Folklore Theorems, Implicit Maps and New Unit Root Limit Theory

By

Peter C. B. Phillips

January 2011

COWLES FOUNDATION DISCUSSION PAPER NO. 1781

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Folklore Theorems, Implicit Maps and New Unit Root Limit Theory*

Peter C. B. Phillips

Yale University, University of Auckland
University of Southampton & Singapore Management University

December 20, 2010

Abstract

The delta method and continuous mapping theorem are among the most extensively used tools in asymptotic derivations in econometrics. Extensions of these methods are provided for sequences of functions, which are commonly encountered in applications, and where the usual methods sometimes fail. Important examples of failure arise in the use of simulation based estimation methods such as indirect inference. The paper explores the application of these methods to the indirect inference estimator (IIE) in first order autoregressive estimation. The IIE uses a binding function that is sample size dependent. Its limit theory relies on a sequence-based delta method in the stationary case and a sequence-based implicit continuous mapping theorem in unit root and local to unity cases. The new limit theory shows that the IIE achieves much more than bias correction. It changes the limit theory of the maximum likelihood estimator (MLE) when the autoregressive coefficient is in the locality of unity, reducing the bias and the variance of the MLE without affecting the limit theory of the MLE in the stationary case. Thus, in spite of the fact that the IIE is a continuously differentiable function of the MLE, the limit distribution of the IIE is not simply a scale multiple of the MLE but depends implicitly on the full binding function mapping. The unit root case therefore represents an important example of the failure of the delta method and shows the need for an implicit mapping extension of the continuous mapping theorem.

Keywords: Binding function, Delta method, Exact bias, Implicit continuous maps, Indirect inference, Maximum likelihood.

JEL Classification: C23

*Some of the limit results in this paper were presented in the author's Durbin lecture at UCL in May 2009, at the Singapore SESG meetings in August 2009, and at the Nottingham Granger Memorial Conference in May 2010. The author acknowledges partial research support from the Kelly Fund and the NSF under Grant Nos. SES 06-47086 and SES 09-56687.

1 Introduction

One of the folklore theorems of statistics is the delta method, a rigorous treatment first appearing in Cramér's (1946) treatise although the history of the method is certainly much more distant. The method appeared in early econometric texts (e.g., Klein, 1953) and its use in asymptotic derivations in econometrics is now almost universal. Equally important in econometric asymptotics, especially since the uptake of function space limit theory in the 1980s, is the continuous mapping theorem whose history also stretches into antiquity, an early source being the Mann and Wald (1943) article on stochastic order notation.

Whilst these methods appear almost everywhere in econometrics, there are some cases where the methods do not apply directly. Particularly important examples arise when a problem involves sample functions that depend on the sample size or when the quantity of interest appears in an implicit functional form. In some cases the methods fail but with some modification can be made to work. In other cases, a new theorem is required to obtain the limit theory. There appears to be no systematic discussion of these issues in the literature, although there is some discussion in the statistical literature of extensions to the continuous mapping theorem for sequences of functions.

The primary goal of the present work was to find the limit distribution of the indirect inference estimator in a simple first order autoregression. This estimator is effective in bias correction, which can be a major problem in autoregression, so the method is of considerable interest in that context. There are also manifestations of this problem and indirect inference alternatives in continuous time finance in diffusion equation estimation. In that context, Phillips and Yu (2009) use indirect inference to price contingent claims in derivative markets and show that this method removes bias and often reduces the variance of option price estimates that are based on maximum likelihood.

Investigation of the autoregressive model implementation of indirect inference reveals that the usual delta method gives the correct solutions in stationary and explosive cases but that the method fails in unit root and near unit root cases. Since the latter cases are most important in practical work, the failure has major implications. The explanation for the failure lies partly in the sample size dependence of the functional that defines the indirect inference estimator, partly in the implicit functional form that the estimator takes, and partly in the breakdown of linear approximation. All these issues need to be confronted in order to obtain the correct limit theory.

The problem is of wider significance because of the growing use of simulation based methods in the construction of extremum estimators in econometrics. Indirect inference (developed in Smith, 1993, and Gourieroux, Monfort and Renault, 1993) is a primary example of such a method. Other examples where sample based functionals arise in econometrics are median unbiased estimation (Andrews, 1992), simulated method of moments (McFadden, 1989), and simulated scores (Hajivassiliou and McFadden 1998), among a growing number of other methods. Particularly when a new procedure depends on the sampling distribution of another estimator, as in the case of median unbiased estimation, the new limit distribution may be fundamentally affected by the properties of the implied distributional transformation, much as its finite sample distribution is affected. It is, in effect, only when the transformation is locally linear in a suitably sized shrinking neighborhood that the limit distribution follows straightforwardly from usual rules such as the delta method.

The present paper introduces these issues, provides some discussion and limit results that extend the delta method and continuous mapping theorem, and applies the ideas in the context of indirect inference (II) limit theory for the first order autoregression. It is shown that the II estimator has a new form of limit distribution when the autoregressive coefficient is in the locality of unity. In effect, II not only removes the bias in the maximum likelihood estimator but also changes its limiting distributional shape in way that reduces variance. This is an instance where the delta method completely fails in the region of unity but a suitably extended version of the delta method applies in the stationary case.

The paper is organized as follows. Section 2 provides some new mapping theorems that extend the usual delta method to sequences of functions and the continuous mapping theorem to sequences of implicit mappings. Both results are useful in considering simulation based estimation procedures where sample based functionals appear in extremum estimation problems. Section 3 describes the indirect inference approach and Section 4 analyzes the use of this method in a first order autoregression, derives the analytic form of the binding function and develops comprehensive asymptotic expansion formulae for stationary, near unit root and explosive cases. Section 5 derives the limit distribution of the indirect inference estimator, applying an extended delta method in the stationary case and an implicit continuous mapping theorem in the unit root and local to unity case, showing that for these parameters the limit theory is a nonlinear functional of the standard unit root and near unit root asymptotics. Section 6 concludes and discusses various extensions. Some new integral asymptotic expansions are given in the Appendix, together with proofs of all the main results in the paper.

2 Mapping theorems and Examples

2.1 Extending the delta method to sequences of functions

While the ideas underlying the delta method have a long history, it seems that the original rigorous development was presented by Cramér (1946). Cramér's discussion included moments (p. 353), the limit distribution (p. 366), the multivariate case (p. 358), and more notably because it is seldom referenced the case where the leading term fails because of a zero first derivative and the variance is of smaller order than $O\left(n^{-1}\right)$, leading to possibly nonnormal limit theory and a higher rate of convergence. In the latter case, Cramér (p. 415) provides an illustration based on the distribution of the multiple correlation coefficient in the null correlation case, where the limit distribution is chi squared. Some related failures of standard methods of expansion and linearization were considered by Sargan (1983). Simple examples of such cases are sometimes mentioned in texts on asymptotic statistical theory, for instance that of van de Vaart (2000). Functional versions of the delta method are also commonly used in semiparametric and nonparametric applications.

For the purposes of this paper, it is sufficient to work in the finite dimensional case. To fix ideas, we use the framework of van de Vaart (2000, chapter 3). Let T_n be a random sequence in $\mathbb{R}^m$ for which $d_n\left(T_n-\theta\right) \Rightarrow T$ as $n \rightarrow \infty$ for some numerical sequence $d_n \rightarrow \infty$. In the usual case $d_n=\sqrt{n}$ and T is Gaussian. Let $\varphi: \mathbb{R}^m \rightarrow \mathbb{R}^p$ be a map that is continuously differentiable at θ with derivative matrix φ'_θ. Then

$$d_n \left(\varphi\left(T_n\right) - \varphi\left(\theta\right)\right) \Rightarrow \varphi'_\theta T. \tag{1}$$

In effect, $d_n \left(\varphi\left(T_n\right) - \varphi\left(\theta\right)\right)$ behaves asymptotically as $n \to \infty$ like the linear functional $\varphi'_\theta T$, which van de Vaart writes as the linear map $\varphi'_\theta\left(T\right)$. The validity of the result relies critically on the validity of a linear approximation at θ as $n \to \infty$. The same critical condition applies in the function space case. In simple applications, the matrix φ'_θ has full row rank and the distribution of T is non degenerate in the sense that its support has positive Lebesgue measure in $\mathbb{R}^m$. Rank deficiencies lead to different rates of convergence and different limit results in the null subspaces. The limit results may be further complicated by the presence of different rates of convergence in the elements of T_n. Many such examples arise in econometrics, especially with models involving trend functions of different orders, such as in systems with cointegrated regressors (Park and Phillips, 1988; Phillips, 1988), systems with slowly varying trend regressors or nonlinear trends (Phillips, 2007; Pollard and Radchenko, 2006), and systems with co-explosive processes (Phillips and Magdalinos, 2008; Nielsen, 2009). These types of complications have been extensively studied in time series econometrics.

But what happens when the function $\varphi = \varphi_n$ also depends on the sample size n? Some very important cases of this type arise in econometrics with the use of simulation based estimators. In this case, an extended delta method for sequences seems well within reach. I could find no general reference in the statistical literature but such results have almost certainly been used before in some asymptotic arguments. A formal statement seems worthwhile.

Consider the special case of a sequence of scalar functions φ_n of a single random sequence T_n. This case will be sufficient for our purposes in the present work but can be substantially generalized. If the functions φ_n are continuously differentiable and their derivatives φ'_n behave with regular variation in the vicinity of the limit θ (relative to the rate of convergence d_n of T_n) then we might expect some version of (1) with a rescaled rate of convergence to hold. The following result is verified by a direct mean value argument.

Theorem 1 *Suppose φ_n has continuous derivatives φ'_n with $\varphi'_n\left(\theta\right) \neq 0$ for all n. Suppose also that the sequence $\{\varphi'_n\}$ is asymptotically locally relatively equicontinuous at θ in the sense that given $\delta > 0$ there exists a sequence $s_n \to \infty$ such that $\frac{s_n}{d_n} \to 0$ and for which as $n \to \infty$*

$$\sup_{|s_n(x-\theta)|<\delta} \left| \frac{\varphi'_n\left(x\right) - \varphi'_n\left(\theta\right)}{\varphi'_n\left(\theta\right)} \right| \to 0. \tag{2}$$

Then

$$\frac{d_n}{\varphi'_n\left(\theta\right)} \left(\varphi_n\left(T_n\right) - \varphi_n\left(\theta\right)\right) \Rightarrow T. \tag{3}$$

As the proof of theorem 1 shows, the conditions effectively require that we may standardize and center the sequence of functions $\varphi_n\left(T_n\right)$ of T_n so that $\frac{d_n}{\varphi'_n(\theta)}\left(\varphi_n\left(T_n\right) - \varphi_n\left(\theta\right)\right)$ is asymptotically linear in T_n in a wide enough neighborhood of θ. If this linearization condition fails, then we need to take further shape characteristics into account in determining the limit theory for $\varphi_n\left(T_n\right)$, just as in the usual delta method asymptotics.

According to the definition of asymptotic local relative equicontinuity, the shrinking neighborhood system of θ may depend on φ_n in order that (2) holds. In particular, the condition requires the existence of some shrinking neighborhood system

$$N_\delta^{s_n}(\theta) = \{x \in \mathbb{R} : |s_n (x - \theta)| < \delta,\ \delta > 0\}, \tag{4}$$

for which (2) holds. So the rate of shrinkage around θ generally depends on the asymptotic behavior of the sequence of functions φ_n. The width of $N_\delta^{s_n}(\theta)$ is $O\left(s_n^{-1}\right)$ and must be large enough to include the local region around θ of $O_p\left(d_n^{-1}\right)$ which contains T_n, at least as $n \to \infty$, which is assured by the rate condition $s_n/d_n \to 0$.

The requirement (2) is stronger than the continuity of φ_n' at θ and different from equicontinuity of $\{\varphi_n'\}$ at θ, which requires a fixed rather than shrinking neighborhood of θ and ignores relative behavior. There is no requirement in the theorem that either $\varphi_n(x)$ or $\varphi_n'(x)$ converge. However, as a consequence of (2), the ratio $\frac{\varphi_n'(x) - \varphi_n'(\theta)}{\varphi_n'(\theta)}$ converges to zero uniformly in a local shrinking neighborhood of θ. Notably, the rate of convergence of $\varphi_n(T_n) - \varphi_n(\theta)$ is modified by the nonrandom sequence $\varphi_n'(\theta)$. When $\varphi_n = \varphi$ for all n, the limit result reduces to the usual delta method formula where $\varphi_n'(\theta) = \varphi'(\theta)$ is a simple slope coefficient. In the general case, the role of $\varphi_n'(\theta)$ changes to that of a slope coefficient combined with a rate of convergence adjustment that takes account of the dependence of the sequence φ_n on n. Just as the usual delta method requires a non-degenerate slope, theorem 1 also requires that $\varphi_n'(\theta) \neq 0$, at least for large enough n. If this condition does not hold, then a higher order version of the result (3) may hold (see Example 3 and the discussion below).

Example 2 *Consider the following sequence of functions*

$$\varphi_n(x) = a n^{-\beta} \sin(n^\alpha x), \quad \varphi_n'(x) = a n^{\alpha-\beta} \cos(n^\alpha x),$$

where a is a constant and α is such that $\frac{n^\alpha}{d_n} \to 0$. Suppose that $d_n T_n \Rightarrow T$ as $n \to \infty$ for some sequence $d_n \to \infty$, so that $T_n \to_p 0$. Observe that $\varphi_n'(0) = a n^{\alpha-\beta}$, and that the sequence $\{\varphi_n'\}$ is locally relatively equicontinuous in the shrinking neighborhood $x \in N_\delta^{s_n} = \{(-\delta/s_n, \delta/s_n) : n^\alpha/s_n \to 0\}$ for some $\delta > 0$ because

$$\sup_{|x-y|<\delta/s_n} |\cos(n^\alpha x) - \cos(n^\alpha y)| \leq n^\alpha \frac{\delta}{s_n} \to 0,$$

provided $n^\alpha/s_n \to 0$. Theorem 1 then implies that $\frac{d_n}{n^{\alpha-\beta}}\varphi_n(T_n) \Rightarrow aT$. Alternatively, since $n^\alpha T_n \to_p 0$ we have the same result by the direct calculation

$$\frac{d_n}{n^{\alpha-\beta}}\varphi_n(T_n) = \frac{a \sin(n^\alpha T_n)}{n^\alpha T_n} d_n T_n \Rightarrow aT.$$

In this example the numerical sequence s_n defining the shrinking neighborhood system $N_\delta^{s_n}$ depends on the form of φ_n because of the condition that $n^\alpha/s_n \to 0$, so the neighborhood shrinks faster than $n^{-\alpha}$. Since $\frac{n^\alpha}{d_n} \to 0$ we can choose the width of the shrinking neighborhood of $\theta = 0$ in such a way that $\frac{n^\alpha}{s_n} + \frac{s_n}{d_n} \to 0$, for example by setting $s_n = n^\alpha \log K_n$ where $K_n = d_n/n^\alpha$. Note that neither $\varphi_n(x)$ nor $\varphi_n'(x)$ converges.

In the above example if the parameter α in $\varphi_n(x)$ is such that $\frac{n^\alpha}{d_n} \to c \in (0,\infty)$, then the continuous mapping theorem applies, rather than an extended delta method. In particular,

$$n^\beta \varphi_n(T_n) = a \sin\left(\frac{n^\alpha}{d_n} d_n T_n\right) \Rightarrow a \sin(cT).$$

The extended delta method fails because higher order terms matter and (2) fails. We can use a full Taylor development to get the same limit result. In particular,

$$\begin{aligned} n^\beta \varphi_n(T_n) &= n^\beta \sum_{j=0}^\infty \frac{\varphi_n^{(j)}(0)\, T_n^j}{j!} = \sum_{k=0}^\infty \frac{(-1)^k n^{\alpha(2k+1)} (d_n T_n)^{2k+1}}{d_n^{2k+1} (2k+1)!} \\ &= \sum_{k=0}^\infty \frac{(-1)^k \{c + o(1)\}^{2k+1} (d_n T_n)^{2k+1}}{(2k+1)!}, \end{aligned} \tag{5}$$

and all terms in the series contribute to the limit distribution. Using the uniform convergence of the series and the Skorohod representation on a probability space for which $d_n T \to_{a.s.} T$, we deduce that on this space

$$n^\beta \varphi_n(T_n) \to_{a.s.} \sum_{k=0}^\infty \frac{(-1)^k c^{2k+1} T^{2k+1}}{(2k+1)!} = \sin(cT), \tag{6}$$

so that weak convergence holds on the original space. Note that when $\frac{n^\alpha}{d_n} \to c > 0$ and $\frac{s_n}{d_n} \to 0$, it follows that $\frac{n^\alpha \delta}{s_n} \to \infty$ for all $\delta > 0$, from which we may deduce that for large enough n

$$\sup_{|x|<\delta/s_n} |\cos(n^\alpha x) - 1| = 1.$$

So asymptotic local relative equicontinuity of $\varphi_n'(x)$ fails in this case.

Example 3 *As in Example 2, suppose $d_n T_n \Rightarrow T$ as $n \to \infty$ for some $d_n \to \infty$. Consider the sequence*

$$\varphi_n(x) = a n^{-\beta} \cos\left(n^\alpha x + b n^{-\gamma}\right), \quad \varphi_n'(x) = -a n^{\alpha-\beta} \sin\left(n^\alpha x + b n^{-\gamma}\right),$$

where a and b are non-zero constants, α is such that $\frac{n^{\alpha+\gamma}}{d_n} \to 0$, and $\gamma > 0$. Now

$$\varphi_n'(0) = -a n^{\alpha-\beta} \sin\left(b n^{-\gamma}\right) = -ab n^{\alpha-\beta-\gamma} \left\{1 + o\left(n^{-1}\right)\right\},$$

so that $\varphi_n'(0) \to 0$ as $n \to \infty$ if $\alpha - \beta - \gamma < 0$. Then

$$\begin{aligned} \frac{\varphi_n'(x) - \varphi_n'(0)}{\varphi_n'(0)} &= \frac{\sin(n^\alpha x + b n^{-\gamma}) - \sin(b n^{-\gamma})}{\sin(b n^{-\gamma})} \\ &= b^{-1} \left\{n^\gamma \sin\left(n^\alpha x + b n^{-\gamma}\right) - n^\gamma \sin\left(b n^{-\gamma}\right)\right\} \left\{1 + o\left(n^{-1}\right)\right\} \end{aligned}$$

and

$$\sup_{|x|<\delta/s_n} \left|n^\gamma \sin\left(n^\alpha x + b n^{-\gamma}\right) - n^\gamma \sin\left(b n^{-\gamma}\right)\right| \le n^{\alpha+\gamma} \frac{\delta}{s_n} \to 0.$$

So $\{\varphi_n'\}$ is locally relatively equicontinuous as $n \to \infty$ in the shrinking neighborhood $x \in N_\delta^{s_n} = \{(-\delta/s_n, \delta/s_n) : n^{\alpha+\gamma}/s_n \to 0\}$ for $\delta > 0$. Theorem 1 then implies that

$$\frac{d_n}{n^{\alpha-\beta-\gamma}}\left(\varphi_n\left(T_n\right) - \varphi_n\left(\theta\right)\right) \Rightarrow -abT.$$

Alternatively, since $n^\alpha T_n \to_p 0$ we have by Taylor expansion

$$\frac{d_n}{n^{\alpha-\beta-\gamma}}\left(\varphi_n\left(T_n\right) - a\cos\frac{b}{n^\theta}\right) = -ab\frac{\sin\left(bn^{-\gamma}\right)}{bn^{-\gamma}}d_nT_n\left\{1 + o_p\left(1\right)\right\} \Rightarrow -abT.$$

Again, s_n depends on the form of φ_n because of the condition that $n^{\alpha+\gamma}/s_n \to 0$, so the neighborhood shrinks faster than $O\left(n^{-\alpha-\gamma}\right)$. The width of $N_\delta^{s_n}$ can be chosen by setting $s_n = n^{\alpha+\gamma}\log K_n$ where $K_n = d_n/n^{\alpha+\gamma}$. In this example since $b \neq 0$ and $\gamma > 0$, the derivative $\varphi_n'\left(0\right)$ converges to zero as $n \to \infty$ when $\alpha < \beta + \gamma$.

When $b = 0$ in the above example, we have $\varphi_n'\left(0\right) = 0$ and the (first order) extended delta method does not apply. But a higher order version is applicable. In particular, a higher order Taylor calculation leads to the result

$$\frac{d_n^2}{n^{2\alpha-\beta}}\left(\varphi_n\left(T_n\right) - a\right) \Rightarrow -\frac{a}{2}T^2.$$

Example 4 *Consider the sequence*

$$Z_n = \varphi_n\left(Y_n\right) = \frac{e^{nY_n}}{\sqrt{n}\left(1 + Y_n\right)}, \quad \text{where } Y_n = X_n^2 \text{ and } \sqrt{n}X_n \Rightarrow X.$$

So $Y_n \to_p 0$ and $nY_n \Rightarrow X^2 =: Y$. Then by direct transformation and neglecting $o_p\left(1\right)$ terms we have by the continuous mapping theorem

$$\sqrt{n}Z_n = \sqrt{n}\varphi_n\left(Y_n\right) = e^{nY_n}\left\{1 + O_p\left(n^{-1}\right)\right\} \Rightarrow e^Y.$$

The usual delta method fails because the first derivative $\varphi_n'\left(0\right) = n^{1/2} + O\left(n^{-1/2}\right)$ is unbounded, linear approximation breaks down and higher order derivatives of $\varphi_n\left(y\right)$ are important. The full Taylor development yields

$$\sqrt{n}Z_n = \sum_{j=0}^\infty \frac{\varphi_n^{(j)}\left(0\right)\left(nY_n\right)^j}{n^{j-1/2}j!},$$

and noting that $\varphi_n^{(j)}\left(0\right) = n^{j-1/2}\left\{1 + O\left(n^{-1}\right)\right\}$ we obtain by an argument similar to (5)-(6)

$$\sqrt{n}Z_n = \sqrt{n}\varphi_n\left(Y_n\right) = \sum_{j=0}^\infty \frac{g_n^{(j)}\left(0\right)\left(nY_n\right)^j}{n^{j-1/2}j!} = \sum_{j=0}^\infty \frac{\left(nY_n\right)^j}{j!} = e^{nY_n} \Rightarrow e^Y,$$

leading again to a continuous map. The full Taylor development gives the correct result but relies on the fact that the function $\varphi_n\left(y\right) = \frac{e^{ny}}{\sqrt{n}(1+y)}$ is analytic over $y \in [0, \infty)$.

2.2 Extending the continuous mapping theorem to implicit maps

If X_n is a random sequence for which $X_n \Rightarrow X$ on a certain probability space and g is a measurable mapping on that space that is continuous except for a set D_g for which the limit measure $P(X \in D_g) = 0$, then $Y_n = g(X_n) \Rightarrow g(X)$. There are well known extensions of this theorem that hold for sequences of functions g_n for which $g_n(X_n) \Rightarrow g(X)$. The result is to be expected if g_n converges uniformly to g. Topsoe (1967) gives a simple and powerful result due to Rubin (undated) according to which if the set $E = \{x : g_n(x_n) \to g(x) \ \forall x_n \to x\}$ has probability one under the limit measure P, then $X_n \Rightarrow X$ implies $g_n(X_n) \Rightarrow g(X)$. See Billingsley (1968) and van de Vaart and Wellner (2000, theorem 1.11.1) for a precise statement, related results and some discussion. The Rubin condition corresponds to a form of asymptotic equicontinuity of $\{g_n\}$ almost everywhere under the limit measure - see van de Vaart and Wellner (2000) and Sweeting (1986). For probability measures on $\mathbb{R}$, if $E = \mathbb{R}$ and the functions g_n are continuous and converge uniformly to g, then $g_n(x_n) \to g(x) \ \forall x_n \to x$ and $P(E) = 1$, so Rubin's condition is assured by uniform convergence on compact sets of $\mathbb{R}$.

Our interest in the current work concerns the limit distribution of random sequences that are determined inversely by sequences of equations of the form

$$X_n = f_n(Y_n), \tag{7}$$

or implicitly by sequences of functions such as

$$h_n(X_n, Y_n) = 0. \tag{8}$$

To my knowledge, there are no limit results for such implicitly defined sequences in the literature. However, given the Rubin-Topsoe result, a limit theory would be expected provided a sequence of globally unique inverse functions exists for (7) and a corresponding sequence of globally unique implicit functions exists for (8) and these sequences are asymptotically equicontinuous in the Rubin sense. .

Conditions for global inverse and global implicit functions have been determined in the mathematics literature since Hadamard (1906) and discussed in economics since Samuelson (1953). Global results of this type are now known for quite general functions on normed spaces (see, for example, Cristea, 2007; and Sandberg, 1980). A variety of conditions can be used to ensure univalence, including monotonicity and P matrix conditions on the Jacobian (see Parthasarathy, 1983, for a review of results up to the early 1980s.) For present purposes in this paper, it will be sufficient to employ results for the real line, where monotonicity is a sufficient condition. The following result uses a one dimensional global implicit function theorem (Ge and Wang, 2002) and will often be convenient in econometric applications. It has been extended by Zhang and Ge (2006) using a Gerschgorin bound condition on the Jacobian to give a global implicit function theorem for mappings in Euclidean spaces of arbitrary dimension.

Lemma 5 *Assume* $f : \mathbb{R}^{m+1} \to \mathbb{R}$ *is continuously differentiable and there exists a constant* $d > 0$ *such that* $\left|\frac{\partial}{\partial y} f(x, y)\right| > d$ *for all* $(x, y) \in \mathbb{R}^m \times \mathbb{R}$. *Then there exists a unique continuously differentiable function* $g : \mathbb{R}^m \to \mathbb{R}$ *such that* $f(x, g(x)) = 0$.

For such an implicit function f with unique solution $y = g(x)$, an implied continuous mapping theorem follows immediately, viz. $Y_n = g(X_n) \Rightarrow g(X)$ whenever $X_n \Rightarrow X$. More generally, suppose we have a sequence of continuously differentiable implicit functions $f_n(x, y)$ which satisfy the monotonicity condition $\left|\frac{\partial}{\partial y} f_n(x, y)\right| > d$ for all $(x, y) \in \mathbb{R}^m \times \mathbb{R}$ and some $d > 0$. Then, there exists a corresponding sequence of unique continuously differentiable solution functions g_n. If these functions satisfy the Rubin asymptotic equicontinuity condition, then $X_n \Rightarrow X$ implies $g_n(X_n) \Rightarrow g(X)$. An application of this type of limit theory is given later in the paper in deriving indirect inference limit theory for the unit root case.

Example 6 *A nontrivial example is given by the function*

$$x = y + e^y =: h(y) \tag{9}$$

whose unique solution is given by

$$y = g(x) := x - W(e^x) \tag{10}$$

where W is Lambert's W function (i.e. the solution of $z = We^W$). The function (9) is graphed in Fig. 1 and is monotonic with derivative bounded above zero. It follows directly that if $X_n \Rightarrow X$ then

$$Y_n = g(X_n) \Rightarrow g(X) = X - W\left(e^X\right).$$

Now consider the sequence of functions $h_n(y) = y + \sum_{j=0}^{n} y^j/j!$ which converges uniformly on compact subsets of $\mathbb{R}$ to $h(y) = y + e^y$. Corresponding to $\{h_n\}$ for large enough n we have a sequence of unique inverse functions $\{g_n\}$ which converges uniformly on compact subsets of $\mathbb{R}$ to the continuous function $g(y) = x - W(e^x)$. Then, $X_n \Rightarrow X$ implies $Y_n = g_n(X_n) \Rightarrow g(X)$.

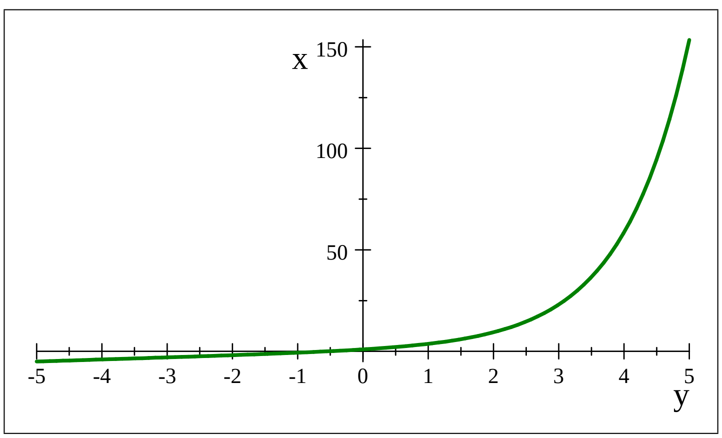

Fig. 1 Graph of $x = y + e^y$ whose solution $y = x - W(e^x)$ involves Lambert's W function.

3 Indirect Inference Estimation

The idea of indirect inference is to use simulated data to determine characteristics such as population moments and to map their dependence on underlying parameters of interest in a manner that is useful in econometric estimation and inference. Like the delta method, this idea has a long history[1]. Practical implementation became possible with advances in computational capability that enabled a sufficiently large number of data generations and replications of a statistical procedure to capture parameter dependencies well enough for them to be used to improve estimation and inference. Typical uses are to estimate parameters indirectly via their dependence on other parameters, which may be easier to estimate, or to use the simulations indirectly for calibration purposes, for example in measuring and correcting bias in estimation.

To fix ideas, a parametric model is simulated to produce H data trajectories $\left\{\tilde{y}^h(\theta)\right\}_{h=1}^H$ for a given parametric value θ. The number of observations in each trajectory $\tilde{y}^h(\theta)$ is chosen to be the same as the number of observations in the observed data set to ensure finite sample calibration accuracy. Suppose $Q_n\left(\beta; y\right)$ is an objective function constructed from the actual data (y) for the estimation of some pseudo parameter β by means of the extremum criterion

$$\hat{\beta}_n = \arg\min{}_{\beta} Q_n(\beta; y).$$

The corresponding estimator based on the h^{th} simulated path for some given θ is

$$\tilde{\beta}_n^h\left(\theta\right) = \arg\min{}_{\beta} Q_n(\beta; \tilde{y}^h(\theta)).$$

Indirect inference estimation of the original parameter θ proceeds by way of calibrating θ to $\hat{\beta}_n$ (or some function of $\hat{\beta}_n$) according to an additional criterion of the form

$$\breve{\theta}_{n,H} = \arg\min{}_{\theta} \left\| \hat{\beta}_n - \frac{1}{H}\sum_{h=1}^{H} \tilde{\beta}_n^h\left(\theta\right) \right\|, \tag{11}$$

for some metric $\left\|\cdot\right\|$. As $H \to \infty$, we anticipate that $H^{-1}\sum_{h=1}^{H} \tilde{\beta}_n^h\left(\theta\right) \to_p \mathrm{E}\tilde{\beta}_n^h\left(\theta\right) =: b_n\left(\theta\right)$. Since H can be made arbitrarily large in implementation, the procedure effectively amounts to calibrating $b_n\left(\theta\right)$, which is called the binding function, so that

$$\breve{\theta}_n = \arg\min{}_{\theta} \left\| \hat{\beta}_n - b_n\left(\theta\right) \right\|. \tag{12}$$

If $b_n\left(\theta\right)$ is invertible, then we have $\breve{\theta}_n = b_n^{-1}\left(\hat{\beta}_n\right) =: f_n\left(\hat{\beta}_n\right)$. The estimator $\hat{\theta}_n$ is therefore determined indirectly by way of the binding function $b_n\left(\theta\right)$ and the estimator $\hat{\beta}_n$. In some applications of indirect inference, such as the one considered in the next section of the paper, the pseudo parameter β corresponds with the original parameter θ and the procedure seeks to adjust the estimator according to some aspect of its sampling properties such as its mean or median.

[1] For instance, Durbin indicated early consideration of such possibilities in an ET Interview (Phillips, 1988) and Sargan (1976) mentions ideas of Barnard related to the bootstrap, both in the 1950s.

Importantly, the dependence of the estimator $\breve{\theta}_n$ on the data is via $\hat{\beta}_n$ and the sequence of binding functions b_n. In general, b_n depends on the finite sample distribution of the data through the exact finite sample functional involved in the criterion. In the case above, the functional is the finite sample mean function $\mathrm{E}\tilde{\beta}_n^h(\theta)$. But it could also be another characteristic of the distribution like the median. The implicit dependence of $\breve{\theta}_n$ on the sequence of functions b_n means that the asymptotic distribution of $\breve{\theta}_n$ cannot be deduced simply by the delta method. As shown above, it is necessary to take into account the properties of the sequence b_n in determining the rate of convergence and the limit theory. The remainder of this paper will look carefully at this problem in a special case that shows how the mapping sequence can play a critical role in shaping the limit theory.

4 First Order Autoregression

4.1 Bias and bias correction

Suppose we wish to estimate the parameter ρ in the simple autoregression

$$y_t = \rho y_{t-1} + u_t, \quad t = 1, ..., n \tag{13}$$

from observations $y = \{y_t\}_{t=0}^n$ where u_t is *iid* $N\left(0, \sigma^2\right)$. Various conditions may be placed on the initial value y_0 and these affect finite sample behavior and may also affect the limit theory when ρ is in the neighborhood of unity or in the explosive region (see Phillips and Magdalinos, 2010, for a recent treatment and the references therein). Such initialization effects are not the concern of the present paper, so we will simply assume that $y_0 = 0$. However, the indirect inference approach is easily adapted to take into account different initializations. Also, it is often convenient to focus on the case where $\rho > 0$ since analogous mirror image results hold for $\rho < 0$.

Standard estimation procedures such as maximum likelihood (ML) and least squares (LS) produce downward biased coefficient estimators of ρ in finite samples when $\rho > 0$. Let $\hat{\rho}_n$ be the ML estimate of ρ under Gaussianity, assuming the initialization y_0 is fixed. White (1961) and Shenton and Johnson (1965; hereafter SJ) gave asymptotic expansions of the bias in terms of powers of n^{-1} as $n \to \infty$. Different expansions were obtained for the case $|\rho| < 1$ and the case $\rho = 1$. In more recent work, Shenton and Vinod (1995) gave integral forms for the bias function for stationary and unit root ρ and developed a high order closed expression for the asymptotic expansion of the bias. Some related work giving analytic moment expressions is contained in Vinod and Shenton (1996) again for these parameter values. Extensions to models with non-Gaussian errors were derived in Bao (2007) for the stationary case. All of this research has a bearing on the estimation of continuous time models from discrete data, where similar problems of estimation bias for the mean reversion parameter arise but can be more severe (Tang and Chen, 2009; Yu, 2009). This bias is particularly important because of its implications for derivative pricing in finance (Phillips and Yu, 2005; 2009).

The next section develops comprehensive bias expressions for $\hat{\rho}_n$ and asymptotic representations that cover stationary, unit root and explosive ρ. This development is needed because the asymptotic formulae required to characterize the limit theory of the indirect inference estimator of ρ rely on full analytic specification of the binding function over all potential values of ρ. Fig. 2 shows the bias function $\mathrm{E}(\hat{\rho}_n|\rho) - \rho$ of the ML estimate of ρ in the Gaussian model (13) for various sample sizes n. The downward bias for $\rho > 0$ is evidently greatest near unity and rapidly diminishes as ρ exceeds unity. The bias function is clearly very nonlinear, as noted by MacKinnon and Smith (1998), who performed simulations in the case of an AR(1) model with a fitted intercept. Most importantly, it has a rapidly changing derivative in the vicinity of unity.

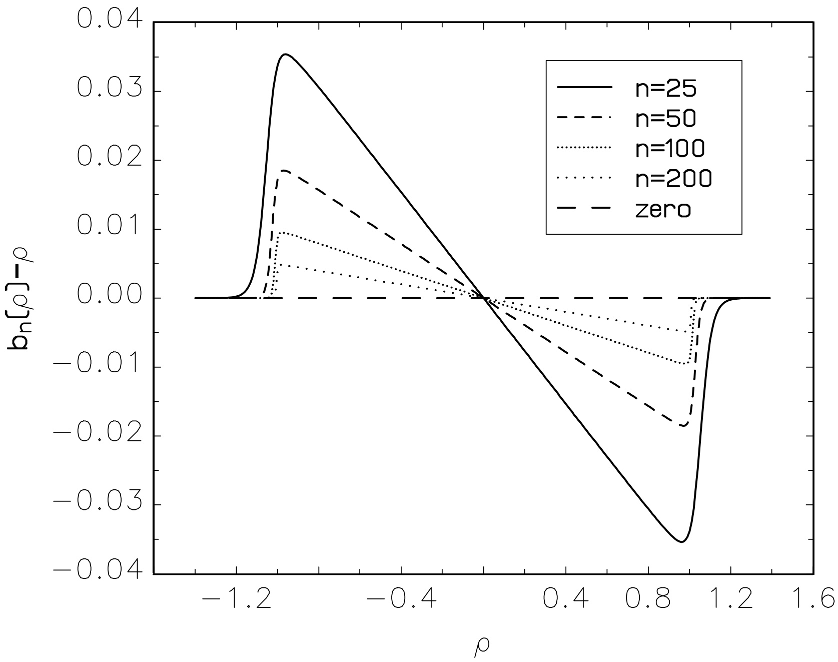

Fig. 2 The exact bias function $b_n(\rho) - \rho = \mathrm{E}(\hat{\rho}_n|\rho) - \rho$ of the ML estimator $\hat{\rho}_n$ for various n based on (19) and (20).

The indirect inference method for fitting ρ takes this bias function into account and was explored in Gouriéroux et al (2000, 2010) by Monte Carlo. As explained above, the approach uses simulations to calibrate the bias function and requires neither an explicit form of the bias nor a bias expansion formula. The simulation results reported in Gouriéroux et al (2000) show that the indirect inference method works as well as the median unbiased estimator of Andrews (1993) when $H = 15,000$ and the calibration estimator is the MLE. Both methods are dependent on the validity of the assumed data distribution for correct calibration through the finite sample binding formula.

As in (12), when the number of replications $H \to \infty$ the indirect inference estimator of ρ satisfies

$$\breve{\rho}_n = \arg\min{}_{\rho} \left|\hat{\rho}_n - E\left(\tilde{\rho}_n^h(\rho)\right)\right| = \arg\min{}_{\rho} \left|\hat{\rho}_n - b_n(\rho)\right|, \tag{14}$$

where $b_n(\rho) = E(\tilde{\rho}_n^h(\rho))$ is the binding function for the MLE $\hat{\rho}_n$. When b_n is invertible

$$\breve{\rho}_n = b_n^{-1}(\hat{\rho}_n) := f_n(\hat{\rho}_n). \tag{15}$$

From SJ (1965), the binding function is known to have the following asymptotic expansion as $n \to \infty$

$$b_n(\rho) = \begin{cases} \rho - \frac{2\rho}{n} + O\left(n^{-2}\right) & |\rho| < 1 \\ \rho - \frac{1.7814}{n} + O\left(n^{-2}\right) & \rho = 1 \end{cases}, \tag{16}$$

which is evidently discontinuous at $\rho = 1$. The numerical value -1.7814 is the mean of the limit distribution of $n\left(\hat{\rho}_n - 1\right)$ and bias persists in the limit when $\rho = 1$. The discontinuity in (16) reflects the discontinuity in the asymptotic distribution theory around unity and manifests this deeper issue in the asymptotics. In contrast, the binding function $b_n(\rho)$ itself is continuous and indeed continuously differentiable for all n, as is apparent in Fig. 3a.

Fig. 3b shows the binding function for $n = 5,000$ in a narrow band around unity to indicate the behavior of the function in this vicinity for very large values of n. The function is below the 45^o line for all ρ with a slope that is less than unity for stationary ρ but that increases and exceeds unity for ρ around unity while rapidly returning to virtually coincide with the 45^o line for explosive ρ. In order to accomplish this smooth transition, the derivative of the binding function is below unity for $\rho < 1$, virtually unity for $\rho > 1$ but greater than unity in the immediate vicinity of $\rho = 1$. As is apparent from Figs. 3a and 3b, the binding function $b_n(\rho)$ is monotone. But Fig. 2 shows that the bias function $b_n(\rho) - \rho$ has a derivative that quickly changes sign in the neighborhood of unity. So a linear approximation to $b_n(\rho)$ is completely inadequate around unity even for very large n.

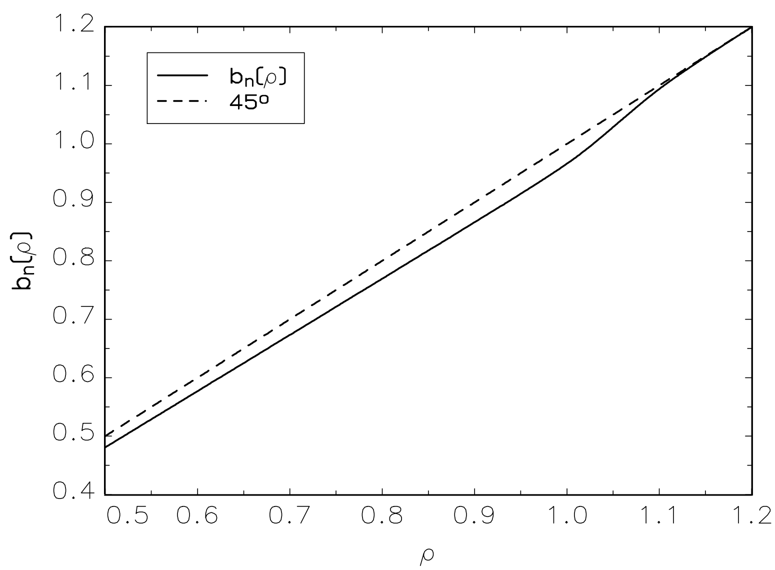

Fig. 3a Graph of the binding function $b_n(\rho)$ of the MLE $\hat{\rho}_n$ for $n = 100$.

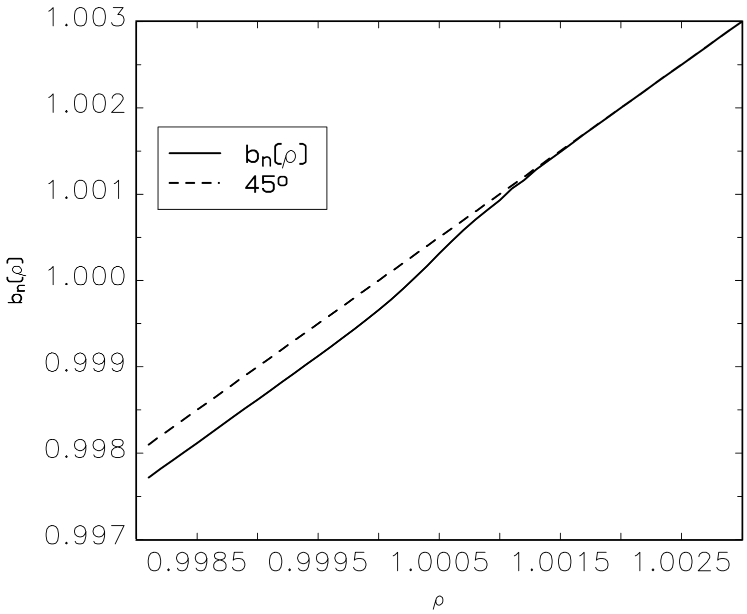

Fig. 3b Graph of the binding function $b_n(\rho)$ of the MLE $\hat{\rho}_n$ around unity for $n = 5000$.

Since the inverse binding function f_n is continuously differentiable with a non zero first derivative at ρ, routine application of the delta method suggests that

$$d_n(\breve{\rho}_n - \rho) \sim f_n'(\rho) d_n(\hat{\rho}_n - \rho). \tag{17}$$

When $|\rho| < 1$, we have $d_n = \sqrt{n}$ and by standard theory $\sqrt{n}(\hat{\rho}_n - \rho) \Rightarrow N(0, 1-\rho^2)$. In this case, as shown in the following section, $f_n'(\rho) = 1 + O(n^{-1})$ and, given $\delta > 0$ and a sequence $s_n \to \infty$ such that $s_n/\sqrt{n} \to 0$, we have

$$\sup_{|s_n(x-\rho)|<\delta} \left| \frac{f_n'(x) - f_n'(\rho)}{f_n'(\rho)} \right| \to 0, \quad \text{as } n \to \infty.$$

Then by theorem 1 it follows that $\sqrt{n}(\breve{\rho}_n - \rho) \Rightarrow N(0, 1-\rho^2)$. The main effect of indirect inference in the stationary AR(1) case therefore is to provide a finite sample bias correction to the estimator, while the asymptotic distribution of $\breve{\rho}_n$ is identical to the MLE.

However, when ρ is in the local vicinity of unity as $n \to \infty$, the linear representation (17) breaks down and it is necessary to take into account the precise features of the binding function $b_n(\rho)$ around unity to determine the correct limit theory. The asymptotics are complex and require much more detailed asymptotic representations of $b_n(\rho)$. These are provided in the following sections.

4.2 The binding function formula

The following theorem extends a result in Shenton and Vinod (1996; hereafter SV). It describes the binding function for the regions $|\rho| \le 1$ and $|\rho| > 1$.

Theorem 7 *For model (13) the binding function $b_n(\rho) = E(\hat{\rho}_n)$ for the ML estimator $\hat{\rho}_n$ is given by*

$$b_n(\rho) = \begin{cases} \rho + \frac{1}{2}\frac{\partial}{\partial\rho}\left\{\int_0^1 x^{(n-5)/2}\left(1-\rho^2x^2\right)^{3/2} F_n^{-1/2}dx\right\} & |\rho| \le 1 \\ \rho + \frac{1}{2}\frac{\partial}{\partial\rho}\left\{\int_1^\infty x^{(n-5)/2}\left(\rho^2x^2-1\right)^{3/2} G_n^{-1/2}dx\right\} & |\rho| > 1 \end{cases} \tag{18}$$

where

$$F_n = F_n(x;\rho) = 1 - \rho^2 x + (1-x)\, x^{2n-1}\rho^{2n},$$
$$G_n = G_n(x;\rho) = \rho^2 x - 1 + (x-1)\, x^{2n-1}\rho^{2n}.$$

Remarks

1. The proof of theorem 7 follows SJ (1965) and SV (1996) in using results for ratios of quadratic forms in normal variates. SV develop the integral representation (18) for the case $|\rho| \le 1$. The present result extends that work to the explosive case and provides explicit representations of the bias for $|\rho| \le 1$ and $|\rho| > 1$. These representations are then used to develop a complete set of asymptotic expansions which facilitate the development of the limit theory for the indirect inference estimator.

2. Explicit formulae for (18) are derived in the proof of theorem 7. For $|\rho| \le 1$ (see (52))

$$\begin{aligned} b_n(\rho) = \rho &- \frac{3\rho}{2}\int_0^1 x^{(n-1)/2}\left(1-\rho^2x^2\right)^{1/2} F_n^{-1/2}dx + \frac{\rho}{2}\int_0^1 x^{(n-3)/2}\left(1-\rho^2x^2\right)^{3/2} F_n^{-3/2}dx \\ &- \frac{n\rho^{2n-1}}{2}\int_0^1 x^{(5n-7)/2}\left(1-\rho^2x^2\right)^{3/2} F_n^{-3/2}(1-x)\,dx, \end{aligned} \tag{19}$$

and for $|\rho| > 1$ (see (54))

$$\begin{aligned} b_n(\rho) = \rho &+ \frac{3\rho}{2}\int_1^\infty x^{(n-1)/2}\left(\rho^2x^2-1\right)^{1/2} G_n^{-1/2}dx - \frac{\rho}{2}\int_1^\infty x^{(n-3)/2}\left(\rho^2x^2-1\right)^{3/2} G_n^{-3/2}dx \\ &- \frac{n\rho^{2n-1}}{2}\int_1^\infty x^{(5n-7)/2}\left(\rho^2x^2-1\right)^{3/2} G_n^{-3/2}(x-1)\,dx. \end{aligned} \tag{20}$$

These expressions are continuous through $\rho = 1$, as shown in Figs. 3a and 3b. An alternate form of (20) that is convenient for computation is (see (55))

$$\begin{aligned} b_n(\rho) = \rho &+ \frac{3\rho}{2}\int_0^1 y^{(n-5)/2}\left(\rho^2-y^2\right)^{1/2} H_n^{-1/2}(y;\rho)\,dy \\ &- \frac{\rho}{2}\int_0^1 y^{(5n-7)/2}\left(\rho^2-y^2\right)^{3/2} H_n^{-3/2}(y;\rho)\,dy \\ &- \frac{n\rho^{2n-1}}{2}\int_0^1 y^{(n-5)/2}\left(\rho^2-y^2\right)^{3/2} H_n^{-3/2}(y;\rho)(1-y)\,dx, \end{aligned}$$

where $H_n(y;\rho) = \left(\rho^2 - y\right) y^{2n-1} + (1-y)\,\rho^{2n}$.

4.3 Asymptotic bias expansions

As discussed earlier, taking asymptotic expansions of the bias function leads to discontinuities that reflect fundamental differences in the limit theory as $n \to \infty$. The technical reason for these discontinuities stems from the presence of terms such as ρ^{2n} in the binding function $b_n(\rho)$, which behave differently depending on whether $|\rho| < 1$, $|\rho| > 1$, or $\rho = 1 + c/n$. To provide a comprehensive analysis, we consider each of these domains separately. The following result summarizes the main cases of interest.

Theorem 8 *For fixed* ρ

$$b_n(\rho) = \begin{cases} \rho - \frac{2\rho}{n} + O\left(n^{-2}\right) & |\rho| < 1 \\ \pm 1 \mp \frac{1.7814}{n} + O\left(n^{-2}\right) & \rho = \pm 1 \\ \rho + O\left(|\rho|^{-n}\right) & |\rho| > 1 \end{cases} . \tag{21}$$

For $\rho = 1 + c/n$ *with* $c < 0$

$$b_n(1 + \frac{c}{n}) = 1 + \frac{c}{n} - \frac{3}{4n} \int_0^1 y^{-\frac{3}{4}} \ell(y, c)^{-1/2} dy + \frac{1}{4n} \int_0^1 y^{-\frac{3}{4}} \ell(y, c)^{-3/2} dy + \frac{e^{2c}}{8n} \int_0^1 y^{\frac{1}{4}} \ell(y, c)^{-3/2} \log y dy + O\left(n^{-2}\right). \tag{22}$$

For $\rho = 1 + c/n$ *with* $c > 0$

$$b_n(1 + \frac{c}{n}) = 1 + \frac{c}{n} + \frac{3}{4n} \int_0^\infty e^{\frac{1}{4}w} k^+(w; c)^{1/2} dw - \frac{1}{4n} \int_0^\infty e^{\frac{1}{4}w} k^+(w; c)^{3/2} dw - \frac{e^{2c}}{8n} \int_0^\infty e^{\frac{5}{4}w} k^+(w; c)^{3/2} w dw + O\left(n^{-2}\right). \tag{23}$$

In the above formulae

$$\ell(y, c) := \frac{4c + \log y}{4c + 2\log y} + \frac{\log y}{4c + 2\log y} y e^{2c}, \text{ and } k^+(w; c) := \frac{4c + 2w}{4c + w + e^{2c} w e^w}. \tag{24}$$

The error orders in (22) and (23) hold uniformly for c *in compact sets of* $\mathbb{R}$.

Remarks

1. The results for $|\rho| < 1$ and $\rho = \pm 1$ are well known. The result for $|\rho| > 1$ appears to be new, as are the results for the local to unity cases. The latter results are particularly useful in deriving the limit distribution of the indirect inference estimator, as we discuss later. The distinction between $c < 0$ and $c > 0$ arises because of the formulation of the binding function $b_n(\rho)$ in these two cases and the manner in which the asymptotic expansions are obtained. These issues are expanded on below.

2. When $c \nearrow 0$ and $c \searrow 0$, the bias function expansions (22) and (23) converge to the same value. So this local formulation, just like the function b_n, is continuous. In particular, we have

$$\begin{aligned}
\lim_{c \nearrow 0} b_n(1+\frac{c}{n}) &= 1 - \frac{3}{4n} 2^{1/2} \int_0^1 y^{-\frac{3}{4}} \{1+y\}^{-1/2} dy + \frac{2^{3/2}}{4n} \int_0^1 y^{-\frac{3}{4}} \{1+y\}^{-3/2} dy \\
&+ \frac{2^{1/2}}{4n} \int_0^1 y^{\frac{1}{4}} \{1+y\}^{-3/2} \log y dy + O\left(n^{-2}\right) \\
&= 1 + \frac{1}{n} \left\{ -\frac{3}{4} 2^{0.5} (3.7081) + \frac{2^{1.5}}{4} (3.2683) - \frac{2^{1/2}}{4} (0.45077) \right\} + O\left(n^{-2}\right) \\
&= 1 - \frac{1.7814}{n} + O\left(n^{-2}\right), \qquad (25)
\end{aligned}$$

and

$$\begin{aligned}
\lim_{c \searrow 0} b_n(1+\frac{c}{n}) &= 1 + \frac{3}{4n} \int_0^\infty e^{\frac{1}{4}w} \left\{ \frac{2w}{w+we^w} \right\}^{1/2} dw - \frac{1}{4n} \int_0^\infty e^{\frac{1}{4}w} \left\{ \frac{2w}{w+we^w} \right\}^{3/2} dw \\
&- \frac{1}{8n} \int_0^\infty e^{\frac{5}{4}w} \left\{ \frac{2w}{w+we^w} \right\}^{3/2} w dw + O\left(n^{-2}\right) \\
&= 1 + \frac{1}{n} \left\{ \frac{3}{4} 5.2441 - \frac{1}{4} 1.2441 - \frac{43.2278}{8} \right\} + O\left(n^{-2}\right) \\
&= 1 - \frac{1.7814}{n} + O\left(n^{-2}\right). \qquad (26)
\end{aligned}$$

3. Further, when $\rho = 1 + c/n$ with $c < 0$ and $c \searrow -\infty$, we have

$$\begin{aligned}
b_n(\rho) &= \rho - \frac{3\rho}{4n} \int_0^1 y^{-\frac{3}{4}} dy + \frac{\rho}{4n} \int_0^1 y^{-\frac{3}{4}} + O\left(n^{-2} + \frac{1}{|c|}\right) \\
&= \rho - \frac{2\rho}{n} + O\left(n^{-2} + \frac{1}{|c|}\right),
\end{aligned}$$

corresponding to the case of fixed $|\rho| < 1$; and when $\rho = 1 + c/n$ with $c > 0$ and $c \nearrow \infty$, we have

$$b_n(\rho) = \rho + O\left(e^{-c}\right),$$

corresponding to the fixed $\rho > 1$ case.

4. An alternative form of the binding function when $\rho = 1 + c/n$ and $c < 0$ is useful and is given in the following result.

Corollary 9 *For $\rho = 1 + c/n$ with $c < 0$, (22) also has the form*

$$\begin{aligned}
b_n(1+\frac{c}{n}) &= 1 + \frac{c}{n} - \frac{3}{4} \int_0^\infty e^{-\frac{1}{4}v} k^-(v;c)^{1/2} dv + \frac{1}{4} \int_0^\infty e^{-\frac{1}{4}v} k^-(v;c)^{3/2} dv \\
&- \frac{e^{2c}}{8} \int_0^\infty e^{-\frac{5}{4}v} k^-(v;c)^{3/2} v dv + O\left(n^{-2}\right), \qquad (27)
\end{aligned}$$

where

$$k^-(v;c) := \frac{4c - 2v}{4c - v - e^{2c} v e^{-v}}. \qquad (28)$$

5 Indirect inference limit theory

The indirect inference estimator of ρ is defined implicitly in terms of the binding function of the ML estimator $\hat{\rho}_n$, so that $\hat{\rho}_n = b_n(\breve{\rho})$, where b_n is given by (19) and (20). We write this implicit relationship for $\breve{\rho}$ as

$$\begin{aligned}\hat{\rho}_n &= b_n(\breve{\rho}) \\ &= b_n(\breve{\rho}; |\breve{\rho}| \leq 1) + b_n(\breve{\rho}; |\breve{\rho}| > 1),\end{aligned}$$

where

$$\begin{aligned} b_n(\rho; |\rho| \leq 1) = \rho &- \frac{3\rho}{2}\int_0^1 x^{(n-1)/2}\left(1-\rho^2x^2\right)^{1/2} F_n^{-1/2}dx \\ &+ \frac{\rho}{2}\int_0^1 x^{(n-3)/2}\left(1-\rho^2x^2\right)^{3/2} F_n^{-3/2}dx - \frac{n\rho^{2n-1}}{2}\int_0^1 x^{(5n-7)/2}\left(1-\rho^2x^2\right)^{3/2} F_n^{-3/2}(1-x)\,dx, \end{aligned} \tag{29}$$

and

$$\begin{aligned} b_n(\rho; |\rho| > 1) = \rho &+ \frac{3\rho}{2}\int_1^\infty x^{(n-1)/2}\left(\rho^2x^2-1\right)^{1/2} G_n^{-1/2}dx \\ &- \frac{\rho}{2}\int_1^\infty x^{(n-3)/2}\left(\rho^2x^2-1\right)^{3/2} G_n^{-3/2}dx - \frac{n\rho^{2n-1}}{2}\int_1^\infty x^{(5n-7)/2}\left(\rho^2x^2-1\right)^{3/2} G_n^{-3/2}(x-1)\,dx. \end{aligned} \tag{30}$$

To find the limit distribution of $\breve{\rho}$ we use asymptotic formulae for the binding function $b_n(\rho)$ and its derivatives. We consider the stationary and near unit root cases separately.

5.1 Stationary case

When $|\rho| < 1$, the extended delta method of theorem 1 is applicable. To show this, consider the first derivative of the binding function. As is clear from (29), when $|\rho| < 1$ the final term in the binding function expression is $O(\rho^n)$. The first derivative of this function is of the same order and since it is dominated by the other terms it is neglected in the calculations below. For $|\rho| < 1$ we therefore have

$$
\begin{aligned}
b_n'(\rho) &= 1 - \frac{\partial}{\partial\rho}\left\{\frac{3\rho}{2}\int_0^1 x^{(n-1)/2}\left(1-\rho^2x^2\right)^{1/2}F_n^{-1/2}dx\right\} \\
&+ \frac{\partial}{\partial\rho}\left\{\frac{\rho}{2}\int_0^1 x^{(n-3)/2}\left(1-\rho^2x^2\right)^{3/2}F_n^{-3/2}dx\right\} + O\left(\rho^n\right) \\
&= 1 - \frac{3}{2}\int_0^1 x^{(n-1)/2}\left(1-\rho^2x^2\right)^{1/2}F_n^{-1/2}dx + \frac{3\rho^2}{2}\int_0^1 x^{(n-5)/2}\left(1-\rho^2x^2\right)^{-1/2}F_n^{-1/2}dx \\
&+ \frac{3\rho}{4}\int_0^1 x^{(n-1)/2}\left(1-\rho^2x^2\right)^{1/2}F_n^{-3/2}\frac{\partial}{\partial\rho}F_n dx \\
&+ \frac{1}{2}\int_0^1 x^{(n-3)/2}\left(1-\rho^2x^2\right)^{3/2}F_n^{-3/2}dx - \frac{3\rho^2}{2}\int_0^1 x^{(n-7)/2}\left(1-\rho^2x^2\right)^{1/2}F_n^{-3/2}dx \\
&- \frac{3\rho}{4}\int_0^1 x^{(n-3)/2}\left(1-\rho^2x^2\right)^{3/2}F_n^{-5/2}\frac{\partial}{\partial\rho}F_n dx + O\left(\rho^n\right).
\end{aligned} \tag{31}
$$

Now $F_n = 1-\rho^2x+(1-x)\,x^{2n-1}\rho^{2n}$ and

$$
\frac{\partial}{\partial\rho}F_n = -2\rho x + 2n\,(1-x)\,x^{2n-1}\rho^{2n-1} = -2\rho x + O\left(n\rho^{2n-1}\right), \tag{32}
$$

so that substituting (32) into (31) and using (40) of lemma 10, we deduce that for $|\rho|<1$

$$
b_n'(\rho) = 1 + O\left(n^{-1}\right).
$$

It follows that, given $|\rho|<1$, for all $\delta>0$ and any sequence $s_n\to\infty$ for which $s_n/n^{1/2}\to 0$, we have

$$
\sup_{s_n|r-\rho|<\delta}\left|\frac{b_n'(\rho)-b_n'(r)}{b_n'(r)}\right| \to 0.
$$

Writing $\breve{\rho} = b_n^{-1}(\hat{\rho}_n) = f_n(\hat{\rho}_n)$ and using the fact that $f_n'(r) = 1/b_n'(r)$ and

$$
\frac{f_n'(r)-f_n'(\rho)}{f_n'(\rho)} = \frac{b_n'(\rho)-b_n'(r)}{b_n'(r)},
$$

it follows that

$$
\sup_{s_n|r-\rho|<\delta}\left|\frac{f_n'(r)-f_n'(\rho)}{f_n'(\rho)}\right| \to 0.
$$

Hence, by theorem 1,

$$
\sqrt{n}\,(\breve{\rho}-\rho) \sim \frac{1}{b_n'(\rho)}\sqrt{n}\,(\hat{\rho}_n-\rho) \sim \sqrt{n}\,(\hat{\rho}_n-\rho) \Rightarrow N\left(0,1-\rho^2\right).
$$

5.2 Unit root case

The unit root case is considerably more complex because of the implicit determination of $\breve{\rho}$ via the mapping $\hat{\rho}_n = b_n(\breve{\rho})$. No explicit functional form for the inverse mapping $\breve{\rho} = b_n^{-1}(\hat{\rho}_n)$ is available, although series expressions may be obtained using Lagrange inversion. Instead of an explicit inverse map, it turns out that we can directly manipulate the expression to accommodate standardized and centred versions of the ML estimator $\xi_n^{ml} = n(\hat{\rho}_n - \rho)$ and the II estimator $\xi_n^{ii} = n(\breve{\rho} - \rho)$. The transformed mapping may be used to deduce the limit theory for ξ_n^{ii} using an implicit function version of the continuous mapping theorem. This approach is applicable when $\rho = 1$ and when $\rho = 1 + c/n$.

We start with the bias expressions for $\hat{\rho}_n$ in the near integrated case $\rho = 1 + c/n$. We need to allow for both $c \leq 0$ and $c > 0$. So we combine (22), or its alternative form (27), with (23). Then, in general for $\rho = 1 + c/n$, we have

$$b_n(\rho) = \rho + \frac{1}{n} g(c) + O\left(n^{-2}\right), \tag{33}$$

where

$$g(c) = g^-(c) 1_{\{c \leq 0\}} + g^+(c) 1_{\{c > 0\}} \tag{34}$$

with

$$g^-(c) = -\frac{3}{4}\int_0^\infty e^{-\frac{1}{4}v} k^-(v;c)^{1/2} dv + \frac{1}{4}\int_0^\infty e^{-\frac{1}{4}v} k^-(v;c)^{3/2} dv - \frac{e^{2c}}{8}\int_0^\infty e^{-\frac{5}{4}v} k^-(v;c)^{3/2} v dv, \tag{35}$$

for $c \leq 0$ and

$$g^+(c) = \frac{3}{4}\int_0^\infty e^{\frac{1}{4}w} k^+(w;c)^{1/2} dw - \frac{1}{4}\int_0^\infty e^{\frac{1}{4}w} k^+(w;c)^{3/2} dw - \frac{e^{2c}}{8}\int_0^\infty e^{\frac{5}{4}w} k^+(w;c)^{3/2} w dw, \tag{36}$$

for $c > 0$ where $k^-(v;c)$ and $k^+(w;c)$ are defined in (24) and (28).

In view of earlier results, the equation error in (33) holds uniformly for c in compact sets of $\mathbb{R}$. Observe that $k^-(v;c) > 0$ for all $c \leq 0$ over $v \in (0,\infty)$, and $k^+(w;c) > 0$ for all $c > 0$ over $w \in (0,\infty)$. Hence, the integrands in (35) and (36) are real and well defined. Moreover, as shown in (25) and (26), when $c \to \pm 0$, we get $g^-(0) = g^+(0) = -1.7814$ and the function $g(c)$ is continuous through $c = 0$.

The derivatives of the binding function $b_n(\rho)$ in the vicinity of unity have a different form from when $|\rho| < 1$. In particular, terms involving ρ^{2n} in (29) and (32) are no longer exponentially small. Calculations reveal that for $\rho = 1 + c/n$, the derivatives take the following form

$$b_n^{(j)}(\rho) = \begin{cases} 1 - \frac{1}{2^{1/2}}\int_0^1 y^{1/4}\left(1 + e^{2c}y\right)^{-3/2} \log y dy + O\left(n^{-1}\right) & j = 1 \\ \frac{(2n)^{j-1}}{2^{1/2}}\int_0^1 y^{1/4}\left(1 + e^{2c}y\right)^{-3/2} \log y dy \{1 + o(1)\} & j > 1 \end{cases}$$

and therefore satisfy $b_n^{(j)}(1) = O\left(n^{j-1}\right)$, so that second and higher derivatives are unbounded at $\rho = 1$ as $n \to \infty$. This corresponds to the rapidly changing form of the bias function $b_n(\rho) - \rho$ in the vicinity of unity that is evident in Fig. 2. As a result, the extended delta method fails for ρ in the immediate vicinity of unity. Note, in particular, that for some intermediate value $\tilde{r}$ between r and ρ we have

$$\sup_{s_n|r-\rho|<\delta}\left|\frac{b_n'(\rho) - b_n'(r)}{b_n'(r)}\right| = \sup_{s_n|r-\rho|<\delta}\left|\frac{b_n^{(2)}(\tilde{r})(\rho - r)}{b_n'(r)}\right| = O_p\left(n \times \frac{\delta}{s_n}\right),$$

which is divergent for all sequences $s_n \to \infty$ for which $s_n/n \to 0$. Hence the sequence b_n (and by implication b_n^{-1}) is not asymptotically locally relatively equicontinuous and theorem 1 does not apply. One way of addressing this failure in the delta method is to attempt a full Taylor representation of $b_n(\rho)$ and Lagrange inversion, as in the examples discussed in Section 2. However, a more direct approach turns out to be possible using the relation (33).

Since $\hat{\rho}_n = b_n(\breve{\rho})$, we have by direct substitution of the centred and scaled estimates $\xi_n^{ml} = n(\hat{\rho}_n - \rho)$ and $\xi_n^{ii} = n(\breve{\rho} - \rho)$ into (33)

$$\xi_n^{ml} = \xi_n^{ii} + g\left(\xi_n^{ii}\right) + O\left(n^{-1}\right) =: h\left(\xi_n^{ii}\right) + O\left(n^{-1}\right), \tag{37}$$

where $h(c) = c + g(c)$ and $g(c)$ is given by (34). Equation (37) defines a sequence of implicit mappings that determine ξ_n^{ii} in terms of ξ_n^{ml}. The functions g and h are independent of n. In the limit as $n \to \infty$, we have $\xi_n^{ml} \sim h\left(\xi_n^{ii}\right)$. So, the limit function h implicitly determines the limit distribution of ξ_n^{ii}.

The limit function h is graphed in Fig. 4. This function is monotonic and continuous (including the point $c = 0$ - see the argument below) and a continuous inverse function h^{-1} exists by virtue of lemma 5. The shape of the limit function h is remarkably similar to that of the binding function b_n shown in Fig. 3.

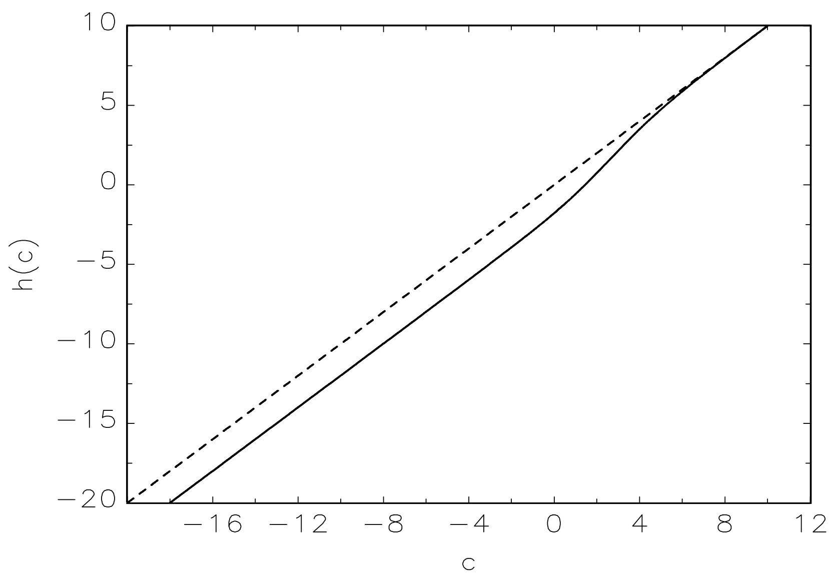

Fig. 4 The limit function h (solid line) in the implicit map $\xi^{ml} = h\left(\xi^{ii}\right)$ relating the indirect inference limiting variate ξ^{ii} to the limiting ML variate $\xi^{ml} = \int W dW / \int W^2$, shown against the 45° line (broken line).

The remaining argument is straightforward. According to standard theory (Phillips, 1987), $\xi_n^{ml} \Rightarrow \xi^{ml} = \int_0^1 W dW / \int_0^1 W^2$, where W is a standard Brownian motion. By the Skorohod representation theorem, we may enlarge the probability space with distributionally equivalent random sequences for which $\xi_n^{ml} \to_{a.s.} \xi^{ml}$. On this space by the continuity of the inverse map h^{-1} we deduce that $\xi_n^{ii} \to_{a.s} \xi^{ii}$, where ξ^{ii} is the solution of

$\xi^{ml} = h\left(\xi^{ii}\right)$. Hence, in the original space we have $\xi_n^{ii} \Rightarrow \xi^{ii}$, as $n \to \infty$ by the implicit continuous mapping theorem. Thus, the limit distribution of ξ^{ii} in the unit root case is given by

$$n\left(\breve{\rho} - 1\right) \Rightarrow h^{-1}\left(\int_0^1 W dW / \int_0^1 W^2\right), \tag{38}$$

where h^{-1} is the inverse function of $h\left(c\right) = c + g\left(c\right)$ and $g\left(c\right)$ is given in (34).

The distribution of the centred and scaled indirect inference estimator ξ_n^{ii} is shown in Fig. 5 against that of the maximum likelihood estimator ξ_n^{ml}. The differences are immediately evident from the figure. The distribution of ξ_n^{ii} is much less biased than ξ_n^{ml}, as we would expect from the criterion function, but it is also much more concentrated that that of the ξ_n^{ml} estimator. Whereas the bulk of the distribution of ξ_n^{ml} is to the left of the origin, the distribution of ξ_n^{ii} leans to the explosive side of the origin while still retaining a long left hand tail. Thus, the functional transformation h^{-1} changes the shape as well as the location of the limit distribution of the ML estimator.

Since the binding function b_n and limit function h are monotonic, tests and confidence intervals based on the IIE $\breve{\rho}$ and the MLE $\hat{\rho}$ are asymptotically equivalent. But, in finite samples there are differences. For example, when $|\rho| < 1$, $\hat{\rho}$ and $\breve{\rho}$ have the same $N\left(\rho, \frac{1-\rho^2}{n}\right)$ asymptotic distribution but tests of $\rho = \rho^0$ and confidence intervals for ρ based on the nominal asymptotics differ. A similar point applies in the case of mildly explosive asymptotics (Phillips and Magdalinos, 2007). Unit root tests based on the test statistics $Z_{\breve{\rho}} = n\left(\breve{\rho} - 1\right)$ and $Z_{\hat{\rho}} = n\left(\hat{\rho} - 1\right)$ are also asymptotically equivalent, as are confidence intervals constructed by inverting these tests using the local to unit limit theory, as in Stock (1991). But finite sample tests and confidence intervals based on the nominal asymptotics differ[2].

5.3 Local to unity case

In this case, the true value is $\rho = 1 + c/n$ and $\xi_n^{ml} = n\left(\hat{\rho}_n - \rho\right) \Rightarrow \xi^{ml} := \int_0^1 J_c dW / \int_0^1 J_c^2$, where W is a standard Brownian motion and $J_c\left(\cdot\right) = \int_0^{\cdot} e^{c(\cdot - s)} dW\left(s\right)$ is a linear diffusion (Phillips, 1987: Chan and Wei, 1987). Since (33) continues to hold for all c we have $\hat{\rho}_n = b_n\left(\breve{\rho}\right)$ with $b_n(\rho) = \rho + \frac{1}{n} g\left(c\right) + O\left(n^{-2}\right)$. Then, setting $\breve{c} = n\left(\breve{\rho} - 1\right)$ we have $\hat{\rho}_n = b_n\left(\breve{\rho}\right) = \breve{\rho} + \frac{1}{n} g\left(\breve{c}\right) + O\left(n^{-2}\right)$ and

$$\begin{aligned} n\left(\hat{\rho}_n - \rho\right) &= n\left(\breve{\rho} - \rho\right) + g\left(n\left(\breve{\rho} - \rho\right) + n\left(\rho - 1\right)\right) + O\left(n^{-1}\right) \\ &= n\left(\breve{\rho} - \rho\right) + g\left(n\left(\breve{\rho} - \rho\right) + c\right) + O\left(n^{-1}\right). \end{aligned} \tag{39}$$

Substituting $\xi_n^{ml} = n\left(\hat{\rho}_n - \rho\right)$ and $\xi_n^{ii} = n\left(\breve{\rho} - \rho\right)$ into (39), we find that

$$\xi_n^{ml} = \xi_n^{ii} + g\left(\xi_n^{ii} + c\right) + O\left(n^{-1}\right),$$

[2] For example, if $f_{L,\alpha}^{ii}$ is the lower α percentile of the limit variate $\xi^{ii} = h^{-1}\left(\int_0^1 W dW / \int_0^1 W^2\right)$, then a one sided nominal 100α% test will reject if $Z_{\breve{\rho}} < f_{L,\alpha}^{ii}$, that is if $\hat{\rho} < b_n\left(1 + f_{L,\alpha}^{ii}/n\right)$ or $Z_{\hat{\rho}} < n\left\{b_n\left(1 + f_{L,\alpha}^{ii}/n\right) - 1\right\} = f_{L,\alpha}^{ii} + g\left(f_{L,\alpha}^{ii}\right) + O\left(n^{-1}\right) = h\left(f_{L,\alpha}^{ii}\right) + O\left(n^{-1}\right) = f_{L,\alpha}^{ml} + O\left(n^{-1}\right)$, where $f_{L,\alpha}^{ml}$ is the lower α percentile of the distribution of the limit variate ξ^{ml}.

and hence

$$\xi_n^{ml} + c = \left(\xi_n^{ii} + c\right) + g\left(\xi_n^{ii} + c\right) + O\left(n^{-1}\right) = h\left(\xi_n^{ii} + c\right) + O\left(n^{-1}\right).$$

Proceeding as in the unit root case, we deduce that

$$n\left(\breve{\rho} - \rho\right) \Rightarrow h^{-1}\left(\int_0^1 J_c dW / \int_0^1 J_c^2 + c\right) - c,$$

so the limit distribution of the indirect inference estimator is given by the inverse of the same implicit mapping h as in the unit root case. Only the intercept $(-c)$ and argument functional $\int_0^1 J_c dW / \int_0^1 J_c^2 + c$ of h^{-1} change according to the value of the localizing coefficient c.

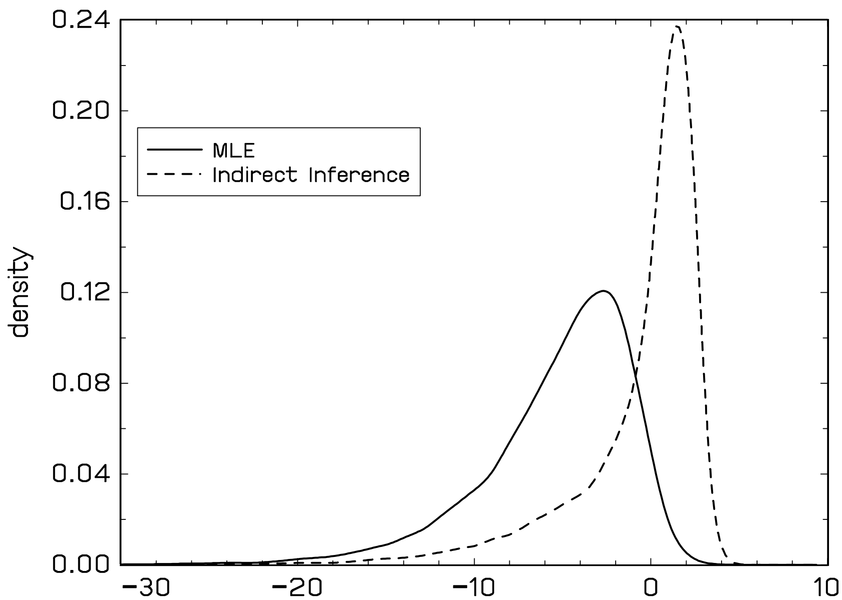

Fig. 5 Densities of $n\left(\xi_n^{ml} - 1\right)$ (solid line) and $n\left(\xi_n^{ii} - 1\right)$ (broken line) for $n = 500$.

6 Conclusions and Extensions

The present work shows how simulation based estimation procedures like indirect inference can complicate limit theory by virtue of the introduction of sample sized dependent functionals into the estimators. These functionals usually serve an important function because of the manner in which they intentionally capture and correct for (possibly undesirable) finite sample features of more basic estimation procedures like maximum likelihood or quasi maximum likelihood. One of the resulting complications is that conventional delta method arguments may fail because of the presence of a sequence of functions rather than some fixed function in the definition of the estimator. Stronger conditions on the sequence of functions are required to validate the standard approach. Another complication is that the estimating function equations may only determine the estimator implicitly, so that it is necessary to work with implicit mappings and global inversion to define the estimator sequence. A final complication and one that is potentially the most significant is that the sequence of functions may influence the limit distribution theory in a material way, affecting the shape characteristics of the distribution as well as simple matters such as location and scale. The indirect inference estimator of the autoregressive coefficient is shown to be affected in this way for values of the coefficient in the usual $O\left(n^{-1}\right)$ vicinity of unity. The resulting limit theory provides both a bias correction and a variance reduction to the maximum likelihood estimator in this vicinity, opening the way to other procedures which have similar properties without compromising the limit theory for the stationary case, such as the fully aggregated estimator of Han, Phillips and Sul (2009).

Given the prolific nature of simulation-based techniques in econometrics, it seems evident that in many cases econometric estimators and inferential procedures will rely on sample-sized based functionals. In such cases, it will generally be necessary to use some version of the extended delta method in asymptotic derivations. These methods are likely to become more numerous in future econometric work as cases of greater complexity are studied using simulation-based methods. Of course, most of these applications will not involve the type of additional difficulties that arise in the limit binding function of the unit root case where nonlinearities in the function persist in the limit and implicit maps are involved. Nonetheless, these additional complexities may arise in some cases of practical importance where simulation-based methods are used in vector time series systems with some unit roots.

The AR(1) case considered here is the prototype for all models with an autoregressive unit root. Practical cases typically involve more variables and parameters. In such cases, it becomes necessary to deal with multivariate asymptotics, possible degeneracies in the limit theory, and the development of binding function algebra for vector autoregressive systems with possible unit roots. Generalization of the extended delta method to multivariate functions and functionals that allow for degeneracies therefore seems worthwhile to accommodate these applications. Similarly, the implicit mapping theorem may be usefully extended to multivariate and functional inverse and implicit function theorems. These may be necessary in dealing with nonstationary time series systems where indirect inference methods are used. More immediate applications of the results here are to dynamic panel data models and continuous time systems where indirect inference methods have been employed to correct bias and to price derivative securities. These extensions and applications seem worthy of consideration in future research.

7 Appendix:

7.1 Some useful integral asymptotic expansions

The following lemmas provide some results on asymptotic expansions of integrals that are useful in the main arguments of the paper. In particular, these results are used to develop bias expansions for $b_n(\rho)$ for three separate fixed ρ cases ($|\rho| < 1$, $\rho = \pm 1$, and $|\rho| > 1$) and to show asymptotic behavior in the local to unity case where $\rho = 1 + \frac{c}{n}$ for some fixed c. These integral asymptotic expansion formulae are likely to have applications in other contexts.

Lemma 10 *Let $F_n = F_n(x;\rho) = 1 - \rho^2 x + (1-x)x^{2n-1}\rho^{2n}$ and suppose $a_1, a_2, \gamma > 0$. Then as $n \to \infty$*

$$\int_0^1 x^{a_1 n + a_4}\left(1-\rho^2 x^2\right)^{\alpha} F_n^{-\beta} dx = \frac{\left(1-\rho^2\right)^{\alpha-\beta}}{a_1 n} + O\left(n^{-2}\right), \quad |\rho| < 1, \tag{40}$$

$$\int_0^1 x^{a_1 n + a_4}(1+x)^{\alpha}\left(1+\gamma x^{a_2 n + a_3}\right)^{\beta} dx = \frac{2^{\alpha}}{a_2 n}\int_0^1 y^{\frac{(a_1-a_2)}{a_2}}(1+\gamma y)^{\beta} dy + O\left(n^{-2}\right), \tag{41}$$

$$\begin{aligned} & n\int_0^1 x^{a_1 n + a_4}(1+x)^{\alpha}\left(1+\gamma x^{a_2 n + a_3}\right)^{\beta}(1-x)\,dx \\ &= -\frac{2^{\alpha}}{a_2^2 n}\int_0^1 y^{\frac{(a_1-a_2)}{a_2}}(1+\gamma y)^{\beta}\log y\,dy + O\left(n^{-2}\right). \end{aligned} \tag{42}$$

Proof of Lemma 10. To prove (40), note first that ρ^{2n} is exponentially small for $|\rho| < 1$. Then, $F_n = 1 - \rho^2 x + O\left(\rho^{2n}\right)$. Setting $y = x^{a_1 n + a_4}$, we have $dy = (a_1 n + a_4)x^{a_1 n + a_4 - 1}dx = (a_1 n + a_4)\,y^{\frac{a_1 n + a_4 - 1}{a_1 n + a_4}}dx$ and upon transformation

$$\begin{aligned} & \int_0^1 x^{a_1 n + a_4}\left(1-\rho^2 x^2\right)^{\alpha} F_n^{-\beta} dx \\ &= \frac{1}{a_1 n + a_4}\int_0^1 y^{1-\frac{a_1 n + a_4 - 1}{a_1 n + a_4}}\left(1-\rho^2 y^{\frac{2}{a_1 n + a_4}}\right)^{\alpha}\left(1-\rho^2 y^{\frac{1}{a_1 n + a_4}}\right)^{-\beta} dy \\ &= \frac{\left(1-\rho^2\right)^{\alpha-\beta}}{a_1 n + a_4}\int_0^1 y^{\frac{1}{a_1 n + a_4}} dy\left\{1 + O\left(n^{-1}\right)\right\} \\ &= \frac{\left(1-\rho^2\right)^{\alpha-\beta}}{a_1 n} + O\left(n^{-2}\right), \end{aligned}$$

since $y^{\frac{b}{a_1 n + a_4}} = 1 + \frac{b}{a_1 n + a_4}\log y + O\left(n^{-2}\right)$ for all $b \neq 0$ and $\left|\int_0^1 y^a \log y\,dy\right| < \infty$ for all $a \geq -1$.

To prove (41), set $y = x^{a_2 n + a_3}$, so that $dy = (a_2 n + a_3)x^{a_2 n + a_3 - 1}dx = a_2 n y^{\frac{a_2 n + a_3 - 1}{a_2 n + a_3}}dx$ and upon transformation

$$\begin{aligned} \int_0^1 x^{a_1 n + a_4}(1+x)^{\alpha}\left(1+\gamma x^{a_2 n + a_3}\right)^{\beta} dx &= \frac{1}{a_2 n + a_3}\int_0^1 y^{\frac{(a_1-a_2)n + a_4 - a_3 + 1}{a_2 n + a_3}}\left(1 + y^{\frac{1}{a_2 n + a_3}}\right)^{\alpha}(1+\gamma y)^{\beta} dy \\ &= \frac{2^{\alpha}}{a_2 n}\int_0^1 y^{\frac{(a_1-a_2)}{a_2}}(1+\gamma y)^{\beta} dy + O\left(n^{-2}\right), \end{aligned}$$

since $y^{\frac{b}{a_2 n+a_3}} = 1 + \frac{b}{a_2 n+a_3}\log y + O\left(n^{-2}\right)$ for all $b \neq 0$ and $\int_0^1 y^a (1+y)^\beta |\log y|\, dy < \infty$ for all $a > 0$.

To prove (42), the same approach leads to

$$\begin{aligned}
& n\int_0^1 x^{a_1 n+a_4}(1+x)^\alpha \left(1+\gamma x^{a_2 n+a_3}\right)^\beta (1-x)\, dx \\
&= \frac{n}{a_2 n+a_3}\int_0^1 y^{\frac{(a_1-a_2)n+a_4-a_3+1}{a_2 n+a_3}} \left(1+y^{\frac{1}{a_2 n+a_3}}\right)^\alpha (1+\gamma y)^\beta \left(1-y^{\frac{1}{a_2 n+a_3}}\right) dy \\
&= -\frac{2^\alpha}{a_2^2 n}\int_0^1 y^{\frac{(a_1-a_2)}{a_2}} (1+\gamma y)^\beta \log y dy + O\left(n^{-2}\right). \qquad (43)
\end{aligned}$$

Observe that, using the transformation $w = -\log y$, we have

$$\int_0^1 y^{a-1} |\log y|^b\, dy = \int_0^\infty e^{-aw} w^b dw < \infty \qquad (44)$$

for all $a > 0$ and $b > -1$, which ensures that (43) is finite. ■

Lemma 11 *For $\rho = 1 + \frac{c}{n}$, with c fixed, $F_n = 1 - \rho^2 x + (1-x)\, x^{2n-1}\rho^{2n}$, $a_1 > 0$, and $\alpha - \beta > -1$, we have*

$$\begin{aligned}
& \int_0^1 x^{a_1 n+a_4}\left(1-\rho^2 x^2\right)^\alpha F_n^{-\beta} dx \\
&= \begin{cases} \frac{2^\alpha}{2n}\int_0^1 y^{\frac{(a_1-2)}{2}} \left(1+e^{2c}y\right)^{-\beta} dy + O\left(n^{-2}\right) & \alpha = \beta \\ \frac{2^\alpha}{4n}\int_0^1 y^{\frac{(a_1-2)}{2}} \left(1+e^{2c}y\right)^{-\beta} (-\log y)^{\alpha-\beta} dy + O\left(n^{-2}\right) & \alpha \neq \beta \end{cases}.
\end{aligned}$$

Proof of Lemma 11. Since $\rho^{2n} = \left(1+\frac{c}{n}\right)^{2n} = e^{2c}\left\{1 + O\left(n^{-1}\right)\right\}$ we have

$$F_n(x,\rho) = 1 - x + (1-x)\, x^{2n-1} e^{2c} + O\left(n^{-1}\right) = (1-x)\left(1+e^{2c}x^{2n-1}\right) + O\left(n^{-1}\right). \qquad (45)$$

Using Lemma 10 we obtain

$$\begin{aligned}
& \int_0^1 x^{a_1 n+a_4}\left(1-\rho^2 x^2\right)^\alpha F_n^{-\beta} dx \\
&= \int_0^1 x^{a_1 n+a_4}\left(1-x^2\right)^\alpha (1-x)^{-\beta}\left(1+e^{2c}x^{2n-1}\right)^{-\beta} dx \left\{1+O\left(n^{-1}\right)\right\} \\
&= \int_0^1 x^{a_1 n+a_4}(1+x)^\alpha (1-x)^{\alpha-\beta}\left(1+e^{2c}x^{2n-1}\right)^{-\beta} dx \left\{1+O\left(n^{-1}\right)\right\} \\
&= \begin{cases} \int_0^1 x^{a_1 n+a_4}(1+x)^\alpha \left(1+e^{2c}x^{2n-1}\right)^{-\beta} dx \left\{1+O\left(n^{-1}\right)\right\} & \alpha = \beta \\ \int_0^1 x^{a_1 n+a_4}(1+x)^\alpha (1-x)^{\alpha-\beta}\left(1+e^{2c}x^{2n-1}\right)^{-\beta} dx \left\{1+O\left(n^{-1}\right)\right\} & \alpha \neq \beta \end{cases} \\
&= \begin{cases} \frac{2^\alpha}{2n}\int_0^1 y^{\frac{(a_1-2)}{2}} \left(1+e^{2c}y\right)^{-\beta} dy + O\left(n^{-2}\right) & \alpha = \beta \\ \frac{2^\alpha}{4n}\int_0^1 y^{\frac{(a_1-2)}{2}} \left(1+e^{2c}y\right)^{-\beta} (-\log y)^{\alpha-\beta} dy + O\left(n^{-2}\right) & \alpha \neq \beta \end{cases},
\end{aligned}$$

the final integral being finite in view of (44) when $\alpha - \beta > -1$. ■

Lemma 12 *If $a_1 > 0$, then as $n \to \infty$*

$$\int_0^1 x^{a_1 n + a_2} \left(1 - \rho^2 x^2\right)^{\alpha} \left(1 - \rho^2 x\right)^{-\beta} dx = \frac{\left(1 - \rho^2\right)^{\alpha - \beta}}{a_1 n} + O\left(n^{-2}\right).$$

Proof of Lemma 12. Integrating by parts we have

$$\begin{aligned}
&\int_0^1 x^{a_1 n + a_2} \left(1 - \rho^2 x^2\right)^{\alpha} \left(1 - \rho^2 x\right)^{-\beta} dx \\
&= \left[\frac{x^{a_1 n + a_2 + 1}}{a_1 n + a_2 + 1} \left(1 - \rho^2 x^2\right)^{\alpha} \left(1 - \rho^2 x\right)^{-\beta}\right]_0^1 \\
&+ \frac{2\alpha\rho^2}{a_1 n + a_2 + 1} \int_0^1 x^{a_1 n + a_2 + 2} \left(1 - \rho^2 x^2\right)^{\alpha - 1} \left(1 - \rho^2 x\right)^{-\beta} dx \\
&- \frac{\beta\rho^2}{a_1 n + a_2 + 1} \int_0^1 x^{a_1 n + a_2 + 2} \left(1 - \rho^2 x^2\right)^{\alpha - 1} \left(1 - \rho^2 x\right)^{-\beta - 1} dx \\
&= \frac{\left(1 - \rho^2\right)^{\alpha - \beta}}{a_1 n} + O\left(n^{-2}\right).
\end{aligned}$$

■

7.2 Proofs of the main results

Proof of Theorem 1. By the mean value theorem

$$\varphi_n\left(T_n\right) - \varphi_n\left(\theta\right) = \varphi_n'\left(T_n^*\right)\left(T_n - \theta\right),$$

for some T_n^* on the line segment connecting T_n and θ. Hence

$$\frac{d_n}{\varphi_n'\left(\theta\right)}\left(\varphi_n\left(T_n\right) - \varphi_n\left(\theta\right)\right) = \left\{1 + \frac{\varphi_n'\left(T_n^*\right) - \varphi_n'\left(\theta\right)}{\varphi_n'\left(\theta\right)}\right\} d_n\left(T_n - \theta\right).$$

Since $\left|T_n^* - \theta\right| \leq \left|T_n - \theta\right| = O_p\left(d_n^{-1}\right)$ and $\frac{s_n}{d_n} \to 0$, it follows that $s_n\left|T_n^* - \theta\right| = o_p(1)$. Then

$$\frac{\varphi_n'\left(T_n^*\right) - \varphi_n'\left(\theta\right)}{\varphi_n'\left(\theta\right)} \to_p 0$$

by local relative equicontinuity (2) in a shrinking neighborhood of radius $O\left(s_n^{-1}\right)$, giving the required result. ■

Proof of Lemma5. See Ge and Wang (2002, Lemma 1). ■

Proof of Theorem 7. As indicated, the structure of the proof follows SJ (1965) and SV (1996) by considering ratios of quadratic forms in normal variates. The starting point is to write the density and moments of $\hat{\rho}_n = \sum_{t=1}^n y_t y_{t-1} / \sum_{t=1}^n y_{t-1}^2 = U/V$ in terms of the joint moment generating function $m(u,q)$ of the quadratic forms (U,V). White (1958) showed that $m(u,q) = D_n^{-1/2}$ where $D_n = D_n(u,q)$ is a determinant that satisfies the second order difference equation

$$D_n = \left(1 + \rho^2 + 2q\right) D_{n-1} - (\rho + u)^2 D_{n-2}, \qquad D_0 = D_1 = 1.$$

Then by direct calculation (see SJ p.3) we have the following expression for the bias function

$$E(\hat{\rho}_n - \rho) = \int_0^\infty \frac{\partial}{\partial \rho} D_n\left(q\right)^{-1/2} dq, \tag{46}$$

where the determinant $D_n\left(q\right) = D_n\left(0, q\right)$ is evaluated explicitly as

$$D_n\left(q\right) = A\theta^n + \left(1 - A\right)\rho^{2n}\theta^{-n}, \quad A = \frac{\theta - \rho^2}{\theta^2 - \rho^2}, \tag{47}$$

$$\theta = \theta\left(q\right) = \left(1 + \rho^2 + 2q + \sqrt{\Delta}\right)/2,$$

$$\Delta = \left(1 + \rho^2 + 2q\right)^2 - 4\rho^2. \tag{48}$$

Observe that the following inequalities hold

$$\begin{aligned}
\theta = \theta\left(q\right) &= \left(1 + \rho^2 + 2q + \sqrt{\Delta}\right)/2 \geq 0, \\
\Delta &= \left(1 - \rho^2\right)^2 + 4q^2 + 4q\left(1 + \rho^2\right) \geq \left(1 - \rho^2\right)^2 \geq 0, \\
\theta - \rho^2 &= \left(1 - \rho^2 + 2q + \sqrt{\Delta}\right)/2 \geq q \geq 0, \\
\theta - \rho &= \left(1 - \rho\right)^2 + 2q + \sqrt{\Delta} \geq 0, \\
\theta + \rho &= \left(1 + \rho\right)^2 + 2q + \sqrt{\Delta} \geq 0, \\
\theta^2 - \rho^2 &= \left(\theta - \rho\right)\left(\theta + \rho\right) \geq 0
\end{aligned}$$

for all $q \geq 0$. It follows that the determinant (47) is positive for all $q > 0$ and the integral (46) is defined for all ρ.

Write the binding function as

$$b_n(\rho) = E(\hat{\rho}_n) = \rho + \int_0^\infty \frac{\partial}{\partial \rho} D_n\left(q\right)^{-1/2} dq = \rho - \frac{1}{2}\int_0^\infty D_n\left(q\right)^{-3/2} \frac{\partial D_n\left(q\right)}{\partial \rho} dq.$$

Define $x = 1/\theta$ and $C = 1 + \rho^2 + 2q$, so that

$$x = \frac{2}{1 + \rho^2 + 2q + \sqrt{\Delta}} = \frac{2}{C + \sqrt{\Delta}} = 2\frac{C - \sqrt{\Delta}}{C^2 - \Delta} = \frac{C - \sqrt{\Delta}}{2\rho^2}, \tag{49}$$

since $\Delta = \left(1 + \rho^2 + 2q\right)^2 - 4\rho^2 = C^2 - 4\rho^2$. It follows from (49) that

$$C + \Delta^{1/2} = \frac{2}{x} \text{ and } C - \Delta^{1/2} = 2\rho^2 x,$$

so that $C = 1/x + \rho^2 x$, leading to

$$q = \frac{1}{2}\left(C - 1 - \rho^2\right) = \frac{1}{2}\left(1/x + \rho^2 x - 1 - \rho^2\right) = \frac{\left(1 - x\right)\left(1 - \rho^2 x\right)}{2x}.$$

We write

$$q = \frac{\left(1 - x\right)\left(1 - \rho^2 x\right)}{2x} = \begin{cases} \frac{(1-x)\left(1-\rho^2 x\right)}{2x} & x \in (0, 1], \quad |\rho| \leq 1 \\ \frac{(x-1)\left(x\rho^2 - 1\right)}{2x} & x \in [1, \infty), \quad |\rho| > 1 \end{cases},$$

with derivative

$$\frac{dq}{dx} = -\frac{\left(1-\rho^2 x^2\right)}{2x^2} \quad \begin{cases} <0 & x\in(0,1], \text{ for } |\rho|\le 1 \\ >0 & x\in[1,\infty), \text{ for } |\rho|>1 \end{cases}, \tag{50}$$

so $q = q(x)$ is monotonic over the two domains of x in each case with $q \in [0,\infty)$. We may therefore change the variable of integration in (46) from q to x with corresponding changes in the domain of integration depending on the value of ρ as specified in (50). For $\rho = 1$, either domain may be used.

Using this change of variable, we have

$$\begin{aligned} A &= \frac{\theta-\rho^2}{\theta^2-\rho^2} = \frac{\frac{1}{x}-\rho^2}{\frac{1}{x^2}-\rho^2} = \frac{x-\rho^2x^2}{1-\rho^2x^2} = \frac{x\left(1-\rho^2 x\right)}{1-\rho^2x^2}, \\ 1-A &= 1-\frac{x-\rho^2x^2}{1-\rho^2x^2} = \frac{1-x}{1-\rho^2x^2}, \end{aligned}$$

and then

$$\begin{aligned} D_n(q) &= \frac{\left(1-\rho^2x\right)}{1-\rho^2x^2}\frac{1}{x^{n-1}} + \frac{1-x}{1-\rho^2x^2}\rho^{2n}x^n \\ &= \begin{cases} \frac{1-\rho^2x+(1-x)x^{2n-1}\rho^{2n}}{(1-\rho^2x^2)x^{n-1}} & = \frac{F_n(x;\rho)}{(1-\rho^2x^2)x^{n-1}}, \quad |\rho|\le 1 \\ \frac{\rho^2x-1+(x-1)x^{2n-1}\rho^{2n}}{(\rho^2x^2-1)x^{n-1}} & = \frac{G_n(x;\rho)}{(\rho^2x^2-1)x^{n-1}}, \quad |\rho|>1 \end{cases}, \end{aligned}$$

where

$$\begin{aligned} F_n(x;\rho) &:= 1-\rho^2x+(1-x)\,x^{2n-1}\rho^{2n}, \\ G_n(x;\rho) &:= \rho^2x-1+(x-1)\,x^{2n-1}\rho^{2n}. \end{aligned}$$

For $|\rho|\le 1$, we have

$$\begin{aligned} E(\hat\rho_n-\rho) &= \frac{\partial}{\partial\rho}\int_0^\infty D_n(q)^{-1/2}\,dq, \\ &= \frac{\partial}{\partial\rho}\int_0^1 \left(\frac{F_n(x;\rho)}{(1-\rho^2x^2)\,x^{n-1}}\right)^{-1/2}\frac{\left(1-\rho^2x^2\right)}{2x^2}dx \\ &= \frac{1}{2}\frac{\partial}{\partial\rho}\left\{\int_0^1 x^{(n-5)/2}\left(1-\rho^2x^2\right)^{3/2}F_n(x;\rho)^{-1/2}\,dx\right\}. \end{aligned} \tag{51}$$

To evaluate (51), note that

$$\begin{aligned} &\frac{\partial}{\partial\rho}\left\{\int_0^1 x^{(n-5)/2}\left(1-\rho^2x^2\right)^{3/2}F_n^{-1/2}dx\right\} \\ &= \frac{3}{2}\int_0^1 x^{(n-5)/2}\left(-2\rho x^2\right)\left(1-\rho^2x^2\right)^{1/2}F_n^{-1/2}dx \\ &-\frac{1}{2}\int_0^1 x^{(n-5)/2}\left(1-\rho^2x^2\right)^{3/2}F_n^{-3/2}\left\{-2\rho x+2n(1-x)\,x^{2n-1}\rho^{2n-1}\right\}dx, \end{aligned}$$

so that for $|\rho| \leq 1$ we have

$$
\begin{aligned}
b_n(\rho) &= \rho + \frac{1}{2}\frac{\partial}{\partial\rho}\left\{\int_0^1 x^{(n-5)/2}\left(1-\rho^2x^2\right)^{3/2} F_n^{-1/2}(x;\rho)\,dx\right\} \\
&= \rho + \frac{3}{4}\int_0^1 x^{(n-5)/2}\left(-2\rho x^2\right)\left(1-\rho^2x^2\right)^{1/2} F_n^{-1/2}dx \\
&- \frac{1}{4}\int_0^1 x^{(n-5)/2}\left(1-\rho^2x^2\right)^{3/2} F_n^{-3/2}\left\{-2\rho x + 2n(1-x)x^{2n-1}\rho^{2n-1}\right\}dx \\
&= \rho - \frac{3\rho}{2}\int_0^1 x^{(n-1)/2}\left(1-\rho^2x^2\right)^{1/2} F_n^{-1/2}dx \\
&+ \frac{\rho}{2}\int_0^1 x^{(n-3)/2}\left(1-\rho^2x^2\right)^{3/2} F_n^{-3/2}dx - \frac{n\rho^{2n-1}}{2}\int_0^1 x^{(5n-7)/2}\left(1-\rho^2x^2\right)^{3/2} F_n^{-3/2}(1-x)\,dx.
\end{aligned} \tag{52}
$$

For $|\rho| > 1$, we have

$$
\begin{aligned}
E(\hat{\rho}_n - \rho) &= \frac{\partial}{\partial\rho}\int_0^\infty D_n(q)^{-1/2}\,dq, \\
&= \frac{\partial}{\partial\rho}\int_1^\infty \left(\frac{G_n(x;\rho)}{\left(\rho^2x^2-1\right)x^{n-1}}\right)^{-1/2}\frac{\left(\rho^2x^2-1\right)}{2x^2}dx \\
&= \frac{1}{2}\frac{\partial}{\partial\rho}\int_1^\infty x^{(n-5)/2}\left(\rho^2x^2-1\right)^{3/2} G_n(x:\rho)^{-1/2}\,dx,
\end{aligned} \tag{53}
$$

and by direct evaluation

$$
\begin{aligned}
&\frac{\partial}{\partial\rho}\left\{\int_1^\infty x^{(n-5)/2}\left(\rho^2x^2-1\right)^{3/2} G_n^{-1/2}dx\right\} = \frac{3}{2}\int_1^\infty x^{(n-5)/2}\left(2\rho x^2\right)\left(\rho^2x^2-1\right)^{1/2} G_n^{-1/2}dx \\
&- \frac{1}{2}\int_1^\infty x^{(n-5)/2}\left(\rho^2x^2-1\right)^{3/2} G_n^{-3/2}\left\{2\rho x + 2n(x-1)x^{2n-1}\rho^{2n-1}\right\}dx \\
&= \frac{3}{2}\int_1^\infty x^{(n-5)/2}\left(2\rho x^2\right)\left(\rho^2x^2-1\right)^{1/2} G_n^{-1/2}dx - \rho\int_1^\infty x^{(n-3)/2}\left(\rho^2x^2-1\right)^{3/2} G_n^{-3/2}dx \\
&- n\rho^{2n-1}\int_1^\infty x^{(n-5)/2}\left(\rho^2x^2-1\right)^{3/2} G_n^{-3/2}(x-1)x^{2n-1}dx.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
b_n(\rho) &= \rho + \frac{1}{2}\frac{\partial}{\partial\rho}\left\{\int_1^\infty x^{(n-5)/2}\left(\rho^2x^2-1\right)^{3/2} G_n^{-1/2}dx\right\} \\
&= \rho + \frac{3}{4}\int_1^\infty x^{(n-5)/2}\left(2\rho x^2\right)\left(\rho^2x^2-1\right)^{1/2} G_n^{-1/2}dx \\
&- \frac{\rho}{2}\int_1^\infty x^{(n-3)/2}\left(\rho^2x^2-1\right)^{3/2} G_n^{-3/2}dx \\
&- \frac{n\rho^{2n-1}}{2}\int_1^\infty x^{(n-5)/2}\left(\rho^2x^2-1\right)^{3/2} G_n^{-3/2}(x-1)x^{2n-1}dx.
\end{aligned}
$$

Hence, the binding formula for $|\rho| > 1$ is

$$b_n(\rho) = \rho + \frac{3\rho}{2}\int_1^\infty x^{(n-1)/2}\left(\rho^2x^2-1\right)^{1/2}G_n^{-1/2}dx$$
$$-\frac{\rho}{2}\int_1^\infty x^{(n-3)/2}\left(\rho^2x^2-1\right)^{3/2}G_n^{-3/2}dx$$
$$-\frac{n\rho^{2n-1}}{2}\int_1^\infty x^{(5n-7)/2}\left(\rho^2x^2-1\right)^{3/2}G_n^{-3/2}(x-1)\,dx \tag{54}$$

Transforming using $y = 1/x$, and noting that

$$G_n\left(\frac{1}{y};\rho\right) = \rho^2\frac{1}{y} - 1 + \left(\frac{1}{y}-1\right)y^{-2n+1}\rho^{2n}$$
$$= \frac{\left(\rho^2-y\right)y^{2n-1} + (1-y)\,\rho^{2n}}{y^{2n}} =: \frac{H_n(y;\rho)}{y^{2n}},$$

we have the alternate form

$$b_n(\rho) = \rho + \frac{3\rho}{2}\int_0^1 y^{-(n-1)/2}\frac{\left(\rho^2-y^2\right)^{1/2}}{y}H_n^{-1/2}(y;\rho)\,y^{n-2}dy$$
$$-\frac{\rho}{2}\int_0^1 y^{-(n-3)/2}\frac{\left(\rho^2-y^2\right)^{3/2}}{y^3}H_n^{-3/2}(y;\rho)\,y^{3n-2}dy$$
$$-\frac{n\rho^{2n-1}}{2}\int_0^1 y^{-(5n-7)/2}\frac{\left(\rho^2-y^2\right)^{3/2}}{y^3}H_n^{-3/2}(y;\rho)\,y^{3n-3}(1-y)\,dy$$
$$= \rho + \frac{3\rho}{2}\int_0^1 y^{(n-5)/2}\left(\rho^2-y^2\right)^{1/2}H_n^{-1/2}(y;\rho)\,dy$$
$$-\frac{\rho}{2}\int_0^1 y^{(5n-7)/2}\left(\rho^2-y^2\right)^{3/2}H_n^{-3/2}(y;\rho)\,dy$$
$$-\frac{n\rho^{2n-1}}{2}\int_0^1 y^{(n-5)/2}\left(\rho^2-y^2\right)^{3/2}H_n^{-3/2}(y;\rho)(1-y)\,dx \tag{55}$$

■

Proof of Theorem 8. (i) Case $|\rho| < 1$

Using $F_n = 1-\rho^2x+(1-x)\,x^{2n-1}\rho^{2n} = 1-\rho^2x+O\left(\rho^{2n}\right)$ and $n\rho^n = o\left(n^{-2}\right)$ we have for $|\rho| < 1$ from (52) and Lemma 12

$$b_n(\rho) = \rho - \frac{3\rho}{2}\int_0^1 x^{(n-1)/2}\left(1-\rho^2x^2\right)^{1/2}F_n^{-1/2}dx + \frac{\rho}{2}\int_0^1 x^{(n-3)/2}\left(1-\rho^2x^2\right)^{3/2}F_n^{-3/2} + o\left(n^{-1}\right)$$
$$= \rho - \frac{3\rho}{2}\int_0^1 x^{(n-1)/2}\left(1-\rho^2x^2\right)^{1/2}\left(1-\rho^2x\right)^{-1/2}dx$$
$$+\frac{\rho}{2}\int_0^1 x^{(n-3)/2}\left(1-\rho^2x^2\right)^{3/2}\left(1-\rho^2x\right)^{-3/2}dx + o\left(n^{-2}\right)$$
$$= \rho - \frac{2\rho}{n} + O\left(n^{-2}\right),$$

giving the well-known asymptotic bias formula for $\hat{\rho}_n$ in the stationary case.

(ii) Case $\rho = \pm 1$

When $\rho = 1$, we have $F_n(x;1) = 1 - x + (1-x)x^{2n-1} = (1-x)\left(1 + x^{2n-1}\right)$ and so from (52)

$$\begin{aligned} b_n(1) &= 1 - \frac{3}{2}\int_0^1 x^{(n-1)/2}\frac{(1+x)^{1/2}}{(1+x^{2n-1})^{1/2}}dx + \frac{1}{2}\int_0^1 x^{(n-3)/2}\frac{(1+x)^{3/2}}{(1+x^{2n-1})^{3/2}}dx \\ &\quad - \frac{n}{2}\int_0^1 x^{(5n-7)/2}\frac{(1+x)^{3/2}}{(1+x^{2n-1})^{3/2}}(1-x)\,dx. \end{aligned}$$

Using (41) and (42), we have

$$\int_0^1 x^{(n-1)/2}\frac{(1+x)^{1/2}}{(1+x^{2n-1})^{1/2}}dx = \frac{2^{1/2}}{2n}\int_0^1 y^{-\frac{3}{4}}(1+y)^{-1/2}dy + O\left(n^{-2}\right),$$

$$\int_0^1 x^{(n-3)/2}\frac{(1+x)^{3/2}}{(1+x^{2n-1})^{3/2}}dx = \frac{2^{3/2}}{2n}\int_0^1 y^{-\frac{3}{4}}(1+y)^{-3/2}dy + O\left(n^{-2}\right),$$

$$n\int_0^1 x^{(n-1)/2}\frac{(1+x)^{3/2}}{(1+x^{2n-1})^{3/2}}(1-x)\,x^{2n-3}dx = -\frac{2^{3/2}}{4n}\int_0^1 y^{\frac{1}{4}}(1+y)^{-3/2}\log y dy + O\left(n^{-2}\right).$$

It follows that

$$\begin{aligned} b_n(1) &= 1 - 3\frac{2^{1/2}}{4n}\int_0^1 y^{-\frac{3}{4}}(1+y)^{-1/2}dy + \frac{2^{3/2}}{4n}\int_0^1 y^{-\frac{3}{4}}(1+y)^{-3/2}dy \\ &\quad - \frac{1}{2}\left\{-\frac{2^{3/2}}{4n}\int_0^1 y^{\frac{1}{4}}(1+y)^{-3/2}\log y dy\right\} + O\left(n^{-2}\right) \end{aligned} \tag{56}$$

and numerical evaluation of the integrals gives

$$\begin{aligned} b_n(1) &= 1 - \frac{3}{4n}2^{0.5}(3.7081) + \frac{2^{1.5}}{4n}(3.2683) - \frac{2^{1/2}}{4n}(0.45077) + O\left(n^{-2}\right) \\ &= 1 - \frac{1.7814}{n} + O\left(n^{-2}\right), \end{aligned} \tag{57}$$

corresponding to the result found by SJ (1965) for the unit root case using different methods. The numerical value -1.7814 is the mean of the limit distribution of $n(\hat{\rho}_n - 1)$ when $\rho = 1$.

Similar calculations apply when $\rho = -1$, in which case we have

$$\begin{aligned} b_n(-1) &= -1 + \frac{3}{4n}2^{0.5}(3.7081) - \frac{2^{1.5}}{4n}(3.2683) + \frac{2^{1/2}}{4n}(0.45077) + O\left(n^{-2}\right) \\ &= -1 + \frac{1.7814}{n} + O\left(n^{-2}\right), \end{aligned} \tag{58}$$

giving the mirror image of (57).

(iii) Case $\rho = 1 + \frac{c}{n}$, $c < 0$

Next consider the local to unity case with $\rho = 1 + \frac{c}{n}$ and $c < 0$. The relevant expression for the bias is

$$b_n(\rho) = \rho - \frac{3\rho}{2}\int_0^1 x^{(n-1)/2}\left(1-\rho^2x^2\right)^{1/2} F_n^{-1/2}dx + \frac{\rho}{2}\int_0^1 x^{(n-3)/2}\left(1-\rho^2x^2\right)^{3/2} F_n^{-3/2}dx$$
$$- \frac{n\rho^{2n-1}}{2}\int_0^1 x^{(5n-7)/2}\left(1-\rho^2x^2\right)^{3/2} F_n^{-3/2}(1-x)\,dx. \tag{59}$$

As before, set $y = x^{2n-1}$ so that $dy = (2n-1)\,x^{2n-2}dx = (2n-1)\,y^{\frac{2n-2}{2n-1}}dx = (2n-1)\,y^{1-\frac{1}{2n-1}}dx$. Then, using $F_n = 1-\rho^2x + (1-x)\,x^{2n-1}\rho^{2n}$, we have for the first integral in (59)

$$\begin{aligned}
&\int_0^1 x^{(n-1)/2}\left(1-\rho^2x^2\right)^{1/2} F_n^{-1/2}dx \\
&= \frac{1}{2n-1}\int_0^1 y^{\frac{n+1}{4n-4}-1}\left(1-\rho^2y^{\frac{2}{2n-1}}\right)^{1/2}\left\{\left(1-\rho^2y^{\frac{1}{2n-1}}\right)+\left(1-y^{\frac{1}{2n-1}}\right)y\rho^{2n}\right\}^{-1/2}dy \\
&= \frac{1}{2n-1}\int_0^1 y^{-\frac{3}{4}+\frac{1}{2n-2}}\left(1-\rho^2y^{\frac{2}{2n-1}}\right)^{1/2}\left\{\left(1-\rho^2y^{\frac{1}{2n-1}}\right)+\left(1-y^{\frac{1}{2n-1}}\right)y\rho^{2n}\right\}^{-1/2}dy.
\end{aligned}$$

Since $y^{\frac{b}{2n+a}} = 1 + \frac{b}{2n+a}\log y + O\left(n^{-2}\right)$ and $\rho^{2n} = \left(1+\frac{c}{n}\right)^{2n} = e^{2c}\left\{1+O\left(n^{-1}\right)\right\}$, it follows that

$$\begin{aligned}
&\int_0^1 x^{(n-1)/2}\left(1-\rho^2x^2\right)^{1/2} F_n^{-1/2}dx \\
&= \frac{1}{2n-1}\int_0^1 y^{-\frac{3}{4}}\left(1-\rho^2y^{\frac{2}{2n-1}}\right)^{1/2}\left\{\left(1-\rho^2y^{\frac{1}{2n-1}}\right)+\left(1-y^{\frac{1}{2n-1}}\right)y\rho^{2n}\right\}^{-1/2}dy\left\{1+O\left(n^{-1}\right)\right\} \\
&= \frac{1}{2n}\int_0^1 y^{-\frac{3}{4}}\left\{\frac{1-\rho^2y^{\frac{1}{2n-1}}}{1-\rho^2y^{\frac{2}{2n-1}}}+\frac{1-y^{\frac{1}{2n-1}}}{1-\rho^2y^{\frac{2}{2n-1}}}y\rho^{2n}\right\}^{-1/2}dy + O\left(n^{-2}\right) \\
&= \frac{1}{2n}\int_0^1 y^{-\frac{3}{4}}\left\{\frac{1-\left(1+\frac{2c}{n}\right)\left(1+\frac{1}{2n}\log y\right)}{1-\left(1+\frac{2c}{n}\right)\left(1+\frac{1}{n}\log y\right)}-\frac{\frac{1}{2n}\log y}{1-\left(1+\frac{2c}{n}\right)\left(1+\frac{1}{n}\log y\right)}y\rho^{2n}\right\}^{-1/2}dy + O\left(n^{-2}\right) \\
&= \frac{1}{2n}\int_0^1 y^{-\frac{3}{4}}\left\{\frac{-\frac{2c}{n}-\frac{1}{2n}\log y}{-\frac{2c}{n}-\frac{1}{n}\log y}-\frac{\frac{1}{2n}\log y}{-\frac{2c}{n}-\frac{1}{n}\log y}ye^{2c}\right\}^{-1/2}dy + O\left(n^{-2}\right) \\
&= \frac{1}{2n}\int_0^1 y^{-\frac{3}{4}}\left\{\frac{4c+\log y}{4c+2\log y}+\frac{\log y}{4c+2\log y}ye^{2c}\right\}^{-1/2}dy + O\left(n^{-2}\right).
\end{aligned} \tag{60}$$

The second integral in (59) may be reduced in the same way. Again, setting $y = x^{2n-1}$, with $dy = (2n-1)\, y^{1-\frac{1}{2n-1}} dx$, and $y^{\frac{b}{2n+a}} = 1 + \frac{b}{2n+a} \log y + O\left(n^{-2}\right)$, we have

$$
\begin{aligned}
&\int_0^1 x^{(n-3)/2} \left(1-\rho^2 x^2\right)^{3/2} F_n^{-3/2} dx \\
&= \frac{1}{2n-1} \int_0^1 y^{-\frac{3}{4}-\frac{1/4}{2n-2}} \left(1-\rho^2 y^{\frac{2}{2n-1}}\right)^{3/2} \left\{ \left(1-\rho^2 y^{\frac{1}{2n-1}}\right) + \left(1-y^{\frac{1}{2n-1}}\right) y\rho^{2n} \right\}^{-3/2} dy \\
&= \frac{1}{2n} \int_0^1 y^{-\frac{3}{4}} \left\{ \frac{1-\rho^2 y^{\frac{1}{2n-1}}}{1-\rho^2 y^{\frac{2}{2n-1}}} + \frac{1-y^{\frac{1}{2n-1}}}{1-\rho^2 y^{\frac{2}{2n-1}}} y\rho^{2n} \right\}^{-3/2} dy + O\left(n^{-2}\right) \\
&= \frac{1}{2n} \int_0^1 y^{-\frac{3}{4}} \left\{ \frac{1-\left(1+\frac{2c}{n}\right)\left(1+\frac{1}{2n}\log y\right)}{1-\left(1+\frac{2c}{n}\right)\left(1+\frac{1}{n}\log y\right)} - \frac{\frac{1}{2n}\log y}{1-\left(1+\frac{2c}{n}\right)\left(1+\frac{1}{n}\log y\right)} y\rho^{2n} \right\}^{-3/2} dy + O\left(n^{-2}\right) \\
&= \frac{1}{2n} \int_0^1 y^{-\frac{3}{4}} \left\{ \frac{-\frac{2c}{n}-\frac{1}{2n}\log y}{-\frac{2c}{n}-\frac{1}{n}\log y} - \frac{\frac{1}{2n}\log y}{-\frac{2c}{n}-\frac{1}{n}\log y} ye^{2c} \right\}^{-3/2} dy + O\left(n^{-2}\right) \\
&= \frac{1}{2n} \int_0^1 y^{-\frac{3}{4}} \left\{ \frac{4c+\log y}{4c+2\log y} + \frac{\log y}{4c+2\log y} ye^{2c} \right\}^{-3/2} dy + O\left(n^{-2}\right). \qquad (61)
\end{aligned}
$$

Finally, for the third integral in (59) we have in the same fashion

$$
\begin{aligned}
&\int_0^1 x^{(5n-7)/2} \left(1-\rho^2 x^2\right)^{3/2} F_n^{-3/2} (1-x)\, dx \\
&= \frac{1}{(2n-1)^2} \int_0^1 y^{\frac{(5n-7)}{4n-2}-1+\frac{1}{2n-1}} \left(1-\rho^2 y^{\frac{2}{2n-1}}\right)^{3/2} \left\{ \left(1-\rho^2 y^{\frac{1}{2n-1}}\right) + \left(1-y^{\frac{1}{2n-1}}\right) y\rho^{2n} \right\}^{-3/2} \log y dy + O\left(n^{-3}\right) \\
&= \frac{1}{4n^2} \int_0^1 y^{\frac{(n-3)}{4n-2}} \left(1-\rho^2 y^{\frac{2}{2n-1}}\right)^{3/2} \left\{ \left(1-\rho^2 y^{\frac{1}{2n-1}}\right) + \left(1-y^{\frac{1}{2n-1}}\right) y\rho^{2n} \right\}^{-3/2} \log y dy + O\left(n^{-3}\right) \\
&= \frac{1}{4n^2} \int_0^1 y^{\frac{1}{4}} \left\{ \frac{1-\rho^2 y^{\frac{1}{2n-1}}}{1-\rho^2 y^{\frac{2}{2n-1}}} + \frac{1-y^{\frac{1}{2n-1}}}{1-\rho^2 y^{\frac{2}{2n-1}}} y\rho^{2n} \right\}^{-3/2} \log y dy + O\left(n^{-3}\right) \\
&= \frac{1}{4n^2} \int_0^1 y^{\frac{1}{4}} \left\{ \frac{1-\left(1+\frac{2c}{n}\right)\left(1+\frac{1}{2n}\log y\right)}{1-\left(1+\frac{2c}{n}\right)\left(1+\frac{1}{n}\log y\right)} - \frac{\frac{1}{2n}\log y}{1-\left(1+\frac{2c}{n}\right)\left(1+\frac{1}{n}\log y\right)} y\rho^{2n} \right\}^{-3/2} \log y dy + O\left(n^{-3}\right) \\
&= \frac{1}{4n^2} \int_0^1 y^{\frac{1}{4}} \left\{ \frac{-\frac{2c}{n}-\frac{1}{2n}\log y}{-\frac{2c}{n}-\frac{1}{n}\log y} - \frac{\frac{1}{2n}\log y}{-\frac{2c}{n}-\frac{1}{n}\log y} ye^{2c} \right\}^{-3/2} \log y dy + O\left(n^{-3}\right) \\
&= \frac{1}{4n^2} \int_0^1 y^{\frac{1}{4}} \left\{ \frac{4c+\log y}{4c+2\log y} + \frac{\log y}{4c+2\log y} ye^{2c} \right\}^{-3/2} \log y dy + O\left(n^{-3}\right). \qquad (62)
\end{aligned}
$$

Combining results (60) - (62) gives the following approximation to the binding function for $\rho = 1 + \frac{c}{n}$ with $c < 0$

$$
\begin{aligned}
b_n(\rho) &= \rho - \frac{3\rho}{2}\int_0^1 x^{(n-1)/2}\left(1-\rho^2x^2\right)^{1/2} F_n^{-1/2}dx + \frac{\rho}{2}\int_0^1 x^{(n-3)/2}\left(1-\rho^2x^2\right)^{3/2} F_n^{-3/2}dx \\
&- \frac{n\rho^{2n-1}}{2}\int_0^1 x^{(5n-7)/2}\left(1-\rho^2x^2\right)^{3/2} F_n^{-3/2}\left(1-x\right)dx \\
&= \rho - \frac{3\rho}{4n}\int_0^1 y^{-\frac{3}{4}}\left\{\frac{4c+\log y}{4c+2\log y} + \frac{\log y}{4c+2\log y}ye^{2c}\right\}^{-1/2} dy \\
&+ \frac{\rho}{4n}\int_0^1 y^{-\frac{3}{4}}\left\{\frac{4c+\log y}{4c+2\log y} + \frac{\log y}{4c+2\log y}ye^{2c}\right\}^{-3/2} dy \\
&+ \frac{\rho^{2n-1}}{8n}\int_0^1 y^{\frac{1}{4}}\left\{\frac{4c+\log y}{4c+2\log y} + \frac{\log y}{4c+2\log y}ye^{2c}\right\}^{-3/2} \log y dy + O\left(n^{-2}\right). \qquad (63)
\end{aligned}
$$

Observe the sign change in the last term because of the transformation in the integrand that involves $\log y$, which is negative for $y \in (0,1)$, so the whole expression is negative. When $c \to -\infty$ the approximation (63) has the reduced form

$$
\begin{aligned}
b_n(\rho) &= \rho - \frac{3\rho}{4n}\int_0^1 y^{-\frac{3}{4}}dy + \frac{\rho}{4n}\int_0^1 y^{-\frac{3}{4}} + O\left(n^{-2} + \frac{1}{|c|}\right) \\
&= \rho - \frac{\rho}{2n}\left[\frac{y^{\frac{1}{4}}}{1/4}\right]_0^1 + O\left(n^{-2}\right) = \rho - \frac{2\rho}{n} + O\left(n^{-2} + \frac{1}{|c|}\right),
\end{aligned}
$$

as in the case $|\rho| < 1$. On the other hand, when $c = 0$ we have

$$
\begin{aligned}
b_n(1) &= 1 - \frac{3}{4n}\int_0^1 y^{-\frac{3}{4}}\left\{\frac{1}{2}+\frac{1}{2}y\right\}^{-1/2} dy + \frac{1}{4n}\int_0^1 y^{-\frac{3}{4}}\left\{\frac{1}{2}+\frac{1}{2}y\right\}^{-3/2} dy \\
&+ \frac{1}{8n}\int_0^1 y^{\frac{1}{4}}\left\{\frac{1}{2}+\frac{1}{2}y\right\}^{-3/2} \log y dy + O\left(n^{-2}\right) \\
&= 1 - \frac{3}{4n}2^{1/2}\int_0^1 y^{-\frac{3}{4}}\left\{1+y\right\}^{-1/2} dy + \frac{2^{3/2}}{4n}\int_0^1 y^{-\frac{3}{4}}\left\{1+y\right\}^{-3/2} dy \\
&+ \frac{2^{1/2}}{4n}\int_0^1 y^{\frac{1}{4}}\left\{1+y\right\}^{-3/2}\log y dy + O\left(n^{-2}\right) \\
&= 1 - \frac{1.7814}{n} + O\left(n^{-2}\right),
\end{aligned}
$$

corresponding to (56). Thus, (63) encompasses both the stationary and unit root cases at the limits of the domain of definition for $c < 0$.

(iv) Case $\rho = 1 + \frac{c}{n}$, $c > 0$

We start with the local to unity case $\rho = 1 + c/n$ with $c > 0$ and later consider the fixed $\rho > 1$. The binding function formula for $\rho > 1$ is

$$\begin{aligned} b_n(\rho) = \rho &+ \frac{3\rho}{2}\int_1^\infty x^{(n-1)/2}\left(\rho^2x^2-1\right)^{1/2} G_n^{-1/2}dx \\ &- \frac{\rho}{2}\int_1^\infty x^{(n-3)/2}\left(\rho^2x^2-1\right)^{3/2} G_n^{-3/2}dx \\ &- \frac{n\rho^{2n-1}}{2}\int_1^\infty x^{(n-5)/2}\left(\rho^2x^2-1\right)^{3/2} G_n^{-3/2}\left(x-1\right)x^{2n-1}dx. \end{aligned} \tag{64}$$

We proceed to take each term in turn. As before, set $y = x^{2n-1}$ so that

$$dy = (2n-1)\,x^{2n-2}dx = (2n-1)\,y^{\frac{2n-2}{2n-1}}dx = (2n-1)\,y^{1-\frac{1}{2n-1}}dx,$$

and use the expansion $y^{\frac{1}{2n-1}} = 1 + \frac{1}{2n-1}\log y + O\left(n^{-2}\right)$. Then, using $G_n = \rho^2x - 1 + (x-1)\,x^{2n-1}\rho^{2n}$ and for $\rho = 1 + \frac{c}{n}$ with $c > 0$, the integral in the second term of (64) is

$$\begin{aligned} &\int_1^\infty x^{(n-1)/2}\left(\rho^2x^2-1\right)^{1/2} G_n^{-1/2}dx \\ &= \int_1^\infty x^{(n-1)/2}\left\{\frac{\rho^2x^2-1}{\rho^2x-1+(x-1)\,x^{2n-1}\rho^{2n}}\right\}^{1/2}dx \\ &= \int_1^\infty y^{\frac{(n-1)/2}{2n-1}}\left\{\frac{\rho^2y^{\frac{2}{2n-1}}-1}{\rho^2y^{\frac{1}{2n-1}}-1+\left(y^{\frac{1}{2n-1}}-1\right)y\rho^{2n}}\right\}^{1/2}\frac{dy}{(2n-1)\,y^{\frac{2n-2}{2n-1}}} \\ &= \frac{1}{2n}\int_1^\infty y^{-\frac{3}{4}}\left\{\frac{\left(1+\frac{2c}{n}\right)\left(1+\frac{2}{2n}\log y\right)-1}{\left(1+\frac{2c}{n}\right)\left(1+\frac{1}{2n}\log y\right)-1+\frac{\rho^{2n}}{2n}y\log y}\right\}^{1/2}dy \\ &= \frac{1}{2n}\int_1^\infty y^{-\frac{3}{4}}\left\{\frac{4c+2\log y}{4c+\log y+\rho^{2n}y\log y}\right\}^{1/2}dy \end{aligned}$$

Use the transformation $w = \log y$ so that $w \in [0,\infty)$ and $dy = e^w dw$, giving

$$\begin{aligned} &\frac{1}{2n}\int_1^\infty y^{-\frac{3}{4}}\left\{\frac{4c+2\log y}{4c+\log y+\rho^{2n}y\log y}\right\}^{1/2}dy \\ &= \frac{1}{2n}\int_0^\infty e^{\frac{1}{4}w}\left\{\frac{4c+2w}{4c+w+\rho^{2n}we^w}\right\}^{1/2}dw. \end{aligned} \tag{65}$$

Proceeding in the same way with the integral in the third term of (64) we have

$$
\begin{aligned}
&\int_1^\infty x^{(n-3)/2}\left(\rho^2x^2-1\right)^{3/2}G_n^{-3/2}dx\\
&=\int_1^\infty x^{(n-3)/2}\left\{\frac{\rho^2x^2-1}{\rho^2x-1+(x-1)\,x^{2n-1}\rho^{2n}}\right\}^{3/2}dx\\
&=\int_1^\infty y^{\frac{(n-3)/2}{2n-1}}\left\{\frac{\rho^2y^{\frac{2}{2n-1}}-1}{\rho^2y^{\frac{1}{2n-1}}-1+\left(y^{\frac{1}{2n-1}}-1\right)y\rho^{2n}}\right\}^{3/2}\frac{dy}{(2n-1)\,y^{\frac{2n-2}{2n-1}}}\\
&=\frac{1}{2n}\int_1^\infty y^{-\frac{3}{4}}\left\{\frac{\left(1+\frac{2c}{n}\right)\left(1+\frac{2}{2n}\log y\right)-1}{\left(1+\frac{2c}{n}\right)\left(1+\frac{1}{2n}\log y\right)-1+\frac{\rho^{2n}}{2n}y\log y}\right\}^{3/2}dy\left\{1+O\left(n^{-1}\right)\right\}\\
&=\frac{1}{2n}\int_1^\infty y^{-\frac{3}{4}}\left\{\frac{4c+2\log y}{4c+\log y+\rho^{2n}y\log y}\right\}^{3/2}dy\left\{1+O\left(n^{-1}\right)\right\}\\
&=\frac{1}{2n}\int_0^\infty e^{\frac{1}{4}w}\left\{\frac{4c+2w}{4c+w+\rho^{2n}we^{w}}\right\}^{3/2}dw. \qquad (66)
\end{aligned}
$$

Finally, the integral in the fourth term of (64) is

$$
\begin{aligned}
&\int_1^\infty x^{(n-5)/2}\left(\rho^2x^2-1\right)^{3/2}G_n^{-3/2}\,(x-1)\,x^{2n-1}dx\\
&\int_1^\infty x^{(n-5)/2}\left\{\frac{\rho^2x^2-1}{\rho^2x-1+(x-1)\,x^{2n-1}\rho^{2n}}\right\}^{3/2}(x-1)\,x^{2n-1}dx\\
&=\int_1^\infty y^{\frac{(n-5)/2}{2n-1}}\left\{\frac{\rho^2y^{\frac{2}{2n-1}}-1}{\rho^2y^{\frac{1}{2n-1}}-1+\left(y^{\frac{1}{2n-1}}-1\right)y\rho^{2n}}\right\}^{3/2}\frac{\left(y^{\frac{1}{2n-1}}-1\right)ydy}{(2n-1)\,y^{\frac{2n-2}{2n-1}}}\\
&=\frac{1}{2n}\int_1^\infty y^{\frac{1}{4}}\left\{\frac{\left(1+\frac{2c}{n}\right)\left(1+\frac{2}{2n}\log y\right)-1}{\left(1+\frac{2c}{n}\right)\left(1+\frac{1}{2n}\log y\right)-1+\frac{\rho^{2n}}{2n}y\log y}\right\}^{3/2}\left(\frac{1}{2n}\log y\right)dy\left\{1+O\left(n^{-1}\right)\right\}\\
&=\left(\frac{1}{2n}\right)^2\int_1^\infty y^{\frac{1}{4}}\left\{\frac{4c+2\log y}{4c+\log y+\rho^{2n}y\log y}\right\}^{3/2}\log y\,dy\left\{1+O\left(n^{-1}\right)\right\}\\
&=\left(\frac{1}{2n}\right)^2\int_0^\infty e^{\frac{5}{4}w}\left\{\frac{4c+2w}{4c+w+\rho^{2n}we^{w}}\right\}^{3/2}wdw\left\{1+O\left(n^{-1}\right)\right\}. \qquad (67)
\end{aligned}
$$

Combining (65) - (67) in (64) we get for $\rho=1+\frac{c}{n}$ with $c>0$

$$
\begin{aligned}
b_n(\rho)=\rho&+\frac{3\rho}{4n}\int_0^\infty e^{\frac{1}{4}w}\left\{\frac{4c+2w}{4c+w+\rho^{2n}we^{w}}\right\}^{1/2}dw\\
&-\frac{\rho}{4n}\int_0^\infty e^{\frac{1}{4}w}\left\{\frac{4c+2w}{4c+w+\rho^{2n}we^{w}}\right\}^{3/2}dw\\
&-\frac{\rho^{2n-1}}{8n}\int_0^\infty e^{\frac{5}{4}w}\left\{\frac{4c+2w}{4c+w+\rho^{2n}we^{w}}\right\}^{3/2}wdw+O\left(n^{-2}\right).
\end{aligned}
$$

Hence the bias function to $O\left(n^{-1}\right)$ in this case of local to unity on the explosive side of unity is

$$
\begin{aligned}
b_n(1+\frac{c}{n}) = 1 + \frac{c}{n} + \frac{3}{4n}\int_0^\infty e^{\frac{1}{4}w}\left\{\frac{4c+2w}{4c+w+\rho^{2n}we^w}\right\}^{1/2}dw \\
-\frac{1}{4n}\int_0^\infty e^{\frac{1}{4}w}\left\{\frac{4c+2w}{4c+w+\rho^{2n}we^w}\right\}^{3/2}dw \\
-\frac{\rho^{2n}}{8n}\int_0^\infty e^{\frac{5}{4}w}\left\{\frac{4c+2w}{4c+w+\rho^{2n}we^w}\right\}^{3/2}wdw + O\left(n^{-2}\right). \qquad (68)
\end{aligned}
$$

(v) Case $|\rho| > 1$

We now turn to the case of fixed $\rho > 1$. The relevant bias expression is from (54)

$$
\begin{aligned}
b_n(\rho) = \rho + \frac{3\rho}{2}\int_1^\infty x^{(n-1)/2}\left(\rho^2x^2-1\right)^{1/2}G_n^{-1/2}dx \\
-\frac{\rho}{2}\int_1^\infty x^{(n-3)/2}\left(\rho^2x^2-1\right)^{3/2}G_n^{-3/2}dx \\
-\frac{n\rho^{2n-1}}{2}\int_1^\infty x^{(n-5)/2}\left(\rho^2x^2-1\right)^{3/2}G_n^{-3/2}\left(x-1\right)x^{2n-1}dx.
\end{aligned}
$$

We examine the order of magnitude of each term in turn as $n \to \infty$. For the first term

$$
\begin{aligned}
\int_1^\infty x^{(n-1)/2}\left(\rho^2x^2-1\right)^{1/2}G_n^{-1/2}dx &= \int_1^\infty x^{(n-1)/2}\left\{\frac{\rho^2x^2-1}{\rho^2x-1+\left(x-1\right)x^{2n-1}\rho^{2n}}\right\}^{1/2}dx \\
&= \frac{1}{\rho^n}\int_1^\infty \frac{x^{(n-1)/2}}{x^{n-1/2}}\left\{\frac{\rho^2x^2-1}{\left(x-1\right)+\frac{\rho^2x-1}{x^{2n-1}\rho^{2n}}}\right\}^{1/2}dx \\
&\leq \frac{B}{\rho^n}\int_1^\infty \frac{1}{x^{n/2}}dx = O\left(n^{-1}\rho^{-n}\right).
\end{aligned}
$$

In a similar way, the second term is

$$
\begin{aligned}
\int_1^\infty x^{(n-3)/2}\left(\rho^2x^2-1\right)^{3/2}G_n^{-3/2}dx &= \int_1^\infty x^{(n-3)/2}\left\{\frac{\rho^2x^2-1}{\rho^2x-1+\left(x-1\right)x^{2n-1}\rho^{2n}}\right\}^{3/2}dx \\
&= O\left(n^{-1}\rho^{-3n}\right).
\end{aligned}
$$

The third term is

$$
\begin{aligned}
&\frac{n\rho^{2n-1}}{2}\int_1^\infty x^{(n-5)/2}\left(\rho^2x^2-1\right)^{3/2}G_n^{-3/2}\left(x-1\right)x^{2n-1}dx \\
&=\frac{n\rho^{2n-1}}{2}\int_1^\infty x^{(n-5)/2}\left\{\frac{\rho^2x^2-1}{\rho^2x-1+\left(x-1\right)x^{2n-1}\rho^{2n}}\right\}^{3/2}\left(x-1\right)x^{2n-1}dx \\
&=\frac{n\rho^{-n-1}}{2}\int_1^\infty \frac{x^{(n-5)/2+2n-1}}{x^{3n-3/2}}\left\{\frac{\rho^2x^2-1}{\left(x-1\right)+\frac{\rho^2x-1}{x^{2n-1}\rho^{2n}}}\right\}^{3/2}\left(x-1\right)dx \\
&=\frac{n\rho^{-n-1}}{2}\int_1^\infty \frac{1}{x^{n/2+2}}\left\{\frac{\rho^2x^2-1}{\left(x-1\right)+\frac{\rho^2x-1}{x^{2n-1}\rho^{2n}}}\right\}^{3/2}\left(x-1\right)dx \\
&=O\left(\rho^{-n}\right),
\end{aligned}
$$

It follows that $b_n(\rho)=\rho+O\left(\rho^{-n}\right)$, showing that the bias is exponentially small (and negative) for $\rho>1$. A similar result holds when $\rho<-1$, in which case the bias is exponentially small and positive. ■

8 References

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

Bao, Y. (2007). "The approximate moments of the least squares estimator for the stationary autoregressive model under a general error distribution." *Econometric Theory,* 23. 1013-1021.

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: J. Wiley.

Chan, N. H. and C. Z. Wei (1987). "Asymptotic Inference for Nearly Nonstationary AR(1) Processes, *Annals of Statistics* 15, 1050–1063.

Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press.

Gouriéroux, C. A., A. Monfort, and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics,* 8S, 85-118.

Gouriéroux, C., P. C. B. Phillips, and J. Yu: "Indirect Inference for Dynamic Panel Models", *Journal of Econometrics*, 2010 (forthcoming)

Gourieroux, C., E. Renault and N. Touzi (2000). "Calibration by Simulation for Small Sample Bias Correction," In R.S. Mariano, T. Schuermann and M.Weeks (eds.), *Simulation-Based Inference in Econometrics: Methods and Applications*, Cambridge University Press, p. 328-358.

Cristea, M. (2007). "A note on global implicit function theorem", *Journal of Inequalities in Pure and Applied Mathematics*, 8, 1-28.

Ge, S. S. C. Wang (2002). “Adaptive NN control of uncertain nonlinear pure-feedback systems,” *Automatica*, 38, 671-682.

Guiérrez, C., C. Biasi and E. L. dos Santos (2009). “Global inverse mapping theorems”. *Cadernos de Matemática*, 10, 9-18.

Hadamard, J. (1906). “Sur les transformations ponctuelles”, *Bulletin de la Société Mathématique de France*, 34, 71-84.

Hajivassiliou, V. and D. McFadden (1998). “The method of simulated scores for the estimation of LDV models,” *Econometrica*, 66, 863-896.

Han, C., P. C. B. Phillips and D. Sul, (2009). “Uniform Asymptotic Normality in Stationary and Unit Root Autoregression”, Working paper, Yale University.

Ichiraku, S. (1985). “A note on global implicit function theorems”, *IEEE Transactions on Circuits and Systems*, CAS-32, 503-505.

Klein, L. R. (1953). *A Textbook of Econometrics.* Prentice Hall.

MacKinnon, J. G. and A. A. Smith (1998). “Approximate bias correction in econometrics,” *Journal of Econometrics*, 85, 205-230.

McFadden, D. (1989). “A method of simulated moments for estimation of discrete response models without numerical integration,” *Econometrica*, 57, 995-1026.

Mann, H. B. and A. Wald (1943). “On stochastic limit and order relationships,” *Annals of Mathematical Statistics*, 14, 217-226.

Nielsen, B. (2009). “Singular vector autoregressions with deterministic terms: Strong consistency and lag order determination.” University of Oxford working paper.

Park, J. Y. and P. C. B. Phillips (1988). “Statistical Inference in Regressions With Integrated Processes: Part 1,” *Econometric Theory* 4, 468–497.

Phillips, P. C. B. (1987). “Towards a Unified Asymptotic Theory for Autoregression,” *Biometrika* 74, 535–547.

Phillips, P. C. B. (1988). “Multiple regression with integrated processes.” In N. U. Prabhu, (ed.), *Statistical Inference from Stochastic Processes, Contemporary Mathematics* 80, 79–106.

Phillips, P. C. B. (1988). “ET Interview: Professor J. Durbin”, *Econometric Theory,* 4, 125–158.

Phillips, P.C.B. (2007). "Regression with Slowly Varying Regressors and Nonlinear Trends," *Econometric Theory,* 23, 557-614.

Phillips, P. C. B., T. Magdalinos (2007), "Limit Theory for Moderate Deviations from a Unit Root," *Journal of Econometrics* 136, 115-130.

Phillips P. C. B. and T. Magdalinos (2008). "Limit theory for explosively cointegrated systems", *Econometric Theory*, 24, 865-887.

Phillips, P. C. B., and T. Magdalinos "Unit Root and Cointegrating Limit Theory when Initialization is in the Infinite Past", *Econometric Theory*, 25, 2009, 1682-1715.

Phillips P. C. B. and J. Yu (2005). "Jackkifing bond option pricing". *Review of Financial Studies*, 18, 707-742.

Phillips P. C. B. and J. Yu (2009). "Simulation-based estimation of contingent-claims prices". *Review of Financial Studies*, 22, 3669-3705.

Pollard, D. and P. Radchenko (2006). "Nonlinear least squares estimation", *Journal of Multivariate Analysis*, 97, 548-562.

Rubin, H. "Topological properties of measures on topological spaces." Unpublished.

Samuelson, P. A. (1953). "Prices of factors and goods in general equilibrium," *Review of Economic Studies,* 21, 1-20.

Sargan, J. D. (1976). Econometric Estimators and the Edgeworth Approximation. *Econometrica*, 44, 421-448.

Sargan, J. D. (1983). "Identification and lack of identification". *Econometrica,* 51, 1605-1633.

Shenton, L. R. and W. L. Johnson (1965). "Moments of a serial correlation coefficient," *Journal of the Royal Statistical Society, Series B*, 27, 308-320.

Shenton, L. R. and H. D. Vinod. "Closed forms for asymptotic bias and variance in autoregressive models with unit roots," *Journal of Computational and Applied Mathematics*, 61, 231-243.

Smith, A. A. (1993). "Estimating nonlinear time series models using simulated vector autoregression," *Journal of Applied Econometrics*, 8, S63-S84.

Stock, J. H. (1991). "Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series," *Journal of Monetary Economics*, 28, 435-459.

Sweeting, T. J. (1986). "On a converse to Scheffé's theorem," *Annals of Statistics*, 14, 1252-1256.

Tang, C. Y. and S. X. Chen (2009). "Parameter estimation and bias correction for diffusion processes." *Journal of Econometrics*, 149, 65-81.

Topsoe, F. (1967). "Preservation of weak convergence under mappings," *Annals of Mathematical Statistics*, 38, 1661-1665.

van der Vaart, A. W. (2000). *Asymptotic Statistics,* Cambridge University Press.

van der Vaart and J.A. Wellner (2000), *Weak Convergence and Empirical Processes*, Springer: New York.

Vinod, H. D. and L. R. Shenton (1996). “Exact moments for autoregressive and random walk models for a zero or stationary initial value.” *Econometric Theory*, 12, 481-499.

White, J. S. (1961). “Asymptotic expansions for the mean and variance of the serial correlation coefficient”, *Biometrika*, 48, 85-94.

Yu, J. (2009). “Bias in the estimation of the mean reversion parameter in continuous time models.” Working Paper.

Zhang,W. and S. S. Ge (2006). “A global implicit function theorem without initial point and its application to control of non-affine systems of high dimension”. *Journal of Mathematical Analysis and Applications*, 313, 251-261.