Massachusetts Institute of Technology
Department of Economics
Working Paper Series

Second Order Expansion of T-Statistic in Autoregressive Models

Anna Mikusheva

Working Paper 07-26
October 12, 2007

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

Second Order Expansion of t-statistic in Autoregressive Models

by Anna Mikusheva [1]

MIT, Department of Economics

Abstract

The purpose of this paper is to receive a second order expansion of the t-statistic in AR(1) model in local to unity asymptotic approach. I show that Hansen's (1998) method for confidence set construction achieves a second order improvement in local to unity asymptotic approach compared with Stock's (1991) and Andrews' (1993) methods.

Key Words: autoregressive process, confidence set, local to unity asymptotics, uniform convergence

1 Introduction

The paper deals with inferences about the persistence parameter (AR coefficient) ρ in AR(1) models. The classical Wald confidence interval typically has low coverage in finite samples, especially if the true value of ρ is close to unity as it happens for most of macroeconomic time series. 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 (CLT and Law of Large Numbers) do not hold uniformly over the interval $\rho \in (0, 1)$, rather the convergence becomes slower as ρ approaches 1, and the both laws do not hold 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. According to Mikusheva (2007) and Andrews and Guggenberger (2007a,b) local to unity asymptotics leads to uniform inferences on ρ, whereas classical asymptotics does not.

There are at least three methods that can be used to construct asymptotically correct confidence set for ρ: method based on the local to unity asymptotic approach

[1] The paper constitutes the second part of my PhD dissertation at Harvard University. I am grateful to my thesis committee Jim Stock, Marcelo Moreira and Gary Chamberlain for helpful comments. E-mail for correspondence: amikushe@mit.edu

(Stock (1991)), parametric grid bootstrap (Andrews (1993)) and non-parametric grid bootstrap (Hansen (1999)). The validity of the methods was proved in Mikusheva (2007).

This paper compares three methods on a ground of accuracy of asymptotic approximation 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 I address is the speed of the convergence. It is well known that Hansen's grid bootstrap achieves a second order refinement in classical asymptotic approach, whereas the two other methods (Andrews' and Stock's) do not. I address the same question in local to unity asymptotic approach. My answer is that the non-parametric grid bootstrap (Hansen's method) achieves the 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 $O(n^{-1/2})$ speed of convergence in local to unity asymptotic approach. To compare the three method I find 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 well-known Edgeworth expansion.

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

And finally, opposed to the Edgeworth' expansion which came from expanding characteristic function, my expansion comes from stochastic embedding and strong approximation principle. The same idea was used in a very inspirational work by

Park(2003a). He obtained a second order expansion of the Dickey-Fuller t-statistic for testing a unit root. The expansion I obtain is "probabilistic" one. That is, I 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. I also show that under additional moment assumptions it leads to a second order "distributional" expansion.

The distributional expansion allows me to show that Hansen's grid bootstrap achieves the second order improvement in local to unity setting compared to Andrews' method and the local to unity asymptotic distribution. The intuition for non-parametric grid bootstrap improvement is the classical one - Hansen's grid bootstrap uses the information about the distribution of error terms.

The paper contributes to the literature on bootstrapping autoregressive processes and closes the discussion on making inferences on persistence in AR(1) model. Here some of the known results on bootstrap of AR models: Bose(1988) showed in classical asymptotics that the usual bootstrap provides the second order improvement compared to the OLS asymptotic distribution. However, Basawa et al(1991) showed the usual bootstrap fails (has asymptotically wrong size) if the true process has a unit root. Their result can be easily generalized to local to unity sequences. Park (2003b) showed that the usual bootstrap achieves higher accuracy than the asymptotic normal approximation of the t-statistic for weakly integrated sequences (for sequences with AR coefficient converging to the unit root with a speed slower than $1/n$). The intuition behind Park's result is that the ordinary bootstrap uses the information about closeness of the AR coefficient to the unit root. His expansion is non-standard and the reason of bootstrap improvement is also not usual (usually bootstrap achieves higher efficiency due to usage information about the distribution of error term).

I get many ideas from a paper by Park (2003a), where he proves the second order improvement of bootstrapped tests for the unit root. He found an asymptotic expansion of t-statistic for the unit root in terms of functionals of Brownian motion. My expansions for local to unity sequences will bear the similar idea to his.

The rest of the paper is organized in the following way. Section 2 introduces nota-

tions. 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 essay. All proofs are left to the Appendix.

2 Notations and preliminary results

Let us have a process

$$y_j = \rho y_{j-1} + \varepsilon_j, \quad j = 1, ..., n \tag{1}$$

We assume that $y_0 = 0$. Error terms ε_j are iid with mean zero, unit variance and finite absolute moment of order r. The procedure of testing and constructing confidence sets is based on the t-statistics. Let

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

be the t-statistic for testing the true value of the parameter ρ using the sample $\{y_j\}_{j=1}^{n}$.

The classical asymptotic approach states that for every fixed $|\rho| < 1$ as n increases to infinity we have

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

According to local to unity asymptotic approach if $\rho_n = 1 + c/n, C \geq 0$

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

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

As it was shown in Mikusheva(2007) the classical asymptotic approximation is not uniform. In particular, if z_α is the α-quantile of standard normal distribution, then

$$\lim_{n\to\infty} \inf_{|\rho|<1} P_\rho\{z_{\alpha/2} < t(y, \rho, n) < z_{1-\alpha/2}\} < 1 - \alpha.$$

As a result, the usual OLS confidence set would have a poor coverage in finite samples if we allow ρ to be arbitrary close to the unit root.

A local to unity asymptotic approach on the contrary is uniform (Mikusheva(2007), Theorem 2). Namely,

$$\lim_{n\to\infty} \sup_{\rho\in[0,1]} \sup_{x} |P_\rho\{t(y,\rho,n) \leq x\} - F^c_{n,\rho}(x)| = 0$$

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

The use of 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 need to test a set of hypothesis $H_0 : \rho = \rho_0$ (in practice the testing could be performed over a fine grid of values of ρ_0). A test compares t-statistic $t(y, \rho_0, n)$ with critical values which are quantiles of the distribution of $F^c_{n,\rho_0}(x)$. The acceptance set is an asymptotic confidence set.

Two alternatives to the procedure above are Andrews' parametric grid bootstrap and Hansen's non-parametric grid bootstrap. The method differ in the choice of critical values. In particular, in Andrews' grid bootstrap critical values are taken as quantiles of finite sample distribution of the t-statistic in a model with normal errors: $F^N_{n,\rho_0}(x) = P_{\rho_0}\{t(z,\rho_0,n) \leq x\}$. Here z_t is AR(1) process with the AR coefficient ρ_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 $y^*_t = \rho_0 y^*_{t-1} + e^*_t$, where e^*_t are sampled from the residuals of the initial OLS regression, then $F^*_{n,\rho_0}(x) = P_{\rho_0}\{t(y^*, \rho_0, n) \leq x\}$.

Previously, Mikusheva (2007) proved that all three methods are uniformly asymptotically correct. My goal is to explore the second order properties of the methods in local to unity asymptotic approach. I will show that Hansen's bootstrap provides the second order improvement in local to unity asymptotic approach. That is, I consider a sequence of models $\rho = \rho_n = \exp\{c/n\}$ as n increases to infinity (this sequence of models is called "nearly integrated" process). The goal is to obtain the second order expansion of $t(y, \rho_n, n)$ along this sequence of models. The next section would be devoted to probabilistic expansion.

3 Stochastic embedding.

Assumptions A. Assume that error terms ε_j are i.i.d. with mean zero, variance $\sigma^2 = 1$ and $E|\varepsilon_j|^r < \infty$ for some $r > 2$.

According to Skorokhod embedding scheme, there exists a Brownian motion w and a sequence of iid variables τ_i on an extended probability space such that the sequence of error terms have the same distribution as a sequence of stopped Brownian motion:

$$\{\varepsilon_j\}_{j=1}^{\infty} =^d \left\{ w\left(\sum_{i=1}^{j} \tau_i\right) - w\left(\sum_{i=1}^{j-1} \tau_i\right)\right\}. \tag{2}$$

It also known that $E\tau_j = \sigma^2 = 1, E|\tau_j|^{r/2} < K_r E|\varepsilon_j|^r$, where K_r is an absolute constant. We define $T_{n,j} = \frac{1}{n}\sum_{i=1}^{j} \tau_i$. 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 $B_n(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]} v_j = (w_n(t), V_n(t), U_n(t))$. Park(2003a) proved that $B_n \to^d B = (w, V, U)$, where B is a Brownian motion with covariance matrix Σ given by

$$\Sigma = \begin{pmatrix} 1 & \mu_3/3\sigma^3 & \mu_3/\sigma^3 \\ \mu_3/3\sigma^3 & \kappa/\sigma^4 & (\mu_4 - 3\sigma^4 + 3\kappa)/12\sigma^4 \\ \mu_3/\sigma^3 & (\mu_4 - 3\sigma^4 + 3\kappa)/12\sigma^4 & (\mu_4 - \sigma^4)/\sigma^4 \end{pmatrix}. \tag{3}$$

Here $E\varepsilon_j^2 = \sigma^2 = 1, E\varepsilon_j^3 = \mu_3, E\varepsilon_j^4 = \mu_4, E(\tau_j - \sigma^2)^2 = \kappa$.

Park(2003a)also proved that B_n and B can be defined on the same probability space in such a way that $B_n \to^{a.s.} B$. Let $N(t) = w(1+t) - w(1)$, $M(t)$ be a Brownian motion independent on w.

Theorem 1 *Let $\rho_n = 1 + c/n, c \le 0$. Assume that ε_j satisfy set of assumptions A with $r \ge 8$, then one has the following probabilistic expansions:*

(a)

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

(b)

$$\frac{1}{n}\sum_{k=1}^{n} y_{k-1}\varepsilon_k = \int_0^1 J_c(x)dw(x) + n^{-1/4}J_c(1)M(V) +$$

$$+\frac{1}{\sqrt{n}}\left(-c\int_0^1\int_0^t e^{c(t-s)}J_c(s)dV(s)dw(t) + J_c(1)N(V) + \frac{1}{2}M^2(V) + \frac{1}{2}U\right) + o_p(\frac{1}{\sqrt{n}})$$

(c)

$$\frac{1}{n^2}\sum y_k^2 = \int_0^1 J_c^2(x)dx - \frac{2c}{\sqrt{n}}\int_0^1 J_c(x)\int_0^x e^{c(x-s)}J_c(s)dV(s)dx-$$
$$-\frac{1}{\sqrt{n}}\int_0^1 J_c^2(x)dV(x) + \frac{1}{\sqrt{n}}J_c^2(1)V + o_p(\frac{1}{\sqrt{n}})$$

(d)

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

(e)

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

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

$$g = \frac{1}{\sqrt{\int_0^1 J_c^2(x)dx}}\left(-c\int_0^1\int_0^t e^{c(t-s)}J_c(s)dV(s)dw(t) + J_c(1)N(V) + \tfrac{1}{2}M^2(V) + \tfrac{1}{2}U\right)$$
$$+t^c\frac{1}{2\sqrt{\int_0^1 J_c^2(x)dx}}\left(2c\int_0^1 J_c(x)\int_0^x e^{c(x-s)}J_c(s)dV(s)dx + \int_0^1 J_c(x)dV(x) - J_c(1)V\right)$$

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 ξ_n (whose distribution is known or could be simulated) with accuracy $o(n^{-1/2})$ in probability: $P\{|\xi_n - t(y, \rho_n, n)| > \epsilon n^{-1/2}\} \to 0$. Probabilistic expansions are not of interest by themselves (since they are abstract constructions), rather they are building blocks in getting 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 ε_j, namely on the first four moments of ε_j and some characterization of non-normality κ (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 easily simulated.

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

Remark 2 *If ε_j are normally distributed, then $V(t) \equiv 0$ and $w(\cdot)$ is independent of $U(\cdot)$. It implies that $t = t^c + \frac{1}{2\sqrt{n}} \frac{U}{\sqrt{\int J_c^2(x)dx}} + o_p(n^{-1/2})$, where U is independent on w. So, according to this probabilistic expansion Stock's and Andrews' methods 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 $o(n^{-1/2})$.*

4 Distributional expansion

For making inferences we need asymptotic theory to approximate the unknown 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 ξ_n with known distribution depending on a vector of parameters ψ (the distribution can be simulated if ψ is known) such that $t(y, n, \rho_n) = \xi_n + o_p(n^{-1/2})$ for $\rho_n = 1 + c/n$. That is,

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

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

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

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, 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 normal 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 Edgeworth expansion does not exist since the limiting distribution is not normal. In this section I show that under some moment conditions our proba-

bilistic expansion corresponds to a distributional expansion. Namely,

$$\sup_x |P_{\rho_n}\{t(y,n,\rho_n) \le x\} - P\{\xi_n \le x\}| = o(T^{-1/2}).$$

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

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

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

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

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

One can notice there is no "unique" distributional expansion even if we require that G_n is a cdf. This surprising fact is explained in the note below.

Remark 3 *Let $G_n(x) = P\{\xi_n < x\}$ be a cdf and assume that η has normal distribution and is independent of σ- algebra $\mathcal{A}$. Let ξ_n and F be measurable with respect to $\mathcal{A}$. If G_n satisfies the distributional approximation (4), then $\widetilde{G}_n(x) = P\{\xi_n + F\frac{1}{\sqrt{n}}\eta < x\}$ would also satisfy it. That is, the additional term (which is of probabilistic order of $O_p(n^{-1/2})$) has distributional impact of order $o(n^{-1/2})$. This point was made by Park(2003a). 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 $\sum y_{j-1}\varepsilon_j$ has term of order $O_p(n^{-1/4})$. This term has distributional impact of order $O(n^{-1/2})$. The idea of the statement is totally parallel to the note 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 4 *Combining Notes 2 and 3 one get the following. If error terms are normally distributed then we have a distributional equivalence*

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

That is the difference between quantiles constructed in Stock's and Andrews' methods 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 got that the distribution of t-statistic $t(y, n, \rho_n)$ could be approximated by a sequence of functions $G_n(x) = P\{t^c + \frac{1}{n^{1/4}} f + \frac{1}{\sqrt{n}} g\}$, where f and g are functionals of Brownian motions $B(\cdot)$ (covariance structure is described in (3), it depends on $\sigma^2, \mu_3, \mu_4, \kappa, c$) and M (independent of B).

The bootstrapped statistic has totally the same form, since it uses the "true value"(not estimator) of ρ(or c). The only difference between the initial distribution of t-statistic and the grid bootstrapped distribution of t-statistic is different distribution of error term.

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

where $G_n^*(x) = P\{t^c + \frac{1}{n^{1/4}} f^* + \frac{1}{\sqrt{n}} g^*\}$ with f^* and g^* are functionals of B^*, M(the same functionals, covariance structure of B^* depends on $\widehat{\sigma}^2, \widehat{\mu}_3, \widehat{\mu}_4, \widehat{\kappa}$)

The next subsection states that the parameter vector $(\widehat{\sigma}^2, \widehat{\mu}_3, \widehat{\mu}_4, \widehat{\kappa})$ converge almost surely to $(\sigma^2, \mu_3, \mu_4, \kappa)$ at a speed of $O_p(n^{-1/2})$, which would be enough to say that the second order terms in expansions of initial and grid bootstrapped statistics coincide up to the order of $o(n^{-1/2})$.

Theorem 3 *Let us have an AR(1) process (1) with $y_0 = 0$ and error terms satisfying Assumptions A with $r \geq 1$. 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 i.i.d. sample from centered and normalized residuals from the initial regression. Then

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

Theorem 3 states that Hansen's grid bootstrap provides the second order improvement compared with Andrews' and Stock's methods in local to unity asymptotic approach. The intuition for that is the usual one. 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 whose parameters 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 local to unity parameter c. The grid bootstrap procedure uses the "true" value of c.

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

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 $\psi = (\sigma^2, \mu_3, \mu_4, \kappa)$ could be well approximated by a sample analog (moments of residuals) $\widehat{\psi} = (\widehat{\sigma}^2, \widehat{\mu}_3, \widehat{\mu}_4, \widehat{\kappa})$.

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

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

The convergence of the third and forth sample moments of residuals to their population analogues with a speed of $O(n^{-1/2})$ is the usual statement. For that we need to require enough moments of error term, 8 moments should be enough.

One parameter, κ, as was discussed above is not intrinsic (it depends on a way the Skorokhod embedding was realized). The fact that $\widehat{\kappa} \to^p E\tau^2$ with speed $O(n^{-1/2})$ is non-trivial mainly because most of known constructions are not explicit and the dependence of moments of τ on distribution of ε is not evident. By messy calculation I got that in the initial Skorokhod construction published in Skorokhod's book(1965) $E\tau^2 = \frac{5}{3}E\xi^4$. That would imply the speed of convergence we need.

6 Appendix. Proofs of results

We use the following results from Park (2003a):

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

If $r \geq 8$, *then*

(a)

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

where $V = V(1)$.

About convergence of stochastic integrals:

Lemma 3 *(Kurtz and Protter) 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 $\mathcal{F}_t^n$ 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 $N_n(t)$ denotes 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 $P\{\tau_n^\alpha \leq \alpha\} \leq 1/\alpha$ and $\sup_n E[[M_n]_{t\wedge\tau_n^\alpha} + T_{t\wedge\tau_n^\alpha}(A_n) < \infty$.*

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

Proof of Theorem 1

(a)

$$\frac{y_k}{\sqrt{n}} - J_c(T_{n,k}) = \sum_{j=1}^{k} e^{c\frac{k-j}{n}} \left(w\left(T_{n,j}\right) - w\left(T_{n,j-1}\right)\right) - \sum_{j=1}^{k} \int_{T_{n,j-1}}^{T_{n,j}} e^{c(T_{n,k}-s)} dw(s) =$$

$$= \sum_{j=1}^{k} \left(e^{c\frac{k-j}{n}} - e^{c(T_{n,k}-T_{n,j})}\right) \left(w\left(T_{n,j}\right) - w\left(T_{n,j-1}\right)\right) +$$

$$+ \sum_{j=1}^{k} \int_{T_{n,j-1}}^{T_{n,j}} \left(e^{c(T_{n,k}-T_{n,j})} - e^{c(T_{n,k}-s)}\right) dw(s) =$$

$$= -\frac{c}{\sqrt{n}} \int_0^{k/n} e^{c(k/n-s)} J_c(s) dV(s) + \widetilde{R}_{1,n} + R_{2,n} + R_{3,n},$$

where we have the following lines of reasoning:

$$\sum_{j=1}^{k} \left(e^{c\frac{k-j}{n}} - e^{c(T_{n,k}-T_{n,j})}\right) \left(w\left(T_{n,j}\right) - w\left(T_{n,j-1}\right)\right) =$$

$$= c \sum_{j=1}^{k} e^{c\frac{k-j}{n}} \left(\frac{k-j}{n} - (T_{n,k} - T_{n,j})\right) \left(w\left(T_{n,j}\right) - w\left(T_{n,j-1}\right)\right) + \widetilde{R}_{1,n} =$$

$$= -\frac{c}{\sqrt{n}} \sum_{j=1}^{k} e^{c\frac{k-j}{n}} \left(V_n(k) - V_n(j)\right) \left(w\left(T_{n,j}\right) - w\left(T_{n,j-1}\right)\right) + \widetilde{R}_{1,n} =$$

$$= -\frac{c}{\sqrt{n}} \sum_{j=1}^{k} e^{c\frac{k-j}{n}} \sum_{i=j+1}^{k} \left(V_n(i) - V_n(i-1)\right) \left(w\left(T_{n,j}\right) - w\left(T_{n,j-1}\right)\right) + \widetilde{R}_{1,n} =$$

$$= -\frac{c}{\sqrt{n}} \sum_{i=1}^{k} \sum_{j=1}^{i-1} e^{c\frac{k-j}{n}} \left(V_n(i) - V_n(i-1)\right) \left(w\left(T_{n,j}\right) - w\left(T_{n,j-1}\right)\right) + \widetilde{R}_{1,n} =$$

$$= -\frac{c}{\sqrt{n}} \sum_{i=1}^{k} e^{c\frac{k-i}{n}} \left(V_n(i) - V_n(i-1)\right) y_{i-1} + \widetilde{R}_{1,T} =$$

$$= -\frac{c}{\sqrt{n}} \int_0^{k/n} e^{c(k/n-s)} J_c(s) dV(s) + \widetilde{R}_{1,n} + R_{2,n}$$

$$|\widetilde{R}_{1,n}| \le \sum_{j=1}^{k} e^{c\frac{k-j}{n}} \left| 1 - e^{c(T_{n,k}-T_{n,j}-\frac{k-j}{n})} - c\left(\frac{k-j}{n} - (T_{n,k} - T_{n,j})\right)\right| \cdot$$

$$\cdot \left|w\left(T_{n,j}\right) - w\left(T_{n,j-1}\right)\right| \le$$

$$\leq \sum_{j=1}^{k} c^2 e^{c\frac{k-j}{n}} \left(\frac{k-j}{n} - (T_{n,k} - T_{n,j}) \right)^2 |w(T_{n,j}) - w(T_{n,j-1})| = R_{1,n}$$

$$R_{2,n} = \frac{c}{\sqrt{n}} \sum_{i=1}^{k} e^{c\frac{k-i}{n}} (V_n(i) - V_n(i-1))\, y_{i-1} - \frac{c}{\sqrt{n}} \int_0^{k/n} e^{c(k/n-s)} J_c(s) dV(s)$$

$$R_{3,n} = \sum_{j=1}^{k} \int_{T_{n,j-1}}^{T_{n,j}} \left(e^{c(T_{n,k}-T_{n,j})} - e^{c(T_{n,k}-s)} \right) dw(s)$$

(b)

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

where we used statement (a). In the next theorem we would need to estimate the distributional impact of the o_p term, so, I keep track of them

$$R_{4,n} = \sum_{k=1}^{n} \left(\frac{y_{k-1}}{\sqrt{n}} - J_c(T_{n,k-1}) \right) \frac{\varepsilon_k}{\sqrt{n}} + \frac{1}{\sqrt{n}} c \int_0^1 \int_0^t e^{c(t-s)} J_c(s) dV(s) dw(t) =$$

$$= \sum_{k=1}^{n} (\widetilde{R}_{1,n} + R_{2,n} + R_{3,n}) \frac{\varepsilon_k}{\sqrt{n}} + R_{5,n}$$

$$R_{5,n} = -\frac{c}{\sqrt{n}} \sum_{k=1}^{n} \int_0^{k/n} e^{c(k/n-s)} J_c(s) dV(s) \frac{\varepsilon_k}{\sqrt{n}} + \frac{c}{\sqrt{n}} \int_0^1 \int_0^t e^{c(t-s)} J_c(s) dV(s) dw(t)$$

So

$$\frac{1}{n} \sum_{k=1}^{n} y_{k-1} \varepsilon_k = \sum_{k=1}^{n} J_c(T_{n,k-1}) \frac{\varepsilon_k}{\sqrt{n}} - \frac{c}{\sqrt{n}} \int_0^1 \int_0^t e^{c(t-s)} J_c(s) dV(s) dw(t) + R_{4,n}$$

Now

$$\sum_{k=1}^{n} J_c(T_{n,k-1}) \frac{\varepsilon_k}{\sqrt{n}} - \int_0^1 J(s) dws = \int_1^{T_{n,n}} J_c(s) dw(s) - \sum_{k=1}^{n} \int_{T_{n,k-1}}^{T_{n,k}} (J_c(s) - J_c(T_{n,k-1})) dw(s)$$

By definition of O-U process $J(s) = w(s) + c \int_0^s J(t) dt$:

$$\sum_{k=1}^{n} \int_{T_{n,k-1}}^{T_{n,k}} (J_c(s) - J_c(T_{n,k-1})) dw(s) = \sum_{k=1}^{n} \int_{T_{n,k-1}}^{T_{n,k}} (w(s) - w(T_{n,k-1})) dw(s) +$$

$$+ c \sum_{k=1}^{n} \int_{T_{n,k-1}}^{T_{n,k}} (B(s) - B(T_{n,k-1})) dw(s),$$

where $B(s)=\int_0^s J_c(t)dt$. Let $R_{6,n}=\sum_{k=1}^n\int_{T_{n,k-1}}^{T_{n,k}}(B(s)-B(T_{n,k-1}))dw(s)$. By definition $\int_{T_{n,k-1}}^{T_{n,k}}(w(s)-w(T_{n,k-1}))dw(s)=\frac{\varepsilon_k^2}{2n}+\frac{T_{n,k}-T_{n,k-1}}{2}$, as a result

$$\sum_{k=1}^n\int_{T_{n,k-1}}^{T_{n,k}}(J_c(s)-J_c(T_{n,k-1}))dw(s)=\frac{1}{2\sqrt{n}}(U(1)+V(1))+R_{6,n}+o_p(n^{-1/2})$$

Now the last. We know that $d(J_c^2(x))=2J_cdw+2cJ_c^2dx+dx$, so

$$\int_1^{T_{n,n}}J_c(s)dw(s)=\frac{1}{2}\left(J_c^2(T_{nn})-J_c^2(1)\right)-\int_1^{Tnn}(cJ_c^2(x)+\frac{1}{2})dx=$$

$$=J_c(1)\left(J_c(T_{nn})-J_c(1)\right)+\frac{1}{2}\left(J_c(T_{nn})-J_c(1)\right)^2-\frac{1}{\sqrt{n}}V\cdot\left(cJ_c^2(1)+\frac{1}{2}\right)+R_{7,n}=$$

$$=J_c(1)\left(w(T_{nn})-w(1)+\frac{c}{\sqrt{n}}J_c(1)V\right)+\frac{1}{2}\frac{\left(w(T_{nn})-w(1)\right)^2}{2}-\frac{V}{\sqrt{n}}\left(cJ_c^2(1)+\frac{1}{2}\right)+\widetilde{R}_{8,n}=$$

$$=J_c(1)\left(w(T_{nn})-w(1)\right)+\frac{1}{2}\left(w(T_{nn})-w(1)\right)^2-\frac{1}{2\sqrt{n}}V+\widetilde{R}_{8,n}=$$

$$=n^{-1/4}J_c(1)M(V)+n^{-1/2}\left(J_c(1)N(V)+\frac{1}{2}M^2(V)-\frac{1}{2}V\right)+\widetilde{R}_{8,n}+o_p(n^{-1/2}).$$

Here the last equality is due to statement (a) of Lemma 2. The definition of error terms is

$$R_{7,n}=-c\int_1^{Tnn}(J_c^2(x)-J_c^2(1))dx;$$

$$\widetilde{R}_{8,n}=R_{7,n}+cJ_c(1)\int_1^{Tnn}(J_c(x)-J_c(1))dx=R_{7,n}+cJ_c(1)R_{8,n}.$$

As a result,

$$\frac{1}{n}\sum_{k=1}^n y_{k-1}\varepsilon_k=\int_0^1 J_c(x)dw(x)+n^{-1/4}J_c(1)M(V)+$$

$$+\frac{1}{\sqrt{n}}\left(-c\int_0^1\int_0^t e^{c(t-s)}J_c(s)dV(s)dw(t)+J_c(1)N(V)+\frac{1}{2}M^2(V)+\frac{1}{2}U\right)+o_p(\frac{1}{\sqrt{n}})$$

(c) Using the statement of part (a)

$$\frac{1}{n}\sum\left(\frac{y_k^2}{n}-J_c^2(T_{nk})\right)=\frac{1}{n}\sum\left(\frac{y_k}{\sqrt{n}}-J_c(T_{nk})\right)\left(2J_c(T_{nk})+\frac{y_k}{\sqrt{n}}-J_c(T_{nk})\right)=$$

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

$$=-\frac{2c}{\sqrt{n}}\int_0^1 J_c(x)\int_0^x e^{c(x-s)}J_c(s)dV(s)dx+R_{9,n}+o_p(n^{-1/2}),$$

where

$$R_{9,n} = \frac{1}{\sqrt{n}} \left(\frac{1}{n} \sum_{k=1}^{n} B_1(k/n) - \int_0^1 B_1(x)dx \right),$$

here

$$B_1(x) = -cJ_c(x) \int_0^x e^{c(x-s)} J_c(s) dV(s).$$

As a result,

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

Now,

$$\frac{1}{n} \sum J_c^2(T_{n,k-1}) - \int_0^1 J_c^2(x)dx = \sum J_c^2(T_{n,k-1}) \left(\frac{1}{n} - (T_{nk} - T_{n,k-1}) \right) -$$

$$- \sum \int_{T_{n,k-1}}^{T_{nk}} (J_c^2(t) - J_c^2(T_{n,k-1}))dt + \int_1^{T_{nn}} J_c^2(t)dt.$$

Let us consider each term separately:

$$\sum J_c^2(T_{n,k-1}) \left(\frac{1}{n} - (T_{nk} - T_{n,k-1}) \right) = -\frac{1}{\sqrt{n}} \int_0^1 J_c^2(x)dV(x) + R_{10,n},$$

where

$$R_{10,n} = -\frac{1}{\sqrt{n}} \left(\sum J_c^2(T_{n,k-1})(V_n(k/n) - V_n(k-1/n)) - \int_0^1 J_c^2(x)dV(x) \right).$$

$$- \sum \int_{T_{n,k-1}}^{T_{nk}} (J_c^2(t) - J_c^2(T_{n,k-1}))dt = R_{11,n} = o_p(n^{-1/2})$$

$$\int_1^{T_{nn}} J_c^2(t)dt = \frac{1}{\sqrt{n}} J_c^2(1)V + R_{7,n}.$$

Summing up:

$$\frac{1}{n^2} \sum y_k^2 = \int_0^1 J_c^2(x)dx - \frac{2c}{\sqrt{n}} \int_0^1 J_c(x) \int_0^x e^{c(x-s)} J_c(s) dV(s) dx -$$

$$- \frac{1}{\sqrt{n}} \int_0^1 J_c^2(x)dV(x) + \frac{1}{\sqrt{n}} J_c^2(1)V + o_p(\frac{1}{\sqrt{n}})$$

(d) Using the statement of part (a)

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

$$= -\frac{c}{\sqrt{n}}\int_0^1\int_0^x e^{c(x-s)}J_c(s)dV(s)dx + R_{9,n} + o_p(n^{-1/2}).$$

As a result,

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

Now,

$$\frac{1}{n}\sum J_c(T_{n,k-1}) - \int_0^1 J_c(x)dx = \sum J_c(T_{n,k-1})\left(\frac{1}{n} - (T_{nk} - T_{n,k-1})\right) -$$

$$-\sum\int_{T_{n,k-1}}^{T_{nk}}(J_c(t) - J_c(T_{n,k-1}))dt + \int_1^{T_{nn}} J_c(t)dt$$

Let me consider each term separately

$$\sum J_c(T_{n,k-1})\left(\frac{1}{n} - (T_{nk} - T_{n,k-1})\right) = -\frac{1}{\sqrt{n}}\int_0^1 J_c(x)dV(x) + R_{12,n},$$

where

$$R_{12,n} = -\frac{1}{\sqrt{n}}\left(\sum J_c(T_{n,k-1})(V_n(k/n) - V_n(k-1/n)) - \int_0^1 J_c(x)dV(x)\right).$$

$$-\sum\int_{T_{n,k-1}}^{T_{nk}}(J_c(t) - J_c(T_{n,k-1}))dt = R_{13,n}.$$

$$\int_1^{T_{nn}} J_c(t)dt = J_c(1)V + R_{8,n}.$$

Summary

$$\frac{1}{n^{3/2}}\sum y_k = \int_0^1 J_c(x)dx - \frac{c}{\sqrt{n}}\int_0^1\int_0^x e^{c(x-s)}J_c(s)dV(s)dx -$$

$$-\frac{1}{\sqrt{n}}\int_0^1 J_c(x)dV(x) + \frac{1}{\sqrt{n}}J_c(1)V + o_p(\frac{1}{\sqrt{n}})$$

Part (e) follows from (a)-(d) and Taylor expansion.

Now we need to check that for all i we have $R_{i,n} = o_p(n^{-1/2})$. Statements for $R_{2,n}, R_{5,n}, R_{10,n}$ and $R_{12,n}$ follows from Lemma 3 on convergence of stochastic integrals. From convergence of non-stochastic integrals we have $R_{9,n} = o_p(n^{-1/2})$.

Terms $R_{11,n}$ and $R_{13,n}$ have a structure of $\sum_{k=1}^n \xi_{k,n}$, where $\xi_{k,n}$ are i.i.d. across k and distributionally equal to $\xi_{i,k} \sim \int_0^{\tau/n} C(t)dt$ for $dC(t) = c_1(t)dt + c_2(t)dw(t)$ with $c_1 \in L_1, c_2 \in L_2$. Then $E\xi_{1,n} = E(\int_0^{\tau/n}\int_0^t c_1(t)dt) \leq const\cdot n^{-2}$, and

$$E\xi_{1,n}^2 = E(\int_0^{\tau/n}\int_0^t E\int_0^u c_2^2(s)dsdudt) \leq const\cdot n^{-3}.$$

As a result, both terms are $O_p(1)$. We can also notice that Chebyshev's inequality implies that they are distributionally $o(n^{-1/2})$.

Terms $R_{3,n}$ and $R_{6,n}$ also have a structure of $\sum_{k=1}^{n} \xi_{k,n}$, where $\xi_{k,n}$ are i.i.d. across k. Here $\xi_{i,n} \sim \int_0^{\tau/n} C(t)dw(t)$ with $dC(t) = c_1 dt$. It is easy to see that $E\xi_{1,n} = 0$ and $E\xi_{1,n}^2 \leq const \cdot n^{-3}$. It implies that terms are probabilistically and distributionally $o(n^{-1/2})$.

Terms $R_{7,n}$ and $R_{8,n}$ have form of $\int_1^{T_n n}(D(s)-D(1))ds$ where $dD(s) = d_1 dw + d_2 dt$. It's easy to see that they are $o_p(n^{-1/2})$. □

Lemma 4 (Park(2003a)) *If $r > 4$ then we might choose B_n and B such that*

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

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

Proof of Theorem 2. We need to check that all terms $R_{i,n}$ used in the proof of Theorem 1 is distributionally of order $o(n^{-1/2})$.

In the proof of Theorem 1 we already showed that terms $R_{3,n}$, $R_{6,n}$, $R_{11,n}$ and $R_{13,n}$ are distributionally $o(n^{-1/2})$.

Terms $R_{2,n}, R_{5,n}, R_{10,n}$ and $R_{12,n}$ 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 4.

Terms $R_{7,n}$ and $R_{8,n}$ have form of $|\int_1^{T_n n}(D(s)-D(1))ds| \leq \sup_{0\leq t\leq 1}|D(t)|\cdot|T_n n - 1|$ which is distributionally $o^{-1/2}$.

Proof of Lemma 1.

First, I show that for Skorokhod's construction presented in Skorokhod's book (1965) we have $E\tau^2 = \frac{5}{3}E\xi^4$.

Let $\tau_{a,b}$ is the smallest root of the equation $(w(t) - a)(w(t) - b) = 0$. Then

$$Ee^{-\lambda\tau_{a,b}} = \frac{\sinh b\sqrt{2\lambda} - \sinh a\sqrt{2\lambda}}{\sinh(b-a)\sqrt{2\lambda}}, \tag{5}$$

and

$$(-1)^k \frac{d^k}{d\lambda^k} Ee^{-\lambda\tau_{a,b}}\Big|_{\lambda=0} = E\tau_{a,b}^k. \tag{6}$$

In the construction from Skorokhod's book (1965) the stopping time is defined as

$$\tau = \inf_t \{(w(t) - \varepsilon)(w(t) - G(\varepsilon))\} = \tau_{\varepsilon,G(\varepsilon)},$$

where ε is independent of w and the function G is defined by $\int_{G(x)}^{x} y dF(y) = 0, F(x) = P\{\varepsilon \leq x\}$. Then $w(\tau)$ has the same distribution as ε.

We can notice that

$$E\tau^2 = E\left(E\left(\tau^2|\varepsilon\right)\right) = E\left(\tau^2_{\varepsilon,G(\varepsilon)}\right).$$

The last could be calculated using equation (6) for moments of $\tau_{a,b}$ and the explicit formula for the characteristic function (5). We also use the following two facts: $G(G(x)) = x$ and $G(x)dF(G(x)) = xdF(x)$. By tedious but straightforward calculations one can obtain the formula $E\tau^2 = \frac{5}{3}E\xi^4$.

Since $E\varepsilon^8 < \infty$ by using Chebyshev's inequality one can get the statement of the lemma.

Proof of Theorem 3. Theorem 2 states the distributional expansions for the t-statistic and its grid bootstrapped analog

$$\sup_x |P\{t(y,n,\rho) \leq x\} - G_n(x)| = o(n^{-1/2}), \tag{7}$$

here $G_n(x) = P\{t^c + \frac{1}{n^{1/4}}f + \frac{1}{\sqrt{n}}g\}$, where f and g are functionals of Brownian motions $B(\cdot)$. The covariance structure of B is described in (3), it depends on $\sigma^2, \mu_3, \mu_4, \kappa, c$. Brownian motion M is independent of B. It could be seen from the proof of Theorem 2 that the term $o(n^{-1/2})$ in equation (7) is bounded by a constant depending on the eights moment of the approximated error term μ_8 times $n^{-1/2-\delta}$ for some $\delta > 0$.

For almost any realization of an infinite sequence of error terms $(\varepsilon_1, ..., \varepsilon_n, ...)$ and its finite subsequence of the length n we would have

$$\sup_x |P\{t(y^*, n, \rho) \leq x\} - G_n^*(x)| \leq Const(\widehat{\mu}_8)n^{-1/2-\delta},$$

here $G_n^*(x) = P\{t^c + \frac{1}{n^{1/4}}f^* + \frac{1}{\sqrt{n}}g^*\}$ where f^* and g^* are the same functionals of B^*, M^*. The covariance structure of B^* depends on $\widehat{\sigma}^2, \widehat{\mu}_3, \widehat{\mu}_4, \widehat{\kappa}$ Since $r \geq 8$, then $Const(\widehat{\mu}_8)n^{-1/2-\delta} = o(n^{-1/2})$ $a.s.$ As a result,

$$\sup_x |P\{t(y^*, n, \rho) \leq x\} - G_n^*(x)| = o(n^{-1/2}) \quad a.s., \tag{8}$$

Combining equations (7) and (8) with Lemma 1 one obtains the statement of Theorem 3.

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 P. Guggenberger (2007a): "Hybrid and Size-corrected Subsample Methods ", Cowles Foundation Discussion Paper 1606.

Andrews D.W.K. and P. Guggenberger (2007b): "The Limit of Finite Sample Size and a Problem with Subsampling", Cowles Foundation Discussion Paper 1605.

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

Hansen B.E. (1999): "The Grid Bootstrap and the Autoregressive Model," *The Review of Economics and Statistics*, 81(4), 594-607.

Mikusheva A. (2007):"Uniform inference in autoregressive models," *Econometrica*, 75(5), 1411-1452.

Park J.Y. (2003a): "Bootstrap Unit root tests," *Econometrica*, 71(6), 1845-1895.

Park J.Y. (2003b): "Weak Unit Roots," unpublished manuscript.

Phillips P.C.B. (1987): "Toward a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74(3), 535-547.

Skorokhod A.V. (1965): "Studies in the Theory of Random Processes," *Addison-Wesley Publishing Company, Reading, Massachusetts*

Stock J. (1991): "Confidence Intervals for the Largest Autoregressive Root in US Macroeconomic Time Series," *Journal of Monetary Economics*, 28, 435-459.