

# **EFFICIENCY AND ROBUSTNESS IN SUBSAMPLING FOR DEPENDENT DATA**

Michael Sherman\*  
Texas A&M University

## Abstract:

There are two main approaches to subsampling for dependent data: model based and model free. In the former the dependence structure is modelled in terms of a few unknown parameters and independent errors. In the latter the observed series is divided into blocks, and these blocks are used to capture the dependence in the original series. It is reasonable to assume that the model based approach is superior under the correct model and inferior under misspecification of the correct model. We formalize this tradeoff between efficiency and robustness and examine how it plays out in the Autoregressive-Moving Average class of models.

Keywords: Resampling, Dependence, Nonparametric, Model-Free, Time-Series, Bootstrap

---

\* Michael Sherman is Assistant Professor in the Department of Statistics, Texas A&M University, College Station, TX 77843

## 1: INTRODUCTION

There are two principle approaches to subsampling for dependent sequences: model based and model free. In the model based approach the dependence structure is modelled in terms of a few unknown parameters and independent errors. For example, in an autoregressive model, the autoregressive-moving average (ARMA) parameters are estimated (e.g., by least squares), yielding residuals. The residuals can then be bootstrapped, and pseudo time series can be formed from the estimated model. Some examples of this approach from the literature are the cases of autoregression (Bose, 1988), moving average (Bose, 1990), linear dynamic models (Freedman, 1984), prediction intervals (Thombs, 1990), explosive autoregression, (Basawa et al., 1989) and unstable autoregression (Datta, 1996).

In the model free approach the observed series is divided up into blocks, and these blocks are used to capture the dependence in the original series. There has been a rapidly growing literature on these methods. The two main model free approaches are the Blockwise Bootstrap (BB) and the Subseries approach (SA). In the BB, blocks are resampled and “glued” together to obtain a series which mimics the original data structure. This method has been studied by Künsch (1989), Liu and Singh (1992a), and has been shown to be second order correct (under modifications) by Götze and Künsch (1990) and Lahiri (1991). In the SA the blocks themselves are taken as replicates of the original data structure. Methods based on the SA have been studied, e.g., by Carlstein (1986), who proved  $L_2$  convergence and studied optimal choice of block length, by Politis and Romano (1994) who constructed general confidence intervals, by Sherman and Carlstein (1996), who use the SA to obtain diagnostic information about the sampling distribution of a statistic, and by Hall and Jing (1994), who show that the SA can be rendered second order correct by a “calibration technique”. There is also a large volume of work on the SA in the operations research literature, e.g., by Meketon and Schmeiser (1984) and Goldsman and Schruben (1984).

The goal in this note is to compare the model based (MB) and model free (MF) procedures. It is reasonable to assume that the MB method is superior under the correct model and inferior under misspecification of the model. We compare by defining “efficiency” and “robustness” in this dependent data setting, and study the efficiency and robustness of the two procedures. It turns out that a large class of MB and MF procedures have similar asymptotic behavior enabling us to label some MB procedures “efficient” and a large class of MF procedures “robust”. This terminology in subsampling procedures is due to Liu and Singh (1992b) who studied resampling algorithms in the case of linear regression models with independent errors. The efficient algorithms exploit the homoskedasticity of the errors, while the robust algorithms allow for heteroskedasticity. In our setting the efficient MB algorithms exploit the “known” dependence structure, while the MF algorithms work reasonably well for a large class of stationary dependence structures.

Within the  $ARMA(p, q)$  class of models we are able to see the tradeoff between robustness and

efficiency in an important special case. Theoretical and numerical results indicate that MB methods need long series to realize their efficiency and that this efficiency can be erased by relatively “little” misspecification.

Beyond misspecification there is an additional problem with MB methods: it is not always clear what the correct MB algorithm is. For example, Bose (1988, 1990) presents very different resampling algorithms in the autoregressive and moving average settings. The algorithm in the moving average case requires theoretical development. Thus, in addition to misspecification of the dependence structure, there is the problem of determining the correct algorithm to employ. In contrast, the MF methods can be used in an omnibus fashion without regard to the specific dependence structure.

In some cases, the parameter of interest is a parameter in the model and it may appear that the only approach is a MB one. It is reasonable, however, that after estimating the model parameter we may want to gauge the estimate’s variability using a MF method. In any case, this can at least give a check to the accuracy of a MB method.

In Section 2, we introduce the estimators we consider and define robustness and efficiency in this context and discuss the various estimators. Section 3 discusses a particular example in detail, while in Section 4 we consider contiguous alternatives to an assumed model. All derivations are in the Appendix.

## 2: MODEL BASED AND MODEL FREE ESTIMATORS OF VARIANCE

We observe the sequence  $\{X_i : 1 \leq i \leq n\}$ , assumed to come from a stationary time series, and compute a statistic of interest  $s_n := s_n(X_1, \dots, X_n)$ , which estimates a parameter,  $\theta$ . We are interested in the distribution of  $s_n$ , and in particular  $Var(s_n)$ , or the standardized limit of this variance,  $V$ , where  $nVar(s_n) \rightarrow V$ .

An example of MB subsampling is the following. Assume the observations are generated by the AR(1) process:

$$X_i = \phi X_{i-1} + \epsilon_i,$$

where  $\{\epsilon_i\}$  are *i.i.d.* according to some distribution function  $F$  with  $Var(\epsilon) = \sigma^2 = 1$ , without loss of generality.

Let  $\hat{\phi}$  be the least squares estimator of  $\phi$ ,  $\hat{\epsilon}_1 = X_1$ ,  $\hat{\epsilon}_i = X_i - \hat{\phi}X_{i-1}$ ,  $i = 2, \dots, n$ , the resulting residuals, and  $\tilde{\epsilon}_i := \hat{\epsilon}_i - \bar{\hat{\epsilon}}$ , the centered residuals. Let  $\{\tilde{\epsilon}_i^*\}$  be *i.i.d.* from  $\tilde{F}_n$ , the empirical distribution function of  $\tilde{\epsilon}_1, \dots, \tilde{\epsilon}_n$ . Then, the MB bootstrapped series is:

$$X_1^* = \tilde{\epsilon}_1^*, \quad X_i^* = \hat{\phi}X_{i-1}^* + \tilde{\epsilon}_i^*,$$

for  $i = 2, \dots, n$ , with corresponding bootstrap replicate  $s_n^*$ . Repeating this operation  $B$  times gives the MB estimator of  $V$ :

$$\hat{V}_{MB} = n \sum_{j=1}^B (s_n^{*j} - \bar{s}_n^*)^2 / B, \quad (1)$$

where  $\bar{s}_n^* = \sum_{j=1}^B s_n^{*j} / B$ .

The validity of this procedure has been established in the case where  $s_n = \hat{\phi}$  by Bose (1988) (who showed higher order accuracy for the stationary case, ( $|\phi| < 1$ )) when estimating the distribution of  $\hat{\phi}$ . Consistency of this algorithm has also been demonstrated by Basawa et al. (1989) in the explosive case ( $|\phi| > 1$ ), and by Basawa et al. (1991) and Datta (1996) in the unstable case ( $|\phi| = 1$ ).

Let  $B_b^m := (X_m, \dots, X_{m+b-1})$ ,  $1 \leq m \leq n - b + 1 =: N$ , denote the ‘‘subseries’’ or ‘‘blocks’’ of length  $b$  starting with the  $m$ 'th observation. The important point is that for each  $m$ ,  $B_b^m$  maintains the same dependence structure as the original sequence, whatever that structure is. Thus, it seems that from the corresponding subseries replicates,  $s_b^m := s_b(B_b^m)$ , we can form an estimator of  $nVar(s_n)$ .

A natural MF estimator of  $nVar(s_n)$ , based on the use of all possible subseries of length  $b$ , is:

$$\hat{V}_O := b \sum_{i=1}^N \frac{(s_b^i - \bar{s}_n)^2}{N}, \quad (2)$$

where  $\bar{s}_n := \sum_{i=1}^N s_b^i / N$ .

It should be noted that various authors use slightly different constants in the definition of  $\hat{V}_O$ , but all are reasonably close for small values of the ratio  $b/n$ .

To compare the Efficiency and Robustness of  $\hat{V}_{MB}$  and of  $\hat{V}_O$  we give the following definitions: An estimator of  $V$ ,  $\hat{V}_E$ , is said to be efficient if

$$n^{1/2}(\hat{V}_E - V) \xrightarrow{D} M, \quad (a)$$

where  $E(M) = 0$ , with  $Var(M) < \infty$ , and some known (up to a finite number of parameters which can be estimated  $n^{1/2}$  consistently) underlying dependence structure and independent errors.

An estimator of  $V$ ,  $\hat{V}_R$ , is said to be robust if

$$n^{1/3}(\hat{V}_R - V) = O_p(1) \quad (b)$$

for any underlying dependence structure satisfying general conditions on the strength of dependence. In order to clarify these definitions consider the following:

Example:

Consider the statistic  $s_n = \bar{X}_n$  computed from the family of processes  $\{X_i\}$  such that

$$\gamma_1 := \sum_{i=-\infty}^{\infty} i\gamma(i) := \sum_{i=-\infty}^{\infty} iCov(X_0, X_i) < \infty.$$

Song and Schmeiser (1995) show that for a large class of MF estimators,  $\hat{V}_{MF}$ , it holds that

$$Bias(\hat{V}_{MF}) = -(1/b)c_b\gamma_1 + o(1/b),$$

and that

$$Var(\hat{V}_{MF}) = (b/n)c_vV^2 + o(b/n),$$

where  $c_b$  and  $c_v$  are constants that depend on the specific estimator being used. Thus:

$$b_{opt} := \left(\frac{2c_b^2\gamma_1^2}{c_vV^2}\right)^{1/3}n^{1/3}, \quad (3)$$

and if  $b = Cn^{1/3}$  for some constant  $C$  then we have

$$n^{1/3}(\hat{V}_{MF} - V) = O_p(1).$$

The following estimators are robust whenever  $b$  is chosen to be a constant times  $n^{1/3}$ :

1. (b) holds with  $\hat{V}_R = \hat{V}_O$  defined in equation (2) with  $c_b = 1$  and  $c_v = 4/3$  by Meketon and Schmeiser (1984), see also Cox and Lewis (1961).
2. Assume  $n = bk$  for some integer  $k$ . Choose blocks  $B_1^*, \dots, B_k^*$  by resampling randomly, with replacement, from among  $B_b^m, b = 1, \dots, N$ . Let  $B_i^*$  denote the  $i$ 'th sampled block,  $(X_{i1}^*, \dots, X_{ib}^*)$ , the block bootstrap series is  $\xi^* := (X_{11}^*, \dots, X_{1b}^*, X_{21}^*, \dots, X_{2b}^*, \dots, X_{k1}^*, \dots, X_{kb}^*)$  and  $s(\xi^*)$  is the corresponding replicate. If this is performed  $B$  times then the block bootstrap variance estimator of  $V$  is:

$$\hat{V}_{BB} := n \sum_{j=1}^B \frac{(s(\xi_j^*) - \bar{s})^2}{B},$$

where  $\bar{s} := \sum_{j=1}^B s(\xi_j^*)/B$ . (b) holds for  $\hat{V}_R = \hat{V}_{BB}$ , with  $c_b = 1$  and  $c_v = 4/3$ , by Kunsch (1989).

3. (b) holds with  $\hat{V}_R = \hat{V}_N$ , the estimator based on using all  $k$  nonoverlapping blocks of length  $b$ , i.e.,

$$\hat{V}_N := b \sum_{i=1}^k \frac{(s_b^{ib} - \bar{s})^2}{k},$$

where  $\bar{s} := \sum_{i=1}^k s_b^{ib}/k$ . In this case  $c_b = 1$  and  $c_v = 2$  by Carlstein (1986).

4. (b) holds for estimators based on standardized time series with  $c_b = 3$  and  $c_v = 2$ , by Schruben (1983) while a method combining  $\hat{V}_N$  and standardized time series satisfies (b) with  $c_b = 2$  and  $c_v = 1$ , by Goldsman and Schruben (1984).

There is some debate as to which of these estimators is preferable. Due to its relatively low bias and variance constants,  $c_b$  and  $c_v$ , and to its ease of implementation we consider only  $\hat{V}_O$  in the sequel.

To see which estimators are efficient consider the case of an assumed  $ARMA(p, q)$  model where  $p$  and  $q$  are correctly specified, and the  $p + q$  parameters are estimated by Maximum Likelihood. Let  $\theta$  denote the  $p + q$  vector of parameters. The parameter estimate  $\hat{\theta}$  is a  $n^{1/2}$  consistent estimator of  $\theta$  (e.g., Brockwell and Davis, 1987, Chapter 8), and thus if  $V = h(\theta)$  for a known differentiable function  $h(\cdot)$  then  $\hat{V} := h(\hat{\theta})$  is an efficient estimator of  $V$ . E.g., in an AR(1) model with parameter  $\phi$  we have  $V = (1 - \phi)^{-2}$  and thus  $\hat{V} := (1 - \hat{\phi})^{-2}$  is an efficient estimator of  $V$ . We show (in Appendix III) that the MB estimator defined in (1) satisfies (a) with  $\hat{V}_E = \hat{V}_{MB}$  when the MB estimator of variance is computed from all possible subsamples, i.e.,  $B = n^n$ .

There are two issues raised by the above definitions and discussion: 1) For what sample sizes does the efficiency of an estimator really payoff and 2) How wrong can the assumed model be before the MB estimator loses out to a MF estimator. In the next section we study this question in detail within the ARMA(1,1) class of models.

We note that in the independent data setting, the properties of Maximum Likelihood estimators under misspecification have been studied by, e.g., McCulloch (1988) and White (1982).

### 3: ROBUSTNESS AND EFFICIENCY IN THE ARMA(1,1) CLASS

Consider the ARMA(1,1) model:

$$X_i - \phi X_{i-1} = \epsilon_i + \theta \epsilon_{i-1},$$

where  $|\phi| < 1$ ,  $|\theta| < 1$ ,  $E(\epsilon_i) = 0$ , and  $E(\epsilon_i^2) = 1$ , without loss of generality. In this case, it is shown in Appendix I that  $V = ((1 + \theta)/(1 - \phi))^2$ .

We now state the behavior of the MB bootstrap under the assumption that the process generating the data is ARMA(1,0)=AR(1), i.e.,  $\theta = 0$ . Under the MB algorithm discussed in Section 2 we show (Appendix II):

$$\hat{V}_{MB} = nVar_*(\bar{X}_n^*) = (1 - \hat{\phi})^{-2} \sum_{i=1}^n (\hat{\epsilon}_i^2/n) + O_p(n^{-1}),$$

where  $nVar_*(\cdot)$  denotes (1) when  $B = n^n$ , i.e., all possible bootstrap samples are employed. Appendix III demonstrates that (a) holds with  $\hat{V}_E = \hat{V}_{MB}$  under an AR(1) model, i.e., when  $\theta = 0$ .

In general, we show in Appendix IV that under an assumed AR(1) model:

$$nVar(\bar{X}_n^*) \xrightarrow{p} \{1 + \theta^2 - \theta^2/Var(X)\} \{1 - \phi - \theta/Var(X)\}^{-2}. \quad (4)$$

We see that this is equal to  $V = (1 - \phi)^{-2}$  if  $\theta = 0$ , in which case  $\hat{V}_{MB}$  is efficient. Note that when  $\theta \neq 0$ ,  $\hat{V}_{MB}$  is not only not efficient, but fails to be robust, and in fact  $\hat{V}_{MB}$  does not tend to  $V$  as  $n \rightarrow \infty$ . Table 1 gives the values of  $V$  and  $plim \hat{V}_{MB} =: \tilde{V}$  for various values of  $\phi$  and  $\theta$ , as well as the ratio of the two. We see that for fixed  $\phi$ , as  $\theta$  increases the MB bootstrap performs worse, while for fixed  $\theta$  the MB bootstrap performs worse as  $\phi$  increases. The former is not surprising while the latter point is somewhat distressing. A small value of  $\theta$ , e.g.  $\theta = .2$ , introduces significant bias to  $\hat{V}_{MB}$  when  $\phi = .5$ .

It is reasonable to raise the question as to how likely it is to misspecify an ARMA(1,1) model as an AR(1) model. If it is unlikely to make this error then there is not much point in determining the results of this type of misspecification. The likelihood of incorrectly specifying the model depends upon (among other things) the user modelling the data as well as the energy the user devotes to the task of identification. A modest approach to modelling time series data is to plot the autocovariance function (ACF) and the partial autocovariance function (PACF) as discussed in Example 8.2.1 of Brockwell and Davis (1987). Given in Figures 1 and 2 are the ACF's and PACF's for each of 6 time series of length  $n = 200$ . These series were shown to several people knowledgeable in time series methodology and many did not find the plots strongly incompatible with data generated by an AR(1) process. The fact is that three of the plots were generated by an AR(1) process, while the remaining three were generated by an ARMA(1,1) model with  $\theta = .2$  (in all cases  $\phi = .5$ ). Even with this information, not a single viewer of the plots could correctly identify the six series. If knowledgeable users cannot correctly identify these models, then this type of misspecification may be fairly common. For longer series lengths, e.g.,  $n = 1000$ , it will become easier to detect a fixed misspecification. For this reason in Section 4, we study the behavior of  $\hat{V}_{MB}$  under contiguous alternatives.

In order to employ  $\hat{V}_O$  a choice of the block length,  $b$ , is necessary. The best choice of  $b$  is given in (3), but depends on the unknown object of interest,  $V$ . A "hybrid" MB-MF approach to choice of block length was proposed by Carlstein (1986): Under an AR(1) model (e.g.) one can explicitly calculate the asymptotically optimal  $b$  as a function of the AR(1) parameter  $\phi$  (see Appendix VI for details in the ARMA(1,1) case). Plugging in an estimator of  $\phi$  then gives a choice of  $b$ , from which  $V$  can be estimated. This is a MB method of choosing  $b$  but a MF variance estimator is employed using that chosen  $b$ . Recently, more MF choices of  $b$  have been proposed by Hall, Horowitz, and Jing (1995) and Song and Schmeiser (1995). The method of the latter works only when  $s_n = \bar{X}_n$  and requires knowledge of the bias and variance constants  $c_b$  and  $c_v$  which depend on the specific estimator being used, while the former applies without change for more general statistics and for a large class of estimators. For these reasons we use the databased choice of Hall, Horowitz, and Jing (1995). They proposed the algorithm for the estimator  $\hat{V}_{BB}$ . In Appendix V we give an analogous algorithm for the estimator  $\hat{V}_O$ .

We now examine the finite sample behavior of MB and MF estimators, both under correct specification of the model and under misspecification. Table 2 gives the comparison between the MF and MB estimators of  $V$  in the case when  $\theta = 0$  for three different values of the AR(1) parameter  $\phi$ , .2, .5, .8, and two different choices of  $B =$  the number of bootstrap replications. The target values,  $V$  are given in the first four rows of Table 1. The optimal choice of block length,  $b$ , for the MF method are for  $n = 200$ ,  $b = 4, 8$ , and 18, for  $\phi = .2, .5, .8$ , and for  $n = 1000$ ,  $b = 6, 14, 31$ . Empirical evidence shows that the databased estimate is typically negatively biased (for positive  $\phi$ ) for the optimal  $b$ . Each row is based on the results of 1000 simulations.

It can be seen from these simulations that under the correct model the MB estimator of  $V$  is very close to unbiased for  $V$  while the MF method has a significant negative bias. The bias, however, decreases for the larger series length,  $n = 1000$ , as does the variance in approximate accordance with what is predicted. What is somewhat surprising, however, is that although the MB estimator has a faster rate of convergence to  $V$  (due to its efficiency) the series of length  $n = 200$  does not demonstrate this at all, and the effect of the efficiency is still relatively weak for  $n = 1000$ . The MB estimator is more variable in all cases, although for  $n = 1000$  the MB is relatively more stable and thus the asymptotics are starting to take hold. Even letting  $B \rightarrow \infty$  would not give the MB estimator smaller variability than the MF estimator in the case where  $n = 200$ . This can be seen by assuming that the error variance  $\sigma^2 = 1$  is estimated perfectly, computing the asymptotic variance of the estimator  $(1 - \hat{\phi})^{-2}$  using the delta method and the knowledge that  $nVar(\hat{\phi}) \rightarrow 1 - \phi^2$ .

We next consider the robustness of the MF method under misspecification. In all cases the MB method assumes an AR(1) model for the data. We first compute the MF estimators. The results are given in Table 3. The MF estimator performs as well under the ARMA(1,1) model as under the AR(1) model. This is to be expected as neither the choice of block length,  $b$ , nor the estimator itself is based on the underlying dependence structure. Further, as predicted, both the bias and the variance of the MF estimator shrink for the larger series length,  $n = 1000$ . Table 4 gives the results for the MB estimator under misspecification. It can be seen that the MB estimator performs more poorly under misspecification as was to be expected from equation (4). Note that  $\theta = .2$  is enough misspecification to make the MB bias comparable to that of the MF method, while the MB estimator is much more variable, even for  $n = 1000$ .

#### 4: Contiguous Alternatives

It is useful to consider the relative performance of MB procedures to MF procedures under alternatives which are “close” to the true model, i.e., contiguous alternatives. Consider the ARMA(1,1) model where for each  $i$ ,  $1 \leq i \leq n$ , the observations are generated from the model:

$$X_i - \phi X_{i-1} = \epsilon_i + \theta_n \epsilon_{i-1},$$

where the errors,  $\epsilon_i$ , are as described at the beginning of Section 3, and  $\theta_n := \theta/n^{1/2}$ . The limiting

variance,  $V$ , is still equal to  $(1 - \phi)^{-2}$  in this case. We show (Appendix VII) that

$$n^{1/2}(\hat{V}_{MB} - V) \xrightarrow{D} M + N(2\theta(1 + \phi)(1 - \phi)^{-2}, 4(1 + \phi)(1 - \phi)^{-5}), \quad (5)$$

where,  $E(M) = 0$ , and thus  $\hat{V}_{MB}$  does not satisfy (a) under contiguous alternatives. Note, however, that  $\hat{V}_{MB}$  does satisfy (b) under this type of misspecification, and is in fact more efficient than any of the MF estimators discussed in Section 2 (although the results of Section 3 indicate that long series are necessary to realize this increased efficiency). These results contrast with those in the independence setup where a large class of estimators satisfy (a). Having possibly arbitrary dependence structure leads us to redefine what is reasonable to require of an estimator.

Conclusion:

We have defined efficiency and robustness for subsampling variance estimators under dependence. We have studied model based and model free subsampling algorithms in both the case where the model is correctly specified and where the model is misspecified. We have seen that within the  $ARMA(p, q)$  class of models, for modest series length ( $n = 200$ ), little if anything is gained by using a model based estimator under the correct model, although the gains increase as the series length increases ( $n = 1000$ ). Under mild misspecification, model based variance estimators have comparable bias to model free estimators while being much more variable. Nevertheless, we see that under contiguous alternatives, model based estimators are still asymptotically superior to model free estimators, although very long series are necessary to realize this superiority.

### Appendix: Derivation of Results

I: *Derivation of the limit  $V$  in the ARMA(1,1) setting:*

As in Brockwell and Davis (1987) (BD), let  $z^k \epsilon_i := \epsilon_{i-k}$  denote the backward shift operator.

Then:

$$X_i = (1 + \theta z) \sum_{k=1}^{\infty} \phi^k z^k \epsilon_i =: \sum_{k=0}^{\infty} \psi_k \epsilon_{i-k}, \text{ where}$$

$$\psi_0 = 1, \psi_i = (1 + \theta/\phi)\phi^i, \text{ for } i > 0,$$

and thus by BD, Theorem 7.1.2, we have:

$$V = \left( \sum_{i=0}^{\infty} \psi_i \right)^2 \text{Var}(\epsilon_i) = (1 + \theta)^2 / (1 - \phi)^2.$$

II: *Derivation of the MB Bootstrap estimate of  $V$ :*

Using the facts that  $E_*(\tilde{\epsilon}_i^*) = 0$ , and  $E_*(X_i^* X_{i+h}^*) = \hat{\phi}^h \sum_{j=0}^{i-1} \hat{\phi}^{2j} \sum_{i=1}^n (\tilde{\epsilon}_i^2/n)$ , and using arithmetic, it follows that  $n \text{Var}(\bar{X}_n^*) = (1 - \hat{\phi})^{-2} \sum_{i=1}^n (\tilde{\epsilon}_i^2/n) + O_p(n^{-1})$ .

III: *Efficiency of the MB estimator of variance under an AR(1) model:*

By BD, Example 8.8.1 we have:

A)

$$n^{1/2}(\hat{\phi} - \phi) \xrightarrow{D} N(0, (1 - \phi^2))$$

and thus, using Taylor Expansion we have that

$$n^{1/2}((1 - \hat{\phi})^{-2} - (1 - \phi)^{-2}) \xrightarrow{D} N(0, (4(1 + \phi)/(1 - \phi)^5)).$$

B) To show that  $n^{1/2}(\sum_{i=1}^n (\tilde{\epsilon}_i^2/n) - 1) \xrightarrow{D} M$ , with  $E(M) = 0$  note that  $n^{1/2}[\sum_{i=1}^n (\tilde{\epsilon}_i^2/n) - 1] = n^{1/2}[\sum_{i=1}^n \hat{\epsilon}_i^2/n - \sum_{i=1}^n \epsilon_i^2/n] + n^{1/2}[\sum_{i=1}^n (\epsilon_i^2/n) - 1] - n^{1/2}[\bar{\epsilon}^2] = [1] + [2] + [3]$ , say.

[1] =  $-2n^{1/2}(\hat{\phi} - \phi) \sum_{i=1}^n X_i X_{i-1}/n + n^{1/2}(\hat{\phi}^2 - \phi^2) \sum_{i=1}^n X_{i-1}^2/n$ . Each of the preceding terms converges to a zero mean normal distribution by the CLT and by the Delta Method.

[2] converges to a mean zero random variable by the Central Limit Theorem, while algebraic manipulations show that  $\bar{\epsilon} = (1 - \hat{\phi})\bar{X}_n + (\hat{\phi}X_n)/n$ , and thus [3] =  $O_p(n^{-1/2})$ . Combining A and B show that (a) holds and thus that  $\hat{V}_{MB}$  is efficient.

IV: *Inconsistency of MB bootstrap estimator of variance under misspecification:*

$\hat{\phi} = \phi + \sum_{i=1}^{n-1} (\theta \epsilon_i + \epsilon_{i+1})(X_{i-1} \phi + \theta \epsilon_{i-1} + \epsilon_i) / \sum_{i=1}^{n-1} X_{i-1}^2 \xrightarrow{p} \phi + \theta / \text{Var}(X)$ , by the ergodic theorem, where

$$\text{Var}(X) = \sum_{i=0}^{\infty} \psi_i^2 = (1 + (\theta + \phi)^2)/(1 - \phi^2).$$

Using similar algebraic manipulations and the ergodic theorem shows that:

$$\sum_{i=1}^n \hat{\epsilon}_i^2/n \xrightarrow{p} 1 + \theta^2 - \theta^2/\text{Var}(X), \text{ and (using the results of Appendices II and III)}$$

$$n\text{Var}_*(\bar{X}_n^*) \xrightarrow{p} (1 + \theta^2 - \theta^2/\text{Var}(X))(1 - \phi - \theta/\text{Var}(X))^{-2}, \text{ and thus that (4) holds.}$$

V: *Algorithm for choosing subseries length for the MF estimator:* (Hall, Horowitz, and Jing, 1995)

- 1) Choose a pilot value for  $b$ ,  $b = b^*$ , say, and calculate  $\hat{V}_O$  using  $b = b^*$ .
- 2) For some  $m$ , consider  $X_m^i, i = 1, \dots, (n - m + 1)$ , all possible series of length  $m$ . For the  $i$ 'th series of length  $m$ , let  $\hat{V}_{m^*}^i$  denote the variance estimator computed from the series  $X_m^i$  using a block length of  $m^*$ , and define:

$$\hat{b}_m = \text{argmin}_{m^*} \sum_{i=1}^{n-m+1} (\hat{V}_{m^*}^i - \hat{V}_O)^2 / (n - m + 1).$$

This is the empirical estimate of the  $b_m$  that minimizes  $MSE(\hat{V}_O)$  for a sequence of length  $m$ .

- 3) Compute  $\hat{b}_n = (n/m)^{1/3} \hat{b}_m$ .

The algorithm could be iterated if desired. In other words, we could set  $b^*$  in step 1) equal to  $\hat{b}_n$  in step 3) and again obtain  $\hat{b}_m$  in step 2, and thus a new  $\hat{b}_n$ . In the simulations the algorithm converged in the single iteration approximately 90 percent of the time, and for this reason no iteration was performed. In all cases we used an initial  $b^* = n/10$  and  $m = n/20$ . Results appear to be relatively insensitive to these choices.

VI: *Optimal block length in the ARMA(1,1) setting:*

We illustrate computations of optimal block length for the MF estimator defined in (2). We need the constants in the definition of  $b_{opt}$  in (3):

$$c_b = 1, c_v = 4/3, \text{ and } V = (1 + \theta)^2/(1 - \phi)^2.$$

Finally,

$$\gamma(k) = \sum_{j=0}^{\infty} \psi_j \psi_{j+k} = \{(\phi + \theta)^2/(1 - \phi^2) + (1 + \theta/\phi)\} \phi^k \text{ and thus,}$$

$$\gamma_1 := 2 \sum_{k=1}^{\infty} \gamma(k) = 2(\phi + \theta)(1 + \phi)/(1 - \phi)^3.$$

Putting these into the formula for  $b_{opt}$  in (3) gives:

$$b_{opt} := \{(6(\phi + \theta)^2/((1 + \theta)^4(1 - \phi^2)^2)\}^{1/3} n^{1/3}.$$

It is interesting to note that for fixed  $\theta$  the optimal block length increases as  $\phi$  increases, but that for fixed  $\phi$ , the block length is concave in  $\theta$ . This is because the ratio  $V/\text{Var}(X_1)$  is not monotone in  $\theta$ .

VII: *Nonefficiency of MB under contiguous alternatives:*

Using the expression for  $\hat{\phi}$  in the first line of IV, we find that  $n^{1/2}(\hat{\phi} - \phi) \xrightarrow{D} N(\theta(1 - \phi^2), (1 - \phi^2))$  and thus using Taylor expansion we see that

$$n^{1/2}((1 - \hat{\phi})^{-2} - (1 - \phi)^{-2}) \xrightarrow{D} N(2\theta(1 + \phi)(1 - \phi)^{-2}, (4(1 + \phi)/(1 - \phi)^5)).$$

Using the decomposition in IIIB we find that [1]  $\xrightarrow{D} \text{Var}(X)[-2\phi N(\theta(1 - \phi^2), 1 - \phi^2) + N(2\phi\theta(1 - \phi^2), 4\phi^2(1 - \phi^2))]$ , [2]  $\xrightarrow{D} N(0, 1)$ , and [3]  $= O_p(n^{-1/2})$ . Combining these results gives (5).

Table 1: Aymptotic Variance of the Sample Mean and Model Based Bootstrap Estimate Under an AR(1) Assumption

		$V$	$\tilde{V}$	Ratio
$\theta=0$	$\phi = 0$	1.00	1.00	1.00
	$\phi = .2$	1.56	1.56	1.00
	$\phi = .5$	4.00	4.00	1.00
	$\phi = .8$	25.0	25.0	1.00
$\theta=.2$	$\phi = 0$	1.44	1.53	0.94
	$\phi = .2$	2.25	2.55	0.89
	$\phi = .5$	5.76	7.07	0.82
	$\phi = .8$	36.0	47.6	0.76
$\theta=.5$	$\phi = 0$	2.25	2.92	0.77
	$\phi = .2$	3.52	4.93	0.71
	$\phi = .5$	9.00	14.0	0.64
	$\phi = .8$	56.3	95.8	0.59
$\theta=.8$	$\phi = 0$	3.24	4.76	0.68
	$\phi = .2$	5.06	7.96	0.64
	$\phi = .5$	13.0	22.4	0.58
	$\phi = .8$	81.0	152	0.53

Table 2: Model Based and Model Free Variance Estimates Under an AR(1) Model

		Model Free		
		Mean	Var.	MSE
n=200	$\phi = .2$	1.39(.01)	.060	.090
	$\phi = .5$	3.12(.02)	.531	1.30
	$\phi = .8$	16.3(.19)	35.9	111
n=1000	$\phi = .2$	1.47(.01)	.022	.030
	$\phi = .5$	3.56(.02)	.237	.431
	$\phi = .8$	20.4(.13)	17.2	38.4
		Model Based (B=100)		
		Mean	Var.	MSE
n=200	$\phi = .2$	1.57(.01)	.148	.148
	$\phi = .5$	4.03(.04)	1.47	1.47
	$\phi = .8$	25.4(.38)	143	143
n=1000	$\phi = .2$	1.54(.01)	.065	.065
	$\phi = .5$	3.95(.02)	.527	.527
	$\phi = .8$	24.7(.19)	35.2	35.3
		Model Based (B=1000)		
		Mean	Var.	MSE
n=200	$\phi = .2$	1.57(.01)	.098	.098
	$\phi = .5$	4.13(.03)	1.13	1.15
	$\phi = .8$	25.4(.35)	124	124
n=1000	$\phi = .2$	1.56(.01)	.025	.025
	$\phi = .5$	4.01(.02)	.274	.274
	$\phi = .8$	25.0(.17)	28.0	28.0

Table 3: Model Free Variance Estimates  
in the ARMA(1,1) Model

			Mean	Var.	MSE
n=200	$\theta=.2$	$\phi = 0$	1.31	.0414	.0583
		$\phi = .2$	1.99	.165	.233
		$\phi = .5$	4.74	1.68	2.72
		$\phi = .8$	26.8	136	221
n=200	$\theta=.5$	$\phi = 0$	2.01	.135	.193
		$\phi = .2$	3.02	.400	.650
		$\phi = .5$	7.11	3.41	6.98
		$\phi = .8$	36.6	196	584
n=200	$\theta=.8$	$\phi = 0$	2.85	.277	.429
		$\phi = .2$	4.31	.821	1.38
		$\phi = .5$	9.78	5.71	16.1
		$\phi = .8$	50.7	352	1270
n=1000	$\theta=.2$	$\phi = 0$	1.36	.0135	.0200
		$\phi = .2$	2.05	.0440	.0840
		$\phi = .5$	5.07	.479	.955
		$\phi = .8$	29.7	38.2	77.9
n=1000	$\theta=.5$	$\phi = 0$	2.08	.0395	.0684
		$\phi = .2$	3.20	.121	.223
		$\phi = .5$	7.78	1.05	2.54
		$\phi = .8$	43.2	64.7	236
n=1000	$\theta=.8$	$\phi = 0$	2.97	.0800	.153
		$\phi = .2$	4.59	.250	.471
		$\phi = .5$	11.0	1.94	5.94
		$\phi = .8$	60.1	115	552

Table 4: Model Based Estimates in the ARMA(1,1) Model  
under a AR(1) Assumption (B=1000)

			Mean	Var.	MSE
n=200	$\theta=.2$	$\phi = 0$	1.53	.092	.100
		$\phi = .2$	2.53	.324	.402
		$\phi = .5$	7.04	4.06	5.70
		$\phi = .8$	47.3	452	580
n=200	$\theta=.5$	$\phi = 0$	2.90	.393	.816
		$\phi = .2$	4.90	1.38	3.28
		$\phi = .5$	13.9	17.0	41.1
		$\phi = .8$	94.9	1841	3330
n=200	$\theta=.8$	$\phi = 0$	4.74	1.11	3.36
		$\phi = .2$	7.93	3.72	12.0
		$\phi = .5$	22.3	44.2	131
		$\phi = .8$	150	4616	9516
n=1000	$\theta=.2$	$\phi = 0$	1.54	.0219	.0300
		$\phi = .2$	2.55	.0741	.164
		$\phi = .5$	7.06	.860	2.55
		$\phi = .8$	47.3	88.0	216
n=1000	$\theta=.5$	$\phi = 0$	2.92	.0913	.540
		$\phi = .2$	4.93	.310	2.30
		$\phi = .5$	14.0	3.57	28.6
		$\phi = .8$	95.1	362	1870
n=1000	$\theta=.8$	$\phi = 0$	4.76	.256	2.57
		$\phi = .2$	7.95	.838	9.19
		$\phi = .5$	22.3	9.26	95.8
		$\phi = .8$	151	915	5815

## REFERENCES

- Basawa, I.V., Mallik, A.K., McCormick, W.P., Reeves, J.H., and Taylor, R.L. (1991). 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. (1989). Bootstrapping Explosive Autoregressions, *Annals of Statistics*, 17, 1479-1486.
- 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. Statist. Math.*, 4, 753-768.
- Brockwell, P.J. and Davis, R.A. (1987). *Time Series, Theory and Methods*, Springer Verlag, N.Y.
- 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.
- Cox, D.R. and Lewis (1961). *The Statistical Analysis of Series of Events*, Methuen, London.
- Datta, S. (1996). On Asymptotic Properties of Bootstrap for AR(1) Processes, *Journal of Statistical Planning and Inference*, to appear.
- Freedman, D. (1984). On Bootstrapping Two-Stage Least Squares Estimates in Stationary Linear Models, *Annals of Statistics*, 12, 827-842.
- Goldsman, D. and Schruben, L. (1984). Asymptotic Properties of Some Confidence Interval Estimators for Simulation Output, *Management Science*, 30, 1217-1225.
- Götze, F. and Künsch, H. (1990). Blockwise Bootstrap for Dependent Observations: Higher Order Approximations for Studentized Statistics (Abstract), *Bull. Inst. Math. Statist*, 19, 443.
- Hall, P., Horowitz, J.L., and Jing, B.Y. (1995). On Blocking Rules for the Bootstrap with Dependent Data, *Biometrika*, 82, 561-574.
- Hall, P. and Jing, B.Y. (1994). On Sample Re-use Methods for Dependent Data, *Research Report*, #SR8-94, Centre for Math. and its Appl., Australian National Univ., Canberra.
- Künsch, H. 1989. The Jackknife and Bootstrap for General Stationary Observations, *Annals of Statistics*, 17, 1217-1241.
- Lahiri, S.N. (1991). Second Order Optimality of Stationary Bootstrap, *Statistics and Probability Letters*, 11, 335-341.
- Liu, R. and Singh, K. (1992a). Moving Blocks Jackknife and Bootstrap Capture Weak Dependence, in *Exploring the Limits of Bootstrap*, eds. Lepage, R. and Billard, L., John Wiley and Sons, N.Y.

- Liu, R. and Singh, K. (1992b). Efficiency and Robustness in Resampling, *Annals of Statistics*, 20, 370-384.
- McCulloch, R.E. (1988). Information and the Likelihood Function in Exponential Families, *The American Statistician*, 42,73-75.
- Meketon, M. and Schmeiser, B.M. (1984). Overlapping Batch Means: Something for Nothing? In *Proc. of the Winter Simulation Conference*, 227-230.
- Politis, D.N. and Romano, J.P. (1994) Large-Sample Confidence Regions Based on Subsamples Under Minimal Assumptions, *Annals of Statistics*, 22, 2031-2050.
- Schruben, L. (1983). Confidence Interval Estimation Using Standardized Time Series, *Operations Research*, 31, 1090-1108.
- Sherman, M. and Carlstein, E. (1996). Replicates Histograms, *Journal of the American Statistical Association*, to appear.
- Song, W.T. and Schmeiser, B.W. (1995). Optimal Mean-Squared-Error Batch Sizes, *Management Science*, 41, 110-123.
- Thombs, L.A. and Schucany, W.R. (1990). Bootstrap Prediction Intervals for Autoregression, *Journal of the American Statistical Association*, 85, 486-492.
- White, H. (1982). Maximum Likelihood Estimation of Misspecified Models, *Econometrica*, 50, 1-25.