

MIT Open Access Articles

*Second Order Expansion of the T-Statistic in AR(1) Models*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Mikusheva, Anna. "Second Order Expansion of the T-Statistic in AR(1) Models." *Econometric Theory* 31.3 (2015): 426–448.

**As Published:** <http://dx.doi.org/10.1017/s0266466614000383>

**Publisher:** Cambridge University Press

**Persistent URL:** <http://hdl.handle.net/1721.1/105179>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike



# Second Order Expansion of the $t$ -statistic in AR(1) Models

by Anna Mikusheva <sup>1</sup>

MIT, Department of Economics

## Abstract

The purpose of this paper is to differentiate between several asymptotically valid methods for confidence set construction for the autoregressive coefficient in AR(1) models. We show that the non-parametric grid bootstrap procedure suggested by Hansen (1999) achieves a second order refinement in the local-to-unity asymptotic approach when compared with a modified version of Stock's (1991) and Andrews' (1993) grid testing procedures. We establish a second order expansion of the  $t$ -statistic in an AR(1) model in the local-to-unity asymptotic approach, which differs drastically from the usual Edgeworth-type expansions by approximating the statistic around a non-standard and non-pivotal limit.

**Key Words:** autoregressive process, confidence set, local-to-unity asymptotics, bootstrap

## 1 Introduction

The paper examines the issue of inferences on the persistence parameter, the autoregressive coefficient  $\rho$ , in AR(1) models. The classic Wald confidence interval typically provides low coverage in finite samples, especially if the true value of  $\rho$  is close to unity, as happens for most of macroeconomic time series. The Wald-type interval is based on classical asymptotic theory, that is, the setup when  $|\rho| < 1$  is considered to be fixed, and the sample size  $n$  converges to infinity. The classical asymptotic laws (Central Limit Theorem and Law of Large Numbers) do not hold uniformly over the interval  $\rho \in (0, 1)$ ; rather, the convergence becomes slower as  $\rho$  approaches 1, and neither law holds for  $\rho = 1$ . An alternative asymptotic approach, local-to-unity asymptotics, considers sequences of models with  $\rho_n = 1 + c/n$  as  $n$  goes to infinity.

---

<sup>1</sup>Financial support from the Castle-Krob Career Development Chair and Sloan Research Fellowship is gratefully acknowledged. I am grateful to Jim Stock, Marcelo Moreira, Gary Chamberlain, Peter Phillips, Denis Chetverikov and three anonymous referees for helpful comments. E-mail for correspondence: amikushe@mit.edu.

According to Mikusheva (2007) and Andrews and Guggenberger (2009, 2010) local-to-unity asymptotics leads to uniform inferences on  $\rho$ , whereas classical asymptotics does not.

There are at least three methods that can be used to construct an asymptotically correct confidence set for  $\rho$ : the method based on the local-to-unity asymptotic approach (a modification of a procedure suggested in Stock (1991)), the parametric grid bootstrap (Andrews (1993)) and the non-parametric grid bootstrap (Hansen (1999)). The validity of these methods was proved in Mikusheva (2007).

This paper compares three methods on the grounds of the accuracy of the asymptotic approximations they provide. All three methods are asymptotically first-order correct, that is, the coverage of the confidence sets uniformly converges to the confidence level as the sample size increases. The question we address here is the speed of the convergence in the local-to-unity asymptotic approach. We show that the non-parametric grid bootstrap (Hansen's method) achieves second-order refinement, that is, the speed of coverage probability convergence is  $o(n^{-1/2})$ , whereas the other two methods in general guarantee only a  $O(n^{-1/2})$  speed of convergence in the local-to-unity asymptotic approach. To compare the three methods we establish an asymptotic expansion of the  $t$ -statistic around its limit in local-to-unity asymptotics.

A second-order distributional expansion is an approximation of the unknown distribution function of the statistic of interest ( $t$ -statistic in our case) by some other function up to the order of  $o(n^{-1/2})$ . One example of a second-order distributional expansion is the first two terms of the well-known Edgeworth expansion.

There are several differences between the expansion obtained in this paper and an Edgeworth expansion. First of all, an Edgeworth expansion is an expansion around a normal or  $\chi^2$  distributions. In our case we expand the  $t$ -statistic around its local-to-unity asymptotic limit, which is a non-normal and non-pivotal distribution. Secondly, it is known that the first two terms of an Edgeworth expansion do not constitute a distribution function themselves. In particular, it can be non-monotonic and non-changing from 0 to 1. One special feature of our expansion is that it approximates the distribution function of the  $t$ -statistic by a *cumulative distribution function* (cdf),

which can be simulated easily.

And finally, as opposed to the Edgeworth expansion - which comes from expanding the characteristic function - our expansion comes from stochastic embedding and a strong approximation principle. The expansion we obtain is a “probabilistic” one. That is, we construct a random variable on the same probability space as the  $t$ -statistic in such a way that the difference between the constructed variable and the  $t$ -statistic is of the order  $o(n^{-1/2})$  in probability. We also show that under additional moment assumptions it leads to a second-order “distributional” expansion. The idea of asymptotically expanding the distribution of the normalized coefficient in the AR(1) with a unit root was first developed in Phillips (1987b) for a Gaussian model, and in Phillips (1987a) for a non-stationary VAR. The same idea was used in Park (2003). He obtained a second-order expansion of the Dickey-Fuller  $t$ -statistic for testing a unit root without assuming normality of error terms.

The distributional expansion allows us to show that Hansen’s grid bootstrap achieves the second-order improvement in the local-to-unity setting when compared to the modified version of Stock’s (1991) and Andrews’ (1993) methods. The intuition for the improvement achieved by the non-parametric grid bootstrap is the classical one - Hansen’s grid bootstrap uses the information about the distribution of error terms, while the other two methods do not. We should be clear that the statement of the second-order superiority of Hansen’s bootstrap has been established in this paper only in the local-to-unity asymptotic framework, and it remains unknown whether the derived asymptotic expansion holds uniformly over all values of  $\rho$ . It seems that the full uniformity result cannot be established by the method used in this paper.

The current paper also discusses that the grid bootstrap does not achieve the asymptotic refinement in a more general  $AR(p)$  case, as in such a case, the statistic becomes asymptotically non-pivotal and depends on the other unknown coefficients describing short-term dynamics.

The paper contributes to the literature on bootstrapping autoregressive processes and closes the discussion on making inferences on persistence in an AR(1) model. Here are some of the known results on bootstrap of AR models: Bose (1988) showed

in classical asymptotics that the usual bootstrap provides second-order improvement compared to the OLS asymptotic distribution. However, Basawa et al. (1991) showed the usual bootstrap fails (is of the wrong size asymptotically) if the true process has a unit root. Their result can be easily generalized to local-to-unity sequences. Park (2006) showed that the usual bootstrap achieves greater accuracy than the asymptotic normal approximation of the  $t$ -statistic for weakly integrated sequences (for sequences with an AR coefficient converging to the unit root at a speed slower than  $1/n$ ). The intuition behind Park's result is that the ordinary bootstrap uses the information about the closeness of the AR coefficient to the unit root. His expansion is non-standard, and the reason for bootstrap improvement is also unusual (usually the bootstrap achieves higher efficiency due to usage information about the distribution of the error term).

The rest of the paper is organized in the following way. Section 2 introduces notation. Section 3 obtains a probabilistic embedding of error terms and a probabilistic expansion of the  $t$ -statistic. Section 4 shows that the probabilistic expansion from the previous section leads to a distributional expansion. Section 5 establishes a similar expansion for a bootstrapped statistic and obtains the main result of the paper on the asymptotic refinement achieved by the non-parametric grid bootstrap. Section 6 discusses the behavior of the grid bootstrap in an  $AR(p)$  case. All proofs are left to the Appendix.

## 2 Notation and preliminary results

Let us have a sample  $\{y_1, \dots, y_n\}$  from an AR(1) process

$$y_j = \rho y_{j-1} + \varepsilon_j, \quad j = 1, \dots, n \quad (1)$$

with  $y_0 = 0$  and  $\rho = \rho_0$ . Let the error terms  $\varepsilon_j$  satisfy Assumptions A below.

**Assumptions A.** Assume that error terms  $\varepsilon_j$  are independent and identically distributed (i.i.d.) random variables with mean zero, variance  $\sigma^2$  and  $\mathbb{E}|\varepsilon_j|^r < \infty$  for some  $r > 2$ .

We consider testing and confidence set construction procedures based on the  $t$ -statistics. Let

$$t(y, \rho_0, n) = \frac{\sum_{j=1}^n (y_j - \rho_0 y_{j-1}) y_{j-1}}{\hat{\sigma} \sqrt{\sum_{j=1}^n y_{j-1}^2}}$$

be the  $t$ -statistic for testing the true hypothesis  $H_0 : \rho = \rho_0$  using the sample  $\{y_j\}_{j=1}^n$ , here  $\hat{\sigma} = \frac{1}{n} \sum_{j=1}^n (y_j - \hat{\rho} y_{j-1})^2$ , and  $\hat{\rho}$  is the OLS estimator of  $\rho$ . The classical asymptotic approach states that for every fixed  $|\rho_0| < 1$  as the sample size,  $n$ , increases to infinity we have

$$t(y, \rho_0, n) \Rightarrow N(0, 1).$$

An alternative asymptotic framework, a local-to-unity asymptotic approach, which is intended to describe the behavior of the statistics when the autoregressive coefficient  $\rho_0$  is very close to the unit root, models the true value of  $\rho_0$  as changing with the sample size, namely,  $\rho = \rho_n = \exp c/n \approx 1 + \frac{c}{n}$ ,  $c \leq 0$ . Under such an assumption, one can show (see Phillips (1987c)), that

$$t(y, \rho_n, n) \Rightarrow \frac{\int_0^1 J_c(x) dw(x)}{\sqrt{\int_0^1 J_c(x) dx}},$$

where  $J_c(x) = \int_0^x e^{c(x-s)} dw(s)$  is an Ornstein-Uhlenbeck (OU) process, and  $w(\cdot)$  is a standard Brownian motion.

As was shown in Mikusheva (2007), the classical asymptotic approximation is not uniform. In particular, if  $z_\alpha$  is the  $\alpha$ -quantile of a standard normal distribution, then

$$\lim_{n \rightarrow \infty} \inf_{|\rho| < 1} \mathbb{P}_\rho \{z_{\alpha/2} < t(y, \rho, n) < z_{1-\alpha/2}\} < 1 - \alpha.$$

As a result, the usual OLS confidence set would provide poor coverage in finite samples if we allow  $\rho$  to be arbitrarily close to the unit root. The local-to-unity asymptotic approach on the contrary is uniform (Mikusheva (2007), Theorem 2). Namely,

$$\lim_{n \rightarrow \infty} \sup_{\rho \in [0, 1]} \sup_x |\mathbb{P}_\rho \{t(y, \rho, n) \leq x\} - F_{n, \rho}^c(x)| = 0,$$

where  $F_{n, \rho}^c(x) = \mathbb{P}\{\int_0^1 J_c(t) dw(t) / \sqrt{\int_0^1 J_c^2(t) dt} \leq x\}$  with  $c = n \log(\rho)$ .

The use of a local-to-unity asymptotic in order to construct a confidence set was suggested by Stock (1991). It can be implemented as a “grid” procedure. One

needs to test a set of hypotheses  $H_0 : \rho = \rho_0$  (in practice the testing could be performed over a fine grid of values of  $\rho_0$ ). A test compares  $t$ -statistic  $t(y, \rho_0, n)$  with critical values that are quantiles of the distribution of  $F_{n, \rho_0}^c(x)$ . The acceptance set is a uniformly asymptotically valid confidence set. We call this method a modified Stock's (1991) method, since the testing procedure stated above slightly differs from the test originally suggested in Stock (1991). The paper by Stock (1991) uses a different statistic, namely, Dickey-Fuller  $t$ -statistic for a unit root test. As discussed in Phillips (2012), Stock's original procedure is poorly centered and has asymptotically zero coverage when applied to a stationary process, while the modified version of it is uniformly asymptotically the correct size. Phillips (2012) explains the bad properties of Stock's original method as a failure of tightness and the escape of probability mass.

Two alternatives to the procedure described above are Andrews' parametric grid bootstrap and Hansen's non-parametric grid bootstrap. The three methods differ in their choices of critical values. In particular, in Andrews' grid bootstrap, critical values are taken as quantiles of a finite-sample distribution of the  $t$ -statistic in a model with normal errors:  $F_{n, \rho_0}^N(x) = \mathbb{P}_{\rho_0}\{t(z, \rho_0, n) \leq x\}$ . Here,  $z_t$  is an AR(1) process with the AR coefficient  $\rho_0$  and normal errors. In Hansen's grid bootstrap we use quantiles of  $F_{n, \rho}^*(x)$ , the finite sample distribution of the  $t$ -statistic for a bootstrapped model with the null imposed. More accurately, let  $y_j^* = \rho_0 y_{j-1}^* + e_j^*$ , where  $e_j^*$  are sampled from the residuals of the initial OLS regression, then  $F_{n, \rho_0}^*(x) = \mathbb{P}_{\rho_0}\{t(y^*, \rho_0, n) \leq x\}$ .

Previously, Mikusheva (2007) proved that all three methods are uniformly asymptotically correct. The goal of this paper is to explore the second-order properties of the methods using the local-to-unity asymptotic approach. Below we show that Hansen's bootstrap provides second-order improvement in the local-to-unity asymptotic approach when compared with Andrews' method and the modification of Stock's method. To prove this we establish a second-order expansion of  $t(y, \rho_n, n)$  in a local-to-unity asymptotic framework.

### 3 Stochastic embedding

According to Skorokhod's embedding scheme (Skorokhod (1965)), the normalized partial sums of the error terms can be realized as a stopped Brownian motion. Namely, there exists a Brownian motion  $w$  and an increasing sequence of stopping times  $T_{j,n}$  on an extended probability space such that

$$\left\{ \frac{1}{\sqrt{n}\sigma} \sum_{i=1}^j \varepsilon_i \right\}_{j=1}^n =^d \{w(T_{n,j})\}_{j=1}^n, \quad (2)$$

where  $T_{n,j} = \frac{1}{n\sigma^2} \sum_{i=1}^j \tau_i$ . It is also known that the random variables  $\tau_i$  are non-negative,  $\mathbb{E}\tau_j = \sigma^2$ ,  $\mathbb{E}|\tau_j|^{r/2} < K_r \mathbb{E}|\varepsilon_j|^r$ , where  $K_r$  is an absolute constant. Since we are interested only in finite-sample *distributions* of statistics generated from  $\{\varepsilon_j\}_{j=1}^n$ , to simplify the notation we assume from now on that  $\frac{\varepsilon_j}{\sigma} = \sqrt{n}(w(T_{n,j}) - w(T_{n,j-1}))$ . Let us consider a sequence of random vectors  $v_j = \left(\frac{\varepsilon_j}{\sigma}, \frac{\tau_j - \sigma^2}{\sigma^2}, \frac{\varepsilon_j^2 - \sigma^2}{\sigma^2}\right)$  and  $B_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nt \rfloor} v_j = (w_n(t), V_n(t), U_n(t))$ . Park (2003) proved<sup>2</sup> that  $B_n \rightarrow^d B = (w, V, U)$ , where  $B$  is a Brownian motion with covariance matrix  $\Sigma$  given by

$$\Sigma = \begin{pmatrix} 1 & \mu_3/3\sigma^3 & \mu_3/\sigma^3 \\ \mu_3/3\sigma^3 & \kappa/\sigma^4 & (\mu_4 - 3\sigma^4 + 3\kappa)/6\sigma^4 \\ \mu_3/\sigma^3 & (\mu_4 - 3\sigma^4 + 3\kappa)/6\sigma^4 & (\mu_4 - \sigma^4)/\sigma^4 \end{pmatrix}. \quad (3)$$

Here  $\mathbb{E}\varepsilon_j^2 = \sigma^2$ ,  $\mathbb{E}\varepsilon_j^3 = \mu_3$ ,  $\mathbb{E}\varepsilon_j^4 = \mu_4$ ,  $\mathbb{E}(\tau_j - \sigma^2)^2 = \kappa$ . Park (2003) also proved that  $B_n$  and  $B$  can be defined on the same probability space in such a way that  $B_n \rightarrow^{a.s.} B$ . Let  $N(t) = w(1+t) - w(1)$ ,  $M(t)$  be a Brownian motion independent on  $w$ . Also denote  $U = U(1)$  and  $V = V(1)$ . We are ready to introduce the second order probabilistic expansions of the  $t$ -statistic.

**Theorem 1** *Let  $\rho_n = \exp\{c/n\}$ ,  $c \leq 0$ . Assume that the  $\varepsilon_j$  satisfy the set of assumptions A with  $r \geq 8$ . Then one has the following probabilistic expansions, that is, there exists a realization of stochastic processes such that:*

---

<sup>2</sup>We use slightly different notation: our third component  $\frac{\varepsilon_j^2 - \sigma^2}{\sigma^2}$  equals to  $\delta_i + 2\eta_i$  in Park's notation. This changes the definition of process  $U(t)$ , which now corresponds to a process, which in Park's notation is referred to as  $V + 2U$ .



(a) the following statement holds uniformly over  $k \in \{1, 2, \dots, n\}$

$$\frac{y_k}{\sigma\sqrt{n}} - J_c(T_{n,k}) = -\frac{c}{\sqrt{n}} \int_0^{k/n} e^{c(k/n-s)} J_c(s) dV(s) + o_p(n^{-1/2});$$

(b)

$$\begin{aligned} \frac{1}{n\sigma^2} \sum_{k=1}^n y_{k-1} \varepsilon_k &= \int_0^1 J_c(x) dw(x) + n^{-1/4} J_c(1) M(V) + \\ &+ \frac{1}{\sqrt{n}} \left( -c \int_0^1 \int_0^t e^{c(t-s)} J_c(s) dV(s) dw(t) + J_c(1) N(V) + \frac{1}{2} M^2(V) - \frac{1}{2} U \right) + o_p\left(\frac{1}{\sqrt{n}}\right); \end{aligned}$$

(c)

$$\begin{aligned} \frac{1}{n^2\sigma^2} \sum_{k=1}^n y_k^2 &= \int_0^1 J_c^2(x) dx - \frac{2c}{\sqrt{n}} \int_0^1 J_c(x) \int_0^x e^{c(x-s)} J_c(s) dV(s) dx - \\ &- \frac{1}{\sqrt{n}} \int_0^1 J_c^2(x) dV(x) + \frac{1}{\sqrt{n}} J_c^2(1) V - \frac{2\mu_3}{3\sqrt{n}\sigma^3} \int_0^1 J_c(t) dt + o_p\left(\frac{1}{\sqrt{n}}\right); \end{aligned}$$

(d)

$$\begin{aligned} \frac{1}{n^{3/2}\sigma} \sum_{k=1}^n y_k &= \int_0^1 J_c(x) dx - \frac{c}{\sqrt{n}} \int_0^1 \int_0^x e^{c(x-s)} J_c(s) dV(s) dx - \\ &- \frac{1}{\sqrt{n}} \int_0^1 J_c(x) dV(x) + \frac{1}{\sqrt{n}} J_c(1) V - \frac{\mu_3}{3\sqrt{n}\sigma^3} + o_p\left(\frac{1}{\sqrt{n}}\right); \end{aligned}$$

(e)

$$t(y, \rho_n, n) = t^c + n^{-1/4} f + n^{-1/2} g + o_p(n^{-1/2}),$$

$$\text{here } t^c = \int_0^1 J_c(x) dw(x) / \sqrt{\int_0^1 J_c^2(x) dx}, \quad f = J_c(1) M(V) / \sqrt{\int_0^1 J_c^2(x) dx},$$

$$\begin{aligned} g &= -\frac{t^c}{2} U + \frac{1}{\sqrt{\int_0^1 J_c^2(x) dx}} \left\{ \left( -c \int_0^1 \int_0^t e^{c(t-s)} J_c(s) dV(s) dw(t) + J_c(1) N(V) + \frac{1}{2} M^2(V) - \frac{1}{2} U \right) \right. \\ &\quad \left. + \frac{t^c}{2} \left( 2c \int_0^1 J_c(x) \int_0^x e^{c(x-s)} J_c(s) dV(s) dx + \int_0^1 J_c(x) dV(x) - J_c(1) V + \frac{2\mu_3}{3\sigma^3} \int_0^1 J_c(t) dt \right) \right\}. \end{aligned}$$

The expansions from Theorem 1 are probabilistic. Namely, we approximate a random variable  $t(y, \rho_n, n)$ , whose distribution is unknown, by another random variable  $\xi_n$  (whose distribution is known or could be simulated) with accuracy  $o(n^{-1/2})$  in probability:  $\mathbb{P}\{|\xi_n - t(y, \rho_n, n)| > \epsilon n^{-1/2}\} \rightarrow 0$ . Probabilistic expansions are not of interest by themselves (since they are abstract constructions); rather, they are building blocks to achieve the distributional expansions described in the next section.

The random variables on the right-hand side are functionals of several Brownian motions  $B(t) = (w(t), V(t), U(t))$  and  $M(t)$ . The covariance matrix of  $B(t)$  depends only on some characteristics  $(\sigma^2, \mu_3, \mu_4, \kappa)$  of the distribution function of  $\varepsilon_j$ , namely on the first four moments of  $\varepsilon_j$  and some characterization of non-normality  $\kappa$  (parameters are defined above).  $M(t)$  is independent of  $B(t)$ . As a result the distribution of the approximating variable depends only on  $\psi = (\sigma^2, \mu_3, \mu_4, \kappa, c)$ . The distribution of the approximating variable can be simulated easily.

**Remark 1** *If one has an exact unit root ( $c = 0$ ), then the expansion is exactly equal to the expansion obtained by Park (2003).*

**Remark 2** *If  $\varepsilon_j$  are normally distributed, then  $V(t) \equiv 0$  and  $w(\cdot)$  is independent of  $U(\cdot)$ . It implies that  $t = t^c + \frac{1}{2\sqrt{n}} \frac{U}{\sqrt{\int J_c^2(x) dx}} - \frac{t^c}{2\sqrt{n}} U + o_p(n^{-1/2})$ , where  $U$  is independent of  $w$ .<sup>3</sup> So, according to this probabilistic expansion Andrews' method and the modification of Stock's method are the same up to an independent summand of order  $O_p(n^{-1/2})$ . I show in the next section that they are the same distributionally up to the order of  $o(n^{-1/2})$ .*

**Remark 3** *The statement of Theorem 1 can be easily generalized to the model with a constant. The corresponding  $t$ -statistic would involve de-meaned processes and would be expanded around its asymptotic limit  $t_\mu^c = \int_0^1 J_c^\mu(x) dw(x) / \sqrt{\int_0^1 (J_c^\mu)^2(x) dx}$ , where  $J_c^\mu(t) = J_c(t) - \int_0^1 J_c(s) ds$  is the de-meaned  $O-U$  process. The resulting expansion has additional terms which are due to estimation of the mean of the process and are derived in part (d) of Theorem 1.*

**Remark 4** *Theorem 1 can be easily generalized to an  $AR(1)$  process with a non-zero starting point.<sup>4</sup> In particular, assume that  $y_0$  is some known random variable. Then  $y_j = \rho^j y_0 + x_j$ , where process  $x_j$  satisfies all assumptions of Theorem 1 and starts from*

---

<sup>3</sup>Peter C.B. Phillips reported (via private communication) a similar expansion for the normalized OLS estimator of  $\rho$  in a Gaussian mode.

<sup>4</sup>I am grateful to Peter C.B. Phillips for pointing this out.

$x_0 = 0$ . Probabilistic expansions for process  $y_t$  have additional terms. In particular,

$$\begin{aligned} \frac{1}{n\sigma^2} \sum_{k=1}^n y_{k-1} \varepsilon_k &= \frac{1}{n\sigma^2} \sum_{k=1}^n x_{k-1} \varepsilon_k + \frac{1}{\sqrt{n}} \left( \sum_{k=1}^n e^{ck/n} \frac{\varepsilon_k}{\sigma\sqrt{n}} \right) \frac{y_0}{\sigma} = \\ &= \frac{1}{n\sigma^2} \sum_{k=1}^n x_{k-1} \varepsilon_k + \frac{y_0}{\sigma\sqrt{n}} \int_0^1 e^{cs} dw(s) + o_p(n^{-1/2}), \end{aligned}$$

$$\begin{aligned} \frac{1}{n^2\sigma^2} \sum_{k=1}^n y_k^2 &= \frac{1}{n^2\sigma^2} \sum_{k=1}^n x_k^2 + \frac{2}{\sqrt{n}} \left( \frac{1}{n} \sum_{k=1}^n e^{ck/n} \frac{y_k}{\sigma\sqrt{n}} \right) \frac{y_0}{\sigma} + o_p(n^{-1/2}) = \\ &= \frac{1}{n^2\sigma^2} \sum_{k=1}^n x_k^2 + \frac{2y_0}{\sigma\sqrt{n}} \int_0^1 e^{cs} J_c(s) ds + o_p(n^{-1/2}). \end{aligned}$$

## 4 Distributional expansion

For making inferences we need asymptotic theory to approximate the unknown finite-sample distribution of the  $t$ -statistic  $t(y, n, \rho_n)$ . In the previous section we established a probabilistic approximation. In particular, we found a sequence of random variables  $\xi_n$  with a known distribution that depends on a vector of parameters  $\psi$  such that  $t(y, n, \rho_n) = \xi_n + o_p(n^{-1/2})$  for  $\rho_n = 1 + c/n$ . That is,

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\rho_n} \left\{ |t(y, n, \rho_n) - \xi_n| > \frac{\epsilon}{\sqrt{n}} \right\} = 0 \text{ for all } \epsilon > 0.$$

The goal of this section is to come up with a distributional expansion. By distributional expansion of the second order we mean a sequence of real-value functions  $G_n(\cdot)$  such that

$$\mathbb{P}_{\rho_n} \{t(y, n, \rho_n) \leq x\} = G_n(x) + o(n^{-1/2}). \quad (4)$$

In general,  $G_n(\cdot)$  is not required to be a cdf of any random variable.

An example of a distributional expansion is the second-order Edgeworth expansion. Initially, an Edgeworth expansion was stated as an approximation to the distribution of normalized sums of random variables. Nowadays, Edgeworth-type expansions have been obtained for many statistics having a normal or chi-squared limiting distribution. Traditionally, Edgeworth-type expansions are obtained from expansions of characteristic functions. It is also known that usually, in Edgeworth expansions,

function  $G_n$  is *not* a cdf of any random variable. In particular,  $G_n$  is not monotonic in many applications.

In our setup an Edgeworth expansion does not exist since the limiting distribution is not normal or chi-squared. In this section we show that under some moment conditions our probabilistic expansion corresponds to a distributional expansion. Namely,

$$\sup_x |\mathbb{P}_{\rho_n} \{t(y, n, \rho_n) \leq x\} - \mathbb{P} \{\xi_n \leq x\}| = o(n^{-1/2}),$$

where  $\xi_n = t^c + n^{-1/4}f + n^{-1/2}g$  from part (e) of Theorem 1. That is, in our case  $G_n(x) = \mathbb{P} \{\xi_n \leq x\}$  is a cdf and depends on the parameter vector  $\psi$ .

**Definition 1 (Park(2003))** *A random variable  $X$  has a distributional order  $o(n^{-a})$  if  $\mathbb{P}\{|X| > n^{-a}\} < n^{-a}$ .*

**Theorem 2** *Let all assumptions of Theorem 1 hold, then all  $o_p(n^{-1/2})$  terms in statements (a)-(e) of Theorem 1 are of distributional order  $o(n^{-1/2})$ .*

**Corollary 1** *If error terms are i.i.d. with mean zero and 8 finite moments, the following distributional expansion holds:*

$$\sup_x |\mathbb{P}\{t(y, \rho_n, n) < x\} - \mathbb{P}\{t^c + n^{-1/4}f + n^{-1/2}g < x\}| = o(n^{-1/2}).$$

One can notice there is no “unique” distributional expansion even if we require that  $G_n$  be a cdf. This surprising fact is explained in the remark below.

**Remark 5** *Let  $G_n(x) = \mathbb{P}\{\xi_n < x\}$  be a cdf and assume that  $\eta$  has a normal distribution and is independent of  $\sigma$ -algebra  $\mathcal{A}$ . Let  $\xi_n$  and  $F$  be random variables measurable with respect to  $\mathcal{A}$ . If  $G_n$  satisfies the distributional approximation (4), then  $\tilde{G}_n(x) = \mathbb{P}\{\xi_n + F \frac{1}{\sqrt{n}}\eta < x\}$  would also satisfy it. That is, the additional term (which is of probabilistic order of  $O_p(n^{-1/2})$ ) has a distributional impact of order  $o(n^{-1/2})$ . This point was made by Park (2003). The idea is that the characteristic function for  $\xi_n + F \frac{1}{\sqrt{n}}\eta$  conditional on  $\mathcal{A}$  is equal to  $e^{it\xi_n}$  up to the order  $O(n^{-1})$ .*

It might seem strange that the probabilistic expansion of  $\sum y_{j-1}\varepsilon_j$  includes a term of order  $O_p(n^{-1/4})$ . This term has a distributional impact of order  $O(n^{-1/2})$ . The idea of the statement is totally parallel to the remark above. Indeed,  $M(V)$  is distributionally  $M(1) \cdot \sqrt{|V|}$ , where  $M(1) \sim N(0,1)$  and is independent of  $B(\cdot) = (w, V, U)$ .

**Remark 6** *Combining Remarks 2 and 5 one comes up with the following. If the error terms are normally distributed then we have a distributional equivalence*

$$\mathbb{P}\{t(z, n, \rho) < x\} = \mathbb{P}\{t^c < x\} + o(n^{-1/2}).$$

*That is, the difference between quantiles constructed in Andrews' (1993) method and the modification of Stock's (1991) method is of the order  $o(n^{-1/2})$ . The two methods achieve the same accuracy up to the second order.*

## 5 Bootstrapped expansion

### 5.1 Embedding for bootstrapped statistic

In section 4 we assert that the distribution of  $t$ -statistic  $t(y, n, \rho_n)$  can be approximated by a sequence of functions  $G_n(x) = \mathbb{P}\{t^c + \frac{1}{n^{1/4}}f + \frac{1}{\sqrt{n}}g \leq x\}$ , where  $f$  and  $g$  are functionals of Brownian motions  $B(\cdot)$  and  $M$ . The covariance structure of  $B$  is described in (3), it depends on  $\psi = (\sigma^2, \mu_3, \mu_4, \kappa)$ , while  $M$  is independent of  $B$ .

The grid bootstrapped statistic (as in Hansen (1999)) has the same form, since it uses the true value (not estimator) of  $\rho_0$  (or  $c$ ). The only difference between the initial distribution of the  $t$ -statistic and the grid bootstrapped distribution of the  $t$ -statistic is a difference in the distribution of the error term. We will show that

$$\mathbb{P}_n^*\{t(y^*, n, \rho) \leq x\} = G_n^*(x) + o(n^{-1/2}) \quad \mathbb{P} - a.s.,$$

here  $\mathbb{P}_n^*$  is the bootstrapped distribution with error terms drawn from  $n$  re-centered residuals, it is conditional on the realization of the initial sample. Function  $G_n^*(x) = \mathbb{P}\{t^c + \frac{1}{n^{1/4}}f^* + \frac{1}{\sqrt{n}}g^* \leq x\}$  is the approximating distribution function, here  $f^*$  and

$g^*$  are functionals of Brownian motions  $B_n^*$  and  $M_n^*$ , where  $M_n^*$  is independent of  $B_n^*$ . The only difference with the initial statistic is that the covariance structure of  $B^*$  depends on the sample moments  $\widehat{\psi}_n = (\widehat{\sigma}_n^2, \widehat{\mu}_{3,n}, \widehat{\mu}_{4,n}, \widehat{\kappa}_n)$  rather than true moments  $\psi$ .

The next subsection states that the parameter vector  $\widehat{\psi}_n$  converges almost surely to  $\psi$  at a speed of  $O_p(n^{-1/2})$ , which would be enough to say that the second-order terms in the expansions of the initial and the grid bootstrapped statistics coincide up to the order of  $o(n^{-1/2})$  almost surely.

**Theorem 3** *Let us have an AR(1) process (1) with  $y_0 = 0$  and error terms satisfying Assumptions A with  $r \geq 8$ . Assume that  $\rho_n = 1 + c/n, c \leq 0$ . Let us consider for every  $n$  a process  $y_j^* = \rho_n y_{j-1}^* + e_j^*, y_0^* = 0$ , where  $e_j^*$  are an i.i.d. sample from centered and normalized residuals from the initial regression. Then*

$$\sup_x |\mathbb{P}\{t(y, n, \rho_n) \leq x\} - \mathbb{P}_n^*\{t(y^*, n, \rho_n) \leq x\}| = o(n^{-1/2}) \quad \mathbb{P} - a.s.$$

Theorem 3 states that Hansen's grid bootstrap provides second-order improvement when compared with Andrews' method and the modification of Stock's method in the local-to-unity asymptotic approach. The intuition for that improvement is typical for the bootstrap. The second-order term depends on the parameters of the distribution of error terms. Those parameters are well approximated by the sampled analogues. The non-parametric grid bootstrap uses sampled residuals, the parameters of which are very close to the population values. As a result, the refinement is achieved. The only parameter (on which the limiting expansion depends) that could not be well estimated is the local-to-unity parameter  $c$ . The grid bootstrap procedure uses the "true" value of  $c$ .

Theorem 3 is a statement obtained using the local-to-unity asymptotic approach. The statement that Hansen's grid bootstrap achieves second-order refinement in the classical asymptotics is easy to make. It could be obtained from an Edgeworth expansion along the lines suggested in Bose (1988). As a result, we should advise applied researchers to choose Hansen's grid bootstrap over Andrews' method and the modification of Stock's method.

## 5.2 Convergence of parameters

This subsection is a part of the proof of Theorem 3 from the previous subsection. Here we show that the parameter vector  $\psi = (\sigma^2, \mu_3, \mu_4, \kappa)$  can be well approximated by a sample analog (moments of residuals)  $\widehat{\psi}_n = (\widehat{\sigma}_n^2, \widehat{\mu}_{3,n}, \widehat{\mu}_{4,n}, \widehat{\kappa}_n)$ . The part of the statement pertaining to the moments of error terms, namely, that vector  $(\widehat{\sigma}_n^2, \widehat{\mu}_{3,n}, \widehat{\mu}_{4,n})$  converges to vector  $(\sigma^2, \mu_3, \mu_4)$  is the standard one and holds if error term  $\varepsilon_j$  has enough moments.

However, one parameter,  $\kappa$ , may potentially depend on the way the Skorokhod embedding has been realized. There are numerous methods to construct a Brownian motion  $w$  and a stopping time  $\tau$  in such a way that the stopped process  $w(\tau)$  has the same distribution as a given mean-zero random variable  $\varepsilon$  and  $\mathbb{E}\tau = \mathbb{E}\varepsilon^2$ . Paper by Obloj (2004) provides a comprehensive survey of different approaches to realize the Skorokhod embedding. Since we are free to choose any construction, from now on I assume that we use the construction stated in Skorokhod (1965). In the proof of Lemma 1 below, I show that the initial Skorokhod construction published in Skorokhod's book (1965) leads to  $\kappa = \mathbb{E}\tau^2 = \frac{5}{3}\mathbb{E}\varepsilon^4$ . This expression ties  $\kappa$  to the fourth moment of the random variable  $\varepsilon$ , as a result, as long as  $\widehat{\mu}_{4,n}$  converges to  $\mu_4$ , the same holds for  $\widehat{\kappa}_n$  and  $\kappa$ .

**Lemma 1** *Let error terms  $\varepsilon_j$  satisfy the set of Assumptions A. Then there is a Skorokhod's embedding for which*

$$\psi - \widehat{\psi}_n = O_p(n^{-1/2}).$$

## 6 Some notes about $AR(p)$ processes

A natural question is whether the results about the second order asymptotic refinement of the grid bootstrap can be generalized to the  $AR(p)$  models. While asymptotic expansions analogous to those obtained in Theorems 1 and 2 can be established in the more general case of  $AR(p)$  models<sup>5</sup>, the result on the second order asymptotic

---

<sup>5</sup>This derivation is beyond the scope of the current paper.

refinement achieved by the grid bootstrap does not hold.

Assume that we have a process written here in the augmented Dickey-Fuller form:

$$y_j = \rho y_{j-1} + \beta_1 \Delta y_{j-1} + \dots + \beta_{p-1} \Delta y_{j-p+1} + \varepsilon_j. \quad (5)$$

Denote  $|\mu_1| \leq |\mu_2| \leq \dots \leq |\mu_p|$  to be the autoregressive roots of process (5). A typical way of modeling a near-unit-root process is to assume that  $\mu_p = 1 + \frac{c}{n}$  while the other roots  $\mu_1, \dots, \mu_{p-1}$  are held fixed and strictly separated from the unit circle ( $|\mu_{p-1}| < \delta < 1$ ).

There are several ways to characterize the persistence of an  $AR(p)$  process. The two most often used characteristics are the largest autoregressive root  $\mu_p$  and the sum of the autoregressive coefficients  $\rho$ . In the case of the unit root testing they coincide: testing for the presence of a unit root  $\mu_p = 1$  is equivalent to testing that the sum of autoregressive coefficients is equal to unity ( $\rho = 1$ ). However, in the case of near unit roots the two characteristics diverge, in particular:

$$\rho = 1 - (1 - \mu_p)(1 - \mu_{p-1}) \dots (1 - \mu_1) = 1 + \frac{c}{n} (1 - \mu_{p-1}) \dots (1 - \mu_1) = 1 + \frac{c^*}{n},$$

where  $c^* = c(1 - \mu_{p-1}) \dots (1 - \mu_1)$ .

There are two very insightful discussions by Phillips (1991) and Andrews and Chen (1994) that provide arguments in favor of using the sum of the autoregressive coefficients,  $\rho$ , as the measure of persistence. In particular, Phillips (1991) points out that the spectrum at zero is equal to  $\frac{\sigma^2}{(1-\rho)^2}$ , while Andrews and Chen (1994) notice that the cumulative impulse response, that is, the sum of all impulse responses, is equal to  $\frac{1}{1-\rho}$ , that is, both these measures of long-run behaviour of a process are directly linked to the sum of coefficients. Below I mention three additional arguments as to why one may prefer to use the sum of coefficients  $\rho$  rather than the largest root  $\mu_p$  as the characteristic of persistence. First, while an estimator of  $\rho$  can be obtained by the usual OLS regression, an estimator of  $\mu_p$  is a very complicated function of the OLS coefficients; this function cannot be written explicitly analytically for cases with  $p \geq 5$ . Second, the largest root is not always well defined and potentially may be a complex number, as with positive probability the two largest (in absolute value) roots



are complex conjugates. Finally, if the order of the autoregression  $p$  is not known and one estimates the OLS regressions of an increasing order ( $p_n \rightarrow \infty$  as  $n \rightarrow \infty$ ), so called sieve regression, then the roots are not consistently estimated (while the OLS estimator for  $\rho$  is still consistent). In particular, Onatski and Uhlig (2012) showed that even if the data is generated from the white noise one will find an asymptotically infinite number of roots concentrated around the unit circle (spurious unit roots).

For the rest of this section we assume that the sum of the AR coefficients,  $\rho$ , is the parameter of interest. The grid bootstrap procedure in such a case was introduced by Hansen (1999) and was proved to be uniformly asymptotically correct by Mikusheva (2007). Below is a brief description of the grid bootstrap for  $AR(p)$ .

Assume that we test the hypothesis  $H_0 : \rho = \rho_0$  using the corresponding  $t$ -statistic in the regression (5). The finite-sample distribution of the  $t$ -statistic depends on  $\rho_0$  and  $n$ , as well as on the unknown nuisance parameters  $\beta_1, \dots, \beta_{p-1}$ . Let us estimate the following OLS regression:

$$y_j - \rho_0 y_{j-1} = \beta_1 \Delta y_{j-1} + \dots + \beta_{p-1} \Delta y_{t-p+1} + \varepsilon_j,$$

that is, we regress the quasi-difference  $y_j - \rho_0 y_{j-1}$  on  $\Delta y_{t-1}, \dots, \Delta y_{t-p+1}$ . Assume that  $\widehat{\beta}_1(\rho_0), \dots, \widehat{\beta}_{p-1}(\rho_0)$  are the OLS coefficients from such a regression. The bootstrapped samples are generated in the following way:

$$y_j^* = \rho_0 y_{j-1}^* + \widehat{\beta}_1(\rho_0) \Delta y_{j-1}^* + \dots + \widehat{\beta}_{p-1}(\rho_0) \Delta y_{t-p+1}^* + \varepsilon_j^*,$$

where  $\varepsilon_j^*$  are simulated from the re-centered residuals. The bootstrapped  $t$ -statistic is obtained by running a regression (5) on the simulated data  $y^*$ . By repeating the simulations many times, one obtains the distribution of the bootstrapped  $t$ -statistic, which she can use to get the critical values for the initial test. These critical values depend on the value  $\rho_0$  tested. To construct a confidence set one has to repeat the procedure for different values of  $\rho_0$ , and choose those for which the corresponding hypothesis is accepted.

Assume that  $\mu_p = 1 + c/n$  while  $\mu_1, \dots, \mu_{p-1}$  are fixed and separated from the unit circle, and the hypothesis of interest  $H_0 : \rho = \rho_0$  holds true, that is,  $\rho_0 =$

$1 - (1 - \mu_p)(1 - \mu_{p-1}) \dots (1 - \mu_1)$ . Then according to Phillips (1987),  $t(y, \rho_0, n) \Rightarrow t^c$ , where  $t^c$  is defined in Theorem 1 (e). Notice that the null hypothesis  $H_0 : \rho = \rho_0$  pins down the value of  $c^* = n(\rho_0 - 1)$ , while the limit distribution depends on the value of  $c$ . Obviously, the values of  $c$  and  $c^*$  are closely related, but the relation between them depends on the other coefficients of the AR process,  $\beta_1, \dots, \beta_{p-1}$ , which are nuisance parameters, that is, unknown parameters not specified by the null hypothesis. From this perspective the  $t$ -statistic for  $\rho$  is not asymptotically pivotal, and thus, it is difficult to expect the bootstrap to provide asymptotic refinement. Notice also that the two cases when the grid bootstrap achieves asymptotic refinement, namely the  $AR(1)$  and testing for the unit root, correspond to the situation when  $\rho = \mu_p$  (or  $c = c^*$ ) and thus the  $t$ -statistic is asymptotically pivotal.

## 7 References

Andrews, D.W.K. (1993): “Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models,” *Econometrica*, 61(1), 139-165.

Andrews, D.W.K. and H-Y. Chen (1994): “Approximately Median-Unbiased Estimation of Autoregressive Models,” *Journal of Business & Economic Statistics*, 12(2), 187-204.

Andrews, D.W.K. and P. Guggenberger (2009): “Hybrid and Size-corrected Subsample Methods,” *Econometrica*, 77(3), 721-762.

Andrews, D.W.K. and P. Guggenberger (2010): “Asymptotic Size and a Problem with Subsampling and with the  $m$  out of  $n$  Bootstrap,” *Econometric Theory*, 26, 426-468.

Breiman, L. (1992) *Probability*, Addison-Wesley series in statistics: Reading, Mass.

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves and R.L. Taylor (1991): “Bootstrapping Unstable First-Order Autoregressive Processes,” *Annals of Statistics*, 19(2), 1098-1101.

Bose, A. (1988): “Edgeworth Correction by Bootstrap in Autoregressions,” *Annals of Statistics*, 16(4), 1709-1722.

- Hansen, B.E. (1999): “The Grid Bootstrap and the Autoregressive Model,” *Review of Economics and Statistics*, 81(4), 594-607.
- Kurtz, T.G. and P. Protter (1991): “Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations,” *Annals of Probability*, 19(3), 1035-1070.
- Mikusheva, A. (2007): “Uniform Inference in Autoregressive Models,” *Econometrica*, 75(5), 1411-1452.
- Obloj, J (2004): “The Skorokhod Embedding Problem and its Offspring,” *Probability Surveys*, 1, 321-392.
- Onatski, A. and H. Uhlig (2012): “Unit Roots in White Noise,” *Econometric Theory*, 28 (3), 485-508.
- Park, J.Y. (2003): “Bootstrap Unit Root Tests,” *Econometrica*, 71(6), 1845-1895.
- Park, J.Y. (2006): “A Bootstrap Theory for Weakly Integrated Processes,” *Journal Of Econometrics*, 133(2), 639-672.
- Phillips, P.C.B. (1987a): “Asymptotic Expansions in Nonstationary Vector Autoregressions,” *Econometric Theory*, 3(1), 45-68.
- Phillips, P.C.B. (1987b): “Time Series Regression with a Unit Root,” *Econometrica*, 55(2), 277-301.
- Phillips, P.C.B. (1987c): “Toward a Unified Asymptotic Theory for Autoregression,” *Biometrika*, 74(3), 535-547.
- Phillips, P.C.B. (1991): “Bayesian Routes and Unit Roots: De Rebus Prioribus Semper Est Disputandum,” *Journal of Applied Econometrics*, 6, 435-473
- Phillips, P.C.B. (2012): “On Confidence Intervals for Autoregressive Roots and Predictive Regression,” *unpublished manuscript*.
- Skorokhod, A.V. (1965): *Studies in the Theory of Random Processes*, Reading, MA: Addison-Wesley.
- Stock, J. (1991): “Confidence Intervals for the Largest Autoregressive Root in US Macroeconomic Time Series,” *Journal of Monetary Economics*, 28, 435-459.

## 8 Appendix. Proofs of results

We use the following results from Park (2003):

**Lemma 2** (Park (2003), Lemma 3.5(a))

If Assumptions A are satisfied with  $r \geq 8$ , then

$$\frac{1}{\sqrt{n}\sigma} \sum_{j=1}^n \varepsilon_j = w(1) + n^{-1/4}M(V) + n^{-1/2}N(V) + o_p(n^{-1/2}),$$

where  $V = V(1)$ .

**Lemma 3 (Park(2003))** If Assumptions A are satisfied with  $r > 4$  then we might choose  $B_n$  and  $B$  such that

$$\mathbb{P} \left\{ \sup_{0 \leq t \leq 1} |B_n(t) - B(t)| > c \right\} \leq n^{1-r/4} C^{-r/2} (1 + \sigma^{-r}) K(1 + \mathbb{E}|\varepsilon_j|^r)$$

for any  $C \geq n^{-1/2+2/r}$ .

About convergence of stochastic integrals:

**Lemma 4** (Kurtz and Protter (1991)) For each  $n$ , let  $(X_n, Y_n)$  be an  $\mathcal{F}_t^n$ - adapted process with sample paths in Skorokhod space  $D$  and let  $Y_n$  be an  $\mathcal{F}_t^n$  semimartingale. Suppose that  $Y_n = M_n + A_n + Z_n$ , where  $M_n$  is a local  $\mathcal{F}_t^n$  martingale,  $A_n$  is  $\mathcal{F}_t^n$ - adapted finite variation process, and  $Z_n$  is constant except for finitely many discontinuities. Let  $N_n(t)$  denote the number of discontinuities of process  $Z_n$  on interval  $[0, t]$ . Suppose that  $N_n$  is stochastically bounded for each  $t > 0$ . Suppose that for each  $\alpha > 0$  there exist stopping times  $\{\tau_n^\alpha\}$  such that  $\mathbb{P}\{\tau_n^\alpha \leq \alpha\} \leq 1/\alpha$  and  $\sup_n \mathbb{E}[[M_n]_{t \wedge \tau_n^\alpha} + T_{t \wedge \tau_n^\alpha}(A_n)] < \infty$ .

If  $(X_n, Y_n, Z_n) \rightarrow^d (X, Y, Z)$  in the Skorokhod topology, then  $Y$  is a semimartingale with respect to a filtration to which  $X$  and  $Y$  are adapted and  $(X_n, Y_n, \int X_n dY_n) \rightarrow^d (X, Y, \int X dY)$  in the Skorokhod topology. If  $(X_n, Y_n, Z_n) \rightarrow (X, Y, Z)$  in probability, then convergence in probability holds in the conclusion.

**Proof of Theorem 1.**

(a) From our stochastic embedding,  $\frac{\varepsilon_j}{\sigma} = \sqrt{n}(w(T_{n,j}) - w(T_{n,j-1}))$ , and the definition of the O-U process,  $J_c(t) = \int_0^t e^{c(t-s)} dw(s)$ , we have that

$$\begin{aligned} & \frac{y_k}{\sigma\sqrt{n}} - J_c(T_{n,k}) = \\ &= \sum_{j=1}^k e^{c\frac{k-j}{n}} (w(T_{n,j}) - w(T_{n,j-1})) - \sum_{j=1}^k \int_{T_{n,j-1}}^{T_{n,j}} e^{c(T_{n,k}-s)} dw(s) = \\ &= \sum_{j=1}^k \left( e^{c\frac{k-j}{n}} - e^{c(T_{n,k}-T_{n,j})} \right) (w(T_{n,j}) - w(T_{n,j-1})) + \\ & \quad + \sum_{j=1}^k \int_{T_{n,j-1}}^{T_{n,j}} (e^{c(T_{n,k}-T_{n,j})} - e^{c(T_{n,k}-s)}) dw(s). \end{aligned}$$

We show below that the first term in the last sum is asymptotically equal to  $-\frac{c}{\sqrt{n}} \int_0^{k/n} e^{c(k/n-s)} J_c(s) dV(s) + o_p(n^{-1/2})$ , while the second term is  $o_p(n^{-1/2})$ . We start with the first term and notice that  $|e^x - 1 - x - x^2| \leq x^3$  for all  $|x| < 1$ . According to Breiman (1992) chapter 13.4, we also have

$$\mathbb{P} \left\{ \max_{1 \leq j \leq k \leq n} \left| (T_{n,k} - T_{n,j}) - \frac{k-j}{n} \right| > 1 \right\} \rightarrow 0.$$

These two statements imply that

$$\begin{aligned} & \sum_{j=1}^k \left( e^{c\frac{k-j}{n}} - e^{c(T_{n,k}-T_{n,j})} \right) (w(T_{n,j}) - w(T_{n,j-1})) = \\ &= c \sum_{j=1}^k e^{c\frac{k-j}{n}} \left( \frac{k-j}{n} - (T_{n,k} - T_{n,j}) \right) (w(T_{n,j}) - w(T_{n,j-1})) + \\ &+ c^2 \sum_{j=1}^k e^{c\frac{k-j}{n}} \left( \frac{k-j}{n} - (T_{n,k} - T_{n,j}) \right)^2 (w(T_{n,j}) - w(T_{n,j-1})) + R_{1,n,k}, \end{aligned}$$

where

$$\begin{aligned} |\tilde{R}_{1,n,k}| &\leq \sum_{j=1}^k c^3 e^{c\frac{k-j}{n}} \left| \frac{k-j}{n} - (T_{n,k} - T_{n,j}) \right|^3 |w(T_{n,j}) - w(T_{n,j-1})| = \\ &= \sum_{j=1}^k c^3 e^{c\frac{k-j}{n}} \left| \frac{1}{\sqrt{n}} V\left(\frac{k}{n}\right) - \frac{1}{\sqrt{n}} V\left(\frac{j}{n}\right) \right|^3 \left| \frac{\varepsilon_j}{\sigma\sqrt{n}} \right| \leq \\ &\leq \frac{1}{n^{3/2}} \left( \sum_{j=1}^k c^3 e^{c\frac{k-j}{n}} \left| \frac{\varepsilon_j}{\sigma\sqrt{n}} \right| \right) \cdot \sup_{0 \leq t, s \leq 1} |V_n(t) - V_n(s)|^3 = O_p(n^{-1}). \end{aligned}$$

The last statement holds uniformly over  $k$ .

$$\begin{aligned}
& c \sum_{j=1}^k e^{c \frac{k-j}{n}} \left( \frac{k-j}{n} - (T_{n,k} - T_{n,j}) \right) (w(T_{n,j}) - w(T_{n,j-1})) = \\
& = \frac{c}{\sqrt{n}} \sum_{j=1}^k e^{c \frac{k-j}{n}} (V_n(k/n) - V_n(j/n)) (w(T_{n,j}) - w(T_{n,j-1})) = \\
& = \frac{c}{\sqrt{n}} \sum_{j=1}^k e^{c \frac{k-j}{n}} \sum_{i=j+1}^k (V_n(i/n) - V_n((i-1)/n)) (w(T_{n,j}) - w(T_{n,j-1})) = \\
& = \frac{c}{\sqrt{n}} \sum_{i=1}^k \sum_{j=1}^{i-1} e^{c \frac{k-j}{n}} (V_n(i/n) - V_n((i-1)/n)) (w(T_{n,j}) - w(T_{n,j-1})) = \\
& = \frac{c}{\sqrt{n}} \sum_{i=1}^k e^{c \frac{k-i}{n}} (V_n(i/n) - V_n((i-1)/n)) \frac{y_i}{\sqrt{n}} = \\
& = \frac{c}{\sqrt{n}} \int_0^{k/n} e^{c(k/n-s)} J_{c,n}(s) dV_n(s) = \\
& = \frac{c}{\sqrt{n}} \int_0^{k/n} e^{c(k/n-s)} J_c(s) dV(s) + o_p(n^{-1/2}).
\end{aligned}$$

The last line in the long equality is due to Lemma 4. Here we used that  $V_n$  is the second component of  $B_n$ ,  $J_{c,n}(i/n) = \frac{y_i}{\sqrt{n}} = \int_0^{i/n} e^{c(\frac{i}{n}-s)} dw_n(s)$ , and  $B_n \rightarrow B$  *a.s.*

The next asymptotic statement can be obtained by analogous considerations:

$$\begin{aligned}
& c^2 \sum_{j=1}^k e^{c \frac{k-j}{n}} \left( \frac{k-j}{n} - (T_{n,k} - T_{n,j}) \right)^2 (w(T_{n,j}) - w(T_{n,j-1})) = \\
& = \frac{c^2}{n} \sum_{j=1}^k e^{c \frac{k-j}{n}} (V_n(k/n) - V_n(j/n))^2 (w(T_{n,j}) - w(T_{n,j-1})) = O_p(n^{-1}).
\end{aligned}$$

The only statement we are left to prove is

$$\sum_{j=1}^k \int_{T_{n,j-1}}^{T_{n,j}} (e^{c(T_{n,k}-T_{n,j-1})} - e^{c(T_{n,k}-s)}) dw(s) = o_p(n^{-1/2}).$$

Notice that random variables  $\xi_{n,j} = \int_{T_{n,j-1}}^{T_{n,j}} (e^{c(1-T_{n,j-1})} - e^{c(1-s)}) dw(s)$  are not correlated across  $j$ ,  $\mathbb{E}\xi_{n,j} = 0$  and  $\mathbb{E}\xi_{n,j}^2 \leq \text{const} \cdot n^{-3}$ . As a result,  $\sum_{j=1}^k \xi_{n,j}$  is  $o_p(n^{-1/2})$  both probabilistically and distributionally uniformly over  $k$ .

(b) Let  $A_n(k/n) = \sqrt{n} \left( \frac{y_{k-1}}{\sigma\sqrt{n}} - J_c(T_{n,k-1}) \right)$ . According to Lemma 4 and statement

(a) proved above, we have

$$\sqrt{n} \sum_{k=1}^n \left( \frac{y_{k-1}}{\sigma\sqrt{n}} - J_c(T_{n,k-1}) \right) \frac{\varepsilon_k}{\sigma\sqrt{n}} = \int_0^1 A_n(t) dw_n(t) \xrightarrow{p} -c \int_0^1 \int_0^t e^{c(t-s)} J_c(s) dV(s) dw(t).$$

Next consider the following equality:

$$\begin{aligned} & \sum_{k=1}^n J_c(T_{n,k-1}) \frac{\varepsilon_k}{\sigma\sqrt{n}} - \int_0^1 J(s)dw(s) = \\ & = \int_1^{T_{n,n}} J_c(s)dw(s) - \sum_{k=1}^n \int_{T_{n,k-1}}^{T_{n,k}} (J_c(s) - J_c(T_{n,k-1}))dw(s). \end{aligned} \quad (6)$$

Here and below we write  $\int_1^{T_{n,n}} J_c(s)dw(s)$  as a short-cut for

$$\int_0^{T_{n,n}} J_c(s)dw(s) - \int_0^1 J_c(s)dw(s) = \begin{cases} \int_1^{T_{n,n}} J_c(s)dw(s), & \text{if } T_{n,n} > 1; \\ -\int_{T_{n,n}}^1 J_c(s)dw(s), & \text{if } T_{n,n} < 1. \end{cases}$$

Consider the second term on the right hand side of equation (6). By definition of the O-U process  $J(s) = w(s) + c \int_0^s J(t)dt$ , as a result, we have

$$\begin{aligned} \sum_{k=1}^n \int_{T_{n,k-1}}^{T_{n,k}} (J_c(s) - J_c(T_{n,k-1}))dw(s) &= \sum_{k=1}^n \int_{T_{n,k-1}}^{T_{n,k}} (w(s) - w(T_{n,k-1}))dw(s) + \\ &+ c \sum_{k=1}^n \int_{T_{n,k-1}}^{T_{n,k}} (D(s) - D(T_{n,k-1}))dw(s), \end{aligned}$$

where  $D(s) = \int_0^s J_c(t)dt$ . According to Ito's lemma and Lemma 3, we obtain the following statement:

$$\int_{T_{n,k-1}}^{T_{n,k}} (w(s) - w(T_{n,k-1}))dw(s) = \frac{\varepsilon_k^2}{2n} - \frac{T_{n,k} - T_{n,k-1}}{2} = \frac{1}{2\sqrt{n}}(U(1) - V(1)) + o_p(n^{-1/2}).$$

Term  $\sum_{k=1}^n \int_{T_{n,k-1}}^{T_{n,k}} (D(s) - D(T_{n,k-1}))dw(s)$  is the sum of uncorrelated identically distributed random variables with mean zero and variance of asymptotic order  $O_p(n^{-3})$ . As a result, this term is of order  $o(n^{-1/2})$  in probability and distributionally.

Now, consider the first term on the right hand side of equation (6). From Ito's lemma we know that  $d(J_c^2(x)) = 2J_c dw + 2cJ_c^2 dx + dx$ , as a result,

$$\int_1^{T_{n,n}} J_c(s)dw(s) = \frac{1}{2} (J_c^2(T_{n,n}) - J_c^2(1)) - \int_1^{T_{n,n}} (cJ_c^2(x) + \frac{1}{2})dx = \quad (7)$$

$$\begin{aligned} &= J_c(1) (J_c(T_{n,n}) - J_c(1)) + \frac{1}{2} (J_c(T_{n,n}) - J_c(1))^2 - \\ &\quad - \frac{1}{\sqrt{n}}V \cdot \left( cJ_c^2(1) + \frac{1}{2} \right) + o_p(n^{-1/2}) = \end{aligned} \quad (8)$$

$$\begin{aligned} &= J_c(1) \left( w(T_{n,n}) - w(1) + \frac{c}{\sqrt{n}}J_c(1)V \right) + \frac{(w(T_{n,n}) - w(1))^2}{2} - \\ &\quad - \frac{V}{\sqrt{n}} \left( cJ_c^2(1) + \frac{1}{2} \right) + o_p(n^{-1/2}) = \end{aligned} \quad (9)$$

$$= J_c(1) (w(T_{n,n}) - w(1)) + \frac{1}{2} (w(T_{n,n}) - w(1))^2 - \frac{1}{2\sqrt{n}} V + o_p(n^{-1/2}) = \quad (10)$$

$$= n^{-1/4} J_c(1) M(V) + n^{-1/2} \left( J_c(1) N(V) + \frac{1}{2} M^2(V) - \frac{1}{2} V \right) + o_p(n^{-1/2}). \quad (11)$$

Going from equation (7) to equation (8), and from (8) to (9) results in remainder terms  $-c \int_1^{T_{n,n}} (J_c^2(x) - J_c^2(1)) dx$  and  $c J_c(1) \int_1^{T_{n,n}} (J_c(x) - J_c(1)) dx$ . Here we used the same convention regarding integrals of the form  $\int_1^{T_{n,n}}$  as described right after equation (6). It is easy to see that the both remainder terms are of order  $o(n^{-1/2})$  in probability and distributionally. Expansion (11) follows from (10) and the statement of Lemma 2.

Putting everything together we arrive at expansion (b) of Theorem 1.

(c) Using the statement of part (a) we have

$$\begin{aligned} \frac{1}{n} \sum \left( \frac{y_k^2}{n\sigma^2} - J_c^2(T_{n,k}) \right) &= \frac{1}{n} \sum \left( \frac{y_k}{\sigma\sqrt{n}} - J_c(T_{n,k}) \right) \left( 2J_c(T_{n,k}) + \frac{y_k}{\sigma\sqrt{n}} - J_c(T_{n,k}) \right) = \\ &= \frac{1}{n} \sum \left( -\frac{c}{\sqrt{n}} \int_0^{k/n} e^{c(k/n-s)} J_c(s) dV(s) + o_p\left(\frac{1}{\sqrt{n}}\right) \right) \left( 2J_c(T_{n,k}) + o_p\left(\frac{1}{\sqrt{n}}\right) \right) = \\ &= -\frac{2c}{\sqrt{n}} \int_0^1 J_c(x) \int_0^x e^{c(x-s)} J_c(s) dV(s) dx + R_n + o_p(n^{-1/2}), \end{aligned}$$

where  $R_n = \frac{1}{\sqrt{n}} \left( \frac{1}{n} \sum_{k=1}^n g(k/n) - \int_0^1 g(x) dx \right)$  and  $g(x) = -c J_c(x) \int_0^x e^{c(x-s)} J_c(s) dV(s)$ .

It is easy to see that  $R_n = o_p(n^{-1/2})$ . As a result,

$$\begin{aligned} \frac{1}{n} \sum \frac{y_k^2}{\sigma^2 n} &= \frac{1}{n} \sum J_c^2(T_{n,k}) - \frac{2c}{\sqrt{n}} \int_0^1 J_c(x) \int_0^x e^{c(x-s)} J_c(s) dV(s) dx + o_p(n^{-1/2}). \\ \frac{1}{n} \sum J_c^2(T_{n,k-1}) - \int_0^1 J_c^2(x) dx &= \sum J_c^2(T_{n,k-1}) \left( \frac{1}{n} - (T_{n,k} - T_{n,k-1}) \right) - \\ &\quad - \sum \int_{T_{n,k-1}}^{T_{n,k}} (J_c^2(t) - J_c^2(T_{n,k-1})) dt + \int_1^{T_{n,n}} J_c^2(t) dt, \quad (12) \end{aligned}$$

here we use the same convention as when we write  $\int_1^{T_{n,n}}$  as described after equation (6).

Let us consider each term in (12) separately. Due to Lemma 4 we have:

$$\begin{aligned} \sum J_c^2(T_{n,k-1}) \left( \frac{1}{n} - (T_{n,k} - T_{n,k-1}) \right) &= -\frac{1}{\sqrt{n}} \int_0^1 J_{c,n}^2(x) dV_n(x) = \\ &= -\frac{1}{\sqrt{n}} \int_0^1 J_c^2(x) dV(x) + o_p(n^{1/2}). \end{aligned}$$



Consider the second summand in equation (12):

$$\begin{aligned}
& \sum \int_{T_{n,k-1}}^{T_{n,k}} (J_c^2(t) - J_c^2(T_{n,k-1})) dt = \\
= & 2 \sum J_c(T_{n,k-1}) \int_{T_{n,k-1}}^{T_{n,k}} (J_c(t) - J_c(T_{n,k-1})) dt + \sum \int_{T_{n,k-1}}^{T_{n,k}} (J_c(t) - J_c(T_{n,k-1}))^2 dt = \\
& = 2 \sum J_c(T_{n,k-1}) \int_{T_{n,k-1}}^{T_{n,k}} (J_c(t) - J_c(T_{n,k-1})) dt + o_p(n^{-1/2}).
\end{aligned}$$

The last asymptotic statement holds because  $\sum \int_{T_{n,k-1}}^{T_{n,k}} (J_c(t) - J_c(T_{n,k-1}))^2 dt$  is a sum of random variables with mean of order  $O(n^{-2})$  and variances of order  $O(n^{-3})$ .

$$\begin{aligned}
& \sum J_c(T_{n,k-1}) \int_{T_{n,k-1}}^{T_{n,k}} (J_c(t) - J_c(T_{n,k-1})) dt = \\
= & \sum J_c(T_{n,k-1}) \int_{T_{n,k-1}}^{T_{n,k}} (w(t) - w(T_{n,k-1})) dt + o_p(n^{-1/2}).
\end{aligned}$$

Lemma A1 part (b) in Park (2003) implies that

$$\mathbb{E} \left( \int_{T_{n,k-1}}^{T_{n,k}} (w(t) - w(T_{n,k-1})) dt \right) = \frac{\mu_3}{3\sigma^3} n^{-3/2}.$$

and  $\mathbb{E} \left( \int_{T_{n,k-1}}^{T_{n,k}} (w(t) - w(T_{n,k-1})) dt \right)^2 = O(n^{-2})$ . As a result,

$$\begin{aligned}
& 2 \sum J_c(T_{n,k-1}) \int_{T_{n,k-1}}^{T_{n,k}} (w(t) - w(T_{n,k-1})) dt = \\
= & \frac{2\mu_3}{3\sigma^3} n^{-3/2} \sum J_c(T_{n,k-1}) + o_p(n^{-1/2}) = \frac{2\mu_3}{\sqrt{n}3\sigma^3} \int_0^1 J_c(t) dt + o_p(n^{-1/2}).
\end{aligned}$$

Now consider the final term in (12). Using the reasoning analogous to that used in the proof of statement (b) and the convention described after equation (6), we obtain

$$\int_1^{T_{n,n}} J_c^2(t) dt = \frac{1}{\sqrt{n}} J_c^2(1)V + o_p(n^{-1/2}).$$

Putting all terms together leads us to statement (c).

(d) Using the statement of part (a)

$$\frac{1}{n} \sum \left( \frac{y_k}{\sigma\sqrt{n}} - J_c(T_{n,k}) \right) = -\frac{c}{\sqrt{n}} \frac{1}{n} \sum \int_0^{k/n} e^{c(k/n-s)} J_c(s) dV(s) + o_p(n^{-1/2}).$$

Lemma 4 implies that

$$\frac{1}{n^{3/2}\sigma} \sum y_k = \frac{1}{n} \sum J_c(T_{n,k}) - \frac{c}{\sqrt{n}} \int_0^1 \int_0^x e^{c(x-s)} J_c(s) dV(s) dx + o_p(n^{-1/2}).$$

We know that the asymptotic limit of  $\frac{1}{n} \sum J_c(T_{n,k-1})$  is  $\int_0^1 J_c(x)dx$ . Consider the higher order terms in this expansion:

$$\begin{aligned} \frac{1}{n} \sum J_c(T_{n,k-1}) - \int_0^1 J_c(x)dx &= \sum J_c(T_{n,k-1}) \left( \frac{1}{n} - (T_{n,k} - T_{n,k-1}) \right) - \\ &\quad - \sum \int_{T_{n,k-1}}^{T_{n,k}} (J_c(t) - J_c(T_{n,k-1}))dt + \int_1^{T_{n,n}} J_c(t)dt \end{aligned} \quad (13)$$

According to Lemma 4, the first summand on the right-hand side in equation (13) equals to

$$\sum J_c(T_{n,k-1}) \left( \frac{1}{n} - (T_{n,k} - T_{n,k-1}) \right) = -\frac{1}{\sqrt{n}} \int_0^1 J_c(x)dV(x) + o_p(n^{-1/2}).$$

Consider the second summand in expansion (13):

$$\begin{aligned} \sum \int_{T_{n,k-1}}^{T_{n,k}} (J_c(t) - J_c(T_{n,k-1}))dt &= \sum \int_{T_{n,k-1}}^{T_{n,k}} (w(t) - w(T_{n,k-1}))dt + \\ + c \sum \int_{T_{n,k-1}}^{T_{n,k}} \int_{T_{n,k-1}}^t J_c(s)dsdt &= \sum \int_{T_{n,k-1}}^{T_{n,k}} (w(t) - w(T_{n,k-1}))dt + o_p(n^{-1/2}). \end{aligned}$$

According to Park (2003) Lemma A1, we have that

$$\mathbb{E} \left( \int_{T_{n,k-1}}^{T_{n,k}} (w(t) - w(T_{n,k-1}))dt \right) = \frac{\mu_3}{3\sigma^3} n^{-3/2},$$

and the variance of  $\int_{T_{n,k-1}}^{T_{n,k}} (w(t) - w(T_{n,k-1}))dt$  is of order  $O_p(n^{-2})$ . So, we have

$\sum \int_{T_{n,k-1}}^{T_{n,k}} (w(t) - w(T_{n,k-1}))dt = \frac{\mu_3}{3\sqrt{n}\sigma^3} + O_p(n^{-1})$ . Finally, the last term in (13) is

$$\int_1^{T_{n,n}} J_c(t)dt = J_c(1)V + o_p(n^{-1/2}).$$

This finishes the proof of statement (d).

Part (e) follows from (a)-(d), a Taylor expansion and from the observation proved in Park (2003) that

$$\hat{\sigma}^2 = \frac{1}{n} \sum \varepsilon_t^2 + O_p(1/n) = \sigma^2 + \frac{1}{\sqrt{n}} \sigma^2 U + o_p(n^{-1/2}).$$

□

**Proof of Theorem 2.** In the proof of Theorem 1 we showed that many remainder terms in our expansions are distributionally of order  $o(n^{-1/2})$ . The only terms for

which we are left to show that they are distributionally of order  $o(n^{-1/2})$ , are those terms for which we previously appealed to the convergence of stochastic integrals (Lemma 4). Here is the comprehensive list of them:

$$\begin{aligned}
R_{2,n,k} &= \frac{c}{\sqrt{n}} \sum_{i=1}^k e^{c \frac{k-i}{n}} (V_n(i/n) - V_n((i-1)/n)) y_{i-1} - \frac{c}{\sqrt{n}} \int_0^{k/n} e^{c(k/n-s)} J_c(s) dV(s); \\
R_{3,n} &= -\frac{c}{\sqrt{n}} \sum_{k=1}^n \int_0^{k/n} e^{c(k/n-s)} J_c(s) dV(s) \frac{\varepsilon_k}{\sqrt{n}} + \frac{c}{\sqrt{n}} \int_0^1 \int_0^t e^{c(t-s)} J_c(s) dV(s) dw(t); \\
R_{4,n} &= -\frac{1}{\sqrt{n}} \left( \sum J_c^2(T_{n,k-1})(V_n(k/n) - V_n(k-1/n)) - \int_0^1 J_c^2(x) dV(x) \right); \\
R_{5,n} &= -\frac{1}{\sqrt{n}} \left( \sum J_c(T_{n,k-1})(V_n(k/n) - V_n(k-1/n)) - \int_0^1 J_c(x) dV(x) \right).
\end{aligned}$$

All these terms have a form of stochastic integrals  $\frac{1}{\sqrt{n}} \int_0^1 \xi(t) d(V(t) - V_n(t))$  or  $\frac{1}{\sqrt{n}} \int_0^1 \xi(t) d(w(t) - w_n(t))$ . Their distributional order would depend on the quadratic variations which have forms of  $\sup_{0 \leq t \leq 1} |V_n(t) - V(t)|^2$  and  $\sup_{0 \leq t \leq 1} |w_n(t) - w(t)|^2$ . The order of the last expressions is determined by Lemma 3.  $\square$

### Proof of Lemma 1.

Let  $\varepsilon$  be a random variable with  $\mathbb{E}\varepsilon = 0$  and  $\mathbb{E}\varepsilon^4 < \infty$ . Then according to Skorokhod's construction as presented in his 1965 book, there exists a Brownian motion  $w$  and a stopping time  $\tau$  such that the stopped Brownian motion  $w(\tau)$  has the same distribution as  $\varepsilon$  and  $\mathbb{E}\tau = \mathbb{E}\varepsilon^2$ . In order to prove Lemma 1 we show that for this specific construction (Skorokhod (1965)) we have  $\mathbb{E}\tau^2 = \frac{5}{3}\mathbb{E}\varepsilon^4$ .

The first step in Skorokhod's construction is the embedding of a random variable  $\xi$  which takes only two values  $a$  and  $b$  with probabilities  $\frac{b}{b-a}$  and  $-\frac{a}{b-a}$  correspondingly, here values  $a$  and  $b$  have opposite signs, and the probabilities are constructed in such a way that  $\mathbb{E}\xi = 0$ . Let  $\tau_{a,b}$  be the smallest root of the equation  $(w(t) - a)(w(t) - b) = 0$ . Then, as shown in Skorokhod (1965, p. 166),  $w(\tau_{a,b})$  has the same distribution as  $\xi$  and the characteristic function for  $\tau_{a,b}$  is

$$\mathbb{E}e^{-\lambda\tau_{a,b}} = \frac{\sinh b\sqrt{2\lambda} - \sinh a\sqrt{2\lambda}}{\sinh(b-a)\sqrt{2\lambda}}. \tag{14}$$

As a result, one can calculate moments of  $\tau_{a,b}$  as

$$(-1)^k \frac{d^k}{d\lambda^k} \mathbb{E}e^{-\lambda\tau_{a,b}} \Big|_{\lambda=0} = \mathbb{E}\tau_{a,b}^k. \tag{15}$$

Let  $F(x) = \mathbb{P}\{\varepsilon \leq x\}$  be the distribution of variable  $\varepsilon$ . Let us define function  $G(x)$  as  $\int_{G(x)}^x y dF(y) = 0$ . Assume also that a Brownian motion  $w$  is independent of  $\varepsilon$ , then the Skorokhod's construction defines  $\tau$  as

$$\tau = \inf\{t : (w(t) - \varepsilon)(w(t) - G(\varepsilon)) = 0\} = \tau_{\varepsilon, G(\varepsilon)}.$$

Skorokhod proves that  $w(\tau)$  has the same distribution as  $\varepsilon$  and  $\mathbb{E}\tau = \mathbb{E}\varepsilon^2$ .

By using formulas (14) and (15), through double differentiation of function (14), one can obtain the following expression

$$\mathbb{E}\tau_{a,b}^2 = \frac{1}{3}(-b^3a + 3b^2a^2 - ba^3) = r(a, b).$$

We can notice that

$$\mathbb{E}\tau^2 = \mathbb{E}(\mathbb{E}(\tau^2|\varepsilon)) = \mathbb{E}[r(\varepsilon, G(\varepsilon))] = \frac{1}{3}\mathbb{E}(-G^3(\varepsilon)\varepsilon + 3G^2(\varepsilon)\varepsilon^2 - G(\varepsilon)\varepsilon^3).$$

Then we make use of the following two facts:  $G(G(x)) = x$  and  $G(x)dF(G(x)) = xdF(x)$  to show that all terms in the last expression result in  $\mathbb{E}\varepsilon^4$ . For example, we can show that  $\mathbb{E}[G^3(\varepsilon)\varepsilon] = -\mathbb{E}\varepsilon^4$ . Indeed,

$$\mathbb{E}[G^3(\varepsilon)\varepsilon] = \int G^3(x)xdF(x) = \int G(x)^3G(x)dF(G(x)) = - \int u^4dF(u) = -\mathbb{E}\varepsilon^4,$$

the sign has changed due to change of limits of integration, since  $u = F(x)$  is a decreasing function of  $x$ . The other terms are done in a similar way. Finally, we arrive to the formula  $\mathbb{E}\tau^2 = \frac{5}{3}\mathbb{E}\varepsilon^4$ .

Since  $\mathbb{E}\varepsilon^8 < \infty$ , by using Chebyshev's inequality one can arrive at the statement of the lemma.  $\square$

**Proof of Theorem 3.** Corollary 1 to Theorem 2 states the distributional expansions for the t-statistic:

$$\sup_x |\mathbb{P}\{t(y, n, \rho_n) \leq x\} - G_n(x)| = o(n^{-1/2}), \quad (16)$$

here  $G_n(x) = \mathbb{P}\{t^c + \frac{1}{n^{1/4}}f + \frac{1}{\sqrt{n}}g \leq x\}$ , where  $f$  and  $g$  are functionals of Brownian motions  $B(\cdot)$ . The covariance structure of  $B$  is described in (3), it depends on  $\psi = (\sigma^2, \mu_3, \mu_4, \kappa)$ .

Examine now the bootstrapped distribution. All statements about them are conditional on realizations of the process  $y$  and formulated as  $\mathbb{P} - a.s.$

First, let us establish the notation. Assume we that have a realization of a sample of size  $n$  of the initial process  $(y_1, \dots, y_n)$ . Bootstrapped errors are generated as i.i.d. from the re-centered residuals  $\{e_1, \dots, e_n\}$ , that is, from the distribution that has moments  $\widehat{\psi}_n = (\widehat{\sigma}_n^2, \widehat{\mu}_{3,n}, \widehat{\mu}_{4,n}, \widehat{\kappa}_n)$ . Here sub-index  $n$  signifies that the sample moments are estimated from a sample of size  $n$ . Let this distribution be described by probability  $\mathbb{P}_n^*$ . Assume we draw a sample  $\varepsilon_1^*, \dots, \varepsilon_m^*$  of size  $m$  from  $\mathbb{P}_n^*$  and produced the bootstrapped sample  $y^* = (y_1^*, \dots, y_m^*)$ . Then Corollary 1 to Theorem 2 establishes the following asymptotic approximation:

$$\sup_x |\mathbb{P}_n^* \{t(y^*, m, \rho_m) \leq x\} - G_{n,m}^*(x)| = o(m^{-1/2}), \quad \mathbb{P} - a.s. \quad (17)$$

here  $G_{n,m}^*(x) = \mathbb{P}\{t^c + \frac{1}{m^{1/4}}f^* + \frac{1}{\sqrt{m}}g^* \leq x\}$  where  $f^*$  and  $g^*$  are functionals of Brownian motions  $B_n^*$  and  $M_n^*$  as described in Theorem 2. The covariance structure of  $B_n^*$  depends on  $\widehat{\psi}_n$  and  $M_n^*$  is independent of  $B_n^*$ .

In order to prove Theorem 3 we will obtain and use the following two statements:

$$\sup_x |\mathbb{P}_n^* \{t(y^*, n, \rho_n) \leq x\} - G_{n,n}^*(x)| = o(n^{-1/2}), \quad \mathbb{P} - a.s., \quad (18)$$

and

$$\sup_x |G_n(x) - G_{n,n}^*(x)| = o(n^{-1/2}), \quad \mathbb{P} - a.s. \quad (19)$$

The latter statement is due to Lemma 1 and the continuity argument. Statements (16), (18) and (19) imply the validity of Theorem 3.

To prove (18) we refine the statement of Corollary 1. Namely, we claim that for any process with error terms satisfying Assumption A with  $r \geq 8$  there exists  $\delta > 0$  such that

$$\sup_x |\mathbb{P}\{t(y, n, \rho_n) \leq x\} - G_n(x)| \leq Const(\mu_8) \cdot n^{-1/2-\delta}, \quad (20)$$

where  $Const(\mu_8)$  is the constant that depends only on the eights moment of the error term  $\mu_8 = \mathbb{E}\varepsilon^8$ . Indeed, all remainder terms described in proofs of Theorems 1 and 2 fall into two categories: sums of independent random variables or the remainders

in the convergence of stochastic integrals (such as  $R_{2,n,k}$ ,  $R_{3,n}$ ,  $R_{4,n}$  and  $R_{5,n}$ ). The refinement of the order of the former comes from a Chebyshev-type inequality. The distributional order of the latter depends on the distributional order of the quadratic variations  $\sup_{0 \leq t \leq 1} |V_n(t) - V(t)|^2$  and  $\sup_{0 \leq t \leq 1} |w_n(t) - w(t)|^2$ , that is established in Lemma 3 with  $r = 8$ .

Refinement (20) of Corollary 1 applied to the bootstrapped distributions gives the following refinement of statement (17):

$$\sup_x |\mathbb{P}_n^* \{t(y^*, m, \rho_m) \leq x\} - G_{n,m}^*(x)| = \text{Const}(\widehat{\mu}_{8,n}) \cdot m^{-1/2-\delta}, \quad \mathbb{P} - a.s.$$

Given that  $\widehat{\mu}_{8,n} \rightarrow \mu_8$  *a.s.* the last statement evaluated at  $m = n$  leads to (18). This finishes the proof of Theorem 3.  $\square$