Edgeworth Expansion of the Parametric Bootstrap t-statistic for Linear Regression Processes with Strongly Dependent Errors

Mosisa Aga

Mathematics Department
Auburn University Montgomery
Montgomery, Alabama 36117
United States
maga@aum.edu

Abstract

The purpose of this paper is to provide a valid Edgeworth expansion for the parametric bootstrap t-statistic of a linear regression process whose error terms are stationary, Gaussian, and strongly dependent time series. Under some sets of conditions on the spectral density function and the parametric values, an Edgeworth expansion of the bootstrap t-statistic of arbitrarily large order of the process is proved to have an error of $o(n^{1-s/2})$ where s is a positive integer. The result is similar to the Edgeworth expansion obtained by Andrews and Lieberman [2002], which was established for the parametric bootstrap t-statistic of the *plug-in maximum likelihood* (PML) estimators of stationary, Gaussian, and strongly dependent processes, but without the linear regression component.

Keywords: Edgeworth Expansion; parametric bootstrap; t-statistic, linear regression, strongly dependent.

2000 Mathematics Subject Classification: 62M10

1. Introduction

Consider a linear regression model $\{X_t = Z_t\beta + \varepsilon_t, t \geq 1\}$, where β is a p vector of deterministic but unknown real numbers, $\{Z_t \in \mathbb{R}^p, t \geq 1, p \geq 1\}$ are non-stochastic regressors, and the error terms $\{\varepsilon_t, t \geq 1\}$ are stationary, Gaussian, and strongly dependent discrete time series. The process $\{\varepsilon_t, t \geq 1\}$ is assumed to have mean zero and and spectral density $f_\theta(\lambda)$ for $\lambda \in (-\pi, \pi)$, where $\theta = (\theta_1, \theta_2, \ldots, \theta_r)' \in \mathbb{R}^r$ and

$$f_\theta(\lambda) = O(|\lambda|^{-2d-\delta}) \tag{1.1}$$

as $|\lambda| \downarrow 0, \forall \delta > 0$, $d \in (0, 1/2)$, and $\theta_1 = d$, referred to as the "long-memory parameter" of the process (see Andrews and Lieberman [2002], pages 4-5).

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ be an observed sample of size n and $\mathscr{E} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ be the corresponding error terms, where for each $i = 1, 2, \ldots, n, X_i = Z_i\beta + \varepsilon_i$. We note that the covariance matrix of $\mathbf{X}$ is the same as the covariance matrix of $\mathscr{E}$.

Let $\mu = (\mu_1, \mu_2, \ldots, \mu_n)$ be the true mean of $\mathbf{X}$. Then, the least square estimate (LSE) $\hat{\beta} = \left(\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p\right)$ of β is given by $\hat{\beta} = V^{-1}\sum_{t=1}^n X_t Z_t$, where $V = \sum_{t=1}^n (Z_t Z_t')$ is a $p \times p$ matrix. Thus, an estimator of μ is $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_n)$, where $\hat{\mu}_t = Z_t'\hat{\beta}, t = 1, 2, \ldots, n$. Let $\mathbf{Z}$ denote the design matrix given by $\mathbf{Z} = (z_{ij})$ for $i = 1, \ldots, n$ and $j = 1, \ldots, p$. We shall assume that the rank of $\mathbf{Z}$ is p. We note that the matrix V is symmetric and positive definite.

The $n \times n$ (Toeplitz) covariance matrix corresponding to $f_\theta(\lambda)$ is denoted by $T_n(f_\theta)$ and has (j,k) element defined by:

$$T_n(f_\theta)_{j,k} = \int_{-\pi}^{\pi} e^{i(j-k)\lambda} f_\theta(\lambda) d\lambda. \tag{1.2}$$

The log-likelihood function is

$$L_n(\theta, \mu) = -\frac{n}{2} ln(2\pi) - \frac{1}{2} ln(det(T_n(f_\theta))) - \frac{1}{2} (X-\mu)' T_n^{-1}(f_\theta)(X-\mu). \tag{1.3}$$

We refer to $L_n(\theta, \hat{\mu})$, where $\hat{\mu}$ is replaced for μ in (1.3) above, as the *plug-in log-likelihood* (PLL) function. Let $Q_n = \mathbf{Z}V^{-1}\mathbf{Z}'$ and let $M_n = I_n - Q_n$, where I_n is the $n \times n$ identity matrix. It is easy to verify that the matrices M_n and Q_n have the following properties: a) Both M_n and Q_n are symmetric. b) $\mathbf{X}'M_n = (\mathbf{X} - \hat{\mu})'$. c) If $U = \mathbf{X} - \mu$, then $M_n\mathbf{X} = M_nU$. d) There exists an $n \times p$ matrix $\mathbf{E}$ such that

$$Q_n = \mathbf{E}\mathbf{E}' \tag{1.4}$$

Using these properties the PLL function can now be written as

$$L_n(\theta, \hat{\mu}) = -\frac{n}{2} ln(2\pi) - \frac{1}{2} ln(det(T_n(f_\theta))) - \frac{1}{2} X'M_n T_n^{-1}(f_\theta) M_n X. \tag{1.5}$$

A *plug-in maximum likelihood* (PML) estimator is an element of the parameter space that maximizes the PLL function (1.5) above. Andrews and Lieberman [2002] (Lemma 9(b), page 29), have established a valid asymptotic expansion for the parametric bootstrap t-statistic of the error term $\{\varepsilon_t, t \geq 1\}$ given above. In this paper we extend this work of Andrews and Lieberman [2002] and obtain an asymptotic expansion of the parametric bootstrap t-statistic of the processes $\{X_t = Z_t\beta + \varepsilon_t, t \geq 1\}$ described above by imposing an additional condition on the regression coefficients and a mild additional condition on the spectral density function.

This paper is an outgrowth of a study first reported by Aga and Sun [2007], which provides higher order improvements of parametric bootstrap confidence intervals of the same model described above. The preliminary results needed and the underlying assumptions of the current paper are contained in Aga and Sun [2007]. Those that are relevant to our discussion are presented here for convenience. We provide a rigorous treatment of the asymptotic expansion of the bootstrap t-statistic in this piece.

The remainder of the paper proceeds as follows. Section 2 provides three assumptions and some preliminaries. Section 3 presents and proves the main Edgeworth expansion result.

2. Assumptions

In addition to the assumptions **I-VI** of Andrews and Lieberman [2002], we shall impose the following assumptions **A1-A3** on the spectral density function, the regression coefficients and the parameter space Θ.

A1. There exists a compact subset Θ_0 of Θ and a constant $\delta > 0$ such that the true value θ_0 lies in the interior of Θ_0 and $f_\theta(\lambda) > \delta$ for all $\theta \in \Theta_0$ and $\lambda \in (-\pi, \pi)$.

A2. The design matrix **Z** is chosen in such a way that for the matrix

$$\mathbf{E} = (e_{ij}), i = 1, ..., n, j = 1, ..., p \qquad (2.1),$$

defined by (1.4) above, there exists a constant $M < \infty$ such that $|e_{ij}| \leq \frac{M}{\sqrt{n}}$ for $1 \leq i \leq n$, $1 \leq j \leq p$. **A1** requires f_θ to be bounded away from zero. This indeed is not a severe restriction on f_θ because for most practical examples of stationary, Gaussian, strongly dependent time series, it is always positive. For the most popular $ARFIMA(p, d, q)$ process for instance, the spectral density function f_θ is given by $f_\theta(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\Psi(e^{i\lambda})|^2}{|\Phi(e^{i\lambda})|^2} |1 - e^{i\lambda}|^{-2d}$, where $\theta = (\sigma_\varepsilon^2; d; \Phi_1, \Phi_1, ..., \Phi_p; \Psi_1, ..., \Psi_q)$. [Brockwell and Davis (1991), Equation (13.2.18).] Clearly, f_θ is positive for this particular model.

A2 puts a restriction on the design matrix **Z**. It is mainly this assumption that enables us to extend the results of Andrews and Lieberman [2002] and that of Lieberman *et al.* [2003] to our current regression model. Theorem 4.2 and Lemmas 4.4-4.5 of Aga and Sun [2007], which are cited in the proof of Theorem 3.1 of section 3 below, are extensions of Theorem 3 and Lemmas 3 and 5, respectively, of Andrews and Lieberman [2002], with a unit vector e_n of theirs replaced by matrix **E** defined in (1.4) above, further complicating the proof of the extension. This complication is resolved partly by making use of Lemma 4.4 of Aga and Sun [2007], which in turn makes use of mainly **A2**, but also **A1**.

One drawback of the result of this paper (and that of Aga and Sun [2007]) is the somewhat severe restriction imposed on the design matrix by **A2**. While Lemma 3.1, page 620 of Aga and Sun [2007] provides a useful example of a design matrix that satisfies this assumption, it is generally unknown whether or not the results of this paper will go through without it.

To state the third assumption we need to introduce the following notations. Let $v = (r_1, r_2, ..., r_q)'$ denote a q-vector of positive integers each less than or equal to $d = dim(\theta)$. We write the real valued q-th order partial derivative of the PLL function indexed by v as

$$L_{n,v} = D_v L_n(\theta, \hat{\mu}) = \frac{\partial^q}{\partial \theta_{r_1} \dots \partial \theta_{r_q}} L_n(\theta, \hat{\mu}) = F_{n,v}(\theta) + X' M_n B_{n,v}(\theta) M_n X \qquad (2.2)$$

where

$$F_{n,v}(\theta) = -\frac{1}{2} D_v ln(det(T_n(f_\theta))) = \sum_{k=1}^{b} a_k tr(\prod_{j=1}^{p_k} T_n^{-1}(f_\theta) T_n(g_{\theta,k,j})) \qquad (2.3)$$

and

$$B_{n,v}(\theta) = -\frac{1}{2} D_v T_n^{-1}(f_\theta) = \sum_{k=1}^{b} a_k (\prod_{j=1}^{p_k} T_n^{-1}(f_\theta) T_n(g_{\theta,k,j}) T_n^{-1}(f_\theta) \qquad (2.4)$$

for some constants b, a_k, and p_k that depend on v and with $g_{\theta,k,j}$ being certain partial derivatives of the spectral density with respect to the components of θ of order q or less. Equations (2.2), (2.3), and (2.4) are the same as those of equations (8.1) and (8.2) of Andrews and Lieberman [2002] except that our M_n is different from theirs and that we used the estimator $\hat{\mu}$ of our sample **X** instead of their sample mean which they denoted by $\bar{X}_n$.

We shall introduce some more notations. Let $V_n(\theta) = (L_{n,v(1)}(\theta), \ldots, L_{n,v(d)}(\theta))$, where each vector $v(j)$ is of the same form as v defined in (2.2)-(2.4) above. $V_n(\theta)$ is what we shall call a vector of log-likelihood derivatives (LLDs). Let $W_n(\theta) = n^{-1/2}(V_n(\theta) - E_\theta V_n(\theta))$. Without loss of generality we may assume that $E_\theta V_n(\theta) = 0$. Let $D_n(\theta) = E[W_n(\theta)W_n(\theta)']$ and $D(\theta) = \lim_{n \to \infty} D_n(\theta)$.

A3. The matrices $D_n(\theta)$ and $D(\theta)$ are non-singular.

3. Edgeworth expansion of the bootstrap t-statistic

As a way of introducing the bootstrap t-statistic, we first define the parametric bootstrap sample and formulate the general set up of bootstrap confidence intervals and tests. (a) By definition, the parametric bootstrap sample $\mathbf{X}^* = (\mathbf{X}_1^*, \ldots, \mathbf{X}_n^*)$ has conditional distribution given $\mathbf{X}$ that is the same as the distribution of the original sample except that the true parameters are $(\hat{\theta}_n, \hat{\mu})$ rather than (θ_0, μ). That is, $\mathbf{X}^*$ consists of Gaussian strongly dependent linear regression random variables with mean $\hat{\mu}$ and spectral density $f_{\hat{\theta}_n}(\lambda)$ conditional on the original sample $\mathbf{X}$. (b) The bootstrap sample mean $\hat{\mu}^*$ is defined by $\hat{\mu}^* = (\hat{\mu}_1^*, \ldots, \hat{\mu}_n^*)$, where, for $t = 1, \ldots, n$, $\hat{\mu}_t^* = Z_t^{*'}\hat{\beta}^*$, $\hat{\beta}^* = V^{*-1}\sum_{t=1}^n \mathbf{X}_t^* \mathbf{Z}_t^*$, and $V^* = \sum_{t=1}^n (Z_t^* Z_t^{*'})$. (c) The bootstrap PLL function $L_n(\theta, \hat{\mu}^*)$ is defined in the same way as the PLL function $L_n(\theta, \hat{\mu})$ (see (1.5) above) but with $\mathbf{X}^*$ and $\hat{\mu}^*$ replacing $\mathbf{X}$ and $\hat{\mu}$, respectively. (d) Let Θ^* denote the set of solutions in the parameter space Θ to the first order conditions for the bootstrap PLL function. The bootstrap estimator $\hat{\theta}_n^*$ can now be defined as that value of θ that maximizes the bootstrap PLL function $L_n(\theta, \hat{\mu}^*)$. Observe that the true parameter of the bootstrap sample is $\hat{\theta}_n$, and hence $\hat{\theta}_n^*$ is a PML estimator of $\hat{\theta}_n$.

Let θ_h denote some element of Θ, the parameter space. Let $\theta_{0,r}$, $\theta_{h,r}$, and $\hat{\theta}_{n,r}$ denote the r-th elements of θ_0, θ_h, and $\hat{\theta}_n$, respectively. Let $\Sigma_{r,r}(\hat{\theta}_n)$ denote the (r,r)-th element of $\Sigma(\hat{\theta}_n)$. Let z_α denote the $1 - \alpha$ quantile of the standard normal distribution. We define the bootstrap t statistic to be

$$T_n^*(\hat{\theta}_{n,r}) = \frac{\sqrt{n}(\hat{\theta}_{n,r}^* - \hat{\theta}_{n,r})}{\Sigma_{r,r}^{1/2}(\hat{\theta}_n^*)} \tag{3.1}.$$

where $\hat{\theta}_{n,r}^*$ denotes the r-th element of $\hat{\theta}_n^*$.

We now state a condition, known as *Condition* $\mathbf{C_s}$, on the parametric values [See Andrews and Lieberman [2002], page 14]. The condition is used in the proof of Theorem 3.1 below.

Condition $\mathbf{C_s}$. Given some integer $s \geq 3$, the sequence of estimators $\{\bar{\theta}_n : n \geq 1\}$ is said to satisfy *Condition* $\mathbf{C_s}$ if for all $\varepsilon > 0$ and all compact subsets Θ_c of Θ, $\sup_{\theta_0 \in \Theta_c} P_{\theta_0}(||\bar{\theta}_n - \theta_0|| > n^{-1/2}ln(n)\varepsilon) = o(n^{1-s/2})$ *as* $n \to \infty$

Another drawback of the results of this paper (apart from that of **A2**, as discussed in section 2 above) is that the PML estimators are required to satisfy *Condition* $\mathbf{C_s}$, which implies that these estimators are consistent. The same drawback occurs in the works of Aga and Sun [2007], Andrews and Lieberman [2002], and in those of Bhattacharya and Gosh [1978]. While Andrews and Lieberman [2002] Lemma 1, page 14, shows that there exists a sequence of estimators that satisfies this condition, it is generally unknown whether or not the result of this paper and those of others are valid without it.

The next lemma is a key ingredient in the proof of the main result of the paper. First we introduce some additional notations.

Let $\Phi(\cdot)$ denote the distribution function of the standard normal distribution. Define $D_{\omega,\eta} = \frac{\partial^q}{\partial \omega_{\eta_1} \ldots \partial \omega_{\eta_q}}$, for $\eta = (\eta_1, \ldots, \eta_q)$. Let $\kappa_{n,s}(\theta)_\eta$ denote the η cumulants of $V_n(\theta)$ when θ is the true

value, where s is as in *Condition* $\mathbf{C_s}$. By definition, $\kappa_{n,s}(\theta)_\eta = i^{-q}D_{\omega,\eta}ln(\varphi_n(\omega,\theta))|_{\omega=0}$, where $i = \sqrt{-1}$. The vecttor $\kappa_{n,s}(\theta)$ is composed of elements $\kappa_{n,s}(\theta)_\eta$ for vectors η of dimension $q \le s$. Let $\bar{\kappa}_{n,s}(\theta) = \frac{\kappa_{n,s}(\theta)}{n}$. By Lemma 4.5(c) of Aga and Sun [2007], the elements of $\bar{\kappa}_{n,s}(\theta)$ are $O(1)$. Let $P_j(\Delta, \bar{\kappa}_{n,s}(\theta))$ be a polynomial in $\Delta = \partial/\partial z$ whose coefficients are polynomials in the elements of $\bar{\kappa}_{n,s}(\theta)$ and for which $P_j(\Delta, \bar{\kappa}_{n,s}(\theta))\Phi(x)$ is an even function of x when j is odd and an odd function of x when j is even for $j = 1, 2, ..., s-2$.

Let $\tau_n(\theta_0)$ denote $n^{1/2}(\hat{\theta}_n - \theta_0)$ or the t-statistic $T_n(\theta_{0,r}) = \frac{n^{1/2}(\hat{\theta}_n - \theta_0)}{\Sigma_{r,r}^{1/2}(\hat{\theta}_n)}$. Let $\frac{1}{n}\bar{V}_n(\theta_0)$ denote the vector $\frac{1}{n}V_n(\theta_0)$ of normalized LLDs augmented to include the vector of expected values of all partial derivatives with respect to θ of order s of $\frac{1}{n}L_n(\theta_0)$.

Lemma 3.1. *For each definition of $\tau_n(\theta_0)$, there is an infinitely differentiable function $F(.)$ that does not depend on θ_0, that satisfies $F(n^{-1}E_{\theta_0}\bar{V}_n(\theta)) = 0$ for all n large and all $\theta_0 \in \Theta_c$, and*

$$\sup_{\theta_0 \in \Theta_c} \sup_{C \in \mathscr{C}_d} |P_{\theta_0}(\tau_n(\theta_0) \in C) - P_{\theta_0}(n^{1/2}F(\frac{1}{n}\bar{V}_n(\theta_0)) \in C)| = o(n^{1-s/2}), \tag{3.2}$$

where $\mathscr{C}_d$ denotes the class of all convex sets in R^d and some compact subset of the parameter space Θ.

Proof. Let $\rho_n(\theta) = \frac{1}{n}L_n(\theta, \hat{\mu})$. To establish (3.2) above we consider two cases. Case (1): when $\tau_n(\theta_0) = n^{1/2}(\hat{\theta} - \theta_0)$, and case (2): when $\tau_n(\theta_0) = T_n(\theta_{0,r})$.

Case (1). Suppose $\tau_n(\theta_0) = n^{1/2}(\hat{\theta} - \theta_0)$. By *Condition* $\mathbf{C_s}$ and **A3**, we have $\inf_{\theta_0 \in \Theta_c} P_{\theta_0}(\hat{\theta}_n \in \Theta^0) = 1 - o(n^{1-s/2})$, where Θ^0 denote the interior of Θ and $\inf_{\theta_0 \in \Theta_c} P_{\theta_0}(\frac{\partial}{\partial \theta}\rho_n(\hat{\theta}_n) = 0) = 1 - o(n^{1-s/2})$. By Taylor expansion of $\frac{\partial}{\partial \theta}\rho_n(\hat{\theta}_n)$ about θ_0 of order $s-1$, there exists $\bar{\theta}_n$ that lies between $\hat{\theta}_n$ and θ_0 such that

$$\begin{aligned} 0 = &\frac{\partial}{\partial \theta}\rho_n(\hat{\theta}_n) = \frac{\partial}{\partial \theta}\rho_n(\theta_0) + \Sigma_{j=1}^{s-2} \frac{1}{j!}D^j \frac{\partial}{\partial \theta}\rho_n(\theta_0)(\hat{\theta}_n - \theta_0, ..., \hat{\theta}_n - \theta_0) \\ &+ \frac{1}{(s-1)!}D^{s-1}\frac{\partial}{\partial \theta}\rho_n(\bar{\theta}_n)(\hat{\theta}_n - \theta_0, ..., \hat{\theta}_n - \theta_0), \end{aligned} \tag{3.3}$$

where $D^j \frac{\partial}{\partial \theta}\rho_n(\theta_0)(\hat{\theta}_n - \theta_0, ..., \hat{\theta}_n - \theta_0)$ denotes $D^j \frac{\partial}{\partial \theta}\rho_n(\theta_0)$ as a j-linear map, whose coefficients are partial derivatives of $\frac{\partial}{\partial \theta}\rho_n(\theta_0)$ of order j, applied to the j-tuple $(\hat{\theta}_n - \theta_0, ..., \hat{\theta}_n - \theta_0)$. Let,

$$\zeta_{1n}(\theta_0) = \frac{1}{(s-1)!}(D^{s-1}\frac{\partial}{\partial \theta}\rho_n(\bar{\theta}_n) - D^{s-1}\frac{\partial}{\partial \theta}\rho_n(\theta_0))(\hat{\theta}_n - \theta_0, ..., \hat{\theta}_n - \theta_0), \; and$$

$$\zeta_{2n}(\theta_0) = \frac{1}{(s-1)!}(D^{s-1}\frac{\partial}{\partial \theta}\rho_n(\theta_0) - ED^{s-1}\frac{\partial}{\partial \theta}\rho_n(\theta_0))(\hat{\theta}_n - \theta_0, ..., \hat{\theta}_n - \theta_0). \tag{3.4}$$

Then, (3.3) above can be written as

$$\begin{aligned} 0 = &\frac{\partial}{\partial \theta}\rho_n(\hat{\theta}) = \frac{\partial}{\partial \theta}\rho_n(\theta_0) + \Sigma_{j=1}^{s-2} \frac{1}{j!}D^j \frac{\partial}{\partial \theta}\rho_n(\theta_0)(\hat{\theta}_n - \theta_0, ..., \hat{\theta}_n - \theta_0) \\ &+ \frac{1}{(s-1)!}ED^{s-1}\frac{\partial}{\partial \theta}\rho_n(\bar{\theta}_n)(\hat{\theta}_n - \theta_0, ..., \hat{\theta}_n - \theta_0) + \zeta_{1n}(\theta_0) + \zeta_{2n}(\theta_0). \end{aligned} \tag{3.5}$$

Note that, by definition, $\frac{1}{n}\bar{V}_n(\theta_0)$ is the column vector whose elements are the non-redundant components of $\frac{\partial}{\partial \theta}\rho_n(\theta_0)$, $D^1 \frac{\partial}{\partial \theta}\rho_n(\theta_0)$, ..., $D^{s-2}\frac{\partial}{\partial \theta}\rho_n(\theta_0)$ plus the components of $ED^{s-1}\frac{\partial}{\partial \theta}\rho_n(\theta_0)$. Let

$e_n(\theta_0) = ((\zeta_{1n}(\theta_0) + \zeta_{2n}(\theta_0))', 0, .., 0)'$ be conformable with $\bar{V}_n(\theta_0)$. Equation (3.5) above can be written as

$$v(\frac{1}{n}\bar{V}_n(\theta_0) + e_n(\theta_0), \hat{\theta}_n - \theta_0) = 0, \tag{3.6}$$

where $v(., .)$ is an infinitely differentiable function that satisfies

$$v(n^{-1}E_{\theta_0}\bar{V}_n(\theta_0), 0) = 0, \tag{3.7}$$

for all $n \geq 1$. (3.6) above holds because $\frac{\partial}{\partial\theta}\rho_n(\hat{\theta}_n) = 0$ and (3.7) holds because if $\hat{\theta}_n - \theta_0 = 0$, then $\hat{\theta}_n = \theta_0$, and hence $\frac{\partial}{\partial\theta}\rho_n(\theta_0) = \frac{\partial}{\partial\theta}\rho_n(\hat{\theta}_n) = 0$. From equation (3.5), we can see by inspection that the function v also satisfies

$$\frac{\partial}{\partial x}v(n^{-1}E_{\theta_0}\bar{V}_n(\theta_0), x)|_{x=0} = n^{-1}E_{\theta_0}\frac{\partial}{\partial\theta\partial\theta'}\rho_n(\theta_0) \tag{3.8}$$

where $x = \hat{\theta}_n - \theta_0$. Using the information matrix equality, the right hand side of (3.8) converges to $-\Sigma^{-1}(\theta_0)$ as $n \to \infty$, and, hence, is negative definite for n large because the later is negative definite by **A3**. Thus, since $v(n^{-1}E_{\theta_0}\bar{V}_n(\theta_0), 0) = 0$, the Implicit Function Theorem can be applied to the function $v(., .)$ at the point $(n^{-1}E_{\theta_0}\bar{V}_n(\theta_0), 0)$. That is, there is an infinitely differentiable function $F(n^{-1}\hat{V}_n(\theta_0) + e_n(\theta_0))$, defined near $n^{-1}E_{\theta_0}\bar{V}_n(\theta_0)$ and that does not depend on n or θ_0 such that

$$v(n^{-1}\bar{V}_n(\theta_0) + e_n(\theta_0), F(\frac{1}{n}\bar{V}_n(\theta_0) + e_n(\theta_0))) = 0, \tag{3.9}$$

where F satisfies $F(\frac{1}{n}E_{\theta_0}\bar{V}_n(\theta_0)) = 0$. Combining (3.5), (3.6), and (3.9) we obtain

$$\inf_{\theta_0 \in \Theta_c} P_{\theta_0}(\hat{\theta}_n - \theta_0 = F(\frac{1}{n}\bar{V}_n(\theta_0) + e_n(\theta_0))) = 1 - o(n^{1-s/2}). \tag{3.10}$$

We now apply Lemma 6 of Andrews and Lieberman [2002], page 27, with $A_n(\theta_0) = n^{1/2}F(\frac{1}{n}\bar{V}_n(\theta_0))$ and $\xi_n(\theta_0) = n^{1/2}(F(\frac{1}{n}\bar{V}_n(\theta_0) + e_n(\theta_0)) - F(\frac{1}{n}\bar{V}_n(\theta_0)))$ to obtain

$$|P_{\theta_0}(n^{1/2}F(\frac{1}{n}\bar{V}_n(\theta_0) + e_n(\theta_0)) \in C) - P_{\theta_0}(n^{1/2}F(\frac{1}{n}\bar{V}_n(\theta_0)) \in C)| = o(n^{-(s-2)/2}), \tag{3.11}$$

uniformly over $\theta_0 \in \Theta_c$ and $C \in \mathscr{C}_{dim(\theta)}$. Equations (3.10) and (3.11) now yield case (1) of the proof of (3.2).

Case (2). Suppose $\tau_n(\theta_0) = T_n(\theta_{0,r})$. Note that $\tau_n(\theta_0)$ is a function of $\hat{\theta}_n$. Taking a Taylor expansion of $\tau_n(\theta_0)/n^{1/2}$ about $\hat{\theta}_n = \theta_0$ to order $s-1$, where the highest-order term involves the expectation of the partial derivatives rather than the partial derivatives themselves, we obtain

$$\tau_n(\theta_0) = n^{1/2}(\widetilde{G}(\frac{1}{n}\bar{V}_n(\theta_0), \hat{\theta}_n - \theta_0) + \widetilde{\zeta}_n(\theta_0)), \tag{3.12}$$

where $\widetilde{G}$ is an infinitely differentiable function that does not depend on θ_0, $\widetilde{G}(\frac{1}{2}\bar{V}_n(\theta_0), 0) = 0$ for large n, $\widetilde{\zeta}_n(\theta_0)$ is the remainder term in the Taylor expansion, and $||\widetilde{\zeta}_n(\theta_0)|| = O(||\hat{\theta}_n - \theta_0||^s)$.

Combining (3.9) with (3.12) gives

$$\tau_n(\theta_0) = n^{1/2}(\widetilde{G}(\frac{1}{n}\bar{V}_n(\theta_0), F(\frac{1}{n}\bar{V}_n(\theta_0) + e_n(\theta_0)) + \widetilde{\zeta}_n(\theta_0)). \tag{3.13}$$

Again we apply Lemma 6 of Andrews and Lieberman [2002] with

$$A_n(\theta_0) = n^{1/2}\widetilde{G}(\frac{1}{n}\bar{V}_n(\theta_0), F(\frac{1}{n}\bar{V}_n(\theta_0)))$$

to obtain

$$\sup_{\theta_0 \in \Theta_c, C \in \mathscr{C}_{dim(\theta)}} | P_{\theta_0}(A_n(\theta_0)) + \widetilde{\zeta}_n(\theta_0)) \in C) - P_{\theta_0}(A_n(\theta_0)) \in C)| = o(n^{1-s/2}). \tag{3.14}$$

Define a function F' by $F'(x) = (\widetilde{G}(x), F(x))$. Then $F'(.)$ is infinitely differentiable and satisfies

$$F'(n^{-1}E_{\theta-0}\bar{V}_n(\theta_0)) = \widetilde{G}(n^{-1}E_\theta\bar{V}_n(\theta_0), F(\frac{1}{n}E_\theta\bar{V}_n(\theta_0))) = \widetilde{G}(\frac{1}{n}E_\theta\bar{V}_n(\theta_0), 0) = 0, \tag{3.15}$$

for all n large. Combining (3.14) and (3.15) gives the result of the theorem for the case $\tau_n(\theta_0) = T_n(\theta_{0,r})$. $\square$

Next we introduce additional notations used in Lemma 3.2 below which in turn is used in Theorem 3.1. The lemma is just Lemma 8 of Andrews and Lieberman [2002] restated here for convenience. For some $\delta > 0$, let $\Theta_c^+ = \{\theta \in \mathbb{R}^d : dist(\theta, \Theta_c) \leq \delta\}$ be a compact subset of Θ, the parameter space, that is slightly larger than Θ_c (where $dist(\theta, \Theta_c) = \inf\{||\theta - \theta_c|| : \theta_c \in \Theta_c\}$). $\square$

Lemma 3.2. *Suppose* $\sup_{\theta_0 \in \Theta_c} P_{\theta_0}(\hat{\theta}_n \notin B(\theta_0, \delta)) = o(n^{-(s-2)/2})$, *where* Θ_c *is a compact subset of* Θ *and* δ *is as in the definition of* Θ_c^+, *and* $\{\lambda_n(\theta) : n \geq 1\}$ *is a sequence of non-random real functions on* Θ_c^+ *that satisfies* $\sup_{\theta \in \Theta_c^+} |\lambda_n(\theta)| = o(n^{-(s-2)/2})$. *Then, for all* $\varepsilon > 0$, $\sup_{\theta_0 \in \Theta_c} P_{\theta_0}(|\lambda_n(\hat{\theta}_n)| > n^{-(s-2)/2}\varepsilon) = o(n^{-(s-2)/2})$.

We now present an Edgeworth expansion for the parametric bootstrap t statistic. We first introduce some additional notations. Let $H(x) = P_{\theta_0}(T_n(\theta_{0,r}) \leq x)$, $\widetilde{H}(x) = \Phi(x) + \sum_{j=1}^{s-2} n^{-j/2}P_j(\Delta, \bar{\kappa}_{n,s}(\theta_0))\Phi(x)$, $H^*(x) = P_{\hat{\theta}_n}^*(T_n^*(\hat{\theta}_{n,r}) \leq x)$, and $\widetilde{H}^*(x) = \Phi(x) + \sum_{j=1}^{s-2} n^{-j/2}P_j(\Delta, \bar{\kappa}_{n,s}(\hat{\theta}_n))\Phi(x)$, where $\widetilde{H}(x)$ and $\widetilde{H}^*(x)$ are formal Edgeworth expansions of $H(x)$ and $H^*(x)$, respectively.

Theorem 3.1. *Suppose Assumption* **I-VI** *of Andrews and Lieberman [2002] and* **A1-A3** *hold, the PML estimators* $\{\hat{\theta}_n \in \hat{\Theta}_n : n \geq 1\}$ *satisfy* Condition **C_s**, *and let* $s \geq 3$ *be an integer. Then, for all* $\varepsilon > 0$,

$$P_{\theta_0}\left(\sup_{x \in \mathbb{R}} |H^*(x) - \widetilde{H}^*(x)| > n^{1-s/2}\varepsilon\right) = o(n^{1-s/2}) \tag{3.16}$$

uniformly over $\theta_0 \in \Theta_c$.

Proof. We first establish the Edgeworth expansion for the t statistic, $T_n(\theta_{0,r})$, which now follows by utilizing the approximation of the t statistic by a smooth function of the LLDs given by Lemma

3.1 above. We need to show for all $\varepsilon > 0$ that

$$sup_{\theta_0 \in \Theta_c} \sup_{x \in \mathbb{R}} |H(x) - \widetilde{H}(x)| = o(n^{1-s/2}) \tag{3.17}.$$

To prove (3.17) above, we observe that,

$$|H(x) - \widetilde{H}(x)| \le |H(x) - P_{\theta_0}(n^{1/2}F(n^{-1}\bar{V}_n(\theta_0)) \le x)| + |P_{\theta_0}(n^{1/2}F(n^{-1}\bar{V}_n(\theta_0)) \le x) - \widetilde{H}(x)|. \tag{3.18}$$

The first term of the right hand side of (3.18) is $o(n^{1-s/2})$ because it is just the statement of (3.2) proved in Lemma 3.1, and the second term is also $o(n^{1-s/2})$ because it is shown in case (1) of the proof of the lemma that $n^{1/2}F(n^{-1}\bar{V}_n(\theta_0))$ possesses an Edgeworth expansion with an error $o(n^{1-s/2})$, and thus (3.17) follows.

Next we establish the Edgeworth expansion of the bootstrap t statistic, $T_n^*(\hat{\theta}_{n,r})$, as given in (3.16). Let

$$\lambda_n(\hat{\theta}_n) = \sup_{x \in \mathbb{R}} |P_{\hat{\theta}_n}^*(T_n^*(\hat{\theta}_{n,r}) \le x) - [\Phi(x) + \sum_{j=1}^{s-2} n^{-j/2} P_j(\Delta, \bar{\kappa}_{n,s}(\hat{\theta}_n))\Phi(x)|. \tag{3.19}$$

(3.16) now follows from Lemma 3.2 above upon checking the conditions of the lemma. The first condition of Lemma 3.2 holds by *Condition* $\mathbf{C_s}$ and the second condition holds by (3.17) above with Θ_c replaced by the compact set Θ_c^+. $\qquad\square$

References

[1] M.Aga, T. Sun, Parametric Bootstrap Confidence Intervals for Linear Regression Processes with long-memory Errors, *Sankhya: The Indian Journal of Statistics*, **69** (2007), Part 4, 615–634.

[2] D. Andrews, D. and O. Lieberman, Higher-order Improvements of the Parametric Bootstrap for Long-memory Gaussian Processes, *Cowles Foundation Discussion Paper No 1378*, (2007)

[3] R.N. Battacharya and J. K. Ghosh On the Validity of the Formal Edgeworth Expansion, *Annals of Statistics*, **6** (1978), 434-451.

[4] P.J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, *Second Edition*, (Springer, New York, 1991).

[5] P. Hall, *The Bootstrap and Edgeworth Expansion*, (Springer, New York, 1992).

[6] O. Lieberman, J. Rousseau, and D.M. Zucker, (2003), *Valid Edgeworth Expansions For the Maximum Likelihood Estimator of the Parameter of a Stationary, Gaussian, Strongly Dependent Processes*, Annals of Statistics, **31** (2003), 586-612.