OPTIMAL CONDITIONAL INFERENCE IN NEARLY-INTEGRATED AUTOREGRESSIVE PROCESSES (JOB MARKET PAPER)

RICHARD K. CRUMP

Dedicated to the memory of S. Lee Crump

Abstract. This paper considers point and interval estimation of the autoregressive parameter in a nearly-integrated first-order autoregression. Under a local-to-unity parameterization, we provide conditional inference procedures based on the Gaussian sufficient statistics. These procedures are optimal in terms of concentration around the true parameter value within a class of conditional estimators. Additionally, the proposed estimators minimize expected loss across a wide range of loss functions. We also generalize the method for models with a nonzero mean and for higher-order autoregressions.

1. Introduction

This paper is concerned with inference on the autoregressive parameter, ρ, in a nearly-integrated first-order autoregression (AR(1)). The degree of persistence in an autoregressive process is of practical importance in macroeconomics and finance where highly-persistent series appear as both independent and dependent variables in many models. Although the primary focus in the econometrics literature has been on tests of the unit-root hypothesis[1], a variety of point and interval estimators have been proposed. In general, these estimators do not enjoy explicit optimality properties[2]. In this paper, we propose confidence bounds and an associated median-unbiased estimator for the autoregressive parameter which enjoy optimality properties within a class of conditional procedures.

Date: November 5, 2008.

Key words and phrases. Nearly-integrated autoregressive process, Local-to-unity asymptotics, Median-unbiased estimation, Conditional prediction, Unit root.

† I would like to thank Michael Jansson for invaluable guidance and support. I would also like to thank Matias Cattaneo, Guido Imbens, Jim Powell, Thomas Rothenberg, and James Stock for helpful comments and suggestions. Jasjeet Sekhon and Rocio Titiunik graciously provided access to the Calgrid cluster.

[1]For surveys of the unit root testing literature see Stock (1994), Phillips and Xiao (1998), and Haldrup and Jansson (2006).

[2]An exception is Gushchin (1995).

A nearly-integrated AR(1) is characterized by a value of ρ close to 1. Thus, we adopt a local-to-unity parameterization, which re-parameterizes the autoregressive parameter, ρ as $\rho_n(c) = 1 + c/n$, where c, the local-to-unity parameter, is now of interest[3]. The restriction of ρ to an n^{-1}-neighborhood of unity (a unit root) is motivated from two different perspectives: first, there is strong empirical evidence that many economic and financial time series are highly persistent, suggesting that the value of ρ is equal to or near unity (Nelson and Plosser (1982), Schotman and van Dijk (1991)); second, it is well known that asymptotic results in an AR(1) model are discontinuous at $\rho = 1$, and this re-parameterization leads to asymptotic results which are continuous in the parameter c. Moreover, asymptotic theory based on the local-to-unity parameterization provides a better approximation for modest sample sizes when ρ is close to 1 despite the fact that standard asymptotic theory is available.

We restrict ourselves to methods based on the output from a least-squares regression (i.e., the Gaussian sufficient statistics). We then proceed conditionally on either of the two sufficient statistics. In one case this involves conditioning on the observed Fisher information. In the other case, the conditional procedure is asymptotically equivalent to procedures which condition on the (properly-scaled) final observation of the series. This has particular relevance to the problem of forecasting where the last observation is generally treated as fixed and so conditional procedures may be appealing. A variety of authors have considered predictive inference conditional on the final observation of an AR(1) process (see the discussion in Section 3). Our procedure satisfies this requirement asymptotically and enjoys optimality properties within this class of conditional procedures. Finally, conditioning on either of these statistics serves the purpose of removing the natural curvature in the Gaussian AR(1) exponential family and allows us to develop optimal procedures.

Many authors have proposed point and interval estimators for autoregressive models under a variety of assumptions with, or without imposing a local-to-unity parameterization. A common approach is to use knowledge of the distributions of available estimators or test statistics and invert these distributions to construct suitable estimators[4]. The work that is closest in spirit to the present paper is that of Stock (1991), Hansen (1999), and Elliott and Stock (2001). In a local-to-unity setting, Stock (1991) proposed interval estimators based on the inversion of the augmented Dickey-Fuller test statistic, while Elliott and Stock (2001) considered both sequential interval estimators and asymptotically valid interval estimators choosing to invert candidate test statistics with good power

[3]See Bobkoski (1983), Cavanagh (1985), Chan and Wei (1987), Phillips (1987), and Chan (1988).

[4]These include Dufour (1990), Stock (1991), Kiviet and Phillips (1992), Rudebusch (1992), Andrews (1993), Andrews and Chen (1994), Hansen (1999), and Elliott and Stock (2001).

properties. Meanwhile, Hansen (1999) introduced a "grid bootstrap" interval estimator which provides correct asymptotic coverage regardless of whether the autoregressive model is highly-persistent or not. Also, in a related context, Andrews (1993) constructs an exactly median-unbiased estimator of ρ based on inversion of the distribution of the least-squares estimator under the assumption of Gaussian errors. Our approach is similar to these papers. The primary difference is that although no uniformly most powerful test exists in general (and consequently no optimal confidence interval based on its inversion), by proceeding conditionally we are able to construct a uniformly most powerful *conditional* test statistic. This statistic may be inverted to obtain a median-unbiased point estimator and confidence bounds that are optimal in terms of concentration around the true parameter value within a certain class of conditional estimators. We will show that this property implies that the estimators minimize expected loss for a wide range of loss functions.

Median-unbiased estimators have the appealing property that the parameter of interest is overestimated or underestimated with about the same probability. This is particularly desirable in the case of an AR(1) as the least-squares estimator is well known to suffer downward bias especially when the autoregressive parameter is near 1. Additionally, median-unbiased estimators are equivariant to any monotonic transformation and so we may provide median-unbiased estimators for other magnitudes of interest. This is an appealing property as errors in estimation may be magnified when forming other expressions of interest (Patterson (2000)). Finally, these inference procedures have correct coverage both conditionally and unconditionally and so we provide alternative options to existing unconditional techniques.

To enable computation of our estimators we provide asymptotic formulas for the relevant density and distribution functions. Thus, our estimators may be implemented with straightforward numerical methods. Preliminary Monte Carlo evidence suggests that the procedure is on par with existing procedures for modest sample sizes. In addition, we are able to generalize the results to incorporate higher-order autoregressions with a non-zero mean. However, a drawback to our proposed procedure is that it does not retain its optimality properties in the presence of a linear time trend.

The paper is organized as follows. Section 2 introduces the model under study and describes the finite-sample properties of the model and sets the stage for the discussion of the asymptotic results. Section 3 reviews the asymptotic properties of our procedure. In Section 4 we consider generalizations of the main results to higher-order autoregressions with a non-zero mean and discuss other points of interest. We evaluate the methods introduced relative to popular competitors in a Monte Carlo experiment summarized in Section 5. Section 6 concludes. All proofs are provided in an Appendix.

2. Model & Finite-Sample Theory

We observe a univariate time series $X := (X_1, \ldots, X_n)$ generated as,

$$X_t = \rho X_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, n, \tag{2.1}$$

where $\rho \in (-1, 1]$ is the parameter of interest and $\{\varepsilon_t\}$ are a sequence of unobserved *i.i.d.* mean-zero error terms. In addition, we make the following assumptions.

Assumption 1. *The initial condition satisfies* $X_0 = x_0 = 0$.

Assumption 2. *The sequence of error terms satisfy* $\varepsilon_t \sim_{iid} \mathcal{N}(0, \sigma^2)$ *for* $t = 1, \ldots, n$ *where* $\sigma^2 \in (0, \infty)$ *is a known parameter.*

In later sections we will relax these assumptions. As discussed in the introduction we also introduce to the model the local-to-unity parametrization,

$$\rho_n(c) := 1 + cn^{-1}, \qquad c \in \mathcal{C} := \mathbb{R}_-.$$

We restrict ρ to an n^{-1}-neighborhood of unity (a unit root)[5]. Under these assumptions the joint distribution of the observed time series may be written as,

$$f_X(x; c) = \eta(x) \exp\left\{ cT_n(x) - \frac{1}{2}c^2 U_n(x) \right\}, \tag{2.2}$$

$$T_n(x) = \frac{1}{n\sigma^2} \sum\nolimits_{t=1}^{n} x_{t-1} \Delta x_t, \qquad U_n(x) = \frac{1}{n^2\sigma^2} \sum\nolimits_{t=1}^{n} x_{t-1}^2. \tag{2.3}$$

By the factorization criterion the sufficient statistics are $T_n(X)$ and $U_n(X)$. Since we have assumed that σ^2 is known, we have two sufficient statistics but only one parameter. Thus, the distribution of X is a member of a $(2, 1)$-curved exponential family (Efron (1975, 1978), Barndorff-Nielsen and Cox (1994)). Curved exponential families generally do not possess a complete sufficient statistic or a monotone-likelihood ratio (MLR). These properties are the primary source of optimality results in full exponential families. For example, in the model under study, Elliott *et al.* (1996) showed that there does not exist a uniformly most powerful test, even asymptotically, for tests of the unit-root null

[5] The specific choice of localizing sequence, n^{-1}, ensures that the ratio of likelihoods for any two values of $c \in \mathcal{C}$ in model (2.1) weakly converges to a random variable which is itself a likelihood ratio. This property is referred to as contiguity and is essentially an asymptotic version of absolute continuity. This enables us to utilize the limits of experiments approach to characterize the asymptotic properties (see Le Cam (1982), Jeganathan (1995), Le Cam and Yang (2000)). It also allows us to demonstrate key properties of our methods in the finite-sample case, and draw clear analogies in the case of large samples.

hypothesis $c = 0$. Elliott and Stock (2001) showed that this result generalizes so that there does not exist an optimal interval estimator for c. The approach taken in this paper is to remove the curvature by conditioning on one of the two sufficient statistics. This allows us to recover the MLR property which leads directly to a uniformly most powerful *conditional* test statistic. We may then invert this test statistic to construct interval estimators.

We might consider conditioning on either $T_n(X)$ or $U_n(X)$ (the observed Fisher information). The results of this section are similar for either choice and so we will provide results for conditioning on $T_n(X)$. We will postpone the discussion of motivation for this conditioning until later sections. In Section 3 we will discuss motivation for conditioning on $T_n(X)$ and later in Section 4 we will discuss conditioning on $U_n(X)$.

In Lemma 2 (in the Appendix) we show that the conditional density of $U_n(X)$ given $T_n(X)$ is of the form,

$$f_n(u_n|\, t_n; c) = g_n(c, t_n) \exp\left\{-\frac{1}{2}c^2 u_n\right\} f_n(u_n|\, t_n; 0)\,. \tag{2.4}$$

Since $c \leq 0$, this family of conditional densities has a monotone-likelihood ratio in u_n. It is this property that allows us to construct confidence bounds (and median-unbiased estimators) with demonstrable optimality properties (Pfanzagl (1970, 1979)). Suppose we observe $S_n(X) := (U_n(X), T_n(X)) = (u_n, t_n)$ then we will show that, $\hat{c}_{n,1-\alpha}$, the solution to the equation[6],

$$F_n(u_n|\, t_n; \hat{c}_{n,1-\alpha}) = \int_0^{u_n} f_n(w|\, t_n; \hat{c}_{n,1-\alpha})\, dw = 1-\alpha, \tag{2.5}$$

is the uniformly most accurate lower confidence bound conditional on $T_n(X) = t_n$ for c with confidence coefficient $1-\alpha$. Alternatively we may write,

$$\hat{c}_{n,1-\alpha}(u_n, t_n) := \{c : F_n(u_n|\, t_n; c) = 1-\alpha\}\,. \tag{2.6}$$

The inversion of the conditional distribution function that is used in equations (2.5) and (2.6) may be summarized as follows: first, the conditional distribution function, $F_n(u_n|\, t_n; c)$, will be shown to be continuous and strictly increasing and so we may define the conditional quantile function $q_{1-\alpha}(t_n, c)$ by the requirement $F_n(q_{n,1-\alpha}(t_n, c)|\, t_n; c) = 1-\alpha$; secondly, by the aforementioned MLR property, $c \mapsto q_{n,1-\alpha}(t_n, c)$ is strictly increasing (see Lemma 4 in the Appendix) and so we may define our estimator, $\hat{c}_{n,1-\alpha}(u_n, t_n)$ by the requirement $q_{n,1-\alpha}(t_n, \hat{c}_{n,1-\alpha}(u_n, t_n)) = u_n$. Then we have that,

$$P_{c,n}(\hat{c}_{n,1-\alpha}(U_n(X), T_n(X)) \leq c|\, t_n) = P_{c,n}(U_n \leq q_{n,1-\alpha}(t_n, c)|\, t_n) = 1-\alpha, \tag{2.7}$$

[6]Throughout the text, expressions involving conditonal expectations will be understood to hold almost surely.

where $P_{c,n}$ is the probability measure associated with the distribution of X with true parameter value c[7]. The first equality follows by the definition of $\hat{c}_{n,1-\alpha}(u_n, t_n)$ and the second equality follows by the definition of $q_{n,1-\alpha}(t_n, c)$. Equation (2.7) implies that $\hat{c}_{n,1-\alpha}(S_n(X))$ is a valid lower-confidence bound for the parameter c conditional on $T_n(X) = t_n$. In fact, by the law of iterated expectations, $\hat{c}_{n,1-\alpha}(S_n(X))$ is also a valid lower confidence bound unconditionally[8].

Uniformly most accurate lower confidence bounds are characterized by their concentration below the true parameter value. More specifically,

$$P_{c,n}\{\hat{c}_{n,1-\alpha}(S_n(X)) \leq r | T_n(X)\} \leq P_{c,n}\{\tilde{c}_{n,l}(X) \leq r | T_n(X)\}, \qquad r < c, \tag{2.8}$$

for all $c \in \mathcal{C}$, where $\tilde{c}_{n,l}(x)$ is any other lower confidence bound with confidence coefficient $1-\alpha$ conditional on $T_n(X) = t_n$. In words, equation (2.8) says that among all lower bounds with conditional confidence coefficient $1-\alpha$, $\hat{c}_{n,1-\alpha}$ has the property that it minimizes the conditional probability of underestimating the true parameter value, c; $\hat{c}_{n,1-\alpha}$ is most concentrated below the true parameter value of any lower confidence bound in this class. Equation (2.8) also implies that the unconditional probability of underestimating c is minimized, but again in the class of conditional lower confidence bounds. We might alternatively be interested in ranking lower confidence bounds by the expected loss from underestimating c. Consider the following class of loss functions:

Definition 1. Let $\mathcal{L}$ be the class of loss functions $\ell_c(\cdot) : \mathbb{R} \to \mathbb{R}_+$, $c \in \mathcal{C}$ which satisfies $\ell_c(d) = L(c, d-c)$ where $L(c, 0) = 0$, and for each c, $L(c, d-c)$ is nondecreasing in $d-c$ for $d-c > 0$ and nonincreasing in $d-c$ for $d-c < 0$[9].

$\mathcal{L}$ is the class of monotone loss functions discussed in, for example, Andrews and Phillips (1987) and Lehmann and Romano (2005, 76). In this class the distinguishing characteristic is that loss is nondecreasing as an estimator takes on values farther away from the true value of the parameter. We also allow the loss function to vary with the value taken on by the parameter c[10]. In order to focus on the loss only from underestimating c we may construct the loss function $\ell_c^*(d) = \ell_c(d)\,1\{d < c\}$ where $\ell_c \in \mathcal{L}$. Because the uniformly most accurate confidence bound is most concentrated

[7]We follow the analagous subscript convention for the expectation operator.

[8]The corresponding results for upper bounds follow analogously and so in the sequel we will, without loss of generality, only provide results for lower confidence bounds.

[9]We are actually able to generalize the class of loss functions to include dependence on the statistic $T_n(X)$. We omit this generalization for notational clarity and revisit this issue in Remark 2.

[10]For example, the class includes weighted squared error loss, $\ell_c(d) = w(c)(d-c)^2$ where the weight function is positive and finite but may take on different values according to the value of c. $\mathcal{L}$ also includes a variety of asymmetric loss functions such as linear exponential loss (linex loss), $\ell_c(d) = b[\exp\{a(d-c)\} - a(d-c) - 1]$ where $a \neq 0$ and $b > 0$.

below c, it has the property that it minimizes expected loss for any $\ell_c^*(d)$ constructed from $\ell_c \in \mathcal{L}$. This is summarized in the following theorem.

Theorem 1. *Suppose Assumptions 1 & 2 hold and* $0 < \alpha < 1/2$. *Then, the solution to equation (2.5),* $\hat{c}_{n,1-\alpha}(S_n(X))$ *is the uniformly most accurate lower confidence bound for* c*, conditional on* $T_n(X)$ *with confidence coefficient* $1-\alpha$. *Moreover, we have that,*

$$E_{c,n}\left[\ell_c^*\left(\hat{c}_{n,1-\alpha}(S_n(X))\right)\middle| T_n(X)\right] \leq E_{c,n}\left[\ell_c^*(\tilde{c}_{n,l}(X))\middle| T_n(X)\right],$$

where $\tilde{c}_{n,l}(x)$ *is any other lower confidence bound with confidence coefficient* $1-\alpha$ *conditional on* $T_n(X)$, $\ell_c^*(d) = \ell_c(d)\,1\{d < c\}$ *and* $\ell_c(\cdot) \in \mathcal{L}$.

A corollary to Theorem 1 is that if we combine this result with the analogous result for upper confidence bounds with each having confidence coefficient $1/2$ then the bounds coincide and the estimator $\hat{c}_n := \hat{c}_{n,1/2}$ is median unbiased. Specifically, $\hat{c}_n$ satisfies,

$$P_{c,n}\{\hat{c}_n(S_n(X)) \geq c | T_n(X)\} \geq 1/2 \quad \text{and} \quad P_{c,n}\{\hat{c}_n(S_n(X)) \leq c | T_n(X)\} \geq 1/2.$$

Median-unbiased estimators have the appealing property that the parameter of interest is overestimated or underestimated with about the same probability. Moreover, in perfect analogy with the summary of uniformly most accurate lower and upper confidence bounds, $\hat{c}_n$ is most concentrated around c among all estimators which are conditionally (on $T_n(X)$) median unbiased. Then applying Theorem 1 we have (see Remark 3 in the Appendix) that

$$E_{c,n}\left[\ell_c(\hat{c}_n(S_n(X)))| T_n(X)\right] \leq E_{c,n}\left[\ell_c(\tilde{c}_n(X))| T_n(X)\right]$$

for all $\ell_c \in \mathcal{L}$, where $\tilde{c}_n(X)$ is any other estimator which is median unbiased conditional on $T_n(X)$.

By the equivariance property of median-unbiased estimators we may also construct corresponding optimal median-unbiased estimators for other parameters of interest such as ρ, or the cumulative impulse-response function, $(1-\rho)^{-1}$, which are monotone functions of the parameter c. This is an appealing property as errors in the estimate of c may be more pronounced when forming other expressions of interest, as discussed in Patterson (2000). Again, these estimators possess optimality properties only in the class of estimators which are median unbiased conditional on the statistic $T_n(X)$, however they retain the property of median unbiasedness when considered as unconditional estimators.

3. Asymptotic Theory

In this section we characterize the asymptotic properties of the model in equation (2.1). The results will be based on the joint and conditional distribution of the asymptotic counterparts to the sufficient statistics $S_n(X)$.

Lemma 1. *Suppose Assumptions 1 & 2 hold. Then,* $(U_n, T_n) \Rightarrow_c (U^c, T^c) =: S^c$ *where*

$$U^c = \int_0^1 B_c(r)^2 \, dr, \qquad T^c = cU^c + \int_0^1 B_c(r) \, dB(r),$$

and $\{B_c(r), 0 \leq r \leq 1\}$ *is defined by the stochastic differential equation,* $dB_c(r) = cB_c(r)\,dr + dB(r)$ *with* $B_c(0) = 0$, *where* $\{B(r), 0 \leq r \leq 1\}$ *is a standard Brownian motion.*

The random process $\{B_c(r), 0 \leq r \leq 1\}$ is known as the Ornstein–Uhlenbeck process and is the continuous time analogue to a highly-persistent autoregression. In Lemma 1, $\Rightarrow_c$ indicates weak convergence with respect to the probability measure $P_{c,n}$. In the special case of a random walk ($c = 0$) we will drop the superscript so we have, $(U_n, T_n) \Rightarrow_0 (U, T) =: S$. The properties of model (2.1) for fixed n suggest that we might generate an asymptotic version of Theorem 1 by constructing our estimator as in equation (2.5), but with $F_n(u_n | t_n; c)$ replaced by the conditional distribution function of U^c given T^c. To do so we must first demonstrate that the exponential family representation is maintained in the limit.

Lemma 2. *Suppose Assumptions 1 & 2 hold.*

(1) *The joint distribution of* $S^c = (U^c, T^c)$ *is a* $(2,1)$*-curved exponential family with density,*

$$f(u, t; c) = \exp\left\{ct - \frac{1}{2}c^2 u\right\} f(u, t; 0),$$

where $f(u, t; 0)$ *is the density of* S^c *when* $c = 0$.

(2) *For* $c < 0$, *the conditional distribution of* U^c *given* $T^c = t$ *is an exponential family with density,*

$$f(u | t; c) = g(c, t) \exp\left\{-\frac{1}{2}c^2 u\right\} f(u | t; 0),$$

where,

$$g(c, t) = \sigma_c \exp\left\{-\frac{c}{2} + \frac{c\tau^2}{2} - \frac{\tau^2}{2\sigma_c^2}\left(\sigma_c^2 - 1\right)\right\}, \qquad \sigma_c^2 = \frac{\exp(2c) - 1}{2c},$$

$$f\left(u|\,t;0\right)=\exp\left\{\frac{1}{2}\tau^{2}\right\}\sum_{j=0}^{\infty}\frac{\left(-\tau^{2}\right)^{j}}{j!}\sum_{k=0}^{\infty}\frac{\left(j+\frac{1}{2}\right)_{k}}{\sqrt{\pi}k!}u^{-\frac{1}{2}\left(\frac{5}{2}+j\right)}\times$$

$$\exp\left\{-\frac{\delta\left(j,k;\tau\right)^{2}}{16u}\right\}D_{3/2+j}\left(\frac{\delta\left(j,k;\tau\right)}{2\sqrt{u}}\right),$$

is the conditional distribution of U given $T=t$ when $c=0$ and

$$\delta\left(j,k;\tau\right)=1+\tau^{2}+4\left(j+k\right),\qquad\tau^{2}=2t+1.$$

$D_{\nu}\left(z\right)$ is the parabolic cylinder function[11] *and $\left(\nu\right)_{k}$ is the Pochhammer symbol, $\left(\nu\right)_{k}:=\Gamma\left(\nu+k\right)/\,\Gamma\left(\nu\right)$ where $\nu,\nu+k\notin\mathbb{Z}_{-}/\left\{0\right\}$.*

The similarities between Lemma 7 (in the Appendix) and Lemma 2 are clear. The main difference is that in the limit we are able to characterize the asymptotic counterparts of $g_{n}\left(c,t_{n}\right)$ and $f_{n}\left(u_{n}|\,t_{n};c\right)$ in a tidy form. Unfortunately, there does not seem to be a comparable expression for the conditional distribution function $F\left(u|\,t;c\right)$ as for $f\left(u|\,t;c\right)$ except in the special case when $c=0$.

Remark 1. *In the unconditional case, even asymptotically, there does not exist a UMP one-sided test of the null hypothesis $c=0$ versus the alternative of $c<0$. However, in the conditional case the asymptotically UMP one-sided test rejects for small values of the statistic U^{c}. The proof of Lemma 2 may be altered slightly to show that the conditional distribution function, $F\left(u|\,t;0\right)$, has the following form,*

$$F\left(u|\,t;0\right)\;=\;2\exp\left\{\frac{\tau^{2}}{2}\right\}\sum_{j=0}^{\infty}\frac{\left(-\tau^{2}\right)^{j}}{j!}\sum_{k=0}^{\infty}\frac{\left(j+\frac{1}{2}\right)_{k}}{\sqrt{\pi}k!}u^{-\left(j/2+1/4\right)}\times$$

$$\exp\left\{-\frac{\delta\left(j,k;,\tau\right)^{2}}{16u}\right\}D_{j-1/2}\left(\frac{\delta\left(j,k;,\tau\right)}{2\sqrt{u}}\right).$$

Thus, if we observe $\left(U^{c},T^{c}\right)=\left(u,t\right)$ we may construct the approximate p-value for the asymptotic unit-root test by $F\left(u|\,t;0\right)$.

In perfect analogy with equation (2.5) we may define $\hat{c}_{1-\alpha}$ as the solution to,

$$F\left(u|\,t;c\right)=\int_{-\infty}^{u}f\left(w|\,t;c\right)dw=1-\alpha, \tag{3.1}$$

and for the special case of the median-unbiased estimator, $\hat{c}:=\hat{c}_{1/2}$. In practice, we will have to calculate this integral numerically. However, the double sums in the definition of the density converge quickly and the density is sufficiently well-behaved that this does not pose a significant problem.

[11]See Borodin and Salminen (2002, 639-40).

By Itô's Lemma, T^c may also be written as $T^c = (B_c(1)^2 - 1)/2$ which is the continuous-time analogue to,

$$2\sigma^2 T_n(x) = \frac{x_n^2}{n} - \frac{1}{n}\sum\nolimits_{t=1}^{n} (\Delta x_t)^2 = \frac{x_n^2}{n} - \sigma^2 + o_p(1),$$

which implies that conditioning on T^c is the asymptotic analogue to conditioning on the appropriately scaled squared value of the final observation of the series Z_n^2, where $Z_n := X_n / (\sigma\sqrt{n}) \Rightarrow B_c(1)$. A symmetry argument can be made to show that conditioning on $B_c(1)^2$ involves no loss of information relative to conditioning on $B_c(1)$; in other words the conditional distribution of U^c given $B_c(1)$ is identical to the conditional distribution of U^c given $B_c(1)^2$ (see Remark 4 in the Appendix). From this result we may show that asymptotically it makes no difference whether we condition on Z_n or $T_n(X)$ and so the large-sample analogue of the small-sample results from Section 2 become relevant to the case of conditioning on the (properly-scaled) final observation. In order to avoid complications having to do with conditional convergence we follow Jansson and Moreira (2006) and note that $\tilde{c}_n$ satisfies,

$$E_{c,n}\left[\left(1\left\{\tilde{c}_n(S_n(X)) \leq c\right\} - (1-\alpha)\right)\middle| Z_n\right] = 0,$$

if and only if,

$$E_{c,n}\left[\left(1\left\{\tilde{c}_n(S_n(X)) \leq c\right\} - (1-\alpha)\right) h(Z_n)\right] = 0, \quad \forall h \in \mathcal{C}_b(\mathbb{R}), \tag{3.2}$$

where $\mathcal{C}_b(\mathbb{R})$ is the class of all bounded, continuous real-valued functions on $\mathbb{R}$. Correspondingly, we say that the sequence $\{\tilde{c}_n\}$ is a conditional asymptotic lower confidence bound if it satisfies,

$$\lim_{n\to\infty} E_{c,n}\left[\left(1\left\{\tilde{c}_n(S_n(X)) \leq c\right\} - (1-\alpha)\right) h(Z_n)\right] = 0 \qquad \forall h \in \mathcal{C}_b(\mathbb{R}). \tag{3.3}$$

That we might consider inference in a first-order autoregressions conditional on the final observation has been pointed out at least as early as Phillips (1979)[12]. Phillips (1979) observes that "...it is this case which is of most interest since, in practice, we do forecast with given final period values of the endogenous variables." Specifically, forecasts for model (2.1) are generally constructed treating the final observation, X_n, as fixed and so it may be argued that conditional inference is more representative of the "experiment" actually conducted. The following theorem shows that the solution to equation (3.1)

[12]A number of authors have considered predictive inference conditional on the final observation in autoregressive models. See Stine (1987), Kabaila (1993), Breidt *et al.* (1995), Barndorff-Nielsen and Cox (1996), Kabaila (1999), Gospodoniv (2002), and Vidoni (2004). In a related context, Elliott (2006) explored the role of the final observation when a practitioner is pre-testing for a unit root.

enjoys asymptotic optimality properties in the class of sequences of estimators which satisfy equation (3.3).

Theorem 2. *Suppose Assumptions 1 & 2 hold,* $0 < \alpha < 1/2$*, and* $\{\tilde{c}_n(\cdot)\}$ *is a sequence of estimators which satisfy equation (3.3) and are uniformly tight. Then,*

$$\liminf_{n\to\infty} E_{c,n}\left[\ell_c^*\left(\tilde{c}_n\left(S_n\left(X\right)\right)\right)\right] \geq \lim_{n\to\infty} E_{c,n}\left[\ell_c^*\left(\hat{c}_{1-\alpha}\left(S_n\left(X\right)\right)\right)\right] = E_c\left[\ell_c^*\left(\hat{c}_{1-\alpha}\left(S^c\right)\right)\right],$$

where $\ell_c^*(d) = \ell_c(d)\,1\{d < c\}$*,* $\ell_c(\cdot) \in \mathcal{L}$*, and* $\ell_c(\cdot)$ *is bounded and discontinuous on a set of probability zero.*

Theorem 2 is the asymptotic analogue to Theorem 1. However, in the asymptotic case we may no longer consider the entire class $\mathcal{L}$, but instead only those loss functions which are bounded and continuous with probability one. However, if we choose $\ell_c(d) = 1\{d \leq r\}$ where $r < c$ then we have that $\hat{c}$ satisfies,

$$\liminf_{n\to\infty} P_{c,n}\left(\tilde{c}_n\left(S_n\left(X\right)\right) \leq r\right) \geq \lim_{n\to\infty} P_{c,n}\left(\hat{c}_{1-\alpha}\left(S_n\left(X\right)\right) \leq r\right) = P_c\left(\hat{c}\left(S^c\right) \leq r\right), \tag{3.4}$$

which may be compared to equation (2.8). Equation (3.4) says that in the class of estimators that are conditional (on Z_n) asymptotic lower confidence bounds as defined by equation (3.3), $\hat{c}_{1-\alpha}$ is asymptotically most concentrated below the true parameter value, c. Moreover, we may follow the same logic as in Section 2 to show that $\hat{c}$ is the (conditional) median-unbiased estimator which is asymptotically most concentrated around c.

Remark 2. *As discussed in Section 2, we may generalize the class of loss function we consider to include dependence on the sufficient statistic* $T_n(X)$*. In particular, because of the asymptotic equivalence between conditioning on* $T_n(X)$ *and* Z_n*, we may consider loss functions of the form,* $\tilde{\ell}_c(d,z)$ *where* $\tilde{\ell}_c(0,z) = 0$*,* $\tilde{\ell}_c(d,0) = 0$ *and for fixed* c *and* z*,* $\tilde{\ell}_c(d,z)$ *is nondecreasing in* $d-c$ *for* $d-c > 0$ *and nonincreasing in* $d-c$ *for* $d-c < 0$*. Then, we may alter the proof of Theorem 2 to obtain,*

$$\liminf_{n\to\infty} E_{c,n}\left[\tilde{\ell}_c^*\left(\tilde{c}_n\left(S_n\right), Z_n\right)\right] \geq \lim_{n\to\infty} E_{c,n}\left[\tilde{\ell}_c^*\left(\hat{c}_{1-\alpha}\left(S_n\right), Z_n\right)\right],$$

where $\tilde{\ell}_c^*(d,z) = \tilde{\ell}_c(d,z)\,1\{d < c\}$*. In particular, a reasonable metric to compare forecasts of an autoregressive series would be to rank forecast procedures by the concentration of the normalized prediction error,* $e_c(d, Z_n) := (d-c)\,Z_n$*, around the value of zero. Thus, we would consider loss functions of the form,* $\tilde{\ell}_c(d,z) = 1\{e_c(d,z) \leq r_0\}$ *where* $r_0 < 0$*, which yields,*

$$\liminf_{n\to\infty} P_{c,n}\left[e_c\left(\tilde{c}_n\left(S_n\right), Z_n\right) \leq r_0\right] \geq \lim_{n\to\infty} P_{c,n}\left[e_c\left(\hat{c}_{1-\alpha}\left(S_n\right), Z_n\right) \leq r_0\right].$$

Thus $\hat{c}_{1-\alpha}$ produces a predictive lower bound that is most concentrated below a normalized prediction error of zero among sequences of estimators which satisfy equation (3.3). Correspondingly, $\hat{c}$ is the median-unbiased estimator that most concentrates the normalized prediction error around zero in the class of asymptotic conditional median-unbiased procedures.

4. Discussion and Extensions

Before considering possible extensions of the procedures introduced in this paper we show that we may replace σ^2 with a consistent estimator without compromising the asymptotic results. If we define the feasible counterpart of $S_n(X)$ as

$$\hat{S}_n\left(X;\hat{\sigma}^2\right) := \left(\hat{U}_n\left(X;\hat{\sigma}^2\right), \hat{T}_n\left(X;\hat{\sigma}^2\right)\right),$$

where we have replaced σ^2 in equation (2.3) by a consistent estimator $\hat{\sigma}^2$.

Theorem 3. *Suppose Assumptions 1 & 2 hold, $0 < \alpha < 1/2$, and that $\hat{\sigma}^2$ is any consistent estimator of σ^2. Then,*

$$\lim_{n\to\infty} E_{c,n}[\ell_c^*(\hat{c}_{1-\alpha}(\hat{S}_n(X;\hat{\sigma}^2)))] = E_c\left[\ell_c^*\left(\hat{c}_{1-\alpha}\left(S^c\right)\right)\right],$$

where $\ell_c^*(d) = \ell_c(d)\, 1\{d < c\}$, $\ell_c(\cdot) \in \mathcal{L}$, and $\ell_c(\cdot)$ is bounded and discontinuous on a set of probability zero.

4.1. Conditioning on Expected Information.

As mentioned in Section 2 we may also remove the curvature in the curved exponential family by conditioning on the sufficient statistic $U_n(X)$ (or in the asymptotic case, U^c). In the finite-sample case this is equivalent to conditioning on the observed Fisher information. An appealing property of proceeding under this conditioning strategy is that we may now consider nearly-explosive models (ρ slightly above or equal to unity) as well, since we no longer require a sign restriction on c to obtain optimality results (i.e., $\mathcal{C} = \mathbb{R}$). However, we no longer have a clear motivation for the conditioning other than for the explicit goal of removing the curvature in the model. Unfortunately, the conditional density does not appear to have a tidy form as in Lemma 2 and so we must use the general formula,

$$f(t|\, u; c) = \frac{f(u,t;c)}{\int_{-\infty}^{\infty} f(u,t;c)\, dt},$$

where the form of $f(u,t;c)$ may be found in Remark 5 in the Appendix. An initial numerical inspection of the properties of this estimator suggest that the numerical properties are unstable and so we leave it to future research.

4.2. **AR$(p+1)$ with Unknown Mean.** In this section we generalize model (2.1) to accommodate more realistic circumstances. To differentiate this section from the previous sections we now observe a univariate time series $Y = (Y_1, \ldots, Y_n)$, generated as,

$$Y_t = \mu + U_t, \qquad (1-\rho L)\,\gamma(L)\,U_t = \varepsilon_t, \qquad t = 1, \ldots, n, \tag{4.1}$$

where $\mu \in \mathbb{R}$ is the mean parameter, $\rho \in (-1, 1]$ is the parameter of interest, $\{\varepsilon_t\}$ are a sequence of unobserved *i.i.d.* mean-zero error terms, and L is the lag operator[13]. In addition to Assumption 2 we make the following assumptions,

Assumption 3. *The pth degree polynomial $\gamma(z) = 1 - \gamma_1 z - \cdots - \gamma_p z^p$ satisfies $\gamma(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$.*

Assumption 4. *The initial conditions satisfy $U_{-i} = o_p(\sqrt{n})$ for $i = 0, \ldots, p$.*

Assumption 3 implies that if $\rho = 0$ then $\{\gamma(L)\,U_t\}$ is causal with respect to $\{\varepsilon_t\}$ (i.e., $\gamma(L)$ may be inverted to produce an infinite-order moving average process with absolutely-summable (and one-summable) coefficients). Thus, Y_t is an AR$(p+1)$ process with nonzero mean and (possibly) one unit-root.

The next step is to discuss the localization techniques we utilize in this section. We follow the presentation in Jansson (2008), implementing the local-to-unity parameterization as in the previous sections along with,

$$\mu = \mu_n(m) = \mu_0 + m, \qquad \gamma(L) = \gamma_n(L; \pi_1, \ldots, \pi_p) = \gamma_0(L) + n^{-1/2}\pi(L), \tag{4.2}$$

where $\gamma_0(L) = 1 - \gamma_{0,1}L - \cdots - \gamma_{0,p}L^p$ is a *known* lag polynomial, $\pi(L) = -\pi_1 L - \cdots - \pi_p L^p$, and $(m, \pi')' = (m, \pi_1, \ldots, \pi_p)'$ are *unknown* nuisance parameters. Without loss of generality we may assume that $\mu_0 = 0$.

Let $P_{c,m,\pi,n}$ and $P_{0,0,0,n}$ denote the distribution of $(Y_1, \ldots, Y_n)$ under the localization parameters (c, m, π) and $(0, 0, 0)$, respectively. Also, let $O_{P_{c,m,\pi,n}}(1)$ and $o_{P_{c,m,\pi,n}}(1)$ indicate that a sequence is $O_p(1)$ and $o_p(1)$ with respect to the sequence of probability measures $\{P_{c,m,\pi,n}\}$. Under this specification we have,

Lemma 3. *Suppose Assumptions 2, 3, & 4 hold in model (4.1).*

(1) *The log-likelihood ratio satisfies,*

$$\begin{aligned} & L(c, m, \pi) - L(0, 0, 0) \\ = \ & cT_n(y) - \frac{1}{2}c^2 U_n(y) - mV_{1n}(y) - \frac{1}{2}m^2 V_2 + \pi' W_{1n}(y) - \frac{1}{2}\pi' W_{2n}(y)\,\pi + o_{P_{0,0,0,n}}(1), \end{aligned}$$

[13] We discuss the case of a linear time trend, $Y_t = \mu + \beta t + U_t$, later in this section.

where

$$T_n(y) = \frac{1}{\sigma^2 n} \sum_{t=p+2}^{n} x_{t-1} \Delta x_t, \qquad U_n(y) = \frac{1}{\sigma^2 n} \sum_{t=p+2}^{n} x_{t-1}^2,$$

$$\sigma^2 V_{1n}(y) = y_1 + \sum_{j=1}^{p} \Delta x_{1+j} \gamma_{0,j}, \qquad \sigma^2 V_2 = 1 + \sum_{j=1}^{p} \gamma_{0,j}^2,$$

$$W_{1n}(y) = \frac{1}{\sigma^2 n^{1/2}} \sum_{t=p+2}^{n} (\Delta y_{p,t-1})' \Delta x_t, \quad W_{2n}(y) = \frac{1}{\sigma^2 n} \sum_{t=p+2}^{n} (\Delta y_{p,t-1})(\Delta y_{p,t-1})',$$

and

$$\Delta y_{p,t-1} = (\Delta y_{t-1}, \ldots, \Delta y_{t-p})', \qquad x_t = \left[\frac{\gamma_0(L)}{L^t}\right]_- L^t y_t,$$

where $[A(L)]_-$ *is defined to be the lag polynomial with terms containing nonnegative powers of* L *dropped.*

(2) *Under the measure* $P_{0,0,0,n}$ *we have,*

$$(U_n(y), T_n(y)) \Rightarrow_{(0,0,0)} (U, T),$$

$$V_{1n}(y) \Rightarrow_{(0,0,0)} V_1 \sim \mathcal{N}(0, V_2), \qquad V_2 := \frac{1}{\sigma^2}\left[1 + \sum_{j=1}^{p} \gamma_{j,0}^2\right],$$

$$W_{1n}(y) \Rightarrow_{(0,0,0)} W_1 \sim \mathcal{N}(0, W_2), \qquad W_{2n}(y) = W_2 + o_{P_{0,0,0}}(1)$$

$$W_2 := E\left[\left(\gamma_0(L)^{-1} \varepsilon_{p,t-1}\right)\left(\gamma_0(L)^{-1} \varepsilon_{p,t-1}\right)'\right],$$

where

$$\varepsilon_{p,t-1} = (\varepsilon_{t-1}, \ldots, \varepsilon_{t-p})',$$

and $\{(U,T), V, W_1\}$ *are pairwise independent.*

(3) *The sequences of probability measures* $P_{c,m,\pi,n}$ *and* $P_{0,0,0,n}$ *are contiguous.*

As a result of this lemma and Lemma 10 (in the Appendix), it makes no difference asymptotically whether we assume that the value of the mean, μ, or the coefficients of the lag polynomial, $\gamma(L)$, are known. This allows us to easily generalize the analogous optimality results from Section 3 to model (4.4). Because we have locally re-parameterized the parameters (μ, γ) the results are local in nature. In order to establish global results we may use discretized estimators of (μ, γ), where the discretization is based on the appropriate neighborhood as a function of n[14]. In the case of μ we may take advantage of the fact that,

$$y_1 = \mu + \varepsilon_1,$$

[14]For a more detailed description of the role of discretization see Le Cam and Yang (2000).

to choose $\hat{\mu} = y_1$. In the case of $\gamma(L)$ the appropriate re-parameterization was with respect to $n^{-1/2}$ and so the appropriate discretization is to utilize a $n^{1/2}$-consistent estimator as our initial estimator and then choose the point on the grid, $n^{-1/2}\mathbb{Z}^p$ closest to the value of this estimator; denote this estimator by $\hat{\gamma}(L)$. Thus we define $\hat{S}_n(Y) := (\hat{U}_n(Y), \hat{T}_n(Y))$ where

$$\hat{T}_n(y) = \frac{1}{\hat{\sigma}^2 n}\sum_{t=p+2}^{n}\hat{x}_{t-1}\Delta\hat{x}_t, \qquad \hat{U}_n(Y) = \frac{1}{\hat{\sigma}^2 n}\sum_{t=p+2}^{n}\hat{x}_{t-1}^2,$$

and $\hat{x}_t = \hat{\gamma}(L)(\hat{y}_t - \hat{\mu})$. Thus, we may proceed similarly as in Section 3. The counterpart to equation (3.3) is,

$$\lim_{n\to\infty} E_{c,m,\pi,n}\left[\left(1\left\{\tilde{c}_n(S_n(Y)) \le c\right\} - (1-\alpha)\right)h(Z_n)\right] = 0 \qquad \forall h \in \mathcal{C}_b(\mathbb{R}), \tag{4.3}$$

where Z_n is defined as in Section 3. It is important to note that, unlike in Section 3, we are no longer considering procedures which condition on the final observation of the observed series, $(Y_1, \ldots, Y_n)$. Instead we are conditioning on the demeaned series less its linear projection, $X_n = \gamma_0(L)(Y_n - \mu)$. However, this series satisfies $(1-\rho L)X_n = \varepsilon_n$ just as in the preceding sections. Thus, we may generalize Theorem 2.

Theorem 4. *Suppose Assumptions 2, 3, & 4 hold in model (4.1),* $0 < \alpha < 1/2$ *and let* $\{\tilde{c}_n(\cdot)\}$ *be a sequence of estimators which satisfy equation (4.3). Then,*

$$\begin{aligned}\liminf_{n\to\infty} E_{c,m,\pi,n}\left[\ell_c^*\left(\tilde{c}_n\left(\hat{S}_n(Y)\right)\right)\right] &\ge \lim_{n\to\infty} E_{c,m,\pi,n}\left[\ell_c^*\left(\hat{c}_{1-\alpha}\left(\hat{S}_n(Y)\right)\right)\right] \\ &= E_{c,m,\pi}\left[\ell_c^*\left(\hat{c}_{1-\alpha}(S^c)\right)\right],\end{aligned}$$

where $\ell_c^*(d) = \ell_c(d)1\{d < c\}$, $\ell_c(\cdot) \in \mathcal{L}$, *and* $\ell_c(\cdot)$ *is bounded and discontinuous on a set of probability zero.*

4.3. **Linear Time Trend.** A drawback to our procedure is that the optimality results of the paper do not extend to the case of the linear time trend. Let our observed data be generated by the model,

$$Y_t = \mu + \beta t + U_t, \qquad (1-\rho L)\gamma(L)U_t = \varepsilon_t, \qquad t = 1, \ldots, n, \tag{4.4}$$

with the localization,

$$\beta = \beta_n(b) = \beta_0 + \frac{\gamma_0(1)}{\sqrt{n}}b,$$

where, without loss of generality, we set $\beta_0 = 0$. Now, rather than the term $cT_n(y) - \frac{1}{2}c^2U_n(y)$ as in Lemma 3 we have,

$$cT_n(y) - \frac{1}{2}c^2U_n(y) + b\left[\frac{1}{\sigma^2\sqrt{n}}\sum_{t=p+2}^{n}\xi_c\left(\frac{t-1}{n}\right)\Delta x_t - c\frac{1}{\sigma^2 n^{3/2}}\sum_{t=p+2}^{n}\xi_c\left(\frac{t-1}{n}\right)x_{t-1}\right],$$

where $\xi_c := 1 - cr$. After conditioning on the term $T_n(y)$, the likelihood ratio will no longer be monotone with respect to the parameter c. This certainly does not preclude the construction of a conditional estimator, however it will no longer enjoy any optimality properties. This result is analogous to the case of point testing the null hypothesis of a unit root ($c = 0$), versus a one-sided alternative ($\bar{c} < 0$) (see Elliott *et al.* (1996)). In this case, the power envelope is unchanged when a mean or serial correlation in the error terms is added, but is lowered when a linear time trend is accommodated.

5. Monte Carlo Evidence

We explored the performance of our estimator relative to popular competitors in a Monte Carlo experiment. The data were generated from equation (2.1) using an initial condition of $x_0 = 0$ and *iid* standard normal errors. We considered two choices for the sample sizes, $n = 200$ and $n = 500$. The other procedures we considered were ordinary least-squares – denoted by "OLS", the grid-α estimator of Hansen (1999) to produce a median-unbiased estimator[15] – denoted by "HAN", and the inversion of the $P(c, 0, \bar{c})$ statistic as discussed in Elliott and Stock (2001) to produce a median-unbiased estimator – denoted by "ES", and $\hat{c}(u_n, t_n)$ – denoted by "HAT". It is important to emphasize that none of the alternative procedures considered are in the class of median-unbiased estimators conditional on the statistic $T_n(X)$ or its asymptotic counterpart. Instead this simulation study is meant to assess the unconditional performance of the HAT estimator relative to popular competitors. To compare performance, we calculated average mean-square error of the parameter estimation error, $\hat{\rho} - \rho$ and average mean-square prediction error, $(\hat{\rho} - \rho)x_n$ over $1,000$ simulations. The results may be found in Tables $1 - 4$ at the end of the paper.

In general, for values of ρ away from unity, the HAT estimator performs poorly. Recall that the conditional distribution function which is inverted to construct $\hat{c}$, relies on the local-to-unity weak limits of the sufficient statistics (U_n, T_n). Thus, for ρ far away from unity, when the local-to-unity parameterization is inappropriate, this choice is not suited for the data-generating process. This also occurs with the ES estimate, whose properties

[15] The grid$-t$ procedure performed similarly to the grid-α procedure and so we omit these results to conserve space. Also, as pointed out in footnote 3 of Hansen (1999) for *i.i.d.* Gaussian errors Hansen's grid-α proceure corresponds to Andrews (1993) procedure.

are based on the local-to-unity asymptotic theory, albeit to a lesser degree than in the case of the HAT estimator. Although not surprising, it is an unappealing property of these two estimators.

In Tables $1-2$ the average MSE for the parameter estimation error has been summarized for both sample sizes. In the case of $n = 200$, ES has the lowest mean-square error for values of ρ near 1, however as expected, as ρ moves away from 1, OLS dominates the other estimators. The HAT estimator performs relatively poorly for the smaller sample size, but improves appreciably for $n = 500$. For the larger sample size, ES performs the best for $\rho = 1$, while no estimator dominates in the range of $\rho \in \{0.99, 0.95, 0.90\}$. Again, as the theory would suggest, OLS performs well when the value of ρ moves away from 1.

The results for the average mean-square prediction error may be found in Tables $3-4$. The results follow a similar pattern as in the first two tables. ES performs the best for higher values of ρ and OLS performs best for lower values of ρ (although as n increases the range of dominance for ES is shortened). When we move to the larger sample size we see again that the HAT estimator's relative performance improves.

These preliminary simulation results suggest that the HAT estimator is on par with its competitors when the value of ρ is close to one, but does not perform well when ρ is away from unity.

6. Conclusion

This paper has proposed an alternative interval and point estimator for the local-to-unity parameter in a nearly-integrated first-order autoregression. The estimators have correct coverage (and are median-unbiased) both conditionally and unconditionally and are optimal in a specific class of conditional procedures. In particular, we have shown that this class of estimators is asymptotically equivalent, under local-to-unity asymptotic methods, to those procedures which condition on the final observation of the series. We have also generalized the results to consider higher-order autoregressive processes with a non-zero mean. A possible avenue for future research is whether me might generalize the results to a wider class of error distributions such as those distributions which may be represented as a scale mixture of normal random variables.

7. Appendix

7.1. Appendix A: Preliminary Lemmas.

Lemma 4. *Suppose Assumptions 2, 1 hold. Then,*

1. *the family, $\{f_n(\cdot|\, t_n; c) : c \in \mathcal{C}\}$ has strictly increasing likelihood ratios.*
2. *$w \mapsto F_n(w|\, t_n; c)$ is continuous and strictly increasing for every $c \in \mathcal{C}$, $t_n \in \mathbb{R}$.*
3. *$c \mapsto F_n(w|\, t_n; c)$ is continuous and strictly decreasing for every $w \in \mathcal{W}$, $t_n \in \mathbb{R}$, where $\mathcal{W} = \{\tilde{w} : 0 < F(\tilde{w}|\, t_n; c) < 1\}$.*
4. *$c \mapsto q_{n,1-\alpha}(t_n, c)$ is continuous and strictly increasing.*

Proof. (1) By Lemma 7,

$$f(w|\, t_n; c_2)/\, f(w|\, t_n; c_1) = (g_n(c_2, t_n)/\, g_n(c_1, t_n)) \exp\left\{-\, w\left(c_2^2 - c_1^2\right)/\, 2\right\},$$

which is strictly increasing in w for $c_1 < c_2$. (2) Now fix $c \in \mathcal{C}$. For any $w \in \mathbb{R}$, and sequence $w_m \to w$ with $w_m \in \mathbb{R}$, $\forall m \in \mathbb{N}$ we have,

$$\lim_{m\to\infty} F(w_m|\, t_n; c) = \lim_{m\to\infty} \int_{-\infty}^{w_m} f(\omega|\, t_n; c)\, d\omega = \lim_{m\to\infty} \int_{-\infty}^{\infty} 1_{(-\infty, w_m]}\{\omega\} \cdot f(\omega|\, t_n; c)\, d\omega,$$

which by the dominated convergence theorem yields,

$$\lim_{m\to\infty} F(w_m|\, t_n; c) = \int_{-\infty}^{\infty} \lim_{m\to\infty} 1_{(-\infty, w_m]}\{\omega\}\, f(\omega|\, t_n; c)\, d\omega = F(w|\, t_n; c).$$

Thus, $F(w|\, t_n; c)$ is a continuous function in w. $F(w|\, t_n; c)$ is strictly increasing by absolute continuity with respect to Lebesgue measure. (3) Fix $w \in \mathbb{R}$. By Lehmann and Romano (2005, Theorem 2.7.1), $F(w|\, t_n; c)$ is a continuous function in c. By Lehmann and Romano (2005, Corollary 3.2.1) $c \mapsto F(w|\, t_n; c)$ is strictly decreasing for any point $w \in \mathcal{W}$. (4) follows by (2) and (3). □

Lemma 5 (Lehmann and Romano (2005, Prob 3.44))**.** *Let $L(\theta, \underline{\theta})$ be nonnegative and nonincreasing in its second argument for $\underline{\theta} < \theta$, and equal to 0 for $\underline{\theta} \geq \theta$. If $\underline{\theta}$ and $\underline{\theta}^*$ are two lower confidence bounds for θ such that*

$$P_\theta\{\underline{\theta} \leq \theta'\} \leq P_\theta\{\underline{\theta}^* \leq \theta'\} \quad \text{for all} \quad \theta' \leq \theta, \tag{7.1}$$

then

$$E_\theta[L(\theta, \underline{\theta})] \leq E_\theta[L(\theta, \underline{\theta}^*)]. \tag{7.2}$$

Proof. We follow the hint in Lehmann and Romano (2005) and define two cumulative distribution functions,

$$G(u) = \frac{P_\theta\{\underline{\theta} \leq u\}}{P_\theta\{\underline{\theta}^* \leq \theta\}}, \qquad G^*(u) = \frac{P_\theta\{\underline{\theta}^* \leq u\}}{P_\theta\{\underline{\theta}^* \leq \theta\}}, \qquad u < \theta,$$

and $G(u) = G^*(u) = 1$ for $u \geq \theta$. Then $G(u) \leq G^*(u)$ by equation (7.1) for all u and

$$\begin{aligned} E_\theta\left[L\left(\theta, \underline{\theta}\right)\right] &= P_\theta\left\{\underline{\theta}^* \leq \theta\right\} \int L(\theta, u)\, dG(u) \\ &\leq P_\theta\left\{\underline{\theta}^* \leq \theta\right\} \int L(\theta, u)\, dG^*(u) \\ &= E_\theta\left[L\left(\theta, \underline{\theta}^*\right)\right]. \end{aligned}$$

The inequality follows by Lemma 6 below since $u \mapsto L(\theta, u)$ is nonincreasing. □

Remark 3. *By similar steps as in the proof of Lemma 5 we may show the analogous result for the uniformly most accurate upper bound. When both the upper and lower uniformly most accurate confidence bounds have confidence coefficient equal to* $1/2$ *then the bounds coincide; call this estimator* $\hat{\theta}$*. Clearly,* $\hat{\theta}$ *is median unbiased and satisfies,*

$$E_\theta[\ell_\theta(\hat{\theta})] \leq E_\theta\left[\ell_\theta\left(\theta^*\right)\right],$$

for all $\ell_\theta(\cdot) \in \mathcal{L}_\theta$ *(defined analogous to* $\mathcal{L}$*), and any other median-unbiased estimator* θ^**.*

Lemma 6 (Lehmann and Romano (2005, Prob 3.40))**.** G_0*,* G_1 *are two cumulative distribution functions on the real line, then* $G_1(x) \leq G_0(x)$ *for all* x *if and only if* $E_0\psi(X) \leq E_1\psi(X)$ *for any nondecreasing function* ψ*.*

Proof. Suppose that $G_1(x) \leq G_0(x)$ for all x. Then, by Lemma 3.4.1 in Lehmann and Romano (2005) there exists two nondecreasing functions g_0 and g_1, and a random variable V, such that $g_0(v) \leq g_1(v)$ for all v and the distributions of $g_0(V)$ and $g_1(V)$ are G_0 and G_1, respectively. Thus,

$$E_0[\psi(X)] = E_V[\psi(g_0(V))] \leq E_V[\psi(g_1(V))] = E_1[\psi(X)].$$

Now suppose that,

$$E_0[\psi(X)] \leq E_1[\psi(X)]$$

for any nondecreasing function $\psi(\cdot)$. Choose $\psi(z) = 1 - 1\{z \leq x\}$ and the result follows. □

7.2. **Appendix B: Proofs.**

Lemma 7. *Suppose Assumptions 1 & 2 hold.*

(1) *The joint distribution of* $S_n(X) := (T_n(X), U_n(X))$ *is a* $(2,1)$*-curved exponential family with density,*

$$f_n(u_n, t_n; c) = \exp\left\{ct_n - \frac{1}{2}c^2 u_n\right\} f_n(u_n, t_n; 0),$$

where $f_n(t_n, u_n; 0)$ *is the density of* $S_n(X)$ *when* $c = 0$*.*

(2) *The conditional distribution of $U_n(X)$ given $T_n(X) = t_n$ is an exponential family with density,*

$$f_n(u_n|\, t_n; c) = g_n(c, t_n)\exp\left\{-\frac{1}{2}c^2 u_n\right\} f_n(u_n|\, t_n; 0),$$

where $g_n(c, t_n)$ is chosen to satisfy $\int_{\mathbb{R}_+} f_n(u_n|\, t_n; c)\, du_n = 1$.

Proof of Lemma 7 . Lemma 7 follows by equation (2.2) and by Lehmann and Romano (2005, Lemma 2.7.2). □

Proof of Theorem 1. By Lemma 4, $F(w|\, t_n; c)$ is continuous in both w and c when the other is fixed. Thus, by Corollary 3.5.1 of Lehmann and Romano (2005) the solution to equation (2.5) is the uniformly most accurate confidence bound for $c \in \mathcal{C}$ conditional on $T_n(X) = t_n$ at the confidence level $1-\alpha$. Since the previous part of the proof was restricted to the partition $T_n(X) = t_n$, to complete of the proof note that $(w, t_n) \mapsto F(w|\, t_n; c)$ is jointly measurable so that $(u_n, t_n) \mapsto \hat{c}_{n,1-\alpha}(u_n, t_n)$ is jointly measurable. For the second part of the theorem, let $\tilde{c}_l(x)$ be any other lower confidence bound with confidence coefficient $1-\alpha$ conditional on $T_n(X) = t_n$. Next, define the two conditional distribution functions,

$$\hat{G}(r) = \frac{P_{c,n}\{\hat{c}_{1-\alpha}(S_n(X)) \le r|\, T_n(X)\}}{P_{c,n}\{\tilde{c}_l(X) \le c|\, T_n(X)\}}, \quad \tilde{G}(r) = \frac{P_{c,n}\{\tilde{c}_l(X) \le r|\, T_n(X)\}}{P_{c,n}\{\tilde{c}_l(X) \le c|\, T_n(X)\}},$$

for $r < c$ and $\hat{G}(r) = \tilde{G}(r) = 1$ for $r \ge c$. Thus, following the proof of Lemma 5, for $\ell_c(\cdot) \in \mathcal{L}$,

$$\begin{aligned} E_{c,n}[\ell_c^*(\hat{c}_{1-\alpha}(S_n(X)))|\, T_n(X)] &= P_{c,n}\{\tilde{c}_l(X) \le c|\, T_n(X)\}\int \ell_c^*(r)\, d\tilde{G}(r) \\ &\le P_{c,n}\{\tilde{c}_l(X) \le c|\, T_n(X)\}\int \ell_c^*(r)\, d\tilde{G}(r) \\ &= E_{c,n}[\ell_c^*(\tilde{c}_l(X))|\, T_n(X)], \end{aligned}$$

and the result follows. □

Proof of Lemma 1: Lemma 1 follows by weak convergence results in Phillips (1987). □

The proofs of Theorems 2 and 3 are greatly simplified by using the limits of experiments approach as introduced in Section 2. To use this approach we first must show that the sequence of probability measures $\{P_{c,n}\}$ is contiguous with respect to the sequence of probability measure $\{P_{0,n}\}$.

Lemma 8. *Suppose Assumptions 1 & 2 hold. Then, the sequence of probability measures $\{P_{c,n}\}$ is contiguous with respect to the sequence of probability measure $\{P_{0,n}\}$.*

Proof of Lemma 8. By Theorem 1 in Le Cam and Yang (2000, 36) it is sufficient to show that under P_n,

$$\log \frac{dP_{c,n}}{dP_n} \Rightarrow_0 \log \frac{dP_c}{dP},$$

where the weak limit is the log-likelihood ratio for some experiment. By equation (2.2) and Lemma 1 we have that,

$$\log \frac{dP_{c,n}}{dP_n} = cT_n(X) - \frac{1}{2}c^2 U_n(X) \Rightarrow_0 cT - \frac{1}{2}c^2 U. \tag{7.3}$$

The right-hand side of equation (7.3) is, by Theorem 7.15 of Lipster and Shiryaev (2001, 279-80), the likelihood ratio of $\{B_c(r) : 0 \le r \le 1\}$ as defined in Lemma 1. □

Remark 4. *If we consider the Laplace transform*[16], *then since* $B_c(0) = 0$, $\{B_c(r), 0 \le r \le 1\} =_d \{-B_c(r), 0 \le r \le 1\}$, *and we have for* $\varkappa > 0$,

$$E\left[\exp\{-\varkappa U_c\} \middle| B_c(1) = a\right] = E\left[\exp\{-\varkappa U_c\} \middle| B_c(1) = -a\right],$$

for some $a \in R$. *Thus,*

$$E\left[\exp\{-\varkappa U_c\} \middle| B_c(1) = a\right] = E\left[\exp\{-\varkappa U_c\} \middle| B_c(1)^2 = a^2\right],$$

by the law of iterated expectations.

Proof of Lemma 2. Lemma 8 establishes contiguity of the relevant sequences of probability measures and Lemma 1 yields $S_n(X) \Rightarrow_0 S$. Thus we may apply Le Cam's Third Lemma (Proposition 1 in Le Cam and Yang (2000, 40)) with the local-to-unity parameterization which yields,

$$f(u, t; c) = \exp\left\{ct - \frac{1}{2}c^2 u\right\} f(u, t; 0).$$

For part (2), we first apply Lemma 2.7.2 in Lehmann and Romano (2005, 48). Next we find explicit forms for $g(c)$ and $f(u|\, t; 0)$. By Lévy (1951) we have,

$$E_0\left[\exp\{-\varkappa U\} \middle| B(1) = \tau\right] = \left(\frac{\sqrt{2\varkappa}}{\sinh\left(\sqrt{2\varkappa}\right)}\right)^{1/2} \exp\left\{-\frac{\tau^2}{2}\left(\sqrt{2\varkappa}\coth\left(\sqrt{2\varkappa}\right) - 1\right)\right\},$$

where $\tau \in \mathbb{R}$ and recall that $(U, B(1))$ are $(U^c, B_c(1))$ with $c = 0$ (i.e., functionals of a standard Brownian motion). Recall that

$$\sinh(z) = \frac{1 - \exp\{-2z\}}{2\exp\{-z\}}, \qquad \coth(z) = 1 + \frac{2\exp\{-2z\}}{1 - \exp\{-2z\}},$$

[16] For nonnegative random variables we may consider the Laplace transform rather than the characteristic function without any loss of generality. See, for example, Kallenberg (2002).

and so,

$$\begin{aligned}&\left(\frac{z}{\sinh(z)}\right)^{1/2}\exp\left\{-\frac{a^2}{2}\left(z\coth(z)-1\right)\right\}\\ =\;&\left(\frac{2z\exp\{-z\}}{1-\exp\{-2z\}}\right)^{1/2}\exp\left\{\frac{a^2}{2}-\frac{a^2}{2}z\right\}\exp\left\{-a^2\frac{z\exp\{-2z\}}{1-\exp\{-2z\}}\right\},\end{aligned}$$

which by expanding the exponential function becomes,

$$\begin{aligned}&=\;\sqrt{2}\exp\left\{\frac{a^2}{2}\right\}\sum_{j=0}^{\infty}\frac{\left(-a^2\right)^j z^{j+1/2}}{j!}\frac{\exp\left\{-z\left(2j+\frac{1}{2}+\frac{a^2}{2}\right)\right\}}{\left(1-\exp\{-2z\}\right)^{j+1/2}}\\ &=\;\sqrt{2}\exp\left\{\frac{a^2}{2}\right\}\sum_{j=0}^{\infty}\frac{\left(-a^2\right)^j}{j!}\sum_{k=0}^{\infty}\frac{\left(j+\frac{1}{2}\right)_k}{k!}\left[z^{j+1/2}\exp\left\{-z\left(2(j+k)+\frac{1}{2}+\frac{a^2}{2}\right)\right\}\right].\end{aligned}$$

If we plug in $z=\sqrt{2\varkappa}$, we obtain,

$$=2^{3/4}\exp\left\{\frac{a^2}{2}\right\}\sum_{j=0}^{\infty}\frac{\left(-a^2\right)^j}{j!}2^{j/2}\sum_{k=0}^{\infty}\frac{\left(j+\frac{1}{2}\right)_k}{k!}\left[\varkappa^{j/2+1/4}\exp\left\{-\sqrt{\varkappa}\sqrt{2}\left(2(j+k)+\frac{1}{2}+\frac{a^2}{2}\right)\right\}\right].$$

By Erdélyi (1954, 246) we have that,

$$\operatorname{lap}_p^{-1}\left(p^{\nu-1/2}\exp\{-\sqrt{\alpha p}\}\right)(q)=2^{-\nu}\pi^{-1/2}q^{-\nu-1/2}\exp\left\{-\frac{\alpha}{8q}\right\}D_{2\nu}\left(\frac{\sqrt{\alpha}}{\sqrt{2q}}\right),$$

where $\operatorname{lap}_p^{-1}(\cdot)$ is the inverse Laplace transform. Then, by term-by-term inversion we have,

$$\begin{aligned}&\operatorname{lap}_\gamma^{-1}\quad\left(E_0\left[\exp\{-\varkappa U_0\}\middle|\,B(1)=\tau\right]\right)\\ =\;&2^{3/4}\exp\left\{\frac{\tau^2}{2}\right\}\sum_{j=0}^{\infty}\frac{\left(-\tau^2\right)^j}{j!}2^{j/2}\sum_{k=0}^{\infty}\frac{\left(j+\frac{1}{2}\right)_k}{k!}2^{-j/2-3/4}\pi^{-1/2}u^{-j/2-5/4}\times\\ &\exp\left\{-\frac{2\left(2(j+k)+\frac{1}{2}+\frac{\tau^2}{2}\right)^2}{8u}\right\}D_{j+3/2}\left(\frac{\left(2(j+k)+\frac{1}{2}+\frac{\tau^2}{2}\right)}{\sqrt{u}}\right)\\ =\;&\exp\left\{\frac{\tau^2}{2}\right\}\sum_{j=0}^{\infty}\frac{\left(-\tau^2\right)^j}{j!}\sum_{k=0}^{\infty}\frac{\left(j+\frac{1}{2}\right)_k}{\sqrt{\pi}k!}u^{-j/2-5/4}\times\\ &\exp\left\{-\frac{\left(4(j+k)+1+\tau^2\right)^2}{16u}\right\}D_{j+3/2}\left(\frac{\left(4(j+k)+1+\tau^2\right)}{2\sqrt{u}}\right).\end{aligned}$$

To find an expression for $g(c)$, note that for for $c<0$, by Borodin and Salminen (2002, 526), we have,

$$f(u|\,t;c)=\sigma_c\exp\left\{\frac{c}{2}+\frac{c\tau^2}{2}\right\}\exp\left\{-\frac{1}{2}\tau^2\left(1-\frac{1}{\sigma_c^2}\tau^2\right)\right\}\exp\left\{-\frac{1}{2}c^2u\right\}f(u|\,t;0),$$

and the result follows. □

Remark 5. *Using the marginal distribution of T we may also derive,*

$$\begin{aligned} f(u,t;0) &= \frac{\sqrt{2}}{\pi}\sum_{j=0}^{\infty}\frac{(-\tau^2)^j}{j!}\sum_{k=0}^{\infty}\frac{\left(j+\frac{1}{2}\right)_k}{k!}u^{-\frac{1}{2}\left(\frac{5}{2}+j\right)}\times \\ &\quad \exp\left\{-\frac{\delta(j,k;\tau)^2}{16u}\right\}D_{3/2+j}\left(\frac{\delta(j,k;\tau)}{2\sqrt{u}}\right). \end{aligned}$$

Proof of Theorem 2. We modify the proof of Theorem 5 in Jansson and Moreira (2006). The proof will proceed in two steps: first, we will define a certain class of estimators and show that $\hat{c}_{1-\alpha}$ is optimal within this class and also satisfies equation (3.3); second, we will show that any other sequence of estimators which satisfies equation (3.3) has a limiting representation that is a member of Ψ^c and thus asymptotically inferior.

First note that $(u,t)\mapsto\hat{c}_{1-\alpha}(u,t)$ is continuous in each argument when the other is fixed. This follows by Lemma 2 and the dominated convergence theorem for the first argument and by Lemma 2 and Remark 1 for the second argument. Therefore, we may apply the continuous-mapping theorem to the results of Lemma 1 to obtain,

$$\left(\hat{c}_{1-\alpha}(S_n(X)),\frac{X_n}{\sigma\sqrt{n}},cT_n(X)-\frac{1}{2}c^2U_n(X)\right)\Rightarrow_0(\hat{c}_{1-\alpha}(S),B(1),\Lambda(c))=:Q_0, \tag{7.4}$$

and

$$\Lambda(c):=cT-\frac{1}{2}c^2U.$$

Next, note that by Lemma 2 for a jointly measurable function $g(\cdot,\cdot)$ we have,

$$\int\int g(u^c,t^c)\,f(u^c,t^c)\,du^cdt^c=E\left[g(S)\exp\{\Lambda(c)\}\right].$$

Recall that $X_n/(\sigma\sqrt{n})\Rightarrow_c B_c(1)$ and $X_n/(\sigma\sqrt{n})\Rightarrow_0 B(1)$. Now, define the following two classes of estimators,

$$\Upsilon^c=\{\tilde{c}:E\left[(1\{\tilde{c}(S)\le c\}-(1-\alpha))\,h(B(1))\exp\{\Lambda(c)\}\right]=0,\ \forall h\in\mathcal{C}_b(\mathbb{R})\},$$

and

$$\Psi^c=\{\tilde{c}(\cdot):E\left[(1\{\tilde{c}(S)\le c\}-(1-\alpha))\,h(T)\exp\{\Lambda(c)\}\right]=0,\ \forall h\in\mathcal{C}_b(\mathbb{R})\}.$$

Since $\hat{c}_{1-\alpha}(\cdot)$ is a lower bound with conditional confidence coefficient $1-\alpha$, it is clear that $\hat{c}_{1-\alpha} \in \Psi^c$. Moreover, by similar arguments as in the proof of Theorem 1,

$$E\left[\ell_c^*\left(\hat{c}_{1-\alpha}(S)\right)\exp\left\{\Lambda(c)\right\}\right] \leq E\left[\ell_c^*\left(\tilde{c}(S)\right)\exp\left\{\Lambda(c)\right\}\right], \qquad \forall \tilde{c} \in \Psi^c. \tag{7.5}$$

By Lemma 9, Υ^c and Ψ^c are equivalent classes of estimators. Thus $\hat{c}_{1-\alpha} \in \Upsilon^c$ and equation (7.5) also holds $\forall \tilde{c} \in \Upsilon^c$. Next, we need to show that $\hat{c}_{1-\alpha}$ satisfies equation (3.3). By equation (7.4) and Le Cam's Third Lemma,

$$\left(\hat{c}_{1-\alpha}\left(S_n(X)\right), \frac{X_n}{\sigma\sqrt{n}}, cT_n(X) - \frac{1}{2}c^2U_n(X)\right) \Rightarrow_c Q_c,$$

where

$$\frac{dQ_c}{dQ_0} = \exp\left\{\Lambda(c)\right\} = \exp\left\{cT - \frac{1}{2}c^2U\right\},$$

is understood as a Radon-Nikodym derivative. Now, since ℓ_c^* is discontinuous on a set of probability zero,

$$\ell_c^*\left(\hat{c}_{1-\alpha}\left(S_n(X)\right)\right) \Rightarrow_0 \ell_c^*\left(\hat{c}_{1-\alpha}(S)\right).$$

Moreover, since ℓ_c^* is bounded then $\left\{\ell_c^*\left(\hat{c}_{1-\alpha}\left(S_n(X)\right)\right)\right\}$ are uniformly integrable. Then by Theorem 3.5 of Billingsley (1999) we have that,

$$\begin{aligned}
&\lim_{n\to\infty} E_{c,n}\left[\left(1\left\{\hat{c}_{1-\alpha}\left(S_n(X)\right) \leq c\right\} - (1-\alpha)\right)h\left(\frac{X_n}{\sigma\sqrt{n}}\right)\right] \\
&= \int \left(1\left\{\hat{c}_{1-\alpha}(S) \leq c\right\} - (1-\alpha)\right)h\left(B(1)\right)dQ_c \\
&= \int \left(1\left\{\hat{c}_{1-\alpha}(S) \leq c\right\} - (1-\alpha)\right)h\left(B(1)\right)\exp\left\{\Lambda(c)\right\}dQ_0 \\
&= E\left[\left(1\left\{\hat{c}_{1-\alpha}(S) \leq c\right\} - (1-\alpha)\right)h\left(B(1)\right)\exp\left\{\Lambda(c)\right\}\right] \\
&= 0,
\end{aligned}$$

since $\hat{c}_{1-\alpha} \in \Upsilon^c$. By similar arguments we have that

$$\lim_{n\to\infty} E_c\left[\ell_c^*\left(\hat{c}_{1-\alpha}\left(S_n(X)\right)\right)\right] = E\left[\ell_c^*\left(\hat{c}_{1-\alpha}(S)\right)\exp\left\{\Lambda(c)\right\}\right].$$

To complete the proof we must show that for any sequence of estimators $\{\tilde{c}_n\}$ which satisfies equation (3.3) and is uniformly tight, that there exists a $\tilde{c} \in \Upsilon^c$ such that,

$$\liminf_{n\to\infty} E_{c,n}\left[\ell_c^*\left(\tilde{c}_n\left(S_n(X)\right)\right)\right] = E_{c,n}\left[\ell_c^*\left(\tilde{c}(S)\right)\exp\left\{\Lambda(c)\right\}\right].$$

First, by properties of the limit inferior we may choose a subsequence, $i(n)$ such that,

$$\lim_{i(n)\to\infty} E_{c,i}\left[\ell_c^*\left(\tilde{c}_i\left(S_i(X)\right)\right)\right] = \liminf_{n\to\infty} E_{c,n}\left[\ell_c^*\left(\tilde{c}_n\left(S_n(X)\right)\right)\right]. \tag{7.6}$$

Now since, $\tilde{c}_i(S_i(X))$ is uniformly tight and $S_i(X) = O_{P_{0,n}}(1)$ we have joint uniform tightness and so by Prohorov's theorem there exists a further subsequence $j = j(i)$ such that along this sequence,

$$(\tilde{c}_j(S_j(X)), S_j(X)) \Rightarrow_0 (\tilde{c}_\infty, S),$$

where $\tilde{c}_\infty$ is a random variable defined on the same probability space as S. Now, by the continuous-mapping theorem,

$$(\ell_c^*(\tilde{c}_j(S_j(X))), S_j(X), cT_j - \frac{1}{2}c^2U_j) \Rightarrow_0 (\ell_c^*(\tilde{c}_\infty), S, \Lambda(c)),$$

we may apply Le Cam's Third Lemma and Theorem 3.5 of Billingsley (1999), and by equation (7.6),

$$\liminf_{n\to\infty} E_{c,n}[\ell_c^*(\tilde{c}_n(S_n(X)))] = \lim_{j(n)\to\infty} E_{c,j}[\ell_c^*(\tilde{c}_j(S_j(X)))] = E[\ell_c^*(\tilde{c}_\infty)\exp\{\Lambda(c)\}].$$

Finally, we need to show that $\tilde{c}_\infty \in \Upsilon^c$. This follows by,

$$\begin{aligned} & E[(1\{\tilde{c}_\infty \le c\} - (1-\alpha))h(B(1))\exp(\Lambda(c))] \\ = & \lim_{j(n)\to\infty} E_{c,j}[(1\{\tilde{c}_j(S_j(X)) \le c\} - (1-\alpha))h(Z_n)] \\ = & \ 0, \end{aligned}$$

since $\{\tilde{c}_n\}$ satisfies equation (3.3). □

Lemma 9. *The two classes of estimators, Υ^c and Ψ^c, as defined in the proof of Theorem 2, are equivalent.*

Proof of Lemma 9. Υ^c is defined by estimators which satisfy,

$$E[(1\{\tilde{c}(S) \le c\} - (1-\alpha))h(B(1))\exp\{\Lambda(c)\}] = 0, \quad \forall h \in \mathcal{C}_b(\mathbb{R}).$$

Now,

$$\begin{aligned} & E[(1\{\tilde{c}(S) \le c\} - (1-\alpha))h(B(1))\exp\{\Lambda(c)\}] = 0, \quad \forall h \in \mathcal{C}_b(\mathbb{R}) \\ \iff & E_c[(1\{\tilde{c}(S^c) \le c\} - (1-\alpha))h(B_c(1))] = 0, \quad \forall h \in \mathcal{C}_b(\mathbb{R}) \\ \iff & E_c[(1\{\tilde{c}(S^c) \le c\} - (1-\alpha)) | B_c(1)] = 0. \end{aligned}$$

Recall that by Itô's Lemma,

$$T^c = \frac{1}{2}\left(B_c(1)^2 - 1\right)$$

and that the conditional distribution of U^c given $B_c(1)$ is identical to the conditional distribution of U^c given $B_c(1)^2$ (see Remark 4 and Lemma 2). Thus,

$$\begin{aligned}
& E_c\left[\left(1\left\{\tilde{c}\left(S^c\right) \leq c\right\}-(1-\alpha)\right) \mid B_c(1)\right]=0 \\
\Longleftrightarrow\quad & E_c\left[\left(1\left\{\tilde{c}\left(U^c, \frac{1}{2}\left(B_c(1)^2-1\right)\right) \leq c\right\}-(1-\alpha)\right) \middle| B_c(1)^2\right]=0 \\
\Longleftrightarrow\quad & E_c\left[\left(1\left\{\tilde{c}\left(U^c, T^c\right) \leq c\right\}-(1-\alpha)\right) \mid T^c\right]=0 \\
\Longleftrightarrow\quad & E_c\left[\left(1\left\{\tilde{c}\left(S^c\right) \leq c\right\}-(1-\alpha)\right) h\left(T^c\right)\right]=0, \quad \forall h \in \mathcal{C}_b(\mathbb{R}) \\
\Longleftrightarrow\quad & E\left[\left(1\{\tilde{c}(S) \leq c\}-(1-\alpha)\right) h(T) \exp \{\Lambda(c)\}\right]=0, \quad \forall h \in \mathcal{C}_b(\mathbb{R}),
\end{aligned}$$

and the result follows. □

Proof of Theorem 3. By consistency of $\hat{\sigma}^2$ we have that,

$$\hat{S}_n\left(X ; \hat{\sigma}^2\right)=S_n(X)+o_p(1).$$

Then, as in the proof of Theorem 2 we note that $(u, t) \mapsto \hat{c}(u, t)$ is continuous and so by the continuous-mapping theorem (since $\ell_c^*(\cdot)$ is discontinuous on a set of probability zero),

$$\ell_c^*(\hat{c}_{1-\alpha}(\hat{S}_n(X ; \hat{\sigma}^2))) \Rightarrow_0 \ell_c^*\left(\hat{c}_{1-\alpha}(S)\right).$$

Then by an application of Le Cam's Third Lemma (Proposition 1 in Le Cam and Yang (2000, 40)) and Theorem 3.5 in Billingsley (1999, 31) we have,

$$\lim_{n \rightarrow \infty} E_c[\ell_c^*(\hat{c}_{1-\alpha}(\hat{S}_n(X ; \hat{\sigma}^2)))]=E\left[\ell_c^*\left(\hat{c}_{1-\alpha}(S)\right) \exp \{\Lambda(c)\}\right]=E_c\left[\ell_c^*\left(\hat{c}_{1-\alpha}\left(S^c\right)\right)\right].$$

□

Proof of Lemma 3. For (1) let us partition the vector $\varepsilon=\left(\left(\varepsilon_1, \ldots, \varepsilon_{p+1}\right),\left(\varepsilon_{p+2}, \ldots, \varepsilon_n\right)\right)$. First note that for $t=p+2, \ldots, n$ we have,

$$\varepsilon_t=(1-\rho L) \gamma(L)\left(y_t-\mu\right).$$

Then,

$$\begin{aligned}
\varepsilon_t &= (1-\rho L) \gamma(L)\left(y_t-\mu\right) \\
&= (1-L+L-\rho L) \gamma(L)\left(y_t-\mu\right) \\
&= \left(1-L-c n^{-1} L\right)\left[\gamma_0(L)+n^{-1 / 2} \pi(L)\right]\left(y_t-\mu\right) \\
&= \Delta x_t-c n^{-1}\left(x_{t-1}-\mu\right)+n^{-1 / 2} \pi(L) \Delta y_t-c n^{-3 / 2} \pi(L)\left(y_{t-1}-\mu\right) \\
&= \Delta x_t-c n^{-1} x_{t-1}+n^{-1 / 2} \pi(L) \Delta y_t+O_{P_{0,0,0}}\left(n^{-1}\right).
\end{aligned}$$

Thus,

$$\begin{aligned}\sum_{t=p+2}^{n}\varepsilon_t^2 &= \sum_{t=p+2}^{n}\left[\Delta x_t - cn^{-1}x_{t-1} + n^{-1/2}\pi(L)\Delta y_t + O_{P_{0,0,0}}\left(n^{-1}\right)\right]^2 \\ &= \sum_{t=p+2}^{n}\left[\Delta x_t\right]^2 + \sum_{t=p+2}^{n}\left(\left[cn^{-1}x_{t-1}\right]^2 + \left[n^{-1/2}\pi(L)\Delta y_t\right]^2\right) \\ &\quad -2\sum_{t=p+2}^{n}\left(cn^{-1}x_{t-1}\Delta x_t - n^{-1/2}\pi(L)\Delta x_t\Delta y_t\right) + o_{P_{0,0,0}}(1).\end{aligned}$$

Next, note that

$$(1-\rho z)\gamma(z) = \sum_{k=0}^{p+1}a_{1,k}z^k - \sum_{k=0}^{p+1}a_{2,k}z^k + o(1),$$

where

$$\begin{aligned}a_{1,0} &= 1,\ a_{2,0}=0, \qquad a_{1,1}=-\gamma_{01},\ a_{2,1}=1 \\ a_{1,k} &= -\gamma_{0k},\ a_{2,k}=-\gamma_{0(k-1)} \qquad k=2,\ldots,p \\ a_{1,p+1} &= \gamma_{0p},\ a_{2,p+1}=0.\end{aligned}$$

Thus, for $t=1,\ldots,p+1$ we have that

$$\varepsilon_1 = y_1 - m, \qquad \varepsilon_t = \Delta x_t + \gamma_{0(t-1)}m, \qquad t=2,\ldots,p+1.$$

Thus,

$$\begin{aligned}&L(c,m,\pi) - L(0,0,0) \\ =\ & -\frac{1}{2\sigma^2}\sum_{t=p+2}^{n}\left(\left(cn^{-1}x_{t-1}\right)^2 + \left(n^{-1/2}\pi(L)\Delta y_t\right)^2 - 2cn^{-1}x_{t-1}\Delta x_t + 2n^{-1/2}\pi(L)\Delta x_t\Delta y_t\right) \\ & -\frac{1}{2\sigma^2}\left[(y_1-m)^2 + \sum_{t=2}^{p+1}\left(\Delta x_t + \gamma_{0(t-1)}m\right)^2\right] + \frac{1}{2\sigma^2}\left[(y_1)^2 + \sum_{t=2}^{p+1}(\Delta x_t)^2\right] + o_{P_{0,0,0}}(1),\end{aligned}$$

and (1) follows. Now consider (2). Convergence and independence of $(U_n(Y), T_n(Y))$ and $(W_{1n}(Y), W_{2n}(Y))$ follow from results in Jeganathan (1991). For $(V_{1n}(y), V_{2n}(y))$, note that under $(c,m,\pi)=(0,0,0)$, $\Delta x_k = \varepsilon_k$ for $k=1,\ldots,p+1$. Pairwise independence follows by Assumption 2. (3) follows by results in Jeganathan (1991). □

Lemma 10. *Suppose Assumptions 2, 3, & 4 hold in model (4.1). Then, the joint distribution of* $S^{c,m,\pi}$ *is,*

$$\begin{aligned} & f(u,t,v,w_1;c,\pi,m,V_2,W_2) \\ = \; & \exp\left\{ct - \frac{1}{2}c^2u\right\} f(u,t;0) \exp\left\{-mv - \frac{1}{2}m^2V_2\right\} f(v;0,V_2) \times \\ & \exp\left\{\pi' w_1 - \frac{1}{2}\pi' w_2 \pi\right\} f(w_1;0,W_2), \end{aligned}$$

and

$$\begin{aligned} & f(u,v,w_1|\,t;c,\pi,m,V_2,W_2) \\ = \; & g(c,t)\exp\left\{-\frac{1}{2}c^2u\right\} f(u|\,t;0) \exp\left\{-mv - \frac{1}{2}m^2V_2\right\} f(v;0,V_2) \times \\ & \exp\left\{\pi' w_1 - \frac{1}{2}\pi' W_2 \pi\right\} f(w_1;0,W_2), \end{aligned}$$

where

$$\begin{aligned} f(v;0,V_2) &= (2\pi V_2)^{-1/2} \exp\left\{-(2V_2)^{-1} v^2\right\}, \\ f(w_1;0,W_2) &= (2\pi)^{-p/2} |W_2|^{-1/2} \exp\left\{-\frac{1}{2}W_1' W_2^{-1} W_1\right\}, \end{aligned}$$

and $g(c,t)$ *and* $f(u|\,t;0)$ *are as in Lemma 2.*

Proof of Lemma 10. By Lemma 3 (1) we have that $L(c,m,\pi) - L(0,0,0) \Rightarrow_{(0,0,0)} \Lambda(c,m,\pi)$. Next, by Lemma 3 (3) we have that the sequences of probability measures are contiguous and so we may appeal to Le Cam's Third Lemma (Proposition 1 in Le Cam and Yang (2000, 40)). An application of Lemma 2.7.2 in Lehmann and Romano (2005, 48) yields,

$$\begin{aligned} & f(u,t,v,w_1;c,\pi,m,V_2,W_2) \\ = \; & \exp\left\{ct - \frac{1}{2}c^2u - mv - \frac{1}{2}m^2V_2 + \pi' w_1 - \frac{1}{2}\pi' W_2 \pi\right\} f(u,t,v,w_1;0,0,0,V_2,W_2). \end{aligned}$$

By the independence results in 3 (2) we may rewrite,

$$f(u,t,v,w_1;0,0,0,V_2,W_2) = f(u,t;0)\, f(v;0,V_2)\, f(w_1;0,W_2),$$

where $f(u,t;0)$, $f(v;0,V_2)$, $f(w_1;0,W_2)$ are $f(u,t;c)$, $f(v;m,V_2)$, $f(w_1;\pi,W_2)$ under the parameter values $c=0$, $m=0$, and $\pi=0$, respectively. Another application of

Lemma 2.7.2 in Lehmann and Romano (2005, 48) yields,

$$\begin{aligned} & f\left(u, v, w_1 | \, t; c, \pi, m, V_2, W_2\right) \\ = \; & C \exp\left\{ct - \frac{1}{2}c^2 u\right\} f\left(u | \, t; 0\right) \exp\left\{-mv - \frac{1}{2}m^2 V_2\right\} f\left(v; 0, V_2\right) \times \\ & \quad \exp\left\{\pi' w_1 - \frac{1}{2}\pi' W_2 \pi\right\} f\left(w_1; 0, W_2\right), \end{aligned}$$

where $C = C\left(c, \pi, m, t\right)$ is a normalizing function. However, by Lemma 2 we know that the appropriate normalizing function for $f\left(u | \, t; 0\right)$ is $g\left(c, t\right)$. Moreover, by standard properties of the (log) normal distribution we have that,

$$\int_{-\infty}^{\infty} \exp\left\{-mv\right\} f\left(v; 0, V_2\right) dv = \exp\left\{\frac{1}{2}m^2 V_2\right\}$$

and

$$\int_{-\infty}^{\infty} \exp\left\{\pi' w_1\right\} f\left(w_1; 0, W_2\right) dw_1 = \exp\left\{\frac{1}{2}\pi' W_2 \pi\right\}.$$

Thus, $C = g\left(c, t\right)$. □

Proof of Theorem 4. This proof follows by similar steps as in the proofs of Theorem 2 and 3. □

8. Simulation Results

ρ	OLS	HAT	GRA	ES
1	0.3006	0.6807	0.3219	0.2170
0.99	0.3768	0.4901	0.4391	0.3258
0.95	0.7655	0.9233	0.8975	0.7622
0.9	1.2260	1.5109	1.3776	1.2630
0.75	2.4324	7.0158	2.5891	3.5756
0.5	3.8169	70.4593	3.9257	20.3303

Table 1. $10^3\times$ Avg MSE Parameter Estimation Error, $n = 200$

ρ	OLS	HAT	GRA	ES
1	0.0489	0.1013	0.0534	0.0351
0.99	0.1018	0.1064	0.1190	0.0935
0.95	0.2842	0.2740	0.3206	0.2914
0.9	0.4780	0.5260	0.5199	0.5415
0.75	0.9587	4.6473	1.0003	2.3671
0.5	1.5401	64.0822	1.5722	18.3329

Table 2. $10^3\times$ Avg MSE Parameter Estimation Error, $n = 500$

ρ	OLS	HAT	GRA	ES
1	0.9642	1.5131	1.0288	0.3082
0.99	0.6691	1.4807	0.7603	0.4819
0.95	0.5175	0.7231	0.5999	0.5082
0.9	0.5199	0.7002	0.5727	0.6637
0.75	0.4968	1.5317	0.5102	1.6938
0.5	0.4538	9.0072	0.4543	5.5514

Table 3. $10^2\times$ Avg MSE Prediction Error, $n = 200$

ρ	OLS	HAT	GRA	ES
1	0.3705	0.4430	0.3964	0.1073
0.99	0.2512	0.4528	0.2830	0.2266
0.95	0.2376	0.2829	0.2619	0.3389
0.9	0.2473	0.2815	0.2631	0.5006
0.75	0.2225	0.9652	0.2273	1.4589
0.5	0.1967	7.9334	0.2001	4.3138

Table 4. $10^2\times$ Avg MSE Prediction Error, $n = 500$

References

[1] Andrews, D. A. (1993): "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models," *Econometrica*, 61(1), 139-165.

[2] Andrews, D.A., & H. Y. Chen (1994): "Approximately Median-Unbiased Estimation of Autoregressive Models," *Journal of Business & Economic Statistics*, 12(2), 187-204.

[3] Andrews, D.A., & P.C.B. Phillips (1987): "Best Median-Unbiased Estimation in Linear Regression with Bounded Asymmetric Loss Functions," *Journal of the American Statistical Association*, 82(399), 886-893.

[4] Barndorff-Nielsen, O. E., & D. R. Cox (1994): *Inference and Asymptotics*. London: Chapman & Hall.

[5] Barndorff-Nielsen, O. E., & D. R. Cox (1996): "Prediction and Asymptotics," *Bernoulli*, 4(2), 319-340.

[6] Billingsley, P. (1999): *Convergence of Probability Measures* (Second Ed.). New York: Wiley.

[7] Bobkoski, M. J. (1983): "Hypothesis Testing in Nonstationary Time Series," Unpublished P.h.D. Thesis, Department of Statistics, University of Wisconsin.

[8] Borodin, A.N., & P. Salminen (1996): *Handbook of Brownian Motion*. Basel: Birkhaüser-Verlag.

[9] Breidt, J.F., Davis, R.A., & W.T.M. Dunsmuir (1995): "Improved Bootstrap Prediction Intervals for Autoregressions," *Journal of Time Series Analysis*, 95(2), 177-200.

[10] Cavanagh, C. (1985): "Roots Local to Unity," Manuscript, Department of Economics, Harvard University.

[11] Chan, N. H. (1988): "The Parameter Inference for Nearly Nonstationary Time Series," *Journal of the American Statistical Association*, 83(403), 857-862.

[12] Chan, N. H., & C.Z. Wei (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *The Annals of Statistics*, 15(3), 1050-1063.

[13] Dufour, J. (1990): "Exact Tests and Confidence sets in Linear Regressions with Autocorrelated Errors," *Econometrica*, 58(2), 475-494.

[14] Efron, B. (1975): "Defining the Curvature of a Statistical Problem (with Applications to Second Order Efficiency)," *The Annals of Statistics*, 3(6), 1189-1242.

[15] Efron, B. (1978): "The Geometry of Exponential Families," *The Annals of Statistics*, 6(2), 362-376.

[16] Elliott, G. (2006): "Unit Root Pre-Testing and Forecasting," Working Paper, Department of Economics, UC-San Diego.

[17] Elliott, G., Rothenberg, T. J. & J. H. Stock (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64(4), 813-836.

[18] Elliott, G., & J. H. Stock (2001): "Confidence Intervals for Autoregressive Coefficients Near One," *Journal of Econometrics*, 103(1-2), 155-181.

[19] Erdélyi, A. (1954): *Tables of Integral Transforms I.* New York: McGraw-Hill.

[20] Gospodinov, N. (2002): "Median Unbiased Forecasts for Highly Persistent Autoregressive Processes," *Journal of Econometrics*, 111(1), 85-101.

[21] Gushchin, A. A., (1995): "On Asymptotic Optimality of Estimators Under the LAQ Condition," *Theory of Probability and its Applications*, 40(2), 261-272.

[22] Haldrup, N., & M. Jansson (2006): "Improving Size and Power in Unit Root Testing," in *Palgrave Handbook of Econometrics, Volume 1: Econometric Theory*, ed. by Mills, T. C., & K. Patterson. New York: Palgrave Macmillan, 252-277.

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

[24] Jansson, M. (2008): "Semiparametric Power Envelopes for Tests of the Unit Root Hypothesis," *Econometrica*, 76(5), 1103-1142.

[25] Jansson M., & M.J. Moreira (2006): "Optimal Inference in Regression Models with Nearly Integrated Regressors," *Econometrica*, 74(3), 681-714.

[26] Jeganathan, P. (1991): "On the Asymptotic Behavior of Least-Squares Estimators in AR Time Series with Roots near the Unit Circle," *Econometric Theory*, 7(3), 269-306.

[27] Jeganathan, P. (1995): "Some Aspects of Asymptotic Theory with Applications to Time Series Models," *Econometric Theory*, 11(5), 818-887.

[28] Kabaila, P. (1993): "On Bootstrap Predictive Inference for Autoregressive Processes," *Journal of Time Series Analysis*, 93(5), 473-484.

[29] Kabaila, P. (1999): "The Relevance Property for Prediction Intervals," *Journal of Time Series Analysis*, 20(6), 655-662.

[30] Kallenberg, O. (2002): "Foundations of Modern Probability," (Second Ed.). New York: Springer-Verlag.

[31] Kiviet, J. F. & G. D. A. Phillips (1992): "Exact Similar Tests for Unit Roots and Cointegration," *Oxford Bulletin of Economics and Statistics*, 54(3), 349-367.

[32] Le Cam, L. (1982): *Asymptotic Methods in Statistical Decision Theory.* New York: Springer-Verlag.

[33] Le Cam, L., & G. L. Yang (2000): *Asymptotics in Statistics: Some Basic Concepts* (Second Ed.). New York: Springer-Verlag.

[34] Lehmann, E. L., & J. P. Romano (2005): *Testing Statistical Hypotheses* (Third Ed.). New York: Springer-Verlag.

[35] Lévy, P. (1951): "Wiener's Random Functions and other Laplacian Random Functions," in *Proc. Second Berkeley Symp. on Math. Statist. and Prob.* Berkeley: Univ. of Calif. Press, 171-187.

[36] Lipster, R. S., & A. N. Shiryaev (2001): *Statistics of Random Processes: I. General Theory* (Second Ed.). New York: Springer-Verlag.

[37] Nelson, C. R., & C. I. Plosser (1982): "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10(2), 139-162.

[38] Patterson, K. (2000): "Finite Sample Biase of the Least Squares Estimator in an AR(p) Model: Estimation, Inference, Simulation and Examples," *Applied Economics*, 32(15), 1993-2005.

[39] Pfanzagl, J. (1970): "Median Unbiased Estimates for M.L.R.-Families," *Metrika*, 15, 30-39.

[40] Pfanzagl, J. (1979): "On Optimal Median Unbiased Estimators in the Presence of Nuisance Parameters," *The Annals of Statistics*, 7(1), 187-193.

[41] Phillips, P.C.B. (1979): "The Sampling Distribution of Forecasts from a First-Order Autoregression," *Journal of Econometrics*, 9(3), 241-261.

[42] Phillips, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74(3), 535–547.

[43] Phillips, P. C. B., & Z. Xiao (1998): "A Primer on Unit Root Testing," *Journal of Economic Surveys*, 12(5), 423-469.

[44] Rudebusch, G.D. (1992): "Trends and Random Walks in Macroeconomic Time Series: A Re-Examination," *International Economic Review*, 33(3), 661-680.

[45] Schotman, P. C., & H. K. van Dijk (1991): "On Bayesian Routes to Unit Roots," *Journal of Applied Econometrics*, 6(4) 387-401.

[46] Stine, R.A. (1987): "Estimating Properties of Autoregressive Forecasts," *Journal of the American Statistical Association*, 82(400), 1072–1078.

[47] Stock, J. H. (1991): "Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series," *Journal of Monetary Economics*, 28(3), 435-459.

[48] Stock, J. H. (1994): "Unit Roots, Structural Breaks and Trends," in *Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. L. McFadden. New York: North-Holland, 2739-2841.

[49] Vidoni, P. (2004): "Improved Prediction Intervals for Stochastic Process Models," *Journal of Time Series Analysis*, 25(1), 137-154.

Department of Economics, University of California at Berkeley
E-mail address: crump@econ.berkeley.edu
URL: http://www.ocf.berkeley.edu/~crump