

Asymptotic Moments of Autoregressive Estimates with a Near Unit Root and Minimax Risk

Bruce E. Hansen*
University of Wisconsin†

November 2013

Abstract

This moments of the asymptotic distribution of the least-squares estimator of the local-to-unity autoregressive model are computed using computationally simple integration. These calculations show that conventional simulation estimation of moments can be substantially inaccurate unless the simulation sample size is very large. We also explore the minimax efficiency of autoregressive coefficient estimation, and numerically show that a simple Stein shrinkage estimator has minimax risk which is uniformly better than least squares, even though the estimation dimension is just one.

*Research supported by the National Science Foundation.

†Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706.

1 Introduction

In a series of seminal contributions, Phillips (1987ab) and Phillips and Perron (1988) developed an asymptotic theory of inference for unit roots in autoregressive models. A core component of this theory is the near unit root model which is parameterized by a localizing parameter c . This model has been the foundation for nearly all subsequent work in non-stationary time series econometric theory.

A key feature of this theory is that it yields simple expression for asymptotic distributions as functions of continuous-time Brownian motions and diffusion processes indexed by c . An inconvenience is that analytic expressions for the distributions are not available. The standard view is that this is not a problem, as the distributions can always be simulated. And indeed numerical calculation of non-stationary asymptotic distributions by simulation is the standard approach. Important examples include MacKinnon (1994)'s calculation of asymptotic critical values and Stock (1991)'s calculation of quantiles for confidence interval construction. A recent example is Phillips (2012b) who examines confidence interval construction.

Following Nabeya (1999), we show that moments of the asymptotic distribution can be calculated by direct integration. This is computationally much simpler (a matter of minutes versus days) and more accurate. As a by-product of our calculations, we find that simulation estimation of near-unit-root distributions for large values of c requires very large sample sizes, much larger than those used in conventional practice.

We also explore the issue of efficient estimation in the near unit root model. Ploberger and Phillips (2012) have recently argued that while the OLS estimator is non-standard, it is minimax efficient in a certain sense. We argue that their argument is incomplete, that it ignores the unbounded nature of estimation variance in the local-to-unit model. We show numerically that a standard Stein shrinkage estimator uniformly dominates the OLS estimator, and can be viewed as dominating OLS in a minimax sense. This result suggests that efficiency is an open question ready to be explored.

The paper is organized as follows. Section 2 introduces the local-to-unit model, its asymptotic moments, and the main theoretical contribution of the paper, which is an expression for the moments in terms of a simple integral. Section 3 presents numerical computation of the moments by both integration and simulation. Section 4 presents a discussion of minimax efficiency. Section 5 introduces the Stein-type shrinkage estimator and contrast its asymptotic risk versus OLS by numerical simulation. Section 6 is a conclusion, and Section 7 contains the proof of Theorem 1.

A Gauss program which creates the numerical work reported in the paper is available on the author's webpage <http://www.ssc.wisc.edu/~bhansen/>

2 Moments of the Asymptotic Distribution

Take the classic AR(1) with a near unit root

$$y_t = \alpha_n y_{t-1} + e_t \quad (1)$$

$$\alpha_n = 1 + c/n \quad (2)$$

with e_t zero mean white noise. Let $\hat{\alpha}_n$ denote the OLS estimator of α . As shown by Chan and Wei (1987) and Phillips (1987),

$$n(\hat{\alpha}_n - \alpha_n) \rightarrow_d \frac{U_c}{V_c}, \quad (3)$$

as $n \rightarrow \infty$, where $U_c = \int_0^1 W_c dW$ and $V_c = \int_0^1 W_c^2$. In this expression, $W(r)$ denotes a standard Brownian motion, and $dW_c(r) = cW_c(r) + dW(r)$ is a standard diffusion process.

Define the r^{th} moment of the asymptotic distribution (3):

$$\mu_r(c) = E \left(\frac{U_c}{V_c} \right)^r. \quad (4)$$

The main theoretical contribution of the paper is a convenient expression for $\mu_r(c)$ as a simple integral.

Theorem 1 For any integer $r \geq 1$ and $c \leq 0$,

$$\mu_r(c) = \sum_{j=0}^r \binom{r}{j} (-c)^{r-j} \int_0^1 g_j(x, c) dx \quad (5)$$

where

$$g_j(x, c) = \frac{2^{3/2-2j} \lambda(x, c) (\lambda(x, c)^2 - c^2)^{j-1}}{(j-1)! (1-x)^2 e^{x/(2-2x)} (1 + e^{-2\lambda(x, c)})^{1/2}} \sum_{\ell=0}^j \binom{j}{\ell} \frac{(-1)^{j-\ell} (2\ell-1)!! \psi(\lambda(x, c))^\ell}{(1 - c\psi(\lambda(x, c)))^{\frac{1}{2}+\ell}}$$

with $\lambda(x, c) = x/(1-x) - c$ and $\psi(u) = \frac{\tanh(u)}{u}$ with $\psi(0) = 1$. The notation $a!!$ is the double factorial defined as $a!! = 1 \cdot 3 \cdots a$ with the convention $a!! = 1$ for $a < 0$.

Theorem 1 restricts the near-unity parameter c to be non-positive, and thus does not cover the locally explosive case. The technical reason for this restriction is due to one of the change-of-variables used in obtaining (5); it could be avoided by alternative manipulations. The representation (5) is particularly convenient, however, as the functions $g_j(x, c)$ (with $c \leq 0$) are free of poles on $[0, 1]$ and thus numerical integration is well behaved.

Theorem 1 gives an integral representation for the exact moments of the local-to-unity asymptotic distribution. This extends Nabeya (1999) who provided an integral representation for the exact moments in the case $c = 0$.

There is a long history of papers investigating asymptotic expansions for asymptotic bias and variance of $\widehat{\alpha}_n$, including White (1961), Shenton and Johnson (1965), and Shenton and Vinod (1995). Most recently, Phillips (2012a, Theorem 3) provides an integral representation of the finite sample bias of $\widehat{\alpha}_n$, and Phillips (2012a, Theorem 4) provides asymptotic expansions for the bias. Theorem 1 above is complementary to these results, as it provides an exact integral representation for the asymptotic local-to-unity model.

3 Calculation of Asymptotic Moments

We calculated the integrals in (5) by numerical integration¹. We divided the interval $[0, 1]$ into 100 intervals of length $1/100$, and over each interval numerically integrated using Gauss-Legendre quadrature with 40 gridpoints in each interval. We calculated the first four moments, and then transformed into conventional cumulants, including the mean $\mu_1(c)$, variance

$$\sigma^2(c) = \mu_2(c) - \mu_1(c)^2,$$

skewness

$$skew(c) = \frac{\mu_3(c) - 3\mu_2(c)\mu_1(c) + 2\mu_1(c)^3}{\sigma^{3/2}(c)}$$

and kurtosis

$$kurtosis(c) = \frac{\mu_4(c) - 4\mu_3(c)\mu_1(c) + 6\mu_2(c)\mu_1(c)^2 - 3\mu_1(c)^4}{\sigma^4(c)}.$$

These four cumulants are reported in Table 1 (for $c = 0$ to $c = -20$ in steps of 1) and in Table 2 (for $c = -40$ to $c = -400$ in steps of 20). The values for $c = 0$ are identical to those reported in Nabeya (1999).

Table 1

¹The computation was done in Gauss using the `intquad1` command.

	Mean	Variance	Skewness	Kurtosis
$c = 0$	-1.781	10.11	-2.270	11.37
$c = -1$	-1.882	11.76	-2.068	9.971
$c = -2$	-1.930	13.54	-1.901	8.887
$c = -3$	-1.954	15.41	-1.759	8.043
$c = -4$	-1.968	17.33	-1.640	7.385
$c = -5$	-1.976	19.27	-1.539	6.886
$c = -6$	-1.981	21.23	-1.453	6.451
$c = -7$	-1.985	23.20	-1.379	6.112
$c = -8$	-1.988	25.18	-1.315	5.831
$c = -9$	-1.990	27.16	-1.258	5.596
$c = -10$	-1.991	29.15	-1.208	5.395
$c = -11$	-1.992	31.13	-1.163	5.223
$c = -12$	-1.993	33.12	-1.123	5.073
$c = -13$	-1.994	35.11	-1.086	4.942
$c = -14$	-1.995	37.11	-1.053	4.826
$c = -15$	-1.995	39.10	-1.023	4.723
$c = -16$	-1.996	41.09	-0.9947	4.631
$c = -17$	-1.996	43.09	-0.9689	4.548
$c = -18$	-1.997	45.08	-0.9449	4.473
$c = -19$	-1.997	47.08	-0.9226	4.405
$c = -20$	-1.997	48.07	-0.9018	4.343

	Mean	Variance	Skewness	Kurtosis
$c = -40$	-1.997	89	-0.6546	3.711
$c = -60$	-1.999	129	-0.5390	3.483
$c = -80$	-2.000	169	-0.4687	3.365
$c = -100$	-2.000	209	-0.4203	3.294
$c = -120$	-2.000	249	-0.3843	3.246
$c = -140$	-2.000	289	-0.3562	3.211
$c = -160$	-2.000	329	-0.3335	3.185
$c = -180$	-2.000	369	-0.3146	3.165
$c = -200$	-2.000	409	-0.2986	3.148
$c = -220$	-2.000	449	-0.2848	3.135
$c = -240$	-2.000	489	-0.2728	3.124
$c = -260$	-2.000	529	-0.2622	3.114
$c = -280$	-2.000	569	-0.2527	3.106
$c = -300$	-2.000	609	-0.2442	3.099
$c = -320$	-2.000	649	-0.2365	3.093
$c = -340$	-2.000	689	-0.2295	3.088
$c = -360$	-2.000	729	-0.2230	3.083
$c = -380$	-2.000	769	-0.2171	3.079
$c = -400$	-2.000	809	-0.2116	3.075

The exact moments can be compared to estimated moments from simulations. The r^{th} finite sample moment is

$$\mu_r(c, n) = E(n(\hat{\alpha}_n - \alpha_n))^r.$$

which approaches $\mu_r(c, n)$ as $n \rightarrow \infty$. Unit root distributions are typically calculated by simulation with large values of n , including $n = 500$ in early papers and $n = 1000$ in later papers. We calculated the same moments (and cummulants) by simulation using 1,000,000 simulation replications and $n = 500$, $n = 1000$, $n = 10,000$, and $n = 100,000$. The results are presents graphically in Figure 1 (for c ranging from -20 to 0) and in Figure 2 (for c ranging from -400 to 0).

Examining the figures, we can see that the simulation moment estimates can be quite poor unless n is very large. The discrepancy is worst for the low order moments. In particular, the simulation estimate of the mean with $n = 500$ and $n = 1000$ is far from accurate even for small values of c . The simulation estimates of the variance are reasonably accurate for small c , but are quite inaccurate for large c , unless n is very large. The simulation estimates of skewness and kurtosis, however, are excellent even for small n .

The errors displayed in the figures show that for reasonable accuracy (except for very small c), the simulation estimate requires setting $n = 100,000$. This is surprisingly large, and much larger than the values used in existing studies. For example, Stock (1992) used $n = 500$ to calculate

the distributions for c as large as $c = -38$ to 6 . Phillips (2012b) used $n = 10,000$ to calculate distributions for c as large as $c = -450$. Our calculations suggest that these values of n are much too small.

To contrast the computation costs, numerical integration is quite quick, with all of the results reported in this paper computed in just a few minutes on an office PC. In contrast, the simulation results took 6 days to compute.

4 Minimax Efficiency

Is the OLS estimator $\hat{\alpha}_n$ efficient for α_n ? Ploberger and Phillips (2012) argue that it is in a certain sense. We re-investigate this question.

We start by reviewing the classic theory of estimation efficiency developed by Hájek (1970, 1972), Le Cam (1982), and van der Vaart (1998) in the locally asymptotic normality (LAN) case. For concreteness and simplicity let's consider a LAN model $f(x, \theta)$ with $\theta \in \Theta \subset \mathbb{R}^k$. If $\hat{\theta}_n$ denotes the MLE from a sample of size n , then

$$\sqrt{n} \left(\hat{\theta}_n - \theta \right) \rightarrow_d Z \sim N(0, J(\theta))$$

where $J(\theta)$ is the inverse of the Fisher information matrix. For any bowl-shaped loss function $\ell(u)$, the asymptotic risk of this estimator is

$$\rho \left(\hat{\theta}, \theta \right) = \lim_{n \rightarrow \infty} E_{\theta} \ell \left(\sqrt{n} \left(\hat{\theta}_n - \theta \right) \right) = E_{\theta} \ell (Z).$$

where E_{θ} means expectation with respect to the model $f(x, \theta)$. For example, with quadratic risk $\ell(u) = u'u$, then $\rho \left(\hat{\theta}, \theta \right) = \text{tr } J(\theta)$.

In this setting, we might ask if there is an alternative estimator $\tilde{\theta}_n$ with smaller risk. This is a treacherous question. Consider the estimator $\tilde{\theta}_n = \theta$. Then $\rho \left(\tilde{\theta}, \theta \right) = \rho(\theta, \theta)$ is minimized and $\tilde{\theta}_n$ has smaller risk than the MLE. This seems disingenuous, as we have constructed an estimator which uses knowledge of the true value of the parameter. But it points to the need to be more careful about what we mean by “smaller risk”.

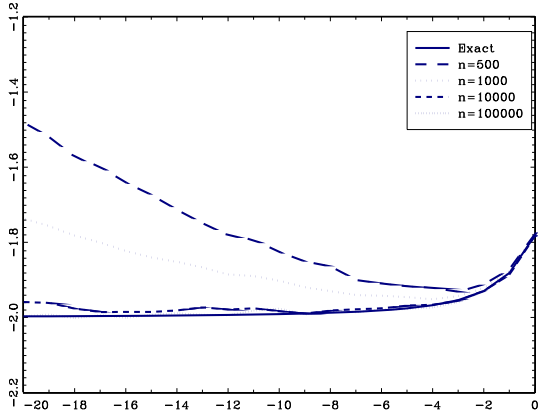
A classic solution to this problem is the minimax criterion: we say an estimator is minimax efficient if it minimizes the maximum risk over (a region of) the parameter space. For $\Gamma \subset \Theta$, we define the maximum asymptotic risk of an estimator $\tilde{\theta}_n$ as

$$\sup_{\theta \in \Gamma} \rho \left(\tilde{\theta}, \theta \right) = \sup_{\theta \in \Gamma} \lim_{n \rightarrow \infty} E_{\theta} \ell \left(\sqrt{n} \left(\tilde{\theta}_n - \theta \right) \right).$$

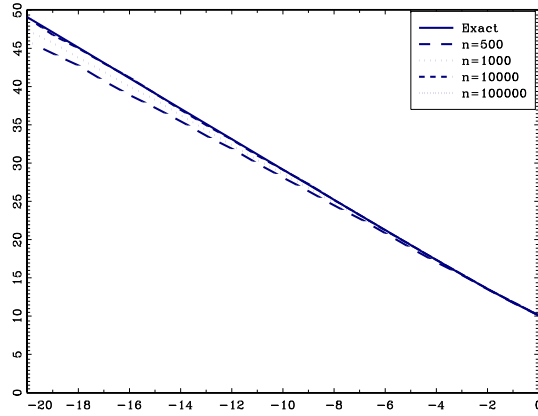
This definition escapes the superefficiency paradox. So long as Γ is not a singleton (contains more than one value of θ) then we cannot artificially set the maximum risk to zero. Essentially, the minimax criterion requires efficient estimators to have uniformly low risk.

There is another difficulty, however. This maximum risk can easily be infinite. For example,

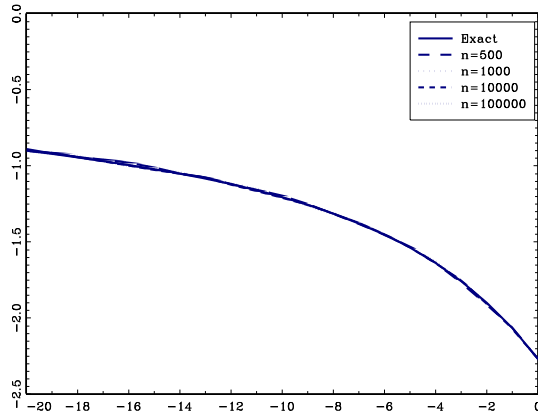
Figure 1: Numerical Integration versus Numerical Simulation, $-20 \leq c \leq 0$



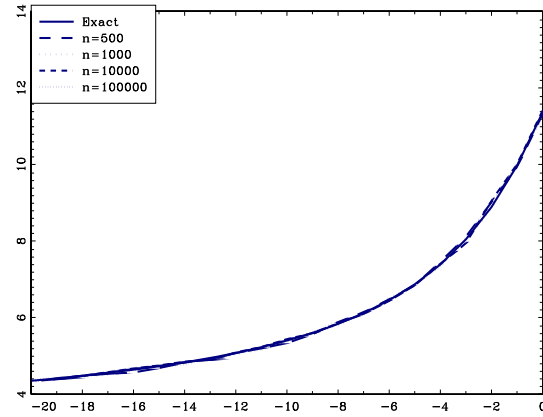
(a) Mean



(b) Variance

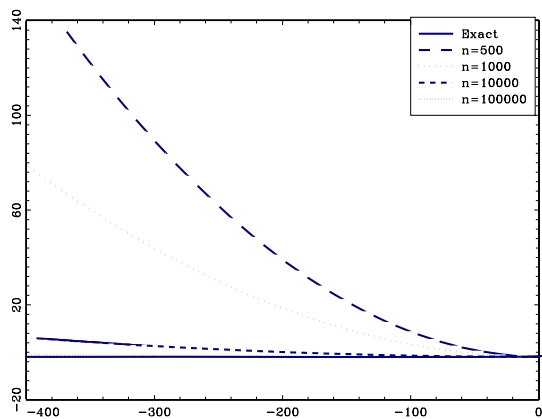


(c) Skewness

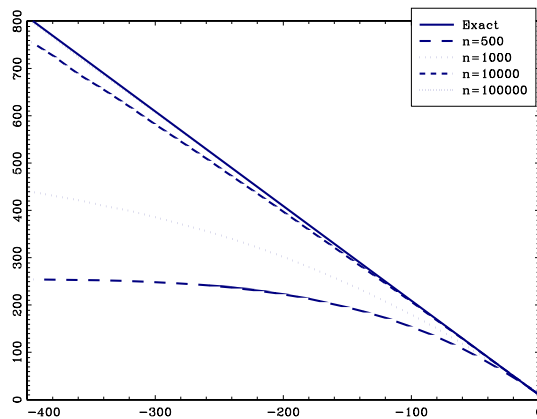


(d) Kurtosis

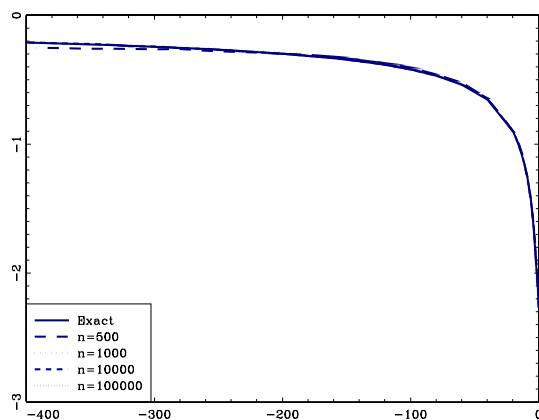
Figure 2: Numerical Integration versus Numerical Simulation, $-400 \leq c \leq 0$



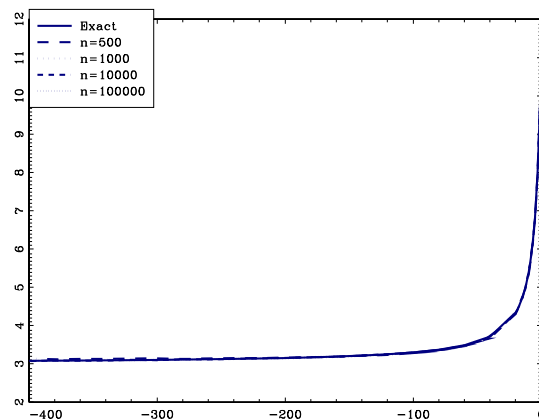
(a) Mean



(b) Variance



(c) Skewness



(d) Kurtosis

suppose $X \sim N(\mu, \sigma^2)$ so that $\theta = (\mu, \sigma^2)$ and $\Theta = \mathbb{R} \times \mathbb{R}^+$, and consider quadratic loss on μ , $\ell(\tilde{\theta} - \theta) = (\tilde{\mu} - \mu)^2$. Then $\rho(\hat{\theta}, \theta) = \sigma^2$ and $\sup_{\theta \in \Theta} \rho(\hat{\theta}, \theta) = \sup_{\sigma^2 > 0} \sigma^2 = \infty$. The problem is that the “worst-case” risk is dominated by the extreme parameter values, and cannot be compensated by good estimation methods.

The solution to this difficulty is to define the maximal risk over a local neighborhood of a parameter value θ . An elegant formulation (see van der Vaart (1998), Chapter 8) reparameterizes using a local parameter space. Define the parameter sequence

$$\theta_n = \theta + n^{-1/2}h$$

where $\theta \in \Theta$ and $h \in \mathbb{R}^k$. We then consider the sequence of probability models indexed by θ_n . In this local reparameterization, for any $h \in \mathbb{R}^k$ the MLE satisfies

$$\sqrt{n}(\hat{\theta}_n - \theta_n) \rightarrow_d Z \sim N(0, J(\theta))$$

and the asymptotic risk equals

$$\rho(\hat{\theta}, \theta) = \lim_{n \rightarrow \infty} E_{\theta_n} \ell(\sqrt{n}(\hat{\theta}_n - \theta_n)) = E_{\theta} \ell(Z).$$

Since the limit is independent of h , the maximal (local) risk of the MLE is thus

$$\sup_{h \in \mathbb{R}^k} \lim_{n \rightarrow \infty} E_{\theta_n} \ell(\sqrt{n}(\hat{\theta}_n - \theta_n)) = E_{\theta} \ell(Z).$$

Furthermore, the famous minimax theorem due to Hájek (see van der Vaart (1998), Theorem 8.11) shows that $E_{\theta} \ell(Z)$ is a lower bound on the maximal risk for any estimator sequence, showing that the MLE is minimax efficient.

Now let’s apply this theory to the local-to-unity model (1)-(2) which is parameterized in terms of the local-to-unity parameter $c \leq 0$ and is local to $\alpha = 1$. The asymptotic risk of the OLS estimator is

$$\begin{aligned} \rho(\hat{\alpha}, \alpha) &= \lim_{n \rightarrow \infty} E_{\alpha_n} \ell(n(\hat{\alpha}_n - \alpha_n)) \\ &= E_c \ell\left(\frac{U_c}{V_c}\right) \\ &= \mu_2(c) \end{aligned}$$

the final equality in the case of quadratic risk, and $\mu_2(c)$ is the second moment defined in (4). It follows that the maximal risk of the OLS estimator is

$$\sup_{c \leq 0} \lim_{n \rightarrow \infty} E_{\alpha_n} \ell(n(\hat{\alpha}_n - \alpha_n)) = \sup_{c \leq 0} \mu_2(c).$$

But this is infinite! The second moment $\mu_2(c)$ is larger than the variance of U_c/V_c , which as shown

in panel (b) of Figures 1 and 2, diverges to infinity as $c \rightarrow -\infty$. Since the maximal risk is unbounded it is not possible to define efficiency in terms of minimizing the maximal risk.

The solution pursued by Phillips and Ploberger (2012) is to restrict the loss function $\ell(u)$ to be bounded, in which case the maximal risk is necessarily finite. However, the fact that the risk is increasing as $c \rightarrow -\infty$ means that the maximal risk will be determined by the extreme values of c . In other words, efficiency improvements for small c will not be captured by a theory which computes maximal risk over unbounded c .

A solution to this dilemma was proposed by Hansen (2013) in the context of LAN models. Instead of defining the maximal risk over all values of c , it can be defined over bounded sets, creating a maximal risk function. Specifically, define the maximal risk function of a sequence of estimators $\tilde{\alpha}_n$ as

$$\rho(C, \tilde{\alpha}, \alpha) = \sup_{C \leq c \leq 0} \lim_{n \rightarrow \infty} E_{\alpha_n} \ell(n(\tilde{\alpha}_n - \alpha_n)).$$

The maximal risk function of the OLS estimator is

$$\rho(C, \hat{\alpha}, \alpha) = \sup_{C \leq c \leq 0} \mu_2(c) = \mu_2(C),$$

the second equality since $\mu_2(c)$ is monotonic in c . [As shown in Figures 1 and 2, both the squared mean and variance are monotonically increasing as c decreases.]

The maximal risk function $\rho(C, \tilde{\alpha}, \alpha)$ can be used to rank the efficiency of estimators. If we have two estimators $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ and we can show that $\rho(C, \tilde{\alpha}_1, \alpha) < \rho(C, \tilde{\alpha}_2, \alpha)$, this means that the maximum risk of $\tilde{\alpha}_1$ is less than that of $\tilde{\alpha}_2$ for $-C \leq c \leq 0$. If this holds for all C then clearly $\tilde{\alpha}_1$ is more efficient than $\tilde{\alpha}_2$.

Furthermore, we can define an estimator $\tilde{\alpha}_n$ of α_n as minimax efficient if its maximal risk function $\rho(C, \tilde{\alpha}, \alpha)$ is the smallest possible for all values of C . Unfortunately this lower bound is unknown, and it is unknown if such an estimator exists.

5 Stein-Type Shrinkage Estimator

In LAN models, Stein-type estimators can achieve efficiency improvements relative to MLE when the estimation dimension is three or greater (Stein (1956, 1981), James and Stein (1961)). The local-to-unity model (1)-(2) only has one parameter and is not LAN, so we should not expect such improvements to hold, but it is intriguing to see what happens.

A Stein-type estimator which shrinks the MLE towards unity is

$$\begin{aligned}\hat{\alpha}_n^* &= 1 + (\hat{\alpha}_n - 1) \left(1 - \frac{s(\hat{\alpha}_n)^2}{(\hat{\alpha}_n - 1)^2} \right)_+ \\ &= \begin{cases} 1 & \text{if } \left| \frac{\hat{\alpha}_n - 1}{s(\hat{\alpha}_n)} \right| \leq 1 \\ \hat{\alpha}_n - \frac{s(\hat{\alpha}_n)^2}{(\hat{\alpha}_n - 1)} & \text{if } \left| \frac{\hat{\alpha}_n - 1}{s(\hat{\alpha}_n)} \right| > 1 \end{cases}\end{aligned}$$

where

$$s(\hat{\alpha}_n) = \left(\frac{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{\alpha}_n y_{t-1})^2}{\sum_{t=1}^n y_{t-1}^2} \right)^{1/2}$$

is the conventional standard error for $\hat{\alpha}_n$. The notation $(a)_+ = 1(a \geq 0)$ is the positive part operator, so that the estimator $\hat{\alpha}_n^*$ takes the ‘‘positive-part’’ form introduced by Baranchik (1964).

The asymptotic distribution of $\hat{\alpha}_n^*$ in the local-to-unity model (1)-(2) is simple to calculate from (3). The maximal risk function is then a function of the asymptotic distribution.

Proposition 1

$$\begin{aligned}n(\hat{\alpha}_n^* - \alpha_n) &\rightarrow_d \left(\frac{U_c}{V_c} + c \right) \left(1 - \frac{V_c}{(U_c + cV_c)^2} \right)_+ - c \\ \rho(C, \hat{\alpha}^*, \alpha) &= \sup_{C \leq c \leq 0} E \left(\left(\frac{U_c}{V_c} + c \right) \left(1 - \frac{V_c}{(U_c + cV_c)^2} \right)_+ - c \right)^2\end{aligned}$$

The asymptotic risk is not a simple function of the moments of (U_c, V_c) , so it cannot be calculated by the exact methods of Theorem 1. Instead, we calculate it by simulation. The results of Section 3 suggest that to obtain accurate results we need to use samples of size $n = 100,000$, and as before, we used 1,000,000 simulation replications.

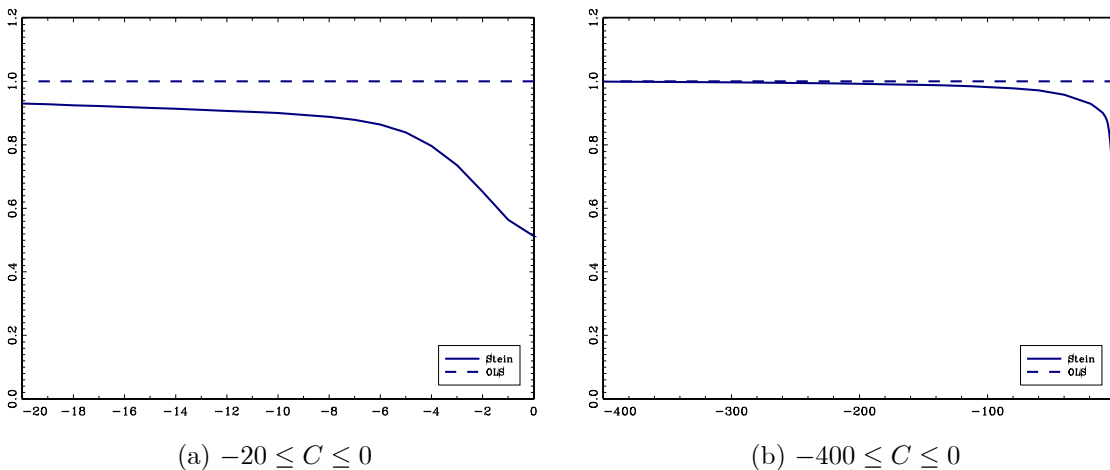
As we are interested in the relative performance of the Stein estimator relative to OLS, we define the relative maximal risk

$$\rho^*(C, \hat{\alpha}^*, \alpha) = \frac{\rho(C, \hat{\alpha}^*, \alpha)}{\rho(C, \hat{\alpha}, \alpha)}.$$

Values less than one indicate improved risk relative to OLS, values over one indicate higher risk than OLS.

The results are presents graphically in Figure 3. The panel on left is shown for C ranging from -20 to 0) and the right panel for C ranging from -400 to 0 . As can be seen, the Stein estimator has uniformly decreased risk relative to OLS. The risk reduction is greatest at $C = 0$. (At $c = 0$, the risk of the Stein estimator is 51% of that of the OLS estimator). The risk reduction remains quite substantial for small C (20% at $C = -4$ and 10% at $C = -10$), but asymptotes to zero. The fact that the relative risk function lies strictly below one for all C below -400 means that there is no value of c for which the Stein estimator does not have lower risk than OLS. Uniformly in the

Figure 3: Relative Asymptotic Risk of Stein Estimator



local-to-unity model, the Stein estimator dominates OLS.

This finding is quite surprising given that this is a one-dimensional problem and classic Stein theory only applies when the dimension is three or higher. It may not be that surprising, however, given that Ploberger (2008) shows that OLS-based unit root tests are not admissible.

It should be emphasized that our finding is numerical; we do not have a formal proof. Given the large number of simulation replications (1 million) and the large range of the local-to-unity parameter explored, the finding appears quite robust. However, based on the numerical evidence alone we cannot exclude the possibility that the relationship will invert for values of C below -400 . Such a numerical exercise does not appear to be fruitful. First, the sample size n would likely need to be increased. We set $n = 100,000$ based on our earlier calculations which showed that this value was needed to obtain good approximations for the mean and variance of the OLS estimator for local-to-unity parameters up to -400 . For values beyond this point this numerical comparison would need to be repeated.

The results of this section are meant to be suggestive, and not guidance for empirical work. We have shown intriguing evidence that a simple shrinkage adjustment can provide major reductions in estimation risk when the local-to-unity parameter is small. This suggests that further research into optimal shrinkage methods could prove fruitful.

6 Conclusion

Many papers have been written about the AR(1) model, and many have used the local-to-unity framework of Chan and Wei (1987) and Phillips (1987). Implementation of the theory typically requires numerical evaluation, and most of the latter uses simulation methods. We have extended earlier work on the exact moments of the unit root model to the local-to-unity framework, and have shown that the moments of the distribution can be easily calculated by numerical integration. Comparing these exact moments with moments from simulated distributions, we have shown that

conventional sample sizes are far too small to provide good approximations. For large local-to-unity parameters, we suggest $n = 100,000$.

We have also explored the theory of efficient estimation in the context of the local-to-unity model. We suggest that the minimax risk should be evaluated locally, as a function of the localizing parameter, and have introduced a simple Stein shrinkage estimator which has lower (numerical) minimax risk than the OLS estimator. This suggests that improvements over OLS are potentially important and feasible.

7 Proof of Theorem 1

The method of proof is a straightforward generalization of the method introduced by Nabeya (1999). It will be useful to start by defining the random variables $(U, V) = (\int_0^1 W dW, \int_0^1 W^2)$, and let $f(u, v)$ denote their joint density function. White (1958) showed that their moment generating function equals

$$\phi(s, t) = E \exp(sU + tV) = \int \exp(su + tv) f(u, v) dudv = e^{-s/2} \left(\cos \sqrt{2t} - s \frac{\sin \sqrt{2t}}{\sqrt{2t}} \right)^{-1/2}.$$

Making the substitutions

$$\sqrt{2t} = i\sqrt{-2t},$$

$$\cos \sqrt{2t} = \cos(i\sqrt{-2t}) = \cosh \sqrt{-2t},$$

and

$$\frac{\tan \sqrt{2t}}{\sqrt{2t}} = \frac{\tan(i\sqrt{-2t})}{i\sqrt{-2t}} = \frac{-i \tan(i\sqrt{-2t})}{\sqrt{-2t}} = \frac{\tanh(\sqrt{-2t})}{\sqrt{-2t}} = \psi(\sqrt{-2t}),$$

we find the alternative expression

$$\phi(s, t) = e^{-s/2} (\cosh \sqrt{-2t})^{-1/2} (1 - s\psi(\sqrt{-2t}))^{-1/2}. \quad (6)$$

Define $U_c^* = \int_0^1 W_c dW_c = U_c + cV_c$. Crump (2008) showed the joint density of (U_c^*, V_c) equals $f_c(u, v) = \exp\left(cu - \frac{c^2 v}{2}\right) f(u, v)$. It follows that their moment generating function equals

$$\begin{aligned} \phi_c(s, t) &= \int \exp(su + tv) f_c(u, v) dudv \\ &= \int \exp\left((s+c)u + \left(t - \frac{c^2}{2}\right)v\right) f(u, v) dudv \\ &= \phi\left(s+c, t - \frac{c^2}{2}\right). \end{aligned} \quad (7)$$

Equation (7) can alternatively be derived from the moment generating function for (U_c, V_c) derived by Phillips (1987). See also Proposition A.1 of Phillips, Magdalinos and Giraitis (2010). It turns

out that the form of expression (7) is convenient for our calculations.

By the binomial expansion,

$$E\left(\frac{U_c}{V_c}\right)^r = E\left(\frac{U_c^*}{V_c} - c\right)^r = \sum_{j=0}^r \binom{r}{j} (-c)^{r-j} E\left(\frac{U_c^*}{V_c}\right)^j. \quad (8)$$

Following Nabeya (1999) and Sawa (1972), the moments in (8) can be expressed as

$$E\left(\frac{U_c^*}{V_c}\right)^j = \frac{1}{(j-1)!} \int_0^\infty t^{j-1} \frac{\partial^j}{\partial s^j} \phi_c(s, -t) \Big|_{s=0} dt. \quad (9)$$

Using (7) and then making the change-of-variables $t = ((z-c)^2 - c^2)/2$ we find that (9) equals

$$\begin{aligned} & \frac{1}{(j-1)!} \int_0^\infty t^{j-1} \frac{\partial^j}{\partial s^j} \phi\left(s+c, -t - \frac{c^2}{2}\right) \Big|_{s=0} dt \\ &= \frac{1}{(j-1)! 2^{j-1}} \int_0^\infty (z-c) ((z-c)^2 - c^2)^{j-1} \frac{\partial^j}{\partial s^j} \phi\left(s+c, -\frac{(z-c)^2}{2}\right) \Big|_{s=0} dz. \end{aligned} \quad (10)$$

Note that this change-of-variables is appropriate when $c \leq 0$ for the transformation is invertible for $t \geq 0$, but it would not be invertible for $c > 0$.

Using (6) and

$$(\cosh(z-c))^{-1/2} = \sqrt{2} e^{(c-z)/2} \left(1 + e^{-2(z-c)}\right)^{-1/2},$$

we can see that

$$\begin{aligned} \phi\left(s+c, -\frac{(z-c)^2}{2}\right) &= e^{-(s+c)/2} (\cosh(z-c))^{-1/2} (1 - (s+c)\psi(z-c))^{-1/2} \\ &= \frac{\sqrt{2}}{e^{z/2} (1 + e^{-2(z-c)})^{1/2}} e^{-s/2} (1 - (s+c)\psi(z-c))^{-1/2}. \end{aligned}$$

Therefore

$$\begin{aligned} \frac{\partial^j}{\partial s^j} \phi\left(s+c, -\frac{(z-c)^2}{2}\right) \Big|_{s=0} &= \frac{\sqrt{2}}{e^{z/2} (1 + e^{-2(z-c)})^{1/2}} \frac{\partial^j}{\partial s^j} \left\{ e^{-s/2} (1 - (s+c)\psi(z-c))^{-1/2} \right\} \Big|_{s=0} \\ &= \frac{2^{1/2-j}}{e^{z/2} (1 + e^{-2(z-c)})^{1/2}} \sum_{\ell=0}^j \binom{j}{\ell} \frac{(-1)^{j-\ell} (2\ell-1)!! \psi(z-c)^\ell}{(1 - c\psi(z-c))^{\frac{1}{2}+\ell}}. \end{aligned}$$

Substituted into (10), and then making the change of variables $z = x/(1 - x)$, we obtain

$$\begin{aligned}
E \left(\frac{U_c^*}{V_c} \right)^j &= \frac{2^{3/2-2j}}{(j-1)!} \int_0^\infty \frac{(z-c) ((z-c)^2 - c^2)^{j-1}}{e^{z/2} (1 + e^{-2(z-c)})^{1/2}} \sum_{\ell=0}^j \binom{j}{\ell} \frac{(-1)^{j-\ell} (2\ell-1)!! \psi(z-c)^\ell}{(1 - c\psi(z-c))^{\frac{1}{2}+\ell}} dz \\
&= \frac{2^{3/2-2j}}{(j-1)!} \int_0^1 \frac{\lambda(x,c) (\lambda(x,c)^2 - c^2)^{j-1}}{(1-x)^2 e^{x/(2-2x)} (1 + e^{-2\lambda(x,c)})^{1/2}} \sum_{\ell=0}^j \binom{j}{\ell} \frac{(-1)^{j-\ell} (2\ell-1)!! \psi(\lambda(x,c))^\ell}{(1 - c\psi(\lambda(x,c)))^{\frac{1}{2}+\ell}} dx \\
&= \int_0^1 g_j(x, c) dx.
\end{aligned}$$

Substituted into (8) we obtain (5). ■

References

- [1] Baranchick, A. (1964): “Multiple regression and estimation of the mean of a multivariate normal distribution,” Technical Report No. 51, Department of Statistics, Stanford University.
- [2] Chan, N.H. and C.Z. Wei (1987): “Asymptotic Inference for Nearly Nonstationary AR(1) Processes,” *Annals of Statistics*, 15, 1050-1063.
- [3] Crump, Richard K. (2008): “Optimal conditional inference in nearly-integrated autoregressive processes,” working paper, Federal Reserve Bank of New York.
- [4] Hájek, J. (1970): “A characterization of limiting distributions of regular estimates,” *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 14, 323-330.,
- [5] Hájek, J. (1972): “Local asymptotic minimax and admissibility in estimation,” *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 175-194.
- [6] Hansen, Bruce E. (2013): “Efficient shrinkage in parametric models,” University of Wisconsin.
- [7] Le Cam, L. (1982): *Asymptotic Methods in Statistical Decision Theory*. New York: Springer-Verlag.
- [8] MacKinnon, James G. (1994): “Approximate asymptotic distribution functions for unit-root and cointegration tests,” *Journal of Business and Economic Statistics*, 12, 167-176.
- [9] Nabeya, Seiji (1999): “Asymptotic moments of some unit root test statistics in the null case,” *Econometric Theory*, 15, 139-149.
- [10] Phillips, Peter C. B. (1987a): “Time series regression with a unit root,” *Econometrica*, 55, 277-302.
- [11] Phillips, Peter C. B. (1987b): “Towards a unified asymptotic theory for autoregression,” *Biometrika*, 74, 535-547.

- [12] Phillips, Peter C. B. (2012a): “Fokelore theorems, implicit maps, and indirect inference,” *Econometrica*, 80, 425-454.
- [13] Phillips, Peter C. B. (2012b): “On confidence intervals for autoregressive roots and predictive regression,” Cowles Foundation Discussion Paper #1879.
- [14] Phillips, Peter C. B. and Pierre Perron (1988): “Testing for a Unit Root in Time Series Regression”, *Biometrika*, 75, 335-346.
- [15] Phillips, Peter C. B., Tassos Magdalinos, and Liudas Giraitis (2010): “Smoothing local-to-moderate unit root theory”, *Journal of Econometrics*, 158, 274-279.
- [16] Ploberger, Werner (2008): “Admissible and non-admissible tests in unit-root-like situations,” *Econometric Theory*, 24, 15-42.
- [17] Ploberger, Werner and Peter C.B. Phillips (2012): “Optimal estimation under nonstandard conditions,” *Journal of Econometrics*, 169, 258-265.
- [18] James, W. and Charles M. Stein (1961): “Estimation with quadratic loss,” *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 361-380.
- [19] Sawa, Takamitsu (1972): “Finite sample properties of k-class estimators,” *Econometrica*, 40, 653-680.
- [20] 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.
- [21] Shenton, L.R. and H. D. Vinod (1995): “Closed forms for asymptotic bias and variance in autoregressive models with unit roots,” *Journal of Computational and Applied Mathematics*, 61, 231-243.
- [22] Stein, Charles M. (1956): “Inadmissibility of the usual estimator for the mean of a multivariate normal distribution,” *Proc. Third Berkeley Symp. Math. Statist. Probab.*, 1, 197-206.
- [23] Stein, Charles M. (1981): “Estimation of the mean of a multivariate normal distribution,” *Annals of Statistics*, 9, 1135-1151.
- [24] Stock, James H. (1991): “Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series,” *Journal of Monetary Economics*, 28, 435-459.
- [25] van der Vaart, Aad W. (1998): *Asymptotic Statistics*, New York: Cambridge University Press.
- [26] White, J.S. (1958): “The limiting distribution of the serial correlation coefficient in the explosive case,” *Annals of Mathematical Statistics*, 29, 1188-1197.
- [27] White, J.S. (1961): “Asymptotic expansions for the mean and variance of the serial correlation coefficient,” *Biometrika*, 48, 85-94.