum_{i=1}^{n} x_{i-1} 1\{T_{n,i-1} \le t < T_{ni}\}$$

and define a continuous martingale

$$Z_n^{\cdot}(t) = \int_0^t C_n(s) dW(s)$$

for $0 \le T_{nn}$. Notice that

$$Z_n(t) = Z_n^{\cdot}(T_{n,i-1})$$

for $(i-1)/n \le t < i/n$. It follows that the quadratic variation $[Z_n^{\cdot}]$ of $Z_n^{\cdot}$ is given by

$$[Z_n^{\cdot}](t) = \int_0^t C_n(s)\, C_n'(s)\, ds = \frac{1}{n\sigma^2} \sum_{i=1}^n x_{i-1} x_{i-1}' 1\{t \le T_{n,i-1}\} + R_n^a(t)$$

where

$$R_n^a(t) = \frac{1}{n\sigma^2} \sum_{i=1}^n \delta_i x_{i-1} x_{i-1}' 1\{t \le T_{n,i-1}\} + \sum_{i=1}^n (t - T_{n,i-1}) 1\{T_{n,i-1} \le t < T_{ni}\}$$

for $0 \le T_{nn}$. Moreover, the quadratic covariation $[W, Z_n^{\cdot}]$ of $Z_n^{\cdot}$ with W becomes

$$[W, Z_n^{\cdot}](t) = \int_0^t C_n(s)\, ds = \frac{1}{n\sigma} \sum_{i=1}^n x_{i-1} 1\{t \le T_{n,i-1}\} + R_n^b(t)$$

where

$$R_n^b(t) = \frac{1}{n\sigma} \sum_{i=1}^n \delta_i x_{i-1} 1\{t \le T_{n,i-1}\} + \sum_{i=1}^n (t - T_{n,i-1}) 1\{T_{n,i-1} \le t < T_{ni}\}$$

for $0 \le T_{nn}$.

We now embed the continuous martingale $Z_n^{\cdot}$, up to a negligible error, into a vector Brownian motion independent of W. Using the representation of the continuous martingale $(W, Z_n^{\cdot\prime})'$ as a stochastic integral with respect to Brownian motion [see, e.g., Theorem 3.9 in Revuz and Yor (1994, p175)], we may have

$$Z_n^{\cdot}(t) = \Gamma^{1/2} W^{\cdot}(t) + R_n(t) \tag{34}$$

where $W^{\cdot}$ is a vector Brownian motion independent of W, and R_n is majorized in probability by

$$\left(\sup_{t \le T_{nn}} |[Z_n^{\cdot}](t) - t\Gamma| \right)^{1/2} + \left(\sup_{t \le T_{nn}} |[W, Z_n^{\cdot}](t)| \right)^{1/2}.$$

Note that we may use a block lower triangular predictable process to represent $(W, Z_n^{\cdot\prime})'$ as a Brownian stochastic integral. The representation (34) for $Z_n^{\cdot}$ is thus possible without changing W. However, we have

$$\mathbf{P}\left\{ \max_{1 \le i \le n} \frac{1}{n} |\delta_i| > c \right\} \le n^{1-r/2} c^{-r/2} K \left(1 + \mathbf{E}|\varepsilon_i|^r\right),$$

$$\mathbf{P}\left\{ \max_{1 \le i \le n} \left| \frac{1}{n} \sum_{k=1}^i \delta_k x_{k-1} x_{k-1}' \right| > c \right\} \le n^{-r/4} c^{-r/2} \upsilon^r K \left[1 + (\mathbf{E}|\varepsilon_i|^r)^2\right],$$

$$\mathbf{P}\left\{ \max_{1 \le i \le n} \left| \frac{1}{n} \sum_{k=1}^i (1 + \delta_k) x_{k-1} \right| > c \right\} \le n^{-r/4} c^{-r/2} (1 + \upsilon^r) K \left[1 + (\mathbf{E}|\varepsilon_i|^r)^2\right],$$

due to Markov's inequality, Lemma A2 and the martingale maximal inequality [see, e.g., Hall and Heyde (1980, Theorem 2.1, p14)]. Therefore, it follows that $\sup_{t\le T_{nn}}|R_n(t)| = o_p(1)$, if $r > 4$ as assumed. Moreover, we have

$$\mathbf{P}\left\{\sup_{t\le T_{nn}}|R_n(t)|>c\right\}\le n^{-r/4}c^{-r}K(1+\upsilon^r)\left[1+(\mathbf{E}|\varepsilon_i|^r)^2\right] \tag{35}$$

which, in particular, implies that $\sup_{t\le T_{nn}}|R_n(t)| = O_p(n^{-1/4})$.

We let

$$Z = \Gamma^{1/2}W^{\cdot}$$

and write for $(i-1)/n \le t < i/n$

$$\begin{aligned} Z_n(t)-Z(t) &= Z_n^{\cdot}(T_{n,i-1})-Z(t) \\ &= [Z_n^{\cdot}(T_{n,i-1})-Z(T_{n,i-1})]-[Z(t)-Z(T_{n,i-1})] \\ &= R_n(T_{n,i-1})-[Z(t)-Z(T_{n,i-1})] \end{aligned} \tag{36}$$

The first term is bounded as shown in (35) above. To obtain the bound for the second term, we first note that

$$|Z(t)-Z(T_{n,i-1})|\le|t-T_{n,i-1}|^{1/2-\epsilon}$$

for any $\epsilon > 0$, due to the Hölder continuity of the Brownian motion sample path, and for $(i-1)/n \le t < i/n$

$$|t-T_{n,i-1}|\le\frac{1}{n}+\left|\frac{1}{n}\sum_{k=1}^{i-1}\delta_k\right|$$

as we may easily deduce by considering three cases $T_{n,i-1} < (i-1)/n$, $(i-1)/n \le T_{n,i-1} < i/n$ and $T_{n,i-1} \ge i/n$ separately. However, it follows from Markov's inequality, Lemma A2 and the martingale maximal inequality that

$$\mathbf{P}\left\{\max_{1\le i\le n}\left|\frac{1}{n}\sum_{k=1}^{i}\delta_k\right|>c\right\}\le n^{1-r/2}c^{-r/2}K(1+\mathbf{E}|\varepsilon_i|^r)$$

and consequently,

$$\mathbf{P}\left\{\max_{1\le i\le n}\sup_{(i-1)/n\le t<i/n}|Z(t)-Z(T_{n,i-1})|>c\right\}\le n^{1-r/2+\epsilon}(c+n^{-1/2})^{-r}K(1+\mathbf{E}|\varepsilon_i|^r). \tag{37}$$

The stated result for Z_n now follows from (35) and (37) using (36) and noticing that $n^{1-r/2+\epsilon} = o(n^{-r/4})$ if $r > 4$ and $\epsilon > 0$ is sufficiently small.

Recall that Z is independent of W, and hence of $(\varepsilon_i, \delta_i, \eta_i)$, which are all defined from W. Therefore, we may embed A_n into a vector Brownian motion A defined in the same probability space as Z. According to Einmahl (1987a, 1989), we may have up to distributional equivalence

$$\max_{1\le i\le n}|A_n(i/n)-A(i/n)| = o(n^{-1/2+2/r}) \text{ a.s.}$$

if $\mathbf{E}|\varepsilon_i|^{r/2}, \mathbf{E}|\delta_i|^{r/2}, \mathbf{E}|\eta_i|^{r/2} < \infty$ for some $r > 4$. It also follows that

$$\sup_{0\le t\le 1} |A_n(t) - A(t)| = o(n^{-1/2+2/r}) \text{ a.s.}$$

since by the uniform continuity of the Brownian motion sample path

$$\sup_{|t-s|\le 1/n} |A(t) - A(s)| \le K(2\log n/n)^{1/2} \text{ a.s.}$$

for any constant $K > 1$ [see, e.g., Hida (1980, Theorem 2.6)]. We therefore have

$$\sup_{0\le t\le 1} |A_n(t) - A(t)| = o(n^{-1/2+r/2}) \text{ a.s.}$$

as long as Assumption 2.1 holds for some $r > 4$. We may also directly deduce from Einmahl (1987b) that

$$\mathbf{P}\left\{\sup_{0\le t\le 1} |A_n(t) - A(t)| > c\right\} \le n^{1-r/4}c^{-r/2}K\left(\mathbf{E}|\varepsilon_i|^{r/2} + \mathbf{E}|\delta_i|^{r/2} + \mathbf{E}|\eta_i|^{r/2}\right)$$

for any $c \ge n^{-1/2+2/r}$, from which and Lemma A2 the stated result for A_n follows immediately. To complete the proof, note that (ξ_i) is completely determined by (ε_i), and therefore, Z_n can obviously be defined in the same probability space as A_n and A. $\square$

Let R_n be a random sequence. We say that R_n is *distributionally* of order $o(n^{-a})$ for some $a > 0$, if and only if

$$\mathbf{P}\left\{|R_n| > n^{-a-\epsilon}\right\} = o(n^{-a})$$

for some $\epsilon > 0$. We may readily deduce various properties of distributional orders defined as such. In particular, it can be easily shown that the sums of random sequences that are distributionally of order $o(n^{-a})$ become distributionally of order $o(n^{-a})$ for any $a > 0$. The following lemma gives the motivation for the definition and some useful results for the distributional orders.

Lemma A4 Let R_n be distributionally of order $o(n^{-a})$ for some $a > 0$.
(a) If $P_n = Q_n + R_n$ and Q_n has density bounded uniformly in n, we have

$$\mathbf{P}\{P_n \le x\} = \mathbf{P}\{Q_n \le x\} + o(n^{-a})$$

uniformly in $x \in \mathbf{R}$.
(b) If S_n has moments finite up to any order and bounded uniformly in n, then R_nS_n is also distributionally of order $o(n^{-a})$.
(c) Let $a > b > 0$, and let $S_n = n^{-b}T_n$. If T_n has finite $(a/b + \epsilon)$-th moment bounded uniformly in n for some $\epsilon > 0$, then R_nS_n is distributionally of order $o(n^{-a})$.
(d) If S_n is distributionally of order $o(n^{-a})$, then so is R_nS_n.
(e) If $P_n = Q_n + R_n$ and Q_n^{-1} has moments finite up to any order and bounded uniformly in n, then we have $P_n^{-1} = Q_n^{-1} + S_n$ with S_n distributionally of order $o(n^{-a})$.

Proof of Lemma A4 For the proof of part (a), we note that

$$\mathbf{P}\{P_n \le x\} \le \mathbf{P}\{Q_n \le x + n^{-a-\epsilon}\} + \mathbf{P}\{|R_n| > n^{-a-\epsilon}\}$$

and

$$\left|\mathbf{P}\{Q_n \le x + n^{-a-\epsilon}\} - \mathbf{P}\{Q_n \le x\}\right| \le K n^{-a-\epsilon}$$

where K is a constant, which majorizes the densities of (Q_n). It therefore follows that $\mathbf{P}\{P_n \le x\} \le \mathbf{P}\{Q_n \le x\} + o(n^{-a})$. A similar argument can be used to show that the opposite inequality, i.e., $\mathbf{P}\{P_n \le x\} \ge \mathbf{P}\{Q_n \le x\} + o(n^{-a})$, also holds true. The result in part (a) is thus established. To prove part (b), we first observe that

$$\mathbf{P}\left\{|R_n S_n| > n^{-a-\epsilon}\right\} \le \mathbf{P}\left\{|R_n| > n^{-a-2\epsilon}\right\} + \mathbf{P}\left\{|S_n| > n^{\epsilon}\right\}$$

and that

$$\mathbf{P}\left\{|S_n| > n^{\epsilon}\right\} = O(n^{-a-\epsilon}) = o(n^{-a})$$

which is due to Markov's inequality, since, in particular, the $(a+\epsilon)/\epsilon$-th moment of S_n exists for any $\epsilon > 0$ and is bounded uniformly in n.

The proof of part (c) is just as easy. The stated result can be easily deduced from that

$$\begin{aligned}\mathbf{P}\left\{|R_n S_n| > n^{-a-\epsilon}\right\} &= \mathbf{P}\left\{n^{-b}|R_n T_n| > n^{-a-\epsilon}\right\} \\ &\le \mathbf{P}\left\{|R_n| > n^{-a-\epsilon}\right\} + \mathbf{P}\left\{|T_n| > n^{b}\right\}\end{aligned}$$

and that

$$\mathbf{P}\left\{|T_n| > n^{b}\right\} \le n^{-a-\epsilon}\mathbf{E}|T_n|^{a/b+\epsilon} = O(n^{-a-\epsilon}) = o(n^{-a}),$$

which holds for some $\epsilon > 0$. For part (d), note that

$$\mathbf{P}\left\{|R_n S_n| > n^{-a-\epsilon}\right\} \le \mathbf{P}\left\{|R_n| > n^{-a-\epsilon}\right\} + \mathbf{P}\left\{|S_n| > n^{-a-\epsilon}\right\},$$

from which the stated result follows immediately.

To prove part (f), we write $P_n = Q_n(1+Q_n^{-1}R_n)$ so that $P_n^{-1} = Q_n^{-1}(1+Q_n^{-1}R_n)^{-1}$. Due to part (b), multiplication by Q_n^{-1} does not change the distributional order of the residual under the given condition. We may therefore set w.l.o.g. that $Q_n = 1$, and consider

$$\frac{1}{1+R_n} = 1 - R_n + \frac{R_n^2}{1+R_n}.$$

Now it suffices to show that $R_n^2/(1+R_n)$ is distributionally of order $o(n^{-a})$. This, however, is rather straightforward. If $|R_n| \le n^{-a-\epsilon}$, we have

$$R_n^2/(1+R_n) \le (n^{-a-\epsilon})^2/(1-n^{-a-\epsilon})$$

and, when n is large enough so that $n^{-a-\epsilon} \le 1/2$,

$$(n^{-a-\epsilon})^2/(1-n^{-a-\epsilon}) \le n^{-a-\epsilon}.$$

Thus it follows that

$$\mathbf{P}\left\{\left|\frac{R_n^2}{1+R_n}\right| > n^{-a-\epsilon}\right\} \le \mathbf{P}\{|R_n| > n^{-a-\epsilon}\}$$

for all n sufficiently large. The proof is therefore complete. □

8.2 Proofs of the Main Results

Proof of Lemma 3.1 See Hall and Heyde (1980, Theorem A.1, p269). □

Proof of Lemma 3.2 That $B_n \to_d B$ directly follows from an invariance principle for martingale difference sequences [see Hall and Heyde (1980, p99)]. The covariance matrix of B can be obtained using the results in Lemma A1, and the orthogonality of (ξ_i) and $(\varepsilon_i, \delta_i, \eta_i)$. Since we have

$$\begin{aligned}\frac{1}{n^{3/2}\sigma}\mathbf{E}\,\varepsilon_i\eta_i &= \mathbf{E}\int_{T_{n,i-1}}^{T_{ni}}[W(t) - W(T_{n,i-1})]dt,\\ \frac{1}{n^2}\mathbf{E}\,\eta_i^2 &= \mathbf{E}\int_{T_{n,i-1}}^{T_{ni}}[W(t) - W(T_{n,i-1})]^2dt,\end{aligned}$$

it follows immediately from part (b) of Lemma A1 that $(1/\sigma)\mathbf{E}\,\varepsilon_i\eta_i = \mu^3/3\sigma^3$ and $\mathbf{E}\,\eta_i^2 = \kappa^4/6\sigma^4$. Furthermore, we may easily get $(1/\sigma)\mathbf{E}\,\varepsilon_i\delta_i$ and $\mathbf{E}\,\delta_i\eta_i$ using the relationship in part (a) of Lemma A1. In fact, we may multiply both sides of the relationship by ε_i/σ and take the expectation and utilize $(1/\sigma)\mathbf{E}\,\varepsilon_i\eta_i = \mu^3/3\sigma^3$ to deduce $(1/\sigma)\mathbf{E}\,\varepsilon_i\delta_i = \mu^3/3\sigma^3$. Moreover, squaring both sides of the relationship and taking expectation yield $\mathbf{E}\,\delta_i\eta_i = (\kappa^4 - 3\sigma^4 - 3\tau^4)/12\sigma^4$. Finally, note that (ξ_i) is a martingale difference sequence, which is uncorrelated with $(\varepsilon_i, \delta_i, \eta_i)$ at all leads and lags, and $(1/\sigma^4)\mathbf{E}\xi\xi' = \Gamma$. □

Proof of Lemma 3.3 Part (a) follows directly from part (a) of Lemma A1 and the discussion following Lemma 3.2. Part (b) is also immediate from Lemma 3.2 and the subsequent remark. Note that we may have stronger results for parts (a) and (b) using the strong approximations in Lemma A3. Part (c) can be easily deduced from part (f) of Lemma A2. □

Proof of Proposition 3.4 Given (13) and (14), both parts (a) and (b) readily follow from Lemma 3.3. □

Proof of Lemma 3.5 We subsequently prove each of parts (a) – (d). Here and elsewhere in the proofs, we use $\epsilon > 0$ to denote any arbitrarily small number. The value of ϵ may vary from line to line. Let V_n and U_n be the partial sum processes of (δ_i) and (η_i), respectively. Therefore, we have in particular

$$V_n(1) = n^{1/2}(T_{nn} - 1).$$

We let $n_i = i/n$ in the subsequent proofs.

Proof of Part (a) We write

$$\frac{1}{n^{1/2}\sigma}\sum_{t=1}^{n}\varepsilon_t = W(T_{nn}) = W(1) + [W(T_{nn}) - W(1)]$$

and let

$$D_n = n^{1/4}[W(T_{nn}) - W(1)] = n^{1/4}\left[W(1 + n^{-1/2}V_n(1)) - W(1)\right]. \tag{38}$$

To obtain the stated result, it now suffices to show that

$$n^{-1/4}D_n = n^{-1/4}M[V(1)] + n^{-1/2}N[V(1)] + o(n^{-3/4+2/r}) \text{ a.s.,} \tag{39}$$

which we set out to do.

Define

$$\begin{aligned} N^{\cdot}(t) &= [W(1+t) - W(1)]1\{t \geq 0\}, \\ N^{\cdot\cdot}(t) &= -[W(1) - W(1-t)]1\{t \geq 0\}, \end{aligned}$$

so that N can be written

$$N(t) = N^{\cdot}(t)1\{t \geq 0\} + N^{\cdot\cdot}(-t)1\{t \leq 0\}.$$

Moreover, we let

$$M_n(t) = M_n^{\cdot}(t)1\{t \geq 0\} + M_n^{\cdot\cdot}(-t)1\{t \leq 0\}$$

where $M_n^\circ = M_n^{\cdot}$ and $M_n^{\cdot\cdot}$ are defined by

$$\begin{aligned} M_n^\circ(t) &= (1 - n^{-1/2})^{-1/2}\left[n^{1/4}N^\circ(n^{-1/2}t) - n^{-1/4}N^\circ(t)\right] \\ &= n^{1/4}N^\circ(n^{-1/2}t) - n^{-1/4}N^\circ(t) + O_p(n^{-1/2}) \end{aligned}$$

respectively from $N^\circ = N^{\cdot}$ and $N^{\cdot\cdot}$. Note that $n^{1/4}N^\circ(n^{-1/2}t), N^\circ(t) = O_p(1)$ uniformly on any compact interval for $N^\circ = N^{\cdot}$ and $N^{\cdot\cdot}$, and $(1 - n^{-1/2})^{-1/2} = 1 + O(n^{-1/2})$.

Now we may write D_n introduced in (38) as

$$D_n = M_n[V_n(1)] + n^{-1/4}N[V_n(1)] + O_p(n^{-1/2}). \tag{40}$$

To establish (39), we first show that M_n can be written as M for every n. The processes $M_n^{\cdot}$ and $M_n^{\cdot\cdot}$ are continuous martingales with quadratic variations

$$[M_n^{\cdot}](t) = [M_n^{\cdot\cdot}](t) = t$$

for all n. Therefore, due to Levy's characterization theorem [see, e.g., Revuz and Yor (1994, pp142-143)], they are standard Brownian motions. Moreover, their quadratic variation vanishes, i.e.,

$$[M_n^{\cdot}, M_n^{\cdot\cdot}](t) = 0$$

for all n. By the Knight's theorem [see, e.g., Revuz and Yor (1994, p175)], therefore, they are standard bivariate Brownian motion for all n. Since the distribution of $(M_n^{\cdot}, M_n^{\cdot\cdot})$ is independent of n, we may designate it as $(M^{\cdot}, M^{\cdot\cdot})$. Accordingly, we also write M instead of M_n in (40).

We now show that $(M^{\cdot}, M^{\cdot\cdot})$ is independent of W on $\mathbf{R}$. It is clear that $M^{\cdot}$ is independent of W on $[0, 1]$, since $N^{\cdot}$ and W on $[0, 1]$ are independent, due to the independent increment property of Brownian motion. To show that $M^{\cdot}$ is independent of W on $(1,\infty)$, we only need to establish the independence of $M^{\cdot}$ and $N^{\cdot}$ since for all $t > 1$

$$W(t) = W(1) + N^{\cdot}(s)$$

with $s = t - 1$. The independence of $M^{\cdot}$ and $N^{\cdot}$, however, follows immediately from that they are Brownian motions and that

$$[M^{\cdot}, N^{\cdot}](t) = 0.$$

We thus have shown that $M^{\cdot}$ is independent of W on $\mathbf{R}$.

The independence of $M^{\cdot\cdot}$ and W on $\mathbf{R}$ can be deduced similarly. Since

$$[M^{\cdot\cdot}, N^{\cdot\cdot}](t) = 0$$

by construction, $M^{\cdot\cdot}$ is independent of $N^{\cdot\cdot}$ (and $W(1)$, in particular). However, we have for all $t \in [0,1]$

$$W(t) = W(1) - N^{\cdot\cdot}(s)$$

with $s = 1 - t$, and therefore, $M^{\cdot\cdot}$ is independent of W on $[0,1]$. The independence also holds between $M^{\cdot\cdot}$ and W on $(1,\infty)$, since for $t > 1$, $W(t)$ can be written as the sum of $W(1)$ and $N^{\cdot}(s)$ with $s = t - 1$. Note that M is also independent of V_n and U_n for all n, since they are all $\mathcal{F}$-measurable, where $\mathcal{F} = \sigma((W(t)_{t\geq 0})$.

To obtain (39), we now show that

$$n^{-1/4}M[V_n(1)] = n^{-1/4}M[V(1)] + o(n^{-3/4+2/r}) \text{ a.s.} \tag{41}$$

Of course, the result in (41) is untrue for a given extended Brownian motion M satisfying the required properties. It is indeed well known to be impossible to have $V_n(1) = V(1) + o(n^{-1/2}\log n)$ a.s. [see, e.g., Einmahl (1989, p21)] unless $V_n(1)$ itself is normally distributed. Here we claim that for each n there exists M satisfying (41) and other distributional requirements, without affecting other expansion results given in Lemma 3.2.

Note that M is a (extended) Brownian motion independent of (V_n, U_n) and (V, U). We may therefore write up to the distributional equivalence

$$n^{-1/4}M[V_n(1)] = n^{-1/4}|V_n(1)|^{1/2}M(1)$$

and

$$n^{-1/4}M[V(1)] = n^{-1/4}|V(1)|^{1/2}M(1)$$

without having to change the expansions of other sample product moments in Lemma 3.2, which are all functionals of (V_n, U_n) whose expansions are represented by (V, U). Consequently, we have

$$\begin{aligned}\left|n^{-1/4}M[V_n(1)] - n^{-1/4}M[V(1)]\right| &\leq n^{-1/4}|M(1)|\left||V_n(1)|^{1/2} - |V(1)|^{1/2}\right| \\ &\leq n^{-1/4}\frac{|M(1)|}{|V_n(1)|^{1/2} + |V(1)|^{1/2}}|V_n(1) - V(1)|,\end{aligned}$$

which establishes (41). The result stated in (39) follows immediately from (40) (with M instead of M_n) and (41), and that

$$|N[V_n(1)] - N[V(1)]| \le |V_n(1) - V(1)|^{1/2-\epsilon} = o(n^{-1/4+1/r+\epsilon}) \text{ a.s.}$$

for any $\epsilon > 0$. The proof is therefore complete. □

Proof of Part (b) We have

$$V_n(n_i) - V_n(n_{i-1}) = n^{1/2}[(T_{ni} - T_{n,i-1}) - 1].$$

Therefore, we may write

$$\begin{aligned} \frac{1}{n^{3/2}\sigma}\sum_{t=1}^{n} w_{t-1} &= \frac{1}{n}\sum_{i=1}^{n} W(T_{n,i-1}) \\ &= \sum_{i=1}^{n} W(T_{n,i-1})(T_{ni} - T_{n,i-1}) \\ &\quad - n^{-1/2}\sum_{i=1}^{n} W(T_{n,i-1})[V_n(n_i) - V_n(n_{i-1})]. \end{aligned} \tag{42}$$

For the first term in (42), we have

$$\begin{aligned} \sum_{i=1}^{n} W(T_{n,i-1})(T_{ni} - T_{n,i-1}) &= \int_0^1 W(t)dt + \int_1^{T_{nn}} W(t)dt \\ &\quad - \sum_{i=1}^{n}\int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]dt. \end{aligned}$$

However, it follows that

$$\begin{aligned} n^{1/2}\int_1^{T_{nn}} W(t)dt &= W(1)V_n(1) + n^{1/2}\int_1^{T_{nn}} [W(t) - W(1)]dt \\ &= W(1)V(1) + o(n^{-1/2+2/r}) \text{ a.s.} \end{aligned} \tag{43}$$

since

$$n^{1/2}(T_{nn} - 1) = V_n(1) = V(1) + o(n^{-1/2+2/r}) \text{ a.s.} \tag{44}$$

and

$$\left|\int_1^{T_{nn}} [W(t) - W(1)]dt\right| \le |T_{nn} - 1|^{3/2-\epsilon} \le n^{-3/4+\epsilon}|V_n(1) - V(1)|^{3/2-\epsilon} \text{ a.s.}$$

for any $\epsilon > 0$, due to the Hölder continuity of the Brownian motion sample path. Moreover, we have from (32) with $k = 1$

$$n^{1/2}\sum_{i=1}^{n}\int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]dt$$

$$= \frac{1}{3n\sigma^3}\sum_{i=1}^n \varepsilon_i^3 - n^{1/2}\sum_{i=1}^n \int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]^2 dW(t)$$
$$= \frac{\mu^3}{3\sigma^3} + \frac{1}{3n\sigma^3}\sum_{i=1}^n (\varepsilon_i^3 - \mu^3) - n^{1/2}\sum_{i=1}^n \int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]^2 dW(t)$$
$$= \frac{\mu^3}{3\sigma^3} + O_p(n^{-1/2}) \tag{45}$$

since, in particular,

$$\mathbf{E}\left(\int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]^2 dW(t)\right)^2 = \mathbf{E}\int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]^4 dt$$
$$= \frac{\mathbf{E}\varepsilon_i^6}{15n^3\sigma^6}$$

due to part (b) of Lemma A1. The asymptotic expansions of the first term in (42) may now readily be obtained from (43) and (45).

For the second term in (42), we have

$$\sum_{i=1}^n W(T_{n,i-1})[V_n(n_i) - V_n(n_{i-1})] = \int_0^1 W_n(t)dV_n(t)$$
$$= \int_0^1 W(t)dV(t) + o_p(n^{-1/2+2/r}). \tag{46}$$

Note that

$$\int_0^1 W_n(t)dV_n(t) - \int_0^1 W(t)dV(t) = \int_0^1 (W_n - W)(t)dV_n(t) + \int_0^1 W(t)d(V_n - V)(t)$$

and that

$$\int_0^1 W(t)d(V_n - V)(t) = W(1)[V_n(1) - V(1)] - \int_0^1 (V_n - V)(t)dW(t).$$

We thus have (46) due to the strong approximation in Lemma A3 and the result by Kurtz and Protter (1991). The stated result now follows immediately. □

Proof of Part (c) The proof of part (c) is similar to that of part (b). We write as in (42)

$$\frac{1}{n^2\sigma^2}\sum_{t=1}^n w_{t-1}^2 = \frac{1}{n}\sum_{i=1}^n W(T_{n,i-1})^2$$
$$= \sum_{i=1}^n W(T_{n,i-1})^2(T_{ni} - T_{n,i-1})$$
$$- n^{-1/2}\sum_{i=1}^n W(T_{n,i-1})^2[V_n(n_i) - V_n(n_{i-1})]. \tag{47}$$

We have

$$\sum_{i=1}^{n} W(T_{n,i-1})^2(T_{ni}-T_{n,i-1}) = \int_0^1 W(t)^2dt + \int_1^{T_{nn}} W(t)^2dt - \sum_{i=1}^{n}\int_{T_{n,i-1}}^{T_{ni}} [W(t)^2 - W(T_{n,i-1})^2]dt$$

and it follows that

$$\begin{aligned} n^{1/2}\int_1^{T_{nn}} W(t)^2dt &= W(1)^2V_n(1) + n^{1/2}\int_1^{T_{nn}} [W(t)^2 - W(1)^2]dt \\ &= W(1)^2V(1) + o(n^{-1/2+2/r}) \text{ a.s.,} \end{aligned}$$

similarly as in (43), and that

$$\begin{aligned} &\sum_{i=1}^{n}\int_{T_{n,i-1}}^{T_{ni}} [W(t)^2 - W(T_{n,i-1})^2]dt \\ &= \sum_{i=1}^{n}\int_{T_{n,i-1}}^{T_{ni}} [W(t)-W(T_{n,i-1})]^2dt + 2\sum_{i=1}^{n} W(T_{n,i-1})\int_{T_{n,i-1}}^{T_{ni}} [W(t)-W(T_{n,i-1})]dt \\ &= n^{-1/2}\frac{2\mu^3}{3\sigma^3 n}\sum_{i=1}^{n} W(T_{n,i-1}) + O_p(n^{-1}) \\ &= n^{-1/2}\frac{2\mu^3}{3\sigma^3}\int_0^1 W(t)dt + O_p(n^{-1}). \end{aligned}$$

Note that

$$\sum_{i=1}^{n}\int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]^2dt = O_p(n^{-1})$$

and that

$$\begin{aligned} &\mathbf{E}\left[\sum_{i=1}^{n} W(T_{n,i-1})\left(\int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]dt - \frac{\mu^3}{3n^{3/2}\sigma^3}\right)\right]^2 \\ &= n^{-3}K\left(\mathbf{E}\varepsilon_i^6\right)\mathbf{E}\left(\sum_{i=1}^{n} W(T_{n,i-1})^2\right) = O(n^{-2}) \end{aligned}$$

due to (32) and part (b) of Lemma A1, where K is some absolute constant. The expansion for the first term in (47) can therefore be easily obtained. For the second term in (47), we note that

$$\begin{aligned} \sum_{i=1}^{n} W(T_{n,i-1})^2[V_n(n_i) - V_n(n_{i-1})] &= \int_0^1 W_n(t)^2dV_n(t) \\ &= \int_0^1 W(t)^2dV(t) + o_p(n^{-1/2+2/r}) \end{aligned}$$

exactly as in (46). The proof for the result in part (c) is thus complete. □

Proof of Part (d) Write

$$\sum_{t=1}^{n} w_{t-1}\varepsilon_t = \frac{1}{2}\left[\left(\sum_{t=1}^{n}\varepsilon_t\right)^2 - \sum_{t=1}^{n}\varepsilon_t^2\right].$$

The stated result now follows easily from part (a) of Lemma 3.3 and part (a) of Lemma 3.5. □

Proof of Proposition 3.6 It follows from (16) that

$$\begin{aligned}
\sum_{t=1}^{n} y_{t-1} &= n\,\pi\sigma\nu + \pi\sum_{t=1}^{n} w_{t-1} - \varpi'\sum_{t=1}^{n} x_{t-1},\\
\sum_{t=1}^{n} y_{t-1}^2 &= \pi^2\sum_{t=1}^{n} w_{t-1}^2 + 2\pi^2\sigma\nu\sum_{t=1}^{n} w_{t-1} + n\,\pi^2\sigma^2\nu^2 + \varpi'\sum_{t=1}^{n} x_{t-1}x_{t-1}'\varpi\\
&\quad - 2\pi\varpi'\sum_{t=1}^{n} x_{t-1}w_{t-1} - 2\pi\sigma\nu\varpi'\sum_{t=1}^{n} x_{t-1},\\
\sum_{t=1}^{n} y_{t-1}\varepsilon_t &= \pi\sum_{t=1}^{n} w_{t-1}\varepsilon_t + \pi\sigma\nu\sum_{t=1}^{n}\varepsilon_t - \varpi'\sum_{t=1}^{n} x_{t-1}\varepsilon_t,\\
\sum_{t=1}^{n} x_{t-1}y_{t-1} &= \pi\sum_{t=1}^{n} x_{t-1}w_{t-1} - \sum_{t=1}^{n} x_{t-1}x_{t-1}'\varpi + \pi\sigma\nu\sum_{t=1}^{n} x_{t-1}.
\end{aligned}$$

Consequently, we have

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n} y_{t-1} = \pi\frac{1}{n^{3/2}}\sum_{t=1}^{n} w_{t-1} + n^{-1/2}\pi\sigma\nu + O_p(n^{-1}), \tag{48}$$

$$\frac{1}{n^2}\sum_{t=1}^{n} y_{t-1}^2 = \pi^2\frac{1}{n^2}\sum_{t=1}^{n} w_{t-1}^2 + n^{-1/2}\left(2\pi^2\sigma\nu\frac{1}{n^{3/2}}\sum_{t=1}^{n} w_{t-1}\right) + O_p(n^{-1}), \tag{49}$$

$$\frac{1}{n}\sum_{t=1}^{n} y_{t-1}\varepsilon_t = \pi\frac{1}{n}\sum_{t=1}^{n} w_{t-1}\varepsilon_t + n^{-1/2}\left(\pi\sigma\nu\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t - \varpi'\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_{t-1}\varepsilon_t\right), \tag{50}$$

$$\frac{1}{n}\sum_{t=1}^{n} x_{t-1}y_{t-1} = \pi\frac{1}{n}\sum_{t=1}^{n} x_{t-1}w_{t-1} - \frac{1}{n}\sum_{t=1}^{n} x_{t-1}x_{t-1}'\varpi + O_p(n^{-1/2}). \tag{51}$$

We may now easily deduce parts (a), (c) and (d) from Lemma 3.5, using (48) – (50).

Due to (51), part (b) follows if we establish

$$\frac{1}{n}\sum_{t=1}^{n} x_{t-1}w_{t-1} = \pi\sigma^2\iota[1 + J(W,W)] + o_p(1)$$

or equivalently,

$$\frac{1}{n}\sum_{t=1}^{n} w_{t-1}u_{t-i} = \pi\sigma^2[1 + J(W,W)] + o_p(1) \tag{52}$$

for all $i = 1, \ldots, p$. This is what we set out to do. We can show after some algebra that

$$\sum_{t=1}^{n} w_{t-1}u_{t-i} = \sum_{t=1}^{n} w_t u_t + \sum_{j=1}^{i-1}\sum_{t=1}^{n} \varepsilon_t u_{t-j} - \sum_{t=1}^{n} \varepsilon_t \sum_{j=0}^{i-1} u_{n-j}.$$

Moreover, it can be deduced that

$$\sum_{t=1}^{n} w_t u_t = \pi \sum_{t=1}^{n} w_t \varepsilon_t + \sum_{j=1}^{p} \pi_j \sum_{t=1}^{n} w_t(u_{t-j-1} - u_{t-j})$$

and that

$$\sum_{t=1}^{n} w_t(u_{t-j-1} - u_{t-j}) = \sum_{t=1}^{n} \varepsilon_{t+1}u_{t-j} - u_{n-j}\sum_{t=1}^{n+1} \varepsilon_t.$$

Consequently, we have

$$\frac{1}{n}\sum_{t=1}^{n} w_{t-1}u_{t-i} = \pi \left(\frac{1}{n}\sum_{t=1}^{n} w_{t-1}\varepsilon_t + \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t^2\right) + R_n$$

where

$$\begin{aligned} R_n &= \frac{1}{n}\sum_{j=1}^{p} \pi_j \sum_{t=1}^{n} \varepsilon_{t+1}u_{t-j} + \frac{1}{n}\sum_{j=1}^{i-1}\sum_{t=1}^{n} \varepsilon_t u_{t-j} \\ &\quad - \frac{1}{n}\sum_{t=1}^{n+1} \varepsilon_t \sum_{j=1}^{p} \pi_j u_{n-j} - \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t \sum_{j=0}^{i-1} u_{n-j} = O_p(n^{-1/2}). \end{aligned} \tag{53}$$

The result in (52) now follows directly from Lemma 3.3(a) and Lemma 3.5(a). The proof is therefore complete. □

Proof of Theorem 3.7 We may deduce from Lemma 3.3 and Proposition 3.6 that

$$\begin{aligned} \frac{P_n}{\pi\sigma^2} &= \frac{1}{n\pi\sigma^2}\sum_{t=1}^{n} y_{t-1}\varepsilon_t \\ &\quad - \frac{\pi}{\sqrt{n}}\left(\frac{1}{n\pi^2\sigma^2}\sum_{t=1}^{n} y_{t-1}x_{t-1}'\right)\left(\frac{1}{n\sigma^2}\sum_{t=1}^{n} x_{t-1}x_{t-1}'\right)^{-1}\left(\frac{1}{\sqrt{n}\sigma^2}\sum_{t=1}^{n} x_{t-1}\varepsilon_t\right) \\ &= J(W,W) + n^{-1/4}WM(V) \\ &\quad + n^{-1/2}\left[\frac{M(V)^2}{2} + WM(V) + \nu W - \frac{V+2U}{2} - \pi(1 + J(W,W))\iota'\Gamma^{-1}Z\right] \\ &\quad + o_p(n^{-1/2}) \end{aligned}$$

and that

$$\frac{Q_n}{\pi^2\sigma^2} = \frac{1}{n^2\pi^2\sigma^2}\sum_{t=1}^{n} y_{t-1}^2 + O_p(n^{-1})$$
$$= I(W^2) + n^{-1/2}\left[W^2V - J(W^2,V) + 2(\nu-\omega)I(W)\right] + o_p(n^{-1/2})$$
$$= I(W^2)\left[1 + n^{-1/2}\frac{W^2V - J(W^2,V) + 2(\nu-\omega)I(W)}{I(W^2)}\right] + o_p(n^{-1/2}).$$

Consequently, it follows that

$$Q_n^{-1} = \frac{1}{\pi^2\sigma^2 I(W^2)}\left[1 - n^{-1/2}\,\frac{W^2V - J(W^2,V) + 2(\nu-\omega)I(W)}{I(W^2)}\right] + o_p(n^{-1/2}),$$
$$Q_n^{-1/2} = \frac{1}{\pi\sigma\sqrt{I(W^2)}}\left[1 - n^{-1/2}\,\frac{W^2V - J(W^2,V) + 2(\nu-\omega)I(W)}{2I(W^2)}\right] + o_p(n^{-1/2}).$$

Moreover, we have from Proposition 3.4 that

$$\sigma_n^{-1} = \sigma^{-1}\left[1 - n^{-1/2}(V+2U)/2\right] + o_p(n^{-1/2}),$$
$$\alpha_n(1)^{-1} = \alpha(1)^{-1}\left[1 + n^{-1/2}\alpha(1)^{-1}\iota'\Gamma^{-1}Z\right] + o_p(n^{-1/2}).$$

Now the stated results follow easily after some tedious, but straightforward, algebra. □

Proof of Corollary 3.8 We first prove that all of the remainder terms of the asymptotic expansions given in Lemmas 3.3 and 3.5 and Propositions 3.4 and 3.6 are distributionally of order $o(n^{-1/2})$, and subsequently show that the error terms in Theorem 3.7 are distributionally of order $o(n^{-1/2})$. The stated results will then follow from the part (a) of Lemma A4. The leading terms F and G of our expansions presented in Theorem 3.7 have bounded densities and finite integral moments of all orders. This is shown in Evans and Savin (1981) and Abadir (1993). Note that all of the expansion terms appearing in the proof of Theorem 3.7 have bounded densities and finite moments up to arbitrary orders, being simple functionals of Brownian motions. Therefore, they satisfy the conditions for S_n and T_n respectively in the parts (b) and (c) of Lemma A4. Furthermore, following Evans and Savin (1981) and Abadir (1993), we may also show that $1/I(W^2)$ has bounded density and finite integral moments of all orders. Consequently, all our expansion terms included in the lower order terms (F_1,F_2) and (G_1,G_2), being products of such terms, have bounded densities and finite moments up to arbitrary orders. Finally, the denominators of F_n and G_n have the expansion terms, the reciprocals of which have bounded densities and finite moments of all orders. This is required to apply the part (e) of Lemma A4.

It is easy to see from the proofs of Lemmas 3.3 and 3.5 and Propositions 3.4 and 3.5 that the remainder terms are majorized by one of the following three types:

(A) $R_n^a = n^{-p}\sup_{t\in[0,1]}|A_n(t) - A(t)|$ with some $p \geq 1/4$,

(B) $R_n^b = n^{-p}\sup_{t\in[0,1]}|Z_n(t) - Z(t)|$ with some $p \geq 7/24$, or

(C) $R_n^c = n^{-p} S_n$ and $\mathbf{E}|S_n|^q < \infty$ uniformly in n with some $p > 1/2$ and $q > 1/(2p-1)$.

If Assumption 2.1 holds with $r > 12$ as assumed here, we may readily show that all three types of the remainder terms introduced here are distributionally of order $o(n^{-1/2})$. We may indeed easily deduce for the type (A) remainder term that

$$\begin{aligned}\mathbf{P}\left\{|R_n^a| > n^{-1/2-\epsilon}\right\} &\leq \mathbf{P}\left\{\sup_{0\leq t\leq 1} |A_n(t) - A(t)| > n^{-1/2+p-\epsilon)}\right\} \\ &\leq n^{1-rp/2+\epsilon}(1+\sigma^{-r})K(1+\mathbf{E}|\varepsilon_i|^r)\end{aligned}$$

due to Lemma A3. However, we have $n^{1-rp/2+\epsilon} = o(n^{-1/2})$ for sufficiently small $\epsilon > 0$ if $r > 12$ and $p \geq 1/4$. Likewise, it also follows from Lemma A3 that

$$\begin{aligned}\mathbf{P}\left\{|R_n^b| > n^{-1/2-\epsilon}\right\} &\leq \mathbf{P}\left\{\sup_{0\leq t\leq 1} |Z_n(t) - Z(t)| > n^{-1/2+p-\epsilon)}\right\} \\ &\leq n^{r/4-rp+\epsilon}(1+v^r)K\left[1+(\mathbf{E}|\varepsilon_i|^r)^2\right]\end{aligned}$$

for the type (B) remainder term. Note that, if $r > 12$ and $p \geq 7/24$ as given, $n^{r/4-rp+\epsilon} = o(n^{-1/2})$ for sufficiently small $\epsilon > 0$. On the other hand, it follows for the type (C) remainder term that

$$\mathbf{P}\left\{|R_n^c| > n^{-1/2-\epsilon}\right\} \leq \mathbf{P}\left\{|S_n| > n^{p-1/2-\epsilon}\right\} \leq n^{-(p-1/2)q+\epsilon}\mathbf{E}|S_n|^q$$

and, since $q > 1/(2p-1)$, we have $n^{-(p-1/2)q+\epsilon} = o(n^{-1/2})$ as required to show. Respectively for $p = 1$ and $p = 3/4$, it suffices to have $q > 1$ and $q > 2$.

For the remainder terms involving $|W_n(1)-W(1)|$, $|V_n(1)-V(1)|$ and $|U_n(1)-U(1)|$, the result for the type (A) remainder term is clearly applicable. The terms including stochastic integrals such as $\int_0^1 (W_n - W)(t)dV_n(t)$ and $\int_0^1 (V_n - V)(t)dW(t)$ can be dealt with similarly, since their stochastic orders are effectively determined by their quadratic variations that are bounded by $\sup_{0\leq t\leq 1}|W_n(t) - W(t)|^2$ and $\sup_{0\leq t\leq 1}|V_n(t) - V(t)|^2$. All our type (A) remainder terms are given with $p \geq 1/4$. Similarly, the result for the type (B) remainder term applies to the remainder term $|Z_n(1) - Z(1)|$, if as in our case $p \geq 7/24$. As the type (C) remainder term, we have the remainder terms consisting of

$$\begin{aligned}S_{1n} &= n^{-1/2}\sum_{j=1}^{p}\sum_{i=1}^{n} u_{i-j}\varepsilon_i, \\ S_{2n} &= n^{-1/2}\sum_{i=1}^{n}(x_{i-1}x_{i-1}' - \Gamma), \\ S_{3n} &= n^{-1/2}\sum_{i=1}^{n}(\varepsilon_i^3 - \mu^3), \\ S_{4n} &= n\sum_{i=1}^{n}\int_{T_{n,i-1}}^{T_{ni}}[W(t)-W(T_{n,i-1})]^2 dW(t),\end{aligned}$$

$$
\begin{aligned}
S_{5n} &= n\sum_{i=1}^{n} W(T_{n,i-1})\left(\int_{T_{n,i-1}}^{T_{ni}} [W(t)-W(T_{n,i-1})]dt - \frac{\mu^3}{3n^{3/2}\sigma^3}\right),\\
S_{6n} &= n^{3/2}\sum_{i=1}^{n} \int_{T_{n,i-1}}^{T_{ni}} [W(t)-W(T_{n,i-1})]^2 dt.
\end{aligned}
$$

Under the assumption $r > 12$, all of $S_{1n}, \ldots, S_{6n}$ satisfy $\mathbf{E}|S_n|^q < \infty$ with the values of q respectively greater than $12, 6, 4, 4, 4, 3$, and the values of p should be greater than or equal to $13/24, 7/12, 5/8, 5/8, 5/8, 1/3$ correspondingly. The condition is met for all our type (C) remainder terms.

The remainder terms in parts (a), (b) and (c) of Lemma 3.3 are majorized respectively by the type (A) remainder terms $|V_n(1)-V(1)|$ and $|U_n(1)-U(1)|$, the type (B) remainder term $|Z_n(1)-Z(1)|$, and the type (C) remainder term with S_{2n} defined above. Both parts (a) and (b) of Proposition 3.4 inherit the remainder terms from Lemma 3.3. The remainder term in part (a) consists of all four terms appearing previously, while part (b) only includes the latter two of those. Part (a) of Lemma 3.5 has the remainder term essentially consisting only of $|V_n(1) - V(1)|$. The remainder terms in parts (b) and (c) of Lemma 3.5 include various additional terms, as well as those appeared earlier. Part (b) has the type (A) remainder terms $|V_n(1) - V(1)|$, $\int_0^1 (W_n - W)(t)dV_n(t)$ and $\int_0^1 (V_n - V)(t)dW(t)$, and type (C) remainder terms with S_{3n} and S_{4n}. Part (c) includes the type (A) remainder terms $|W_n(1) - W(1)|$ and $|V_n(1) - V(1)|$, and the type (C) remainder terms with S_{5n} and S_{6n}. There is no new remainder term in part (d) of Lemma 3.5. Proposition 3.6 does not introduce any new remainder term except the type (C) remainder term with S_{1n} and its trivial variants. The rest remainder terms appearing in parts (a) - (d) of Proposition 3.6 are inherited from our earlier results. Note that we may allow the remainder terms introduced here to be multiplied by a random sequences satisfying the conditions in part (b) or (c) of Lemma A4. Note also that the products of two remainder terms and the expansions for the inverses can be dealt using the results in parts (d) and (e) of Lemma A4. This completes the proof. □

Proof of Lemma 3.9 We let

$$B_n^* = (A_n^{*\prime}, Z_n^{*\prime})'$$

similarly as in (33). Then it follows from Lemma A3 that we may choose the limit Brownian motion $B = (A', Z')'$ satisfying

$$\mathbf{P}^*\left\{\sup_{0\le t\le 1} |A_n^*(t) - A(t)| > c\right\} \le n^{1-r/4}c^{-r/2}(1+\sigma_n^{-r})K(1+\mathbf{E}^*|\varepsilon_i^*|^r) \tag{54}$$

and

$$\mathbf{P}^*\left\{\sup_{0\le t\le 1} |Z_n^*(t) - Z(t)| > c\right\} \le n^{-r/4}c^{-r}(1+\upsilon_n^r)K\left[1+(\mathbf{E}^*|\varepsilon_i^*|^r)^2\right] \tag{55}$$

where υ_n^2 is the sample analogue estimator for υ^2, defined similarly as σ_n^2 for σ^2. Note that we have $\upsilon_n^2 = \sum_{i=0}^\infty \varphi_{ni}^2$ in terms of the coefficients (φ_{ni}) in the MA representation of

(u_t^*). The autoregression (6) is invertible a.s. for large n, and therefore, (φ_{ni}) is absolutely summable a.s.

To obtain the stated result, now it suffices to show that

$$\mathbf{E}^*|\varepsilon_i^*|^r < \infty \text{ a.s.} \tag{56}$$

for some $r > 4$. Given (56), the bootstrap invariance principle $B_n^* \to_{d^*} B^*$ a.s. follows immediately from (54) and (55). To show (56), we write

$$\mathbf{E}^*|\varepsilon_i^*|^r = \frac{1}{n}\sum_{i=1}^n \left|\hat{\varepsilon}_i - \frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i\right|^r \leq K(A_n + B_n + C_n^r)$$

where

$$\begin{aligned}
A_n &= \frac{1}{n}\sum_{i=1}^n |\varepsilon_i|^r, \\
B_n &= \left(\max_{1\leq i\leq p} |\alpha_{ni} - \alpha_i|^r\right) \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^p |u_{i-j}|^r, \\
C_n &= \left|\frac{1}{n}\sum_{i=1}^n \varepsilon_i\right| + \left(\max_{1\leq i\leq p} |\alpha_{ni} - \alpha_i|^r\right) \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^p |u_{i-j}|.
\end{aligned}$$

Note that $\sum_{i=1}^n |\varepsilon_i|^r/n \to_{a.s.} \mathbf{E}|\varepsilon_i|^r$, $\sum_{i=1}^n |u_i|/n \to_{a.s.} \mathbf{E}|u_i|$ and $\sum_{i=1}^n |u_i|^r/n \to_{a.s.} \mathbf{E}|u_i|^r$ by strong laws of large numbers. Also, we have $\sum_{i=1}^n \varepsilon_i/n = O(n^{-1/2}(\log\log n)^{1/2})$ by law of the iterated logarithm. Moreover, we may easily show using the result in, e.g., Móricz (1976) that $\max_{1\leq i\leq p} |\alpha_{ni} - \alpha_i| = o(n^{-1/2}(\log n)^{1/2})$ a.s. The condition in (56) thus holds and the proof is complete. $\square$

Proof of Theorem 3.10 The proof is analogous to that of Theorem 3.7. We just need to show the remainder terms in Theorem 3.7 are now given in terms of $o_p^*(n^{-1/2})$ in place of $o_p(n^{-1/2})$. This follows rather straightforwardly from our earlier results, as we will explain below. We say that remainder term R_n in our expansion is majorized by certain moments and parameters if $\mathbf{E}|R_n|^s$, for some $s > 0$, is bounded by those moments and parameters. If R_n is majorized by some moments and parameters, their bootstrap counterpart R_n^*, say, is majorized by the corresponding sample moments and estimators based on the expectation $\mathbf{E}^*$. Therefore, to show that the bootstrap remainder term R_n^* is $O_p^*(n^{-s})$ for some $s > 0$, it suffices to have the corresponding $n^s|R_n|$ majorized by the moments and parameters whose sample analogue estimators converge a.s.

Our strong approximations in Lemma A3 have the bounds majorized by the parameter v^2 and the moment $\mathbf{E}|\varepsilon_i|^r$. Consequently, we may immediately deduce their bootstrap analogues given in (54) and (55). The bootstrap strong approximations in (54) and (55) in turn yield the bootstrap stochastic orders of the errors in approximating B_n^* by B. The bootstrap stochastic orders of the bootstrap remainder terms bounded by $\sup_{t\in[0,1]} |B_n^*(t) - B(t)|$ can therefore be easily determined.

Other types of the bootstrap remainder terms can also be easily analyzed. For instance, we may show that R_n given in (53) is majorized by $\upsilon^2, \sigma^2, \Gamma$ and ϖ. Note that

$$\mathbf{E}\left(\frac{1}{n}\sum_{j=1}^{p}\pi_j\sum_{t=1}^{n}\varepsilon_{t+1}u_{t-j}\right)^2 = \frac{\sigma^2}{n}\varpi'\Gamma\varpi.$$

Also, we have that

$$\mathbf{E}\left(\frac{1}{n}\sum_{j=1}^{i-1}\sum_{t=1}^{n}\varepsilon_t u_{t-j}\right)^2 \leq (i-1)\sum_{j=1}^{i-1}\mathbf{E}\left(\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t u_{t-j}\right)^2 \leq \frac{p^2}{n}\sigma^2\upsilon^2.$$

Moreover, we have

$$\mathbf{E}\left(\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t\right)^2 = \frac{\sigma^2}{n}, \qquad \mathbf{E}\left(\frac{1}{n}\sum_{t=1}^{n+1}\varepsilon_t\right)^2 = \frac{(n+1)\sigma^2}{n},$$

$$\mathbf{E}\left(\frac{1}{n}\sum_{j=1}^{p}\pi_j u_{n-j}\right)^2 = \frac{\varpi'\Gamma\varpi}{n}, \qquad \mathbf{E}\left(\frac{1}{n}\sum_{j=0}^{i-1}u_{n-j}\right)^2 \leq \frac{p^2\upsilon^2}{n}.$$

Consequently, it follows that

$$\mathbf{E}|R_n| \leq \frac{2\sigma}{n}\left(p\upsilon + \frac{1+\sqrt{2}}{2}(\varpi'\Gamma\varpi)^{1/2}\right)$$

and that $R_n = O_p(n^{-1})$. Correspondingly, we have

$$\mathbf{E}^*|R_n^*| \leq \frac{2\sigma_n}{n}\left(p\upsilon_n + \frac{1+\sqrt{2}}{2}(\varpi_n'\Gamma_n\varpi_n)^{1/2}\right)$$

from which we may deduce that $R_n^* = O_p^*(n^{-1})$

Moreover, if we write

$$\frac{1}{n}\sum_{t=1}^{n}x_{t-1}x_{t-1}' = \Gamma + R_n,$$

then it follows from the part (f) of Lemma A2 that

$$\mathbf{E}|R_n|^{r/2} \leq n^{-r/4}\upsilon^r K\left(\sigma^r + \mathbf{E}|\varepsilon_i|^r\right)$$

and we have $R_n = O_p(n^{-r/4})$. Similarly, we have

$$\mathbf{E}^*|R_n^*|^{r/2} \leq n^{-r/4}\upsilon_n^r K\left(\sigma_n^r + \mathbf{E}^*|\varepsilon_i^*|^r\right)$$

and $R_n^* = O_p^*(n^{-r/4})$. If $r \geq 8$ as assumed, we have $R_n = O_p(n^{-2})$ and $R_n = O_p^*(n^{-2})$. $\square$

Proof of Corollary 3.11 Given Lemma 3.9 and Theorem 3.10 and their proofs, the proof is entirely analogous to that of Corollary 3.8. The details are therefore omitted. □

Proof of Theorem 4.1 We let $u_t = \triangle_c y_t$, so that (u_t) becomes an AR(p) process as earlier, and let

$$\triangle y_t = u_t - \frac{c}{n} y_{t-1}.$$

Moreover, we write

$$\triangle y_t = \sum_{i=1}^{p} \alpha_i \triangle y_{t-i} + \left[\varepsilon_t - \frac{c}{n} \left(y_{t-1} - \sum_{i=1}^{p} \alpha_i y_{t-1-i} \right) \right]. \tag{57}$$

In what follows, we denote respectively by (α_{ni}) and σ_n^2 the least squares estimators of (α_i) and σ^2, and by $(\hat{\varepsilon}_t)$ the fitted residuals, in regression (57).

We first establish that

$$\sum_{t=1}^{n} y_{t-i} y_{t-j} = o(n^2 \log n) \text{ a.s.,} \tag{58}$$

$$\sum_{t=1}^{n} y_{t-i} u_{t-j} = o(n \log n) \text{ a.s.,} \tag{59}$$

which would follow immediately if we show

$$\sum_{t=1}^{n} y_t^2 = o(n^2 \log n) \text{ a.s.,} \tag{60}$$

$$\sum_{t=1}^{n} y_{t-1} u_t = o(n \log n) \text{ a.s.} \tag{61}$$

since, in particular, $\sum_{t=1}^{n} u_{t-i} u_{t-j} = O(n)$ a.s. by strong law of large numbers. Note that

$$\sum_{t=1}^{n} y_{t-1} u_t = \frac{1}{2\alpha} y_n^2 + \frac{1-\alpha^2}{2\alpha} \sum_{t=1}^{n} y_{t-1}^2 - \frac{1}{2\alpha} \sum_{t=1}^{n} u_t^2. \tag{62}$$

The initialization of (y_t) does not affect our result, and for simplicity we assume $y_0 = 0$ a.s. here and in what follows.

It can be readily deduced after recursive substitution that

$$y_i = \sum_{j=1}^{i} u_j - \frac{1-\alpha}{\alpha} \sum_{j=1}^{i-1} \alpha^{i-j} \left(\sum_{k=1}^{j} u_k \right).$$

However, we may show

$$\max_{1 \le i \le n} \left| \sum_{k=1}^{i} u_k \right| = o\left((n \log n)^{1/2} \right) \text{ a.s.}$$

as in the proof of Theorem 6 of Móricz (1976) [i.e., by applying his inequality in the bottom line of page 309 to M_n in place of S_n]. Consequently,

$$\max_{1\le i\le n}\left|\frac{y_i}{\sqrt{n}}\right| = o\left((\log n)^{1/2}\right) \text{ a.s.} \tag{63}$$

We may now easily obtain (60) from (63), and (61) from (62) together with (63).

It follows from (58) and (59) that

$$\sum_{t=1}^{n} \triangle y_{t-i}\triangle y_{t-j} = \sum_{t=1}^{n} u_{t-i}u_{t-j} + o(\log n) \text{ a.s.} \tag{64}$$

and

$$\sum_{t=1}^{n} \triangle y_{t-i}\left[\varepsilon_t - \frac{c}{n}\left(y_{t-1} - \sum_{i=1}^{p}\alpha_i y_{t-1-i}\right)\right] = \sum_{t=1}^{n} u_{t-i}\varepsilon_t + o(\log n) \text{ a.s.} \tag{65}$$

and we have immediately from (64) and (65) that

$$\max_{1\le i\le p}|\alpha_{ni} - \alpha_i| = o(n^{-1/2}(\log n)^{1/2}) \text{ a.s.} \tag{66}$$

as $n\to\infty$. Moreover, since

$$\hat{\varepsilon}_t = \varepsilon_t - \frac{c}{n}\left(y_{t-1} - \sum_{i=1}^{p}\alpha_i y_{t-1-i}\right) - \sum_{i=1}^{p}(\alpha_{ni}-\alpha_i)\triangle y_{t-i},$$

we have

$$\sigma_n^2 \to_{a.s.} \sigma^2$$

as $n\to\infty$, and we may deduce exactly as in the proof of Lemma 3.9 that

$$\mathbf{E}^*|\varepsilon_i^*|^r < \infty \text{ a.s.}$$

due to (63), (64), (65) and (66). The bootstrap invariance principle in Lemma 3.9 thus holds also under the local-to-unity model. The proof is therefore complete. □

Proof of Lemma 5.1 For part (a), we simply note that

$$\frac{1}{n^{1/2}\sigma}\sum_{t=1}^{n}\frac{t}{n}\varepsilon_t = -\frac{1}{n^{3/2}\sigma}\sum_{t=1}^{n} w_{t-1} + \frac{1}{n^{1/2}\sigma}\sum_{t=1}^{n}\varepsilon_t.$$

The stated result then follows directly from Lemma 3.5 and the fact that $W - I(W) = J(\imath, W)$, which can easily be deduced using integration by parts formula.

Let $n_i = i/n$ for $i = 1,\ldots,n$. To prove part (a), we first note that

$$\frac{1}{n^{3/2}\sigma}\sum_{t=1}^{n}\frac{t}{n}w_{t-1} = \frac{1}{n^{3/2}\sigma}\sum_{t=1}^{n}\frac{t-1}{n}w_{t-1} + O_p(n^{-1})$$

and write

$$\frac{1}{n^{3/2}\sigma}\sum_{t=1}^{n}\frac{t-1}{n}w_{t-1} = \frac{1}{n}\sum_{i=1}^{n} n_{i-1}W(T_{n,i-1}) = -A_n + B_n$$

where

$$A_n = \frac{1}{n}\sum_{i=1}^{n}(T_{n,i-1} - n_{i-1})W(T_{n,i-1}) \quad \text{and} \quad B_n = \frac{1}{n}\sum_{i=1}^{n} T_{n,i-1}W(T_{n,i-1}),$$

each of which will be analyzed below.

It is straightforward to deduce that

$$A_n = n^{-1/2}\frac{1}{n^2}\sum_{i=1}^{n} V_n(n_{i-1})W(T_{n,i-1}) = n^{-1/2}I(WV) + o_p(n^{-1/2}). \tag{67}$$

Furthermore, we may write B_n as

$$B_n = I(\imath W) + n^{-1/2}[WV - J(\imath W, V)] - C_n + o_p(n^{-1/2}) \tag{68}$$

where

$$C_n = \sum_{i=1}^{n}\int_{T_{n,i-1}}^{T_{ni}} [tW(t) - T_{n,i-1}W(T_{n,i-1})]dt.$$

To deduce (68), note that

$$\begin{aligned} B_n &= \sum_{i=1}^{n} T_{n,i-1}W(T_{n,i-1})(T_{ni} - T_{n,i-1}) \\ &\quad - n^{-1/2}\sum_{i=1}^{n} T_{n,i-1}W(T_{n,i-1})[(V_n(n_i) - V_n(n_{i-1})] \end{aligned}$$

and

$$\begin{aligned} &\sum_{i=1}^{n} T_{n,i-1}W(T_{n,i-1})(T_{ni} - T_{n,i-1}) \\ &= I(\imath W) + \int_1^{T_{nn}} tW(t)dt - \sum_{i=1}^{n}\int_{T_{n,i-1}}^{T_{ni}} [tW(t) - T_{n,i-1}W(T_{n,i-1})]dt. \end{aligned}$$

Moreover, observe that

$$n^{1/2}\int_1^{T_{nn}} tW(t)dt = WV + o_p(1)$$

and that

$$\sum_{i=1}^{n} T_{n,i-1}W(T_{n,i-1})[(V_n(n_i) - V_n(n_{i-1})] = J(\imath W, V) + o_p(1)$$

due to Kurz and Protter (1992).

Now we write

$$C_n = \sum_{i=1}^n T_{n,i-1} \int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]dt + \sum_{i=1}^n W(T_{n,i-1}) \int_{T_{n,i-1}}^{T_{ni}} (t - T_{n,i-1})dt$$
$$+ \sum_{i=1}^n \int_{T_{n,i-1}}^{T_{ni}} (t - T_{n,i-1})[W(t) - W(T_{n,i-1})]dt$$

and show that

$$C_n = n^{-1/2} \frac{\mu^3}{6\sigma^3} + o_p(n^{-1/2}). \tag{69}$$

Note that

$$n^{1/2} \sum_{i=1}^n T_{n,i-1} \int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]dt$$
$$= \frac{\mu^3}{3\sigma^3} \frac{1}{n} \sum_{i=1}^n T_{n,i-1} + o_p(1) = \frac{\mu^3}{6\sigma^3} + o_p(1),$$

which becomes the leading term in C_n. The rest terms are negligible as we show below.

We have

$$\sum_{i=1}^n W(T_{n,i-1}) \int_{T_{n,i-1}}^{T_{ni}} (t - T_{n,i-1})dt = \frac{1}{2} \sum_{i=1}^n W(T_{n,i-1})(T_{ni} - T_{n,i-1})^2$$
$$= \frac{1}{2n^2} \sum_{i=1}^n W(T_{n,i-1})\Delta_i^2 = O_p(n^{-1}).$$

Moreover, we have

$$\begin{aligned}
&\mathbf{E} \left| \int_{T_{n,i-1}}^{T_{ni}} (t - T_{n,i-1})[W(t) - W(T_{n,i-1})]dt \right| \\
&\leq \mathbf{E} \left(\int_{T_{n,i-1}}^{T_{ni}} (t - T_{n,i-1})^2 dt \right)^{1/2} \left(\int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]^2 dt \right)^{1/2} \\
&\leq \left(\mathbf{E} \int_{T_{n,i-1}}^{T_{ni}} (t - T_{n,i-1})^2 dt \right)^{1/2} \left(\mathbf{E} \int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]^2 dt \right)^{1/2} \\
&= O_p(n^{-5/2})
\end{aligned}$$

since

$$\mathbf{E} \int_{T_{n,i-1}}^{T_{ni}} (t - T_{n,i-1})^2 dt = O(n^{-3}),$$
$$\mathbf{E} \int_{T_{n,i-1}}^{T_{ni}} [W(t) - W(T_{n,i-1})]^2 dt = O(n^{-2}),$$

and therefore,

$$\sum_{i=1}^{n}\int_{T_{n,i-1}}^{T_{ni}}(t-T_{n,i-1})[W(t)-W(T_{n,i-1})]dt=O_p(n^{-3/2}).$$

We thus have established (69). The stated result in part (a) now follows immediately from (67), (68) and (69). The proof is therefore complete. □

Proof of Proposition 5.2 The stated result is immediate from Lemma 5.1 and (16). □

Proof of Theorem 5.3 For time series (z_t), we let $\tilde{z}_t = z_t - \sum_{t=1}^n z_t/n$ for the case $q=0$, and let

$$\tilde{z}_t = z_t - \frac{1}{n}\sum_{t=1}^{n} z_t - \left(\sum_{t=1}^{n}(t-c_n)z_t \Bigg/ \sum_{t=1}^{n}(t-c_n)^2\right)(t-c_n)$$

with $c_n=(n+1)/2$ for the case $q=1$. Define $\tilde{P}_n$ and $\tilde{Q}_n$ by

$$\begin{aligned}
\tilde{P}_n &= \frac{1}{n}\sum_{t=1}^{n}\tilde{y}_{t-1}\tilde{\varepsilon}_t - \frac{1}{n}\left(\sum_{t=1}^{n}\tilde{y}_{t-1}\tilde{x}_{t-1}'\right)\left(\sum_{t=1}^{n}\tilde{x}_{t-1}\tilde{x}_{t-1}'\right)^{-1}\left(\sum_{t=1}^{n}\tilde{x}_{t-1}\tilde{\varepsilon}_t\right),\\
\tilde{Q}_n &= \frac{1}{n^2}\sum_{t=1}^{n}\tilde{y}_{t-1}^2 - \frac{1}{n^2}\left(\sum_{t=1}^{n}\tilde{y}_{t-1}\tilde{x}_{t-1}'\right)\left(\sum_{t=1}^{n}\tilde{x}_{t-1}\tilde{x}_{t-1}'\right)^{-1}\left(\sum_{t=1}^{n}\tilde{x}_{t-1}\tilde{y}_{t-1}\right),
\end{aligned}$$

similarly as P_n and Q_n in (11) and (12). Also, we let

$$\tilde{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n}\tilde{\varepsilon}_t^2 - \frac{1}{n}\left(\sum_{t=1}^{n}\tilde{\varepsilon}_t\tilde{x}_{t-1}'\right)\left(\sum_{t=1}^{n}\tilde{x}_{t-1}\tilde{x}_{t-1}'\right)^{-1}\left(\sum_{t=1}^{n}\tilde{x}_{t-1}\tilde{\varepsilon}_t\right)$$

and define

$$\tilde{\alpha}_n(1) = \alpha(1) - \iota'\left(\sum_{t=1}^{n}\tilde{x}_{t-1}\tilde{x}_{t-1}'\right)^{-1}\left(\sum_{t=1}^{n}\tilde{x}_{t-1}\tilde{\varepsilon}_t\right),$$

which correspond to σ_n^2 and $\alpha_n(1)$ in (13) and (14). Then we may write

$$\tilde{F}_n = \frac{\tilde{P}_n}{\tilde{\sigma}_n\sqrt{\tilde{Q}_n}},\quad \tilde{G}_n = \frac{\tilde{P}_n}{\tilde{\alpha}_n(1)\tilde{Q}_n},$$

correspondingly as F_n and G_n in (15).

For both the cases $q=0$ and $q=1$, it can be easily deduced that

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\tilde{x}_{t-1}\tilde{x}_{t-1}' &= \frac{1}{n}\sum_{t=1}^{n}x_{t-1}x_{t-1}' + O_p(n^{-1}),\\
\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\tilde{x}_{t-1}\tilde{\varepsilon}_t &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}x_{t-1}\varepsilon_t + O_p(n^{-1/2}),\\
\frac{1}{n}\sum_{t=1}^{n}\tilde{\varepsilon}_t^2 &= \frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2 + O_p(n^{-1}).
\end{aligned}$$

Note in particular that

$$\frac{1}{n^3}\sum_{t=1}^{n}(t-c_n)^2 = \frac{1}{3} + O(n^{-1})$$

and

$$\frac{1}{n^{3/2}}\sum_{t=1}^{n}(t-c_n)z_t = \frac{1}{n^{1/2}}\sum_{t=1}^{n}\frac{t}{n}z_t + O_p(n^{-1}) \tag{70}$$

for both $z_t = x_{t-1}$ and ε_t.

Moreover, we have for the case $q = 0$

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\tilde{y}_{t-1}\tilde{\varepsilon}_t &= \frac{1}{n}\sum_{t=1}^{n}y_{t-1}\varepsilon_t - \left(\frac{1}{n^{3/2}}\sum_{t=1}^{n}y_{t-1}\right)\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t\right), \\
\frac{1}{n^2}\sum_{t=1}^{n}\tilde{y}_{t-1}^2 &= \frac{1}{n^2}\sum_{t=1}^{n}y_{t-1}^2 - \left(\frac{1}{n^{3/2}}\sum_{t=1}^{n}y_{t-1}\right)^2, \\
\frac{1}{n}\sum_{t=1}^{n}\tilde{x}_{t-1}\tilde{y}_{t-1} &= \frac{1}{n}\sum_{t=1}^{n}x_{t-1}y_{t-1} - \iota\left(\frac{1}{n^{3/2}}\sum_{t=1}^{n}y_{t-1}\right)\left(\pi\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t\right) + o_p(1),
\end{aligned}$$

and for the case $q = 1$

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\tilde{y}_{t-1}\tilde{\varepsilon}_t &= \frac{1}{n}\sum_{t=1}^{n}y_{t-1}\varepsilon_t - \left(\frac{1}{n^{3/2}}\sum_{t=1}^{n}y_{t-1}\right)\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t\right) \\
&\quad - 3\left(\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{t}{n}y_{t-1} - \frac{1}{2n^{3/2}}\sum_{t=1}^{n}y_{t-1}\right)\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{t}{n}\varepsilon_t\right) + O_p(n^{-1}), \\
\frac{1}{n^2}\sum_{t=1}^{n}\tilde{y}_{t-1}^2 &= \frac{1}{n^2}\sum_{t=1}^{n}y_{t-1}^2 - \left(\frac{1}{n^{3/2}}\sum_{t=1}^{n}y_{t-1}\right)^2 + O_p(n^{-1}) \\
&\quad - 3\left(\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{t}{n}y_{t-1} - \frac{1}{2n^{3/2}}\sum_{t=1}^{n}y_{t-1}\right)^2, \\
\frac{1}{n}\sum_{t=1}^{n}\tilde{x}_{t-1}\tilde{y}_{t-1} &= \frac{1}{n}\sum_{t=1}^{n}x_{t-1}y_{t-1} - \iota\left(\frac{1}{n^{3/2}}\sum_{t=1}^{n}y_{t-1}\right)\left(\pi\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t\right) \\
&\quad - 3\iota\left(\frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{t}{n}y_{t-1} - \frac{1}{2n^{3/2}}\sum_{t=1}^{n}y_{t-1}\right)\left(\pi\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{t}{n}\varepsilon_t\right) + o_p(1),
\end{aligned}$$

which follows from (70) and

$$\frac{1}{n^{5/2}}\sum_{t=1}^{n}(t-c_n)y_{t-1} = \frac{1}{n^{3/2}}\sum_{t=1}^{n}\frac{t}{n}y_{t-1} - \frac{1}{2n^{3/2}}\sum_{t=1}^{n}y_{t-1} + O_p(n^{-1})$$

The stated results now follow easily. □

Proof of Theorem 5.4 Given the results in Theorem 5.3, the proof is entirely analogous with the proofs of Theorem 3.10 and Corollary 3.11. The details are therefore omitted. □

References

Abadir, K.M. (1993). The limiting distribution of the autocorrelation coefficient under a unit root. *Annals of Statistics*, **21**, 1058-1070.

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves and R.L. Taylor (1991a). "Bootstrapping unstable first-order autoregressive processes," *Annals of Statistics*, **19**, 1098-1101.

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves and R.L. Taylor (1991b). "Bootstrap test of significance and sequential bootstrap estimation for unstable first order autoregressive processes," *Communications in Statistics - Theory and Methods*, **20**, 1015-1026.

Chang, Y. and J.Y. Park (2002). "A sieve bootstrap for the test of a unit root," *Journal of Time Series Analysis*, forthcoming.

Courbot, B. (2001). "Rates of convergence in the functional CLT for multidimensional continous time martingales," *Stochastic Processes and Their Applications*, **91**, 57-76.

Datta, S. (1996). "On asymptotic properties of bootstrap for AR(1) processes," *Journal of Statistical Planning and Inference*, **53**, 361-374.

Dickey, D. A. and W. A. Fuller (1979). "Distribution of estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, **74**, 427-431.

Dickey, D.A. and W.A. Fuller (1981). "Likelihood ratio statistics for autoregressive time series with a unit root," *Econometrica*, **49**, 1057-1072.

Einmahl, U. (1987a). Strong invariance principles for partial sums of independent random vectors. *Annals of Probability* **15** 1419-1440.

Einmahl, U. (1987b). A useful estimate in the multidimensional invariance principle. *Probability Theory and Related Fields* **76** 81-101.

Einmahl, U. (1989). Extensions of results of Komlós, Major and Tusnády to the multivariate case. *Journal of Multivariate Analysis* **28** 20-68.

Elliott, G., T.J. Rothenberg and J.H. Stock (1996). "Efficient tests for an autoregressive unit root," *Econometrica*, **64**, 813-836.

Evans, G.B.A. and N.E. Savin (1981). "Testing for unit roots: 1," *Econometrica*, **49**, 753-779.

Evans, G.B.A. and N.E. Savin (1984). "Testing for unit roots: 2," *Econometrica*, **52**, 1241-1269.

Ferretti, N. and J. Romo (1996). "Unit root bootstrap tests for AR(1) models," *Biometrika*, **83**, 849-860.

Fuller, W.A. (1996). *Introduction to Statistical Time Series*, 2nd ed. New York: Wiley.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* New York: Springer-Verlag.

Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and Its Application.* New York: Academic Press.

Hida, T. (1980). *Brownian Motion.* New York: Springer-Verlag.

Horowitz, J. (2001). "The bootstrap," in *Handbook of Econometrics*, Vol. 5, ed. by J.J. Heckman and E.E. Leamer. Amsterdam: Elsevier, 3159-3228.

Kurtz, T.G. and P. Protter (1991). "Weak limit theorems for stochastic integrals and stochastic differential equations," *The Annals of Probability*, **19**, 1035-1070.

Móricz, F. (1976). "Moment inequalities and the strong law of large numbers," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **35**, 298-314.

Nankervis, J.C. and N.E. Savin (1996). "The level and power of the bootstrap t-test in the AR(1) model with trend," *Journal of Business and Economic Statistics*, **14**, 161-168.

Park, J.Y. (2002). "An invariance principle for sieve bootstrap in time series," *Econometric Theory*, **18**, 469-490.

Park, J.Y. and P.C.B. Phillips (1999). "Asymptotics for nonlinear transformations of integrated time series," *Econometric Theory*, **15**, 269-298.

Phillips, P.C.B. (1987). "Time series regression with a unit root," *Econometrica*, **55**, 277-301.

Phillips, P.C.B. and V. Solo (1992). "Asymptotics for linear processes," *Annals of Statistics*, **20**, 971-1001.

Pollard, D. (1984). *Convergence of Stochastic Processes.* New York: Springer-Verlag.

Revuz, D. and Yor, M. (1994). *Continuous Martingale and Brownian Motion*, 2nd ed. New York: Springer-Verlag.

Skorohod, A.V. (1965). *Studies in the Theory of Random Processes.* Reading, MA: Addison-Wesley.

Stock, J.H. (1994). "Unit roots and structural breaks," in *Handbook of Econometrics*, Vol. 4, ed. by R.F. Engle and D. McFadden. Amsterdam: Elsevier, 2739-2841.

Table 1: Rejection Probabilities for Tests with Fitted Mean

		Asymptotic Tests		Bootstrap Tests	
n	β	F_n	G_n	F_n	G_n
		Normal Innovations			
25	0.4	0.063	0.084	0.050	0.055
	0.0	0.060	0.080	0.049	0.053
	–0.4	0.068	0.084	0.054	0.053
50	0.4	0.056	0.064	0.049	0.049
	0.0	0.057	0.065	0.049	0.050
	–0.4	0.057	0.064	0.048	0.050
100	0.4	0.053	0.057	0.050	0.050
	0.0	0.051	0.047	0.048	0.049
	–0.4	0.057	0.059	0.053	0.052
		Shifted Chi-Square Innovations			
25	0.4	0.059	0.083	0.047	0.052
	0.0	0.061	0.080	0.051	0.051
	–0.4	0.060	0.075	0.049	0.050
50	0.4	0.056	0.064	0.049	0.052
	0.0	0.060	0.066	0.052	0.052
	–0.4	0.055	0.064	0.049	0.050
100	0.4	0.052	0.056	0.049	0.049
	0.0	0.053	0.056	0.050	0.049
	–0.4	0.053	0.056	0.049	0.049
		Mixed-Normal Innovations			
25	0.4	0.061	0.084	0.050	0.054
	0.0	0.065	0.082	0.053	0.055
	–0.4	0.062	0.077	0.050	0.049
50	0.4	0.056	0.063	0.050	0.049
	0.0	0.060	0.065	0.053	0.052
	–0.4	0.057	0.064	0.051	0.051
100	0.4	0.057	0.058	0.053	0.051
	0.0	0.052	0.056	0.049	0.049
	–0.4	0.054	0.059	0.050	0.052

Table 2: Rejection Probabilities for Tests with Fitted Time Trend

		Asymptotic Tests		Bootstrap Tests	
n	β	F_n	G_n	F_n	G_n
Normal Innovations					
25	0.4	0.081	0.128	0.051	0.058
	0.0	0.084	0.121	0.054	0.056
	–0.4	0.082	0.112	0.054	0.053
50	0.4	0.063	0.082	0.048	0.051
	0.0	0.066	0.081	0.050	0.051
	–0.4	0.064	0.077	0.048	0.050
100	0.4	0.059	0.068	0.051	0.054
	0.0	0.054	0.061	0.047	0.048
	–0.4	0.061	0.065	0.052	0.051
Shifted Chi-Square Innovations					
25	0.4	0.082	0.130	0.054	0.060
	0.0	0.077	0.116	0.048	0.051
	–0.4	0.073	0.104	0.046	0.047
50	0.4	0.066	0.083	0.051	0.054
	0.0	0.067	0.080	0.052	0.052
	–0.4	0.066	0.078	0.052	0.051
100	0.4	0.057	0.062	0.050	0.050
	0.0	0.058	0.064	0.050	0.050
	–0.4	0.060	0.064	0.053	0.051
Mixed-Normal Innovations					
25	0.4	0.088	0.135	0.055	0.067
	0.0	0.082	0.119	0.052	0.054
	–0.4	0.079	0.109	0.053	0.053
50	0.4	0.065	0.084	0.049	0.055
	0.0	0.064	0.077	0.050	0.048
	–0.4	0.063	0.075	0.048	0.050
100	0.4	0.059	0.065	0.051	0.051
	0.0	0.055	0.063	0.048	0.049
	–0.4	0.056	0.060	0.049	0.048