ON CONFIDENCE INTERVALS FOR AUTOREGRESSIVE ROOTS AND PREDICTIVE REGRESSION

By

Peter C.B. Phillips

September 2012

COWLES FOUNDATION DISCUSSION PAPER NO. 1879

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

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

On Confidence Intervals for Autoregressive Roots and Predictive Regression*

Peter C. B. Phillips
Yale University, University of Auckland,
Singapore Management University & University of Southampton

August 2012

Abstract

A prominent use of local to unity limit theory in applied work is the construction of confidence intervals for autogressive roots through inversion of the ADF t statistic associated with a unit root test, as suggested in Stock (1991). Such confidence intervals are valid when the true model has an autoregressive root that is local to unity ($\rho = 1 + \frac{c}{n}$) but are invalid at the limits of the domain of definition of the localizing coefficient c because of a failure in tightness and the escape of probability mass. Consideration of the boundary case shows that these confidence intervals are invalid for stationary autoregression where they manifest locational bias and width distortion. In particular, the coverage probability of these intervals tends to zero as $c \to -\infty$, and the width of the intervals exceeds the width of intervals constructed in the usual way under stationarity. Some implications of these results for predictive regression tests are explored. It is shown that when the regressor has autoregressive coefficient $|\rho| < 1$ and the sample size $n \to \infty$, the Campbell and Yogo (2006) confidence intervals for the regression coefficient have zero coverage probability asymptotically and their predictive test statistic Q erroneously indicates predictability with probability approaching unity when the null of no predictability holds. These results have obvious implications for empirical practice.

Keywords: Autoregressive root, Confidence belt, Confidence interval, Coverage probability, Local to unity, Localizing coefficient, Predictive regression, Tightness.

JEL classification: C22

*Support is acknowledged from the NSF under Grant No. SES-0956687.

1 Introduction

A primary reason for the introduction of local to unity limit theory was to develop asymptotic power functions for unit root test procedures. This theory facilitated comparisons between different test procedures. The limit theory also provided convenient approximations to the distributions of estimators and tests for models with an autoregressive parameter in the vicinity of unity of the form $\rho = 1 + \frac{c}{n}$, giving approximations that depend on the value of the localizing coefficient c. A prominent application of this theory in empirical work is the construction of confidence intervals for autogressive roots through the inversion of unit root test statistics. The approach was suggested in Stock (1991). It has been recommended and used in later work on confidence interval construction for autoregressive roots (Elliott and Stock, 2001) and in predictive regression tests with persistent regressors (Cavanagh, Elliott and Stock, 1995; Campbell and Yogo, 2006).

Simulations in Hansen (1999) revealed that the inversion procedure proposed by Stock (1991) performed well for ρ in the immediate vicinity of unity but poorly for stationary ρ values distant from unity. The limit theory in Phillips (1987) shows that appropriately centred statistics have limits as $c \to -\infty$ that correspond to the stationary limit theory for fixed $|\rho| < 1$, which suggests that inversion of appropriately defined test statistics should lead to confidence intervals that correspond to those that apply for the stationary region and are based on stationary asymptotics. Mikusheva (2007) recently confirmed this supposition by demonstrating that confidence intervals obtained in this way are valid uniformly for $|\rho| \leq 1$.

On the other hand, inversion procedures based on unit root tests, such as those in Stock (1991) and Elliott and Stock (1991) are not valid uniformly for $|\rho| \leq 1$. The reason for and extent of the failure has not been explored in the existing literature. Since these procedures are recommended in applications and form the basis of empirical work, it is important to understand their properties when they are applied to data with stationary regressors, as they may very well be in predictive regressions of the type considered in Campbell and Yogo (2006).

The present paper contributes to this literature by providing an asymptotic analysis of the properties of confidence intervals obtained by the inversion procedure applied to unit root tests. It is shown that such confidence intervals are invalid at the limits of the domain of definition of the localizing coefficient. In particular, consideration of the boundary case shows that these confidence intervals manifest severe locational bias and width distortion. The asymptotic coverage probability of the intervals is zero in the stationary case as $c \to -\infty$ even though the intervals are wider than those constructed in the usual way under stationarity. Similar consequences are shown to follow when these procedures are used in predictive regression tests of the type considered in Campbell and Yogo (2006). In particular, the commonly used Q test is biased towards accepting predictability and associated confidence intervals for the regressor coefficient asympototically have zero coverage probability in the stationary regressor case. These results have potentially important empirical consequences for practical work given that degrees of persistence in predictive regressors are determined very imprecisely and tests are needed that are robust for a wide range of such regressors. Some alternative approaches that do achieve robustness are discussed in the paper.

2 Boundary Behavior in Confidence Intervals based on Unit Root Test Inversion

It will be sufficient for our purpose to consider the simple AR(1) model

$$x_t = \rho x_{t-1} + u_t, \quad \text{with } u_t \sim iid\left(0, \sigma^2\right), \tag{1}$$

initialized at $x_0 = 0$, with least squares regression estimate $\hat{\rho}$ of ρ and unit root t test $t_{\hat{\rho}} = \frac{\hat{\rho}-1}{\sigma_{\hat{\rho}}}$, with $\sigma_{\hat{\rho}}^2 = \hat{\sigma}^2 / \sum_{t=1}^n x_{t-1}^2$ and $\hat{\sigma}^2 = n^{-1} \sum_{t=1}^n \left(x_t - \hat{\rho} x_{t-1}\right)^2$. In the local to unity case $\rho = \rho_n = 1 + \frac{c}{n}$, and we have the following limit theory for any fixed localizing coefficient c (Phillips, 1987)

$$t_{\hat{\rho}} = \frac{n\left(\hat{\rho} - \rho_n\right) + n\left(\rho_n - 1\right)}{\left\{\hat{\sigma}^2 / \left(n^{-2} \sum_{t=1}^n x_{t-1}^2\right)\right\}^{1/2}} \Longrightarrow \frac{\int J_c dW}{\left(\int J(r)^2 dr\right)^{1/2}} + c\left(\int J_c(r)^2 dr\right)^{1/2} := \tau_c, \tag{2}$$

where $J_c(r) = \int_0^r e^{c(r-s)} dW(s)$ is a linear diffusion, W is standard Brownian motion, and all integrals are over the interval $[0,1]$.

2.1 Confidence belt asymptotics

The method of confidence belts suggested in Stock (1991) proceeds as follows: compute a unit root test statistic such as $t_{\hat{\rho}}$ and use the known asymptotic distribution of that statistic under the alternative, as given in (2) above, to construct a confidence interval for c by inversion of the test. The confidence belts provide a graphical method of performing this operation and can be tabulated and complemented with interpolation to achieve a reasonable degree of accuracy for implementation in practice. Since (2) is the appropriate asymptotic distribution of $t_{\hat{\rho}}$ under the local alternative $\rho_n = 1 + \frac{c}{n}$ to a unit root for all c, it may not be immediately obvious why the procedure fails to deliver confidence intervals with good properties for stationary ρ. The fact that it is not so obvious perhaps explains why the matter seems to have passed unnoticed and unanalyzed for so many years except for a brief observation on poor simulation performance in Hansen (1999) and Mikusheva's (2007, p. 1422) remark to the effect that a modified version of this inversion procedure produces uniform confidence intervals.

To explain the failure we need to develop the asymptotics in (2) further, focusing on a more detailed analysis of behavior at the lower limit of the domain of definition, viz. $c \to -\infty$, which effectively captures the stationary case as shown in Phillips (1987). Upon simple manipulation of the limit (2) we have

$$\tau_c = \frac{(-2c)^{1/2} \int J_c(r)dW(r)}{\left((-2c) \int J_c(r)^2 dr\right)^{1/2}} - \frac{|c|^{1/2}}{2^{1/2}} \left((-2c) \int J_c(r)^2 dr\right)^{1/2}. \tag{3}$$

Phillips (1987) proved that the (centred) first component of (3) satisfies

$$\lambda_c := \frac{\int J_c(r)dW(r)}{\left(\int J_c(r)^2 dr\right)^{1/2}} \Rightarrow N(0,1) \quad \text{as } c \to -\infty,$$

and gave the following results

$$J_c(1)^2 = 1 + 2c \int J_c(r)^2 dr + 2 \int J_c(r)dW(r), \tag{4}$$

$$(-2c)^{1/2} \int J_c(r)dW(r) \Rightarrow \xi = N(0,1), \tag{5}$$

$$(-2c)^{1/2} J_c(1) \Rightarrow \eta = N(0,1), \tag{6}$$

which are useful in developing the theory below. Importantly, (3) and (5) imply that the sequence τ_c (unlike the sequence λ_c) is not tight as $c \to -\infty$. The failure of tightness leads to an escape of probability mass in the limit and this has material consequences in terms of the properties of induced confidence intervals and tests of predictability that are founded on unit root test statistics like (2).

In what follows it is convenient to use a suitably extended probability space where the convergences (5)-(6) may be treated as convergence in probability. Using an expansion of the moment generating function given in

Phillips (1987) we may then write in expansion form

$$(-2c)^{1/2}\int J_c(r)dW(r)=\xi+O_p\left(|c|^{-1/2}\right), \tag{7}$$

from which $\int J_c(r)dW(r)=\frac{\xi}{(-2c)^{1/2}}+O_p\left(|c|^{-1}\right)$. By straightforward computation

$$\begin{aligned}
&\mathbb{E}\left[\left\{(-2c)^{1/2}\int J_c(r)dW(r)\right\}\left\{(-2c)^{1/2}J_c(1)\right\}\right]\\
=\;&\mathbb{E}\left[\left\{(-2c)\int_0^1\int_0^1 J_c(r)e^{c(1-s)}dW(s)dW(r)\right\}\right]\\
=\;&(-2c)\int_0^1\mathbb{E}\left[J_c(r)\right]e^{c(1-r)}dr=0,
\end{aligned}$$

for all c, so that the limit variates ξ and η that appear in (5)-(6) are independent. In this expanded probability space we have

$$(-2c)\int J_c(r)^2dr=1+2\int J_c(r)dW(r)-J_c(1)^2$$

$$=1+2\frac{\xi}{(-2c)^{1/2}}\left\{1+O_p\left(|c|^{-1/2}\right)\right\}-\frac{\eta^2}{-2c}\left\{1+O_p\left(|c|^{-1/2}\right)\right\}. \tag{8}$$

As shown in the Appendix, using (8) and binomial expansion for large $|c|$ in (3) produces the following asymptotic representation of the t ratio τ_c

$$\tau_c=\left\{\frac{1}{2}\xi-\frac{|c|^{1/2}}{2^{1/2}}-\frac{1}{2^{3/2}|c|^{1/2}}-\frac{\varphi}{2^{5/2}|c|^{1/2}}+O\left(\frac{1}{|c|}\right)\right\}\left\{1+O_p\left(|c|^{-1/2}\right)\right\}, \tag{9}$$

where

$$\varphi=3\left(\xi^2-1\right)-\left(\eta^2-1\right)$$

is a weighted linear combination of chi-squared variates, each with unit degree of freedom, and centred to have zero mean. Importantly, note that there is a non-random term $-\frac{|c|^{1/2}}{2^{1/2}}-\frac{1}{2^{3/2}|c|^{1/2}}$ in (9) and the leading distributional term is no longer $\xi\equiv N(0,1)$ but $\frac{1}{2}\xi=N\left(0,\frac{1}{4}\right)$. This change arises because we are dealing with an expansion of the miscentred test statistic $t_{\hat\rho}=\frac{\hat\rho-1}{\sigma_{\hat\rho}}$ not the correctly centred t ratio $t_{\hat\rho,\rho}:=\frac{\hat\rho-\rho_n}{\sigma_{\hat\rho}}\Rightarrow\xi$.

The random variable ξ occurs not only in the limit of the numerator (5) but also in the expansion of the denominator standard error $\sigma_{\hat\rho}$ (as is apparent from (4)), thereby contributing to the dependence between numerator and denominator of the limit variate in the near integrated case. Hence, we end up with a bias term in the approximating (normal) distribution based on the first three terms of (9):

$$\tau_c \sim N\left(-\frac{|c|^{1/2}}{2^{1/2}} - \frac{1}{2^{3/2}|c|^{1/2}}, \frac{1}{4}\right) + O_p\left(\frac{1}{|c|^{1/2}}\right). \tag{10}$$

The distribution (9) delivers approximate percentiles of τ_c based on the local to unity limit theory of τ_c when $|c|$ is large. These percentile functions are needed in the inversion process and produce the confidence belts used in the confidence interval construction.

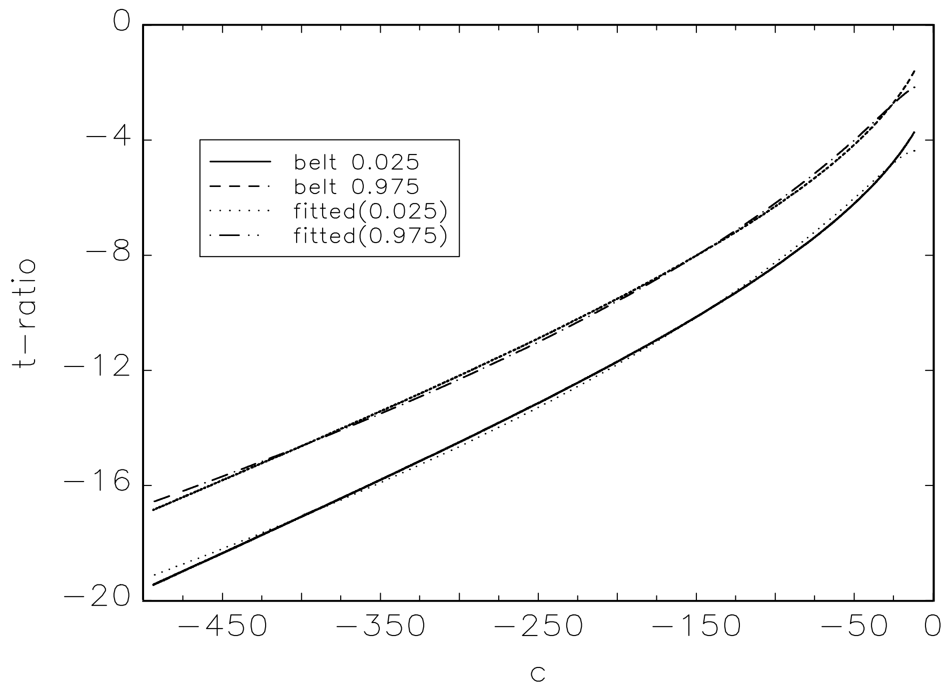

Fig. 1: Confidence belts (at levels 2.5% and 97.5%) based on local to unity ($\rho = 1 + \frac{c}{n}$) limit theory for the t ratio $t_{\hat{\rho}} = \frac{\hat{\rho}-1}{\sigma_{\hat{\rho}}}$, shown against fitted regression curves of the form $f(c) = \alpha_0 + \alpha_1 |c|^{1/2} + \alpha_2 |c|^{-1/2}$, as suggested by the asymptotic functional form of the belts given in (13).

To illustrate, the $2\frac{1}{2}$% and $97\frac{1}{2}$% confidence belts produced from τ_c when the generating mechanism is a local to unity process are shown in Fig. 1 over the very large region $-450 < c < -15$. These confidence belts were computed using 100,000 replications and a grid of 20,000 values of c using the model (1) with Gaussian errors and a sample size of $n = 1,000$. They can be used to execute the inversion of the t ratio τ_c to produce an induced confidence interval for c for any given value of the test statistic $t_{\hat{\rho}}$, as suggested in Stock (1991), where the graphs and tables are given over a smaller region $-40 < c < 5$ that nests the origin. (Table 1 of the implementation paper of Campbell and Yogo (2005) gives the belts and implied confidence sets for c over the slightly wider region $-67 < c < 5$.) In addition to the two confidence belts, Fig. 1 also shows least squares regression curves $\tilde{\alpha}_0 + \tilde{\alpha}_2 |c|^{1/2} + \tilde{\alpha}_3 |c|^{-1/2}$, showing the good conformity of this asymptotic functional form to the two belts over the broad region $-450 < c < -15$.

2.2 Induced confidence intervals for c and ρ

Using confidence belts obtained numerically in the manner just described, Stock (1991) suggested that a $100(1-\alpha)\%$ confidence set can be constructed as $S(\tau) = \left\{c : f_{L,\alpha/2}(c) \leq \tau \leq f_{U,\alpha/2}(c)\right\}$ where $f_{L,\alpha/2}(c)$ and $f_{U,\alpha/2}(c)$ are the lower $\alpha/2$ and upper $1-\alpha/2$ percentiles of τ_c as a function of c. Taking $f_{L,\alpha/2}(c)$ and $f_{U,\alpha/2}(c)$ to be strictly increasing in c, Stock (1991) numerically inverted the test critical values in the belts to yield the confidence interval

$$\left\{c : f_{U,\alpha/2}^{-1}(\hat{\tau}) \leq c \leq f_{L,\alpha/2}^{-1}(\hat{\tau})\right\} \tag{11}$$

for c, where (11) is calculated for some given observed t ratio $\hat{\tau}$. Importantly, the confidence interval (11) is based on the limit theory for the t ratio τ_c under the assumption of model (1) with a local to unity coeffficient of the explicit form $\rho = 1 + \frac{c}{n}$. Associated with (11) we have the implied confidence interval for ρ, viz,

$$\left\{\rho = 1 + \frac{c}{n} : f_{U,\alpha/2}^{-1}(\hat{\tau}) \leq c \leq f_{L,\alpha/2}^{-1}(\hat{\tau})\right\}. \tag{12}$$

If this confidence interval were uniformly valid in the sense of Mikusheva (2007) then it would apply for all c, including the limiting case where $c \to -\infty$. But uniformity fails because the distribution on which this confidence interval is based is the local to unity asymptotic theory for a unit root test, which is miscentred as $c \to -\infty$. This distribution produces the limits $f_{U,\alpha/2}^{-1}(\hat{\tau})$ and $f_{L,\alpha/2}^{-1}(\hat{\tau})$ in (11) and these are obtained by reading off the values of c that correspond to the observed $\hat{\tau}$ in the calculated confidence belts (computed using the local to unity limit theory, as we have done in Fig. 1).

This process amounts to solving an equation for c based on the form of the confidence belts, for some given $\hat{\tau}$. As the argument above leading to (10) shows, the distribution that produces the confidence belts is approximately normal with mean $-\frac{|c|^{1/2}}{2^{1/2}}-\frac{1}{2^{3/2}|c|^{1/2}}$ and variance $\frac{1}{4}$. This distribution diverges as $c\to-\infty$ and the divergence is mirrored in the behavior of the confidence belts shown in Fig. 1. The reason for the divergence is that the unit root test statistic $t_{\hat{\rho}}$ is miscentred and has a divergent component – the second term of (3). This miscentering ensures, of course, that the unit root test has full power of unity in the limit as $c\to-\infty$ (Phillips, 1987). But it also means that the sequence of distributions is not tight and that the stationary case (when $c\to-\infty$) relies on a divergent distribution. Moreover, the manner of the divergence is nonlinear, as reflected in the mean of the approximating distribution (10) which is nonlinear in c. This failure in tightness and the nonlinearity in c ends up distorting the form and location of the induced confidence intervals obtained from the confidence belts for large c. This phenomenon will become clearer in the argument that follows.

The confidence belts have the explicit asymptotic form

$$\begin{aligned}
&\left\{f_{L,\alpha/2}\left(c\right),f_{U,\alpha/2}\left(c\right)\right\}\\
=&\left\{-\frac{|c|^{1/2}}{2^{1/2}}-\frac{1}{2^{3/2}\left|c\right|^{1/2}}-z_{\alpha/2}\sqrt{\frac{1}{4}},-\frac{|c|^{1/2}}{2^{1/2}}-\frac{1}{2^{3/2}\left|c\right|^{1/2}}+z_{\alpha/2}\sqrt{\frac{1}{4}}\right\}\\
=&\left\{-\frac{|c|^{1/2}}{2^{1/2}}-\frac{1}{2^{3/2}\left|c\right|^{1/2}}-\frac{z_{\alpha/2}}{2},-\frac{|c|^{1/2}}{2^{1/2}}-\frac{1}{2^{3/2}\left|c\right|^{1/2}}+\frac{z_{\alpha/2}}{2}\right\}, \qquad (13)
\end{aligned}$$

where $z_{\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution $N\left(0,1\right)$. Note that this interval has length $2\times\frac{z_{\alpha/2}}{2}=1.96$ for a 95% interval and this length conforms with the vertical distance between the belts shown in Fig. 1. We can proceed to derive the length of the induced confidence intervals of c (and for ρ) and the coverage probability $\mathbb{P}\left\{f_{U,\alpha/2}^{-1}\left(\hat{\tau}\right)\leq c\leq f_{L,\alpha/2}^{-1}\left(\hat{\tau}\right)\right\}$ of the interval when the true ρ is fixed and stationary, i.e., $|\rho|<1$, by considering the corresponding limits of these quantities at the boundary of the local to unity limit theory when $c\to-\infty$.

The asymptotic approximation ((10) that is used in constructing the confidence interval by inversion can be written as

$$\tau_c=-\frac{|c|^{1/2}}{2^{1/2}}-\frac{1}{2^{3/2}\left|c\right|^{1/2}}-\frac{\mathcal{Z}}{2}+O_p\left(\frac{1}{|c|^{1/2}}\right),\quad \text{with } \mathcal{Z}\equiv N\left(0,1\right). \qquad (14)$$

More specifically for the lower limit with percentile $\alpha/2$ (the curve furthest from the origin in Fig. 1 in the 2.5% case)

$$f_{L,\alpha/2}(\hat{\tau}) \sim -\frac{|c|^{1/2}}{2^{1/2}} - \frac{1}{2^{3/2}|c|^{1/2}} - \frac{\mathcal{Z}_U}{2},$$

and for the upper limit with percentile $1-\alpha/2$ (the curve closest to the origin in Fig. 1 in the 97.5% case)

$$f_{U,\alpha/2}(\hat{\tau}) \sim -\frac{|c|^{1/2}}{2^{1/2}} - \frac{1}{2^{3/2}|c|^{1/2}} - \frac{\mathcal{Z}_L}{2},$$

where $\mathcal{Z}_U = z_{\alpha/2}$ and $\mathcal{Z}_L = -z_{\alpha/2}$ are the upper and lower symmetric $\alpha/2$ percentiles of $\mathcal{Z}$. Inverting equation (14) for c we find that

$$|c|^{1/2} + \frac{1}{2|c|^{1/2}} = -2^{1/2}\left\{\tau_c + \frac{\mathcal{Z}}{2}\right\} + O_p\left(\frac{1}{|c|^{1/2}}\right). \qquad (15)$$

Solving (14) for $\tau_c = \hat{\tau}$ at these percentiles we have the following expressions[1] for the lower and upper limits $c_L = f^{-1}_{U,\alpha/2}(\hat{\tau})$ and $c_U = f^{-1}_{L,\alpha/2}(\hat{\tau})$:

$$\begin{aligned} |c_L|^{1/2} + \frac{1}{2|c_L|^{1/2}} &= 2^{1/2}\left\{-\hat{\tau} - \frac{\mathcal{Z}_L}{2}\right\} + O_p\left(\frac{1}{|c|^{1/2}}\right), \\ |c_U|^{1/2} + \frac{1}{2|c_U|^{1/2}} &= 2^{1/2}\left\{-\hat{\tau} - \frac{\mathcal{Z}_U}{2}\right\} + O_p\left(\frac{1}{|c|^{1/2}}\right). \end{aligned}$$

Hence

$$|c_L| + \frac{1}{4|c_L|} + 1 \simeq 2\left\{-\hat{\tau} - \frac{\mathcal{Z}_L}{2}\right\}^2, \quad |c_U| + \frac{1}{4|c_U|} + 1 \simeq 2\left\{-\hat{\tau} - \frac{\mathcal{Z}_U}{2}\right\}^2,$$

so that, up to an error of $O\left(|c|^{-1/2}\right)$ and using the fact that $c = -|c|$ for all large $|c|$, we have[2]

$$\begin{aligned} c_L &= -2\left\{\hat{\tau}^2 + \mathcal{Z}_L\hat{\tau} + \frac{\mathcal{Z}_L^2}{4}\right\} - 1 = -2\left(\hat{\tau}^2 + \mathcal{Z}_L\hat{\tau}\right) - \frac{\mathcal{Z}_L^2}{2} - 1, \quad (16) \\ c_U &= -2\left\{\hat{\tau}^2 + \mathcal{Z}_U\hat{\tau} + \frac{\mathcal{Z}_U^2}{4}\right\} - 1 = -2\left(\hat{\tau}^2 + \mathcal{Z}_U\hat{\tau}\right) - \frac{\mathcal{Z}_U^2}{2} - 1. \quad (17) \end{aligned}$$

[1] Note that for large $|c|$ we have $c = -|c|$ and $\hat{\tau} < 0$. Also $-\mathcal{Z}_L > -\mathcal{Z}_U$ and so $-\hat{\tau} - \frac{\mathcal{Z}_L}{2} > -\hat{\tau} - \frac{\mathcal{Z}_U}{2}$.

[2] For large $|c|$, $\hat{\tau} < 0$, and $\mathcal{Z}_L = -z_{\delta/2} < 0$ while $\mathcal{Z}_U = z_{\delta/2} > 0$. So $\hat{\tau}^2 + \tau\mathcal{Z}_L > \hat{\tau}^2 + \tau\mathcal{Z}_U$ and $c_L < c_U$.

The length of the CI for c is therefore

$$|c_L - c_U| = 2\left|\left(\mathcal{Z}_U - \mathcal{Z}_L\right)\hat{\tau} + \frac{1}{4}\left(\mathcal{Z}_U^2 - \mathcal{Z}_L^2\right)\right|, \tag{18}$$

whose behavior depends on that of the t ratio $\hat{\tau}$.

2.3 Properties in the stationary case

We are interested in the properties of this confidence interval construction at the limits of the domain of definition of the local to unity model corresponding to the stationary case. In that case, under a model with fixed $|\rho| < 1$ we have the following limits for the correctly centred statistics

$$\frac{\sqrt{n}\left(\hat{\rho} - \rho\right)}{\sqrt{1 - \rho^2}} \Rightarrow N\left(0, 1\right), \; t_{\hat{\rho},\rho} = \frac{\hat{\rho} - \rho}{\sigma_{\hat{\rho}}} \Rightarrow N\left(0, 1\right),$$

which results hold for all fixed $|\rho| < 1$ as well as uniformly over all $|\rho| < 1$ for which $n\left(\rho - 1\right) \to 0$, as shown in Giraitis and Phillips (2006). Next, take a probability space for which the convergences apply in probability, so that for fixed $|\rho| < 1$

$$t_{\hat{\rho},\rho} = \frac{\hat{\rho} - \rho}{\sigma_{\hat{\rho}}} = \zeta + o_p\left(1\right), \quad \zeta \equiv N\left(0, 1\right), \tag{19}$$

and then

$$\begin{aligned} t_{\hat{\rho}} &= \frac{\hat{\rho} - 1}{\sigma_{\hat{\rho}}} = \frac{\hat{\rho} - \rho + (\rho - 1)}{\sigma_{\hat{\rho}}} = t_{\hat{\rho},\rho} + \frac{\sqrt{n}\left(\rho - 1\right)}{\left\{\hat{\sigma}^2 / \left(\frac{1}{n}\sum_{t=1}^n x_{t-1}^2\right)\right\}^{1/2}} \\ &= \left\{\zeta + o_p\left(1\right)\right\} + \frac{\sqrt{n}\left(\rho - 1\right)}{\left\{\left(1 - \rho^2\right) + o_p\left(1\right)\right\}^{1/2}} \\ &= -\sqrt{n}\left(\frac{1 - \rho}{1 + \rho}\right)^{1/2} + \zeta + o_p\left(1\right) = -\sqrt{n}\left(\frac{1 - \rho}{1 + \rho}\right)^{1/2} + O_p\left(1\right). \end{aligned} \tag{20}$$

Setting

$$\hat{\tau} = t_{\hat{\rho}} = -\sqrt{n}\left(\frac{1 - \rho}{1 + \rho}\right)^{1/2} + \zeta + o_p\left(1\right) \tag{21}$$

in (18) we have

$$|c_L - c_U| = 2\left|\left(\mathcal{Z}_U - \mathcal{Z}_L\right)\left(-\sqrt{n}\left(\frac{1 - \rho}{1 + \rho}\right)^{1/2} + \zeta + o_p\left(1\right)\right) + \frac{1}{4}\left(\mathcal{Z}_U^2 - \mathcal{Z}_L^2\right)\right|. \tag{22}$$

The length of the confidence interval $[\rho_L, \rho_U]$ for ρ that is implied by inversion from an assumed local to unity model $\rho = 1 + \frac{c}{n}$ is $|\rho_L - \rho_U| = \left|\frac{c_L - c_U}{n}\right|$. When the true ρ is fixed and $|\rho| < 1$ we deduce from (22) that

$$
\begin{aligned}
|\rho_L - \rho_U| &= 2\left|(\mathcal{Z}_U - \mathcal{Z}_L)\left(-\frac{1}{\sqrt{n}}\left(\frac{1-\rho}{1+\rho}\right)^{1/2} + \frac{\zeta}{n} + o_p\left(n^{-1}\right)\right) + \frac{1}{4n}\left(\mathcal{Z}_U^2 - \mathcal{Z}_L^2\right)\right| \\
&= \frac{2}{\sqrt{n}}\left(\frac{1-\rho}{1+\rho}\right)^{1/2} |\mathcal{Z}_U - \mathcal{Z}_L| \\
&\quad \times \left|1 - \left(\frac{1+\rho}{1-\rho}\right)^{1/2}\left[\frac{\zeta}{\sqrt{n}} + \frac{1}{4\sqrt{n}}\left(\mathcal{Z}_U^2 - \mathcal{Z}_L^2\right)\right] + o_p\left(n^{-1/2}\right)\right| \\
&= \frac{2}{\sqrt{n}}\left(\frac{1-\rho}{1+\rho}\right)^{1/2} |\mathcal{Z}_U - \mathcal{Z}_L| \left\{1 + O_p\left(\frac{1}{\sqrt{n}}\right)\right\}.
\end{aligned}
$$

Hence, the implied confidence interval for ρ has the following average length up to an error of $O_p\left(n^{-1}\right)$

$$
\frac{2}{\sqrt{n}}\left(\frac{1-\rho}{1+\rho}\right)^{1/2} |\mathcal{Z}_U - \mathcal{Z}_L| = \frac{4}{\sqrt{n}}\left(\frac{1-\rho}{1+\rho}\right)^{1/2} z_{\alpha/2}, \tag{23}
$$

when $\mathcal{Z}_L = -z_{\alpha/2}$ and $\mathcal{Z}_U = z_{\alpha/2}$ for a central confidence interval.

Under the assumed stationary (true) model with $|\rho| < 1$ the standard confidence interval based on $t_{\hat{\rho},\rho}$ is

$$
\left\{\hat{\rho} \pm z_{\alpha/2}\sigma_{\hat{\rho}}\right\} = \left\{\hat{\rho} \pm z_{\alpha/2}\left(\frac{1-\rho^2}{n} + O_p\left(\frac{1}{n^{3/2}}\right)\right)^{1/2}\right\}, \tag{24}
$$

whose average length up to an error of $O_p\left(n^{-1}\right)$ is

$$
\begin{aligned}
2z_{\alpha/2}\left(\frac{1-\rho^2}{n}\right)^{1/2} &= 2z_{\alpha/2}\frac{(1-\rho)^{1/2}(1+\rho)^{1/2}}{\sqrt{n}} \\
= \frac{2}{\sqrt{n}}\left(\frac{1-\rho}{1+\rho}\right)^{1/2}(1+\rho)\, z_{\alpha/2} &\leq \frac{4}{\sqrt{n}}\left(\frac{1-\rho}{1+\rho}\right)^{1/2} z_{\alpha/2}.
\end{aligned} \tag{25}
$$

It follows that the implied confidence interval obtained by inversion of the local to unity confidence belts for a localizing parameter c based on the unit root test statistic $t_{\hat{\rho}} = \frac{\hat{\rho}-1}{\sigma_{\hat{\rho}}}$ is not equivalent to that of the standard confidence interval that applies in the stationary case when $c \to -\infty$. Hence, this confidence interval is not uniform over ρ.

As is apparent from the inequality (25), the implied confidence interval from inversion of local to unity limit theory has length that is greater than that of the standard interval. The length exceeds that of the standard interval by the factor $2/(1+\rho)$, which exceeds 2 for $\rho < 0$. As $\rho \to 1$, on the other hand, the ratio approaches unity and the lengths of the two confidence intervals are the same, as the limit theory in Phillips (1987) and Giraitis and Phillips (2006) predicts. In effect, the lower boundary of the local to unity domain as $c \to -\infty$ corresponds to the upper boundary (of the mildly integrated region) for which $n(1-\rho) \to \infty$.

Next consider the coverage probability of these intervals. From (16) and (17), the interval for c is

$$[c_L, c_U] = \left[-2\left(\hat{\tau}^2 + \mathcal{Z}_L\hat{\tau}\right) - \left(\frac{\mathcal{Z}_L^2}{2}+1\right), -2\left(\hat{\tau}^2 + \mathcal{Z}_U\hat{\tau}\right) - \left(\frac{\mathcal{Z}_U^2}{2}+1\right)\right], \tag{26}$$

which implies the following confidence interval for ρ

$$[\rho_L, \rho_U] = \left[1 - \frac{2}{n}\left(\hat{\tau}^2 + \mathcal{Z}_L\hat{\tau}\right) - \frac{1}{n}\left(\frac{\mathcal{Z}_L^2}{2}+1\right), 1 - \frac{2}{n}\left(\hat{\tau}^2 + \mathcal{Z}_U\hat{\tau}\right) - \frac{1}{n}\left(\frac{\mathcal{Z}_U^2}{2}+1\right)\right]. \tag{27}$$

The required coverage probability is $\mathbb{P}\{\rho \in [\rho_L, \rho_U]\}$. We find the form taken by this probability in the stationary case $|\rho| < 1$. From (20), we know that $\hat{\tau}$ has the limit behavior $\hat{\tau} = -\sqrt{n}A_\rho^{1/2} + \zeta + o_p(1)$, where $A_\rho = \frac{1-\rho}{1+\rho}$. Hence

$$\begin{aligned}\hat{\tau}^2 + \mathcal{Z}\hat{\tau} &= \left\{-\sqrt{n}A_\rho^{1/2} + \zeta\right\}^2 + \mathcal{Z}\left\{-\sqrt{n}A_\rho^{1/2} + \zeta\right\} + o_p(1) \\ &= nA_\rho - (2\zeta + \mathcal{Z})\sqrt{n}A_\rho^{1/2} + O_p(1).\end{aligned}$$

Then, up to $O_p\left(n^{-1/2}\right)$, the interval (27) has the form

$$[\rho_L, \rho_U] = \left[1 - 2A_\rho + \frac{2\zeta + \mathcal{Z}_L}{\sqrt{n}}A_\rho^{1/2}, 1 - 2A_\rho + \frac{2\zeta + \mathcal{Z}_U}{\sqrt{n}}A_\rho^{1/2}\right], \tag{28}$$

and we can compute the coverage probability using the distribution of $\zeta = N(0,1)$. Setting $\mathcal{Z}_L = -z_{\alpha/2}$ and $\mathcal{Z}_U = z_{\alpha/2}$ for the usual symmetric interval, we have

$$\begin{aligned}&\mathbb{P}_\zeta\{\rho \in [\rho_L, \rho_U]\} \\ =\ &\mathbb{P}_\zeta\left\{1 - 2A_\rho + \frac{2\zeta - z_{\alpha/2}}{\sqrt{n}}A_\rho^{1/2} \le \rho \le 1 - 2A_\rho + \frac{2\zeta + z_{\alpha/2}}{\sqrt{n}}A_\rho^{1/2}\right\} \\ =\ &\mathbb{P}_\zeta\left\{\frac{2\zeta - z_{\alpha/2}}{\sqrt{n}}A_\rho^{1/2} \le \rho + 2A_\rho - 1 \le \frac{2\zeta + z_{\alpha/2}}{\sqrt{n}}A_\rho^{1/2}\right\}\end{aligned}$$

Now

$$\rho + 2A_\rho - 1 = \frac{\rho^2 - 2\rho + 1}{1+\rho} = (1-\rho)\,A_\rho.$$

So we find that

$$\begin{aligned} & \mathbb{P}_\zeta\left\{\frac{2\zeta - z_{\alpha/2}}{\sqrt{n}}A_\rho^{1/2} \le (1-\rho)\,A_\rho \le \frac{2\zeta + z_{\alpha/2}}{\sqrt{n}}A_\rho^{1/2}\right\} \\ = \;& \mathbb{P}_\zeta\left\{2\zeta - z_{\alpha/2} \le \sqrt{n}\,(1-\rho)\,A_\rho^{1/2} \le 2\zeta + z_{\alpha/2}\right\} \\ = \;& \mathbb{P}_\zeta\left\{\frac{\sqrt{n}}{2}\,(1-\rho)\,A_\rho^{1/2} - \frac{z_{\alpha/2}}{2} \le \zeta \le \frac{\sqrt{n}}{2}\,(1-\rho)\,A_\rho^{1/2} + \frac{z_{\alpha/2}}{2}\right\} \\ \to \;& 0, \text{ if } \sqrt{n}\,(1-\rho) \to \infty \text{ as } n \to \infty. \end{aligned} \tag{29}$$

It follows that Stock's (1991) confidence interval based on inverting a unit root test using local to unity limit theory has zero coverage probability asymptotically as $c \to -\infty$ whenever $\sqrt{n}\,(1-\rho) \to \infty$. In particular, the probability that the true value of ρ lies to the right of the interval $[\rho_L, \rho_U]$ is $\mathbb{P}_\zeta\left\{\zeta \le \frac{\sqrt{n}}{2}\,(1-\rho)\,A_\rho^{1/2} + \frac{z_{\alpha/2}}{2}\right\}$, which tends to unity whenever $\sqrt{n}\,(1-\rho) \to \infty$. These asymptotics explain the simulation findings in Hansen (1999), where the Stock confidence intervals were shown to be "poor for $\rho = 0.6$ with an error that increases with the sample size" and with the true value lying to the right of the confidence interval in 100% of the simulations when $n = 240$ and $\rho = 0.6$, precisely as predicted in (29).

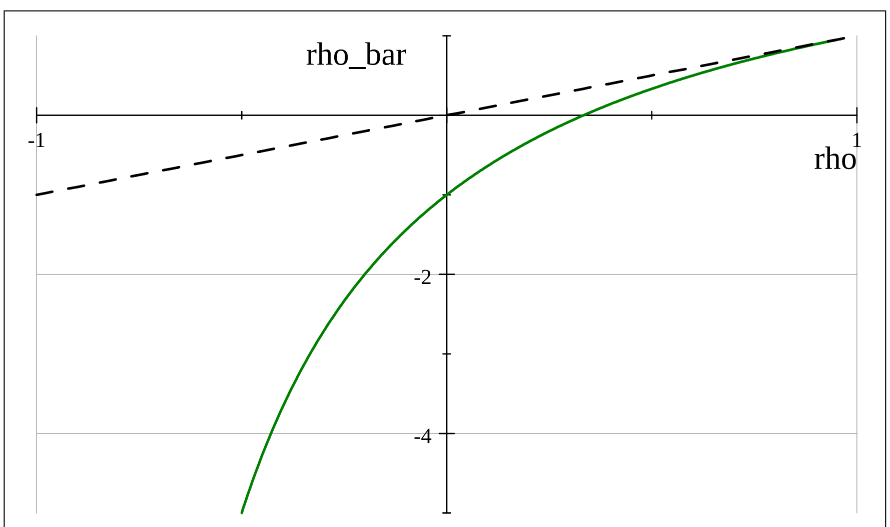

Fig. 2: Plot of $\bar{\rho} = \frac{3\rho-1}{\rho+1}$ (solid green) and 45^o line (dashed black).

The induced confidence interval (28) is centred on

$$\bar{\rho} = 1 - 2A_\rho = \frac{3\rho - 1}{\rho + 1} \tag{30}$$

and the interval shrinks to the pseudo true value $\bar{\rho}$ as $n \to \infty$. Observe that $\rho - \bar{\rho} = \frac{(\rho-1)^2}{\rho+1} > 0$, so that $\bar{\rho} < \rho$ for all $|\rho| < 1$ and $\bar{\rho}$ equals ρ if and only if $\rho = 1$ (see Fig. 2). Note that when the true $\rho = 0$, $\bar{\rho} = -1$ and when $\rho = 0.5$, $\bar{\rho} = 1/3$. So the bias in ρ is substantial for much of the stationary region.

3 Hansen and Mikusheva Constructions

Hansen (1999) and Mikusheva (2007) suggested to construct confidence intervals by performing test inversion with a properly centred t-ratio statistic (Hansen) or a general test function involving a centred numerator and separate denominator components (Mikusheva). These suggestions mirror earlier work by Andrews (1993), under Gaussianity, and bootstrap test inversion methods in the statistical literature (Carpenter, 1999). In place of (2), these approaches effectively amount to working with the statistic

$$t_{\hat{\rho},\rho} = \frac{n\left(\hat{\rho} - \rho\right)}{\left\{\hat{\sigma}^2 / \left(n^{-2}\sum_{t=1}^n x_{t-1}^2\right)\right\}^{1/2}} \Longrightarrow \frac{\int J_c dW}{\left(\int J(r)^2 dr\right)^{1/2}} := \lambda_c, \tag{31}$$

or a coefficient based version of the test instead.

Proceeding in the same way as above for large $|c|$, we find

$$\begin{aligned}
\lambda_c &= \frac{\xi\left\{1 + O_p\left(|c|^{-1/2}\right)\right\}}{\left\{1 + \left(2\frac{\xi}{(-2c)^{1/2}} - \frac{\eta^2}{-2c}\right)\left\{1 + O_p\left(|c|^{-1/2}\right)\right\}\right\}^{1/2}} \\
&= \xi\left\{1 - \frac{\xi}{(-2c)^{1/2}} + O_p\left(\frac{1}{c}\right)\right\}\left\{1 + O_p\left(|c|^{-1/2}\right)\right\} \\
&= \left\{\xi - \frac{1}{(-2c)^{1/2}} - \frac{\xi^2 - 1}{(-2c)^{1/2}} + O_p\left(\frac{1}{c}\right)\right\}\left\{1 + O_p\left(|c|^{-1/2}\right)\right\},
\end{aligned}$$

in place of (9) and

$$\lambda_c \sim N\left(-\frac{1}{2^{1/2}|c|^{1/2}}, 1\right) + O_p\left(\frac{1}{|c|^{1/2}}\right) = N(0,1) + O_p\left(\frac{1}{|c|^{1/2}}\right).$$

in place of (10). The induced confidence interval for c upon inversion of $t_{\hat{\rho},\rho}$ is, up to an error of $O_p\left(|c|^{-1/2}\right)$

$$\begin{aligned}\left\{f_{L,\alpha/2}\left(t_{\hat{\rho},\rho}\right), f_{U,\alpha/2}\left(t_{\hat{\rho},\rho}\right)\right\} &= \left\{-\frac{1}{2^{1/2}|c|^{1/2}} - z_{\alpha/2}, -\frac{1}{2^{1/2}|c|^{1/2}} + z_{\alpha/2}\right\} \\ &\sim \left\{-z_{\alpha/2}, +z_{\alpha/2}\right\}, \quad \text{as } c \to -\infty.\end{aligned}$$

In view of the centred form of the t ratio $t_{\hat{\rho},\rho}$ in (31), the corresponding induced interval for ρ when $c \to -\infty$ is simply $\left\{\hat{\rho} - z_{\alpha/2}\sigma_{\hat{\rho}}, \hat{\rho} + z_{\alpha/2}\sigma_{\hat{\rho}}\right\}$, the same as the classical stationary interval based on normal asymptotics.

The difference between using the unit root test statistic and centred statistic for inversion can be further explained in terms of the simple relationship between the two statistics

$$t_{\hat{\rho},\rho_n} = \frac{\hat{\rho}-1}{\sigma_{\hat{\rho}}} + \frac{1-\rho_n}{\sigma_{\hat{\rho}}} = t_{\hat{\rho}} + \frac{1-\rho_n}{\sigma_{\hat{\rho}}},$$

and noting the difference in the treatment of the second component in the two approaches when $\rho_n = 1 + \frac{c}{n}$. When inverting the local to unity statistic $t_{\hat{\rho}}$, the limit behavior of $\sigma_{\hat{\rho}}$ under local to unity asymptotics is imposed, effectively replacing $\sigma_{\hat{\rho}} = \left\{\hat{\sigma}^2 / \sum_{t=1}^{n} x_{t-1}^2\right\}^{1/2}$ by its limiting version under these asymptotics, viz., $n\left(\int J_c(r)^2 dr\right)^{1/2}$, so that

$$\frac{1-\rho_n}{\sigma_{\hat{\rho}}} \sim n\left(1-\rho_n\right)\left(\int J_c(r)^2 dr\right)^{1/2} = -c\left(\int J_c(r)^2 dr\right)^{1/2},$$

as in (2). As the analysis above shows, it is this term that produces the biased confidence intervals in the limit as $c \to -\infty$. On the other hand, use of the correctly centred statistic automatically bias-adjusts for this quantity, as is apparent in (31).

4 Predictive Regression Tests

In predictive regressions it is now common empirical practice to allow for unknown persistence in the regressor-predictors. Such predictors complicate testing procedures by introducing nonstandard limit theory and dependence on nuisance parameters like the localizing coefficient in a near integrated regressor. Various approaches are now available to deal with these complications. A recent overview is given in Phillips and Lee (2012a).

A popular procedure was implemented in Campbell and Yogo (2006, hereafter CY), following an earlier suggestion by Cavanagh, Stock and Elliott (1995, hereafter CSE). These procedures both use a Bonferroni method in conjunction with Stock's (1991) confidence interval construction for the autoregressive coefficient of the regressor-predictor to produce tests of predictability that are intended to be robust to persistence. Our interest here is in the asymptotic properties of the implied predictability tests as the stationarity region in the regressor is approached to assess whether they are uniform over stationary and local to unity ρ.

We start with a brief outline of the two procedures. Both the Q test of CY and the t ratio test in CSE involve t ratios computed by simple regression in combination with Stock's (1991) confidence interval construction. To fix ideas we consider the standard predictive regression model (without an intercept to simplify matters as there are no differences of any import for our present purposes when an intercept is included)

$$y_t = \beta x_{t-1} + u_{0t} \tag{32}$$

$$x_t = \rho_n x_{t-1} + u_{xt} \tag{33}$$

with $\rho_n = 1 + \frac{c}{n}$ for $c \le 0$ and mds innovations $u_t = (u_{0t}, u_{xt})$ for which

$$E_{\mathcal{F}_{t-1}} u_t = 0,\ E_{\mathcal{F}_{t-1}} \left[u_t u_t' \right] = \begin{bmatrix} \sigma_{00} & \sigma_{0x} \\ \sigma_{x0} & \sigma_{xx} \end{bmatrix} =: \Sigma,$$

and

$$n^{-1/2} \sum_{t=1}^{\lfloor n \cdot \rfloor} u_t \Rightarrow B(\cdot) = \begin{bmatrix} B_0(\cdot) \\ B_x(\cdot) \end{bmatrix},$$

where B is Brownian motion with variance matrix Σ.

Upon fitting (32) by least squares we have the centered decomposition

$$\begin{aligned} \hat{\beta} - \beta &= \frac{\sum_{t=1}^n x_{t-1} u_{0t}}{\sum_{t=1}^n x_{t-1}^2} = \frac{\sum_{t=1}^n x_{t-1} u_{0.xt}}{\sum_{t=1}^n x_{t-1}^2} + \left(\frac{\sigma_{0x}}{\sigma_{xx}} \right) \frac{\sum_{t=1}^n x_{t-1} u_{xt}}{\sum_{t=1}^n x_{t-1}^2} \\ &= \frac{\sum_{t=1}^n x_{t-1} u_{0.xt}}{\sum_{t=1}^n x_{t-1}^2} + \left(\frac{\sigma_{0x}}{\sigma_{xx}} \right) (\hat{\rho} - \rho_n), \end{aligned}$$

where $u_{0.xt} = u_{0t} - \frac{\sigma_{0x}}{\sigma_{xx}} u_{xt}$. Set $\hat{u}_{0t} = y_t - \hat{\beta} x_t$ and $\hat{\sigma}^2 = n^{-1} \sum_{t=1}^n \hat{u}_{0t}^2$. The centred limit theory for $\hat{\rho}$ and $\hat{\beta}$ is

$$n(\hat{\rho} - \rho_n) \Longrightarrow \frac{\int K_c(r) dB_x(r)}{\int K_c(r)^2 dr}, \tag{34}$$

and

$$n\left(\hat{\beta}-\beta\right)=\frac{\frac{1}{n}\sum_{t=1}^{n}x_{t-1}u_{0t}}{\frac{1}{n^2}\sum_{t=1}^{n}x_{t-1}^2}\Longrightarrow\frac{\int K_{xc}(r)dB_{0.x}(r)}{\int K_{xc}(r)^2dr}+\frac{\sigma_{0x}}{\sigma_{xx}}\frac{\int K_c(r)dB_x(r)}{\int K_c(r)^2dr},\tag{35}$$

where $K_c(r)=\int_0^r e^{c(r-s)}dB_x\left(s\right)=\sigma_{xx}^{1/2}\int_0^r e^{c(r-s)}dW_x\left(s\right)=:\sigma_{xx}^{1/2}J_c\left(r\right)$ and W_x is standard Brownian motion. As discussed above, the corresponding unit root t- ratio statistic is

$$t_{\hat{\rho}}=\frac{\hat{\rho}-1}{\sigma_{\hat{\rho}}}\Rightarrow\frac{\int J_c(r)dW_x(r)}{\left(\int J_c(r)^2dr\right)^{1/2}}+c\left(\int J_c(r)^2dr\right)^{1/2}.\tag{36}$$

Defining $\hat{\sigma}_{00}^2=n^{-1}\sum_{t=1}^{n}\left(y_t-\hat{\beta}x_t\right)^2$ and $\sigma_{\hat{\beta}}^2=\hat{\sigma}_{00}^2/\sum_{t=1}^{n}x_{t-1}^2$, the t-ratio test on the regression coefficient β is

$$t_{\hat{\beta}}=\frac{\hat{\beta}-\beta}{\sigma_{\hat{\beta}}}\Longrightarrow\frac{\int J_c(r)dB_{0.x}(r)}{\left(\sigma_{00}\int J_c(r)^2dr\right)^{1/2}}+\frac{\sigma_{x0}}{\sigma_{xx}^{1/2}}\frac{\int J_c(r)dW_x(r)}{\left(\sigma_{00}\int J_c(r)^2dr\right)^{1/2}}\tag{37}$$

$$=\left(\frac{\sigma_{00.x}}{\sigma_{00}}\right)^{1/2}Z+\frac{\sigma_{x0}}{\left(\sigma_{xx}\sigma_{00}\right)^{1/2}}\left(\frac{\int J_c(r)dW_x(r)}{\left(\int J_c(r)^2dr\right)^{1/2}}\right)\tag{38}$$

$$=:\left(1-\delta^2\right)^{1/2}Z+\delta\eta_{LUR}\left(c\right),\tag{39}$$

where $\sigma_{00.x}=\sigma_{00}-\sigma_{x0}^2/\sigma_{xx}$, $\delta=\frac{\sigma_{x0}}{(\sigma_{xx}\sigma_{00})^{1/2}}$, $1-\delta^2=\frac{\sigma_{00.x}}{\sigma_{00}}$, $B_{0.x}(r)=B_0(r)-\frac{\sigma_{x0}}{\sigma_{xx}}B_x\left(r\right)$, $Z\equiv N\left(0,1\right)$, and

$$\eta_{LUR}\left(c\right)=\frac{\int J_c(r)dW_x(r)}{\left(\int J_c(r)^2dr\right)^{1/2}},\tag{40}$$

giving a mixture limit theory in (39) that depends on c and the correlation parameter δ. The limit variates Z and $\eta_{LUR}\left(c\right)$ in (39) are independent in view of the independence of $B_{0.x}$ and B_x. When $\delta=0$, we have standard asymptotic normal inference. When $\delta=1$ we have strong endogeneity and local unit root limit theory (LUR).

CSE and CY address inference when $\delta\neq 0$ by using a Bonferroni method: in effect, finding possible values for c (or ρ) and using the most conservative ones to produce a robust test. The approach can be used with various test statistics. The two methods considered below are those used in CSE and CY.

4.1 The t ratio test $t_{\hat{\beta}}$.

The t ratio statistic $t_{\hat{\beta}}$ is considered in both CSE and CY, although CY recommend for implementation a different t-ratio test called the Q test, which will be discussed next. The mixture limit theory of $t_{\hat{\beta}}$ given in (39) means that tests and confidence intervals need to allow for the unknown value of c. A $100\,(1-\alpha_1)\,\%$ confidence interval (CI) is constructed for c using unit root test inversion as in Stock (1991). For each c in this CI, a $100\,(1-\alpha_2)\,\%$ CI is constructed for β, denoted as $CI_{\beta|c}\,(\alpha_2)$. A CI for β that is free of c is obtained as the union

$$CI_{\beta}\,(\alpha) = \bigcup_{c\in CI_c(\alpha_1)} CI_{\beta|c}\,(\alpha_2)\,.$$

More specifically, the estimate $\hat{\rho}$ is used to find a confidence interval $CI_c(\alpha_1) = [c_L\,(\alpha_1)\,, c_U\,(\alpha_1)]$ for c by inverting a t ratio ADF unit root test statistic for ρ as in Stock (1991) or, in the case of CY, a version of this statistic that improves efficiency in trend removal by quasi-differencing (so-called GLS detrending), which is unnecessary in the present case. Then (given δ or a consistent estimate of δ) the authors use the critical value $d_{t_{\hat{\beta}},c,\frac{1}{2}\alpha_2}$ of $t_{\hat{\beta}} \sim \delta\eta_{LUR}\,(c) + \left(1-\delta^2\right)^{1/2} Z$ at the percentile $\frac{1}{2}\alpha_2$ to find the following CI for β :

$$\begin{aligned} CI_{\beta}(\alpha_1,\alpha_2) &= \left[d_L^{\beta}(\alpha_1,\alpha_2), d_U^{\beta}(\alpha_1,\alpha_2)\right] \\ &:= \left[\min_{c_L(\alpha_1)\le c\le c_U(\alpha_1)} d_{t_{\hat{\beta}},c,\frac{1}{2}\alpha_2}, \max_{c_L(\alpha_1)\le c\le c_U(\alpha_1)} d_{t_{\hat{\beta}},c,1-\frac{1}{2}\alpha_2}\right]. \end{aligned} \tag{41}$$

It follows that as $n \to \infty$

$$\begin{aligned} & \Pr\left(t_{\hat{\beta}} \notin \left[d_L^{\beta}(\alpha_1,\alpha_2), d_U^{\beta}(\alpha_1,\alpha_2)\right]\right) \\ \to\; & \Pr\left(\delta\eta_{LUR}\,(c) + \left(1-\delta^2\right)^{1/2} Z \notin \left[d_L^{\beta}(\alpha_1,\alpha_2), d_U^{\beta}(\alpha_1,\alpha_2)\right]\right) \\ \le\; & \alpha_1+\alpha_2, \text{ (by Bonferroni)} \end{aligned} \tag{42}$$

thereby achieving a test and confidence interval for β that is robust to persistence, as measured by the localizing coefficient c or ρ_n.

We now consider the implications for this predictability test of the results obtained above for test inversion confidence interval construction in the stationary case $|\rho| < 1$. As noted, the interval $CI_c(\alpha_1) = [c_L\,(\alpha_1)\,, c_U\,(\alpha_1)]$ advances progressively towards $-\infty$ as x_t approaches stationarity. Although the implied confidence interval for ρ has zero coverage probability asymptotically, it nonetheless still holds that $c \to -\infty$ for all $c \in [c_L\,(\alpha_1)\,, c_U\,(\alpha_1)]$

when ρ is stationary, as is immediately evident from setting (21) in (26). Hence, the actual coverage probability associated with the interval $CI_c(\alpha_1)$ tends to unity rather than $1-\alpha_1$ as n tends to infinity in the stationary case. In effect, the restriction $c \in [c_L(\alpha_1), c_U(\alpha_1)]$ is vacuous in the stationary case because both limits $c_L(\alpha_1)$ and $c_U(\alpha_1) \to -\infty$ and these limits are dominated by $\hat{\tau}^2$ not the size-determining quantities $\mathcal{Z}_L\hat{\tau}$ or $\mathcal{Z}_U\hat{\tau}$ in (26). The probability mass $\alpha_1 = \mathbb{P}\{c \notin [c_L(\alpha_1), c_U(\alpha_1)]\}$ therefore escapes to zero due to the failure of tightness in the sequence of unit root test statistics from which this interval is constructed.

Moreover, when $c \to -\infty$, we have $\eta_{LUR}(c) \to_d \xi \equiv N(0,1)$ and the limit variate ξ is independent of Z in (39), as remarked above. Hence, for all $c \in [c_L(\alpha_1), c_U(\alpha_1)]$ we actually have

$$\left(1-\delta^2\right)^{1/2} Z + \delta\eta_{LUR}(c) \to_d \left(1-\delta^2\right)^{1/2} Z + \delta\xi \equiv N(0,1) \qquad (43)$$

when ρ is stationary and $n \to \infty$. The confidence interval in the limit for the stationary case then has coverage probability determined only by the controlled level α_2 of the test (39). The CSE test is therefore undersized in the limit (here by the full probability α_1 which is lost in the limit by test inversion and failure of tightness), just as it is also (partially) undersized for finite values of c (because of the Bonferroni bounds). Hence, the CSE confidence interval has excess coverage probability for stationary ρ and has longer length than the usual stationary regression interval. So the CSE confidence intervals are not uniform with the stationary case in the limit and they remain wider than the usual stationary intervals with their nominal coverage level understating the actual coverage probability.

The limit statistic $\eta_{LUR}(c)$ in (40) that is used in the CSE confidence interval is properly centred; and the improperly centred unit root test statistic that is used in the test inversion to create the induced confidence interval for c is effective in revealing that $c \to -\infty$ when ρ is stationary. But the critical values of the CSE test are not based on the correct limit theory (43) in this case, so the test is conservative because of the use of the Bonferroni bounds and the induced confidence interval correspondingly has incorrect coverage probability in the stationary limit.

Fig. 3 shows actual coverage probabilities at a nominal asymptotic level of 90% of the CSE confidence intervals for the predictive regression coefficient β for regressors x_t with AR coefficient $\rho \in \{0.01, 0.02, ..., 0.99\}$, $n = 200$, and endogeneity coefficient $r_{0x} \in \{-0.99, -0.9, -0.6, -.04\}$ where $r_{0x} = \delta = \sigma_{x0}/\left(\sigma_{xx}\sigma_{xx}\right)^{1/2}$. The results are based on 400,000 replications and use model (32) and the confidence belts shown in Fig. 1 for the inversion of the unit root t statistic $t_{\hat{\rho}}$. Additional background computations were needed to tabulate the distribution (43) on a detailed grid of potential values of the localizing coefficient c.

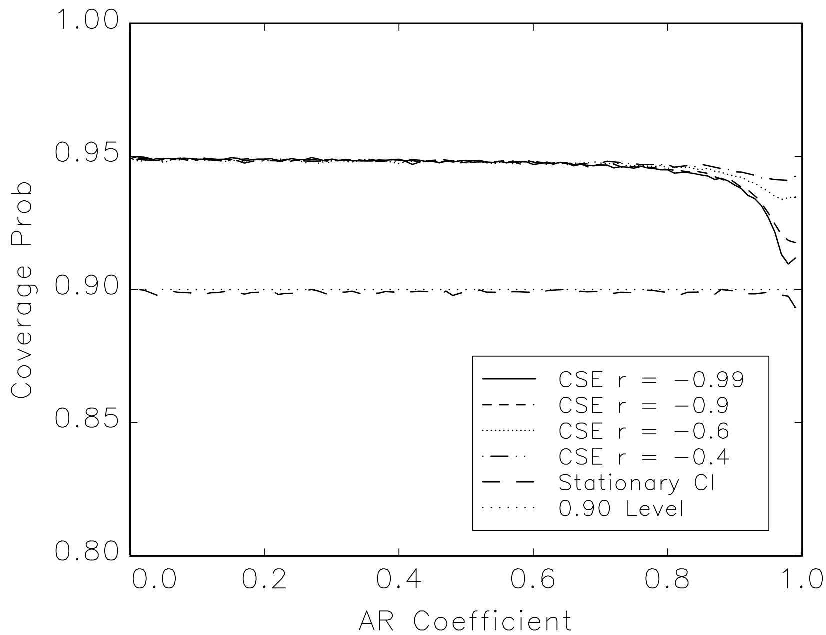

Fig. 3: Coverage probabilities of CSE and stationary confidence intervals for the predictive regression coefficient β plotted against the autoregressive coefficient ρ of x_t, shown for various values of the endogeneity coefficient r_{0x}. The nominal asymptotic level is 90%, sample size is $n = 200$, and the number of replications is $400,000$.

Evidently the coverage probability for the CSE intervals is close to 95% for stationary ρ, as predicted by the asymptotic theory. For ρ close to unity, the coverage probability decreases towards 90% but still reflects undersizing from the use of Bonferroni bounds. The CSE intervals have coverage closest to nominal coverage when ρ is close to unity and there is strong endogeneity ($\delta = -0.99, -09$) in the regression, whereas coverage is close to 95% when $\delta = -0.4$ in that case, again corroborating the asymptotics. Fig. 3 also shows the coverage probability of the usual stationary interval, which is close to the nominal 90% for all values of $\rho < 0.99$ but shows some undercoverage for $\rho = 0.99$.

4.2 The Q test

CY(2006) recommend a different t ratio test, called the Q test, that is based on the augmented regression equation (c.f. Phillips and Hansen, 1990)

$$y_t = \beta x_{t-1} + \frac{\sigma_{0x}}{\sigma_{xx}}(x_t - \rho x_{t-1}) + u_{0.xt}.$$

Specifically, the Q test employs the following coefficient estimator (conditional on ρ and with no need to fit an intercept here)

$$\hat{\beta}(\rho) = \frac{\sum_{t=1}^{n} x_{t-1}\left[y_t - \frac{\hat{\sigma}_{x0}}{\hat{\sigma}_{xx}}(x_t - \rho x_{t-1})\right]}{\sum_{t=1}^{n} x_{t-1}^2},$$

where $\hat{\sigma}_{x0}$ and $\hat{\sigma}_{xx}$ are obtained in the usual way from the least squares residuals in regressions of (32) and (33). The induced confidence interval for β is based on the t statistic $t_{\hat{\beta}(\rho)}$ for $\hat{\beta}(\rho)$ and an asymptotic normal distribution for $t_{\hat{\beta}(\rho)}$ (which is effectively the first term of (38)), together with the confidence interval $[\rho_L, \rho_U]$ for ρ that is calculated using the unit root test inversion process (based on the statistic $t_{\hat{\rho}}$ calculated from the autoregression (33)) which produces an induced confidence interval $[c_L, c_U]$ for c. In the numerical implementation of their test, CY bound the interval $[c_L, c_U]$ to lie within $[-50, 5]$, which arbitrarily restricts the allowable range of c (and hence ρ), inducing bias if the true value lies outside these bounds. This restriction is relaxed for the purpose of the following discussion, which explores the properties of the CY procedure for large $|c|$ and stationary ρ.

The asymptotic form of the induced interval $[\rho_L, \rho_U]$ is given by (28) in the stationary ρ case. From (30), the interval $[\rho_L, \rho_U]$ is asymptotically centred on $\bar{\rho}$ and shrinks to this pseudo value as $n \to \infty$ when $|\rho| < 1$. It follows that the induced Bonferroni confidence interval for β is from CY(equations (15)-(17), pp. 38-39) and conditional[3] on $\hat{\sigma}_{x0} < 0$ and $\rho_U, \rho_L > 0$

$$[\beta_L(\rho_U), \beta_U(\rho_L)] = \left[\hat{\beta}(\rho_U) - z_{\alpha_2/2}\sigma_{\hat{\beta}(\rho)}, \hat{\beta}(\rho_L) + z_{\alpha_2/2}\sigma_{\hat{\beta}(\rho)}\right], \tag{44}$$

[3] If $\hat{\sigma}_{x0} > 0$ and $\rho_U, \rho_L > 0$ the corresponding confidence interval for β should be $[\beta_L(\rho_L), \beta_U(\rho_U)]$. This dependence of the interval on the signs of $\hat{\sigma}_{x0}, \rho_L$, and ρ_U does not appear to be mentioned in CY (2006), so their stated interval only applies when $\hat{\sigma}_{x0} < 0$ and $\rho_U, \rho_L > 0$. CY do assume that the true covariance $\sigma_{x0} < 0$ and for roots ρ local to unity seem to presume that $\rho_U, \rho_L > 0$. Of course, $\hat{\sigma}_{x0} > 0$ with probability greater than zero even when the true covariance $\sigma_{x0} < 0$.

where $\sigma^2_{\hat{\beta}(\rho)} = \hat{\sigma}^2_{00.x} / \sum_{t=1}^n x^2_{t-1} = \hat{\sigma}^2_{00} \left(1 - \hat{\delta}^2\right) / \sum_{t=1}^n x^2_{t-1} \to_p 0$. The interval (44) correspondingly shrinks as $n \to \infty$ to

$$\begin{aligned} [\beta_L(\rho_U), \beta_U(\rho_L)] \quad &\to \quad \bar{\beta} := \beta + \frac{\sigma_{x0}}{\sigma_{xx}} (\rho - \bar{\rho}) \\ &= \quad \beta + \frac{\sigma_{x0}}{\sigma_{xx}} \frac{(\rho - 1)^2}{\rho + 1} \neq \beta \end{aligned}$$

for all $|\rho| < 1$ whenever $\sigma_{x0} \neq 0$. It follows that the CY confidence interval based on the Q test has zero coverage probability in the limit for all stationary $|\rho| < 1$ whenever there is regressor endogeneity ($\sigma_{x0} \neq 0$). Observe that the pseudo true value $\bar{\beta} \lesseqgtr \beta$ according as $\sigma_{x0} \lesseqgtr 0$ and the bias is greater the greater is $\left|\frac{\sigma_{x0}}{\sigma_{xx}}\right|$ and the smaller is ρ. Moreover, the Q test is biased and, when the true $\beta = 0$, the (two sided) test will erroneously indicate predictability with probability approaching unity as $n \to \infty$.[4]

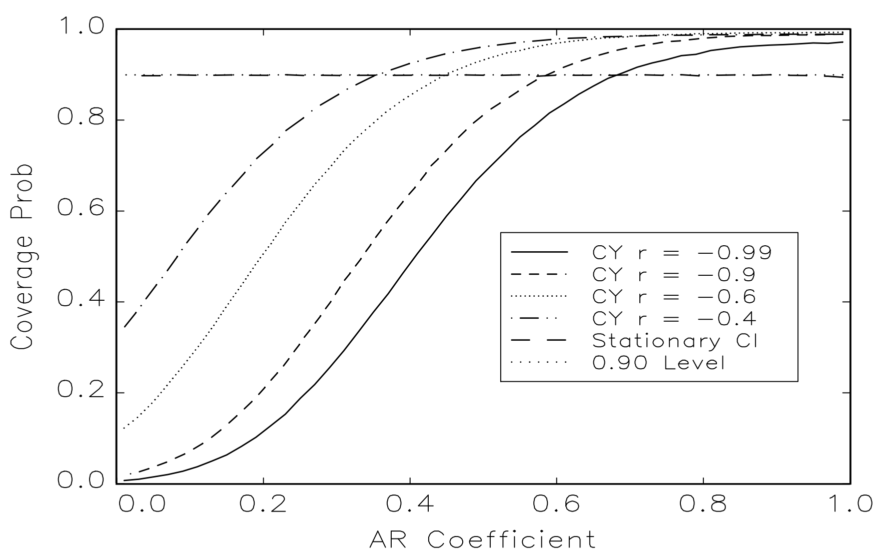

Fig. 4: Coverage probabilities of Campbell-Yogo and stationary confidence intervals for the predictive regression coefficient β plotted against the autoregressive coefficient ρ of x_t, shown for various values of the endogeneity coefficient r_{0x}. The nominal asymptotic level is 90%, sample size is $n = 200$, and the number of replications is 50,000.

[4] One sided tests correspondingly have size unity or zero depending on the direction of the test. For instance, if $\sigma_{x0} < 0$ so that the probability limit $\bar{\beta} < \beta$, we would reject the null $\beta = 0$ in a left sided test against $\beta < 0$ with probability unity in the limit or in a right sided test against $\beta > 0$ with probability zero.

Using the same design as in the simulations for CSE, Fig. 4 shows actual coverage probabilities at a nominal asymptotic level of 90% of the CY confidence intervals for the predictive regression coefficient β for regressors x_t with AR coefficient $\rho \in \{0.01, 0.03, ..., 0.99\}$, $n = 200$, and endogeneity coefficient $r_{0x} \in \{-0.99, -0.9, -0.6, -.04\}$ where $r_{0x} = \delta = \sigma_{x0}/(\sigma_{xx}\sigma_{xx})^{1/2}$. The results are based on 50,000 replications and use model (32) and the confidence belts shown in Fig. 1 for the inversion of the unit root t statistic $t_{\hat{\rho}}$. Evidently the coverage probability monotonically declines with ρ, with sharper declines that approach zero when there is stronger endogeneity in the predictive regression (higher $|\delta_{0x}|$), thereby corroborating the limit theory. The graphs reveal that the CY Q test is typically undersized for ρ close to unity and seriously oversized when ρ is distant from unity. Also shown in Fig. 3 is the coverage probability of the standard regression confidence interval based on stationary x_t with strong endogeneity $r_{0x} = -0.99$. The stationary interval has close to nominal 90% coverage for all values of $\rho \leq 0.99$. Note that with $n = 200$, $\rho = 0.99$ corresponds to $c = -2.0$, so that this value of ρ may be regarded as being in the local to unity range.[5]

5 Simple Extensions

The same results on zero coverage probability and distended length of the induced confidence intervals apply when demeaned and detrended data are used in the construction of unit root t- statistics. For example, in the demeaned case we need only note that

$$(-2c)^{1/2}\tilde{J}_c(r) = (-2c)^{1/2}J_c(r) - (-2c)^{1/2}\int J_c(s)\,ds = (-2c)^{1/2}J_c(r) + O_p\left(\frac{1}{|c|^{1/2}}\right),$$

because $(-2c)^{1/2}\int J_c(s)\,ds$ has zero mean and variance $\frac{1}{-c} + O\left(\frac{1}{|c|^2}\right) \to 0$ as $c \to -\infty$. It follows that

$$\begin{aligned}(-2c)\int \tilde{J}(r)^2 dr &= (-2c)\int J(r)^2 dr - \left((-2c)^{1/2}\int J_c(s)\,ds\right)^2 \\ &= (-2c)\int J(r)^2 dr + O_p\left(\frac{1}{|c|}\right),\end{aligned}$$

[5] In regressions with a fitted mean the autoregressive bias is well known to be greater than in the zero intercept case that is considered here. Correspondingly, the stationary test and confidence interval have more distortion, particularly for the strong endogenous case in the immediate vicinity of unity (see Fig. 2 of Campbell and Yogo, 2006).

and

$$(-2c)^{1/2}\int \tilde{J}_c(r)dW(r) = (-2c)^{1/2}\int J_c(r)dW(r) - (-2c)^{1/2}\int J_c(s)\,dsW(1)$$
$$= \xi + O_p\left(|c|^{-1/2}\right),$$

so the earlier arguments continue to apply with the same error order as $c \to -\infty$.

The results here also hold when we use test statistics based on a local alternative $\bar{\rho} = 1 + \frac{\bar{c}}{n}$ for some fixed $\bar{c} < 0$ rather than the usual unit root statistic. The findings therefore apply to the procedure suggested in Elliott and Stock (2001) involving inversion of a sequence of point optimal tests based on some fixed local alternative.

6 Conclusion

Given the uncertainty over the persistence characteristics of economic and financial data, the results presented here are relevant to much practical empirical work where there is a need for robust inference. Applications of unit root test inversion methods with local to unity asymptotics have been especially recommended for this purpose in the context of predictive regression. But as shown here, these methods are not uniformly robust and can seriously bias inference when the regressors are stationary. Inversion methods that are robust can be constructed, as demonstrated in Hansen (1999) and Mikusheva (2007), and these methods are mainly useful in a context where there is dependence on a single localizing coefficient. Implementation of grid procedures of this type can involve extensive tabulations that may require billions of regressions. To illustrate the scale of the numerical work that can be involved, the background grid computations for the tabulations of the distribution (43) that were required for Fig. 3 involved 4 billion regressions.[6] An additional 160 million regressions were needed to compute the curves shown in Fig. 3.[7]

Other methods, like the IVX instrumental variable method of Magdalinos and Phillips (2009), are also appropriate for inference in predictive

[6]Specifically, the computations involved 20,000 grid points for the localizing coefficient c, an additional 4 grid points for δ (r_{0x}), and 50,000 replications of regressions with a sample size $n = 1000$. These simulations were performed in Gauss and took a week on a machine with 8g RAM and two 2.3ghz processors.

[7]These computations involved 100 grid points for ρ, 4 grid points for δ (r_{0x}), and 400,000 replications of regressions with a sample size $n = 200$.

regressions. The IVX method is particularly useful because it has wide generality and applies to stationary, mildly integrated and local to unity regressors (Kostakis et al, 2012). The method has the advantage of accommodating multiple regressors with varying degrees of persistence, as often occurs in empirical work, as well as mildly explosive roots (Phillips and Lee, 2012b). Implementation is by straightforward linear regression (either a single regression or a few regressions to allow for different weights in the IVX instruments) and the use of standard statistical tables. These features make the method convenient and robust for empirical work.

7 Appendix: derivation of equation (9)

From (3), (7) and (8) we have

$$\begin{aligned}
\tau_c &= \frac{(-2c)^{1/2}\int J_c(r)dW(r)}{\left((-2c)\int J_c(r)^2dr\right)^{1/2}} - \frac{|c|^{1/2}}{2^{1/2}}\left((-2c)\int K_c(r)^2dr\right)^{1/2} \\
&= \frac{\xi\left\{1+O_p\left(|c|^{-1/2}\right)\right\}}{\left\{1+\left(2\frac{\xi}{(-2c)^{1/2}}-\frac{\eta^2}{-2c}\right)\left\{1+O_p\left(|c|^{-1/2}\right)\right\}\right\}^{1/2}} \\
&\quad -\frac{|c|^{1/2}}{2^{1/2}}\left\{1+\left(2\frac{\xi}{(-2c)^{1/2}}-\frac{\eta^2}{-2c}\right)\left\{1+O_p\left(|c|^{-1/2}\right)\right\}\right\}^{1/2} \\
&= \xi\left\{1-\frac{\xi}{(-2c)^{1/2}}+\frac{\eta^2}{-4c}+\frac{3}{8}\left(2\frac{\xi}{(-2c)^{1/2}}\right)^2+o_p\left(\frac{1}{c}\right)\right\}\left\{1+O_p\left(|c|^{-1/2}\right)\right\} \\
&\quad -\frac{|c|^{1/2}}{2^{1/2}}\left\{1+\frac{\xi}{(-2c)^{1/2}}-\frac{\eta^2}{-4c}-\frac{1}{8}\left(2\frac{\xi}{(-2c)^{1/2}}\right)^2+o_p\left(\frac{1}{c}\right)\right\} \\
&= -\frac{|c|^{1/2}}{2^{1/2}}+\left\{\xi\left\{1-\frac{1}{2}\right\}-\frac{\xi^2}{(-2c)^{1/2}}+\frac{|c|^{1/2}}{2^{1/2}}\left(\frac{\eta^2}{-4c}\right)+\frac{|c|^{1/2}}{2^{1/2}}\frac{1}{8}\left(2\frac{\xi}{(-2c)^{1/2}}\right)^2\right\} \\
&\qquad \times\left\{1+O_p\left(|c|^{-1/2}\right)\right\} \\
&= -\frac{|c|^{1/2}}{2^{1/2}}+\left\{\frac{1}{2}\xi-\frac{\xi^2}{2^{1/2}|c|^{1/2}}+\frac{\eta^2}{2^{5/2}|c|^{1/2}}+\frac{1}{2^{5/2}}\frac{\xi^2}{|c|^{1/2}}\right\}\left\{1+O_p\left(|c|^{-1/2}\right)\right\} \\
&= -\frac{|c|^{1/2}}{2^{1/2}}+\left\{\frac{1}{2}\xi-\frac{\xi^2}{2^{1/2}|c|^{1/2}}\left\{1-\frac{1}{4}\right\}+\frac{\eta^2}{2^{5/2}|c|^{1/2}}\right\}\left\{1+O_p\left(|c|^{-1/2}\right)\right\}
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{|c|^{1/2}}{2^{1/2}} + \left\{ \frac{1}{2}\xi - \frac{3\xi^2}{2^{5/2}\,|c|^{1/2}} + \frac{\eta^2}{2^{5/2}\,|c|^{1/2}} \right\} \left\{ 1 + O_p\left(|c|^{-1/2} \right) \right\} \\
&= -\frac{|c|^{1/2}}{2^{1/2}} + \left\{ \frac{1}{2}\xi - \frac{3\xi^2 - \eta^2}{2^{5/2}\,|c|^{1/2}} \right\} \left\{ 1 + O_p\left(|c|^{-1/2} \right) \right\}.
\end{aligned}
$$

We therefore have the following asymptotic representation of the t ratio τ_c for large c

$$
\begin{aligned}
\tau_c &= -\frac{|c|^{1/2}}{2^{1/2}} + \left\{ \frac{1}{2}\xi - \frac{3\xi^2 - \eta^2}{2^{5/2}\,|c|^{1/2}} \right\} \left\{ 1 + O_p\left(|c|^{-1/2} \right) \right\} \\
&= -\frac{|c|^{1/2}}{2^{1/2}} + \left\{ \frac{1}{2}\xi - \frac{2}{2^{5/2}\,|c|^{1/2}} - \frac{3\left(\xi^2 - 1\right) - \left(\eta^2 - 1\right)}{2^{5/2}\,|c|^{1/2}} \right\} \left\{ 1 + O_p\left(|c|^{-1/2} \right) \right\} \\
&= -\frac{|c|^{1/2}}{2^{1/2}} + \left\{ \frac{1}{2}\xi - \frac{1}{2^{3/2}\,|c|^{1/2}} - \frac{\varphi}{2^{5/2}\,|c|^{1/2}} \right\} \left\{ 1 + O_p\left(|c|^{-1/2} \right) \right\},
\end{aligned}
$$

with $\varphi = 3\left(\xi^2 - 1\right) - \left(\eta^2 - 1\right)$, giving the stated result.

8 References

Andrews, D. W. K. (1993). "Exactly median unbiased estimation of first autoregressive unit root models", *Econometrica*, 61, 139-165.

Campbell, J. and M. Yogo (2006). "Efficient tests of stock return predictability," *Journal of Financial Economics*, 81(1), 27-60.

Carpenter, J. (1999). "Test inversion bootstrap confidence intervals", *Journal of the Royal Statistical Society, Series B*, 61, 159-172.

Cavanagh, C., G. Elliott. and J.Stock (1995). "Inference in models with nearly integrated regressors," *Econometric Theory*, 11(05), 1131-1147.

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

Elliott, G. and J. H. Stock (2001). "Confidence intervals for autoregressive coefficients near one", Journal of Econometrics, 103, 155-181.

Giraitis, L. and P. C. B. Phillips (2006). "Uniform limit theory for stationary autoregression," *Journal of Time Series Analysis*, 27, 51-60.

Hansen, B. E. (1999). “The grid bootstrap and the autoregressive model”, *Review of Economics and Statistics*, 81, 594-607.

Kostakis, A., A. Magdalinos and M. Stamatogiannis (2012). “Robust econometric inference for stock return predictability,” *Unpublished Manuscript,* University of Nottingham.

Magdalinos, T. and P. C. B. Phillips (2009). Econometric Inference in the Vicinity of Unity. Yale University, Working Paper.

Mikusheva, A. (2007). “Uniform inference in autoregressive models”, *Econometrica*, 75, 1411-1452.

Phillips, P. C. B. (1987). “Towards a unified asymptotic theory for autoregression,” *Biometrika* 74, 535–547.

Phillips, P. C. B. and B. E. Hansen (1990). “Statistical inference in instrumental variables regression with I(1) processes,” *Review of Economic Studies* 57, 99–125.

Phillips, P. C. B. and T. Magdalinos (2007a), "Limit theory for moderate deviations from a unit root," *Journal of Econometrics* 136, 115-130.

Phillips, P. C. B. and T. Magdalinos (2007b), "Limit theory for moderate deviations from a unit root under weak dependence," in G. D. A. Phillips and E. Tzavalis (Eds.) *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis.* Cambridge: Cambridge University Press, pp.123-162.

Phillips, P. C. B. and J. H. Lee (2012a), “Predictive regression under various degrees of persistence and robust long-horizon regression”, unpublished paper, Yale University.

Phillips, P. C. B. and J. H. Lee (2012b), “Limit theory for VARs with mixed roots near unity”, unpublished paper, Yale University.

Stock, J. (1991). “Confidence intervals for the largest autoregressive root in US macroeconomic time series,” *Journal of Monetary Economics*, 28(3), 435-459.