

Recent Developments in Bootstrapping Time Series

Jeremy Berkowitz and Lutz Kilian*

September 25, 1996

Abstract

In recent years, several new parametric and nonparametric bootstrap methods have been proposed for time series data. Which of these methods should applied researchers use? We provide evidence that for many applications in time series econometrics parametric methods are more accurate, and we identify directions for future research on improving nonparametric methods. We explicitly address the important, but often neglected issue of model selection in bootstrapping. In particular, we emphasize the advantages of the AIC over other lag order selection criteria and the need to account for lag order uncertainty in resampling. We also show that the block size plays an important role in determining the success of the block bootstrap, and we propose a data-based block size selection procedure. Our discussion also highlights the importance of accounting for small-sample bias in autoregressions and some shortcomings of the standard percentile and percentile-t intervals in the time series context.

Key Words: Bootstrap, ARMA, Frequency Domain, Blocks

JEL Classification: C13, C22

* We would like to thank Frank Diebold, Robert Stine and Peter Christoffersen for helpful suggestions. Any remaining errors are ours. Correspondence to: Jeremy Berkowitz, Federal Reserve Board, Stop 61A, Washington, D.C. 20551, Telephone: (202) 736-5581, Email: m1jmb02@FRB.GOV.

1 Introduction

In recent years, many exciting developments have taken place in bootstrapping time series. Advances have proceeded along a number of distinct paths. Some authors have focused on adapting the familiar residual-based resampling approach of Efron (1979) to finite-order ARMA models. More recently, the focus has shifted toward residual based nonparametric methods such as the sieve bootstrap and the Cholesky factor bootstrap, which treat the underlying population model as unknown. Other researchers have explored resampling blocks of time series data. Still another approach has been to develop algorithms which operate in the frequency domain. The advantage of this approach is that in the frequency domain there are iid variables which can be exploited for bootstrapping even when the original data are non-iid.

Given this array of alternative bootstrap methods, which method should applied econometricians use? We observe that these algorithms differ in the extent to which they impose parametric structure on the data. Algorithms which make few parametric assumptions are relatively likely to encompass the true model. However, methods which condition on a particular parametric model afford higher precision. We provide evidence that suggests that for many applications in time series econometrics parametric methods may be preferable. Rather than rejecting nonparametric methods, we identify directions for future improvements.

An important, but often neglected issue in evaluating the performance of bootstrap algorithms is model selection. For lag order selection in autoregressive models, we stress the advantages of the AIC compared to more parsimonious criteria as well as the need to explicitly account for lag order uncertainty. Similar problems arise in nonparametric resampling. In particular, we show that the block size plays an important role in determining the success of the block bootstrap, and we propose an automatic data-based selection procedure. We also discuss the choice of bandwidth selection criteria for frequency domain bootstraps.

Another concern in applied work is the presence of nonstationarities. We argue

that the asymptotic bootstrap theory for nonstationary data is more fully developed for the parametric case. In particular, parametric resampling methods have recently been shown to be valid, even for processes with some explosive roots (Datta (1995)).

Nevertheless, care must be taken in applying parametric bootstrap algorithms. In particular, we emphasize the importance of accounting for small-sample bias in autoregressive models. We also stress that in the time series context percentile intervals may perform poorly and their use should be supported with Monte Carlo evidence. In addition, standard percentile intervals may also require modifications.

In contrast to the discussion in Léger, Politis and Romano (1992), Carlstein (1992), Jeong and Maddala (1993), Young (1994) and Horowitz (1995), our main focus is on inference in the linear stochastic regressor model. Our discussion complements recent work by Li and Maddala (1996). Section 2 reviews parametric and nonparametric residual-based bootstrap algorithms. Section 3 discusses the block bootstrap. Section 4 presents bootstrap algorithms for the frequency domain. Section 5 focuses on the treatment of nonstationary data. Section 6 contains a Monte Carlo study which compares these three approaches to bootstrapping. Section 7 concludes.

2 Residual Based Resampling

Efron's (1979) original bootstrap algorithm required resampling from data which are, in population, independent and identically distributed. In the iid case, one can create artificial repeated samples by random resampling with replacement from the data. If the data display heteroskedasticity or serial correlation, a randomly resampled set of data will not preserve these properties, so that statistics calculated from the resampled data (or from transformations of resampled data) will be inconsistent. Thus, the iid bootstrap fails for time-dependent data. One way to reduce time dependent data to an iid structure is to fit a parametric model.

2.1 Parametric Methods

Early applications of the bootstrap algorithm to time-dependent data assumed that the underlying process follows a stationary finite order autoregression of the form:

$$A(L)y_t = \varepsilon_t \tag{1}$$

where $\varepsilon_t \sim iid$ with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t^2) < \infty$. $Y = (y_1, \dots, y_T)'$ denotes the observed data. $A(L)$ is an invertible polynomial in the lag operator. For example, Efron and Tibshirani (1986) and De Wet and van Wyk (1986) bootstrapped the AR(1) model. Stine (1987) extended the analysis to the AR(p) model and Runkle (1987) to the finite order VAR(p) model. The AR(p) model may be bootstrapped as follows:

1. Determine the order of the AR(p) process.
2. Estimate the parameters $\hat{A}(L)$.
3. Generate bootstrap innovations ε_t^* by resampling with replacement from the empirical residuals $\hat{\varepsilon}_t = \hat{A}(L)y_t$.
4. Generate a random draw for the vector of p initial observations $Y_0^* = (y_1^*, \dots, y_p^*)'$.
5. Generate pseudo-data: $\hat{A}(L)y_t^* = \varepsilon_t^*$ conditional on Y_0^* .
6. Calculate the bootstrap parameter estimates: $\hat{A}^*(L)$.
7. Repeat steps 3-6 many times and build up the empirical distribution of interest.¹

Under some additional regularity conditions, Bose (1988) proves that the bootstrap approximation improves the asymptotic accuracy of the OLS estimates in the AR(p) model from $O(T^{-\frac{1}{2}})$ to $o(T^{-\frac{1}{2}})$ almost surely.

In practice, AR(p) models are almost always estimated by least-squares. If no intercept is included in the regression model, the residuals $\hat{\varepsilon}_t$ must be recentered prior to resampling to ensure that their bootstrap population mean is zero. It is also

¹ For a discussion on choosing the number of bootstrap replications, see Efron and Tibshirani (1993).

common to rescale the empirical residuals by a factor of $[T - p / (T - p - d)]^{1/2}$, where d denotes the number of estimated coefficients. The aim is to give the ε_t^* the desired variance (see Stine (1987), Peters and Freedman (1984)).

The p initial conditions $Y_0 = (y_1, \dots, y_p)'$ like all observations for y_t are stochastic. While the effect of conditioning on a particular set of initial conditions is asymptotically negligible, it is not appropriate to condition on Y_0 in order to generate the bootstrap replicates. One way to randomize Y_0^* is to set $Y_0^* = \hat{\Gamma}^{-1/2}(\hat{\Sigma}_\varepsilon^{-1/2}\varepsilon_t^*)$, where $\hat{\Gamma}$ is the estimate of $E(Y_t Y_t')$ defined by $\hat{A}(L)$ and $Y_t = (y_t, y_{t-1}, \dots, y_{t-p+1})'$ (Lütkepohl (1991), p. 496). This procedure will preserve the second moment structure in the data. The problem with this method is that it requires the estimated process to be stationary. For nonstationary coefficient estimates, the procedure breaks down because $\hat{\Gamma}$ is noninvertible. Even for borderline stationary processes, there is a positive probability that some least-squares estimates will be explosive. A method which does not require matrix inversion is to pick arbitrary values for Y_0^* in the recursion $\hat{A}(L) y_t^* = \varepsilon_t^*$ and to discard the start-up transients for $\{y_t^*\}$. Alternatively, one could build up the initial observations from the estimated moving average representation $Y_0^* = \hat{A}^{-1}(L)\varepsilon_t^*$ as in Rayner (1990), but this requires the truncation of an infinite sum. A third approach which avoids the truncation of an infinite sum and does not require start-up transients is to divide the observed data into $T - p + 1$ overlapping blocks of length p and randomly select one block with replacement for Y_0^* . This block initialization has been used for example in Stine (1987).

An alternative class of parametric models are stationary MA(q) models:

$$y_t = B(L)\varepsilon_t \tag{2}$$

where $B(L)$ denotes a lag polynomial and ε_t and y_t are defined as above. MA(q) models are rarely bootstrapped in econometric practice, but it is straightforward to adapt the bootstrap algorithm for AR(p) models to the present context. Simulation results for the finite order stationary MA(1) model can be found in De Wet and van Wyk

(1986) and Bose (1990). Under further regularity conditions, Bose (1990) proves that the bootstrap approximation of the parameter estimates in moving average models is accurate to the order $o(T^{-\frac{1}{2}})$. In contrast, the asymptotic normal approximation is accurate only to the order $O(T^{-\frac{1}{2}})$.

Chatterjee (1986) applies the bootstrap algorithm to general ARMA(p,q) models of the form:

$$A(L)y_t = B(L)\varepsilon_t \tag{3}$$

where ε_t and y_t are defined as above. $A(L)$, $B(L)$ are invertible polynomials in the lag operator, satisfying the assumption that together they imply $\text{var}(y_t) < \infty$. The ARMA bootstrap algorithm proceeds as follows:

1. Determine the order of the ARMA(p,q) process.
2. Estimate the parameters: $\hat{A}(L)$, $\hat{B}(L)$.
3. Resample from: $\hat{\varepsilon}_t = \hat{B}^{-1}(L)\hat{A}(L)y_t$ (after recentering the $\hat{\varepsilon}_t$ around zero).
4. Choose a large positive integer τ , set $y_t^* = 0$ for $t < -\tau$ and generate iid draws for ε_t^* for $t = -\tau, \dots, T$.
5. Generate pseudo-data: $y_t^* = \hat{A}^{-1}(L)\hat{B}(L)\varepsilon_t^*$ for $t = -\tau, \dots, T$ and retain the last T values of y_t^* .
6. Calculate the bootstrap parameter estimates: $\hat{A}^*(L)$, $\hat{B}^*(L)$.
7. Repeat steps 3-6 many times and build up the empirical distribution of interest.

Under regularity conditions, Kreiss and Franke (1989) prove the asymptotic validity of the bootstrap approximation for ML estimators in the finite-order stationary ARMA model.²

²The parametric bootstrap may be robustified against possible serial correlation in ε_t by re-sampling blocks of residuals using the block methods discussed in section 3 (e.g., Li and Maddala (1993)), or by explicitly modeling the error term (e.g., Lamoureux and Lastrapes (1990)).

2.1.1 Generating Bootstrap-Data in VAR Models

Superficially, the bootstrap algorithm for VAR models is similar to the familiar algorithm for the regression model with fixed regressors. However, in autoregressive models the OLS estimates of the slope coefficients are systematically biased away from their population values. As a result, the standard bootstrap algorithm used by Runkle (1987) may be misleading in small samples. The size of the bias depends on the sample size, the persistence of the data generating process and whether a deterministic time trend is included in the regression.

The tendency of the bootstrap to intensify the deficiencies of the OLS estimator was first observed by Kiviet (1984) in a linear regression model with lagged dependent variables. Pope (1987) and Nicholls and Pope (1988) suggest bias-correcting the slope coefficients "prior to bootstrapping" in order to improve the bootstrap approximation in the vector autoregressive model. They develop closed form expressions for the asymptotic first-order bias of the slope coefficients in the VAR model without a deterministic time trend. Recent work by Kilian (1995) implements and extends Nicholls and Pope's (1988) proposal. Kilian uses resampling to estimate the first-order coefficient bias in the VAR model with and without deterministic time trends.

2.1.2 Pitfalls in Constructing Confidence Intervals in VAR Models

Even if we bias-correct the autoregressive coefficients $\hat{A}(L)$ prior to resampling the data, small-sample bias may cause additional problems in bootstrap inference. This is because the bootstrap estimates $\hat{A}^*(L)$ themselves will be subject to stochastic regressor bias. This bootstrap bias tends to undermine the coverage accuracy of bootstrap confidence intervals for statistics that are functions of $\hat{A}(L)$, regardless of the type of confidence interval used.

There is a widespread perception (e.g, Horowitz (1995), Li and Maddala (1996)) that more accurate finite-sample confidence intervals can be obtained by bootstrap-

ping asymptotically pivotal statistics. A statistic is said to be asymptotically pivotal if its limiting distribution does not depend on any unknowns. Many statistics of interest based on AR(p) and ARMA(p,q) models are asymptotically normal and can be *studentized* to make them asymptotically pivotal. Consider a statistic $\hat{\theta}$. Percentile-t bootstrap intervals are based on the bootstrap approximation $(\hat{\theta}^* - \hat{\theta})/SD(\hat{\theta}^*)$ of the studentized statistic, $(\hat{\theta} - \theta)/SD(\hat{\theta})$.

Unfortunately, in many time series models reliable measures of scale, $SD(\hat{\theta}^*)$, do not exist. In particular, the distribution of the slope coefficients in the VAR model may undergo drastic changes as the dominant root of the process approaches unity. Bias in $\hat{A}^*(L)$ tends to move the estimate $\hat{\theta}^*(\hat{A}^*(L))$ away from its true value, changing its variance. As a result, the percentile-t bootstrap interval is unlikely to perform well without a suitable bias correction prior to estimating the variance. For example, Kilian (1995) reports that the percentile-t interval for VAR impulse response estimates tends to fail spectacularly in small samples. Similar results for the correlation coefficient are well known in the bootstrap literature (Efron (1987)). While the percentile-t interval promises higher order asymptotic accuracy, the asymptotic behavior may be a poor indicator of its accuracy in finite samples.

Note that in the context of hypothesis testing, this pitfall is effectively overcome by resampling under the null (e.g., Zivot and Andrews (1992)) and/or using restricted estimation techniques (e.g., Nankervis and Savin (1996)). Li and Maddala (1996) provide an extensive discussion.

Unfortunately, for interval estimation there is no specific null to refer to. In principle, it is possible to improve the small-sample performance of the percentile-t interval with variance stabilizing transformations, but these transformations are not generally known and have to be simulated. This adds another layer of bootstrapping and makes the percentile-t method computationally burdensome (Efron and Tibshirani (1993)). We conclude that the percentile-t method should not be used blindly without supporting Monte Carlo evidence of its small-sample properties.

Why not directly bootstrap the unstudentized statistic of interest? As with the percentile-t, the percentile interval requires the statistic to be unbiased and scale invariant. This assumption is asymptotically valid for many statistics based on VAR slope coefficients, but it is not reasonable for their small-sample distribution. As a result, the percentile interval can be expected to perform poorly. The bias-corrected (BC) percentile interval discussed in Efron and Tibshirani (1993) does not necessarily remedy this problem either, because it ignores scale effects. Intuitively, shifting the interval endpoints to account for median bias in a scalar distribution fails to account for the changes in the shape of the distribution across the parameter space. The more general BC_α percentile interval of Efron and Tibshirani (1993) is designed to account for bias *and* changes in the variance of the statistic of interest. However, its adaptation to time-dependent data has not been investigated. To date empirical evidence on the coverage accuracy of the percentile and BC intervals is scant. Kilian (1995) finds that these intervals perform erratically for VAR impulse response estimates. He also explores an alternative approach to removing bias and scale effects prior to bootstrapping.

2.1.3 Lag Order Uncertainty in Parametric Models

The bootstrap can only be expected to perform well when the parametric model provides a good approximation to the true model. Determining the correct orders of an ARMA or AR model is thus a crucial issue. Chatterjee (1986), for example, reports simulation results for ARMA(1,1), ARMA(2,0) and ARMA(0,2) models. He compares bootstrap and asymptotic estimates of standard errors. Chatterjee regards the bootstrap results as quite satisfactory, but observes that much of the attraction of this method depends on selecting the right order. He notes that the bootstrap performs poorly if the selected order is not correct.

Recent work by Kilian (1996a) offers some guidance on selecting the lag order. For the AR(p) model, the lag order selection criterion need not be consistent for

the lag order for the bootstrap algorithm to be asymptotically valid. However, it is necessary that the probability of underestimating the true lag order is asymptotically zero. Provided that the range of lag orders considered includes the true lag order, this suggests that a wide range of information-based lag order selection criteria including the Akaike Information Criterion (AIC) are potentially valid criteria.³

Kilian (1996a) also points out that the consequences of bootstrapping an over-parameterized VAR model may be very different from those of bootstrapping an under-parameterized model. This suggests that lag order selection criteria such as the Schwarz Information Criterion (SIC), which are known to be biased downward in small samples, will result in poor bootstrap estimates. Kilian's simulation results confirm that in small and moderate samples the coverage accuracy of bootstrap confidence intervals for VAR impulse response estimates is much closer to nominal coverage for the AIC than for more parsimonious criteria such as the SIC or the Hannan-Quinn Criterion.

Once it is explicitly recognized that the lag order must be estimated, another major difference between bootstrapping a fixed design model and a stochastic regressor model becomes apparent. The standard bootstrap algorithm for AR(p) models conditions on the lag order estimate as though it were the true lag order. Even if the lag order is estimated correctly, the standard algorithm ignores the sampling uncertainty about the lag order estimate and may lead to misleading inferences. Masarotto (1990) and Kilian (1996b) therefore propose a generalization of the bootstrap algorithm for VAR(p) models which reflects the true sampling uncertainty of the lag order estimate. This "endogenous lag order" bootstrap algorithm does not condition on the initial lag order estimate, but re-estimates the lag order in each bootstrap iteration. Extensions of this idea to ARMA(p,q) models are straightforward.

³The asymptotic validity of the AIC for bootstrapping follows from results in Paulsen and Tjøstheim (1985) and Quinn (1988). Also see Pötscher (1991, p. 179).

2.1.4 Conditional Bootstrap Prediction

Applied researchers are often interested in the distribution of forecasts conditional on the last p observations of the sample path. Bootstrapping the conditional distribution requires the last p observations in each bootstrap sample to be identical to the last p observations in the original data. The standard bootstrap algorithm for the AR(p) model is not appropriate for this purpose, because it does not constrain the values of the last p bootstrap observations. To solve this problem, Thombs and Schucany (1990) propose initializing the AR(p) with the last p observations and backcasting the time series using the 'backward representation':

$$A(L^{-1})y_t = w_t \tag{4}$$

where the backward noise w_t is the sequence defined by:

$$w_t = \frac{A(L^{-1})}{A(L)}\varepsilon_t \tag{5}$$

If ε_t is normally distributed, one can use iid resampling of the backward residuals, \hat{w}_t , based on (4) to generate conditional bootstrap sample paths. However, this algorithm is not valid for non-Gaussian innovations ε_t . Findley (1986) shows that in the non-Gaussian AR(1) model, w_t (though uncorrelated) is not iid. Breidt and Davis (1991) prove this result for the AR(p) model.

Breidt, Davis, and Dunsmuir (1992, 1995) observe that the distribution and the dependence structure of w_t are complicated, but that the sequence can be simulated if we rewrite 5 as:

$$A(L)w_t = A(L^{-1})\varepsilon_t, \tag{6}$$

where w_t is an ARMA(p,p) process driven by the iid sequence ε_t . They propose the following algorithm:

1. Determine the lag order of the AR(p) process.
2. Estimate the parameters $\hat{A}(L)$ for the observed data $\{y_t\}$, $t = 1, \dots, T$.
3. Compute $\hat{\varepsilon}_t = \hat{A}(L)y_t$ for $t = p + 1, \dots, T$.
4. Generate a bootstrap realization w_t^* of the backward noise w_t via:

$$\hat{A}(L)w_t^* = \hat{A}(L^{-1})\hat{\varepsilon}_t^*,$$

using standard bootstrap techniques for ARMA(p,q) models.

5. Generate a bootstrap realization $\{y_t^*\}$ of $\{y_t\}$ passing through the last p observations of the sample path via:

$$\begin{aligned} y_t^* &= y_t & t = T, T-1, \dots, T-p+1 \\ y_t^* &= \hat{A}(L^{-1})y_t^* + w_t^* & t = T-p, T-p-1, \dots, 1 \end{aligned} \quad (7)$$

using the sequence of observations for w_t^* from step 4.

6. Calculate the bootstrap estimates $\hat{A}^*(L)$.
7. Repeat steps 4-6 many times and build up the conditional empirical distribution of the h-step ahead forecasts \hat{y}_{T+h} .

McCullough (1994) finds that the conditional forecast distributions implied by the Thombs and Schucany (1990) procedure are very different from those implied by the Breidt et al. procedure.

Kabaila (1993) shows that the conditional sample paths generated by the Breidt et al. procedure are not entirely correct. He proposes an exact bootstrap procedure based on an estimate of the pdf of ε_t . However, as Breidt, Davis, and Dunsmuir (1995) point out, estimating the density in small samples may be problematic, and the gains from using exact rather than approximate conditional sample paths appear small.

2.1.5 Bootstrapping State Space Models

ARMA processes can also be cast in state-space form and consistently estimated using Kalman filter techniques (e.g., Harvey (1989)). Stoffer and Wall (1991) propose a bootstrap algorithm for linear state space models. They show that their algorithm delivers consistent bootstrap standard errors under some regularity conditions. Define the state-space model as:

$$\begin{aligned} s_{t+1} &= F s_t + G x_t + w_t \\ y_t &= H s_t + D x_t + v_t, \end{aligned} \tag{8}$$

where y_t is a $q \times 1$ vector of observed data, s_t is a $p \times 1$ unobserved state vector and x_t is an $r \times 1$ vector of exogenous variables. $F, G, H,$ and D are coefficient matrices. The innovations w_t and v_t are iid with zero mean and nonnegative definite covariance matrices. $E(v_t v_t') = R,$ $E(w_t w_t') = Q,$ and $E(w_t v_t') = 0.$ The model coefficients and correlation structure are assumed to be uniquely parameterized by the vector $\theta.$

The model (8) may, alternatively, be represented in innovations form. Let $s_{t+1|t}$ denote the best linear predictor of $s_{t+1}.$ Then the forecast errors are $\varepsilon_t = y_t - H s_{t|t-1} - D x_t$ with covariance matrix $\Sigma_t = H P_{t|t-1} H' + R.$ The innovations form is:

$$\begin{aligned} s_{t+1|t} &= F s_{t|t-1} + G x_t + F K_t \varepsilon_t \\ y_t &= H s_{t|t-1} + D x_t + \varepsilon_t, \end{aligned} \tag{9}$$

where $K_t = P_{t|t-1} H' \Sigma_t^{-1}$ (the Kalman gain) and where $P_{t|t-1}$ is the covariance matrix of $s_t - s_{t|t-1}.$

Let $\hat{\theta}$ denote the Gaussian ML estimate, and define the $(p+q) \times 1$ vector $\xi_t = [s'_{t+1|t}, y'_t]'$. Stacking the equations in (9) and evaluating at $\hat{\theta}$ results in:

$$\xi_t = \hat{A}\xi_{t-1} + \hat{B}x_t + \hat{C}_t\hat{\varepsilon}_t(\hat{\theta}) \quad (10)$$

where $A = \begin{bmatrix} \hat{F} & 0 \\ \hat{H} & 0 \end{bmatrix}$, $B = \begin{bmatrix} \hat{G} \\ \hat{D} \end{bmatrix}$, $C_t = \begin{bmatrix} \hat{F}\hat{K}_t\hat{\Sigma}_t^{1/2}(\hat{\theta}) \\ \hat{\Sigma}_t^{1/2}(\hat{\theta}) \end{bmatrix}$.

Stoffer and Wall (1991) devise the following bootstrap procedure:

1. Calculate $\hat{\theta} = \arg \max_{\theta} \left[-\sum_{t=1}^T \left\{ \log |\Sigma_t(\theta)| + \varepsilon_t(\theta)' \Sigma_t^{-1}(\theta) \varepsilon_t(\theta) \right\} \right]$. This implies a set of forecast errors, $\hat{\varepsilon}_t(\hat{\theta})$ and the forecast error covariance matrix $\hat{\Sigma}_t(\hat{\theta})$.
2. Generate e_t^* by sampling with replacement from the normalized residuals $\hat{e}_t = \hat{\Sigma}_t^{-1/2}(\hat{\theta})\hat{\varepsilon}_t(\hat{\theta})$.
3. Generate pseudo-data, y_t^* , by substituting e_t^* for \hat{e}_t in equation (10), holding fixed the exogenous variables x_t and the initial conditions.
4. Calculate bootstrap parameter estimates, $\hat{\theta}^*$.
5. Repeat steps 2-4 many times and build up the empirical distribution of interest.

2.2 Nonparametric Methods

The bootstrap algorithms discussed above assume that the true model is a finite order ARMA process with iid innovations. However, these models are at best viewed as approximations. A broader class of models are linear autoregressions of infinite order. If the true model is not finite-ordered, the asymptotic justification of the bootstrap approximation proposed by Bose (1988) and Kreiss and Franke (1992) is no longer valid. We will discuss two bootstrap algorithms designed for this class of processes: the sieve bootstrap and the Cholesky factor bootstrap.

2.2.1 The Sieve Bootstrap

Bühlmann (1995, 1996a) considers a class of linear, infinite dimensional process which can be approximated by a sequence of finite-dimensional autoregressive approximations of order $p(T)$ where $p(T) \rightarrow \infty$ and $p(T) = o(T)$ as $T \rightarrow \infty$. He argues that the standard OLS bootstrap for the AR(p) model may be given a nonparametric interpretation. In particular, he proposes estimating an AR($p(T)$) model using the AIC and generating a bootstrap sample by resampling the residuals of the fitted model. This so-called sieve bootstrap is model free within the class of linear MA(∞) processes with polynomial decay. Bühlmann (1996a) proves that the sieve bootstrap gives correct approximations to the distribution of smooth functions of linear statistics of the data. Under the more restrictive assumption of exponential decay, Paparoditis (1996) proved that this bootstrap procedure delivers an asymptotically valid bootstrap approximation for the autoregressive coefficients and for the moving average representation of the VAR(p) model. Similar procedures have also been proposed in Swanepoel and Van Wyk (1986), Paparoditis and Streitberg (1992), and Kreiss (1992). In related work, Bickel and Bühlmann (1996) propose a smoothed sieve bootstrap for nonlinear, nonregular statistics. Their proposal involves resampling from the smoothed distribution of the empirical residuals of the approximating autoregressive model. Bühlmann (1996b) studies the sieve bootstrap for autoregressive models including a deterministic time trend.

2.2.2 The Cholesky Factor Bootstrap

While the sieve bootstrap effectively reinterprets the familiar parametric AR model as a device for nonparametric estimation, Diebold, Ohanian and Berkowitz (1995) formulate a bootstrap algorithm which does not require conditioning on any particular parametric model of the VARMA type. The context is the vector covariance stationary MA(∞). Any finite realization of length T thus has representation:

$$Y = P\varepsilon, \quad (11)$$

where P is $rT \times rT$ and ε is $rT \times 1$. The nonparametric bootstrap proceeds as follows:

1. Consistently estimate $\text{Cov}(Y)=\Sigma$, by applying a suitable truncation lag rule.
2. Take the Cholesky decomposition: $\hat{P}\hat{P}' = \hat{\Sigma}$.
3. Resample from the normal distribution: $\varepsilon^* \sim N(0, \hat{\Sigma})$.
4. Generate pseudo-data: $y^* = \hat{P}\varepsilon^*$.
5. Calculate bootstrap statistics: $\theta(y^*)$.
6. Repeat steps 3-5 many times and build up the empirical distribution of interest.

Alternatively, in step 3 the residuals may be resampled without imposing Gaussianity by drawing from the empirical distribution of $\hat{\varepsilon} = \hat{P}^{-1}Y$, after rescaling the $\hat{\varepsilon}$ so that they have a variance of 1.

This "Cholesky factor" algorithm is a model-free method for generating pseudo-data focusing on the second moment properties of the observed data. Note that the ARMA(p,q) parametric bootstrap generates pseudo-data from:

$$y_t^* = \hat{A}^{-1}(L)\hat{B}(L)\varepsilon_t^*. \quad (12)$$

The Cholesky factor bootstrap replaces the parametric estimates $\hat{A}^{-1}(L)\hat{B}(L)$ with a non-parametric estimate of the dynamics. Specifically, $\hat{P} = \hat{\Sigma}^{1/2}$ is lower triangular so that,

$$y_t^* = \sum_{j=1}^T \hat{P}_{t,j} \varepsilon_j^*. \quad (13)$$

In order to consistently estimate \hat{P} , the number of autocovariances being estimated must grow with (but slower than) the sample size. This can be achieved by down-weighting the off-diagonal elements of $\hat{\Sigma}$. Selecting a particular sequence of weights

amounts to choosing a bandwidth. This choice will, of course, affect the performance of the bootstrap. Severe downweighting induces bias, while too little downweighting reduces efficiency. Thus, in place of the lag order selection problem in parametric models, the nonparametric Cholesky factor bootstrap requires a bandwidth choice. Data-based bandwidth selection procedures for consistent covariance matrix estimation may be found, for example, in Andrews (1991), Andrews and Monahan (1992) or Newey and West (1994).

3 Resampling Blocks of Data

The bootstrap algorithms discussed in section 2 all transform stationary time series data in a way that gives rise to iid residuals. These residuals may then be resampled with replacement. A different strategy has focused on resampling 'blocks' of contiguous time series observations. Throughout this paper we focus on methods for resampling overlapping blocks of data (*moving blocks*). Resampling overlapping blocks may provide somewhat higher bootstrap estimation efficiency than non-overlapping blocks, although the available evidence indicates that the efficiency gain is small (e.g., Hall, Horowitz and Jing (1995)).

Given a set of observations, y_t , $t = 1, \dots, T$, define $b = T - k + 1$ blocks of data $x_t = (y_t, \dots, y_{t+k-1})$ of length k . Künsch (1989) and Liu and Singh (1992) independently propose resampling with replacement from the blocks (x_1, x_2, \dots, x_b) to form pseudo-data, $(x_1^*, x_2^*, \dots, x_l^*)$ of length $T = lk$. The statistic of interest is then calculated for each of many sets of pseudo-data. The distribution of the bootstrap statistic approximates the asymptotic distribution, as long as the size of the blocks increases with sample size.

Under some conditions on the mixing coefficients of the data process, Künsch (1989) proves that the block bootstrap provides a valid approximation to the unknown distribution of the normalized univariate sample mean. Liu and Singh (1992) prove

the optimal block size tends to increase with the persistence of the time series as measured by the dominant root. Third, figure 1 suggests that the performance of the moving blocks bootstrap tends to be fairly stable in the neighborhood of the optimal block size. This suggests that even a fairly coarse grid for k will provide valuable information.

Other work on block selection includes Hall, Horowitz and Jing (1995) and Bühlmann and Künsch (1996). Hall et al. propose an iterative empirical procedure for determining the optimal block size. Starting with an initial guess for \hat{k} , they first determine the optimal block size for a subseries of the original data of length $m < T$. Using asymptotic expressions for the optimal block length, the result for the subseries is recalibrated, so that it applies to the original full sample size. This procedure may then be iterated until convergence is achieved. However, an important issue not discussed in Hall et al. is the selection of the tuning parameter, m .

Bühlmann and Künsch (1996) propose a data-driven procedure based on the equivalence of the block size to the inverse of the bandwidth of a lag weight estimator of the spectral density at frequency zero. This allows them to apply an iterative plug-in method to select the optimal block size. In practice, their method requires an estimate of the influence function of the statistic of interest.

4 Resampling in the Frequency Domain

A different approach to bootstrapping time series data is to resample in the frequency domain. This research is motivated by noting that, even for non-iid data, there are iid relationships in the frequency domain which can be exploited for bootstrapping. The algorithms considered in this section require the data to be covariance stationary or appropriately detrended prior to bootstrapping.

Ramos (1984) makes use of the fact that taking the discrete Fourier transform of

the covariance stationary data gives rise to iid normal variables. Specifically,

$$\frac{1}{\sqrt{2\pi T}} \sum_{t=0}^{T-1} y_t \exp(i\omega_j t) = a(\omega_j) + b(\omega_j)i \quad (14)$$

where $a(\omega_j)$ and $b(\omega_j)$ are known as the Fourier coefficients. For any finite number of frequencies,

$$\begin{aligned} \frac{a(\omega_j)}{\sqrt{f(\omega_j)}} &\xrightarrow{asy} iid N(0, 1/2) \\ \frac{b(\omega_j)}{\sqrt{f(\omega_j)}} &\xrightarrow{asy} iid N(0, 1/2), \end{aligned} \quad (15)$$

where $f(\omega)$ is the spectral density function (e.g., Brillinger (1981)). The independence holds across frequencies and between the two coefficients at each frequency.

Ramos makes use of the asymptotic independence in the following algorithm:

1. Consistently estimate the spectral density function (s.d.f).
2. Generate pseudo-Fourier coefficients:

$$\begin{aligned} a^*(\omega_j) &= \sqrt{\hat{f}(\omega_j)} z_a(\omega_j) \\ b^*(\omega_j) &= \sqrt{\hat{f}(\omega_j)} z_b(\omega_j) \end{aligned}$$

with $z_a(\omega)$ iid $N(0, 1/2)$, $z_b(\omega)$ iid $N(0, 1/2)$.

3. Calculate pseudo-data by taking the inverse Fourier transform:

$$y_t^* = \sum_{j=0}^{T-1} [a^*(\omega_j) \cos(\omega_j t) + b^*(\omega_j) \sin(\omega_j t)]$$

4. Calculate the statistic of interest from the pseudo-data.
5. Repeat many times and build up the empirical distribution of

the validity of the block bootstrap for m -dependent data. Bühlmann (1994) proves the asymptotic validity of the block bootstrap for statistics given by smooth functionals of sample means of vector valued observations. Related results can be found in Naik-Nimbalkar and Rajarshi (1994). In related work, Lahiri (1995) proposes a modified block bootstrap procedure for normalized sums of heavy-tail dependent variables.

While these results establish asymptotic validity, Lahiri (1991, 1992), Götze and Künsch (1993), and Davison and Hall (1993) prove that the block bootstrap is *second order correct* for a wide class of studentized statistics based on sample means in the multivariate setting. That is, this bootstrap algorithm corrects for the second order term in the Edgeworth expansion which the asymptotic approximation cannot. However, the asymptotic refinements cannot in general be obtained by applying the usual formula for the test statistic to the block-bootstrap data. In fact, without suitable modifications, the rate of convergence by Künsch's method may be worse than the rate of normal approximation (e.g. Hall and Horowitz (1996)).

Politis and Romano (1992a) extend the idea of the block bootstrap to estimates of parameters of the infinite-dimensional joint distribution of a stationary time series. Their procedure is as follows:

1. Resample blocks of data $(x_1^*, x_2^*, \dots, x_l^*)$ from the original blocks (x_1, x_2, \dots, x_b) .
2. Define a function $T(x_i^*)$ on each block, such that the statistic of interest \bar{T} can be written, $\bar{T} = \frac{1}{l} \sum_{i=1}^l T(x_i^*)$.
3. Define blocks of statistics, $B_j = (T(x_i^*), T(x_{i+1}^*), \dots, T(x_{i+m}^*))$.
4. Resample with replacement from the B_j which gives a sequence of statistics $(T(x_1^*), T(x_2^*), \dots, T(x_M^*))$.
5. Calculate the bootstrap statistic $\bar{T}^* = \frac{1}{M} \sum_{i=1}^M T(x_i^*)$.

Politis and Romano (1992a) motivate this 'blocks of blocks' bootstrap in part by showing that Künsch's (1989) original block bootstrap fails in the case of spectral density estimators. The 'blocks of blocks' bootstrap, in contrast, delivers a consistent bootstrap distribution. Related work can be found in Politis and Romano (1992b),

Politis, Romano and Lai (1992), and Bühlmann and Künsch (1995).

Künsch and Carlstein (1990) and Carlstein et al. (1995) observe that pseudo-data generated by concatenating resampled blocks of data will not preserve the dependence structure of the original data near block 'endpoints'. They propose linking the blocks in a way designed to deliver a more natural transition from one block to the next.

Even if the true data process is stationary, a particular draw of pseudo-data may not be. Politis and Romano (1994) propose the 'stationary bootstrap' which guarantees stationary pseudo-data. However, their method requires the choice of additional tuning parameters.

Usually, theoretical work on the moving block bootstrap assumes short-range dependence; that is, the observations are assumed to satisfy some form of mixing conditions with a rapidly decaying mixing coefficient. Lahiri (1993) relaxes this assumption and investigates the behavior of the moving block bootstrap when the data exhibit long-range dependence.

3.1 How to Select the Block Size

Moving blocks bootstrap algorithms require the researcher to choose a block size. Li and Maddala (1996) discuss several rules for block size selection based on specific models or on asymptotic mean-squared error (MSE) considerations. In this section, we propose a data-based procedure for choosing the block size in finite samples.

We observe that choosing a block size involves a tradeoff. As the block size becomes too small, the moving blocks bootstrap destroys the time dependency of the data and its average accuracy will decline. As the block size becomes too large, there are few blocks and pseudo-data will tend to look alike. As a result, the average accuracy of the moving blocks bootstrap also will decline. This suggests that there exists an optimal block size \hat{k} which maximizes accuracy. The proposed procedure *automatically* selects this block size for a given series and statistic of interest, regard-

less of the sample size, persistence or lag structure of the underlying process. We will effectively use the bootstrap as an aid in block size selection. Consider a stationary series $\{y_t\}$ of length T :

1. Approximate the underlying $MA(\infty)$ process by a parametric $ARMA(p,q)$ or $AR(p)$ model.
2. Generate many Monte Carlo trials of length T from this fitted model.
3. For each Monte Carlo trial generate moving blocks bootstrap data $\{y_t^*\}$ for alternative block sizes k .
4. Calculate the statistic of interest for $\{y_t^*(k)\}$.
5. Select the block size k which on average produces the most accurate test statistic, point estimate, or confidence interval across Monte Carlo trials.
6. Use that block size \hat{k} to apply the fully nonparametric moving blocks algorithm to the original data $\{y_t\}$.

We illustrate this procedure for the quarterly time series on the S&P common stock earnings-price ratio (CitiBase code: FSEXP) for 1947.2-1994.3. Based on the AIC, we fit an $AR(2)$ model with intercept to this series. After bias-correcting the autoregressive coefficients using the closed-form expression of Pope (1990) we obtain the estimate:

$$y_t = 0.3814 + 1.2514 y_{t-1} - 0.2902 y_{t-2} + \varepsilon_t, \quad \sigma_\varepsilon^2 = 0.6105$$

Our statistic of interest is the response of the earnings-price ratio to a one-standard deviation shock. We will trace out this impulse response function for 16 quarters after the initial shock.

We generate 1,000 Monte Carlo trials from this model to evaluate the probability content of the 90 percent moving blocks bootstrap confidence intervals for the statistic of interest. For each Monte Carlo trial we generate 1,000 moving blocks bootstrap replications of length $T = 190$ for alternative block sizes $k \in \{4, 12, 24, 36, 48, 60, 72,$

84, 96, 108, 120, 136, 144} using Künsch’s (1989) algorithm. For each bootstrap replication, we resample the data conditional on k , fit an $\text{AR}(\hat{p}^*)$ model to $\{y_t^*\}$ based on the AIC lag order estimate, and construct the implied impulse response function for up to 16 quarters after the initial shock. Then we build up the empirical distributions of the impulse response coefficients and construct nominal 90 percent bootstrap confidence intervals for each time horizon. The probability content of these intervals is evaluated across the 1,000 Monte Carlo trials.

Rather than compare the coverage accuracy of the intervals for each block size by visual inspection of the coverage plots across the time horizon $i = 0, \dots, 16$, we construct a simple statistical summary measure. Assuming a quadratic loss function and equal weights for all time horizons of the impulse response function, we average the squared deviations from nominal coverage across the time horizon and tabulate them as a function of the block size. Then the optimal block size is:

$$\hat{k} = \arg \min_k \left[\frac{1}{17} \sum_{i=0}^{16} (\text{coverage}(k, i) - 0.9)^2 \right].$$

Figure 1 plots the mean squared deviations from nominal coverage as a function of k . As expected the curve follows a U-shape. The global minimum is at $\hat{k} = 48$ (quarters).

This result is of interest for several reasons. First, it shows that the performance of the moving blocks bootstrap can be highly sensitive to the choice of the block size. Second, figure 1 indicates that appropriate block sizes are much larger than some illustrative examples in the literature would suggest. For example, Efron and Tibshirani (1993) consider block sizes of 3 and 5 for bootstrapping the slope coefficient in an $\text{AR}(1)$ model with $T = 48$. For the same model, Künsch (1989) considers $k = 4$ for a data set of length $T = 120$. Our procedure indicates that the optimal block size for macroeconomic time series may be up to 12 times higher than the values sometimes considered for similar models in the literature. Additional examples suggest that

interest.

Despite the frequency domain setting, this algorithm is a generic method for generating pseudo-data. Ramos (1984) and Stine (1985) also discuss a version of this method which does not impose normality.

Ramos (1988) provides a general and thorough treatment of frequency domain resampling plans for linear functionals of the spectral density,

$$\theta_f = \int \theta(\omega)f(\omega)d(\omega), \quad (16)$$

where $\theta(\omega)$ is an even periodic function satisfying some regularity conditions and $f(\omega)$ is the s.d.f. of the process. Examples of statistics of this form include the variance and autocovariances. The corresponding estimates must be a linear functional of the periodogram: $\hat{\theta}_y = \int \theta(\omega)I_y(\omega)d(\omega)$. Ramos therefore proposes the following procedure:

1. Obtain a consistent estimate $\hat{f}(\omega)$.
2. Generate pseudo-data y_t^* from the Gaussian distribution described by $\hat{f}(\omega)$. Since a Gaussian distribution is completely characterized by its s.d.f. (ignoring the mean of the process for convenience), $N(0, \hat{\Sigma})$ may be equivalently written as $N(0, \hat{f}(\omega))$. How can we draw from data from $N(0, \hat{f}(\omega))$? Ramos suggests: Take the first, say, m terms in the Fourier series expansion of $\hat{f}(\omega)$: $\gamma_1, \dots, \gamma_m$. These are the first m covariances of the process, with which the covariance matrix may then be constructed. In Ramos' notation, $\hat{\Sigma} = Top(\gamma_1, \dots, \gamma_m)$. Pseudo-data is then drawn from $N(0, \hat{\Sigma})$.
3. Calculate $\hat{\theta}_{y^*} = \int \theta(\omega)I_{y^*}(\omega)d(\omega)$.
4. Repeat many times and build up the empirical distribution of interest.

Ramos' main result is that under some regularity conditions, including normality,

$$\sqrt{T}(\hat{\theta}_{y^*} - \theta_f) - \sqrt{T}(\hat{\theta}_y - \theta_f) \xrightarrow{p} 0, \quad (17)$$

that is, the bootstrap principle holds.

Note that in both the Ramos (1988) and Gaussian Cholesky factor procedures, the pseudo-data are drawn from $y^* \sim N(0, \hat{\Sigma})$. Ramos estimates $\hat{\Sigma}$ by inverting a consistently estimated spectral density function. In contrast, the Cholesky factor procedure generates consistent estimates $\hat{\Sigma}$ in the time domain by setting $\hat{\Sigma} = \text{Top}(\lambda_0 \hat{\gamma}_0, \dots, \lambda_{T-1} \hat{\gamma}_{T-1})$, where $\text{Top}(\cdot)$ is the operator which creates a toeplitz matrix from a single row of covariances, and where λ_t are decreasing weights. One can show that, if the λ_t form a Bartlett window, the Cholesky factor procedure is equivalent to inverting the s.d.f estimated with a Fejer spectral window in Ramos' procedure.

An algorithm explicitly designed to bootstrap the spectral density function itself is given in Franke and Härdle (1992). They make use of the same asymptotic relationships as Ramos (1984), but in a different way. Equations (15) immediately imply that,

$$\left(\sqrt{2} \frac{a(\omega)}{\sqrt{f(\omega)}} \right)^2 + \left(\sqrt{2} \frac{b(\omega)}{\sqrt{f(\omega)}} \right)^2 \xrightarrow{d} \chi_2^2 \quad (18)$$

and so

$$\frac{2}{f(\omega)} (a(\omega)^2 + b(\omega)^2) \xrightarrow{d} \chi_2^2. \quad (19)$$

But from the definition of Fourier coefficients, equation (19) implies that

$$\frac{2}{f(\omega)} I(\omega) \xrightarrow{d} \chi_2^2, \quad (20)$$

where $I(\omega)$ is the periodogram. Equation 20 holds approximately in finite samples. Thus, Franke and Härdle (1992) suggest the following bootstrap algorithm:

1. Compute the periodogram, $I(\omega)$, and consistently estimate the s.d.f., $\hat{f}(\omega)$.
- 2a. Resample from $\varepsilon^* = \chi_2^2/2$ or
- 2b. Resample from the empirical distribution: $\varepsilon^* = \frac{I(\omega)}{\hat{f}(\omega)}$, where the number of estimated residuals equals the number of frequencies.
3. Calculate the bootstrap periodogram ordinates: $I(\omega)^* = \varepsilon^* \hat{f}(\omega)$.
4. Calculate bootstrap spectral density estimates: $f^*(\omega) = \sum k(\omega) I^*(\omega)$, by smoothing the periodogram ordinates.
5. Repeat steps 2-4 many times and build up empirical distribution of interest.

Franke and Härdle (1992) prove the consistency of the bootstrap distribution of the pivoted s.d.f. They also report the results of a simulation study for an AR(5) model at five discrete frequencies. They find that the bootstrap performs favorably relative to the asymptotics in capturing the finite sample skewness of the distribution of the s.d.f. estimates.

Berkowitz and Diebold (1996) describe a multivariate generalization of the Franke and Härdle (1992) procedure. Equation 20 is, in fact, a special case of the more general result:

$$I_{yy}(\omega) \xrightarrow{d} W_r^c(1, F_{yy}(\omega)),$$

(see, for example, Brillinger (1981)). $I_{yy}(\omega)$ is the $r \times r$ periodogram matrix and $F_{yy}(\omega)$ is the $r \times r$ spectral density matrix of an r -dimensional vector random variable, Y_t . The asymptotic distribution is complex Wishart. A multivariate version of the bootstrap is implemented by noting that

$$F_{yy}^{-1/2}(\omega) I_{yy}(\omega) F_{yy}^{-1/2}(\omega) \xrightarrow{d} iid W_r^c(1, I_r).$$

Thus, step 2 above is replaced with

- 2a. Resample from the empirical distribution: $\varepsilon^* \sim \hat{F}_{yy}^{-1/2}(\omega) I_{yy}(\omega) \hat{F}_{yy}^{-1/2}(\omega)$
 or 2b. from the parametric distribution: $\varepsilon^* \sim W_r^c(1, I_r)$.

The frequency domain methods discussed in this section may be distinguished by the statistics they are designed to bootstrap. Whereas some algorithms are specifically designed for bootstrapping the s.d.f. or linear functionals of the s.d.f., others are omnibus procedures for generating pseudo-data. A common feature is that they generally require a consistent estimate of the s.d.f. and thus a bandwidth choice. However, the effect of the bandwidth choice on the performance of the bootstrap remains an open question. For frequency domain bootstraps which require the calculation of the entire spectral density function (rather than the spectral density at frequency zero), the data-based bandwidth selection procedures of section 2 are not appropriate. Automatic bandwidth selection procedures for the entire s.d.f. have been suggested, for example, by Beltrao and Bloomfield (1987).

5 Bootstrapping Nonstationary Data

If the true process is nonstationary, many of the standard results of the asymptotic validity of the bootstrap approximation no longer apply. For example, the Cholesky factor and the frequency domain bootstrap assume stationarity, as does nearly all the work on moving blocks bootstraps. Lahiri (1992) shows that, under appropriate conditions, the Künsch procedure is second-order correct even for the sample mean of nonstationary data. It is not known to what extent his results generalize to other statistics. For the parametric AR(1) model, Basawa, Mallik, McCormick, Reeves, and Taylor (1991) and Datta (1992) prove that the standard bootstrap algorithm for the level autoregression is invalid if the true model is a random walk model. Intuitively, this result arises because of the discontinuity of the asymptotic distribution at the unit circle and generalizes to all exact unit root VAR models estimated in levels.

If the model is known to contain an exact unit root, resampling remains valid if

we impose that unit root in estimation. In particular, in the absence of cointegration, autoregressive $I(1)$ processes can be estimated in first differences and the coefficients converted to the level representation. Similarly, nonparametric methods continue to be valid if resampling is based on first-differenced data. While nonparametric bootstrap methods can easily deal with $I(1)$ processes, there are no theoretical results to show that nonparametric resampling preserves cointegration relationships in the data. In fact, cointegration itself may be viewed as a parametric notion. Thus, if the data are known to be cointegrated, parametric methods are preferable. For example, if the true process were a cointegrated VAR, one would resample from the ML estimate of a vector error correction model. Since the innovations in that model are identical to the innovations in the level autoregression, the usual bootstrap algorithm is valid. Kilian (1995) reports that in many cases the bootstrap algorithm is quite accurate for vector error correction models, provided the cointegration rank is known. However, the estimation of a linear time trend or the presence of additional roots of large magnitude in the system may lead to similar bias problems as in stationary autoregressions. Li and Maddala (1996) provide a detailed discussion of the special problems involved in bootstrapping the cointegration relationship itself.

In applied work the existence of a unit root or cointegration is rarely known with certainty. For some econometric questions, this is not a problem. For example, if we are interested in approximating the finite sample distribution of a test statistic under the null hypothesis of a unit root, it is correct to simply impose the unit root in estimation. However, not all inference problems involve a unit root null hypothesis. For example, bootstrap confidence intervals do not involve a specific null. Rather, the user faces the choice between ignoring the possible presence of unit roots or relying on the result of pre-tests with low power. One would like to think that the level bootstrap still provides a satisfactory approximation for roots arbitrarily close to unity. However, there is little evidence to support that view. For example, the accuracy of the bootstrap approximation for the level autoregression

can be expected to deteriorate, as the persistence of the process rises (Bose (1988)). Simulation evidence in De Wet and van Wyk (1986) for the AR(1) coefficient seems to confirm that conjecture. We conjecture that this result is largely due to small-sample bias, and that, in light of the arguments in sections 2.1.1. and 2.1.2, the accuracy of the bootstrap could be substantially improved by appropriate bias corrections.

What if the roots of the process are explosive? There are no known results for nonparametric bootstrap methods involving explosive roots, but, interestingly, the parametric bootstrap can be shown to be theoretically valid for such models. Basawa, Mallik, McCormick, and Taylor (1989) prove the validity of the bootstrap in the explosive AR(1) model with finite error variance. Datta (1995) shows that the limit of the bootstrap distribution in the AR(p) model converges to that of the OLS slope coefficient estimate in probability without any additional moment restrictions. In particular, Datta proves that the bootstrap approximation as measured by the Kolmogoroff distance goes to zero almost surely provided $E|\varepsilon_t| < \infty$. Thus for the explosive case the bootstrap principle works, even if the error distribution is heavy tailed.

These results seem of limited practical relevance, but Datta (1995) takes the analysis one step further. He shows that for $E|\varepsilon_t^2| < \infty$, the standard OLS bootstrap offers an asymptotically valid approximation even in the partially explosive AR(p) model (when the characteristic polynomial admits roots both inside and outside the unit circle). Datta assumes that for some k , $1 \leq k \leq p$, exactly k roots of the characteristic polynomial lie inside and the remaining $s = p - k$ roots lie outside the unit circle. Note that for $k = p$ the model is stationary and for $k < p$ it is partially explosive. Exact unit roots are ruled out by assumption. For the OLS estimator, Datta proves that the error in bootstrap approximation (measured in sup norm) converges to zero almost surely. His proof subsumes the stationary case, strengthening Kreiss and Franke's (1992) result about convergence in probability. Ignoring the possibility of purely explosive AR(p) models, Datta's result suggests that, at least asymptoti-

cally, the standard bootstrap is valid for processes with roots arbitrarily close to unity, both inside and outside the unit circle. It thus strengthens the case for bootstrapping level autoregressions.

6 A Monte Carlo Comparison of Bootstrap

Methods

The various bootstrap algorithms of the previous sections differ in how much parametric structure they impose in estimation. For example, Cholesky factor, frequency domain and block bootstrap procedures are completely nonparametric. They do not use any particular model to generate pseudo-data. On the other hand, parametric AR(p) bootstrap procedures make strong assumptions about the form of the time dependency of the data. Which procedures should applied time series econometricians use? We conjecture that for the sample sizes macroeconomists tend to work with, parsimony is essential and thus parametric models may be the only reasonable choice. To verify this conjecture, we compare the accuracy of the parametric AR(p) bootstrap, the Cholesky factor bootstrap, the Ramos (1984) frequency domain bootstrap, and the moving blocks bootstrap algorithm of Künsch (1989) by Monte Carlo simulation.

We consider confidence intervals for the responses of the T-bill rate to a one-standard deviation shock over the subsequent 16 quarters. We believe that this example is of broad interest to applied users. Impulse responses play an important role in macroeconometrics. Further, they share many statistical properties with multi-step ahead forecasts.

In the Monte Carlo experiment, we consider three sample sizes, corresponding to 20 and 40 years worth of quarterly data and 40 years worth of monthly data: $T = 80, 160, 480$. Our data generating process is based on quarterly U.S. T-bill data

(CitiBase code: FYGM3) for 1971.4-1993.4. The AIC suggests an ARMA(2,4) data generating process:

$$y_t = 0.8871 + 0.3499 y_{t-1} + 0.5231 y_{t-2} + 1.0004 \varepsilon_{t-1} - 0.1103 \varepsilon_{t-2} \\ + 0.0021 \varepsilon_{t-3} + 0.3492 \varepsilon_{t-4} + \varepsilon_t,$$

with $\sigma_\varepsilon^2 = 0.7124$. The innovations ε_t are normally distributed.

An important concern in this study is model selection. The AR(p) estimate underlying the parametric bootstrap is based on the minimum of the AIC for $1 \leq p \leq 8$. For the Cholesky factor bootstrap, we use the automatic bandwidth selection procedure of Andrews (1991). For computational reasons, the bandwidth is not permitted to exceed three quarters of the sample size. We use the algorithm of Beltrao and Bloomfield (1987) for the Ramos (1984) bootstrap. For the moving blocks bootstrap we use the procedure outlined in the previous section to select the block size.⁴

For each bootstrap replication $\{y_t^*\}$ we use the AIC to estimate the lag order $\{p^*\}$, fit an AR($\{p^*\}$), and calculate the statistic of interest. The confidence intervals are based on the 0.05 and 0.95 percentile interval endpoints of the empirical distribution of the impulse response coefficient estimates.

Figures 2-4 plot the effective coverage rates of the nominal 90 percent intervals for each method and sample size. For the sample sizes considered the parametric AR(p) bootstrap (labeled AR) interval clearly dominates the other three bootstrap intervals. The Cholesky factor (CHOL), the Künsch (1989) block bootstrap (BLOCK) and the Ramos (1984) frequency domain bootstrap (RAMOS) intervals perform erratically

⁴For computational reasons we determine the optimal block size a priori based on the AR(p) approximation of the actual data. The AIC suggests an AR(6) model for the T-Bill rate data. We bias-correct the autoregressive coefficient estimates using the expressions in Pope (1990). Conditional on this DGP, we find $\hat{k} = 36$ for $T = 80$ and $k \in \{12, 24, 36, 48, 60, 72\}$, $\hat{k} = 60$ for $T = 160$ and $k \in \{12, 24, 36, 48, 60, 72, 84, 96\}$, and $\hat{k} = 120$ for $T = 480$ and $k \in \{80, 120, 160, 200, 240, 280\}$.

and are generally unreliable. However, the Cholesky bootstrap performs somewhat better than the moving blocks bootstrap. The extremely poor performance of the Ramos bootstrap may in part reflect the bandwidth selection criterion employed. It also suggests that the Ramos procedure may not be consistent for our statistic of interest.

With the exception of the Ramos bootstrap, coverage accuracy tends to improve with sample size. However, in absolute terms none of the methods can be considered adequate. Even the AR(p) bootstrap interval consistently falls short of nominal coverage. This coverage deficiency is consistent with the arguments we presented in section 2. Further improvements in coverage accuracy are likely to require interval calibration (Breidt et al. (1995)) or some form of bias correction in the statistic of interest (e.g., Rudebusch (1993), Andrews and Chen (1994)). To illustrate this point, we added results for the same AR(p) bootstrap with additional bias corrections (AR-BC) for the impulse response estimates based on Kilian (1995). The improvement in coverage accuracy is sizable. Similar improvements are likely for the other methods.

7 Conclusion

We reviewed a range of alternative parametric and nonparametric bootstrap algorithms for time-dependent data. These methods differ in the extent to which they impose parametric structure on the data. In highly parsimonious models, parameters are estimated with many degrees of freedom. Bootstrap estimates, in turn, are comparatively accurate. However, parsimony is also likely to increase the "distance" between the fitted and the true model. Conditioning on a misspecified model may cause the misspecification to be propagated (and possibly magnified) through resampling. We provided some preliminary Monte Carlo evidence that for typical sample sizes faced by macroeconomists, parsimony is essential and thus parametric models may be the only reasonable choice. However, care must be exercised to overcome the

drawbacks of bootstrapping parametric models in small samples. In particular, we stressed the choice of lag order selection criteria, the treatment of lag order uncertainty, and the need for bias corrections in small samples.

We also showed that the accuracy of the moving block bootstrap can be highly sensitive to the choice of block size. We proposed an automatic data-based procedure for selecting the block size. We successfully applied this procedure to several economic time series and determined that for macroeconomic time series the optimal block sizes tend to be much larger than those sometimes used in the literature.

While our preliminary Monte Carlo evidence suggests that nonparametric bootstrap methods may perform poorly in small samples, results for other statistics and data generating processes would be useful. In particular, the use of an ARMA data generating process may have biased the results in favor of parametric bootstraps. Additional research is needed to determine whether nonparametric algorithms enjoy special advantages for data generating processes that are not encompassed by the ARMA framework. Moreover, it would be important to obtain a sense of the sample sizes required for reliable inference based on nonparametric methods. That information would be useful for the analysis of higher frequency financial data. However, these caveats cannot obscure that nonparametric methods need to be improved to be of much use in small and moderately large samples.

In particular, the comparatively poor performance of the Cholesky factor, the block bootstrap and the frequency domain bootstrap for the sample sizes considered is surprising. A partial explanation could be the slow rate of convergence of nonparametric methods, but it would be premature to discard these methods. Indeed, Diebold, Ohanian and Berkowitz (1995) present Monte Carlo evidence which suggests that the Cholesky bootstrap delivers far better coverage for spectral density estimates than the asymptotic approximation. The Ramos bootstrap does not appear suitable for impulse response estimates. More theoretical work is needed to establish a class of statistics for which this algorithm may be used.

More research is needed to clarify how the choice of the bandwidth affects the performance of these algorithms. We conjecture that their performance could be improved by refining the bandwidth selection process. In this paper, we used the Andrews (1991) and the Beltraõ and Bloomfield (1987) automatic bandwidth selection criteria. It would be of interest to systematically compare the performance of frequency domain methods to other data-based bandwidth selection criteria (e.g., Andrews and Monahan (1992), Newey and West (1994)) or parametric devices for estimating the spectral density. For example, den Haan and Levin (1996) report that autoregressive estimates of the spectral density at frequency zero outperform kernel-based estimates. In addition, bootstrapping the bandwidth selection process is likely to improve small-sample performance.

Similarly, the performance of the Künsch (1989) moving block bootstrap left much to be desired. In the future, it would be valuable to study more sophisticated block bootstraps such as the linked-block bootstrap of Künsch and Carlstein (1990), the stationary block bootstrap of Politis and Romano (1994) or the asymptotic refinements proposed by Lahiri (1991, 1995) and Götze and Künsch (1993). Moreover, a systematic comparison of the block size selection methods proposed in this paper and elsewhere would be useful.

References

- [1] Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817-858.
- [2] Andrews, D.W.K, and H.Y. Chen (1994), "Approximately Median-Unbiased Estimation of Autoregressive Models," *Journal of Business and Economic Statistics*, 12, 187-204.
- [3] Andrews, D.W.K, and J.C. Monahan (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953-966.
- [4] Basawa, L.V., A.K.Mallik, W.P. McCormick, and R.L. Taylor (1989), "Bootstrapping Explosive Autoregressive Processes," *Annals of Statistics*, 17, 1479-1486.
- [5] Basawa, L.V., A.K.Mallik, W.P. McCormick, J.H. Reeves, and R.L. Taylor (1991), "Bootstrapping Unstable First Order Autoregressive Processes," *Annals of Statistics*, 19, 1098-1101.
- [6] Beltrao, K.I. and P. Bloomfield (1987), "Determining the Bandwidth of a Kernel Spectrum Estimate," *Journal of Time Series Analysis*, 8, 21-38.
- [7] Berkowitz, J. and F.X. Diebold (1996), "Bootstrapping Multivariate Spectra," manuscript, Department of Economics, University of Pennsylvania.
- [8] Bickel, P., and P. Bühlmann (1996), "Mixing Property and Functional Central Limit Theorems for a Sieve Bootstrap in Time Series," manuscript, Department of Statistics, University of California, Berkeley.
- [9] Bose, A. (1988), "Edgeworth Correction by Bootstrap in Autoregressions," *Annals of Statistics*, 16, 1709-1722.

- [10] Bose, A. (1990), "Bootstrap in Moving Average Models," *Annals. Institute of Statistical Mathematics*, 42, 753-768.
- [11] Breidt, F.J., and R.A. Davis (1991), "Time-reversibility, Identifiability, and Independence of Innovations in Stationary Time Series," *Journal of Time Series Analysis*, 13, 377-390.
- [12] Breidt, F.J., R.A. Davis, and W.T.M. Dunsmuir (1992), "On Backcasting in Linear Time Series Models," in *New Directions in Time Series Analysis. Part I*, D. Brillinger et al., eds., Springer-Verlag, New York.
- [13] Breidt, F.J., R.A. Davis, and W.T.M. Dunsmuir (1995), "Improved Bootstrap Prediction Intervals for Autoregressions," *Journal of Time Series Analysis*, 16, 177-200.
- [14] Brillinger, D.R. (1981), *Time Series: Data Analysis and Theory*. San Francisco: Holden-Day.
- [15] Bühlmann, P. (1994), "Blockwise Bootstrapped Empirical Processes for Stationary Sequences," *Annals of Statistics*, 22, 995-1012.
- [16] Bühlmann, P. (1996a), "Sieve Bootstrap for Time Series," to appear in *Bernoulli*.
- [17] Bühlmann, P. (1996b), "Confidence Regions for Trends in Time Series: A Simultaneous Approach with a Sieve Bootstrap," Technical Report 447, Department of Statistics, University of California, Berkeley.
- [18] Bühlmann, P., and H.R. Künsch (1996), "Block Length Selection in the Bootstrap for Time Series," manuscript, Seminar für Statistik, ETH Zürich.
- [19] Bühlmann, P., and H.R. Künsch (1995), "The Blockwise Bootstrap for General Parameters of a Stationary Time Series," *Scandinavian Journal of Statistics*, 22, 35-54.

- [20] Carlstein, E. (1992), "Resampling Techniques for Stationary Time Series: Some Recent Developments," in *New Directions in Time Series Analysis. Part I*, D. Brillinger et al., eds., Springer-Verlag, New York.
- [21] Carlstein, E., K.-A. Do, P. Hall, T. Hesterberg, and H.R. Künsch (1995), "Matched Block Bootstrap for Dependent Data," Technical Report 74, Seminar für Statistik, ETH Zürich.
- [22] Chatterjee, S. (1986), "Bootstrapping ARMA Models: Some Simulations," *IEEE Transactions Systems, Man and Cybernetics*, SMC-16, 294-299.
- [23] Datta, S. (1992), "On Asymptotic Properties of Bootstrap for Unstable AR(1) Processes," Technical Report, Department of Statistics, University of Georgia.
- [24] Datta, S. (1995), "Limit Theory and Bootstrap for Explosive and Partially Explosive Autoregression," *Stochastic Processes and their Applications*, 57, 285-304.
- [25] Davison, A.C., and P. Hall (1993), "On Studentizing and Blocking Methods for Implementing the Bootstrap with Dependent Data," *Australian Journal of Statistics*, 35, 215-224.
- [26] De Wet, T., and J.W.J. van Wyk (1986), "Bootstrap Confidence Intervals for Regression Coefficients when the Residuals are Dependent," *Journal of Statistical Computation and Simulation*, 23, 317-327.
- [27] Den Haan, W.J. and A. Levin (1996), "Inferences from Parametric and Non-parametric Covariance Matrix Estimation Procedures," manuscript, Board of Governors of the Federal Reserve System.
- [28] Diebold, F.X., L.E. Ohanian and J. Berkowitz (1995), "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data," National Bureau of Economic Research Technical Working Paper No. 174.

- [29] Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, 9, 1218-1228.
- [30] Efron, B. (1987), "Better Bootstrap Confidence Intervals," *Journal of the American Statistical Association*, 82, 171-200.
- [31] Efron, B. and R. Tibshirani (1986), "Bootstrap Methods for Standard Errors, Confidence Intervals, and other Measures of Statistical Accuracy," *Statistical Science*, 1, 54-77.
- [32] Efron, B. and R. Tibshirani (1993), *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- [33] Findley, D.F. (1986), "On Bootstrap Estimates of Forecast Mean Square Errors for Autoregressive Processes," in: *Computer Science and Statistics: The Interface*. D.M. Allen, ed., Elsevier Science Publishers, Amsterdam.
- [34] Franke, J. and W. Härdle (1992), "On Bootstrapping Kernel Spectral Estimates," *Annals of Statistics*, 20, 121-145.
- [35] Götze, F., and H.R. Künsch (1993), "Blockwise Bootstrap for Dependent Observation: Higher-Order Approximations for Studentized Statistics," Technical Report, Universität Bielefeld, Germany.
- [36] Hall, P., and J.L. Horowitz (1996), "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators," *Econometrica*, 64, 891-916.
- [37] Hall, P., J.L. Horowitz and B.-Y. Jing (1995), "On Blocking Rules for the Bootstrap with Dependent Data," *Biometrika*, 82, 561-574.
- [38] Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

- [39] Horowitz, J.L. (1995), "Bootstrap Methods in Econometrics: Theory and Numerical Performance," manuscript, Department of Economics, University of Iowa.
- [40] Jeong, J. and G.S. Maddala (1993), "A Perspective on Application of Bootstrap Methods in Econometrics," *Handbook of Statistics*, vol 11.
- [41] Kabaila, P. (1993), "On Bootstrap Predictive Inference for Autoregressive Processes," *Journal of Time Series Analysis*, 14, 473-484.
- [42] Kilian, L. (1995), "Small-Sample Confidence Intervals for Impulse Response Functions," manuscript, Department of Economics, University of Pennsylvania.
- [43] Kilian, L. (1996a), "Impulse Response Analysis in Vector Autoregressions with Unknown Lag Order," manuscript, Department of Economics, University of Pennsylvania.
- [44] Kilian, L. (1996b), "An Endogenous Lag-Order Bootstrap Algorithm with an Application to Impulse Response Analysis," manuscript, Department of Economics, University of Pennsylvania.
- [45] Kiviet, J.F. (1984), "Bootstrap Inference in Lagged-dependent Variable models," University of Amsterdam Working Paper.
- [46] Kreiss, J.P. (1992), "Bootstrap Procedures for $AR(\infty)$ -Processes," in *Bootstrapping and Related Techniques*, Proceedings, Lecture Notes in Economics and Mathematical Systems, eds. K.H. Jöckel, G. Rothe, and W. Sendler. New York: Springer-Verlag.
- [47] Kreiss, J.P., and J. Franke (1992), "Bootstrapping Stationary Autoregressive Moving-Average Models," *Journal of Time Series Analysis*, 13, 287-317.
- [48] Künsch, H.R. (1989), "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217-1241.

- [49] Künsch, H.R., and E. Carlstein (1990), "The Linked Blockwise Bootstrap for Serially Dependent Observations," Technical Report in preparation, Department of Statistics, University of North Carolina, Chapel Hill.
- [50] Lahiri, S.N. (1991), "Second-Order Optimality of Stationary Bootstrap," *Statistics and Probability Letters*, 11, 335-341
- [51] Lahiri, S.N. (1992), "Edgeworth Correction by 'Moving Block' Bootstrap for Stationary and Nonstationary Data", in *Exploring the Limits of Bootstrap*, R. LePage and L. Billard, eds., John Wiley & Sons, New York.
- [52] Lahiri, S.N. (1993), "On the Moving Block Bootstrap under Long Range Dependence," *Statistics and Probability Letters*, 18, 405-413.
- [53] Lahiri, S.N. (1995), "On the Asymptotic Behavior of the Moving Block Bootstrap for Normalized Sums of Heavy-Tail Random Variables," *Annals of Statistics*, 23, 1331-1349.
- [54] Lamoureux, C.G., and W.D. Lastrapes (1990), "Persistence in Variance, Structural Change, and the GARCH Model," *Journal of Business and Economic Statistics*, 8, 225-234.
- [55] Léger, C., D.N. Politis, and J.P. Romano (1992), "Bootstrap Technology and Applications," *Technometrics*, 34, 378-398.
- [56] Li, H., and G.S. Maddala (1993), "Bootstrapping Cointegrating Regressions," paper presented at the Fourth Meeting of the European Conference Series in Quantitative Economics and Econometrics, December 16-18, 1993.
- [57] Li, H., and G.S. Maddala (1996), "Bootstrapping Time Series Models," *Econometric Reviews*, 15, 115-158.
- [58] Liu, R.Y. (1988), "Bootstrap Procedures under Some Non-i.i.d Models," *Annals of Statistics*, 16, 1696-1708.

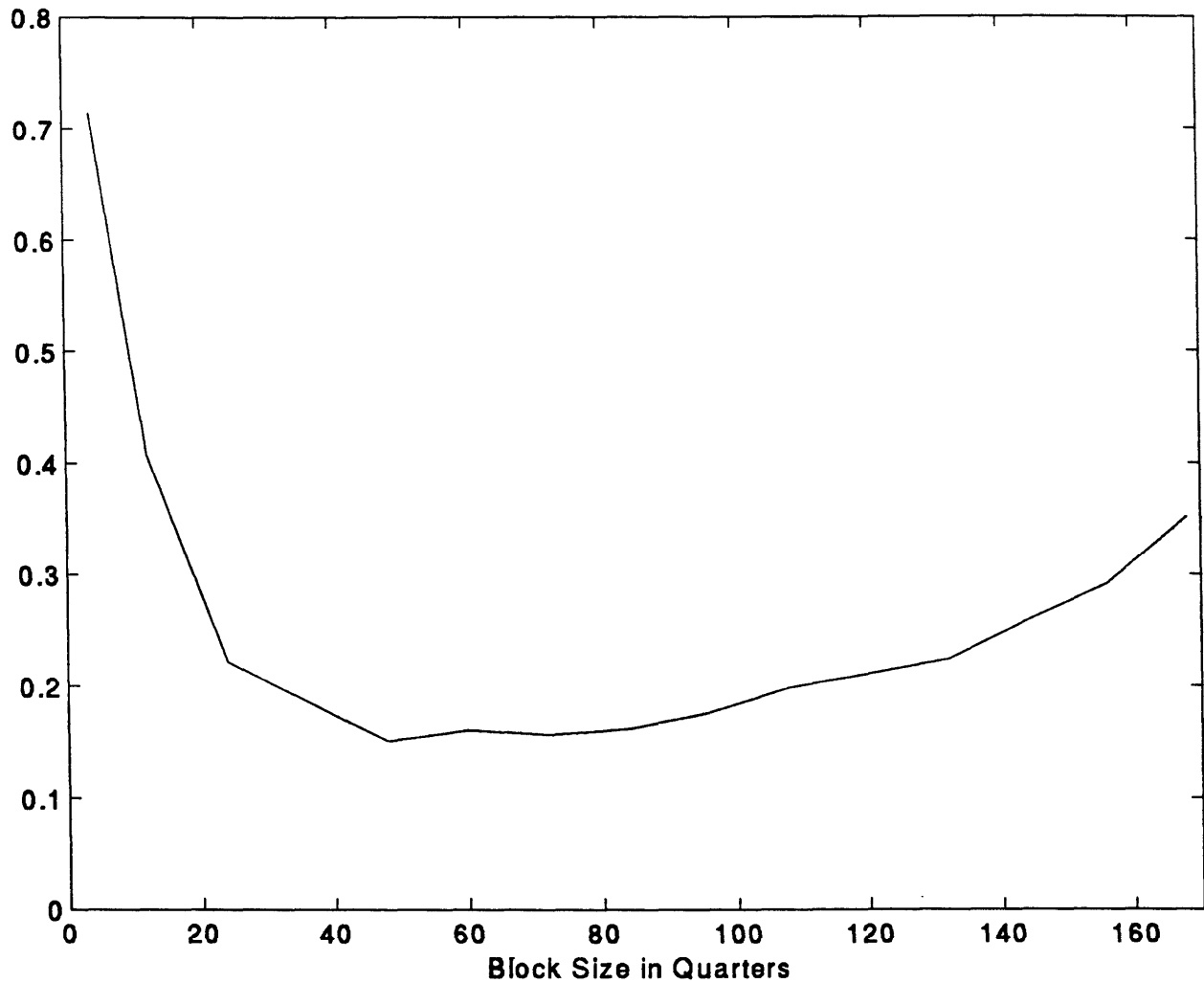
- [59] Liu, R.Y. and K. Singh (1992), "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence", in *Exploring the Limits of Bootstrap*, R. LePage and L. Billard, eds. New York: John Wiley & Sons
- [60] Masarotto, G. (1990), "Bootstrapping Prediction Intervals for Autoregressions," *International Journal of Forecasting*, 6, 229-239.
- [61] McCullough, B.D. (1994), "Bootstrapping Forecast Intervals: An Application to AR(p) Models," *Journal of Forecasting*, 13, 51-66.
- [62] Naik-Nimbalkar, U.V., and M.B. Rajarshi (1994), "Validity of Blockwise Bootstrap for Empirical Processes with Stationary Observations," *Annals of Statistics*, 22, 980-994.
- [63] Nankervis, J.C., and N.E. Savin (1996), "The Level and Power of the Bootstrap t Test in the AR(1) Model with Trend," *Journal of Business and Economic Statistics*, 14, 161-168.
- [64] Newey, W.K. and K.D. West (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631-653.
- [65] Nicholls, D.F. and A.L. Pope (1988), "Bias in the Estimation of Multivariate Autoregressions," *Australian Journal of Statistics*, 30A, 296-309.
- [66] Paparoditis, E. (1996), "Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes," *Journal of Multivariate Analysis*, 57, 277-296.
- [67] Paparoditis, E., and B. Streitberg (1992), "Order Identification in Stationary Autoregressive Moving Average Models: Vector Autocorrelations and the Bootstrap," *Journal of Time Series Analysis*, 13, 415-435.

- [68] Paulsen, J., and D. Tjøstheim (1985), "On the Estimation of Residual Variance and Order in Autoregressive Time Series," *Journal of the Royal Statistical Society B*, 47, 216-228.
- [69] Peters, S.C., and D.A. Freedman (1984), "Some Notes on the Bootstrap in Regression Problems," *Journal of Business and Economic Statistics*, 2, 406-409.
- [70] Pötscher, B.M. (1991), "Effects of Model Selection on Inference," *Econometric Theory*, 7, 163-185.
- [71] Politis, D.N. and J.P. Romano (1992a), "A General Resampling Scheme for Triangular Arrays of α -Mixing Random Variables with Application to the Problem of Spectral Density Estimation," *Annals of Statistics*, 20, 1985-2007.
- [72] Politis, D.N., and J.P. Romano (1992b), "A Nonparametric Resampling Procedure for Multivariate Confidence Regions in Time Series Analysis," in *Computing Science and Statistics, Proceedings of the 22nd Symposium on the Interface*, eds. C. Page and R. LePage, Springer-Verlag, New York
- [73] Politis, D.N. and J.P. Romano (1994), "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89, 1303-1313.
- [74] Politis, D.N., J.P. Romano and T. Lai (1992), "Bootstrap Confidence Intervals for Spectra and Cross-Spectra," *IEEE Transactions on Signal Processing*, 40, 1206-1215.
- [75] Pope, A.L. (1987), "Small-Sample Bias Problems in Time Series," unpublished MSc Thesis, Australian National University.
- [76] Pope, A.L. (1990), "Biases of Estimators in Multivariate Non-Gaussian Autoregression," *Journal of Time Series Analysis*, 11, 249-258.
- [77] Priestley, M.B. (1981), *Spectral Analysis and Time Series*. New York: Academic Press.

- [78] Quinn, B.G. (1988), "A Note on AIC Order Determination for Multivariate Autoregression," *Journal of Time Series Analysis*, 9, 241-245.
- [79] Ramos, E. (1984), "A Bootstrap for Time Series," Qualifying paper, Department of Statistics, Harvard University.
- [80] Ramos, E. (1988), "Resampling Methods for Time Series," Technical Report No. ONR- C-2, Department of Statistics, Harvard University.
- [81] Rayner (1990), "Bootstrapping p Values and Power in the First-Order Autoregression: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 8, 251-263.
- [82] Rudebusch, G.D. (1993), "The Uncertain Unit Root in Real GNP," *American Economic Review*, 83, 264-272.
- [83] Stine, R.A. (1985), "A Frequency Domain Bootstrap for Time Series," manuscript, Department of Statistics, Wharton School, University of Pennsylvania.
- [84] Stine, R.A. (1987), "Estimating Properties of Autoregressive Forecasts," *Journal of the American Statistical Association*, 82, 1072-1078.
- [85] Stoffer, D.S. and K.D. Wall (1991), "Bootstrapping State-Space Models: Gaussian Maximum Likelihood Estimation and the Kalman Filter," *Journal of the American Statistical Association*, 86, 1024-1033.
- [86] Swanepoel, J.W.H and J.W.Y. van Wyk (1986), "The Bootstrap Applied to Power Spectral Density Function Estimation," *Biometrika*, 73, 135-141.
- [87] Thombs, L.A., and W.R. Schucany (1990), "Bootstrap Prediction Intervals for Autoregressions," *Journal of the American Statistical Association*, 85, 486-492.

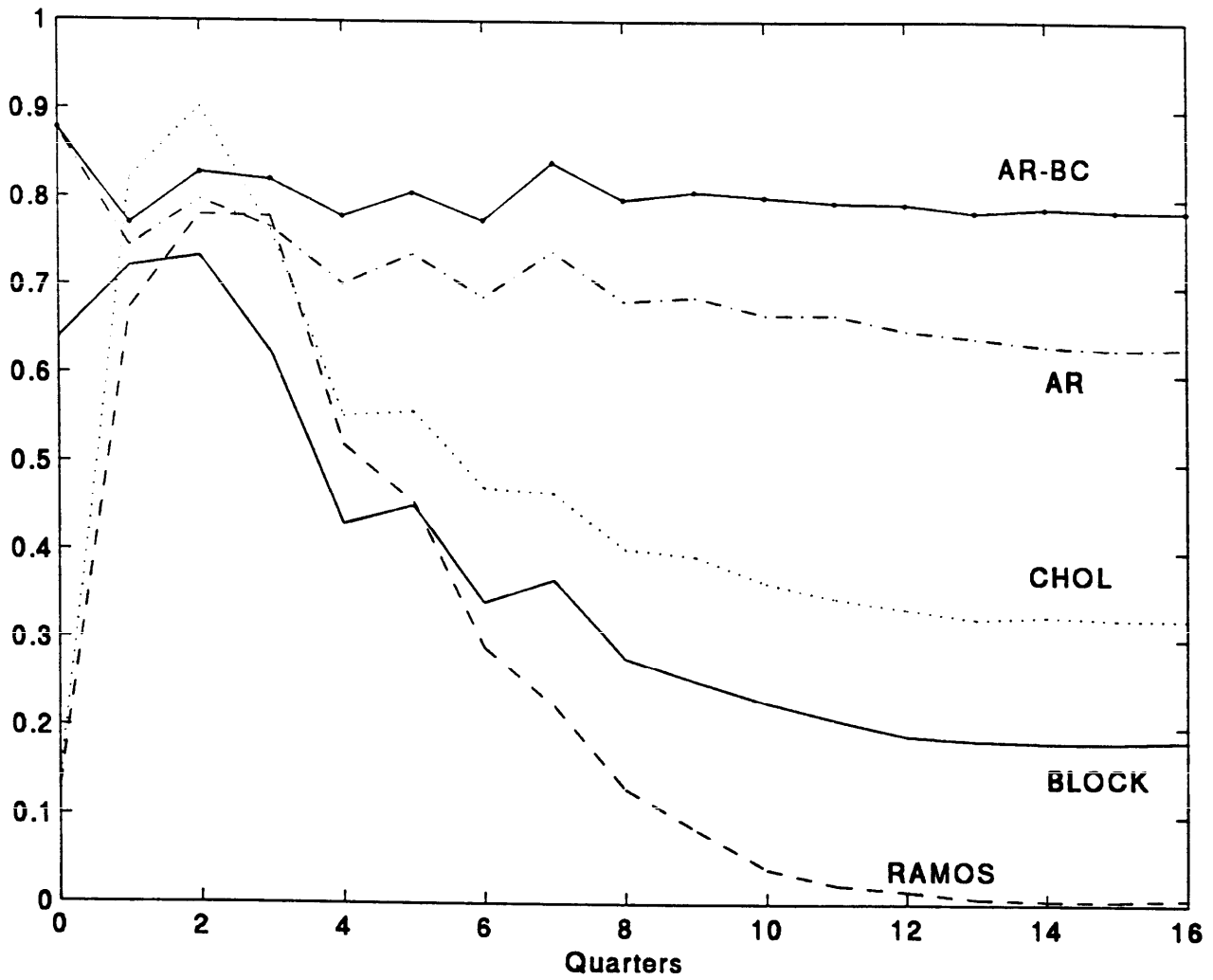
- [88] Young, G.A. (1994), "Bootstrap: More than a Stab in the Dark? " *Statistical Science*, 9, 382-415.
- [89] Zivot, E., and D. Andrews (1992), Further Evidence on the Great Crash, the Oil Price Shock, and The Unit Root Hypothesis," *Journal of Business and Economic Statistics*, 10, 237-250.

Figure 1
Selecting the Optimal Blocksize by Minimizing
Mean Squared Deviations from Nominal Coverage



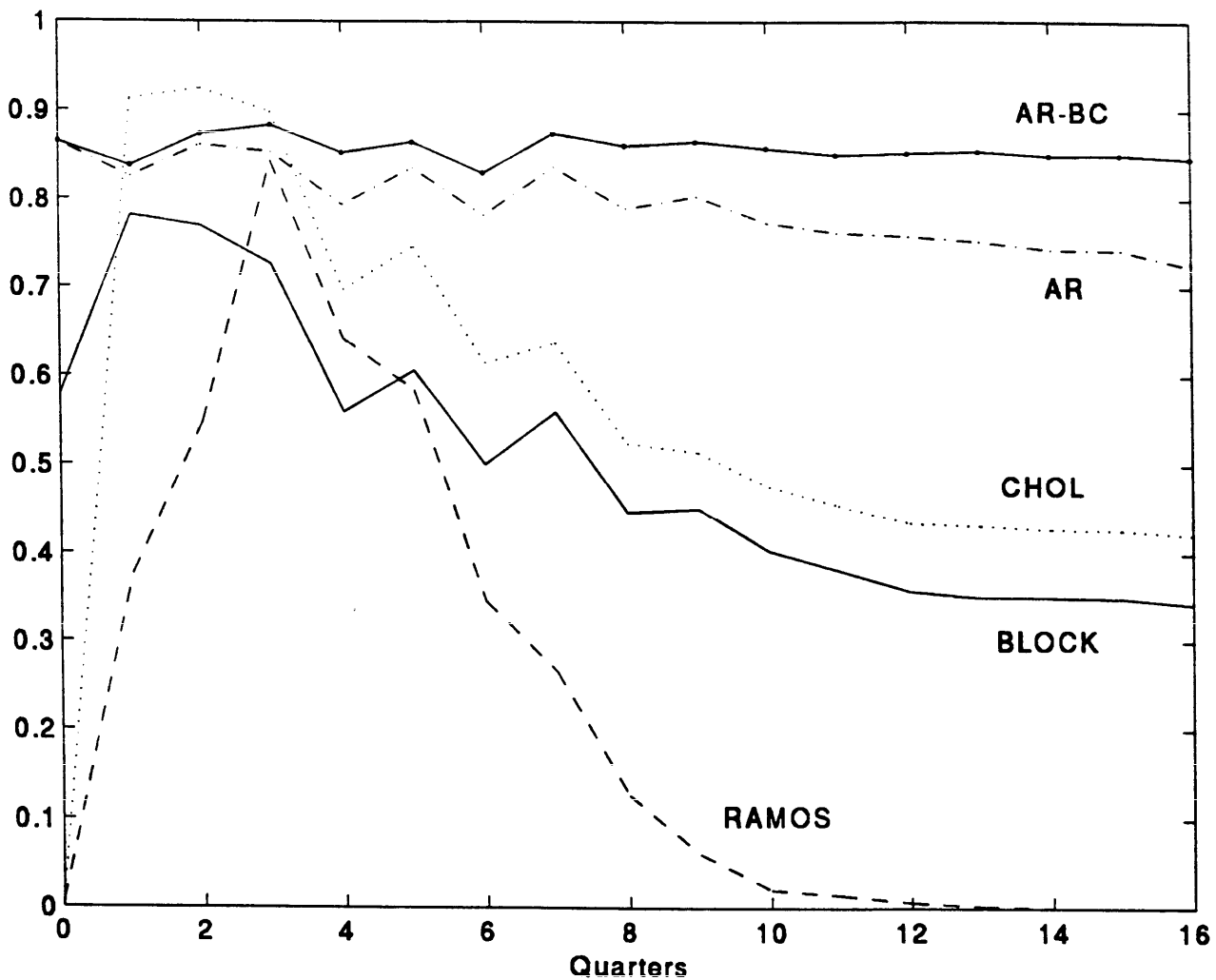
Notes: Quarterly S&P common stock earnings-price ratio, 1947.2 to 1994.3. For each block size, k , mean squared deviation from nominal coverage is defined as $\frac{1}{17} \sum_{i=0}^{16} [\text{coverage}(k,i) - 0.9]^2$, where $i = 0, \dots, 16$ denotes the time horizon of the impulse response.

Figure 2
Effective Coverage of Nominal 90 %
Bootstrap Confidence Intervals for Impulse Responses
T = 80



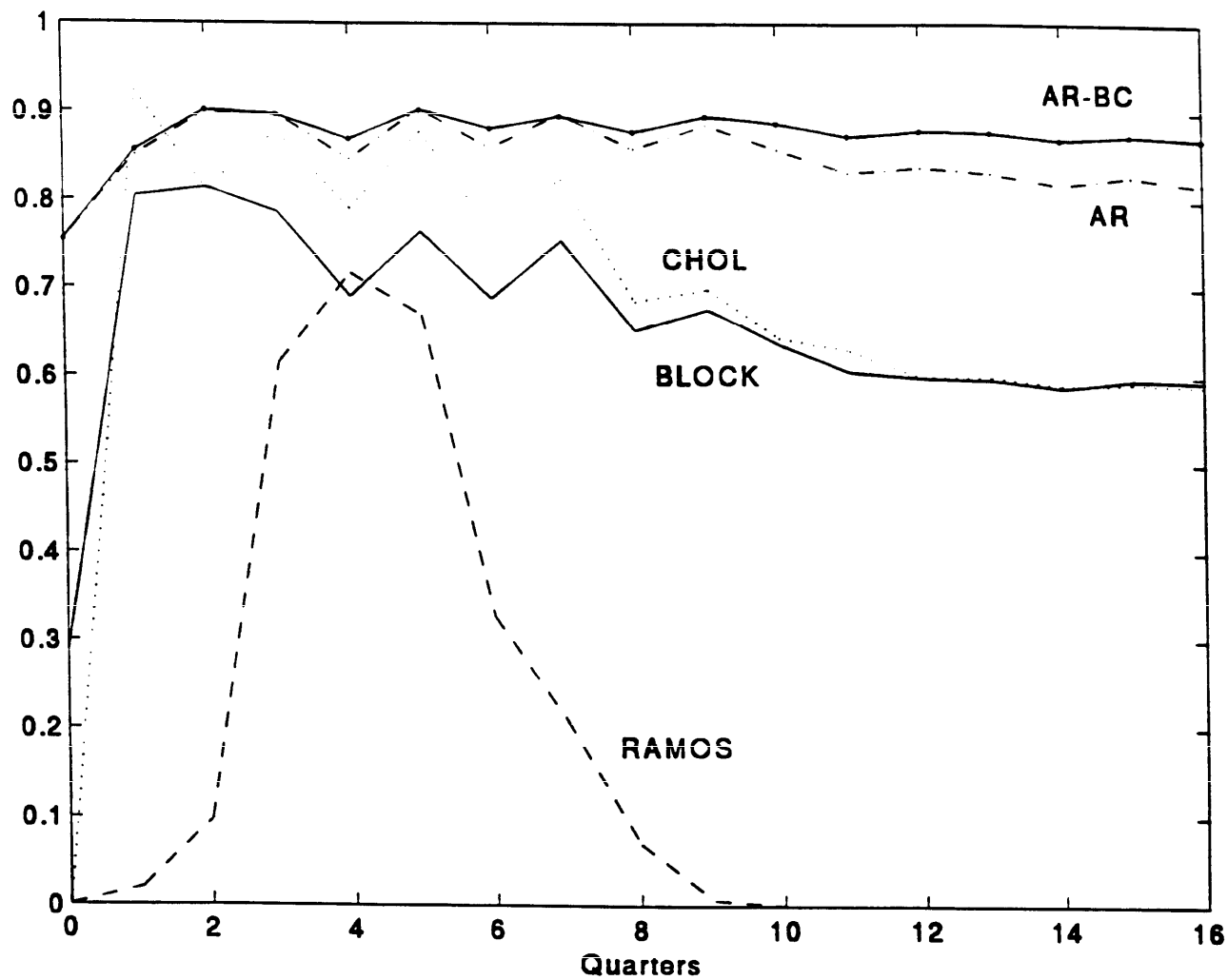
Source: 1,000 Monte Carlo trials for ARMA(2,4)-DGP based on U.S. T-Bill rate for 1971.4-1993.4.

Figure 3
Effective Coverage of Nominal 90 %
Bootstrap Confidence Intervals for Impulse Responses
T = 160



Source: 1,000 Monte Carlo trials for ARMA(2,4)-DGP based on U.S. T-Bill rate for 1971.4-1993.4.

Figure 4
Effective Coverage of Nominal 90 %
Bootstrap Confidence Intervals for Impulse Responses
T = 480



Source: 1,000 Monte Carlo trials for ARMA(2,4)-DGP based on U.S. T-Bill rate for 1971.4-1993.4.