Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: www.tandfonline.com/journals/lecr20

Bootstrapping time series models

G. S. Hongyi Li & Maddala

To cite this article: G. S. Hongyi Li & Maddala (1996) Bootstrapping time series models, Econometric Reviews, 15:2, 115-158, DOI: 10.1080/07474939608800344

To link to this article: https://doi.org/10.1080/07474939608800344

Published online: 21 Mar 2007.

ECONOMETRIC REVIEWS, 15(2), 115-158 (1996)

Bootstrapping Time Series Models

Hongyi Li and G.S. Maddala[1]

Department of Economics
The Ohio State University
Columbus, OH 43210

Abstract

This paper surveys recent development in bootstrap methods and the modifications needed for their applicability in time series models. The paper discusses some guidelines for empirical researchers in econometric analysis of time series. Different sampling schemes for bootstrap data generation and different forms of bootstrap test statistics are discussed. The paper also discusses the applicability of direct bootstrapping of data in dynamic models and cointegrating regression models. It is argued that bootstrapping residuals is the preferable approach. The bootstrap procedures covered include the recursive bootstrap, the moving block bootstrap and the stationary bootstrap.

1 Introduction

The purpose of this paper is to provide a survey of bootstrap procedures applied to time series econometric models and to present some guidelines for

[1]This is a revised version of a paper under the same title presented at the Econometric Society Meetings in Uppsala, August 1993. Helpful comments of anonymous referees are greatfully acknowledged. The constructive comments of one of the referees has led to a rewriting of the paper. The usual disclaimer applies.

empirical researchers in this area. The bootstrap approach has been gaining popularity among theoretical and empirical econometricians. However, there have also been several misapplications of the bootstrap method in econometrics. The purpose of the present paper is to outline some general guidelines and then discuss several problems in bootstrapping time series models in this context.

Most of the inferential procedures available in the analysis of time series data are asymptotic. Although analytic small sample results are available in a few cases, there is currently, no widely applicable and easily accessible method that can be used to make small sample inferences. Methods like Edgeworth expansions involve a lot of algebra and are also applicable in very special cases.

The bootstrap technique introduced by Efron (1979) could possibly be a potential alternative in estimation and inference from time series models in finite samples. However, in time series regressions, the standard bootstrap resampling method designed for independent and identically distributed (*IID*) errors is not applicable because in most situations the assumption of *IID* errors is violated. Correlated errors are not exchangeable, and lagged dependent variables create extra problems in pseudo data generation. Unit root and cointegration regression models create further complications in bootstrap data generation. Finally, to achieve an improvement over the asymptotic results, one needs to work with asymptotically pivotal statistics. This is not usually done.

Jeong and Maddala (1993) provide a general survey of bootstrap methods in econometrics. The present paper will provide a survey of bootstrap procedures with special reference to time series regressions. It is thus a more detailed discussion of Section 4.2 in Jeong and Maddala (1993). The present paper also provides a more organized critique of the work on the application of bootstrap methods in econometrics.

The plan of the paper is as follows. In Section 2 we discuss some general guidelines for using the bootstrap approach. It will be seen that very often these guidelines have been violated. In Section 3 we discuss bootstrap methods for stationary $ARMA(p,q)$ models and the $AR(1)$ model in some detail. In Section 4 we discuss the problem of general error structures and the application of the moving block bootstrap (*MBB*) procedure as well as some recent modifications. Section 5 discusses forecasting problems. Section 6 discusses the problem of unit root testing using bootstrap methods, although, due to space limitations we shall be confined to the Dickey-Fuller framework. Section 7 discusses issues in the generation of bootstrap data in cointegrated regressions. The final section presents the conclusions.

2 General Guidelines for Using the Bootstrap Approach

We shall present some general guidelines that empirical researchers in econometric time series analysis need to follow. Some of these guidelines have been derived for the usual regression models but empirical evidence suggests that they apply to time series models as well. First we shall present a brief review of the bootstrap approach.

The bootstrap method introduced by Efron (1979) is a resampling method. Several resampling methods were in use earlier but they were disparate and Efron made the resampling method a widely applicable technique. For a history of the resampling approach going back to early papers by Barnard (1963) and Hartigan (1969), see Hall (1992).

Let $(y_1, y_2, \cdots, y_n)$ be a random sample from a distribution characterized by a parameter θ. Inference about θ will be based on a statistic T. The basic bootstrap approach consists of drawing repeated samples (with replacement) of size m (which may or may not be equal to n, although it usually is) from $(y_1, y_2, \cdots, y_n)$. Call this sample $(y_1^*, y_2^*, \cdots, y_n^*)$. This is the bootstrap sample. We do this B times. B is the number of bootstrap replications. For each bootstrap sample we compute the statistic T. Call this T^*. The distribution of T^* is known as the bootstrap distribution of T. We use this bootstrap distribution to make inferences about θ. Under some circumstances (to be described later) the bootstrap distribution enables us to make more accurate inferences than the asymptotic distribution of T. The bootstrap method described here is the simplest one that is valid for *IID* observations. Needless to say that when the *IID* assumption is not satisfied this method needs to be modified. These modifications in the context of time series models are described later. Babu (1995) gives an example of modifications in another non-standard case (errors in variables).

Early work on the application of bootstrap methods merely consisted of using the bootstrap distribution to get standard errors. The standard error of T^* was used as a better estimate of the small sample standard error of T than that given by the asymptotic distribution of T. In some complicated models the derivation of asymptotic standard errors (*SE*'s) is extremely complicated and in this case the bootstrap procedure for computation of the *SE*'s were the only ones available. See Veall (1988) for this sort of argument. However, as remarked by Hartigan (1986, p. 76) earlier in these cases one should use the bootstrap standard error to construct a pivotal statistic and then bootstrap this again for the construction of confidence intervals or testing of hypotheses. In those cases where the asymptotic *SE*'s were available, the bootstrap

SE's were justified as a finite sample approximation. However, the standard error is of interest only if the small sample distribution of T is known to be normal. In most econometric applications this is not the case. Thus, using the bootstrap distribution to get standard errors in this situation is useless. Confidence interval estimation and hypothesis testing can be based directly on the bootstrap distribution. For many early applications on the use of just bootstrap standard errors in econometrics, see p. 580 of Jeong and Maddala (1993). It is not necessary to quote these misapplications here.

Even if the asymptotic and bootstrap standard errors are the same in any given example, confidence interval statements from the bootstrap distribution will be different from the confidence interval based on the asymptotic distribution if the bootstrap distribution is skewed. For the autoregressive model $y_t = \phi y_{t-1} + \varepsilon_t$ based on Wolfer's sunspot numbers for 1770-1889, Efron and Tibshirani (1986) got $\widehat{\phi} = 0.815$ with asymptotic $SE = .053$. The bootstrap SE based on 1,000 replications was .055 agreeing nicely with the asymptotic results. However, the bootstrap distribution was skewed to the left. We shall now list some guidelines for confidence interval estimation, hypothesis testing and methods of generation of bootstrap samples.

2.1 Confidence Intervals

Many of the results on the properties of bootstrap confidence intervals, that we shall quote now, have been derived using Edgeworth expansions. The use of Edgeworth expansions to study the properties of bootstrap methods started with the papers by Singh (1981) and Bickel and Freedman (1980). They considered the distribution of the sample mean with known variance. Later in three papers in 1983, 1984 and 1985, Babu and Singh considered the case of unknown variance and the distribution of the studentized mean. See also Abramovich and Singh (1985). See Hall (1992, p. 151) for a history of these methods. As Hall points out (in Appendix V) there are some limitations to the results obtained from Edgeworth expansions. However, many of the conclusions that have been derived from the Edgeworth expansions have been substantiated in the Monte Carlo studies on models of interest to econometricians. Hence, we shall state the main results.

Suppose $\widehat{\theta}$ is a consistent estimator for θ and $\widehat{\theta}^*$ is the bootstrap estimator of θ. We consider the following methods to construct confidence intervals.

(i) Use the asymptotic distribution of $\widehat{\theta}$. The two-sided confidence interval is $\widehat{\theta} \pm z_\alpha SE(\widehat{\theta})$ where $SE(\widehat{\theta})$ is the asymptotic standard error of $\widehat{\theta}$ and z_α is the $(100-\alpha)$ percentile from the standard normal distribution. This interval is a symmetric interval.

(ii) Use the bootstrap distribution of $\hat{\theta}^*$. The two-sided $(100 - 2\alpha)$ confidence interval for θ is $(\hat{\theta} - z^*_{1-\alpha}, \hat{\theta} + z^*_{\alpha})$ where z^*_{α} is the 100α percentile of the distribution of $\hat{\theta}^* - \hat{\theta}$. This is a two-sided equal-tailed interval. It is often non-symmetric. This is called the percentile method. The nominal coverage of this interval is $(100 - 2\alpha)$ and the difference between the actual coverage and nominal coverage is called the coverage error. To improve on the coverage error, Efron suggested two other percentile intervals (see Efron, 1987):

(iii) Bias-corrected (BC) percentile interval.

(iv) Accelerated bias-corrected (BC_a) percentile interval. In addition we have the following other methods.

(v) The percentile-t (or bootstrap-t) method (see Hall, 1988b, 1992). This is the percentile method based on the bootstrap distribution of the t-statistic $t = \frac{\sqrt{n}(\hat{\theta}-\theta)}{s}$ where s^2 is a $\sqrt{n}$ consistent estimator of the variance of $\sqrt{n}(\hat{\theta}-\theta)$. This procedure is often referred to as Studentization. t is said to be 'asymptotically-pivotal'. A pivotal statistic is one whose distribution is independent of the true parameter θ. Hartigan (1986) stressed the importance of using a pivotal statistic. See also Beran (1987, 1988).

These procedures for the construction of confidence intervals are all reviewed in DiCiccio and Romano (1988) and Hall (1988b, 1992) and we shall not repeat the details. The coverage errors have been shown to be

$O(n^{-1/2})$ for (i), (ii) and (iii) and

$O(n^{-1})$ for (iv) and (v). (Where n is the sample size.)

These results apply to coverage errors in each trial. Hall (1988a) shows that under quite general circumstances, the symmetric percentile-t confidence interval has coverage error $O(n^{-2})$. It is also proved that symmetric intervals are not necessarily any wider than equal-tailed intervals. In fact, Hall shows that in the case of the slope parameter in a simple regression, symmetric confidence intervals with coverage between 68% and 97% tend to be shorter than their equal-tailed counterparts. (For the 99% confidence level, the symmetric confidence interval is longer than the two-sided interval.) These results are based on the so-called "small sample asymptotics" (that is Edgeworth expansions) and hold for skew distributions.

The important conclusion that follows from these results is that, using the simple percentile method (ii) cannot be expected to produce an improvement over the asymptotic result (i). Many econometric applications, as we shall see, are based on (ii).

Some particular cases are of interest for us. Hall (1989) shows that in a regression model $y = \alpha + \beta' x + \varepsilon$, in the case of slope parameters, β, the percentile-t produces two-sided intervals with coverage error $O(n^{-2})$ and one-sided intervals with coverage error $O(n^{-3/2})$. For the constant term α or the conditional mean $(\alpha + \beta' x_0)$, the percentile-t has coverage error $O(n^{-1})$ only.

(vi) Beran's pivotal method. This was introduced in Beran (1987) and studied in Beran (1988, 1990). This is an iterative bootstrap method that involves a bootstrap within a bootstrap. Hence it is computationally intensive. This method applied to the symmetrical percentile-t leads to a coverage error $O(n^{-3})$. See Hall (1988a).

Hall (1988b) argues in favor of the percentile-t method in preference to Efron's BC_a. He shows that the difference between the percentile-t and the BC_a limits is $O(n^{-3/2})$. But more importantly, the BC_a method involves tedious analytical corrections, which the bootstrap methods are designed to avoid. In econometric work where the models are more complicated than those considered in these papers, the percentile-t is easier to use than the BC_a method, and hence is the preferable one. Rilstone (1993) compares the BC_a method with percentile-t in a Monte Carlo study and finds that the percentile-t performs better.

Several objections have been raised to the percentile-t method. These are:

(i) It produces bad results if the estimate of the variance is poor. The percentile-t method has been found to perform erratically when a jackknife estimate of the standard error is used. Also, it has been found that when used without using variance-stabilization transformation, the percentile-t method performs very poorly in the construction of a confidence interval for the correlation coefficient. See Hall (1992). In our simulations, (Li and Maddala, 1993), we found that the percentile-t did not perform well for the Johansen procedure. This is again due to the fact that the variance of the variance estimator for the Johansen (1988, 1991) ML estimator is high as compared to that of other asymptotically efficient estimators of the cointegrating vector. This point is discussed in Section 7.

(ii) The percentile-t method is not invariant to transformations. If it produces a confidence interval $(\widehat{a}, \widehat{b})$ for θ, and f is a monotone increasing function of θ, then the percentile-t does not, in general, give $[f(\widehat{a}), f(\widehat{b})]$ as a confidence interval for $f(\theta)$. This is quite troublesome in many econometric applications where Wald tests for non-linear hypotheses are often used and the Wald test statistic depends on how the non-linear constraint is formulated. For instance $H_0 : \beta_1 = -\beta_2^{-1}$ and $H_0 : \beta_1\beta_2 = -1$ are equivalent formulations. By contrast, the likelihood-ratio test and Rao's score test (LM test) are invariant to the different parameterizations. Gregory and Veall (1985) show that there are substantial size distortions with the use of the Wald test statistic, and these differ considerably under different formulations. This size distortion problem and the problem of invariance with the Wald statistic have been emphasized by many others. Horowitz and Savin (1992) argue that the bootstrap based critical values for the Wald test solve the size distortion problem and that in some cases, the power of the Wald test is higher than that of the LR test. Hence, the invariance of the Wald test need not be much of a concern. Jeong and Maddala (1994) also correct the size distortions in the Wald statistic using bootstrap methods.

2.2 Hypothesis Testing

In the statistical literature, hypothesis testing using critical values from bootstrap distributions has received less attention than construction of confidence intervals. (In the econometric literature, it is perhaps the reverse.) There are a few exceptions. Beran (1988) shows that if the asymptotic distribution of the test statistic is pivotal under the null, then under some regularity conditions the size of the bootstrap test has an error of a smaller order than the size of asymptotic theory test. Hinkley (1988) discusses briefly bootstrap tests of significance and this is followed by a more detailed discussion in Hinkley (1989). Hall and Wilson (1991) and Hall (1992) provide two general guidelines in hypothesis testing which, we shall discuss below. These guidelines have been violated in econometric practice but with good reasons. In the econometric literature the importance of using pivotal statistics in hypothesis testing is discussed in Horowitz (1994).

The familiar duality between hypothesis testing and construction of confidence intervals is maintained under bootstrap methodology as well. Given this duality, what we said earlier regarding the importance of (asymptotically) pivotal statistics in the construction of confidence intervals and the orders of approximation for the different methods applies to hypothesis testing as well. Thus, it is important to apply significance tests using (asymptotically) pivotal statistics. Otherwise, one cannot expect much of an improvement over the asymptotic results. This is confirmed, for instance, by the conflicting results in Veall (1986) who finds that the performance of bootstrap is no better than that of asymptotic theory, and Rayner (1991) and Rilstone (1993) who come to the opposite conclusion. The former bootstraps the coefficients $\widehat{\beta}$ and the latter two bootstrap $\widehat{\beta}/SE(\widehat{\beta})$ which is pivotal. Other examples of this will be provided in the following sections.

Two important issues concerning hypothesis testing using bootstrap methods relate to the questions about

(a) what test statistic to bootstrap and

(b) how to generate the bootstrap samples.

We have said that it is important to use a pivotal (or asymptotically pivotal) statistic in hypothesis tests. But there is another issue that has been brought up by Hall and Wilson (1991). Suppose that we want to test the hypothesis $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$. Given an estimator $\widehat{\theta}$ of θ, the usual test procedure would be based on $T = \widehat{\theta} - \theta_0$ and the significance level and p-values are obtained from the distribution of T under H_0. A direct application of the bootstrap procedure would suggest using the bootstrap distribution of $T^* = \widehat{\theta}^* - \theta_0$ instead of the distribution of T. ($\widehat{\theta}^*$ is the value of $\widehat{\theta}$ from the bootstrap sample.) However, Hall (1992, Section 3.12) discusses the bad behavior of the power of this test arguing that T^* does not approximate the null hypothesis when the sample comes from a distribution with parameter θ far away from θ_0. Hall and Wilson, therefore consider another bootstrap procedure based on the empirical distribution of $(\widehat{\theta}^* - \widehat{\theta})$. They compare this with the test procedure based on T and T^*.

Hall and Wilson propose two guidelines for hypothesis testing. The first suggests using the bootstrap distribution of $(\widehat{\theta}^* - \widehat{\theta})$ but not $(\widehat{\theta}^* - \theta_0)$. The second guideline suggests using a properly studentized statistic, that is $(\widehat{\theta}^* - \widehat{\theta})/\widehat{\sigma}^*$ but not $(\widehat{\theta}^* - \widehat{\theta})/\widehat{\sigma}$ or $(\widehat{\theta}^* - \widehat{\theta})$, where $\widehat{\sigma}^*$ is the estimate of $\widehat{\sigma}$ from the bootstrap sample.

Giersbergen and Kiviet (1993b) discuss these two rules in the context of hypothesis testing in regression models. To simplify the exposition, we shall discuss the case of a simple regression

$$y = \beta x + \varepsilon, \quad \varepsilon \sim iid(0, \sigma^2).$$

Let $\widehat{\beta}$ and $\widehat{\sigma}$ be the OLS estimators of β and σ respectively and $\widehat{\varepsilon}$ the OLS residuals. Let ε^* be the bootstrap residuals obtained by resampling $\widehat{\varepsilon}$. The null hypothesis to be tested is $H_0 : \beta = \beta_0$ vs. $H_1 : \beta \neq \beta_0$.

Consider two sampling schemes for the generation of the bootstrap samples:

$$S_1 : y^* = \widehat{\beta} x + \varepsilon^*$$
$$S_2 : y^* = \beta_0 x + \varepsilon^*.$$

Both use ε^* but they differ the way y^* is generated. For each sampling scheme consider two t-statistics

$$T_1 : T(\widehat{\beta}) = (\widehat{\beta}^* - \widehat{\beta})/\widehat{\sigma}^*$$
$$T_2 : T(\beta_0) = (\widehat{\beta}^* - \beta_0)/\widehat{\sigma}^*.$$

Thus four versions of the t-statistic can be defined. Hall and Wilson consider sampling scheme S_1 only and suggest using T_1 only. They do not consider sampling scheme S_2 which is the appropriate one for statistic T_2. Giersbergen and Kiviet suggest, on the basis of a Monte Carlo study of an

$AR(1)$ model, the use of T_2 under sampling scheme S_2 in preference to T_1 under S_1. The main conclusions of the paper are:

(i) Inference based on T_2 under S_1 does not just have low power but in fact has size close to zero. Similarly T_1 under S_2 does not work and should not be used. The resampling scheme should mimic the null distribution of the test statistic to be bootstrapped.

(ii) Using T_1 under S_1 and T_2 under S_2 are equivalent in non-dynamic models. This equivalence extends to the multiparameter case if one bootstraps the appropriate F-statistic. However, in dynamic models this equivalence breaks down in finite samples. The Monte Carlo results suggest that it is better to use T_2 under S_2.

(iii) The limiting distributions of T_1 under S_1 and T_2 under S_2 are identical even with dynamic models. The conclusion that T_2 under S_2 is better is based on small sample performance.

One can also propose another resampling scheme

$$S_3 : y^* = \beta_0 x + \varepsilon_0^*$$

where ε_0^* is the bootstrap sample from $\varepsilon_0 = y - \beta_0 x$ (after centering). Note that in both S_1 and S_2 we use resampling based on the OLS residuals $\widehat{\varepsilon}$. If the null $H_0 : \beta = \beta_0$ is true but the OLS estimator $\widehat{\beta}$ gives a value of β far from β_0, the empirical distribution of the residuals will suffer from a poor approximation of the distribution of the errors under the null. The intuition behind S_3 is as follows. If the null hypothesis is true, $\varepsilon_0 = y - \beta_0 x$ is exactly the true distribution of the regression errors. Hypothesis testing based on this will give (approximately) the correct test size. If the null is not true, then ε_0 is different from the true distribution of the errors. Hypothesis testing will give proper power depending on how far the null is away from the true value of β. Thus, using T_2 under S_3 is better than using T_1 under S_1 or T_2 under S_2 which are identical as shown in Giersbergen and Kiviet. The idea of using restricted regression errors for resampling has been used in Li and Maddala (1993) and Nankervis and Savin (1994).

As we shall discuss in Section 6, for the unit root model, test statistic T_2 under sampling scheme S_2 or S_3 are the only ones that can be justified on asymptotic grounds. It is not appropriate to use T_1 under S_1.

2.3 Methods of Generation of Bootstrap Samples: Residual Based vs. Direct Methods

In the preceding section we discussed some issues about the generation of bootstrap samples and how this is related to the hypothesis test under consideration. The resampling methods were all residual based methods. We shall now discuss issues related to direct resampling of the data.

For the case of random regressors (which he calls the "correlation model" as opposed to the "regression model") Freedman (1981) suggests resampling the pair (y, X), which have a joint distribution with $E(y|X) = X\beta$. In this model the pairs (y_i, X_i) are assumed to be *IID*. Given a sample of size n we can compute $\widehat{\beta}$, the least squares estimator of β. Denote this by $\widehat{\beta}(n)$. Then $\sqrt{n}(\widehat{\beta}(n) - \beta)$ is asymptotically normal with mean 0 and a certain covariance matrix which we shall denote by Ω. Let m be the size of the bootstrap sample and denote the bootstrap estimator by $\widehat{\beta}^*(m)$. Freedman (1981, p. 1226) shows that under some conditions, as $m \to \infty$ and $n \to \infty$

$$\sqrt{m}(\widehat{\beta}^*(m) - \widehat{\beta}(n)) \to N(0, \Omega).$$

What this result shows is that the bootstrap distribution based on direct sampling is useful as an approximation of the asymptotic distribution of $\sqrt{n}(\widehat{\beta}(n) - \beta)$. In cases where the asymptotic distribution can be derived, this result is not of much use except as a consolation that using the bootstrap is alright. In this case the bootstrap method cannot be expected to improve upon the asymptotic result without using a pivotal method. However, in the case of heteroskedasticity of an unknown form, this direct method even without pivoting is useful. In fact, in this case, even if the model is the usual regression model with fixed X's (as opposed to the model with random X's) the direct method of sampling (y_i, X_i) is useful because it gives an estimate of the correct covariance matrix Ω. However, randomly resampling pairs (y_i, X_i) does not impose the restriction $E(u|X) = 0$ where u is the error in the regression equation. It has been found that the numerical performance of the bootstrap can be improved by imposing this restriction. The bootstrap method that imposes this restriction is the "wild bootstrap" introduced by Liu (1988). On the other hand, if one uses White's heteroskedasticity-consistent covariance matrix estimator, one can obtain asymptotically valid pivotal statistics in the presence of heteroskedasticity of an unknown form.

Efron (1981) uses the direct method of resampling the data in a problem involving censored data. He bootstraps the data (x_i, d_i) where $d_i = 1$ if x_i is not censored and $d_i = 0$ if x_i is censored. His model does not have covariates. Some applications of this direct method to censored regression models (that is models with covariates) are reviewed in Jeong and Maddala (1993, pp. 591-4). The problem here is one of non-linearity of an unknown form. But the basic question is: What do you do after generating the bootstrap sample? Suppose you have the data on a latent variable y_i which is observed as a dummy variable $d_i(= 1$ or $0)$ and a set of covariates X_i. We draw the bootstrap sample (d_i^*, X_i^*). What do we do next? If we are going to estimate a logit model, then we can as well estimate the logit model with the actual data, compute the generalized residuals, and resample the generalized residuals to generate bootstrap samples. In this case the direct bootstrap method does not make use of all the information as the method based on generalized residuals. It is only if some general semiparametric method of estimation is used that the direct bootstrap method would be useful. We cannot go into a detailed discussion of the appropriate bootstrap methods for the censored regression model here. But we will argue that the above considerations carry over to time series models as well.

The direct method of sampling the data, rather than the residuals, has not gone unnoticed in econometrics. The earliest use is in Kiviet (1984) which we will discuss later. Veall (1987, p. 205) considers the direct bootstrap approach but rejects it on grounds that it does not embody all the information used in the residual based approach. Li and Maddala (1993) actually use the direct sampling approach combined with the moving block method (discussed in Section 4) to data vectors (y_i, x_i) in cointegrating regression models and find its performance worse than that of the residual based bootstrap. Here we shall review some problems with the use of the direct bootstrap approach in econometric applications once we move out of the framework of the usual regression model. These problems have not received sufficient attention.

Consider first the lagged dependent variable model. Freedman and Peters (1984) consider a q variable model and introduce the recursive bootstrap method. The model is

$$\underset{1\times q}{Y_t} = \underset{1\times q}{Y_{t-1}} \underset{q\times q}{B} + \underset{1\times p}{X_t} \underset{p\times q}{C} + \underset{1\times q}{\varepsilon_t}, \quad t = 1, 2, \cdots, n, \quad Y_0 \text{ known.}$$

ε_t are assumed $iid(0, \Sigma)$. $\widehat{B}$, $\widehat{C}$ are first estimated using a system method (that takes care of contemporaneous correlation among the errors ε_t), and the residuals $\widehat{\varepsilon}_t = Y_t - Y_{t-1}\widehat{B} - X_t\widehat{C}$ are calculated. Bootstrap samples ε_t^* are generated. Since the entire vector of residuals is resampled, this preserves the contemporaneous correlation structure of the errors. The bootstrap sample Y_t^* are generated in a recursive fashion, assuming X_t, and Y_0 are given, using the equation

$$Y_t^* = Y_{t-1}^*\widehat{B} + X_t\widehat{C} + \varepsilon_t^*.$$

Now reestimate the model using (Y_t^*, X_t and Y_0). This method was used in an empirical application with $q = 10$ and $n = 18$. It is shown that the conventional asymptotic standard errors are substantially (by about 40%-50%) below the bootstrap standard errors. The empirical results, of course, merely say that the asymptotic and bootstrap standard errors are different.

There is no way of telling without a Monte Carlo study how biased each is. Freedman and Peters argue that the problem lies in the estimation of the covariance matrix Σ in the conventional procedure, which can be poor with such a limited sample. It is also worth noting that Freedman and Peters concentrated on the standard errors and did not consider bootstrapping the studentized t-statistic, which on asymptotic grounds, is expected to give better confidence intervals, as argued earlier.

Suppose we abstract from the multiple equation issue and concentrate on the estimation of a simple dynamic equation

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + \beta_3 x_{t-1} + \varepsilon_t. \tag{1}$$

For this model Kiviet (1984) presents the results of a Monte Carlo study and finds that the bootstrap does not improve on conventional asymptotic inference. Even though this paper is out of date (and was never published) there are a few points in this paper worth noting which are useful for further work. Kiviet considered two error structures: One in which ε_t are $IN(0,1)$ and the other in which ε_t are $IL(0,1)$ where L is the Laplace distribution. Also the problem may have been caused by the use of the percentile method which is known to give no improvement over asymptotic results. The percentile-t bootstrap as discussed earlier, should do better. One suspects from Kiviet's Monte Carlo study that the substantial discrepancies observed by Freedman and Peters between the asymptotic and bootstrap standard errors may have been caused by the high dimensionality of their system ($q = 10$). However, one needs to investigate this by a Monte Carlo study.

It is interesting to note that Kiviet considers the recursive method suggested by Freedman and Peters (1984) but he also considers a direct resampling scheme whereby the data vector $(y_t, y_{t-1}, x_t, x_{t-1})$ is bootstrapped. This did not give any improvement over asymptotic results either. What this suggests is that the important issue is not residual sampling vs. direct sampling of data but one of percentile vs. percentile-t. In fact this dynamic model was analyzed by Giersbergen and Kiviet (1993a) using the percentile-t with substantial improvement over the asymptotic results (although an iterative percentile-t that they suggest, did better), as far as the confidence interval construction is concerned.

There are, in fact, more problems with the direct method in this case. Suppose the errors ε_t in (1) are $AR(1)$. Then there is a serial correlation bias in the OLS parameters in (1). Suppose that the information that ε_t in (1) are $AR(1)$ is used in the estimation of the model from the bootstrap sample, then this information is incorporated in the residual method but is not used in the direct method of generation of the bootstrap sample by resampling the data.

Thus, we feel that the direct method ignores some important information utilized in the residual based sampling in the generation of bootstrap samples of dynamic models.

Our basic argument is that whatever information about the error term is used in the estimation of the model from the bootstrap samples, should also be used in the generation of the bootstrap samples. This is done in the residual based bootstrap generation but not in the method of bootstrapping the data. This issue is not important in the case of models with no lagged dependent variables, where one can bootstrap the data and use White's heteroskedasticity-consistent covariance matrix to get asymptotic pivotal statistics and thus get asymptotic refinements using bootstrap methods. But in the case of lagged dependent variables with serially correlated errors, there is also a problem of consistency of the resulting estimators if the order of serial correlation is not properly taken into account.

Turning next to the application of the direct method to the cointegrated regression models, the problems are similar. Suppose that y_t and x_t are $I(1)$ and we have the regression equation

$$y_t = \beta x_t + u_t. \tag{2}$$

Suppose we bootstrap the data (y_t, x_t) and estimate equation (2) by OLS. If u_t is also $I(1)$, then it is well known that equation (2) is a spurious regression. But there is no way of knowing this if we use the direct bootstrap method, without first testing whether (2) is indeed a cointegration relationship.

Suppose we initially apply cointegration tests to equation (2) and make sure that equation (2) is a meaningful regression relationship. This implies that y_t and x_t are $I(1)$ and u_t is $I(0)$. (An $I(0)$ variable is stationary and an $I(1)$ variable is stationary in first differences.) But the direct bootstrap method does not use the information that u_t is $I(0)$. This point is discussed further in Section 7. In fact, in Li and Maddala (1993) the direct method was used in several Monte Carlo studies of the comparison of moving block procedure with asymptotic methods and it was found that the performance of the direct method was slightly worse than the one based on bootstrapping the residuals, although the direct method provided an improvement over asymptotic results. It should be noted that Li and Maddala considered bootstrapping a t-statistic in both cases and hence that is not an issue here.

We have reviewed several cases where the direct method of bootstrapping the data has been used. We have also outlined some problems in the use of the direct method as compared to bootstrapping residuals. These problems have not been appreciated in econometric work and it is often thought that the direct method, which is the simplest, is universally applicable.

3 Structured Time Series Models: The Recursive Bootstrap

By structured time series models we mean stationary $ARMA(p,q)$ models with known p and q. We shall consider stationary models first. One particular type of nonstationarity that has received considerable attention in the econometric literature is the unit root model which we shall discuss in Section 6. The $AR(1)$ model and regression models with $AR(1)$ errors will be discussed in detail later.

3.1 *ARMA* Models

Consider the stationary $AR(p)$ model

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + e_t, \quad e_t \sim iid(0, \sigma^2). \tag{3}$$

Given data on $n+p$ observations $(y_{1-p}, \cdots, y_0, y_1, \cdots, y_n)$ our objective is to get confidence intervals for the parameters a_i or some smooth function $h(a_1, a_2, \cdots, a_p)$ of the parameters a_i. First, we estimate (3) by OLS based on n observations $(y_1, y_2, \cdots, y_n)$ (the observations $(y_{1-p}, \cdots, y_0)$ are used as initial values) and we get the least squares residuals $\widehat{e}_t$. Define the centered and scaled residuals

$$\widetilde{e}_t = \left(\widehat{e}_t - \frac{1}{n}\sum \widehat{e}_t\right)\left(\frac{n}{n-p}\right)^{1/2}.$$

The rescaling has been suggested by Bickel and Freedman (1983) because the residuals tend to be smaller than the true errors. (They do acknowledge that scaling is just an ad hoc solution.) Then resample $\widetilde{e}_t$ with replacement to get the bootstrap residuals e_t^*. Next we construct the bootstrap sample

$$y_t^* = \sum_{i=1}^{p} \widehat{a}_i y_{t-i}^* + e_t^*.$$

This is done recursively using $y_t^* = y_t$ for $t = 1-p, \cdots, 0$ in each bootstrap iteration. This recursive method was first used in Freedman and Peters (1984) and subsequently by many others. Efron and Tibshirani (1986) also use this recursive method for bootstrapping the estimates from $AR(1)$ and $AR(2)$ models.

Bose (1988), using Edgeworth expansions, shows that the least squares estimates $\hat{a}_i$ can be bootstrapped with accuracy $o(n^{-1/2})$ a.s. thereby improving the normal approximation error of $O(n^{-1/2})$. In a subsequent paper, Bose (1990) shows that the bootstrap approximation of the parameter estimates in an invertible MA process is also accurate $o(n^{-1/2})$ a.s. He also reports some limited simulation results with the $AR(1)$ and $MA(1)$ models under the normal and the centered exponential distributions for the errors. Chatterjee (1986) also provides simulations with the bootstrap methods for the parameters in $ARMA$(1,1), $AR(2)$, and $MA(2)$ models under a stable paretian error distribution which includes the normal and the Cauchy as special cases. He concentrates on the standard errors only and concludes that in the case of the $MA(2)$ and $ARMA$(1,1) models, the bootstrap estimates of the standard errors of the MA parameters are overestimated compared to the small sample standard errors obtained by simulation.

It is well known that estimation of MA time series models is not as straightforward as the estimation of the AR models. See, for instance, Davidson (1981). These problems carry over to the application of bootstrap methods to MA models. Kreiss and Franke (1992) suggest M-estimators for $ARMA$(1,1) with error distributions which are normal, double-sided exponential and contaminated normal, respectively. Note that, since each model has a different error distribution it is not possible to judge the effect of different error distributions in each case. They compare the bootstrap distributions with the normal approximation. The former are skewed whereas the latter is symmetric. Note, however, that they report two bootstrap approximations with only two samples, that is two random samples from the true model are generated, and the parameters estimated in each case are bootstrapped. In this sense this is not a complete Monte Carlo study. (It is based on just two draws.)

There is a need for a more thoroughly done Monte Carlo study on MA and $ARMA$ models where there is a comparative evaluation of coverage errors of bootstrap methods and asymptotic methods. Furthermore this comparison needs to be done using studentized statistics. In the study by Chatterjee (1986), for instance, the concentration is on standard errors of the estimators of the parameters. In the study by Kreiss and Franke (1992) the concentration is on comparing skewness. Although these problems are interesting, they do not tell the whole story about a comparison of different methods for tests of hypotheses or construction of confidence intervals.

One study that compares the coverage probabilities of bootstrap confidence intervals with that of asymptotic procedures is by Wet and Wyk (1986). The $AR(1)$ part of their study is not of much interest because, as we shall review later, their study has been surpassed by subsequent work on this model. However, their $MA(1)$ study might be of interest, because it uses studentized statistics and symmetric bootstrap intervals for the $MA(1)$ model.

Wet and Wyk consider a time trend model with $AR(1)$ and $MA(1)$ error structures (which has the same form as in Eriksson (1983) who studied its asymptotics):

$$x_t = \alpha + \beta(t - \bar{t}) + \varepsilon_t, \quad t = 1, 2, \cdots, n$$

$$\begin{pmatrix} \varepsilon_t \\ \varepsilon_{t-s} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \gamma_s \\ \gamma_s & \sigma^2 \end{pmatrix} \right).$$

Define the correlation $\rho_s = \gamma_s/\sigma^2$. $\{\varepsilon_t\}$ can be either $AR(1)$

$$\varepsilon_t = \phi\varepsilon_{t-1} + \eta_t$$

or $MA(1)$

$$\varepsilon_t = \omega\eta_{t-1} + \eta_t$$

where $\eta_t \sim iidN(0, \sigma_0^2)$. η_t can also have a non-normal distribution. Let $(\widehat{\alpha}, \widehat{\beta})$ be the OLS estimate of (α, β), $\{\widehat{\varepsilon}_t\}$ be the OLS residuals. Define

$$\widehat{\gamma}_s = (n - s)^{-1} \sum_{t=1}^{n-s} \widehat{\varepsilon}_t \widehat{\varepsilon}_{t+s}, \quad s^2 = (n - 2)^{-1} \sum_{t=1}^{n} \widehat{\varepsilon}_t^2.$$

For the $AR(1)$ model, the estimated asymptotic variances (which can be found in Eriksson, 1983) of $(\widehat{\alpha}, \widehat{\beta})$ are:

$$var(\widehat{\alpha}) = n^{-1}s^2(1 + \widehat{\phi})(1 - \widehat{\phi})^{-1}$$

and

$$var(\widehat{\beta}) = (n^3 - n)^{-1}12s^2(1 + \widehat{\phi})(1 - \widehat{\phi})^{-1}$$

where $\widehat{\phi} = s^{-2}\widehat{\gamma}_1$ and for the $MA(1)$ model

$$var(\widehat{\alpha}) = n^{-1}(s^2 + 2\widehat{\gamma}_1)$$

and

$$var(\widehat{\beta}) = (n^3 - n)^{-1}12(s^2 + 2\widehat{\gamma}_1).$$

In what follows, all $var(\cdot)$ expressions are estimated asymptotic variances. Based on the asymptotic normality of $(\widehat{\alpha}, \widehat{\beta})$, the $(100-2\gamma)$ percent confidence intervals for (α, β) are

$$\widehat{\alpha} \pm z_\gamma(var(\widehat{\alpha}))^{1/2} \quad \text{and} \quad \widehat{\beta} \pm z_\gamma(var(\widehat{\beta}))^{1/2}$$

where z_γ is the upper γ quantile of the standardized normal distribution. To construct the bootstrap confidence interval, the authors follow the recursive method. Next the bootstrap distributions $t^*_{\alpha 1}, \cdots, t^*_{\alpha B}$ and $t^*_{\beta 1}, \cdots, t^*_{\beta B}$ are obtained where the t-statistics are defined as

$$t^*_\alpha = (\alpha^* - \widehat{\alpha})(var(\alpha^*))^{-1/2} \quad \text{and} \quad t^*_\beta = (\beta^* - \widehat{\beta})(var(\beta^*))^{-1/2}$$

where α^*, β^*, $var(\alpha^*)$ and $var(\beta^*)$ are estimated from bootstrap samples. Note that the t-statistics bootstrapped follow the two guidelines of Hall and Wilson discussed in the previous section. Take the absolute values of the above t-statistics and obtain the γ-*th* quantile, $z^*_\gamma(\alpha)$ and $z^*_\gamma(\beta)$. The bootstrap $(100 - \gamma)$ percent confidence intervals are hence

$$\widehat{\alpha} \pm z^*_\gamma(\alpha)(var(\alpha^*))^{1/2} \quad \text{and} \quad \widehat{\beta} \pm z^*_\gamma(\beta)(var(\beta^*))^{1/2}.$$

The simulation results (based on sample sizes 10 to 60) presented in Wet and Wyk (1986) show that as far as the confidence intervals are concerned, the bootstrap confidence intervals have better coverage than those based on asymptotic theory. This is one study that bootstraps the t-statistic in $MA(1)$ model, and also uses the symmetric bootstrap interval. Needless to say, a more thorough analysis is also called for in this case (including bootstrap data generation using the null and changing the test statistics accordingly).

We shall now turn to the $AR(1)$ model which has been extensively analyzed.

3.2 The $AR(1)$ Model

The simplest $AR(1)$ model is given by

$$y_t = \beta y_{t-1} + u_t$$

with $y_0 = 0$, $u_t \sim iid(0, \sigma^2)$ and $-\infty < \beta < \infty$. We can divide β into three regions. When $|\beta| < 1$ the process $\{y_t\}$ is stationary. When $|\beta| = 1$ the process is unstable. This is known as the unit root model. When $|\beta| > 1$ the process is explosive. Rubin (1950) showed that the OLS estimator $\widehat{\beta}$ of β is consistent in the range $(-\infty, \infty)$. However, the asymptotic distributions of $\widehat{\beta}$ in the different ranges are different. In the stationary case the asymptotic distribution is normal. For the explosive case, Anderson (1959) shows that the limiting distribution is a Cauchy distribution. The asymptotic distribution in the different cases with different normalizing constants are tabulated in Table 2.1 of Fuller (1985).

The unit root and explosive cases are discussed in Section 6. Here we shall discuss the stationary case. The $AR(1)$ model with intercept is

$$y_t = \alpha + \beta y_{t-1} + u_t, \quad u_t \sim iid(0, \sigma^2). \tag{4}$$

However, since the distribution of the OLS estimator $\widehat{\beta}$ of β is invariant to

α and σ^2, we might as well set $\alpha = 0$ and $\sigma^2 = 1$ and consider the model

$$y_t = \beta y_{t-1} + u_t, \quad u_t \sim iid(0,1). \tag{5}$$

Dufour (1990) and Andrews (1993) have developed exact inference procedures for the $AR(1)$ parameter but these depend on the normality assumption of the errors. Hence bootstrap methods which are robust to distributional assumptions of the errors hold promise. The procedure for the generation of the bootstrap samples is the recursive procedure described in the previous section. However, after computing the least squares residuals $\widehat{u}_t$ and the bootstrap residuals u_t^*, we use β_0 but not $\widehat{\beta}$ in generating y_t^*. This is sampling scheme S_2 discussed in Section 2.2 earlier.

The first bootstrap application of this model is by Rayner (1990). He uses the percentile-t (or the bootstrap-t method) and is concerned with the problem of testing the hypothesis $H_0 : \beta = \beta_0$. He does not examine the percentile method, but examines the studentized statistic

$$t = (\widehat{\beta} - \beta_0)/SE(\widehat{\beta}).$$

He considers approximating the distribution of this t-statistic by the bootstrap distribution of t^* which is defined as

$$t^* = (\beta^* - \beta_0)/SE(\beta^*)$$

where β^* and $SE(\beta^*)$ are the values of $\widehat{\beta}$ and $SE(\widehat{\beta})$ computed from the bootstrap sample. This violates the Hall and Wilson rule which says that $\widehat{\beta}$ but not β_0 should be used in t^*. However, Rayner modifies the sampling rule by using β_0 but not $\widehat{\beta}$ in generating y^*. (See the discussion in Section 2.2 earlier. That is, he uses the t-statistic T_2 with sampling rule S_2.)

Rayner finds that the use of the student-t approximation is not satisfactory, particularly for high values of β and that the bootstrap-t performs very well in samples of sizes 5-10, even when mixtures of normal distributions are used for the errors. He also argues that the bootstrap approximates the power well for samples as small as 5-10. Finally, for the starting value y_0, he says, it is important to use a random starting value (from the equilibrium distribution of y_t).

Rayner's argument that the bootstrap sample generation should be consistent with the null hypothesis being tested is correct (see the discussion in Section 2.2 earlier). So is his use of the bootstrap-t instead of the percentile method. However, his conclusion about the sample sizes (as small as 5-10) at which the performance of the bootstrap is good is very surprising. So is his conclusion regarding the robustness of the procedure to departures from normality with small sample sizes. Subsequent studies have failed to confirm these conclusions.

Giersbergen and Kiviet (1993a) do a Monte Carlo study with the same model as Rayner. In addition to the percentile-t they also consider the usual percentile method and an iterative percentile-t. The sampling scheme used was T_1 and S_1 in the case of percentile and percentile-t methods and T_2 and S_2 in the case of the iterative percentile-t method. As expected, the percentile method did not perform any better than the asymptotic theory. The percentile-t method performed much better but even for a sample size $n = 40$, the percentile-t gave a significance level of .08 compared to a nominal significance level of .05. In their study, the iterative percentile-t did better giving a significance level of around .05 even for samples of size $n = 20$. As far as robustness to non-normality is concerned, they investigated a shifted χ^2 distribution – shifted to have mean zero, truncated t distributions, and $ARCH$ disturbances. The performance of all the methods deteriorated, particularly with $ARCH$ disturbances but the iterative percentile-t performed best. As mentioned in Section 2.3, these authors also discuss an $AR(1)$ model with an exogenous regressor and again find the performance of the iterative percentile-t the best. Due to space limitation, we shall not discuss their iterative percentile-t method.

There are a few differences in the way the bootstrap sample is generated in the Giersbergen-Kiviet study. They condition the sample generation on y_0, unlike the case in Rayner's study. They also adjust the least squares residuals by centering and scaling (see the description in the previous section).

Nankervis and Savin (1994) also replicate Rayner's study using Rayner's method of treating y_0. They consider three models: (i) The $AR(1)$ model discussed here, (ii) A linear trend model with $AR(1)$ errors which we shall discuss in Section 3.3 and (iii) The unit root model which we shall discuss in Section 6.

First Nankervis and Savin use the test statistic T_2 with sampling scheme S_3 (see Section 2.2) but not S_2 as done by Rayner. Thus, their results are not exactly comparable to Rayner's. They find that the bootstrap-t test has the correct level for sample size of 10 when the error distribution is normal but it suffers substantial size distortions when the errors follow a mixture normal or lognormal distributions, although the distortions are less than with the usual t-test. However, for samples of size 50, the bootstrap-t test has the correct size even for these non-normal distributions. Also, the empirical power of the bootstrap-t test is identical to that of the usual t-test assuming normality. Thus, there is no loss of power in using the bootstrap-t test even when the errors have a normal distribution. We shall skip more details, since this is a very restricted model.

In summary, the bootstrap-t test performs well in having the correct size and good power, for samples of size 50 (lower if errors are normal). The results presented by Rayner are not likely to be extended to non-normal errors because the sample sizes he considers are very low (5-10). The major differences between the three studies by Rayner, Giersbergen-Kiviet, and Nankervis-Savin, are in the treatment of y_0 and the way any information of the form $|\beta| < 1$ is exploited in the construction of confidence regions.

3.3 Regression Models with $AR(1)$ Errors

In the estimation of regression models with autocorrelated errors, a number of Monte Carlo studies have unanimously concluded that if the explanatory variables are strongly trended and the degree of serial correlation is high, the generalized least squares (GLS) standard errors of the coefficients will be substantially understated (thus leading to the overrejection of the null hypotheses). See Kwok and Veall (1988) for the references which, to conserve space, we shall not repeat here.

Veall (1986) investigated whether the bootstrap method solves this problem. He concludes that "there is no evidence that bootstrap procedures will improve inference". His pessimistic conclusion, however, is based on the fact that he used the percentile method. Rilstone (1993) replicates Veall's study exactly, and investigates Efron's BC_a method, the percentile-t (or bootstrap-t) method and an iterated bootstrap method. The BC_a method did not provide much of an improvement over the results obtained by Veall, but the bootstrap-t provided substantial improvement and gave approximately the correct test sizes. The iterated bootstrap was not worth the extra computational effort. All in all, the bootstrap-t is to be recommended agreeing with the conclusion of Hall noted in Section 2.2 earlier.

Kwok and Veall (1988) argue that the problem with the GLS method noted earlier was perhaps due to the fact that all standard estimators of ρ (Cochrane-Orcutt, Durbin, ML) tend to be biased downwards. They suggest using a jackknife estimator of ρ originally due to Quenouille which is

$$\tilde{\rho} = 2\hat{\rho} - (\hat{\rho}_1 + \hat{\rho}_2)/2$$

when $\hat{\rho}$, $\hat{\rho}_1$, $\hat{\rho}_2$ are estimators of ρ form the whole sample and two half-samples (splitting the entire sample into two). If this estimator gives a value ≥ 1 they suggest using a trimmed estimator (replacing the upper limit at say .999). Kwok and Veall found that the estimated GLS using the half-sample jackknife provided better inference than conventional GLS. For instance for $\rho = 0.9$, and nominal size of 5%, using the ML estimation method, the conventional

GLS gave an empirical size of 21.1% whereas the jackknife based GLS gave an empirical size of 5.7%.

Rayner (1991) also replicates Veall's study using the Prais-Winsten estimator method and the half-sample jackknife estimator of ρ. Unlike Veall he uses the percentile-t method. Rayner's simulation results suggest that the bootstrap – in combination with a bias reduction method such as the half-sample jackknife – substantially corrects the problem discussed earlier with GLS estimated standard errors. Rilstone (1993, p. 338 fn.) on the other hand rejected the jackknife estimator of ρ on the grounds that when the true ρ was 0.9 it gave $\tilde{\rho} > 1$ in about 25% of the cases.

In all these studies, however, the attention is on getting the test size right. There is no discussion of the power of the tests.

Nankervis and Savin (1994) also consider this model but instead of specifying x_t as a trending regressor, they specify it as t. The model therefore, is

$$y_t = \alpha_0 + \alpha_1 t + u_t, \quad u_t = \beta u_{t-1} + e_t \tag{6}$$

where $e_t \sim iid(0, \sigma^2)$, $t = 1, 2, \cdots, n$. They work with the reduced form of this model which is

$$y_t = \gamma + \delta t + \beta y_{t-1} + e_t \tag{7}$$

where $\gamma = (\alpha_0(1-\beta) + \alpha_1\beta)$ and $\delta = \alpha_1(1-\beta)$. They again use the test statistic T_2 and sampling scheme S_3 to test the hypothesis $H_0 : \beta = \beta_0$. Their major findings for this model are:

(1) The sample sizes required for the bootstrap-t to work are much higher than for the $AR(1)$ model described in the previous section. For $n = 20$ or 50 there were substantial size distortions. For $n = 100$, these size distortions usually disappear.

(2) In general, the symmetric t-test outperforms the equal-tailed test (see our discussion in Section 2.1).

(3) The bootstrap-t does not perform as well when the errors have a Cauchy distribution.

Unlike the studies quoted earlier, their study also compares the power of the bootstrap-t test with that of a comparable t-test. The general conclusion is that there is no loss in power in using the bootstrap-t test.

We need not go into all the details here. The general conclusion that emerges from these studies is that the bootstrap-t test can be used in regression models with $AR(1)$ errors and trending regressors although a reasonably large sample $n \geq 100$ is required for the bootstrap method to work. This is true for a variety of non-normal errors except for some heavy-tailed distributions (like the Cauchy in which case the bootstrap-t does not perform well).

These results possibly carry over for higher order AR processes for the errors. The case of MA and $ARMA$ errors needs further investigation. We conjecture that the sample sizes required for the bootstrap to work well (provide the correct size of tests) would be much higher in these models, although it is possible that bootstrap methods would provide some improvement over conventional methods at lower sample sizes. See the evidence in Wet and Wyk (1986) quoted in Section 3.1.

4 General Error Structures: The Moving Block Bootstrap and Extensions

As we discussed in the previous section, application of the residual based bootstrap methods is straightforward if the error distribution is specified to be an $ARMA(p,q)$ process with known p and q. However, if the structure of serial correlation is not tractable or is misspecified, the residual based methods will give inconsistent estimates (if lagged dependent variables are present in the system). Other approaches which do not require fitting the data into a parametric form have been developed to deal with general dependent time series data. Carlstein (1986) first discussed the idea of bootstrapping blocks of observations rather than the individual observations. The blocks are non-overlapping. Künsch (1989) and Liu and Singh (1992) (the paper was available as a discussion paper in 1988) independently introduced a more general bootstrap procedure, the moving block bootstrap (MBB) which is applicable to stationary time series data. In this method the blocks of observations are overlapping.

The methods of Carlstein (non-overlapping blocks) and Künsch (overlapping blocks) both divide the data of n observations into blocks of length l and select b of these blocks (with repeats allowed) by resampling with replacement all the possible blocks. Let us for simplicity assume $n = bl$. In the Carlstein procedure, there are just b blocks. In the Künsch procedure there are $n-l+1$ blocks. The blocks are $L_k = \{x_k, x_{k+1}, \cdots, x_{k+l-1}\}$ for $k = 1, 2, \cdots, (n-l+1)$. For example with $n = 6$ and $l = 3$ suppose the data are: $x_t = \{3, 6, 7, 2, 1, 5\}$. The blocks according to Carlstein are $\{(3, 6, 7), (2, 1, 5)\}$. The blocks according to Künsch are $\{(3, 6, 7), (6, 7, 2), (7, 2, 1), (2, 1, 5)\}$. Now draw a sample of two blocks with replacement in each case. Suppose, the first draw gave $(3, 6, 7)$. The probability of missing all of $(2, 1, 5)$ is $1/2$ in Carlstein's scheme and $1/4$ in the moving block scheme. Thus there is a higher probability of missing entire blocks in the Carlstein scheme. For this reason, it is not popular, and is not often used. Our own experience with Carlstein's non-overlapping block method is that it gave very erratic results as the block length was varied. The *MBB* did better.

The literature on blocking methods is mostly on the estimation of the sample mean and its variance, although Liu and Singh (1992) talk about the applicability of the results to more general statistics, and Künsch (1989, p. 1235) discusses the $AR(1)$ and $MA(1)$ model. In all these studies the bootstrapping is done by sampling blocks of the data. In our study, Li and Maddala (1993), since it was a regression model we used blocks of the residuals. The block sampling of the data was also tried, but its performance was slightly worse (for the reasons given earlier in Section 2.3). Many of the rules suggested for the optimal block length can possibly be adapted for regression models with $(n - p)$ substituted for n, where p is the number of regressors.

4.1 Problems with *MBB*

There are some important problems worth noting about the moving block bootstrap procedure.

(1) The pseudo time series generated by the moving block method is not stationary, even if the original series $\{x_t\}$ is stationary. For this reason, Politis and Romano (1994) suggest the stationary bootstrap method. This involves sampling blocks of random length, where the length of each block has a geometric distribution. They show that the pseudo time series generated by the stationary bootstrap method is indeed stationary.

The application of moving block method to $I(1)$ variables has more problems. Suppose that $\{x_t\}$ is $I(1)$. Then it is not necessarily true that the pseudo data $\{x_t^*\}$ generated by the moving block bootstrap is also $I(1)$.

(2) The mean $\overline{x}_n^*$ of the moving block bootstrap is biased in the sense that $E(\overline{x}_n^*|x_1, x_2, \cdots, x_n) \neq \overline{x}_n$. See the result (iii) of Theorem 6 in Liu and Singh (1992, p. 241). Politis and Romano (1994) show that, in contrast, for the stationary bootstrap procedures, $E(\overline{x}_n^*|x_1, x_2, \cdots, x_n) = \overline{x}_n$.

(3) The moving block bootstrap estimator of the variance of $\sqrt{n}\overline{x}_n$ is also biased. Davison and Hall (1993) argue that this creates problems in using the percentile-t method with the moving block bootstrap. They suggest that the usual estimator $\widehat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2$ be modified to $\widetilde{\sigma}^2 = n^{-1}\sum_{i=1}^{n}\{(x_i - \overline{x}_n)^2 + \sum_{k=1}^{i-1}\sum_{i=1}^{n-k}(x_i - \overline{x}_n)(x_{i+k} - \overline{x}_n)\}$. With this modification the bootstrap-t can improve substantially on the normal approximation. The reason for this bias in the estimator of the variance is that the block bootstrap method damages the dependence structure of the data. Unfortunately this formula is valid only

for the variance of $\sqrt{n}\bar{x}_n$. For more complicated problems there is no such simple correction available. Hence, in the study of Li and Maddala (1993) no such corrections were applied. However, the Monte Carlo studies showed that the bootstrap-t provided considerable improvement over asymptotic results.

(4) In a subsequent paper, Hall and Horowitz (1994) investigate this problem in the context of tests based on GMM estimators. They argue that because the blocking methods do not replicate the dependence structure of the original data, it is necessary to develop special bootstrap versions of the test statistics and these must have the same distribution as the sample version of the test statistics through $O_p(n^{-1})$. They derive the bootstrap versions of the test statistics with Carlstein's blocking scheme (non-overlapping blocks) but argue that Künsch's blocking scheme is more difficult to analyze owing to its use of overlapping blocks.

In the case of cointegration tests based on the moving block scheme, as studied in Li and Maddala (1993), the derivation of the appropriate bootstrap versions of the test statistics is still more complicated. Although the use of the bootstrap version of the usual test statistics cannot be theoretically justified, the Monte Carlo results unequivocally indicate considerable improvement over the asymptotic results.

4.2 Optimal Length of Blocks

There is some discussion on the optimal length of the blocks and the several rules that have been suggested are based on different criteria. However, the rules are useful as rough guides to selecting the optimal sized blocks.

First, the number of blocks should be lower under the non-overlapping blocks rule of Carlstein than in the moving blocks rule of Künsch. In the stationary bootstrap approach, where blocks of random length are sampled, the average length of a block is $1/p$, where p is the parameter of the geometric distribution. Thus, $1/p$ should play the same role as the parameter l in the moving block bootstrap. Politis and Romano (1994) argue that the application of stationary bootstrap is less sensitive to the choice of p than the application of moving block bootstrap is to the choice of l.

There is some discussion of optimal block lengths in the papers by Carlstein (1986), Künsch (1989) and more detailed discussion in Hall and Horowitz (1993). The rules are suggestive but putting some numbers in them we get a rough idea of block sizes to consider.

Carlstein is interested in minimizing the *MSE* of the block bootstrap estimate of the variance of a general statistic $t(x_1, x_2, \cdots, x_n)$ (e.g. a trimmed mean or a robust estimate of scale). He argues that as the block size increases the bias goes down but the variance goes up. Also, as the dependency among the x_i gets stronger, we need a longer block size. Based on these considerations, he derives the optimal block size l^* for the $AR(1)$ model $x_t = \phi x_{t-1} + \varepsilon_t$. His rule is:

$$l^* = (2\phi/(1-\phi^2))^{2/3} n^{1/3}.$$

Note that as ϕ increases l^* increases. For $n = 200$ and $\phi = .5, .8$ and $.9$ we get $l^* = 7.08, 15.81$ and 26.18 respectively. In practice these numbers are rounded, e.g. for $\phi = .9$ we might consider 8 blocks of length 25.

As for Künsch, it is commonly believed that he suggested the optimal number of blocks to be proportional to $n^{1/3}$, i.e. $l^* \propto n^{2/3}$ (see p. 1226). This implies much longer blocks. But he also suggested to use "subjective judgement based on sample correlations". This is perhaps a better rule than the other widely quoted one.

Hall and Horowitz (1993) derive rules taking into account the *MSE* in the estimation of variance as in Carlstein. They argue that the rules are similar both for the moving block scheme (they call this Künsch's rule) and the non-overlapping block scheme (they call this Carlstein's rule). Further more they say the rule is the same for *MSE* of the estimate of bias. Their rule is

$$l = (3/2)^{1/3} n^{1/3} \rho^{-2/3} \quad \text{under Künsch's rule}$$

$$l = n^{1/3} \rho^{-2/3} \quad \text{under Carlstein's rule}$$

where $\rho = \{\gamma(0) + 2 \sum_{j=1}^{\infty} \gamma(j)\}\{|\sum_{j=1}^{\infty} j\gamma(j)|\}^{-1}$ and $\gamma(j)$ equals the covariance of x_t at lag j. For the $AR(1)$ process $x_t = \phi x_{t-1} + \varepsilon_t$, $\rho = (1-\phi^2)/\phi$. For the $MA(1)$ process $x_t = \varepsilon_t + \theta\varepsilon_{t-1}$, $\rho = (1+\theta)^2/\theta$.

We have computed l^* for the two processes. They are for $n = 200$.

	AR(1)			MA(1)	
ϕ	K	C	θ	K	C
0.5	5.11	4.46	0.5	2.46	2.15
0.8	11.40	9.96	0.8	2.63	2.30
0.9	18.88	16.49	0.9	2.65	2.32

Note: K = Künsch's rule, C = Carlstein's rule.

5 Forecasting Problems

The traditional method of forecasting in time series models is based on the Box-Jenkins methods for *ARMA* or *ARIMA* models. The asymptotic standard errors in this method are derived on the assumption of normality of the errors and in addition the sampling variability of the estimated coefficients is not taken account.

An early approach to this problem of sampling variability of the coefficients is the stochastic simulation approach in Fair (1980, 1984). This approach is similar in some respects to the bootstrap approach. Fair's approach accounts for four main sources of uncertainty: (i) the error term, (ii) the coefficient estimates, (iii) the exogenous variables and (iv) the possible misspecification of the model.

Peters and Freedman (1985) suggest a bootstrap method for estimating the standard errors (we shall discuss the limitation of this later) of multi-step ahead forecasts and to choose among competing specifications. One major difference between Fair's approach and the bootstrap approach is that Fair assumes a normal distribution for the errors and a limiting normal distribution for the parameter estimates. The bootstrap approach is, however, distribution-free. One the other hand, tests for model misspecification in the Fair approach are absent in the bootstrap approach of Peters and Freedman. To focus on the differences, we shall discuss briefly both approaches.

5.1 The Stochastic Simulation and Bootstrap Approaches

To focus on the issues, consider the $AR(p)$ equation

$$y_t = \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2). \tag{8}$$

Given data on $(y_{1-p}, \cdots, y_0, y_1, \cdots, y_n)$, equation (8) can be estimated by OLS to get estimates of $\alpha = (\alpha_1, \cdots, \alpha_p)$ and σ^2. Call these $a = (a_1, \cdots, a_p)$ and s^2. Denote by e_t the estimated residuals.

The forecasting equation is

$$\widehat{y}_t = a_1 y_{t-1} + \cdots + a_p y_{t-p}. \tag{9}$$

This equation can be used to generate multi-step forecasts $\widehat{y}_{n+1}, \widehat{y}_{n+2}, \cdots, \widehat{y}_{n+h}$. We just use the observed values and where observed values are not available we use the forecast values, e.g.

$$\widehat{y}_{n+2} = a_1 \widehat{y}_{n+1} + \cdots + a_p y_{n-p+2}.$$

So far there is no difference between the Fair approach and the bootstrap approach.

Now, in the Fair approach we assume that a has a multivariate normal distribution $N(\alpha, \Sigma)$ which can be estimated as $N(a, S)$. Draw an observation e^* from the distribution $N(0, s^2)$ and an observation a^* from the distribution

$N(a, S)$. (We shall denote the simulated values by $*$.) For each of the forecast periods, the forecast is obtained by using a^* in place of a in (9) and adding e^* to it (error). We do this say N times. We thus have N value of y^*_{n+h}. Let the mean of these values be μ^*_{n+h} and variance σ^{*2}_{n+h}. Then an estimate of the forecast variance of y^*_{n+h} is σ^{*2}_{n+h} and the forecast error is $y^*_{n+h} - \mu^*_{n+h}$. We can also use the square of this forecast error as an estimate of the forecast variance. Call this $\widetilde{\sigma}^2_{n+h}$. Denote the difference between the two as d_{n+h}

$$d_{n+h} = \widetilde{\sigma}^2_{n+h} - \sigma^{*2}_{n+h}.$$

Fair argues that if the model is correctly specified, $E(d_{n+h}) = 0$. Otherwise, in general, $E(d_{n+h})$ is likely to be positive (see Fair, 1984), although this is not always true (see footnote p. 361, Fair, 1980). Note that in Fair's approach, although he calculates the mean and variance of the forecasts, he has actually the distribution of the forecasts from which he can construct confidence intervals (using the percentile method).

The difference between the stochastic simulation approach and the bootstrap approach consists not in just the normality assumption but in what exactly is resampled. In the bootstrap approach of Peters and Freedman (1985), again (5.1) is estimated, but the actual errors e_t are resampled to get the bootstrap sample e^*_t. Now using the initial values $y_{1-p}, \cdots, y_0$ and $(a_1, \cdots, a_p)$ as given, a bootstrap sample of past and future values $y^*_1, \cdots, y^*_n, y^*_{n+1}, \cdots, y^*_{n+h}$ is generated by a recursive method. Now model (8) is re-fitted to the pseudo data $y^*_1, \cdots, y^*_n$ and used to forecast the pseudo data of the future $y^*_{n+1}, \cdots, y^*_{n+h}$. In this artificial world the errors of forecast are directly observable. The distribution of these errors is used to approximate the distribution of the unobservable errors in the real forecasts.

Peters and Freedman also suggest how the bootstrap forecasting approach can be used to select between two models. (They illustrate this with an empirical example.) The procedure with two models is as follows:

(a) use model 1 to generate pseudo data and pseudo future. But fit and forecast with model 2. Compute the pseudo error of forecasts.

(b) now interchange the models. If for instance model 1 forecasts well both on its own assumptions and on those of model 2, then it is preferred. This procedure is similar to the forecast encompassing approach (suggested by David Hendry). For a recent example and references, see Ericsson and Marquez (1993) where dynamic non-linear models are used for multi-step forecasts.

It should be noted that Fair's approach of checking the model for misspecification using forecast errors can also be used with the bootstrap distribution of forecast errors.

5.2 Modifications of the Simple Bootstrap Prediction Approach

There are several problems that have been pointed out with the Peters-Freedman approach. We shall mention these problems and skip the details which can be found in the papers cited.

First, we talked of the merits of studentization with regard to confidence intervals and tests of hypotheses (Section 2). Why not use it in forecasting as well? This has been discussed in Kabaila (1993, pp. 477-9) who argues that studentization can be applied to predictive inference also with beneficial results, and presents some simulation results.

The next issue is the way the bootstrap sample is generated. The forecast for y_{n+h}, it should be noted, does not depend on $y_1, y_2, \cdots, y_n$ but on $y_{n-p+1}, \cdots, y_n$ only (that is the last p observations). If this is the case, the way the bootstrap sample y_t^* is generated should take account of this. Note that Peters and Freedman conditioned the bootstrap sample y_t^* on $y_{1-p}, \cdots, y_0$. Instead it should be conditioned on the last p observed y_t. Using these as given one generates y_t^* by the backward recursion corresponding to (8). This point has been made by Findley (1986), Stine (1987), and Thombs and Schucany (1990). The backward recursions are, however, not straightforward for non-Gaussian processes.

The method of generating the bootstrap sample conditional on the initial p observations (also used by Efron and Tibshirani, 1986) has been found to work well when getting the distribution of the least squares estimates of the coefficients α_i in (8). However, when it comes to prediction problems, the conditioning should be on the last p observations. The problem of bootstrap prediction inference for a non-Gaussian autoregressive process of order p conditional on the last p observations is discussed in Kabaila (1993). It is also discussed in Thombs and Schucany (1990) but Kabaila claims that their solution is ad hoc.

Another issue raised by Findley (1986) is that the potential contribution of bootstrap methods in prediction problems is limited. His argument is that the theoretical mean square forecast error from an estimated model is the sum of two components: (i) the mean square error of the optimal predictor and (ii) mean square difference between the optimal forecast and the estimated forecast. (ii) is of order $1/n$ where n is the sample size. (i) is easily obtainable without bootstrap. What the bootstrap does is to get a small sample estimate of (i) +(ii). But since (ii) is negligible, Findley argues the potential for the bootstrap method is limited.

Findlay's arguments should not deter us from using the bootstrap method for prediction problems. The big advantage of the bootstrap in the prediction context is in the case of prediction intervals given non-normal errors. In the case of confidence intervals one would expect the standard formulas to work well, in large samples, because of the central limit theorem. But for prediction intervals the forecast period disturbance also enters the calculation and if this disturbance is non-normal, inference based on the usual asymptotic theory could work very badly in comparison with the bootstrap non-parametric approach. In practice, many studies have found substantial difference between the prediction intervals obtained by using asymptotic theory and bootstrap methods.

On a theoretical level, there are results on the coverage probabilities of confidence intervals and prediction intervals based on bootstrap methods. The differences are due to the above-mentioned problem of the forecast period disturbance term. Bai and Olshen (1988) argue that whereas bootstrap-t confidence intervals are correct to $O(n^{-1})$, the prediction intervals are only correct to $O(n^{-1/2})$. We also quoted Hall (1988a) in Section 2 regarding bootstrap-t intervals for prediction from a regression model being correct to $O(n^{-1})$ whereas bootstrap-t confidence intervals are correct to $O(n^{-2})$.

5.3 Some Other Examples of Bootstrap Prediction Intervals

There are several applications of bootstrap procedures in predictive inference. Here we shall comment on a few studies. In all these studies, note that the confidence intervals are of the percentile type.

Veall (1987) and Bernard and Veall (1987) use bootstrap methods to get forecast intervals of peak electricity demand. The forecast intervals are wider than those that do not take account of parameter uncertainty and uncertainty about the exogenous variables. In Bernard and Veall (1987), the bootstrap method is extended to cover the forecasting equations for the exogenous variables as well. It is necessary to take uncertainty about the exogenous variables into account. Freedman and Peters did not address this problem.

In Masarotto (1990), the main issue discussed is determining the order p in the $AR(p)$ process using criteria like the AIC and Hannan-Quinn criterion. But several questions of how the bootstrap approach can be used simultaneously for model selection and prediction are left unanswered. These need to be investigated, but this is beyond the scope of this paper.

Runkle (1987) uses the bootstrap approach to get confidence intervals for variance decompositions in VAR systems. The application of the bootstrap approach is straightforward but it does not lead to any new conclusions. The major problem is that both point and interval estimates of variance decompositions in this model change markedly with minor changes in the model's specification. Thus, there is not much scope for the use of the bootstrap approach in this case.

6 Bootstrapping Unit Root Tests

The literature on unit root testing is enormous and it is not possible for us to consider all the issues in bootstrapping the different unit root test procedures in this paper. (This is being done in another paper.) The issues relate, among other things, to devising more powerful tests than the standard Dickey-Fuller test and its extensions, considering stationarity as null vs. unit root as null, and using the Bhargava (1986) type structural approach instead of the Dickey-Fuller type reduced form approach. The conflict between the structural vs. reduced form approaches has also been noted in the Bayesian approach to unit root testing and Peter Phillips argues in favor of the Bhargava approach. For an exposition of Phillips' arguments and some recomputations, see Zivot (1993).

In spite of its defects, we shall follow the Dickey-Fuller reduced form approach in what follows. This is because we wish to concentrate on the major issues pointed out in Section 2 relating to the type of test statistics to use and the type of sampling scheme to be used. There is some confusion in the econometric literature that needs to be cleared up. For instance, Basawa et al. (1991a) prove that sampling scheme S_1 is not appropriate in the unit root case. Basawa et al. (1991b) use test statistic $n(\widehat{\beta}^* - 1)$ with sampling scheme S_3. Ferretti and Romo (1994) establish the result that test statistic $n(\widehat{\beta}^* - 1)$ with sampling scheme S_2 can also be used in bootstrap tests of unit roots. Note that this is similar to the procedure used by Rayner (1990) for the stationary $AR(1)$ model, except that it is not in the pivotal form. We have compared both the Basawa et al. (1991b) and Ferretti and Romo schemes of getting bootstrap samples and did not notice any difference, although more detailed investigation of this is under way.

Consider an $AR(1)$ unit root model:

$$y_t = \beta y_{t-1} + \varepsilon_t, \quad \beta = 1, \quad t = 1, 2, \cdots, n \tag{10}$$

where $y_0 = 0$, $\varepsilon_t \sim iid(0, \sigma^2)$. In this case, the OLS estimator $\widehat{\beta}$ of β is a function of the standard Wiener process $W(r)$ and has a nonnormal limiting distribution

$$\left(\sum_{t=1}^{n} y_{t-1}^2\right)^{1/2} (\widehat{\beta} - 1) \xrightarrow{d} \frac{\sigma}{2}\left((W(1))^2 - 1\right)\left(\int_0^1 (W(r))^2 dr\right)^{-1/2}. \tag{11}$$

Thus, conventional tests based on normal asymptotic theories are not valid. The associated tests in this case are the Dickey-Fuller coefficient test $\widehat{\rho}$ and t-test $\widehat{\tau}$, which have the following limiting distributions

$$\widehat{\rho} = n(\widehat{\beta} - 1) \Rightarrow \frac{1}{2}\left((W(1))^2 - 1\right)\left(\int_0^1 (W(r))^2 dr\right)^{-1} \tag{12}$$

and

$$\widehat{\tau} \Rightarrow \frac{1}{2}\left((W(1))^2 - 1\right)\left(\int_0^1 (W(r))^2 dr\right)^{-1/2} \tag{13}$$

respectively, where $\widehat{\tau} = (\widehat{\beta} - 1)/SE(\widehat{\beta})$. See Dickey and Fuller (1979). The critical values of these tests have been tabulated by simulation methods by Dickey and reported in Fuller (1976).

To construct the bootstrap test corresponding to (12), we start with the OLS estimation of (10), compute $\widehat{\varepsilon}_t$ and get the bootstrap sample ε_t^*. We now generate y_t^* using ε_t^* and $\widehat{\beta}$ by a recursive procedure. What Basawa et al. (1991a) show is that the limiting distribution of $\left(\sum_{t=1}^{n} y_{t-1}^{*2}\right)^{1/2} (\widehat{\beta}^* - \widehat{\beta})$ is not (12). The limiting distribution of $n(\widehat{\beta}^* - \widehat{\beta})$ turns out to be random and does not coincide with (12) even if the error distribution is assumed to be normal.

However, for the sampling scheme S_3, which is resampling the restricted residuals $(y_t - y_{t-1})$ after centering, Basawa et al. (1991b) show the following which we shall state without proof.

Theorem 1 Along almost all sample paths $\{y_1, y_2, \cdots\}$ we have, under $H_0 : \beta = 1$,

$$n(\widehat{\beta}^* - 1) \xrightarrow{d^*} \frac{1}{2}\left((W(1))^2 - 1\right)\left(\int_0^1 (W(r))^2 dr\right)^{-1}, \quad \text{as } n \to \infty, \tag{14}$$

where $\xrightarrow{d^*}$ denotes the distributional convergence, conditionally on $(y_1, y_2, \cdots)$.

Corollary 1 Along almost all sample paths $\{y_1, y_2, \cdots\}$,

$$\left(\sum_{t=1}^{n} y_{t-1}^{*2}\right)^{1/2} (\widehat{\beta}^* - 1) \xrightarrow{d^*} \frac{\sigma}{2}\left((W(1))^2 - 1\right)\left(\int_0^1 (W(r))^2 dr\right)^{-1/2}. \tag{15}$$

Note that Corollary 1 is easily adapted to account for the asymptotic distribution of the bootstrap-t test statistic.

As an alternative, consider the resampling scheme S_2, that is the unrestricted OLS regression errors are used to generate ε_t^* and the pseudo data y_t^* are generated using the null $H_0 : \beta = 1$. The asymptotic validity of this bootstrap method is established in Ferretti and Romo (1994). We restate their result below in Theorem 2.

Theorem 2 Under the null, we have that

$$\left(\sum_{t=1}^{n} y_{t-1}^{*2}\right)^{1/2} (\widehat{\beta}^{*}-1) \xrightarrow{a.s.} \frac{\sigma}{2}\left((W(1))^2-1\right)\left(\int_0^1 (W(r))^2 dr\right)^{-1/2}, \quad n \to \infty \tag{16}$$

for almost all samples $\{y_1, y_2, \cdots\}$.

It is important to note two points. First Basawa et al. (1991a) do not prove the invalidity of sampling scheme S_2. Thus, the use of S_2 is not in conflict with their results. Second, both sampling schemes S_2 and S_3 are valid in the unit root case provided the test statistic is of the form T_2, that is use a suitably normalized function of $(\widehat{\beta}^* - \beta_0)$.

Turning next to the explosive case $|\beta| > 1$, Basawa et al. (1989) also show that if the ε_t are IN, conditional on $(y_1, y_2, \cdots, y_n)$

$$T_n = \widehat{\sigma}_n^{-1}\left(\sum_{t=1}^{n} y_{t-1}^{*2}\right)^{1/2} (\widehat{\beta}^* - \widehat{\beta}) \xrightarrow{d} N(0,1)$$

for all sample paths, where

$$\widehat{\sigma}_n^2 = n^{-1}\sum_{t=1}^{n}(y_t - \widehat{\beta} y_{t-1})^2.$$

One thing to notice is that it is $\widehat{\sigma}_n$ that is used here but not $\widehat{\sigma}_n^*$. We have not investigated the issue of why this is valid in the explosive case except to note that this is in conflict with the Hall and Wilson guidelines discussed in Section 2.

Nankervis and Savin (1994) present simulation results on bootstrap tests for unit roots in the trend model

$$y_t = \gamma + \delta t + \beta y_{t-1} + \varepsilon_t$$

under different error distributions. They use sampling scheme S_3 and test statistic T_2. In general the conclusions regarding sample sizes and empirical significance levels for the unit root case are the same as for the trend stationary model discussed in Section 3. The bootstrap-t test had, in general, the same power as the Dickey-Fuller test, although for some non-normal distributions, its performance was slightly better than that of the Dickey-Fuller test.

Earlier, we stated that we did not find much difference in the form of the bootstrap-t tests under sampling schemes S_2 and S_3. This was the case for normally distributed errors and for the three models:

$$\begin{aligned}
\text{(i) } y_t &= \beta y_{t-1} + \varepsilon_t, \quad \beta = 1 \\
\text{(ii) } y_t &= \alpha + \beta y_{t-1} + \varepsilon_t, \quad \alpha = 0, \quad \beta = 1 \\
\text{(iii) } y_t &= \alpha + \delta t + \beta y_{t-1} + \varepsilon_t, \quad \delta = 0, \quad \beta = 1.
\end{aligned}$$

In all cases the powers for the Dickey-Fuller test as well as two bootstrap-t tests using sampling schemes S_2 and S_3 were about the same and in all cases the powers deteriorated going from model (i) to (ii) and from (ii) to (iii). The models considered here (as well as the models considered in Nankervis and Savin, 1993) are too restrictive to be useful in practice because they are all based on the assumption that ε_t are *IID*. Extensions of the bootstrap approach to cases where ε_t are AR processes (the ADF test) and other unit root tests are under investigation. Schwert (1987) analyzed many U.S. macroeconomic time series and found that although they appeared to be nonstationary, they contained a significant $MA(1)$ coefficient in their $ARIMA$ specification. In view of this we do not want to make any definite recommendations on the merits of sampling schemes S_2 and S_3.

One final point concerns the use of pivotal statistics. There are two Dickey-Fuller tests: the coefficient test (12) and the t-test (13). When it comes to the bootstrap approach there is again the question of whether to consider the coefficient test or the t-test. We have looked into this issue and found the t-test to be only marginally better, although this conclusion is highly tentative. The case for considering pivotal statistics may not be as strong for the unit root model (as in the stationary models).

7 Issues in Bootstrapping Cointegrating Regressions

To focus on the issues concerning the bootstrapping of cointegrating regressions consider the simple system consisting of two variables y_1 and y_2 which are both $I(1)$ and the cointegrating regression

$$y_{1t} = \beta y_{2t} + u_t, \quad t = 1, 2, \cdots, n \tag{17}$$

where u is $I(0)$. As is well known the least squares estimator $\hat{\beta}$ of β is super-consistent (it converges to the true value at rate n rather than $\sqrt{n}$) but the asymptotic distribution of $\hat{\beta}$ involves nuisance parameters arising from

(i) endogeneity: correlation between y_2 and u

(ii) serial correlation in u.

Several corrections for the problems of endogeneity and serial correlation have been proposed in the literature. See e.g. Johansen (1988, 1991), Phillips and Hansen (1990), Phillips (1991), Saikkonnen (1991), Park (1992) and so on. However, all these methods involve estimation of equations different from equation (17) and enable the derivation of asymptotically pivotal statistics that are needed for the proper application of bootstrap methods. This was what was done in Li and Maddala (1993). Note that estimation of (17) does not lead to asymptotically pivotal statistics, except in the special case where the problem of endogeneity and serial correlation are absent.

7.1 Generation of Bootstrap Data for cointegrated Regressions

Suppose that there is no endogeneity problem nor the serial correlation problem. Then the OLS estimator $\widehat{\beta}$ of β in (17) has an asymptotic distribution that is nuisance parameter free and one can apply the bootstrap procedure. Even then it is better to bootstrap the pivotal t-statistic rather than $\widehat{\beta}$ itself to get confidence intervals for β. The bootstrap-t confidence intervals are more accurate, as discussed in Section 2, than those based on the bootstrap distribution of $\widehat{\beta}$.

Note that the bootstrap methods applicable to regression models and as used in Vinod and McCullouch (1995) cannot be used here. Nor is bootstrapping (y_1, y_2) directly, as suggested by Freedman (1981) for the stochastic regression model, and by Efron and Gong (1983), a valid procedure for the reasons following. The important thing to note is that y_2 is not just stochastic. It is also $I(1)$. Also (17) is a cointegration relationship. Hence, what we do is get $\widehat{u}_t$ by estimating (17) by OLS. Note that $\widehat{\beta}$ is super-consistent. We also get another set of residuals $\widehat{v}_t = \Delta y_{2t}$. After centering these residuals, we bootstrap the pairs $(\widehat{u}_t, \widehat{v}_t)$. Now we construct y_{2t}^* using the recursive method and y_{1t}^* using $\widehat{\beta}, u_t^*$, and y_{2t}^* in (17). This method uses the information that y_2 is $I(1)$ and that (17) is a cointegrating relationship. Note also that bootstrapping $\widehat{u}_t$ as in a regression model is not enough. This is the correct residual based procedure of constructing the bootstrap sample in this case. Directly bootstrapping the data (y_1, y_2) does not use the information that y_2

is $I(1)$ and (17) is a cointegration relationship. Note that this problem does not arise in the censored regression model considered by Efron (1981) or the examples considered in Efron and Gong (1983).

What if there is no endogeneity but there is serial correlation in the errors u in equation (17)? In this case as Phillips and Park (1988) showed, the OLS and GLS estimators are asymptotically equivalent. However, to obtain the valid t-statistic, which we need for the purpose of bootstrap-t confidence intervals, we need to calculate the asymptotic variance of $\widehat{\beta}$. This is described in Phillips and Park (1988) and we need not repeat it here. As for bootstrap data generation when the structure of autocorrelation is not known, we bootstrap blocks of $(\widehat{u}_t, \widehat{v}_t)$ as discussed in Li and Maddala (1993). If the errors are assumed to be $AR(1)$ process, we use recursive methods as in Li (1994). All this is not valid in the presence of endogeneity.

To focus on these issues consider the estimation of the consumption function viewed as a two-variable problem. Suppose we are interested in estimating the consumption function

$$c_t = \alpha + \beta y_t + u_t \tag{18}$$

and we are interested in estimating the (long-run) marginal propensity to consume as in Vinod and McCullouch (1995). The inferential procedure will depend on whether we regard c_t and y_t as stationary or unit root processes (around a trend). In either case, as pointed out by Haavelmo, more than 50 years ago, there is the problem of simultaneity bias in the OLS estimation of β. If c_t and y_t are $I(1)$ and (18) is treated as a cointegration relationship, then the OLS estimator $\widehat{\beta}$ of β does not involve an asymptotic bias, because of the super-consistency property of $\widehat{\beta}$. However, as discovered by Entorf (1992) in his simulations, the Haavelmo bias does not disappear except for very large samples. Super-consistency is based on the fact that the noise term in the cointegration relation (18) vanishes with respect to the signal as $n \to \infty$. If we assume a constant noise-signal ratio, then this leads to a persistent or even constant simultaneous equation bias. Irrespective of the asymptotic bias argument, the asymptotic distribution of $\widehat{\beta}$ involves parameters arising from endogeneity. If one ignores the endogeneity and serial correlation problems as done in Vinod and McCullouch, a pivotal statistic is readily available and one can simply use the bootstrap-t method. One does not need the complicated double bootstrap of Beran (1988). Also, as pointed out earlier, bootstrapping just $\widehat{u}_t$ is not correct in the cointegration model.

7.2 Results of Some Monte Carlo Studies

The Monte Carlo study by Li (1994) involved a model with no endogeneity. The serial correlation was of the $AR(1)$ type so that $u_t = \rho u_{t-1} + \varepsilon_t$. In this case after obtaining $\widehat{\varepsilon}_t$ by the usual least squares procedures, he resamples $(\widehat{\varepsilon}_t, \widehat{v}_t)$ jointly to get bootstrap samples (ε_t^*, v_t^*). Now recursive methods are used to get u_t^* and generation of y_{1t}^* and y_{2t}^* proceeds as described earlier. Li finds that the percentile-t method works well (in terms of giving true coverage) but the percentile method performs poorly. Note that the data generation uses $\widehat{\beta}$ but not the value β_0 specified by the null hypothesis. The t-statistic used in the paper is not stated but presumably it was based on $(\beta^* - \widehat{\beta})/\sigma^*$. In the study by Li and Maddala (1993) the data generation was based on β_0 and the t-statistic based on $(\beta^* - \beta_0)/\sigma^*$. Thus it is sampling scheme S_3 and the t-statistic T_2. In Li (1994) the sampling scheme is S_1 and t-statistic T_1. Earlier (Section 6) we said that this was not valid for a unit root model. However, for the cointegration model it appears that it is a valid procedure. The coefficient β is a regression coefficient. The only difference is that y_1 and y_2 are $I(1)$ variables.

The cointegration regression considered in Li (1994) is rather simple in that the regressor is strictly exogenous and the errors are $AR(1)$. In Li and Maddala (1993), various bootstrap methods, including the recursive bootstrap, the moving block bootstrap and the stationary bootstrap methods, are discussed in estimation and inference for cointegrating regression models in the presence of both endogeneity and serial correlation. They compared the performance of several popular asymptotically efficient estimators of cointegrating regressions, such as the fully modified OLS by Phillips and Hansen (1990), the canonical cointegrating regressions by Park (1992), and the vector error correction model by Johansen (1988), with the corresponding bootstrap methods in very general model specifications. The bootstrap methods are used for both bias correction and correction of size distortions. We shall describe one method of generating the bootstrap sample data. The moving block bootstrap procedure based on the fully modified OLS (FM) estimator for a cointegrating regression takes the following steps

1. Estimate (17) by FM, which gives $\widehat{\beta}_{FM}$. Calculate the associated t-statistic $\widehat{t}$ for testing the null $H_0 : \beta = \beta_0$.

2. Form moving block pairs $\{\widehat{u}_t, \cdots, \widehat{u}_{t+k-1}, \widehat{v}_t, \cdots, \widehat{v}_{t+k-1}\}$ from t to $t+k-1$ and $t = 1, \cdots, n-k+1$, where $\widehat{u}_t = y_{1t} - \beta_0 y_{2t}$ and $\widehat{v}_t = \Delta y_{2t}$.

3. Draw blocks $\{u_t^*, \cdots, u_{t+k-1}^*, v_t^*, \cdots, v_{t+k-1}^*\}$ randomly with replacement from the original moving blocks. Construct $y_{2t}^* = y_{2t-1}^* + v_t^*$ and $y_{1t}^* = \beta_0 y_{2t}^* + u_t^*$.

4. Estimate (17) using the bootstrap sample by FM. Calculate the associated t-statistic t_1^* for testing the null $H_0 : \beta = \beta_0$.

5. Repeat step 3 and step 4 B times. Obtain the bootstrap parameter distribution $\beta_1^*, \beta_2^*, \cdots, \beta_B^*$ and the distribution of the t-statistic $t_1^*, t_2^*, \cdots, t_B^*$.

6. The 2.5% quantiles (at the two ends) of $t_1^*, t_2^*, \cdots, t_B^*$, t_L^* and t_H^* are calculated. Reject the null if $\widehat{t} > t_H^*$ or $\widehat{t} < t_L^*$,where $\widehat{t}$ is the t-statistic from the original sample.

Since the pairs $\{\widehat{u}_t, \widehat{v}_t\}$ are resampled, by construction, the correlation between the innovations of the regressor and the regression errors are preserved. However, as remarked earlier in Section 4, the MBB does not replicate the dependence structure of the original data.

The overall evidence indicates that while the asymptotic procedures are associated with biases and serious size distortions, the bootstrap procedures are useful in reducing significantly the test size distortions. The only case where the bootstrap-t method did not provide any substantial improvement over the asymptotic procedures was in the Johansen method. This was because of the high variance of the variance of the Johansen ML estimate for estimator compared to that of other asymptotically efficient procedures. We did not compare the power of the different bootstrap based tests, nor did we investigate the behavior of the different procedures under non-normal distributions. These are under investigation and will be reported in another paper.

One important point is worth noting in this context. As pointed out by Hall and Horowitz (1994) and as discussed earlier in Section 4, since the moving block method does not replicate the dependence structure of the data, one needs to modify the test statistics themselves when applying them to the bootstrap samples. However, the development of bootstrap versions of test statistics for the cointegration model with the MBB scheme is extremely complicated. Meanwhile, it is important to note that even though the use of standard test statistics with bootstrap samples is not justified on theoretical grounds, the evidence in Li and Maddala (1993) shows that they provide substantial improvement over asymptotic results (particularly with the stationary bootstrap). Until more theoretical developments occur, it is best to use the procedures outlined there.

8 Conclusions

In this paper we have presented the main results on estimation of confidence intervals, tests of statistical hypotheses, and use of the moving block methods, that should be of interest to applied econometricians. The importance of using pivotal statistics and maintaining the appropriate correspondence between the sampling scheme and the test statistic used are emphasized. The pitfalls in the blind use of the direct bootstrap methods on the raw data instead of the properly defined residuals in the case of cointegrated relationships is pointed out.

It is our hope that the above mentioned aspects of the application of bootstrap methods in time series models, in particular the unit root and cointegration models, would be helpful to applied researchers in avoiding some pitfalls in the use of the bootstrap methods. Given the availability of several computer programs (bootstrap methods are available on SHAZAM and possibly several others) and the ease with which the methods can be applied, there is ample scope for misapplication of bootstrap methods. As noted in Jeong and Maddala (1993, p. 603) it is easy to jump on the computer and mechanically apply a certain bootstrap procedure when in fact the structure of the model suggests some other procedure for bootstrap data generation. It is also important to think about what statistic to bootstrap which depends on the particular problem and procedure for studentization. For this reason it is important to avoid some readily available canned programs.

References

Abadir, K.M. (1993). OLS bias in a non-stationary autoregression, *Econometric Theory*, 9, 81-93.

Abramovich, L. and Singh, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Annals of Statistics*, 13, 116-132.

Anderson, T.W., (1959). On asymptotic distribution of estimates of parameters of stochastic differences equations. *Ann. Math. Statist.*, 30, 676-687.

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

Babu, G.J. (1995). Bootstrap for non-standard cases. *Journal of Statisitcal Planning and Inference*, 43, 197-203.

Bai, C. and Olshen, R. (1988). Discussion of Hall's comparison of bootstrap confidence intervals. *Annals of Statistics*, 16, 953-956.

Barnard, G.A. (1963). Contribution to discussion. *Journal of the Royal Statistical Society*, B, 25, 294.

Basawa, I.V., Mallik, A.K., McCormick, W.P. and Taylor, R.L., (1989). Bootstrapping explosive autoregressive processes. *Annals of Statistics*, 17, 1479-1486.

Basawa, I.V., Mallik, A.K., McCormick, W.P. and Taylor, R.L., (1991a). Bootstrapping unstable first order autoregressive processes. *Annals of Statistics*, 19, 1098-1101.

Basawa, I.V., Mallik, A.K., McCormick, W.P. and Taylor, R.L., (1991b). Bootstrapping test of significance and sequential bootstrap estimation for unstable first order autoregressive processes. *Commun. Statist. -Theory Meth.*, 20, 1015-1026.

Bernard, J.T. and Veall, M.R. (1987). The probability distribution of future demand. *Journal of Business and Economic Statistics*, 5, 417-424.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika*, 74, 457-468.

Beran, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements. *J. Amer. Statist. Assoc.*, 83, 687-697.

Beran, R. (1990). Refining bootstrap simultaneous confidence sets. *J. Amer. Statist. Assoc.*, 85, 417-426.

Bhargava, A. (1986). On the theory of testing for unit roots in observed time series. *Review of Economic Studies*, 53, 369-384.

Bickel, P.J. and Freedman, D.A. (1980). On Edgeworth expansions and the bootstrap. Unpublished manuscript.

Bickel, P.J. and Freedman, D.A. (1983). Bootstrapping regression models with many parameters. In: A Festschrift for Erich Lehmann. Ed. by Bickel, P., Doksum, K. and Hodges, J.L. Belmont, California: Wadsworth, pp. 28-48.

Bose, A., (1988). Edgeworth correction by bootstrap in autoregressions. *Annals of Statistics*, 16, 1709-1722.

Bose, A., (1990). Bootstrap in moving average models. *Ann. Inst. Stat. Math.*, 42, 753-768.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Annals of Statistics*, 14, 1171-1179.

Chatterjee, S., (1986). Bootstrapping ARMA models: some simulations. *IEEE Transactions*, SMC-16, 294-299.

Davidson, J.E.H. (1981). Problems with the estimation of moving-average processes. *Journal of Econometrics*, 19, 295-310.

Davison, A.C. and Hall, P. (1993). On studentizing and blocking methods for implementing the bootstrap with dependent data. *Australian Journal of Statistics*, 35, 212-224.

DiCiccio, T. and Romano, J. (1988). A review of bootstrap confidence intervals (with discussion). *Journal of the Royal Statistical Society*, B, 50, 338-354.

Dickey, D.A. and Fuller, W.A. (1979). Distribution of the autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74, 427-431.

Dickey, D.A. and Fuller, W.A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49, 1057-1072.

Dufour, J.M. (1990). Exact tests and confidence sets in linear regressions with autocorrelated errors. *Econometrica*, 58, 475-494.

Efron, B., (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7, 1-26.

Efron, B. (1981). Censored data and the bootstrap. *Journal of the American Statistical Association*, 76, 312-319.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82, 171-200.

Efron, B. and Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross validation. *American Statistican*, 37, 36-48.

Efron, B. and Tibshirani, R., (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1, 54-77.

Engle, R. and Granger, C. W. J. (1987), Cointegration and error correction: representations, estimation and testing. *Econometrica*, 55, 252-276.

Entorf, H. (1992). Random Walks with drift, simultaneous equation errors, and small samples: simulating the birds-eye view. INSEE discussion paper, #9222.

Ericsson, N.R. and Marquez, J. (1993). Encompassing the forecasts of U.S. trade balance models. *Review of Economics and Statistics*, 75, 19-31.

Eriksson, B., (1983). On the construction of confidence limits for the regression coefficients when the residuals are dependent. *J. Statist. Comput. Simul.*, 17, 297-309.

Fair, R.C. (1980). Estimating the expected predictive accuracy of econometric models. *International Economic Review*, 21, 355-378.

Fair, R.C. (1984). Specification, Estimation and Analysis of Macroeconomic Models. Harvard University Press, Cambridge.

Ferretti, N. and Romo, J., (1994). Unit root bootstrap tests for AR(1) models. Working paper, Division of Economics, Universidad Carlos III de Madrid.

Findley, D.F., (1986). On bootstrap estimates of forecast mean square errors for autoregressive processes. Computer Science and Statistics: The Interface. Allen, D.M. (ed.), North-Holland: Elsevier Science Publishers B. V., 11-17.

Freedman, D.A., (1981). Bootstrapping regression models. *The Annals of Statistics*, 9, 1218-1228.

Freedman, D.A., (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. *The Annals of Statistics*, 12, 827-842.

Freedman, D.A. and Peters, S.C., (1984). Bootstrapping a regression equation: some empirical results. *Journal of the American Statistical Association*, 79, 97-106.

Fuller, W.A., (1976). Introduction to Statistical Time Series. Wiley, New York.

Fuller, W.A. (1985). Nonstationary autoregressive time series. In E.J. Hannan, P.R. Krishniah and M.M. Rao (eds). Handbook of Statistics, Vol. 5, (North Holland, pp. 1-23).

van Giersbergen, N.P.A. and Kiviet, J.F. (1993a). A Monte Carlo comparison of asymptotic and various bootstrap inference procedures in first-order dynamic models. Paper presented at the European Meetings of the Econometric Society, Uppsala, Aug. 1993.

van Giersbergen, N.P.A. and Kiviet, J.F. (1993b). How to implement bootstrap hypothesis testing in regression models. Paper presented at EC^2 Conference, Oxford, Dec. 1993.

Gregory, A.W. and M.R. Veall (1985). Formulating Wald tests of nonlinear restrictions. *Econometrica*, 53, 1465-1468.

Hall, P. (1988a). On symmetric bootstrap confidence intervals. *Journal of the Royal Statistical Society*, B, 50, 35-45.

Hall, P. (1988b). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.*, 16, 927-953.

Hall, P. (1989). Unusual properties of bootstrap confidence intervals in regression problems. *Probability Theory and Related Fields*, 81, 247-273.

Hall, P. (1992). The Bootstrap and Edgeworth Expansion. (Springer-Verlag, New York).

Hall, P and Horowitz, J.L. (1993). Corrections and blocking rules for the block bootstrap with dependent data. Working Paper #93-11, Department of Economics, University of Iowa.

Hall, P and Horowitz, J.L. (1994). Bootstrap critical values for tests based on generalized method of moments estimators. Forthcoming in *Econometrica*.

Hall, P. and S.R. Wilson (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics*, 47, 757-762.

Hartigan, J.A. (1969). Using subsample values as typical values. *Journal of the American Statistical Association*, 64, 1303-1317.

Hartigan, J.A. (1986). Comment on the paper by Efron and Tibshirani. *Statistical Science*, 1, 75-76.

Hinkley, D. V. (1988), Bootstrap methods (With Discussion). *Journal of the Royal Statistical Society*, B, 50, 321-337.

Hinkley, D.V. (1989). Bootstrap significance tests. Proceedings of the 47th Session of the International Statistical Institute, Vol. 3, pp. 65-74.

Horowitz, J.L. (1994). Bootstrap-based critical values for the information matrix test. *Journal of Econometrics*, 61, 395-411.

Horowitz, J.L. and N.E. Savin (1992). Non-invariance of the Wald test: The bootstrap to the rescue. Manuscript, Univ. of Iowa.

Jeong, J. and Maddala, G.S., (1993). A perspective on application of bootstrap methods in econometrics. Handbook of Statistics, Vol. 11, 573-610. (North Holland Publishing Co.)

Jeong, J. and Maddala, G.S., (1994). Testing the rationality of survey data using the weighted double-bootstrapped method of moments. Forthcoming *Review of Economics and Statistics.*

Johansen, S., (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12, 231-255.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregression models. *Econometrica*, 59, 1551-1580.

Kabaila, P. (1993). On bootstrap predictive inference for autoregressive processes. *Journal of Time Series Analysis*, 14, 473-484.

Kiviet, J.F., (1984). Bootstrap inference in lagged-dependent variable models. Working Paper, University of Amsterdam.

Kreiss, J.-P. and Franke, J., (1992). Bootstrapping stationary autoregressive moving-average models. *Journal of Time Series Analysis*, 13, 297-317.

Künsch, H.R., (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17, 1217-1241.

Kwok, B. and Veall, M.R., (1988). The Jackknife and regression with AR(1) errors. *Economics Letters*, 26, 247-252.

Lahiri, S.N., (1991). Second order optimality of stationary bootstrap. *Statistics & Probability Letters*, 11, 335-341.

Lahiri, S.N., (1992). Edgeworth correction by 'moving block' bootstrap for stationary and nonstationary data. Exploring the Limits of Bootstrap. LePage, R. and Billard, L. (eds), New York: John Wiley & Sons, Inc., 183-214.

Lahiri, S.N., (1993). On the moving block bootstrap under long range dependence. *Statistics & Probability Letters*, 18, 405-413.

Leger, C., Politis, D.N. and Romano, J.P., (1992). Bootstrap technology and applications. *Technometrics*, 34, 378-398.

Li, Y., (1994). Bootstrapping cointegrating regressions. *Economics Letters*, 32, 229-233.

Li, H. and Maddala, G.S., (1993). Bootstrapping cointegrating regressions. Paper presented at the Fourth Meeting of the European Conference Series in Quantitative Economics and Econometrics: , Dec. 16-18, 1993.

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

Liu, R.Y. and Singh, K., (1992). Moving blocks jackknife and bootstrap capture weak dependence. Exploring the Limits of Bootstrap. LePage, R. and Billard, L. (eds), New York: John Wiley & Sons, Inc., 225-248.

Masarotto, G., (1990). Bootstrap prediction intervals for autoregressions. *International Journal of Forecasting*, 6, 229-239.

Nankervis, J.C. and Savin, N.E., (1988). The Student's-t approximation in a stationary first order autoregressive model. *Econometrica*, 56, 119-145.

Nankervis, J.C. and Savin, N.E., (1994). The level and power of the bootstrap t-test in the AR(1) model with trend. Manuscript, Department of Economics, University of Surrey and University of Iowa.

Park, J.Y., (1992). Canonical cointegrating regressions. *Econometrica*, 60, 119-143.

Peters, S.C. and Freedman, D.A. (1985). Using the bootstrap to evaluate forecasting equations. *Journal of Forecasting*, 4, 251-262.

Phillips, P. C. B., (1991). Optimal inference in cointegrated systems. *Econometrica*, 59, 283-306.

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

Phillips, P. C. B. and Joon Y. Park (1988), Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors. *Journal of the American Statistical Association*, 83, 111-115.

Politis, D. N. and Romano, Joseph P. (1994), The stationary bootstrap. *Journal of American Statistical Association*, 89, 1303-1313.

Quenouille, M. H., (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society*, B, 11, 68-83.

Rayner, R.K., (1990). Bootstrapping p values and power in the first-order autoregression: a Monte Carlo investigation. *Journal of Business & Economic Statistics*, 8, 251-263.

Rayner, R.K., (1991). Resampling methods for tests in regression models with autocorrelated errors. *Economics Letters*, 36, 281-284.

Rilstone, P. (1993). Some improvements for bootstrapping regression estimators under first-order serial correlation. *Economics Letters*, 42, 335-339.

Rubin, H. (1950). Consistency of maximum likelihood estimates in the explosive case. In: Statistical Inference in Dynamic Economic Models (T. C. Koopmans, ed.) 356-364. Wiley, New York.

Runkle, D.E. (1987). Vector auto-regression and reality. *Journal of Business & Economic Statistics*, 5, 437-454.

Saikkonen, P., (1991). Asymptotically efficient estimation of cointegrating regressions. *Econometric Theory*, 7, 1-21.

Schwert, G.W. (1987). Effects of model specification on tests for unit roots in macroeconomic data. *Journal of Monetary Economics*, 20, 73-103.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Annals of Statistics*, 9, 1187-1195.

Thombs, L.A. and Schucany, W.R. (1990). Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association*, 85, 486-492.

Veall, M.R., (1986). Bootstrapping regression estimators under first-order serial correlation. *Economics Letters*, 21, 41-44.

Veall, M.R., (1987). Bootstrapping the probability distribution of peak electricity demand. *International Economic Review*, 28, 203-212.

Veall, M.R., (1988). Discussion of Hall's paper. *Annals of Statistics*, 16, 979-981.

Vinod, H.D. and McCullough, B.D., (1995). Estimating cointegration parameters: an application of the double bootstrap. *Journal of Statistical Planning and Inference*. 43, 147-156.

Wet, T. and Wyk, J.W.J., (1986). Bootstrap confidence intervals for regression coefficients when the residuals are dependent. *J. Statist. Comput. Simul.*, 23, 317-327.

Zivot, E. (1993). A Bayesian analysis of the unit root hypothesis within an unobserved components model. Discussion paper 93-15, University of Washington, Seattle.