Fordham University
Department of Economics
Discussion Paper Series

A New Solution to Time Series Inference in Spurious Regression Problems

Hrishikesh D. Vinod
Fordham University, Department of Economics

Discussion Paper No: 2010-01
March 2010

Department of Economics
Fordham University
441 E Fordham Rd, Dealy Hall
Bronx, NY 10458
(718) 817-4048

A New Solution to Time Series Inference in Spurious Regression Problems

H. D. Vinod *

Abstract

Phillips (1986) provides asymptotic theory for regressions that relate nonstationary time series including those integrated of order 1, $I(1)$. A practical implication of the literature on spurious regression is that one cannot trust the usual confidence intervals. In the absence of prior knowledge that two series are cointegrated, it is therefore recommended that after carrying out unit root tests we work with differenced or detrended series instead of original data in levels. We propose a new alternative for obtaining confidence intervals based on the Maximum Entropy bootstrap explained in Vinod and López-de-Lacalle (2009). An extensive Monte Carlo simulation shows that our proposal can provide more reliable conservative confidence intervals than traditional, differencing and block bootstrap (BB) intervals.

KEY WORDS: bootstrap; simulation; confidence intervals. JEL Codes: C12, C15, C22, C51

*Professor of Economics, Fordham University, Bronx, New York, USA 10458. url: http://www.fordham.edu/economics/vinod E-Mail: vinod@fordham.edu. I thank Bruce McCullough for comments. This paper is a part of my invited talk at an international bootstrap conference in March 2010 at Texas A&M University.

1 Introduction

We discuss a Monte Carlo simulation of a new solution to Granger and Newbold's spurious regression problem by using confidence intervals based on the Maximum Entropy bootstrap (meboot) explained with examples in Vinod (2008), Vinod and López-de-Lacalle (2009) and Vinod (2010). Vinod (2006) provides theoretical justification for meboot.

Before describing our new solution to spurious regression problem, let us review the problem in the context of a simple bivariate regression:

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t, \tag{1}$$

where the original variables in levels are (integrated) nonstationary random walk series, $(x_t, y_t) \sim I(1)$. Ordinary least squares (OLS) coefficients are: $(\hat{\beta}_0, \hat{\beta}_1)$. If errors are serially correlated, $\epsilon_t = \rho\epsilon_{t-1} + \epsilon'_t$ is an autoregressive process of order 1, AR(1). One tests the null hypothesis that $\rho = 0$ by the Durbin-Watson (DW) statistic. Granger and Newbold define spurious regression as occurring when $R^2 > DW$.

Some authors including Hamilton (1994) define spurious regression more narrowly as occuring when ϵ_t is nonstationary, such as when $\rho = 1$, making it a random walk. Then OLS point estimates are inconsistent. Using the narrow view, Hamilton explains the asymptotic theory of spurious regression from Phillips (1986), whereby the sampling distributions of OLS coefficients are non-standard. Hence the usual Student's t tests and confidence intervals (CI) are unreliable. If $\epsilon_t \sim I(0)$, the OLS point estimates are known to be 'super consistent,' converging at the rate T instead of the usual $\sqrt{T}$ as the sample size $T \to \infty$. Then, one should retain the OLS point estimates and seek improved confidence intervals.

The narrow view can be misleading if a practitioner uses unit root test rejecting the null hypothesis $\rho = 1$, to conclude that spurious regression problem is absent and relies on 95% CI based on Student's t. Our simulation reveals that coverage probabilities for OLS are much lower than 0.95.

Hamilton (1994) lists three 'cures' denoted here by (c1) to (c3) to the spurious regression problem: (c1) Include y_{t-1} and x_{t-1} as additional regressors. Along similar lines Stock and Watson (1993) suggest 'dynamic' OLS, where several additional regressors involve leads and lags of first differences: $\Delta y_t, \Delta x_t$. (c2) Difference the data before estimation. (c3) Following Blough (1992) estimate (1) by generalized least squares (GLS) after Cochrane-Orcutt correction for AR(1) errors.

We refer to (c2) as 'OLSdiff' model defined by:

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \nu_t, \tag{2}$$

If OLSdiff is derived from (1), we have the special case where ($\nu_t = \epsilon_t - \epsilon_{t-1}$) and $\beta_0 = 0$. Differencing strategy is subject to at least four drawbacks:

(d1) If ϵ_t are iid, the differencing transformation makes ν_t follow a moving average process.

(d2) Efficiency of OLS is reduced because of induced MA errors, and also because of a reduction in the overall variability in the differenced data.

(d3) Original subject matter specification is destroyed and there is a possibility of misspecification, Hamilton (1994). Although Nelson and Plosser (1982) claimed that most macro economic series are differece stationary, this issue is far from settled. If the series are 'trend stationary,' rather than 'difference stationary, differencing will lead to misspecification. Standard inference on the slope and intercept can be distinct for the two types.

(d4) If the series is subject to structural change or other shifts, differencing across these changes may not be appropriate.

Subject to the above drawbacks, our simulation confirms the result that differencing vastly improves the coverage of the true parameter β_1 when OLS is used to estimate (2).

Statistical relations are generally based on some subject matter theoretical propositions. For example, economic theory might postulate a relation

(1) among levels of inflation and unemployment variables. Now, (2) may not be equivalent to (1) in all cases. When there are other variables present, the differencing of all variables might stretch the economic interpretation more than an economist can accept. Thus the cures (c1) and (c2) changing the specification (1) might be problematic. This paper suggests a fourth cure which does not change the specification, but uses meboot in conjunction with a unit root test on residuals $\hat{\epsilon}_t$. Sections 2 and 3 describe our extensive Monte Carlo simulations.

If $\epsilon_t \sim I(0)$, we suggest using meboot on OLS. If $\epsilon_t \sim I(1)$, we suggest using the meboot on Cochrane-Orcutt type GLS as explained in subsection 2.2. The following subsection provides an introduction to the main ideas of the meboot algorithm.

1.1 Introduction to the meboot algorithm

Normal theory inference assumes that the observed statistic such as a sample mean $\bar{x}$ (*invariant* to reordering of x_t) is one realization from an infinitely large normal population. By analogy, time series inference developed in 1930's by Wiener, Kolmogorov and Khintchine (WKK), among others, imagined an infinite ensemble Ω as a population of stationary ($I(0)$) time series with the observed series x_t as its element. If the lag operator is defined as $Lx_t = x_{t-1}$, the analogy with normal theory inference is facilitated if the time series is *invariant* with respect to lag shifts of any order. WKK refined the invariance ideas in a stationary time series model which satisfies the ergodic theorem (time averages equal ensemble averages), before the era of modern computers.

Unfortunately, converting a typical short time series in social sciences, x_t, to satisfy the covariance stationary model is inconvenient and sometimes destroys the original specification. Vinod (2006) bypasses the WKK model by using a computer algorithm to create an approximation to the large Ω ensemble itself. This paper enhances that work with a detailed Monte Carlo simulation.

Note that traditional independent and identically distributed (iid) bootstrap directly shuffles individual data points *with replacement* to create a large number $J = 999$, say of resamples. Since the iid bootstrap fails for m-dependent time series data, one must use the block bootstrap (BB) explained by various authors including Davison and Hinkley (1997), Liu (1988), and Lahiri (2003). Since m-dependent means that roughly speaking blocks of size m are independent, BB shuffles blocks of data.

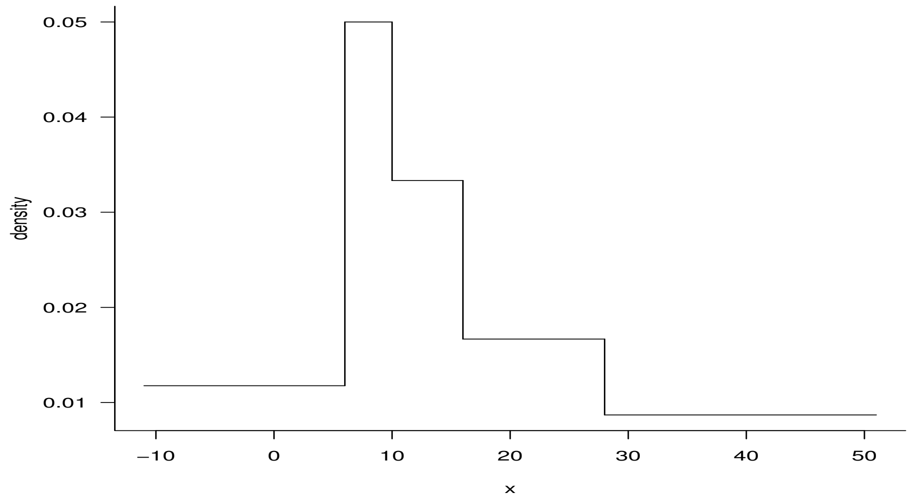
The meboot provides a new alternative to BB, applicable to possibly nonstationary time series $x_t, t == 1, \ldots, T$. The seven steps of the meboot algorithm are explained in Vinod and López-de-Lacalle (2009). Vinod (2010) views the algorithm with a slightly new angle, in terms of two-way mappings based on simultaneous sorting of two columns of a matrix from the time domain (t-dom) and numerical magnitudes or 'values' domain (v-dom) and back.

The first step of the algorithm sorts a matrix with the first column containing $t = 1, \ldots, T$ and second column containing x_t to obtain the usual order statistics $x_{(t)}$ in the second column, while remembering the first column of sorted t subscripts. This is a one-one onto bijection mapping: (t-dom)$\rightarrow$ (v-dom). Lemma 2 in Vinod (2010) shows that the map is linear, being represented by a matrix multiplication.

Unlike iid or BB, meboot admits resample values from a small neighborhood of $x_{(t)}$ in the (v-dom). Vinod and López-de-Lacalle (2009) use a simple example of five observations: $x_t = (4, 12, 36, 20, 8)$, to illustrate these neighborhoods using averages of adjacent order statistics: $x_{(t)} = (4, 8, 12, 20, 36)$. The maximum entropy (ME) principle is used to claim that the ME density is 'uniform.' For the simple example, the ME density is shown to consist of five half open uniform density intervals: $U(-11, 6] * U(6, 10] * U(10, 16] * U(16, 28] * U(28, 51]$, illustrated in Figure 1.

The meboot algorithm further requires the ME density to satisfy two constraints: 1) Mass-preserving: On average, a fraction $1/T$ of the mass of the probability distribution must lie in each small interval. 2) Mean-

Figure 1: Toy ME density for $x_t = (4, 12, 36, 20, 8)$



.

preserving: This constraint numerically ensures that the time average equals the ensemble average.

The algorithm goes on to construct a *cumulative* ME density based on a density similar to the one in Figure 1 in the (v-dom). Now, one independently selects a large number ($J = 999$, say) of resamples from the ME density, just like the iid bootstrap selecting a large number of resamples from the empirical cumulative density function (ECDF) defined from x_t in the (t-dom).

Finally, the algorithm maps all iid resamples from (v-dom) to (t-dom) by using the *sorted* t subscripts of the first step in the first column and iid resamples from the ME density in the second column. This is a simultaneous sort of the matrix with a focus on the first column, which recovers the time subscript, as it were. An important property of all resampled series: $x_{t,j}, j = 1, \ldots, J$ is that Spearman's rank correlation coefficient between the original data x_t and $x_{t,j}$ is always near unity. In future work it might be possible to relax this by considering rolling windows over short time intervals. Koutris et al. (2008) use the meboot algorithm with rolling windows to propose a new test for stationarity.

While meboot is relatively new and designed for strongly dependent, evolutionary, nonstationary times series, BB is designed for mildly dependent stationary series. Since one is curious to compare the performance of meboot with BB, we include such a comparison in some Monte Carlo experiments. Since BB is not really applicable to the nonstationary times series mainly considered in this paper, we also include some fractionally integrated (long memory) and autoregressive series for which BB is applicable in our comparisons. We still find that BB performs poorly, consistent with Spanos and Kourtellos (2002).

2 Monte Carlo Simulation of Spurious Regression

This section focuses on spurious regression (1) involving two series: $(y_t, x_t) \sim I(1)$, but $\epsilon_t \sim I(0)$, implying that OLS remains super consistent. Let us first choose $T = 100$ and create our simulated regressor data $x_t, t = 1, \ldots T$ as follows. It is well known that starting with normal random deviates x_{0t}, a cumulative sum: $x_t = \Sigma_{j=1}^{t} x_{0j}$, directly creates an $I(1)$ series.

Our $y_t \sim I(1)$ data are created by following a definition based on (1). We define:

$$y_t = 1 + 2x_t + u_t, \tag{3}$$

where the assumed true coefficients are: $\beta_0 = 1$ and $\beta_1 = 2$, and u_t are simulated AR(1) series with various choices of the autoregressive parameter $\rho \neq 0$, such that our simulations satisfy $R^2 > DW$, ensuring that we are dealing with spurious regressions.

Noriega and Ventosa-Santaularia (2006) discuss asymptotic theory confirming the presence of spurious regression despite 'breaks' in the levels and trends of the series. Since breaks in the data are shown to matter, even asymptotically, our simulation will incorporate artificial breaks via impulse or step modifications to x_{0t} before computing their cumulative sums. In a handful of experiments we choose to modify x_{0t} by impulse or step inputs, *after* computing their cumulative sums. Our conclusions (details not reported for brevity) regarding coverage performance of meboot, OLS and BB are unchanged even if the cumulative sum is computed before the impulse or step modifications.

We modify initial standard normal deviates x_{0t} as:

(a) No modifications.

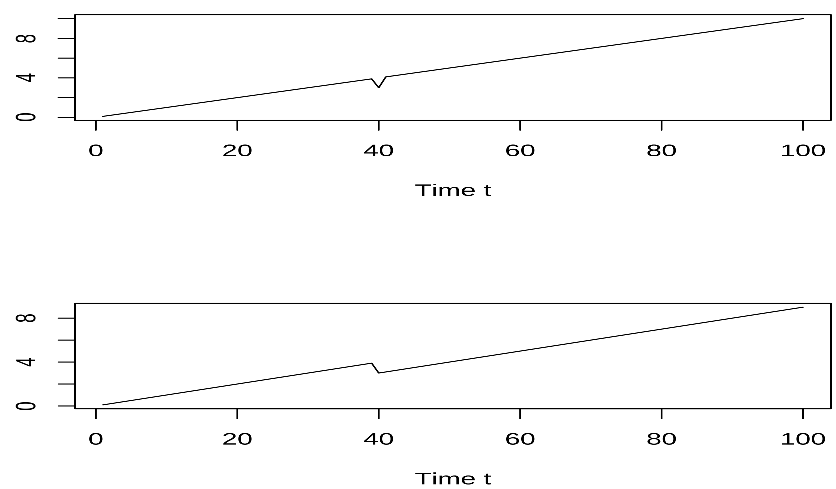
(b) Impulse input modifications of three values of randomly chosen sizes from the uniform density $SIZ_i \in [U_1, U_2]$ at randomly chosen three locations. Impulse modifications add SIZ_i only to the three values where modification is suggested by the random choice mechanism.

(c) Step (up or down) input modifications of three values of randomly chosen step sizes chosen from the uniform density $SIZ_s \in [U_1, U_2]$ at randomly chosen three locations. The step modifications add SIZ_s not only to the three values where modification is suggested by the random choice mechanism, but also to all larger (subsequent) values of j in the time series $x_{0j}, j \in [t, T]$.

It is convenient to refer to these modifications as (a) to (c) in our figures and tables. Figure 2 where the original series x_t is simply the sequence 0.1 to 10, with time $t = 1, \ldots, 100$. Both the impulse and step input modifications are of size -1 made at the location $t = 40$ only and displayed. It is not necessary to display the line of type (a) with no modification. This figure should clarify the distinction between our three modifications (a) to (c) described above.

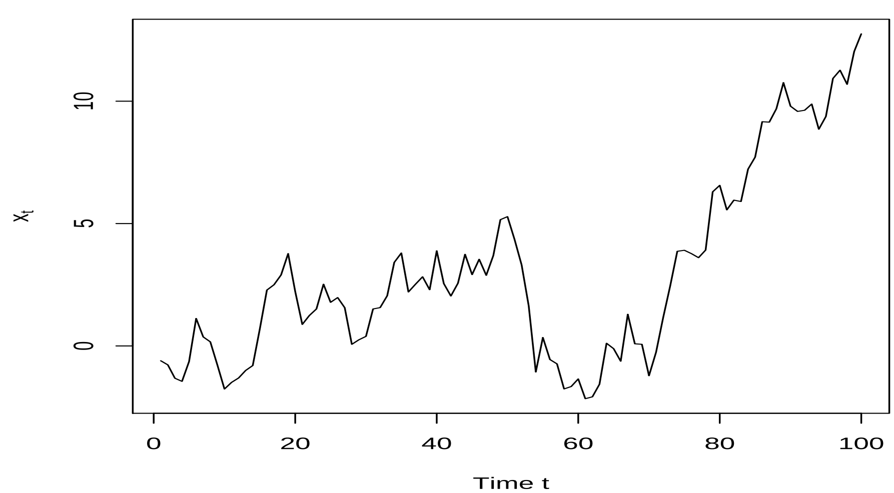
Random walk series of type (a) are illustrated by Figure 3. Impulse input modified series of type (b) are illustrated by Figure 4. Step input modified series of type (c) when the size shifts are all positive are illustrated by Figure 5, and those when the size shift can be negative are illustrated by Figure 6. These figures show that our simulation is working with plausible time series.

Figure 2: Typical modifications to artificial series $x_t = 0.1, 0.2, \ldots, 10$ at the location $t = 40$ of size -1. The top figure has impulse modification of type (b), and the bottom figure has step modification of type (c).



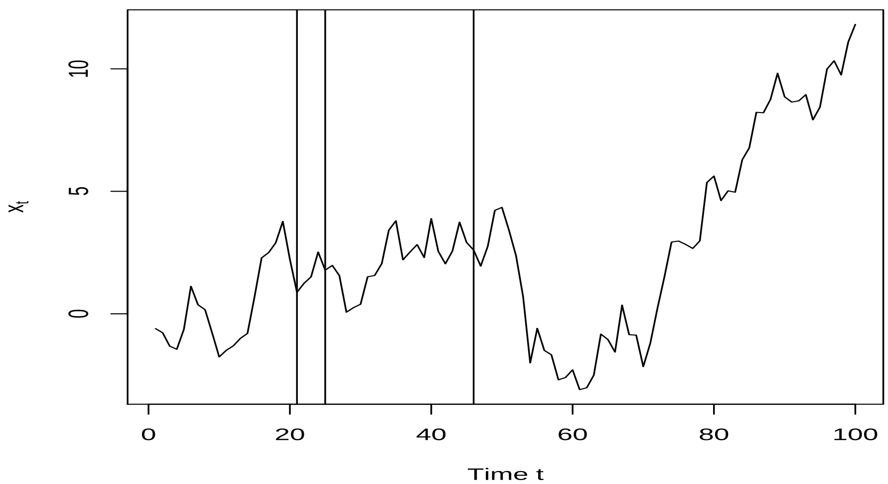
.

Figure 3: Typical time series for the random walk process.



.

Figure 4: Typical time series for three *impulse* perturbations to random walk at randomly chosen three time points of sizes -0.96, 1.87, and -0.94 $\in [-1, 3]$ respectively, indicated by added vertical axes.



.

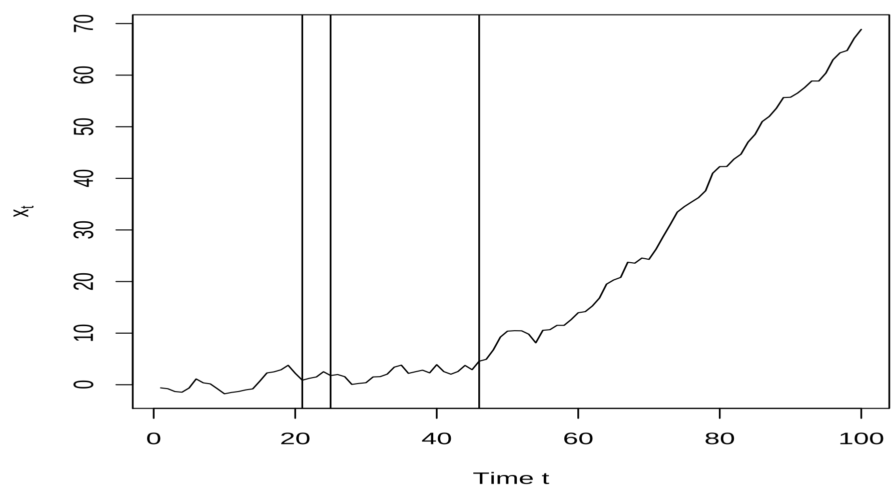
Now, we are ready to use the R package called 'meboot', Vinod and López-de-Lacalle (2009) to construct $J = 999$ resamples of x_t, y_t series. Unit root tests on resampled x_t, y_t series further confirm that they are $I(1)$ despite the impulse or step modifications, and hence remain subject to the spurious regression problem.

Next, we regress y_t on x_t for each resample to yield J coefficient estimates $b_j^*, j = 1, \ldots J$. Our simulation reports the following four types bootstrap confidence intervals. Denote by b the statistic, by β the (generic) parameter, by α the Type I error, and by $(1 - \alpha)$ the confidence level.

pctile) Naive percentile method based on ordering $b_j^*, j = 1, \ldots, J$ values from the smallest to the largest as $b_{(j)}^*, j = 1, \ldots J$. If $J = 999, \alpha = 0.05$ $(J+1)(\alpha/2) = 25$ and $(J+1)(1-\alpha/2) = 975$. Hence the 'pctile' interval is given by the order statistics: $[b_{(25)}^*, b_{(975)}^*]$.

bpctile) This interval improves upon the 'pctile' interval by working on a

Figure 5: Typical time series for *positive step* perturbations to random walk at randomly chosen three time points of sizes 1.01, 1.96 and 1.02 $\in [1, 2]$ respectively, indicated by added vertical axes.



.

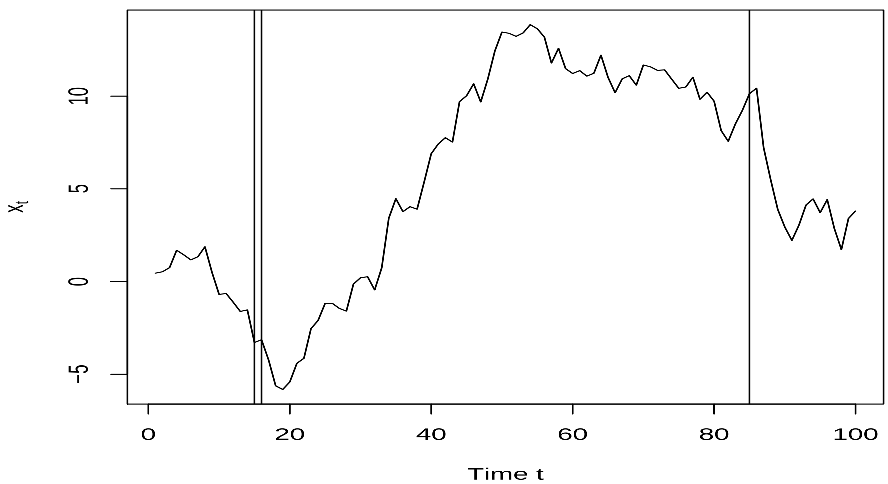
transformed scale to force the distribution of b^* to be symmetric, without knowing that transformation explicitly. See Davison and Hinkley (1997) p. 202.

norm) The 'norm' interval uses a normal approximation to the distribution of b based on bootstrap estimates b^* of the bias and variance described by Davison and Hinkley (1997) p. 14.

basic) The 'basic' interval uses the following basic notion to better approximate the 'norm' interval. Instead of directly using b^* to approximate the unknown β, the observable deviations $b^* - b$ are likely to be better at approximating the unknown deviations $b - \beta$. See Vinod (1993) p. 635 and Davison and Hinkley (1997) p. 28.

Our tables report four types of CI under the headings: 'pctile,' 'bptile,' 'norm' and 'basic'. Since the true value of the slope $\beta_1 = 2$, its CI should

Figure 6: Typical time series for *negative step* perturbations to random walk at randomly chosen three time points of sizes 0.71, 0.04 and -0.38 $\in [-1, 2]$ respectively, indicated by added vertical axes.



.

include (cover) the true value 2. Our experiments repeat the creation of CI $N = 500, 1000$ times. Normal theory yields the CI for two estimators: (i) OLS, (ii) OLSdiff (applying OLS to differenced data). We resample $J = 999$ times to construct CI for the two bootstrap estimators: (iii) BB and (iv) meboot. If $N_k, (k = i, ii, iii, iv)$, denotes the number of times the true slope is included inside the CI proposed by the k-th method, its coverage probability is (N_k/N). Note that when $k = iii, iv$ we have 'pctile,' 'bptile,' 'norm' and 'basic' distinct CI estimates and hence as many distinct coverage probabilities in our tables. The simulation also includes non-stationary $u_t \sim ARIMA(1, 1, 0)$ process making OLS estimate $\hat{\beta}_1$ inconsistent in a separate subsection.

Recall that we modify initial x_t by impulse or step defined above as (a) to (c). These random modifications depend on a range of sizes and the AR(1) parameter ρ used in the definitions of u_t in (3). We use row names to identify individual experiments in terms of these choices in all our tables. In addition to coverage probabilities, we report CI widths. Hence each experiment needs two lines in our tables.

Table 1 has two panels, upper and lower. Batches of two lines refer to a single experiment with $N = 1000, 500$ and $T = 150, 100$ in the upper and lower panels, respectively. We find that using a larger sample size $T = 150$ and larger $N = 1000$ in the upper panel does not offer any great advantage, while slowing the execution. Using R version 2.10.1 and meboot version 1.1-1 on a Dell Pentium 4 PC running at 2.9 Ghz, each simulation with $N = 500, T = 100$ took about three hours to finish. The odd numbered lines of Table 1 report coverage probabilities over the N evaluations of confidence intervals, and all even numbered lines report corresponding 95% confidence interval widths.

All tables describing the simulation experiments have three sets of columns. The first set identifies the nature of experiment indicating the type (a) to (c) of random walk modification including the limits U_1, U_2 within which the size of the modification must lie and the ρ used to define the AR(1) regression errors. The second set entitled OLS contains two columns for traditional OLS intervals on levels and on differenced data (denoted as $k = i, ii$ above).

Column headings 'OLS' and 'OLSdiff' in Table 1 contain results for models (1) and (2), respectively. Note that the CI widths for OLS are too low in each experiment, while OLS coverage probabilities are generally far less than 0.95. This confirms that spurious regression problem cannot be ignored, even when $\epsilon_t \sim I(0)$. By contrast, both the widths are large and coverage probabilities are at least close to 0.95 for 'OLSdiff', based on OLS applied to differenced data. This supports the current practice of differencing $I(1)$ series, except that differencing is subject to the drawbacks (d1) to (d4) listed earlier.

The remaining column headings in Table 1 refer to the four ways of getting the 'meboot' 95% confidence intervals as described above, whose (somewhat unintuitive) headings are based on the R package 'boot', Canty and Ripley

Table 1: Simulation results using 'meboot' with odd numbered lines having coverage probabilities for β_1 and even numbered lines having corresponding 95% confidence interval widths.

		OLS		meboot			
Line	$\rho, [U_1, U_2]$	OLS	OLSdiff	pctile	bpctile	norm	basic
1	0.5,[1, 3](c)	0.7640	0.9790	0.9990	0.9990	0.9980	1.0000
2	width(c)	0.0374	0.3025	0.3957	0.3971	0.3632	0.3971
3	0.5,[1, 3](b)	0.7680	0.9480	0.9910	0.9920	0.9900	0.9910
4	width(b)	0.0899	0.3734	0.2619	0.2629	0.2490	0.2629
5	0.6,[1, 3](c)	0.6720	0.9800	0.9960	0.9960	0.9940	0.9980
6	width(c)	0.0636	0.3656	0.4453	0.4468	0.4074	0.4468
7	0.5,[1, 3](c)	0.7480	0.9780	1.0000	1.0000	0.9960	0.9980
8	width(c)	0.0604	0.3775	0.4603	0.4619	0.4212	0.4619
9	0.5,[-1, 2](b)	0.7420	0.9780	0.9900	0.9900	0.9840	0.9860
10	width(b)	0.0927	0.4386	0.3797	0.3810	0.3497	0.3810
11	0.5,[-1, 2](b)	0.7280	0.9460	0.9740	0.9740	0.9680	0.9740
12	width(b)	0.1398	0.4699	0.3267	0.3279	0.3030	0.3279
13	0.1,(a)	0.9300	0.9660	1.0000	1.0000	1.0000	1.0000
14	width(a)	0.0950	0.5327	0.3461	0.3473	0.3192	0.3473

(a) random walk (RW), (b) impulse modified RW, (c) step modified RW.

(2009). Note that coverage probabilities all exceed 0.95 for all four bootstrap intervals for all experiments in Table 1.

We conducted a similar experiment focusing on the intercept and the ability of various methods to cover the true value 1 chosen in the simulation. We find that differencing leads to poor coverage properties for the intercept. In two experiments (four lines similar to those in Table 1) the coverage probability of the 'OLSdiff' column is zero. All one thousand intervals (in two experiments of $N = 500$ each, created after omitting 5% or 50 out of 999 extreme values) completely miss the true value 1 in our simulation. Since one is generally not interested in the intercept, we omit those results for brevity. If the researcher happens to be interested in the intercept, the traditional solution of differencing the data appears to be far worse than our meboot. Table 13 will discuss coverage of the intercept β_0 later.

Table 2 reports on a similar Monte Carlo experiment using the block bootstrap (BB). We construct confidence intervals for typical spurious regression problems where $R^2 > DW$ holds, using an analogous experimental design. The block size is chosen to be 5 in the upper panel of Table 2. The BB literature makes it clear that it is not suitable for nonstationary data such as in our nonstationary model. However, it is rather surprising that the coverage probabilities are all zero for all simulations. In other words, in thousands of tries, the BB intervals obtained after omitting 50 out of 999 realizations to determine 95% confidence intervals, we could not cover the true slope value of 2 even once. During the same experiments the OLS and differenced OLS intervals did cover the value over 60 and 95% of times, respectively.

The lower panel of Table 2 reports the performance of BB bootstrap with the block size 10. It might be possible to manipulate block sizes and other choices to nudge the coverage probabilities to higher values. Yet it is clear that the BB is not recommended for nonstationary data. This particular simulation is admittedly unfair to the BB method, since we are applying it to nonstationary data knowing fully well that BB is designed for m-dependent stationary data. Following section will consider cases where one or more of the regressors is stationary, while retaining the spurious regressions' rule of thumb: $R^2 > DW$.

2.1 Effect of Changing Confidence Levels on Coverage and Widths

The ME density approximates the population density of a population of time series subject to the following limitations.

L1] The ME density in the (v-dom) is a patchwork of uniform densities illustrated in Figure 1, which cannot become analogous to the familiar Normal density $\in (-\infty, \infty)$ used for traditional inference.

L2] The ME density as implemented in the meboot package has endpoint truncations.

L3] Theil and Laitinen (1980) prove that the variance of the ME density is:

$$\sigma_{ME}^2 = \sigma^2 - \frac{1}{4T}\sum_{i=1}^{T-1}(x_{i+1} - x_i)^2 - \frac{1}{24T}\sum_{i=2}^{T-1}(x_{i+1} - x_{i-1})^2, \tag{4}$$

where σ^2 denotes the variance of x_t. Note that both terms involving summations in (4) are strictly positive and therefore the variance of the ME density is smaller: $\sigma_{ME}^2 < \sigma^2$.

These three limitations might be why Table 1 coverage probabilities are all exceeding 0.95, when the *nominal* confidence level for bootstrap interval computations is set at 0.95. This subsection explores choosing 0.85 or 0.80 as *nominal* confidence levels to achieve the desired 0.95 coverage probability.

Table 3 reports the results for the case where we set the confidence level at 0.85. Table 4 has the confidence level set at 0.80. These Tables show that it might be useful to modify the usual level from 0.95 to 0.80 or 0.85 to make the coverage probability closer to the usual 95%.

Note that the widths reported along the even numbered lines in Table 3 are generally larger than the ones reported in Table 4. In these experiments the confidence level was set at 0.85 and 0.80 only for the meboot case. The confidence level was retained at 0.95 for the two OLS cases, since the OLS inference does not use our ME density. Note that the direction of changes in confidence interval widths along with coverage probabilities show that the simulation is behaving as expected.

These experiments suggest that setting the nominal confidence level 0.80 for the ME density bootstrap algorithms gives the desired type I error of $\alpha = 0.05$. We conclude this subsection by noting that it is possible to choose a suitable nominal confidence level for meboot intervals for achieving the desired type I errors. However, since 95% CI coverges exceed 0.95, a practitioner need not specify a lower confidence level unless the CI widths are too long.

2.2 Nonstationary errors and inconsistent OLS

Spurious regression can be narrowly defined in the sense that regression errors are random walk or nonstationary with unbounded variance, Hamilton (1994). In this case OLS estimation is inconsistent. Note that $(\hat{\beta}_1 - \beta_1)$ equals a ratio whose numerator is: $Num = (1/T)\sum_{t=1}^{T}(x_t - \bar{x}_t)\epsilon_t$, and whose denominator is: $Den = (1/T)\sum_{t=1}^{T}(x_t - \bar{x}_t)^2$. Since both x_t and ϵ_t have stochastic trends or are random walks, the law of large numbers as well as the central limit theorem fail, and the ratio converges to a random number instead of converging to zero as $T \to \infty$.

This subsection simulates the ARIMA(1,1,0) errors case. Hence the discussion here is relevant only if $\epsilon \sim I(1)$ is confirmed by unit root tests on regression residuals.

Note that Table 5 based on ARIMA(1,1,0) errors has a nominal 95% confidence level setting for both meboot and OLS. Only when meboot coverage probabilities exceed the 0.95 setting, we can say that meboot is working well here for the nonstationary errors case. Note that the coverage probabilities under 'pctile' (meboot naive bootstrap intervals) are (97, 93, 81, 76, 96)% which are below 0.95 in majority of cases suggesting that meboot 'pctile' is not working well here. All other intervals are also not working very well here. The performance of block bootstrap (BB) in this case is found to be much worse than meboot with coverage probabilities all zero (detailed table is omitted for brevity).

Note that Table 6 based on ARIMA(1,1,0) errors has 80% confidence level only for meboot while keeping the level for OLS and differenced OLS at the usual 95% level, similar to Table 4 of the previous subsection. Only when meboot coverage probabilities exceed the 0.80 setting, we can say that meboot is working well here for the nonstationary errors case. Note that the coverage probabilities under 'pctile' (meboot naive bootstrap intervals) are (89, 78, 57, 53, 80)% which are below 0.80 in majority of cases suggesting that meboot 'pctile' is not working well here. The 'norm' intervals are also not working well here. But 'bptile' and 'basic' coverage probabilities exceeding 0.80 in all cases, except for line number 7, where it is 0.77 (perhaps close enough to 0.8). We find that meboot is certainly an improvement over OLS in levels and over block bootstrap, but not over differenced OLS. This means that meboot is partly working, while it's performance does leave something to be desired in situations where OLS is used even when OLS is known to be inconsistent.

Now consider a simulation where generalized least squares (GLS) is used [recall Hamilton's cure (c3) above] instead of OLS, without changing the original specification. Following Blough (1992) we use of Cochrane-Orcutt type correction for AR(1) errors even though actual regression errors (known in our simulation) are ARIMA(1,1,0) not AR(1). It is of interest to know whether a similar correction using a bi-diagonal matrix from page 425 of Vinod (2008) helps here. Compared to the option of adding lagged variables, an appeal of this cure to me is that it does not change economists' original specification. Using a more efficient GLS estimator, instead of OLS, allows for nonspherical errors without changing the specification.

A simulation using the GLS is obviously more computer intensive, since each estimate needs two steps. First step estimates the residuals and fits an AR(1) model to the residuals to estimate ρ. Second step estimates the GLS regression using the estimated AR(1) parameter. Table 7 shows that meboot does work when AR(1) errors are corrected by using Cochrane-Orcutt type GLS based on estimated AR(1) coefficient for residuals. The last two columns for meboot entitled 'norm' and 'basic' report coverage probabilities all exceeding 0.95. Perhaps, this is the first ever simulation of the Blough's suggestion which took about eleven hours to complete.

A simulation similar to Table 6 shows the following meboot coverage probabilities: (Line 1: 0.70, 0.94, 1.00, 0.99), (Line 3: 0.65, 0.90, 0.97, 0.96), (Line 5: 0.64, 0.88, 0.96, 0.96), (Line 7: 0.82, 0.99, 1.00, 1.00). It is clear that except for naive pctile interval in the first column, all meboot intervals exceed the assigned value 0.80. This suggests that if inconsistency of OLS is present due to nonstationary regression errors, Cochrane-Orcutt type correction for AR(1) errors does help meboot intervals to become conservatively reliable. That is, generalized least squares 80% (GLS) mebbot confidence intervals (except pctile) also cover the true slope with probability exceeding 0.80.

3 Monte Carlo Simulation of Two Regressor Models

Even though the theory of spurious regressions focuses on equations of type (1), its practical impact goes far beyond models with only one regressor. Hence, our simulations will be incomplete unless we consider a two regressor model discussed in this section.

First, let us focus on the following spurious regression:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 \tau_t + \epsilon_t, \tag{5}$$

where the original variable x_t in levels remains (integrated) nonstationary series with modifications (a) to (c), as in the previous section, while including an additional regressor for time $\tau_t = 1, \ldots, T$. The simulation will use an equation similar to (3), $y_t = 1 + 2x_t + 3\tau_t + u_t$, with the AR(1) generated u_t as before and setting the true value $\beta_2 = 3$. Since τ_t is also nonstationary, the block bootstrap is not expected to work well here.

Table 8 considers the coverage for the main slope coefficient when the second regressor as the time variable is present for the $T = 100$ case. The coverage probabilities for all cases in the last four columns exceed 0.95, suggesting that meboot remains conservatively reliable, except that the widths are relatively large. It is interesting that OLS applied to differenced data (popular in econometrics) has reasonable coverage in columns entitled "OLS-diff", although slightly less than 0.95 in some cases. By contrast, the OLS on levels data in column entitled "OLS" cover only about (66, 76, 77, 92)%

times respectively, far short of the 95% needed. Next we consider the same situation as this Table, except with shorter time series with $T = 50$.

Table 9 considers the coverage for the main slope coefficient when the second regressor as the time variable is present for the $T = 50$ case. The coverage probabilities for all cases in the last four columns exceed 0.95, suggesting that meboot remains conservatively reliable for the $T = 50$ case, except that the widths are large. Of course, widths can be reduced by choosing smaller confidence levels (< 0.95). However, OLS applied to differenced data has reasonable coverage in the column entitled "OLSdiff". The OLS on levels data in column entitled "OLS" cover only about (69, 77, 79, 78, 92)% times respectively, short of the 95% needed when $T = 50$. Next, let us consider the case where we compute the coverage probabilities for the coefficient of τ_t.

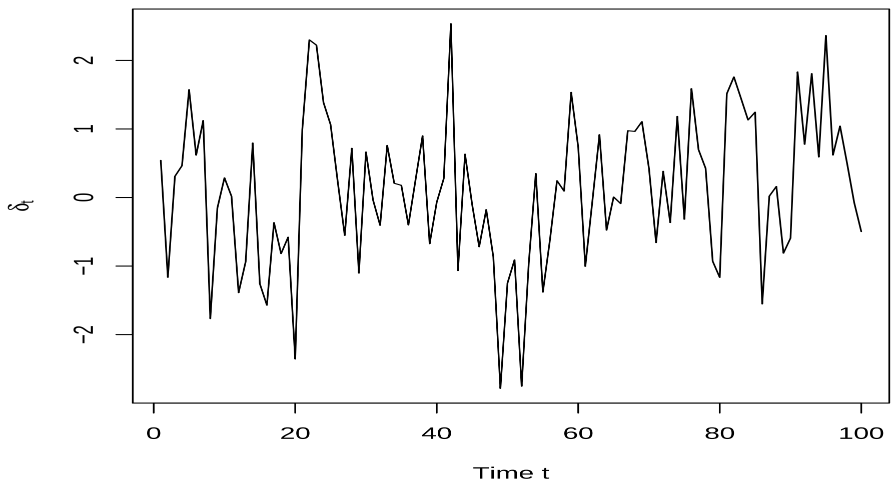
Table 10 considers the case when the focus is on the coverage for the coefficient for the second regressor τ_t in (5) or the time variable. The coverage probabilities for all cases in the last four columns exceed 0.95, suggesting that the meboot remains conservative and reliable at the cost of larger widths. Note that OLS applied to differenced data has zero coverage in the column entitled "OLSdiff". This suggests that these OLS confidence intervals do not cover the true value of 3 almost at all. Again, as before, the OLS on levels data in column entitled "OLS" cover only about (66, 76, 75, 93)% times respectively, far short of the 95% needed. Thus OLS inference remains unreliable for the model (5), an extension of (1).

Next, we focus on the following spurious regression:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 \delta_t + \epsilon_t, \tag{6}$$

where the original variable x_t in levels remains (integrated) nonstationary series with modifications (a) to (c), similar to those of the previous section along with an additional regressor $\delta_t \sim FI(d = 0.4)$, a fractionally integrated long memory variable with the long memory parameter $d = 0.4$, implying that it is stationary. Figure 7 is a typical plot for δ_t, where no trending is visible.

Figure 7: Typical time series for long memory fractionally integrated $\delta_t \sim FI(d = 0.4)$.



.

Since δ_t is stationary, the block bootstrap may have a better chance to work well for the model in (6).

Table 11 considers the coverage for the main slope coefficient when the second regressor $\delta_t \sim FI(d = 0.4)$ the fractionally integrated long memory variable is present with $T = 100$. The coverage probabilities for all cases in the last four columns exceed 0.95, suggesting that meboot remains conservatively reliable. Again, OLS applied to differenced data has reasonable coverage in the column entitled "OLSdiff", although it is a bit short of 0.95 along two rows. By contrast, the OLS on levels data in column entitled "OLS" cover only about (71, 74, 76, 93)% times respectively, far short of the 95% needed. It is of interest to know whether the performance of block bootstrap improves when the regressor is stationary, although not $m-$dependent. We use the shorter time series $T = 50$ to improve the chances for BB.

Table 12 considers the coverage for the main slope coefficient when the second regressor $\delta_t \sim FI(d = 0.4)$ is present for the $T = 50$ case. The coverage probabilities for all cases in the last four columns are far short of 0.95 while using larger confidence interval widths. Hence, BB remains unacceptable for the $T = 50$ case, even if the fractionally integrated regressor is stationary. It is interesting that OLS applied to differenced data has good coverage in the column entitled "OLSdiff". By contrast, the OLS on levels data in column entitled "OLS" cover only about (74, 79, 78, 76, 94)% times respectively, short of the 95% needed when $T = 50$. Although coverage probabilities for the slope β_1 remain extremely low, they are a bit better than the zeros found earlier.

What if we change the focus on the coverage probabilities for the intercept? Table 13 considers the coverage for the intercept when the second regressor $\sim FI(d = 0.4)$ is present for the $T = 50$ case. The coverage probabilities for all cases in the last four columns are generally above 0.4, but far short of 0.95, suggesting that BB remains unacceptable (perhaps not as bad as Table 12 where the slope is the focus of interest) for the intercept in the $T = 50$ case, even if the fractionally integrated regressor is stationary. The slightly improved coverage is purchased at the cost of much wider BB intervals. In the lower panel where we have forced the first regressor x_t of (6) is stationary from a simulated AR(1) model having $\rho = 0.5, 0.1$ autoregressive parameters, respectively. Thus BB is seen to be unsatisfactory for models having one variable long memory δ_t and the other variable stationary AR(1) with the autoregressive parameter $|\rho| < 1$.

Now, OLS applied to differenced data has zero coverage in the column entitled "OLSdiff" in the upper panel but good coverage in the lower panel. By contrast, the OLS on levels data in column entitled "OLS" cover only about (74, 77, 79, 80, 94)% times in the upper panel and (79, 93)% in the lower panel, still short of the 95% needed.

4 Summary and Final Remarks

Since the block bootstrap assumes stationary data, the absence of a bootstrap for state-dependent nonstationary data has been a long standing gap, only recently filled by the Maximum Entropy bootstrap (meboot) described in Vinod (2006). This paper begins with a short introduction to the ME density and the meboot algorithm.

More important, this paper reports several Monte Carlo experiments where the regressor x_t is constructed as a random walk series with and without random impulse or step contaminations of random sizes at randomly chosen three points for sample sizes $T = 50, 100$. Our graphs reveal that all series in our experiments are plausible as data series. The simulated y_t equals $1 + 2x_t + 3z_t + u_t$, where z_t is mostly absent till Section 3, where it is either time itself or a long-memory (fractionally integrated) series.

The simulation mostly focuses on the known slope (=2) of x_t and its confidence intervals (CI) estimated by four methods. Normal theory yields the CI for two estimators: (i) OLS, (ii) OLSdiff (applying OLS to differenced data). We resample $J = 999$ times to construct CI for the two bootstrap estimators: (iii) BB and (iv) meboot. If $N_k, (k = i, ii, iii, iv)$, denotes the number of times the true slope is included inside the CI proposed by the k-th method, its coverage probability is (N_k/N), by considering $N = 500$ experiments.

A practitioner may define spurious regression problems by the rule of thumb: $R^2 > DW$. However, it is important to distinguish between two cases. First, when residuals u_t (used in creating y_t) are stationary, then OLS is consistent. Second, when $u_t \sim ARIMA(1, 1, 0)$, or otherwise nonstationary, then OLS is inconsistent. Using unit root testing, it is a simple matter to assess whether regression residuals are stationary or not.

Our extensive Monte Carlo simulations, within the limitations of any simulation, suggest that meboot confidence intervals are reliable, and can be generally recommended when OLS is consistent. They outperform the competition and have the advantage that one need not change the original specification (in levels). If OLS is inconsistent due to nonstationary errors, the performance of meboot remains good, (always superior to OLS and BB intervals) but not always superior to OLSdiff. Our Section 1 lists four drawbacks of OLSdiff.

We include (perhaps a first ever) simulatoin of Blough's suggestion (cited by Hamilton (1994)) to solve spurious regression by using GLS based on Cochrane-Orcutt type correction for AR(1) regression errors. The simulation suggests that meboot applied to GLS can be recommended.

The advantage of meboot over other methods for dealing with spurious regression in the literature is that the original economic specification need not be changed, even when GLS is used. Thus, we can begin to free the researcher from always having to use differencing when the available data have near unit roots or other forms of nonstationarity such as long memory. In the modern era of computers we are not bound by the onerous requirement of the 1930's WKK model for time series inference, sometimes forcing us to transform all our series into stationary series.

References

S. R. Blough. Spurious regressions with ar(1) correction and unit root pretest. Johns Hopkins University, 1992.

A. Canty and B. Ripley. *boot: Bootstrap R (S-Plus) Functions. R package version 1.2-41.*, 2009.

A C Davison and D V Hinkley. *Bootstrap Methods and Their Application.* Cambridge: Cambridge University Press, 1997.

J. D. Hamilton. *Time Series Analysis.* Princeton University Press, 1994.

A. Koutris, M. S. Heracleous, and A. Spanos. Testing for nonstationarity using maximum entropy resampling: A misspecification testing perspective. *Econometric Reviews*, 27(4):363–384, 2008.

S. N. Lahiri. *Resampling Methods for Dependent data*. Springer Verlag, New York, 2003.

R. Y. Liu. Bootstrap procedures under some non-iid models. *The Annals of Statistics*, 16(4):1696–1708, 1988.

C R Nelson and C I Plosser. Trends and random walks in macroeconomic time series. *Journal of Monetary Economics*, 10:139–162, 1982.

A. E Noriega and D. Ventosa-Santaularia. Spurious regression under broken-trend stationarity. *Journal of Time Series Analysis*, 27(8):671–684, 2006.

P.C.B. Phillips. Understanding spurious regressions in econometrics. *Journal of Econometrics*, 33:311–340, 1986.

A Spanos and A Kourtellos. Model validation and resampling. Working paper, Economics Department, Virginia Tech., 2002.

J H Stock and M W Watson. A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica*, 61:783–820, 1993.

H Theil and K Laitinen. Singular moment matrices in applied econometrics. In PR Krishnaiah, editor, *Multivariate Analysis – V*, pages 629–649. North USA-Holland Publishing Co., New York, USA, 1980.

H D Vinod. Bootstrap methods: Applications in econometrics. In G S Maddala, C R Rao, and H D Vinod, editors, *Handbook of Statistics: Econometrics*, volume 11, chapter 23, pages 629–661. North Holland, Elsevier, New York, 1993.

H D Vinod. Maximum entropy ensembles for time series inference in economics. *Journal of Asian Economics*, 17(6):955–978, 2006.

H. D. Vinod. Superior estimation and inference avoiding heteroscedasticity and flawed pivots: R-example of inflation unemployment trade-off. In H. D. Vinod, editor, *Advances in Social Science Research Using R*, pages 39–63. Springer, New York, 2010.

H. D. Vinod and Javier López-de-Lacalle. Maximum entropy bootstrap for time series: The meboot R package. *Journal of Statistical Software*, 29(5): 1–19, 2009. URL `http://www.jstatsoft.org/v29/i05/`.

Hrishikesh D. Vinod. *Hands-on Intermediate Econometrics Using R: Templates for Extending Dozens of Practical Examples*. World Scientific, Hackensack, NJ, 2008. URL `http://www.worldscibooks.com/economics/6895.html`. ISBN 10-981-281-885-5.

Table 2: BB simulation Results, odd numbered lines having coverage probabilities for β_1 and even numbered lines having corresponding 95% confidence interval widths. Upper and Lower panels have block sizes 5, 10.

		OLS		Block Bootstrap			
Line	$\rho, [U_1, U_2]$	OLS	OLSdiff	pctile	bpctile	norm	basic
1	0.6,[1, 3](c)	0.69	0.96	0.00	0.00	0.00	0.00
2	width(c)	0.06	0.36	1.63	1.63	1.62	1.63
3	0.5,[1, 3](c)	0.76	0.96	0.00	0.00	0.00	0.00
4	width(c)	0.09	0.44	1.62	1.63	1.62	1.63
5	0.6,[-1, 2](c)	0.69	0.96	0.00	0.00	0.00	0.00
6	width(c)	0.06	0.36	1.63	1.63	1.62	1.63
7	0.5,[-1, 2](b)	0.77	0.95	0.00	0.00	0.00	0.00
8	width(b)	0.14	0.47	1.59	1.60	1.60	1.60
9	0.5,[-1, 2](b)	0.77	0.95	0.00	0.00	0.00	0.00
10	width(b)	0.14	0.47	1.59	1.60	1.60	1.60
11	0.1,(a)	0.91	0.95	0.00	0.00	0.00	0.00
12	width(a)	0.12	0.55	1.58	1.58	1.58	1.58
1	0.5,[-1, 2](b)	0.80	0.98	0.00	0.00	0.00	0.00
2	width(b)	0.14	0.47	2.04	2.04	2.04	2.04
3	0.5,[-1, 2](b)	0.80	0.98	0.00	0.00	0.00	0.00
4	width(b)	0.14	0.47	2.04	2.04	2.04	2.04
5	0.1,(a)	0.94	0.95	0.00	0.00	0.00	0.00
6	width(a)	0.12	0.55	2.01	2.01	2.01	2.01

(a) random walk (RW), (b) impulse modified RW, (c) step modified RW.

Table 3: Meboot simulation, odd numbered lines having coverage probabilities for the second slope β_2, and even numbered lines having corresponding 85% confidence interval widths. T=100.

		OLS		meboot			
Line	$\rho, [U_1, U_2]$	OLS	OLSdiff	pctile	bpctile	norm	basic
1	0.6,[1, 3](c)	0.6720	0.9800	0.9780	0.9780	0.9600	0.9600
2	width(c)	0.0636	0.3656	0.2965	0.2971	0.2992	0.2971
3	0.5,[-1, 2](c)	0.7420	0.9780	0.9400	0.9400	0.9640	0.9520
4	width(c)	0.0927	0.4386	0.2521	0.2525	0.2568	0.2525
5	0.5,[-1, 2](b)	0.7280	0.9460	0.8740	0.8760	0.8960	0.8740
6	width(b)	0.1410	0.4674	0.2154	0.2158	0.2230	0.2158
7	0.5,(a)	0.7380	0.9680	0.9020	0.9040	0.8940	0.8760
8	width(a)	0.1398	0.4699	0.2158	0.2162	0.2225	0.2162
9	0.1,(a)	0.9320	0.9540	0.9800	0.9800	0.9900	0.9840
10	width(a)	0.1228	0.5444	0.2156	0.2160	0.2225	0.2160

(a) random walk (RW), (b) impulse modified RW, (c) step modified RW.

Table 4: Meboot simulation, odd numbered lines having coverage probabilities for the second slope β_2, and even numbered lines having corresponding 80% confidence interval widths. T=100.

		OLS		meboot			
Line	$\rho, [U_1, U_2]$	OLS	OLSdiff	pctile	bpctile	norm	basic
1	0.6,[1, 3](c)	0.6720	0.9800	0.9620	0.9620	0.9480	0.9380
2	width(c)	0.0636	0.3656	0.2522	0.2526	0.2664	0.2526
3	0.5,[-1, 2](c)	0.7420	0.9780	0.9080	0.9080	0.9420	0.9280
4	width(c)	0.0927	0.4386	0.2148	0.2152	0.2286	0.2152
5	0.5,[-1, 2](b)	0.7280	0.9460	0.8200	0.8220	0.8640	0.8320
6	width(b)	0.1410	0.4674	0.1845	0.1848	0.1985	0.1848
7	0.5,(a)	0.7380	0.9680	0.8420	0.8440	0.8500	0.8200
8	width(a)	0.1398	0.4699	0.1845	0.1847	0.1981	0.1847
9	0.1,(a)	0.9320	0.9540	0.9580	0.9580	0.9800	0.9760
10	width(a)	0.1228	0.5444	0.1842	0.1845	0.1980	0.1845

(a) random walk (RW), (b) impulse modified RW, (c) step modified RW.

Table 5: Meboot simulation, odd numbered lines having coverage probabilities for the second slope β_2, and even numbered lines having corresponding 95% confidence interval widths. T=100. ARIMA(1,1,0) nonstationary errors.

		OLS		meboot			
Line	$\rho, [U_1, U_2]$	OLS	OLSdiff	pctile	bpctile	norm	basic
1	0.6,[1, 3](c)	0.1400	0.8300	0.9650	0.9650	0.9650	0.9750
2	width(c)	0.0403	0.0391	0.4540	0.4555	0.4156	0.4555
3	0.5,[-1, 2](c)	0.2400	0.9000	0.9300	0.9300	0.9050	0.9150
4	width(c)	0.0474	0.0436	0.3637	0.3649	0.3345	0.3649
5	0.5,[-1, 2](b)	0.2600	0.9400	0.8100	0.8100	0.7750	0.7850
6	width(b)	0.0785	0.0462	0.3161	0.3173	0.2934	0.3173
7	0.5,(a)	0.2050	0.9400	0.7550	0.7550	0.7300	0.7500
8	width(a)	0.0750	0.0457	0.3228	0.3240	0.2989	0.3240
9	0.1,(a)	0.2550	0.9500	0.9600	0.9600	0.9500	0.9550
10	width(a)	0.0445	0.0401	0.3172	0.3184	0.2949	0.3184

(a) random walk (RW), (b) impulse modified RW, (c) step modified RW.

Table 6: Meboot simulation, odd numbered lines having coverage probabilities for the second slope β_2, and even numbered lines having corresponding 80% confidence interval widths. T=100. ARIMA(1,1,0) nonstationary errors.

		OLS		meboot			
Line	$\rho, [U_1, U_2]$	OLS	OLSdiff	pctile	bpctile	norm	basic
1	0.6,[1, 3](c)	0.1700	0.8420	0.8920	0.9560	0.9620	0.9660
2	width(c)	0.0404	0.0393	0.2588	0.4565	0.4167	0.4565
3	0.5,[-1, 2](c)	0.2180	0.9000	0.7800	0.9380	0.9120	0.9200
4	width(c)	0.0483	0.0433	0.2102	0.3742	0.3430	0.3742
5	0.5,[-1, 2](b)	0.2420	0.9280	0.5660	0.8260	0.7900	0.8080
6	width(b)	0.0742	0.0458	0.1811	0.3214	0.2971	0.3214
7	0.5,(a)	0.1960	0.9500	0.5340	0.7740	0.7440	0.7680
8	width(a)	0.0770	0.0461	0.1829	0.3249	0.3002	0.3249
9	0.1,(a)	0.2560	0.9540	0.7980	0.9420	0.9360	0.9420
10	width(a)	0.0451	0.0403	0.1821	0.3237	0.2993	0.3237

(a) random walk (RW), (b) impulse modified RW, (c) step modified RW.

Table 7: Meboot simulation, odd numbered lines having coverage probabilities for the second slope β_2, and even numbered lines having corresponding 95% confidence interval widths. T=100. ARIMA(1,1,0) nonstationary errors corrected with GLS based on estimated AR(1).

		OLS		meboot			
Line	$\rho, [U_1, U_2]$	OLS	OLSdiff	pctile	bpctile	norm	basic
3	0.5,[-1, 2](c)	0.2840	0.9560	0.9420	0.9420	0.9960	0.9920
4	width(c)	0.0800	0.0621	0.6727	0.6749	0.6455	0.6749
5	0.5,[-1, 2](b)	0.3140	0.9480	0.9000	0.9020	0.9680	0.9580
6	width(b)	0.1084	0.0655	0.5756	0.5777	0.5470	0.5777
7	0.5,(a)	0.2980	0.9460	0.8820	0.8820	0.9640	0.9640
8	width(a)	0.1108	0.0663	0.5851	0.5871	0.5550	0.5871
9	0.1,(a)	0.3560	0.9520	0.9860	0.9860	0.9980	0.9980
10	width(a)	0.0627	0.0589	0.5736	0.5757	0.5404	0.5757

(a) random walk (RW), (b) impulse modified RW, (c) step modified RW.

Table 8: Meboot simulation, odd numbered lines having coverage probabilities for the second slope β_2, and even numbered lines having corresponding 95% confidence interval widths. Time τ_t is second regressor, T=100.

		OLS		meboot			
Line	$\rho, [U_1, U_2]$	OLS	OLSdiff	pctile	bpctile	norm	basic
1	0.6,[1, 3](c)	0.6620	0.9580	1.0000	1.0000	1.0000	1.0000
2	width(c)	0.0880	0.4063	2.7837	2.7915	2.7831	2.7915
3	0.5,[-1, 2](c)	0.7620	0.9480	1.0000	1.0000	1.0000	1.0000
4	width(c)	0.1372	0.4545	3.9282	3.9394	3.9193	3.9394
5	0.5,[-1, 2](b)	0.7720	0.9480	1.0000	1.0000	1.0000	1.0000
6	width(b)	0.1976	0.4744	4.9447	4.9583	4.9373	4.9583
7	0.1,(a)	0.9220	0.9300	1.0000	1.0000	1.0000	1.0000
8	width(a)	0.1725	0.5508	4.8640	4.8780	4.8510	4.8780

(a) random walk (RW), (b) impulse modified RW, (c) step modified RW.

Table 9: Meboot simulation, odd numbered lines having coverage probabilities for β_1 and even numbered lines having corresponding 95% confidence interval widths. Time τ_t as second regressor, T=50.

		OLS		meboot			
Line	$\rho, [U_1, U_2]$	OLS	OLSdiff	pctile	bpctile	norm	basic
1	0.6,[1, 3](c)	0.6940	0.9800	1.0000	1.0000	1.0000	1.0000
2	width(c)	0.1959	0.5962	3.2411	3.2501	3.2412	3.2501
3	0.5,[-1, 2](c)	0.7740	0.9420	1.0000	1.0000	1.0000	1.0000
4	width(c)	0.3036	0.6865	4.1685	4.1809	4.1514	4.1809
5	0.5,[-1, 2](b)	0.7880	0.9380	1.0000	1.0000	0.9960	0.9960
6	width(b)	0.3788	0.6962	4.7206	4.7336	4.6935	4.7336
7	0.5,(a)	0.7800	0.9460	1.0000	1.0000	0.9960	0.9960
8	width(a)	0.3785	0.6989	4.7435	4.7554	4.7309	4.7554
9	0.1,(a)	0.9180	0.9580	0.9960	0.9960	1.0000	1.0000
10	width(a)	0.3610	0.8202	4.7178	4.7306	4.6920	4.7306

(a) random walk (RW), (b) impulse modified RW, (c) step modified RW.

Table 10: Meboot simulation, odd numbered lines having coverage probabilities for β_2 coefficient of τ_t or time and even numbered lines having corresponding 95% confidence interval widths.

		OLS		meboot			
Line	$\rho, [U_1, U_2]$	OLS	OLSdiff	pctile	bpctile	norm	basic
1	0.6,[1, 3](c)	0.6600	0.0000	0.9920	0.9920	1.0000	0.9980
2	width(c)	0.0308	0.0177	1.4048	1.4086	1.3991	1.4086
3	0.5,[-1, 2](c)	0.7620	0.0000	1.0000	1.0000	1.0000	0.9980
4	width(c)	0.0269	0.0168	0.8271	0.8294	0.8226	0.8294
5	0.5,[-1, 2](b)	0.7480	0.0000	1.0000	1.0000	1.0000	0.9920
6	width(b)	0.0245	0.0164	0.5070	0.5085	0.5025	0.5085
7	0.1,(a)	0.9260	0.0000	0.9960	0.9960	1.0000	0.9980
8	width(a)	0.0211	0.0190	0.4911	0.4925	0.4866	0.4925

(a) random walk (RW), (b) impulse modified RW, (c) step modified RW.

Table 11: Meboot simulation, odd numbered lines having coverage probabilities for β_2 and even numbered lines having corresponding 95% confidence interval widths. Second regressor $\delta_t \sim FI(d = 0.4)$.

		OLS		meboot			
Line	$\rho, [U_1, U_2]$	OLS	OLSdiff	pctile	bpctile	norm	basic
1	0.6,[1, 3](c)	0.7060	0.9640	0.9940	0.9940	0.9900	0.9920
2	width(c)	0.0642	0.3623	0.4521	0.4536	0.4178	0.4536
3	0.5,[-1, 2](c)	0.7380	0.9440	0.9980	0.9980	0.9920	0.9960
4	width(c)	0.0952	0.4421	0.3952	0.3966	0.3678	0.3966
5	0.5,[-1, 2](b)	0.7560	0.9400	0.9860	0.9860	0.9820	0.9840
6	width(b)	0.1397	0.4646	0.3644	0.3656	0.3416	0.3656
7	0.1,(a)	0.9300	0.9560	1.0000	1.0000	1.0000	1.0000
8	width(a)	0.1224	0.5481	0.3636	0.3649	0.3417	0.3649

(a) random walk (RW), (b) impulse modified RW, (c) step modified RW.

Table 12: BB simulation, odd numbered lines having coverage probabilities for β_1 and even numbered lines having corresponding 95% confidence interval widths. Second regressor $\delta_t \sim FI(d = 0.4)$, Block size 10, $T = 50$.

		OLS		Block Bootstrap			
Line	$\rho, [U_1, U_2]$	OLS	OLSdiff	pctile	bpctile	norm	basic
1	0.6,[1, 3](c)	0.7420	0.9720	0.0020	0.0020	0.0020	0.0020
2	width(c)	0.1504	0.5336	2.5862	2.5925	2.5920	2.5925
3	0.5,[-1, 2](c)	0.7860	0.9640	0.0100	0.0100	0.0180	0.0160
4	width(c)	0.2069	0.6399	2.6618	2.6682	2.6853	2.6682
5	0.5,[-1, 2](b)	0.7780	0.9560	0.0080	0.0080	0.0280	0.0280
6	width(b)	0.2761	0.6760	2.6911	2.6977	2.7235	2.6977
7	0.5,(a)	0.7620	0.9640	0.0160	0.0160	0.0320	0.0320
8	width(a)	0.2795	0.6831	2.6934	2.6996	2.7264	2.6996
9	0.1,(a)	0.9380	0.9660	0.0180	0.0180	0.0320	0.0280
10	width(a)	0.2659	0.8069	2.6721	2.6787	2.6965	2.6787

(a) random walk (RW), (b) impulse modified RW, (c) step modified RW.

Table 13: BB simulation, odd numbered lines having coverage probabilities for the intercept β_0 and even numbered lines having corresponding 95% confidence interval widths. Second regressor$\sim FI(d = 0.4)$, Block size 10, $T = 50$ Lower panel focuses on β_1 with $x_t \sim AR(1)$ having $\rho = 0.5, 0.1$ values.

		OLS		Block Bootstrap			
Line	$\rho, [U_1, U_2]$	OLS	OLSdiff	pctile	bpctile	norm	basic
1	0.6,[1, 3](c)	0.7360	0.0000	0.5440	0.5440	0.6520	0.5440
2	width(c)	1.1079	0.7112	29.9814	30.0486	31.0534	30.0486
3	0.5,[-1, 2](c)	0.7660	0.0000	0.4920	0.4920	0.5600	0.5080
4	width(c)	1.0817	0.6905	18.3226	18.3625	18.8602	18.3625
5	0.5,[-1, 2](b)	0.7860	0.0000	0.4500	0.4500	0.4840	0.4620
6	width(b)	1.1652	0.6690	12.8144	12.8424	13.2051	12.8424
7	0.5,(a)	0.7960	0.0000	0.4420	0.4420	0.4840	0.4640
8	width(a)	1.1722	0.6690	12.7379	12.7652	13.1309	12.7652
9	0.1,(a)	0.9420	0.0000	0.4720	0.4740	0.4920	0.4720
10	width(a)	1.0858	0.7835	12.2340	12.2599	12.6057	12.2599
11	x_t ∼AR(0.5)	0.7940	0.9700	0.0180	0.0200	0.0340	0.0400
12	width	0.3710	0.6730	2.6125	2.6190	2.6366	2.6190
13	x_t ∼AR(0.1)	0.9340	0.9480	0.0180	0.0180	0.0220	0.0220
14	width	0.3372	0.7777	2.6305	2.6371	2.6456	2.6371

(a) random walk (RW), (b) impulse modified RW, (c) step modified RW.