

INFERENCE FOR AUTOCORRELATIONS IN THE POSSIBLE PRESENCE OF A UNIT ROOT

BY DIMITRIS N. POLITIS, JOSEPH P. ROMANO AND MICHAEL WOLF

University of California, Stanford University and Universitat Pompeu Fabra

First Version received July 2001

Abstract. We consider the problem of making inference for the autocorrelations of a time series in the possible presence of a unit root. Even when the underlying series is assumed to be strictly stationary, the robustness against a unit root is a desirable property to ensure good finite-sample coverage in case the series has a near unit root. In addition to discussing a confidence interval for the autocorrelation at a given lag, we also consider a simultaneous confidence band for the first k autocorrelations. We suggest the use of the subsampling method applied to properly studentized statistics, which results in confidence intervals and bands with asymptotically correct coverage probability. An application to practical model selection is given, while a simulation study examines finite-sample performance.

Keywords. Autocorrelations; confidence band; confidence interval; integrated series; subsampling; unit root.

1. INTRODUCTION

Consider time series data X_1, X_2, \dots, X_T , where $\{\dots, X_{-1}, X_0, X_1, \dots\}$ is a sequence of random variable with mean zero. It is assumed that either $\{X_t\}$ or $\{\Delta X_t = X_t - X_{t-1}\}$ is strictly stationary, so that we are either in the stationary or in the unit root case. Consider the j th order autocorrelation defined as

$$\rho(j) = \lim_{t \rightarrow \infty} \frac{EX_t X_{t+j}}{EX_t^2}.$$

Of course, in the stationary case the limit is superfluous. In the unit root case, however, the limit is generally needed and then equal to 1. Our objective is two-fold. First, we would like to construct a robust confidence interval for $\rho(j)$ for some fixed integer $j \geq 1$. Second, we are interested in a robust simultaneous confidence band, or more generally a joint confidence region, for the vector $\rho = (\rho(1), \dots, \rho(k))'$.

Here, robustness means that we want confidence sets with asymptotically correct coverage probability whether we are in the stationary case or not. Note that the robustness against a unit root is a desirable property even if the underlying series is supposed to be stationary, say on economic grounds (such as dividend yields, for instance). The reason is that if a method does not work

asymptotically when the series has a unit root, it is likely to exhibit poor finite-sample properties when the series has a near unit root. To illustrate this point, consider the problem of making inference for the autoregressive parameter in the simple AR(1) model:

$$X_t = \rho(1)X_{t-1} + \epsilon_t.$$

Obviously, in this model, the first-order autocorrelation is equal to the autoregressive parameter $\rho(1)$. When $|\rho(1)| < 1$, the time series is strictly stationary and a normal approximation can be used to construct a confidence interval for $\rho(1)$ with asymptotically correct coverage probability. However, this interval exhibits poor coverage in finite samples when $\rho(1)$ is very close to 1 and the samples size is not very large. For example, when $T = 200$, $\rho(1) = 0.99$, and the innovations are i.i.d. standard normal, the estimated coverage of this interval with normal level 0.95 is only 0.82; see Politis *et al.* (1999, Table 12.1). The reason is that, when $\rho(1) = 1$, the series has a unit root and the normal approximation is no longer valid.

The problem of making inference for autocorrelations is nontrivial. Even if the series is stationary and far from integrated, Romano and Thombs (1996) show that asymptotically valid inference can, outside of simple linear models, basically only be achieved by resampling methods. Indeed, the limiting distribution of the estimator of $\rho(j)$, the j th-order sample autocorrelation, is normal with mean zero but the corresponding variance appears intractable. If in addition one would like to allow for the possibility of the series being integrated, matters complicate much further. The limiting distribution of the j th-order autocorrelation under a unit root is nonstandard, and the rate of convergence is T as opposed to \sqrt{T} ; consequently, the block bootstrap fails in this case. Although the rate problem can be avoided by properly studentizing the estimator, the problem of different types of limiting distributions remains. The approach we suggest is to construct confidence bands based on the subsampling method.

The remainder of the paper is organized as follows. Section 2 details how to construct a subsampling confidence interval for a specific autocorrelation $\rho(j)$. Section 3 generalizes to the construction of a confidence band for the first k autocorrelations. Section 4 discusses the choice of the block size, which is an important model parameter. Section 5 examines finite sample performance via a simulation study. Finally, Section 6 concludes.

2. THE SUBSAMPLING CONFIDENCE INTERVAL

This section details the confidence interval for the j th-order autocorrelation, for some fixed integer $j \geq 1$, that we propose. To start out, we present the distributional assumptions on the underlying time series. For this reason, we define the new series $\{U_t\}$ to be identical to the series $\{X_t\}$ if $\{X_t\}$ is strictly stationary; otherwise, that is, if $\{X_t\}$ is not stationary but $\{\Delta X_t\}$ is stationary, we define $U_t = \Delta X_t$.

ASSUMPTION 1. Assume that the stationary series $\{U_t\}$ is strong mixing with mixing coefficients $\alpha_U(\cdot)$, and that it satisfies the following conditions: $E(U_t) = 0$, $E|U_t|^\beta < \infty$ for some $\beta > 2$, $f_U(0) > 0$, where $f_U(\cdot)$ denotes the spectral density of $\{U_t\}$, and $\sum_{k=1}^\infty \alpha_U(k)^{1-2/\beta} < \infty$.

The estimator of $\rho(j)$ we suggest is the one commonly used in the unit root literature, based on an OLS regression of X_{t+j} on X_t . It has the form

$$\hat{\rho}_T(j) = \frac{\sum_{t=1}^{T-j} X_t X_{t+j}}{\sum_{t=1}^{T-j} X_t^2}.$$

Its OLS standard error is given by

$$\hat{\rho}_T(j) = \sqrt{\frac{s_T^2(j)}{\sum_{t=1}^{T-j} X_t^2}} \quad \text{where} \quad s_T^2(j) = \frac{1}{T-j-1} \sum_{t=1}^{T-j} (X_{t+j} - \hat{\rho}_T(j)X_t)^2.$$

This leads to a studentized statistic defined as

$$\hat{\xi}_T(j) = \frac{\hat{\rho}_T(j) - \rho(j)}{\hat{\sigma}_T(j)}.$$

If the sampling distribution of $\hat{\xi}_T(j)$, or alternatively the one of $|\hat{\xi}_T(j)|$ were known, confidence intervals for $\rho(j)$ could be constructed based on the appropriate quantile(s) of that distribution. Knowing the distribution of $\hat{\xi}_T(j)$ would allow for the construction of one-sided and equal-tailed two-sided intervals. Knowing the distribution of $|\hat{\xi}_T(j)|$ would allow for the construction of symmetric two-sided intervals. Symmetric confidence intervals often show improved coverage probability in comparison with equal-tailed ones when the asymptotic distribution is normal; for example, see Hall (1988) and Politis *et al.* (1999, Ch. 10). Nevertheless, equal-tailed intervals give some information on the skewness of the underlying distribution which may also be interesting.

To this end, denote the underlying probability mechanism by P and let

$$J_T(x, P) = \text{Prob}_P\{\hat{\xi}_T(j) \leq x\},$$

and

$$J_{T,|\cdot|}(x, P) = \text{Prob}_P\{|\hat{\xi}_T(j)| \leq x\}.$$

Based on the block of data $\{X_t, \dots, X_{t+b-1}\}$ of size b , compute the j th-order autocorrelation and the corresponding standard error denoted by $\hat{\rho}_{b,t}(j)$ and $\hat{\sigma}_{b,t}(j)$ respectively. Next, define the quasi-studentized statistic

$$\hat{\xi}_{b,t}(j) = \frac{\hat{\rho}_{b,t}(j) - \hat{\rho}_T(j)}{\hat{\sigma}_{b,t}(j)}.$$

We use the term ‘quasi-studentized’ to indicate that in the numerator we do not subtract the true value of $\rho(j)$ but the corresponding estimate based on the entire sample.

Then, the subsampling approximation to $J_T(x, P)$ is given by

$$L_{T,b}(x) = \frac{1}{T - b + 1} \sum_{t=1}^{T-b+1} 1\{\hat{\xi}_{b,t}(j) \leq x\},$$

while the subsampling approximation to $J_{T,|\cdot|}(x, P)$ is given by

$$L_{T,b,|\cdot|}(x) = \frac{1}{T - b + 1} \sum_{t=1}^{T-b+1} 1\{|\hat{\xi}_{b,t}(j)| \leq x\},$$

where $1\{\cdot\}$ is the indicator function. The following theorem shows that constructing symmetric confidence intervals for $\rho(j)$ based on the quantiles of $L_{T,b}$ and/or $L_{T,b,|\cdot|}$ is asymptotically justified.

THEOREM 1. *Assume Assumption 1 and that $b/T \rightarrow 0$ and $b \rightarrow \infty$ as $T \rightarrow \infty$. Let $c(1 - \alpha)$ be a $1 - \alpha$ quantile of the distribution $L_{T,b}$, that is, let $c(1 - \alpha) = \inf\{x: L_{T,b}(x) \geq 1 - \alpha\}$. Then, the equal-tailed confidence interval*

$$CI_{ET} = [\hat{\rho}_T(j) - \hat{\sigma}_T(j)c(1 - \alpha/2), \hat{\rho}_T(j) + \hat{\sigma}_T(j)c(\alpha/2)]$$

has asymptotic coverage probability of $1 - \alpha$.

Let $c_{|\cdot|}(1 - \alpha) = \inf\{x: L_{T,b,|\cdot|}(x) \geq 1 - \alpha\}$. Then, the symmetric confidence interval

$$CI_{SYM} = [\hat{\rho}_T(j) - \hat{\sigma}_T(j)c_{|\cdot|}(1 - \alpha), \hat{\rho}_T(j) + \hat{\sigma}_T(j)c_{|\cdot|}(1 - \alpha)]$$

has asymptotic coverage probability of $1 - \alpha$.

PROOF. We focus on the case of approximating $J_{T,|\cdot|}(x, P)$ by $L_{T,b,|\cdot|}(x)$; the case of approximating $J_T(x, P)$ by $L_{T,b}(x)$ is similar.

Arguing along the lines of Corollary 12.2.1 of Politis *et al.* (1999), it is sufficient to verify the following three conditions, whether we are in the stationary or in the unit root case: for some sequence $\tau_T = T^\gamma$ with $\gamma > 0$,

- (i) $J_{T,|\cdot|}(P)$ converges weakly to a nondegenerate continuous limiting distribution.
- (ii) $\tau_T(\hat{\rho}_T(j) - \rho(j))$ converges weakly to a random variable having a nondegenerate distribution.
- (iii) $\tau_T\hat{\sigma}_T(j)$ converges weakly to a random variable having a (possibly degenerate) distribution which puts no mass at zero.

Stationary case. Let $\tau_T = T^{1/2}$. It follows from Romano and Thombs (1996) that $\tau_T(\hat{\rho}_T(j) - \rho(j))$ converges weakly to Y , where Y has a normal distribution with

mean zero and some positive variance $\kappa^2(j)$. Next, $\tau_T \hat{\sigma}_T(j)$ converges in probability, and hence weakly, to the constant

$$v(j) = \sqrt{\frac{E(X_{1+j} - \rho(j)X_1)^2}{EX_1^2}}$$

Finally, by Slutsky's theorem, $J_{T,|\cdot|}(P)$ converges weakly to the distribution of $|Z|$ where Z follows a normal distribution with mean zero and variance $\kappa^2(j)/v^2(j)$. *Unit root case.* Let $\tau_T = T$. Conditions (ii)–(iii) then follow from Phillips (1986, 1987). Moreover, it follows from the same work that $\hat{\xi}_T(j)$ converges weakly to random variable V having a nondegenerate distribution. Hence, condition (i) is implied by the continuous mapping theorem where the limiting distribution is the one of $|V|$. Note that all limiting distributions are nonstandard but well-defined.

REMARK 1. The limiting variance $\kappa^2(j)/v^2(j)$ in the stationary case is in general not equal to 1 and therefore $\hat{\xi}_T$ is not a proper studentized statistic in the strictest of senses. However, the validity of the subsampling method does not depend on $\kappa^2(j)/v^2(j)$ being equal to 1 and we shall use the term ‘studentized’ in an appropriately loose sense; note that it might be more appropriate, although less familiar, to use the term ‘self-normalized statistic’ instead.

REMARK 2. As mentioned in the Introduction, Romano and Thombs (1996) have shown that, outside of semi-parametric models such as ARMA models, the only viable option is to resort to resampling/subsampling methods in order to construct valid confidence intervals for the autocorrelations of stationary time series. For example, in the stationary case, confidence intervals could be constructed based on the moving blocks bootstrap (Künsch, 1989) or on the stationary bootstrap (Politis and Romano, 1994). However, the discontinuity in the type of the limiting distribution when the underlying series moves from strictly stationary to having a unit root causes these bootstrap intervals to fail. Hence, it appears that the subsampling-based approach we propose is the only one yielding asymptotically consistent confidence intervals for $\rho(j)$ in full generality, whether the series is stationary or unit-root integrated.

3. THE SUBSAMPLING CONFIDENCE BAND

This section details the confidence band for the first k autocorrelations (for some fixed $k \geq 2$) that we propose. Using the same notation as in the previous section, define $\hat{\rho}_T = (\hat{\rho}_T(1), \dots, \hat{\rho}_T(k))'$ and $\hat{\xi}_T = (\hat{\xi}_T(1), \dots, \hat{\xi}_T(k))'$. If the sampling distribution of $\|\hat{\xi}_T\|$ were known, a joint confidence region for $\rho = (\rho(1), \dots, \rho(k))'$ could be computed; here, $\|\cdot\|$ is any norm on \mathbb{R}^k . When $\|\cdot\|$ is chosen to be $\|\cdot\|_\infty$, that is, the sup norm, the resulting confidence region will actually be a simultaneous confidence band for the sequence $(\rho(1), \dots, \rho(k))$.

Let

$$J_{T,\|\cdot\|}(x, P) = \text{Prob}_P\{\|\hat{\xi}_T\| \leq x\}.$$

Based on the block of data $\{X_t, \dots, X_{t+b-1}\}$ of size b , compute $\hat{\xi}_{b,t}(j)$ for $j = 1, \dots, k$ and let $\hat{\xi}_{b,t} = (\hat{\xi}_{b,t}(1), \dots, \hat{\xi}_{b,t}(k))'$. Then, the subsampling approximation to $J_{T,\|\cdot\|}(x, P)$ is given by

$$L_{T,b,\|\cdot\|}(x) = \frac{1}{T-b+1} \sum_{t=1}^{T-b+1} 1\{\|\hat{\xi}_{b,t}\| \leq x\},$$

where $1\{\cdot\}$ is the indicator function. The following theorem shows that constructing a confidence region for ρ based on the quantiles of $L_{T,b,\|\cdot\|}$ is asymptotically justified. For ease of notation, define a (data-dependent) distance between $\hat{\rho}_T$ and an arbitrary autocorrelation vector $\tilde{\rho}$ as

$$d_{T,\|\cdot\|}(\hat{\rho}_T, \tilde{\rho}) = \|([\hat{\rho}_T(1) - \tilde{\rho}(1)]/\hat{\sigma}_T(1), \dots, [\hat{\rho}_T(k) - \tilde{\rho}(k)]/\hat{\sigma}_T(k))'\|.$$

THEOREM 2. *Assume Assumption 1 and that $b/T \rightarrow 0$ and $b \rightarrow \infty$ as $T \rightarrow \infty$. Let $c_{\|\cdot\|}(1 - \alpha) = \inf\{x: L_{T,b,\|\cdot\|}(x) \geq 1 - \alpha\}$. Then, the confidence region*

$$\text{CR} = \{\tilde{\rho} : d_{T,\|\cdot\|}(\hat{\rho}_T, \tilde{\rho}) \leq c_{\|\cdot\|}(1 - \alpha)\} \tag{1}$$

has asymptotic coverage probability of $1 - \alpha$.

PROOF. Let $J_T(P)$ denote the k -dimensional sampling distribution of $\hat{\xi}_T$ with distribution function

$$J_T(y, P) = \text{Prob}_P\{\hat{\xi}_T \leq y\}, \quad \text{for } y \in \mathbb{R}^k.$$

The corresponding subsampling approximation is then given by

$$L_{T,b}(y) = \frac{1}{T-b+1} \sum_{t=1}^{T-b+1} 1\{\hat{\xi}_{b,t} \leq y\}.$$

We now claim that in order to prove the theorem it is enough to show that

- (I) $J_T(P)$ converges weakly to a nondegenerate continuous limiting law $J(P)$.
- (II) $L_{T,b}(y) \rightarrow J(y, P)$ in probability for every continuity point y of $J(\cdot, P)$.

To see why, note that by the continuous mapping theorem, (I) implies that $J_{T,\|\cdot\|}(P)$ converges weakly to $J_{\|\cdot\|}(P)$, where $J_{\|\cdot\|}(P)$ is the distribution of $\|V\|$ when V is a random vector having distribution $J(P)$. Moreover, by the continuous mapping theorem again, (II) implies that $L_{T,b,\|\cdot\|}(x) \rightarrow J_{\|\cdot\|}(x, P)$ in probability for every continuity point x of $J_{\|\cdot\|}(\cdot, P)$, including $c_{T,\|\cdot\|}(1 - \alpha)$. The latter, by a standard argument, is sufficient for the asymptotic consistency of the confidence region (1).

A straightforward multivariate extension of Theorem 12.2.2 of Politis *et al.* (1999) shows that conditions (I)–(II) can be verified by demonstrating that, for some sequence $\tau_T = T^\gamma$ with $\gamma > 0$,

- (i) $J_T(P)$ converges weakly to a nondegenerate continuous limiting distribution.
- (ii) $\tau_T(\hat{\rho}_T(j) - \rho(j))$ converges weakly to a random variable having a nondegenerate distribution, for $i \leq j \leq k$.
- (iii) $\tau_T\hat{\sigma}_T(j)$ converges weakly to a random variable having a (possibly degenerate) distribution which puts no mass at zero, for $1 \leq j \leq k$.

Stationary case. Let $\tau_T = T^{1/2}$. It follows from Romano and Thombs (1996) that $\tau_T(\hat{\rho}_T - \rho)$ converges weakly to Y , where Y has a normal distribution with mean zero and some positive-definite covariance matrix Σ . As before, for arbitrary j , $\tau_T\hat{\sigma}_T(j)$ converges in probability, and hence weakly, to the constant $v(j)$. Let Ψ be the $k \times k$ diagonal matrix with j th element $1/v(j)$. Then, by Slutsky’s theorem, $\hat{\xi}_T$ converges weakly to the normal distribution with mean zero and covariance matrix $\Psi\Sigma\Psi$, which verifies (i). Conditions (ii)–(iii) are proved in the same way as in the proof of Theorem 1.

Unit root case. Let $\tau_T = T$. As stated in proof of Theorem 1, the weak convergence of each entry $\hat{\xi}_T(j)$ to a nondegenerate limiting distribution follows by Phillips (1986, 1987); see also Hall (1989). Using the standard functional limit theorem machinery, it can be easily seen that therefore the vector $\hat{\xi}_T$ also has nondegenerate limiting distribution. Conditions (ii)–(iii) are proved in the same way as in the proof of Theorem 1.

REMARK 3. The theorem allows for the construction of a confidence region for ρ based on an arbitrary norm $\|\cdot\|$. Note that in general it may difficult to write down the region in closed form. However, for the special choice of the sup norm $\|\cdot\|_\infty$, the confidence region yields a confidence band. First, recall that for a vector $a \in \mathbb{R}^k$,

$$\|a\|_\infty = \sup_{1 \leq j \leq k} |a_j|.$$

Then, a simple calculation shows that for this choice of norm, the confidence region (1) simplifies to the band

$$[\hat{\rho}_T(j) \pm \hat{\sigma}_T(j) * c_{\|\cdot\|_\infty}(1 - \alpha)], \quad j = 1, \dots, k,$$

which can be easily visualized and interpreted.

REMARK 4. We have detailed the construction of the confidence band for the parameter $\rho = (\rho(1), \dots, \rho(k))'$, for a given integer k . An analogous construction would obviously also apply when the parameter of interest is $\rho = (\rho(q + 1), \dots, \rho(q + k))'$, for two given integers q and k .

We now give an application of Theorem 2 and Remark 3 to the subject of model selection, in particular choosing the order q when fitting the familiar MA(q) model

$$X_t = \sum_{i=0}^q \theta_i Z_{t-i}$$

to the data; here θ_0 is taken to be 1, and $\{Z_t\}$ is an uncorrelated (but not necessarily independent) time series with mean zero. The usual approach is to build this model up by increasing the order q one step at a time until a ‘satisfactory’ model is obtained. Note that an MA(q) model for $\{X_t\}$ is essentially equivalent to the autocorrelations of $\{X_t\}$ vanishing for lags greater than q ; for example, see Priestley (1981) or Brockwell and Davis (1991). Therefore, a ‘satisfactory’ model may be defined as the one with smallest order (say \hat{q}) for which $\hat{\rho}(j)$ is not significantly different from zero for $j > \hat{q}$.

The usual way of accomplishing this objective is the use of confidence bands around zero that are based on Bartlett’s formula; for example, the $\pm 1.96/\sqrt{T}$ confidence bands that are typically included with a correlogram for comparison purposes are supposed to help the practitioner decide whether an independent model for $\{X_t\}$ is satisfactory, that is, an MA(0) model with i.i.d. innovations $\{Z_t\}$.

Nevertheless, if the $\{Z_t\}$ are uncorrelated but not independent, the confidence bands based on Bartlett’s formula are *not* valid for checking the validity of an MA(q) model with $q = 0$ or higher. Thus, we resort to the proposed subsampling confidence bands. The procedure is as follows:

- Fix an order of interest, say q , and an integer $k \geq 2$.
- Consider the statistic $\hat{\rho}_T = (\hat{\rho}_T(q+1), \dots, \hat{\rho}_T(q+k))'$ as an estimator of the parameter $\rho = (\rho(q+1), \dots, \rho(q+k))'$.
- Construct the approximate $1 - \alpha$ confidence band around zero of the type $\pm \hat{\sigma}(j)c_{\|\cdot\|_\infty}(1 - \alpha)$, valid for $j = q+1, q+2, \dots, q+k$.
- If at least one of the statistics $\hat{\rho}_T(q+1), \dots, \hat{\rho}_T(q+k)$ is outside the confidence band, reject the hypothesis that MA(q) is a satisfactory model at level α .

4. CHOICE OF THE BLOCK SIZE

The practical problem in constructing the subsampling confidence sets is the choice of the block size b . Note that the asymptotic requirements $b/T \rightarrow 0$ and $b \rightarrow \infty$ as $T \rightarrow \infty$ give little guidance. We therefore propose the following calibration method. The method will be illustrated for the construction of a confidence interval for the j th autocorrelation $\rho(j)$. An analogous idea can be used to construct a confidence band for the first k autocorrelations.

The goal is to compute a $1 - \alpha$ confidence interval for $\rho(j)$ for some fixed j . In finite samples, a subsampling interval will typically not exhibit coverage probability exactly equal to $1 - \alpha$; moreover, the actual coverage probability generally depends

on the block size b . Indeed, one can think of the actual coverage level $1 - \lambda$ of a subsampling confidence interval as a function of the block size b , conditional on the underlying probability mechanism P , the nominal confidence level $1 - \alpha$, and the sample size T . The idea is now to adjust the ‘input’ b in order to obtain the actual coverage level close to the nominal one. Hence, one can consider the block size calibration function $g : b \rightarrow 1 - \lambda$. If $g(\cdot)$ were known, one could construct an ‘optimal’ confidence interval by finding \tilde{b} that minimizes $|g(b) - (1 - \alpha)|$ and use \tilde{b} as the block size; note that $|g(b) - (1 - \alpha)| = 0$ may not always have a solution.

Of course, the function $g(\cdot)$ depends on the underlying probability mechanism P and is therefore unknown. We now propose a semi-parametric bootstrap method to estimate it. The idea is that in principle we could simulate $g(\cdot)$ if P were known by generating data of size T according to P and computing subsampling confidence intervals for θ for a number of different block sizes b . This process is then repeated many times and for a given b one estimates $g(b)$ as the fraction of the corresponding intervals that contain the true parameter. The method we propose is identical except that P is replaced by an estimate \hat{P}_T .

We suggest to use an AR(p) model

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t$$

combined with block-resampling the residuals as the estimate \hat{P}_T , giving rise to a sequence X_1^*, \dots, X_T^* in the following manner

ALGORITHM 1. (Sampling from the estimated model)

1. Estimate the AR(p) model, yielding fitted parameters $\hat{\phi}_1, \dots, \hat{\phi}_p$ and residuals $\hat{\epsilon}_{p+1}, \dots, \hat{\epsilon}_T$.
2. Block-bootstrap the residuals to get $\epsilon_{p+1}^*, \dots, \epsilon_T^*$
3. $X_t^* = X_t$ for $t = 1, \dots, p$.
4. $X_t^* = \hat{\phi}_1 X_{t-1}^* + \dots + \hat{\phi}_p X_{t-p}^* + \epsilon_t^*$ for $t = p + 1, \dots, T$

REMARK 5. Alternatively, an ARMA(p, q) model could be used. (In fact, any low-order model that captures most of the linear dependence structure will do in practice.) If desired, one can determine p (and/or q) by any of the popular model selection criteria such as AIC or BIC.

REMARK 6. We consider it important to block-bootstrap the residuals $\hat{\epsilon}_t$ rather than using Efron’s bootstrap (that is, treating the residuals as approximately i.i.d.) in order to capture and ‘leftover’, nonlinear dependence. Take the simple case of an uncorrelated but dependent sequence such as $X_t = Z_t Z_{t-1}$ and Z_t i.i.d. standard normal. In that case an AR(p) or ARMA(p, q) model will not capture any of the dependence but block-bootstrapping the residuals will. Analogous reasoning applies to linear models with uncorrelated but dependent innovations.

Having specified how to generate data from the estimated model \hat{P}_T , we next detail the algorithm to determine the block size \tilde{b} to be used in practice.

ALGORITHM 2. (Choice of the block size)

1. Fix a selection of reasonable block sizes b between limits b_{low} and b_{up} .
2. Generate L pseudo sequences $X_{l,1}^*, \dots, X_{l,T}^*$, $l = 1, \dots, L$, according to Algorithm 1. For each sequence, $l = 1, \dots, L$, and for each b , compute a subsampling confidence interval $\text{CI}_{l,b}$ for $\rho(j)$.
3. Compute $\hat{g}(b) = \#\{\hat{\rho}_T(j) \in \text{CI}_{l,b}\}/L$.
4. Find the value \tilde{b} that minimizes $|\hat{g}(b) - (1 - \alpha)|$.

REMARK 7. Algorithm 1 is by an order of magnitude more expensive than the computation of the final subsampling interval once the block size has been determined. While it is advisable to choose the selection of candidate block sizes in Step 2 as fine as possible (ideally, include every integer between b_{low} and b_{up}), this may computationally not be feasible, especially in simulation studies. In those instances, a coarse grid should be employed.

REMARK 8. An analogous method can be used for the construction of a confidence band for the first k autocorrelations. The modifications are obvious and left to the reader.

5. SIMULATION STUDY

The goal of this section is to highlight the small sample performance of the proposed subsampling method. We will focus on a symmetric confidence interval for the first autocorrelation $\rho(1)$. The data generating process (DGP) we consider is the ARMA(1, 1) model given by

$$X_t = \phi X_{t-1} + \epsilon_t + \theta \epsilon_{t-1},$$

where $\{\epsilon_t\}$ is a sequence of strictly stationary, uncorrelated variables with mean zero. The parameter values included in the study are $\phi = 1, 0.95$ and 0.8 , and $\theta = 0.8, 0$ and -0.8 respectively. The innovations are either $\epsilon_t = Z_t$ or $\epsilon_t = Z_t Z_{t-1}$, where the Z_t are i.i.d. standard normal. Note that the latter specification for ϵ_t results in an uncorrelated but dependent sequence. The sample size is $n = 128$.

For the data-dependent choice of the block size, we fit an AR(p) model to each generated data set, where the order p is determined by the BIC criterion. Also, we use the stationary bootstrap of Politis and Romano (1994) with average block size $b_{\text{Boot}} = 10$ on the residuals.

We estimate coverage probabilities of nominal 90% and 95% confidence intervals based on 1000 repetitions per scenario. The results are presented in Tables I and II. The estimated coverage probabilities are given for several fixed block sizes and also for the data-dependent block size \tilde{b} determined by Algorithm 2; note that the fixed block sizes listed constitute the input block sizes for Algorithm 2.

One can see the following. The best fixed block size generally depends very much on the parameters ϕ and θ , so that the choice of a good block size in

practice is both important and nontrivial. However, our algorithm to determine the data-dependent block size in practice seems to work satisfactorily. The estimated coverage levels are in general close to the nominal level except when $\theta = -0.8$, which is a well-known problematic case in the literature. For example, when $\phi = 1$ and $\theta = -0.8$, it comes to a near ‘cancellation’ of the unit root, so that the observed time series might appear white noise when indeed it is integrated. Apparently, much larger sample sizes are needed in such scenarios in order for the inference to become reliable. Note, however, that is also true for other methods, such as unit-root tests.

6. CONCLUSIONS

In this paper, we have investigated the problem of constructing asymptotically consistent confidence intervals and confidence bands for autocorrelations of a time series in the possible presence of a unit root. It is desirable to have confidence

TABLE I
ESTIMATED COVERAGE PROBABILITIES OF SYMMETRIC SUBSAMPLING CONFIDENCE INTERVALS FOR $\rho(1)$
WITH NOMINAL LEVELS 0.90 AND 0.95

| Target | $b = 5$ | $b = 15$ | $b = 25$ | $b = 35$ | \hat{b} |
|------------------------------|---------|----------|----------|----------|-----------|
| $\phi = 1, \theta = 0.8$ | | | | | |
| 0.90 | 0.99 | 0.93 | 0.88 | 0.85 | 0.89 |
| 0.95 | 1.00 | 0.97 | 0.94 | 0.89 | 0.95 |
| $\phi = 1, \theta = 0$ | | | | | |
| 0.90 | 0.95 | 0.85 | 0.81 | 0.78 | 0.88 |
| 0.95 | 0.99 | 0.92 | 0.87 | 0.83 | 0.94 |
| $\phi = 1, \theta = -0.8$ | | | | | |
| 0.90 | 0.04 | 0.03 | 0.05 | 0.05 | 0.07 |
| 0.95 | 0.15 | 0.06 | 0.07 | 0.07 | 0.15 |
| $\phi = 0.95, \theta = 0.8$ | | | | | |
| 0.90 | 1.00 | 0.96 | 0.92 | 0.88 | 0.91 |
| 0.95 | 1.00 | 0.98 | 0.95 | 0.93 | 0.95 |
| $\phi = 0.95, \theta = 0$ | | | | | |
| 0.90 | 0.96 | 0.88 | 0.82 | 0.78 | 0.91 |
| 0.95 | 0.99 | 0.93 | 0.89 | 0.86 | 0.95 |
| $\phi = 0.95, \theta = -0.8$ | | | | | |
| 0.90 | 0.76 | 0.74 | 0.72 | 0.70 | 0.75 |
| 0.95 | 0.87 | 0.81 | 0.78 | 0.75 | 0.85 |
| $\phi = 0.8, \theta = 0.8$ | | | | | |
| 0.90 | 0.99 | 0.92 | 0.87 | 0.82 | 0.89 |
| 0.95 | 1.00 | 0.97 | 0.92 | 0.89 | 0.94 |
| $\phi = 0.8, \theta = 0$ | | | | | |
| 0.90 | 0.94 | 0.87 | 0.82 | 0.77 | 0.89 |
| 0.95 | 0.98 | 0.92 | 0.87 | 0.83 | 0.94 |
| $\phi = 0.8, \theta = -0.8$ | | | | | |
| 0.90 | 0.91 | 0.84 | 0.80 | 0.75 | 0.86 |
| 0.95 | 0.93 | 0.91 | 0.87 | 0.82 | 0.93 |

Notes: The sample is $n = 128$ and the residuals are $\epsilon_t = Z_t$, where the Z_t are i.i.d. standard normal.

TABLE II
ESTIMATED COVERAGE PROBABILITIES OF SYMMETRIC SUBSAMPLING CONFIDENCE INTERVALS FOR $\rho(1)$
WITH NOMINAL LEVELS 0.90 AND 0.95

| Target | $b = 5$ | $b = 15$ | $b = 25$ | $b = 35$ | \tilde{b} |
|------------------------------|---------|----------|----------|----------|-------------|
| $\phi = 1, \theta = 0.8$ | | | | | |
| 0.90 | 1.00 | 0.96 | 0.93 | 0.89 | 0.92 |
| 0.95 | 1.00 | 0.98 | 0.96 | 0.94 | 0.96 |
| $\phi = 1, \theta = 0$ | | | | | |
| 0.90 | 1.00 | 0.95 | 0.89 | 0.85 | 0.92 |
| 0.95 | 1.00 | 0.98 | 0.94 | 0.91 | 0.96 |
| $\phi = 1, \theta = -0.8$ | | | | | |
| 0.90 | 0.84 | 0.44 | 0.30 | 0.23 | 0.78 |
| 0.95 | 0.95 | 0.64 | 0.45 | 0.36 | 0.90 |
| $\phi = 0.95, \theta = 0.8$ | | | | | |
| 0.90 | 1.00 | 0.96 | 0.93 | 0.89 | 0.91 |
| 0.95 | 1.00 | 0.99 | 0.97 | 0.94 | 0.96 |
| $\phi = 0.95, \theta = 0$ | | | | | |
| 0.90 | 1.00 | 0.94 | 0.89 | 0.84 | 0.92 |
| 0.95 | 0.99 | 0.93 | 0.89 | 0.86 | 0.95 |
| $\phi = 0.95, \theta = -0.8$ | | | | | |
| 0.90 | 0.89 | 0.75 | 0.72 | 0.67 | 0.80 |
| 0.95 | 0.96 | 0.83 | 0.79 | 0.74 | 0.90 |
| $\phi = 0.8, \theta = 0.8$ | | | | | |
| 0.90 | 0.99 | 0.92 | 0.88 | 0.84 | 0.88 |
| 0.95 | 1.00 | 0.97 | 0.93 | 0.90 | 0.94 |
| $\phi = 0.8, \theta = 0$ | | | | | |
| 0.90 | 0.99 | 0.83 | 0.80 | 0.75 | 0.87 |
| 0.95 | 1.00 | 0.93 | 0.88 | 0.83 | 0.93 |
| $\phi = 0.8, \theta = -0.8$ | | | | | |
| 0.90 | 0.91 | 0.75 | 0.74 | 0.71 | 0.81 |
| 0.95 | 0.99 | 0.85 | 0.80 | 0.78 | 0.90 |

Notes: The sample is $n = 128$ and the residuals are $\epsilon_t = Z_t Z_{t-1}$, where the Z_t are i.i.d. standard normal.

sets that are robust against a unit root even in the case when the time series is assumed to be stationary. The reason is that when the series has a near unit root a robust method is expected to yield better finite sample performance compared with a method which breaks down in the presence of unit root; for example, see Politis *et al.* (1990, Ch. 12) for some evidence in the context of simple AR(1) models.

We suggested the use of the subsampling technique in order to construct asymptotically valid confidence sets. By subsampling appropriate studentized statistics, a confidence interval or a confidence band can be computed that have the right coverage probability in the limit, whether the underlying time series is stationary or not. Having a robust confidence band has an immediate application to the practical issue of model selection.

The main problem in applying the method is the choice of the block size in practice. To this end, we proposed a fully automatic, data-dependent algorithm. Some simulation studies showed good performance of the method with a moderate sample size.

ACKNOWLEDGEMENT

Research of the author Michael Wolf supported by DGES grant BEC2001-1270 and by the Barcelona Economics Program of CREA.

NOTE

Corresponding author: Dimitris N. Politis, Department of Mathematics, University of California, San Diego, La Jolla, CA 92093, USA, Tel: 858 534-5861, Fax: 858 534-5273, E-mail: politis@euclid.ucsd.edu

REFERENCES

- BROCKWELL, P. and DAVIS, R. A. (1991) *Time Series: Theory and Methods*, 2nd edn. New York: Springer.
- HALL, P. (1988) On symmetric bootstrap confidence intervals. *Journal of the Royal Statistical Society, Series B* 50, 35–45.
- HALL, A. (1989) Testing for a unit root in the presence of moving average errors. *Biometrika* 76, 49–56.
- KÜNSCH, H. R. (1989) The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–41.
- PHILLIPS, P. C. B. (1986) Understanding spurious regressions in econometrics. *Journal of Econometrics* 33, 311–40.
- PHILLIPS, P. C. B. (1987) Times series regressions with a unit root. *Econometrica* 55, 277–301.
- POLITIS, D. N. and ROMANO, J. P. (1994) The stationary bootstrap. *Journal of the American Statistical Association* 89, 1303–13.
- POLITIS, D. N. ROMANO, J. P. and WOLF, M. (1999) *Subsampling*. New York: Springer.
- PRIESTLEY, M. B. (1981) *Spectral Analysis and Time Series*. New York: Academic Press.
- ROMANO, J. P. and THOMBS, L. A. (1996) Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association* 91, 590–600.