Working Paper 01-2
Statistics and Econometrics Series 1
January 2001

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34-91) 624-9849

IMPROVED NONPARAMETRIC CONFIDENCE INTERVALS IN TIME SERIES REGRESSIONS

Joseph P. Romano and Michael Wolf*

Abstract

Confidence intervals in time series regressions suffer from notorious coverage problems. This is especially true when the dependence in the data is noticeable and sample sizes are small to moderate, as is often the case in empirical studies. This paper proposes a method that combines prewhitening and the studentized bootstrap. While both prewhitening and the studentized bootstrap each provides improvement over standard normal theory intervals, one can achieve a further improvement by conjoining them in an appropriate way. As a side note, it is stressed that symmetric confidence intervals equal-tailed ones, since they exhibit improved coverage accuracy. We propose concrete ways to deal with the issues of block size, choice of kernel, and choice of bandwidth. The improvements in small sample performance are supported by a simulation study.

Keywords: Bootstrap; confidence intervals; studentization; time series regressions; prewhitening.

*Romano, Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A., Phone: 650-7236326, e-mail: romano@stat.stanford.edu. Wolf, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid. C/ Madrid, 126 28903 Madrid. Spain. Ph: 34-91-624.98.63, Fax: 34-91-624.98.49, e-mail: wolf@est-econ.uc3m.es; Research of the first author funded by the National Science Foundation, DMS 9704487. Research of the second author partly funded by the Spanish "Dirección General de Enseñanza Superior" (DGES), reference number PB98-0025.

Improved Nonparametric Confidence Intervals in Time Series Regressions

Joseph P. Romano *
Department of Statistics
Stanford University
Stanford, CA 94305, U.S.A.

Michael Wolf
Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
28903 Getafe, SPAIN

Working Paper 01-2, Statistics and Econometrics Series 1, January 2001

Abstract

Confidence intervals in time series regressions suffer from notorious coverage problems. This is especially true when the dependence in the data is noticeable and sample sizes are small to moderate, as is often the case in empirical studies. This paper proposes a method that combines prewhitening and the studentized bootstrap. While both prewhitening and the studentized bootstrap each provides improvement over standard normal theory intervals, one can achieve a further improvement by conjoining them in an appropriate way. As a side note, it is stressed that symmetric confidence intervals equal-tailed ones, since they exhibit improved coverage accuracy. We propose concrete ways to deal with the issues of block size, choice of kernel, and choice of bandwidth. The improvements in small sample performance are supported by a simulation study.

JEL CLASSIFICATION NOS: C14, C15, C22, C32.

KEY WORDS: Bootstrap, Confidence Intervals, Studentization, Time Series Regressions, Prewhitening.

*Joseph P. Romano, Phone: (650) 723-6326, Fax: (650) 725-8977, E-mail: romano@stat.stanford.edu.

Research of the first author funded by the National Science Foundation, DMS 9704487; Research of the second author partly funded by the Spanish "Dirección General de Enseñanza Superior" (DGES), reference number PB98-0025.

1 Introduction

Regressions in macroeconomics and finance typically involve explanatory variables and/or error variables that exhibit serial dependence. It is well-known that standard regression theory does not apply to such settings but appropriate asymptotic theory for time series regression has been developed and is routinely applied in practice; for example, see Hannan (1970) or White (1984). Over the last decade, the literature has shown that the finite sample properties of standard (or normal theory) methods are often lacking in practice; a common phenomenon is that confidence intervals undercover and that hypothesis tests overreject. We will focus on the construction of confidence intervals in the remainder of the paper; however, the ideas also apply to hypothesis tests.

A number of recent proposals have been made to construct confidence intervals that exhibit improved coverage accuracy. The two most common proposals, arguably, are prewhitening (Andrews and Mohanan, 1992; Newey and West, 1994) and bootstrapping (Fitzenberger, 1997). As far as the bootstrap is concerned, it seems that the importance of using a studentized bootstrap in order to achieve asymptotic refinements over normal theory has not yet been realized in the time series regression literature. This importance, in the context of smooth functionals of the mean, was pointed out by Davison and Hall (1993) and Götze and Künsch (1996).

We propose a new approach which combines prewhitening and the studentized bootstrap to yield an even further improvement. Previous theory suggests that both prewhitening and the use of a studentized bootstrap each offers improved coverage properties. By combining these two approaches and offering concrete suggestions on implementation (such as choice of block size, kernel, and bandwidth parameters), we believe we provide a useful method that is safe to apply in practice. Formal theory would require tedious and lengthy Edgeworth expansions, and even in i.i.d. situations, typically only provide rates of convergence for coverage error and length of intervals. Although such theory lends theoretical support in our fairly complicated time series regression setting, we validate our claims by heuristics and a simulation study, with theoretical work to follow. In addition, it is stressed that when two-sided confidence intervals are to be constructed, one should employ symmetric intervals as opposed to equal-tailed intervals, since the former tend to enjoy better coverage properties.

The remainder of the paper is organized as follows. Section 2 presents the model and the inference problem. In Section 3 some previous methods are briefly reviewed. Section 4 gives the details of our method. We also discuss how to choose the block size which is an important model parameter of any nonparametric time series bootstrap. Finite sample performance is examined via a simulation study in Section 5. Finally, Section 6 summarizes our findings. All tables appear in Section 7 at the end of the paper.

2 The Model

We consider the following regression model

$$y_t = X_t'\beta + \epsilon_t,\ t = 1, \ldots, T, \tag{1}$$

where $\beta \in \mathbb{R}^p$. The ordinary least squares (OLS) estimator for β is given by

$$\hat{\beta}_T = \left(\sum_{t=1}^{T} X_t X_t' \right)^{-1} \sum_{t=1}^{T} X_t y_t.$$

A critical assumption to ensure its consistency is $E(X_t \epsilon_t) = 0,\ t = 1, \ldots, T$, that is, the regressors are uncorrelated with the error term. This assumption is typically implied by economic considerations, such as a rational expectation model. Another key assumption is a weak dependence condition on the $\{(X_t', \epsilon_t)'\}$ process. We assume an α-mixing condition, which is standard in the literature; for example, see White (1984). Alternatively, other weak dependence conditions could be employed. Finally, some moment conditions on the $\{X_t' \epsilon_t\}$ and $\{\epsilon_t\}$ processes and some heteroscedasticity bounds are needed to ensure asymptotic normality of the OLS estimator $\hat{\beta}_T$. For detailed sufficient sets of mixing, moments, and heteroscedasticity bounds conditions, the reader is referred to White (1984, Chapter V) or Politis et al. (1999, Theorem 3.4), among others. One of such sets shall be assumed in the sequel of the paper. Note that our method will generally not work under nonstandard asymptotics, one example being the situation where one of the regressors has a unit root.

Interest focuses on constructing confidence intervals for a real-valued parameter $\theta = a'\beta$, where a is a fixed and known $p \times 1$ vector. Quite often θ will simply be a particular regression coefficient of interest. The restriction to real-valued parameters is made mainly to give a natural setting for the construction of confidence intervals. Of course, other parameters can be considered as well. For example, one might be interested in a general linear function $R\beta$, where R is a fixed and known $l \times p$ matrix, an important special case being $R = I_p$, the $p \times p$ identity matrix. If the parameter of interest is multivariate, one often uses hypothesis tests, since it is somewhat difficult to explicitly write down a multivariate confidence region. The method proposed can be easily adapted to testing problems; see Remark 4.1.

3 Background

In this section, two previous approaches for making inference in time series regressions are reviewed. Some details are presented that will be relevant later on.

3.1 Normal Theory

Under nonrestrictive regularity conditions, $\sqrt{T}(\hat{\beta}_T - \beta) \stackrel{\mathcal{L}}{\Longrightarrow} N(0, \Sigma)$, where Σ is a positive-definite $p \times p$ matrix and $\stackrel{\mathcal{L}}{\Longrightarrow}$ denotes convergence in distribution (or convergence in law).

This implies $\sqrt{T}(a'\hat{\beta}_T - \theta) \stackrel{\mathcal{L}}{\Longrightarrow} N(0, a'\Sigma a)$, which would allow one to construct an asymptotic normal theory confidence interval for θ if Σ were known. Unfortunately, Σ depends on the unknown underlying probability mechanism. The standard way of making inference is therefore to consistently estimate the limiting covariance matrix Σ by an estimator $\hat{\Sigma}_T$ and to pretend the distribution of $\sqrt{T}(a'\hat{\beta}_T - \theta)$ is given by $N(0, a'\hat{\Sigma}_T a)$, that is, to use the plug-in principle. As is well-known,

$$\Sigma = \lim_{T\to\infty} \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1} \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} E\epsilon_s X_s (\epsilon_t X_t)' \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1}.$$

Since the series $\{X_t\}$ is observed, consistent estimation of Σ only requires a consistent estimator of $J_T \equiv (1/T)\sum_{s=1}^{T}\sum_{t=1}^{T} E\epsilon_s X_s (\epsilon_t X_t)'$. By change of variables, and introducing the notation $V_t = \epsilon_t X_t$, the estimand J_T can be rewritten as

$$J_T = \sum_{j=-T+1}^{T-1} \Gamma_T(j) + \Gamma_T(j)',$$

where

$$J_T = \begin{cases} \Gamma_T(j) = \frac{1}{T}\sum_{t=j+1}^{T} EV_t V_{t-j} & \text{if } j \geq 0 \\ \Gamma_T(j) = \frac{1}{T}\sum_{t=-j+1}^{T} EV_{t+j} V_t & \text{if } j < 0. \end{cases}$$

A natural estimator for $\Gamma_T(j)$ is given by

$$\hat{\Gamma}_T(j) = \frac{1}{T}\sum_{t=j+1}^{T} \hat{V}_t \hat{V}_{t-j}, \text{ where } \hat{V}_t = X_t(y_t - X_t'\hat{\beta}_T).$$

Unfortunately, simply replacing $\Gamma_T(j)$ by $\hat{\Gamma}_T(j)$ in the formula for J_T would lead to an inconsistent estimator; for example, see Priestley (1981, Chapter 6). The most common way to arrive at consistent estimators for J_T is to consider kernel estimators of the form

$$\hat{J}_T = \hat{J}_T(S_T) = \frac{T}{T-p}\sum_{j=-T+1}^{T-1} k\left(\frac{j}{S_T}\right)\hat{\Gamma}_T(j), \tag{2}$$

yielding

$$\hat{\Sigma}_T = \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1} \hat{J}_T \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1}. \tag{3}$$

In (2), $k(\cdot)$ is a real-valued kernel, S_T is a bandwidth parameter, and the factor $T/(T-p)$ is a small-sample-degree-of-freedom adjustment for estimating the p-dimensional parameter β. The kernel $k(\cdot)$ determines the weight that is given to the $\hat{\Gamma}_T(j)$ and typically satisfies the three conditions $k(0) = 1$, $k(\cdot)$ is continuous at 0, and $\lim_{x\to\pm\infty} k(x) = 0$. For a number of popular kernels, see Priestley (1981, Chapter 6) or Andrews (1991), among others.

An important feature of a kernel is its characteristic exponent $1 \leq q \leq \infty$, determined by the smoothness of the kernel at the origin. Note that the bigger q, the smaller is the asymptotic bias of a kernel variance estimator; on the other hand, only kernels with $q \leq 2$

yield estimates that are guaranteed to be positive semi-definite in finite samples. Most of the commonly used kernels have $q = 2$, such as the Parzen, Tukey-Hanning, and Quadratic-Spectral (QS) kernels, while the Bartlett kernel has $q = 1$ and the Truncated kernel has $q = \infty$. For a broader discussion on this issue, see Priestley (1981, Chapter 6).

Once a particular kernel $k(\cdot)$ has be chosen for application, one is left to pick the bandwidth S_T. Several data-dependent methods, based on various asymptotic optimality criteria, are available to choose S_T; for example, see Andrews (1991) and Newey and West (1994). Note that the 'optimal' bandwidth generally depends on the underlying stochastic mechanism generating the data, as well as the choice of the kernel $k(\cdot)$.

Given the kernel estimator $\hat{J}_T$, it is now straightforward to compute the estimator $\hat{\Sigma}_T$ and construct normal theory confidence intervals for θ as outlined above. However, these intervals tend to work unsatisfactorily in small samples, especially when the dependence structure of the underlying data is strong; again, see Andrews (1991) and Newey and West (1994). The method of prewhitening, dating back to Press and Tukey (1956), has been suggested to better this situation and it works as follows. It is well-known that covariance matrix estimation corresponds to spectral density matrix estimation at frequency zero and that it is easier to estimate spectral density matrices of white noise processes than those of heavily dependent processes. The idea then is to first transform the $\{\hat{V}_t\}$ process to make it look more like white noise (that is, to *prewhiten* it), to then estimate the spectral density matrix of the transformed process at frequency zero, and to finally *recolor* the estimated matrix. The transformation is usually done by passing the $\{\hat{V}_t\}$ process through a linear filter, often taken to be a Vector Autogressive (VAR) model, and the recoloring is achieved by applying the inverse transformation to the estimated spectral density matrix of the filtered process. To be more specific, first, one estimates a dth order VAR model (by OLS, say) for $\hat{V}_t$:

$$\hat{V}_t = \sum_{r=1}^{d} \hat{A}_r \hat{V}_{t-r} + \hat{U}_t.$$

Second, one computes a standard covariance matrix estimator, call it $\hat{J}_T^{\hat{U}}$, as described above but based on the VAR residual vectors $\{\hat{U}_t\}$ rather than the original vectors $\{\hat{V}_t\}$. Third, one recolors $\hat{J}_T^{\hat{U}}$ to obtain a prewhitened estimator of J_T:

$$\hat{J}_{T,PW} = \hat{D}\hat{J}_T^{\hat{U}}\hat{D}', \quad \text{where} \quad \hat{D} = \left(I_p - \sum_{r=1}^{d} \hat{A}_r\right)^{-1}, \tag{4}$$

yielding

$$\hat{\Sigma}_{T,PW} = \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1} \hat{J}_{T,PW} \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1}. \tag{5}$$

Note that the matrix $(I_p - \sum_{r=1}^{d} \hat{A}_r)$ in (4) may not be invertible, in which case a general inverse can be used or an invertibility adjustment (e.g., Andrews and Monahan, 1992) can be made to compute $\hat{D}$. Instead of VAR models, other linear filters could be used in the prewhitening and recoloring process.

According to empirical studies in Andrews and Monahan (1992) and Newey and West (1994), among others, confidence intervals for θ based on $\hat{\Sigma}_{T,PW}$ tend to exhibit better coverage properties in finite samples than those based on $\hat{\Sigma}_T$.

3.2 The Bootstrap

Given the success of the bootstrap in regression settings with independent observations (Wu, 1986), it is natural to apply an appropriate bootstrap method to time series regressions. Such a method has to take into account the time series structure of both the regressors and the error variables. In the absence of semi-parametric structural models for these variables, such as AR or VAR models with i.i.d. innovations, the common approach is to resample blocks of data. For ease of notation, let $Z_t = (X_t', y_t)'$, so the observed data is $\{Z_1, \ldots, Z_T\}$; also denote the true probability mechanism by P. The most popular time series bootstrap is the moving blocks bootstrap due to Künsch (1989) and Liu and Singh (1992). It considers overlapping blocks of size b, namely $Y_t = \{Z_t, \ldots, Z_{t+b-1}\}$, $t = 1, \ldots, T-b+1$. Assuming for the moment $T = lb$, the method selects l blocks Y_t^* at random and with replacement from the $T-b+1$ available blocks and concatenates them to arrive at the pseudo sequence $\{Y_1^*, \ldots, Y_l^*\} = \{Z_1^*, \ldots, Z_T^*\}$; in case T is not a multiple of b, one would do the same with the smallest l such that $T < lb$ and then truncate the pseudo sequence at T observations. Denote by P_T^* the bootstrap distribution (conditional on the observed data) of the pseudo sequence $\{Z_1^*, \ldots, Z_T^*\}$. Let $\hat{\beta}_T^*$ be the OLS estimator of β computed from the pseudo sequence and let $\hat{\theta}_T^*$ be the corresponding linear combination $a'\hat{\beta}_T^*$. A straightforward bootstrap approximation of the sampling distribution of the OLS estimator $\hat{\beta}_T$ is then

$$\mathcal{L}_P\{\hat{\theta}_T - \theta\} \approx \mathcal{L}_{P_T^*}\{\hat{\theta}_T^* - \hat{\theta}_T\}. \tag{6}$$

Here, the general notation $\mathcal{L}_Q\{W\}$ stands for the law of a random variable W when the probability mechanism is Q.

The relation (6) can now be used to construct an approximate confidence interval for θ. This particular bootstrap approximation is usually referred to as the hybrid bootstrap (Hall, 1992) or the basic bootstrap (Davison and Hinkley, 1997); we shall use the latter notation henceforth. Two problems with the moving blocks bootstrap are that it does not produce stationary pseudo sequences even if the original sequence is stationary and that $\hat{\theta}_T$ generally is not equal to $\theta(P_T^*)$, the parameter θ corresponding to the bootstrap distribution P_T^*, due to 'edge effects'. To fix these shortcomings, Politis and Romano (1994) suggested the stationary bootstrap where the data are 'wrapped' in a circle and the block lengths are random i.i.d. according to a geometric distribution. The result are bootstrap sequences that are stationary and avoid edge effects.

Fitzenberger (1997) applied the basic moving blocks bootstrap to time series regressions. However, its finite sample performance did not appear superior to normal theory intervals.

The reason for this 'disappointment' has its roots in the much-studied simpler setting of the mean for i.i.d. observations. It is well-known (e.g., Hall, 1992) that, in this setting, the basic bootstrap does not provide an asymptotic refinement over the CLT normal interval in the sense that both are only first order correct. To achieve second order correctness, a more sophisticated method such as the studentized or the BC_a bootstrap has to be employed. This result carries over to the dependent case. Davison and Hall (1993) and Götze and Künsch (1996), abbreviated by DH and GK, respectively, in the remainder of this paper, considered inference for the mean (and smooth functions of the mean) in the context of stationary, dependent observations. They showed that the basic bootstrap is only first order correct while the studentized bootstrap provides an asymptotic refinement, at least under regularity conditions that ensure an Edgeworth expansion; detailed sets of such conditions can be found in DH and GK. Note that it turns out to be important that the studentization is done in a certain way; see the discussion below. Moreover, the moving blocks bootstrap distribution needs to be centered around the mean of the bootstrap distribution rather than the sample mean. Unfortunately, the theory covering the mean case for dependent data, based on nontrivial Edgeworth expansions, is already very complicated, so that a corresponding result for time series regressions would be lengthy and quite technical. However, since the mean is a special case of an OLS estimator, it is natural to expect better performance from a studentized bootstrap in general regression settings as well.

The studentized bootstrap, together with the proper centering, leads to the approximation

$$\mathcal{L}_P\{(\hat{\theta}_T - \theta)/\hat{\sigma}_T\} \approx \mathcal{L}_{P_T^*}\{(\hat{\theta}_T^* - \theta(P_T^*))/\hat{\sigma}_T^*\}, \tag{7}$$

which again can be used to construct a confidence interval for θ. Here, $\hat{\sigma}_T$ is an estimator of the standard deviation of $\hat{\theta}_T$ and $\hat{\sigma}_T^*$ is an estimator of the standard deviation of $\hat{\theta}_T^*$. Often, and also in the particular method that we will suggest in Section 4, the two estimators have the same form, one being applied to the original sequence and the other being applied to the bootstrap sequence. A natural choice for our application is a kernel estimator and then $\hat{\sigma}_T = \sqrt{a'\hat{\Sigma}_T a}$ and $\hat{\sigma}_T^* = \sqrt{a'\hat{\Sigma}_T^* a}$. In the latter expression, $\hat{\Sigma}_T^*$ is the kernel estimator of Σ applied to the bootstrap sequence.

Remark 3.1 In certain instances the forms of $\hat{\sigma}_T$ and $\hat{\sigma}_T^*$ may be different. For example, when the bootstrap sequence is generated by the moving blocks bootstrap, one can exploit the particular dependence structure of such a sequence (it being generated by concatenating i.i.d. blocks of data) to arrive at the following 'natural' variance estimator $\hat{\sigma}_T^*$ (DH and GK): Assuming for simplicity that $T = lb$, where b is the block size used to construct the moving blocks bootstrap sequence, define

$$\hat{V}_t^* = X_t^*(y_t^* - X_t^*\hat{\beta}_T^*), \quad t = 1, \ldots, T,$$

$$\zeta_j = \frac{1}{\sqrt{b}} \sum_{t=1}^{b} V_{(j-1)b+t}^*, \quad j = 1, \ldots, l,$$

and

$$\hat{\Sigma}^*_{T,MB} = \left(\frac{1}{T}\sum_{t=1}^{T} X^*_t (X^*_t)'\right)^{-1} \frac{1}{l}\sum_{j=1}^{l}(\zeta_j - \bar{\zeta})^2 \left(\frac{1}{T}\sum_{t=1}^{T} X^*_t (X^*_t)'\right)^{-1}.$$

Then, the 'natural' variance estimator is given by

$$\hat{\sigma}^*_T = \sqrt{a'\hat{\Sigma}^*_{T,MB}a}. \tag{8}$$

On the other hand, DH and GK suggest to use a kernel estimator based on the Truncated kernel for the choice of $\hat{\sigma}_T$.

The reasoning for choosing different forms of $\hat{\sigma}_T$ and $\hat{\sigma}^*_T$ in this context is as follows. DH and GK showed that in order to obtain asymptotic refinements over normal theory it is crucial to use low bias variance estimators in the studentization. They then picked the variance estimators with smallest bias for both the original time series as well as the bootstrap generated time series. With this choice, the bootstrap approximation (7) can be shown to have error $O_P(n^{-3/4+\epsilon})$ where ϵ is a small number (DH and GK); in contrast, the approximation (6) has error larger than $O_P(n^{-1/2})$ as does normal theory.

Remark 3.2 An issue that arises when two-sided bootstrap confidence intervals are desired is whether to construct equal-tailed or symmetric intervals; note that normal theory intervals are both equal-tailed and symmetric by nature. In the setting of independent data and under appropriate regularity conditions, it has been repeatedly pointed out that symmetric bootstrap confidence intervals enjoy improved coverage accuracy (e.g., Beran, 1987; Hall, 1988). Such considerations carry over to dependent data settings. Hence, if coverage is of prime concern, it is preferable to construct symmetric intervals based on the bootstrap approximation of the two-sided sampling distribution function

$$\mathcal{L}_P\{|\hat{\theta}_T - \theta|/\hat{\sigma}(\hat{\theta}_T)\} \approx \mathcal{L}_{P^*_T}\{|\hat{\theta}^*_T - \theta(P^*_T)|/\hat{\sigma}^*(\hat{\theta}^*_T)\}.$$

4 Studentization with a Prewhitened Kernel Variance Estimator

4.1 The General Method

The previous section presented two approaches to improve upon standard normal theory intervals, namely using a prewhitened kernel estimator to construct normal theory intervals, and employing a studentized bootstrap. It is an obvious idea to combine these two approaches to obtain an even further improvement. The method is simple and consists of using a studentized bootstrap in conjunction with a prewhitened kernel variance estimator.

Consider the general studentized bootstrap approximation (7). As outlined in Remark 3.1, DH and GK use a different studentization for the bootstrap sequence as compared

to the original sequence in order to get variance estimators with minimal variance in both worlds. However, a closer inspection of the proofs of DH and GK reveals that $\hat{\sigma}_T$ and $\hat{\sigma}_T^*$ can be taken to be of the same form and an asymptotic refinement over the normal approximation still be retained. Indeed, if the (common) variance estimator is based on a kernel with characteristic component $q = 2$, an error of $O_P(n^{-2/3+\epsilon})$ is achieved.

Given that the method of DH and GK has an error of $O_P(n^{-3/4+\epsilon})$, why do we suggest to use a common variance estimator? There are two reasons. First, and less importantly, the Truncated kernel used by DH and GK in computing $\hat{\sigma}_T$ can lead to negative variance estimates in finite samples. By using a kernel with characteristic component $q = 2$ this problem is avoided. Second, we expect to improve upon DH and GK by using a common *prewhitened* kernel variance estimator. A corresponding theoretical result would be extremely technical and would seem to have to be limited to certain classes of probability mechanisms. Rather, along the lines of Andrews and Monahan (1992) and Newey and West (1994), we shall promote our method by heuristic arguments combined with simulation evidence. Heuristics say that prewhitened kernel variance estimators tend to perform better than standard kernel variance estimators and this edge should carry over to the studentized bootstrap. The simulation evidence will be presented in Section 5.

For the implementation of the method, one must choose a prewhitened kernel variance estimator and a suitable time series bootstrap. As far as the kernel is concerned, it was seen necessary to employ a kernel with $q \geq 2$ to achieve asymptotic refinements. On the other hand, only a kernel with $q \leq 2$ is guaranteed to yield positive semi-definite variance estimates in finite samples. We therefore suggest a kernel with $q = 2$. Among these, the so-called Quadratic Spectral (QS) kernel enjoys some attractive optimality properties; for example, see Priestley (1983, Chapter 6) or Andrews (1991). We shall use this kernel in the remainder of the paper.

As for a suitable bootstrap method, it is clear that a generally valid time series bootstrap needs to be employed. In the absence of a plausible (semi-)parametric data generating mechanism, a nonparametric method should be employed. Natural choices are the moving blocks bootstrap and the stationary bootstrap. We propose the latter for two reasons. First, due to the avoidance of 'edge effects', one has $\theta(P_T^*) = \hat{\theta}_T$ for the stationary bootstrap while for the moving blocks bootstrap the parameter of the bootstrap distribution would require an extra computation. Second, both bootstrap methods depend on a model parameter that governs the block size but the stationary bootstrap appears less sensitive to its choice (Politis and Romano, 1994; Davison and Hinkley, 1997, Section 8.2).

Remark 4.1 We have proposed a method to construct confidence intervals for a real-valued parameter $\theta = a'\beta$, where a is a fixed known $p \times 1$ vector. At times, interest might instead focus on a multivariate parameter $\theta = R\beta$, where R is a fixed and known $l \times p$ matrix. In this instance, it is more natural to consider hypothesis tests of the sort $H_0 : R\beta = r_0$, since multivariate confidence regions are somewhat difficult to write down. Note

that our method has its analogue in the construction of hypothesis tests. The test can be performed by approximating the sampling distribution of the Wald test statistic under the null hypothesis. Our method works by using a prewhitened kernel variance estimator in computing the Wald statistic both for the original sequence and for the bootstrap sequences. The approximation, under the null hypothesis, then is the following

$$\mathcal{L}_P\{(R\hat{\beta}_T - r_0)'(R\hat{\Sigma}_{T,PW}R')^{-1}(R\hat{\beta}_T - r_0)\} \approx_{H_0}$$
$$\mathcal{L}_{P^*_{T,0}}\{(R\hat{\beta}^*_T - r_0)'(R\hat{\Sigma}^*_{T,PW}R')^{-1}(R\hat{\beta}^*_T - r_0)\}.$$

Note that when the bootstrap is used for the purposes of hypothesis testing, it is advisable that the bootstrap law $P^*_{T,0}$ satisfy the constraints of the null hypothesis (e.g., Politis et al., 1999, Section 1.8). For our application, this means that $R\beta(P^*_{T,0}) = r_0$. Obviously, this cannot be achieved by simply resampling blocks of the observed data. One way of enforcing the null hypothesis in $P^*_{T,0}$ is the following. Denote by $\tilde{\beta}_T$ the constrained least squares estimators based on the observed data and satisfying $R\tilde{\beta}_T = r_0$. Also, let $\tilde{\epsilon}_t = y_t - X_t'\tilde{\beta}_T$, $\tilde{\epsilon}_{t,0} = \tilde{\epsilon}_t - T^{-1}\sum_{i=1}^T \tilde{\epsilon}_i$, and $y_{t,0} = X_t'\tilde{\beta}_T + \tilde{\epsilon}_{t,0}$. Then, $P^*_{T,0}$ resamples blocks from the 'null data' $(X_1', y_{1,0})', \ldots, (X_T', y_{T,0})'$.

4.2 Choice of the Block Size

The application of a time series bootstrap requires a choice of the block size. The moving blocks uses blocks of fixed length b_{MB}, while the stationary bootstrap draws blocks of random sizes according to a geometric distribution with parameter $0 < p < 1$. With little loss of generality, we can assume that the average block size $b_{SB} = 1/p$ is an integer. In this way, both bootstrap methods can be handled by a unique parameter b. Asymptotic theory typically only requires that $b \to \infty$ and that $b/T \to 0$ as $T \to \infty$; for example, see Künsch (1989) and Politis and Romano (1994). However, these requirements are of little practical help. The choice of the block size is a difficult but important problem, comparable to the choice of the bandwidth for kernel variance estimators. Unfortunately, in the relevant literature this problem is quite often either ignored or delayed to future research. The aim of this paper is to propose an inference method for time series regressions that is not only of academic interest but will also find the approval of practitioners. Therefore, we feel the need to provide at least a reasonable ad hoc method that can be used in practice, though we are unable to completely solve this difficult problem (and it appears unlikely that a 'perfect' solution will ever be found).

A notable exception in the literature, dealing explicitly with the problem of choosing the block size, is Hall et al. (1996). They showed that, for the moving blocks bootstrap, the optimal block size (minimizing the asymptotic mean squared error or MSE) depends significantly on context and is given by $C(P)n^{1/k}$, where $C(P)$ is a constant and $k = 3$, 4, or 5 for the contexts of variance estimation, estimation of a one-sided distribution function, or estimation of a two-sided distribution function, respectively. The constant $C(P)$ depends

on the underlying joint distribution P and the context but a way is suggested to estimate it in practice. The problem with trying to adopt their approach for our purposes is two-fold. First, all the asymptotic MSE calculations are done for the basic moving blocks bootstrap and thus would no longer be valid for the studentized moving blocks bootstrap. Second, for the estimation of a distribution function, $C(P)$ depends on the argument of that function, that is, on x in $F(x) = Prob_P\{Z \leq x\}$ for a general random variable Z. Since our interest is in estimating a quantile of a distribution function, it seems that the corresponding x would first have to be found in some recursive fashion.

Instead, we will propose a method which can be applied to an arbitrary bootstrap method, whether studentized or not, and which immediately tackles the task of estimating a specific quantile as opposed to estimating the distribution function at a given point. To this end, we suggest to use a *calibration* method, a concept dating back to Loh (1987, 1988, 1991). One can think of the actual coverage level $1-\lambda$ of a time series bootstrap confidence interval as a function of the block size b, conditional on the underlying probability mechanism P, the nominal confidence level $1-\alpha$, and the sample size T. The idea is now to adjust the 'input' b in order to obtain the actual coverage level close to the desired one. Hence, one can consider the block size calibration function $g : b \to 1-\lambda$. If $g(\cdot)$ were known, one could construct an 'optimal' confidence interval by finding $\tilde{b}$ that minimizes $|g(b) - (1-\alpha)|$ and use $\tilde{b}$ as the block size of the time series bootstrap; note that $|g(b) - (1-\alpha)| = 0$ may not always have a solution.

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

Algorithm 4.1 (Choice of the Block Size)

1. Fit a semi-parametric model $\hat{P}_T$ to the observed data $(X_1', y_1)', \ldots, (X_T', y_T)'$.
2. Fix a selection of reasonable block sizes b.
3. Generate K pseudo sequences $((X_1^*)', y_1^*)_k' \ldots, ((X_T^*)', y_T^*)_k'$, $k = 1, \ldots, K$, according to $\hat{P}_T$. For each sequence, $k = 1, \ldots, K$, and for each b, compute a confidence interval $\mathrm{CI}_{k,b}$.
4. Compute $\hat{g}(b) = \#\{\theta(\hat{P}_T) \in \mathrm{CI}_{k,b}\}/K$.
5. Find the value of $\tilde{b}$ that minimizes $|\hat{g}(b) - (1-\alpha)|$.

The choice of a semi-parametric model rather than a nonparametric time series bootstrap in generating the pseudo sequences is motivated by the fact that the latter would require a block size of its own. The role of the semi-parametric model in Algorithm 4.1 can be compared to the role of the semi-parametric model in the prewhitening for kernel variance estimation. Even if the model is misspecified, it should contain some information on the dependence structure of the true mechanism P that can be exploited to estimate $g(\cdot)$. In practice we suggest to employ a VAR model, whose order could be estimated by one of the well-known information criteria, say, in conjunction with bootstrapping the estimated residuals.

Remark 4.2 Note that Algorithm 4.1 is essentially a double bootstrap and therefore computationally more expensive, by an order of magnitude, than the application of the bootstrap method once the block size has been determined.

Remark 4.3 If a bootstrap method is used for hypothesis testing rather than confidence interval construction, an analogous algorithm can be used by focusing on the significance level of the test rather than on the confidence level of the interval. Note that in this case the semi-parametric model $\hat{P}_T$ should be replaced by a semi-parametric model $\hat{P}_{T,0}$ which satisfies the constraints of the the null hypothesis; the remaining details are straightforward.

5 Simulation Study

The purpose of this section is to compare the small sample performance of various methods to construct confidence intervals in time series regressions via a simulation study. Performance is measured in terms of estimated coverage probability of nominal 95% and 90% intervals. The methods included in the study are normal theory intervals as well as basic and studentized bootstrap intervals. A few words regarding the various methods are in order.

The normal theory intervals use the QS kernel, both for the standard interval and for the prewhitened (PW) interval. The prewhitening is done using a VAR(1) model. The automatic choice of bandwidth is the one of Andrews (1991). We also tried the one of Newey and West (1994) but the differences were not meaningful and so the corresponding results are not reported.

As was discussed in Subsection 3.2, one can improve upon the equal-tailed basic bootstrap confidence intervals by both studentizing and symmetrizing. To judge the magnitude of the corresponding improvements, we include equal-tailed basic, equal-tailed studentized, and symmetric studentized intervals in the study. Here, the studentization is done as suggested by DH and GK. This means that the variance estimator corresponding to the observed sequence is based on the rectangular kernel, with bandwidth equal to the block size,

and the variance estimator corresponding to the bootstrap sequence is given by (8). In case the rectangular kernel variance estimate corresponding to the original sample is negative, we switch to the QS kernel with the automatic bandwidth choice of Andrews (1991).

Finally, we include the interval based on our method of Subsection 4.1. For the reasons discussed there, the time series bootstrap employed is the stationary bootstrap of Politis and Romano (1994). The prewhitened kernel variance estimator is based on the QS kernel together with a VAR(1) model for the prewhitening process. Note that we only compute the symmetric interval to limit the number of interval types included in the study to a reasonable one.

The following abbreviations are used to label the various confidence interval types.

- NT: normal theory interval
- NT-PW: prewhitened normal theory interval
- BO-ET: equal-tailed basic bootstrap interval
- BO-SD-ET: equal-tailed studentized (DH and GK) bootstrap interval
- BO-SD-SYM: symmetric studentized (DH and GK) bootstrap interval
- BO-SD-PW: symmetric studentized bootstrap interval using a prewhitened kernel variance estimator

The experimental design for our simulation study is as follows. The regression model under consideration is

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t + \epsilon_t, \quad t = 1, \ldots, T, \\ x_t &= 0.7 x_{t-1} + u_t, \quad u_t \text{ i.i.d } \sim N(0,1), \quad t = 1, \ldots, T. \end{aligned}$$

The parameter of interest is $\theta = \beta_1$. Without loss of generality, the true regression parameter vector is set to $(\beta_0, \beta_1)' = (0,0)'$. Five data generating processes (DGP's) for the residual process $\{\epsilon_t\}$ are employed. In what follows, $(v_{1,t}, v_{2,t})'$ i.i.d. $\sim N(0, I_2)$ is independent of $\{u_t\}$. The first specification is

$$\text{MA(1):} \quad \epsilon_t = v_{1,t} + 0.7 v_{1,t-1}, \tag{9}$$

that is, an MA(1) model with normal innovations. The second specification allows for one-dependent residuals with a marginal exponential law

$$\text{EXP:} \quad \epsilon_t = \frac{(v_{1,t} + 0.7 v_{1,t-1})^2 + (v_{2,t} + 0.7 v_{2,t-1})^2}{2(1 + 0.49)} - 1. \tag{10}$$

The third specification considered is a Markov regime-switching model (Hamilton, 1989). Here, d_t is an unobserved state variable which takes on the values 0 or 1 with Markovian transition probabilities

$$\begin{aligned} Prob\,(d_t = 1|d_{t-1} = 1) &= 0.9, \qquad Prob\,(d_t = 0|d_{t-1} = 1) = 0.1 \\ Prob\,(d_t = 1|d_{t-1} = 0) &= 0.25, \qquad Prob\,(d_t = 0|d_{t-1} = 0) = 0.75. \end{aligned}$$

Then, the residual process is defined by

$$\text{MARKOV:} \quad \epsilon_t = 0.25 v_{1,t} + d_t - E(\epsilon_t). \tag{11}$$

The fourth specification is an AR(1) model with normal innovations

$$\text{AR(1):} \quad \epsilon_t = 0.7\epsilon_{t-1} + v_{1,t}. \tag{12}$$

A variation, yielding conditionally heteroskedastic residuals, is the last specification

$$\text{AR(1)-HET:} \quad \tilde{\epsilon}_t = 0.7\tilde{\epsilon}_{t-1} + v_{1,t}, \quad \epsilon_t = |x_t|\tilde{\epsilon}_t. \tag{13}$$

As in Andrews (1991), the computational burden is reduced by transforming both $(y_1, \ldots, y_T)$ and $(x_1, \ldots, x_T)$ to mean zero vectors before running the regression; the same is done to the bootstrap samples. Note that, after the transformation, we are left with a univariate regressor, which much simplifies the computations. The transformation is close to the identity transformation with a high probability and should affect the simulation results very little. We also ran some 'untransformed' simulations for several scenarios to check this intuition and the results were identical up to simulation error.

Sample sizes included are $T = 50$, $T = 100$, and $T = 200$. Block sizes for the bootstrap methods are $b = 3$ and $b = 5$ when $T = 50$ and $b = 5$ and $b = 10$ when $T = 100$ or $T = 200$, respectively. The number of bootstrap replications is $K = 1,000$. Estimated coverage probabilities are based on 1,000 replications per scenario, with all interval types being computed on the same replications. Note that the accuracy of these estimates was improved by the use of the control variate technique (Davidson and MacKinnon, 1981). To this end, for each scenario the true finite sampling distribution of the OLS estimator $\hat{\beta}_1$ was approximated from 50,000 replications.

Remark 5.1 It would have been interesting to include an automatic choice of the block size like Algorithm 4.1 into the simulation study. Unfortunately, this was not possible. Even with the two fixed block sizes, each of Tables 1–5 represents roughly a full day of computing time using stand-alone C++ code on a supercomputer HP-Convex Exemplar SPP S2000. Keep in mind that the algorithm to determine the block size is computationally more demanding by an order of magnitude, so a corresponding table would take several months!

The results are listed in tables Tables 1– 5. The findings can be summarized as follows.

- In accordance with Davison and Hall (1993) and Götze and Künsch (1996), BO-ET does not improve upon NT.
- In accordance with Andrews and Monahan (1992) and Newey and West (1994), NT-PW improves upon NT. Note that this is true for all DGP's so it is confirmed that prewhitening is useful even when the parametric filter, here a VAR(1) model, is misspecified.
- In accordance with Davison and Hall (1993) and Götze and Künsch (1996), BO-SD-ET generally improves upon both BO-ET and NT.
- In accordance with Beran (1987) and Hall (1988), among others, BO-SD-SYM improves upon BO-SD-ET.
- BO-SD-SYM is better than NT-PW at some times and worse at other times.
- As suggested by our reasoning in Section 4.1, BO-SD-PW generally improves upon both NT-PW and BO-SD-SYM.
- The choice of the block size matters for the bootstrap methods but BO-SD-PW appears the least sensitive; this is probably due to the fact that it utilizes the stationary bootstrap and in accordance with Politis and Romano (1994) and Davison and Hinkley (1997, Section 8.2).

The last summary point, and the detailed results of Tables 1–5, demonstrate that the choice of the block size is indeed important in applying block bootstrap methods, although its importance diminishes as the sample size increases. While we suggested Algorithm 4.1 to deal with this choice in practice, it is, unfortunately, computationally too expensive to be incorporated in the general simulation study; see Remark 5.1. However, we wanted to shed at least *some* light on its finite sample performance. Therefore, we carried out a separate small-scale simulation to address this problem. The semi-parametric model in Algorithm 4.1 was a VAR(1) with bootstrapping the fitted residuals. For the latter, the stationary bootstrap with a small block size ($b = 3$) was used to capture some left-over dependence in the residuals in case the VAR(1) is misspecified. We considered the DGP's EXP and AR(1) with sample sizes $T = 50$ and $T = 100$ and the corresponding block sizes included in our main study, that is, $b = 3$ and $b = 5$ when $T = 50$, as well as $b = 5$ and $b = 10$ when $T = 100$. The interval types included in the study are BO-ET and BO-SD-PW. Given the results in Tables 3 and 4, we know that the optimal (of the two) block size is always the smaller one. Based on 100 replications for each scenario, we kept track of how frequently each block size was chosen for every of the four interval combinations (2 confidence levels and 2 interval types). Since Algorithm 4.1 is essentially a double bootstrap method and thus computationally very demanding, we had to limit the number of bootstrap samples to $K = 300$ in each layer.

The results are presented in Table 6; note that its generation required about as much computing time as Tables 1–5 combined. It is seen that our method is more likely to pick the better block size in all scenarios. These findings are encouraging, although, admittedly, it would be of further interest to directly simulate the performance of bootstrap methods with data-dependent choice of block size. In fact, a data-dependent choice of block size can result in a performance comparable and sometimes even superior to that of the best fixed block size; for example, see Politis et al. (1999, Section 9.5).

6 Conclusions

In this paper, a new method to construct confidence intervals in time series regressions was proposed. The motivation was to combine the improvements upon standard normal theory intervals that can be achieved by using a prewhitened kernel variance estimator on the one hand and a studentized bootstrap on the other hand. The crux of the method is to use a prewhitened kernel variance estimator in the bootstrap studentization. While this is a simple idea, it is of practical relevance and has not been realized elsewhere. The existing literature dealing with the bootstrap in time series regressions either neglects studentizing (e.g., Fitzenberger, 1997) or does not utilize prewhitened kernel variance estimators when studentizing (Davison and Hall, 1993; Götze and Künsch, 1996).

The small sample performance of various confidence interval types was examined via a simulation study. Based on the necessarily limited results of this study, our method indeed seems to improve upon both prewhitened kernel intervals and the studentized bootstrap suggested by Davison and Hall (1993) and Götze and Künsch (1996). In addition, it was seen that symmetric bootstrap confidence intervals enjoy improved coverage accuracy over equal-tailed bootstrap intervals.

There are two main avenues for further research, both currently blocked by the speed of commonly available computers. First, the main problem in applying block bootstraps method in time series regressions lies in choosing the block size. We proposed a heuristic algorithm to deal with this problem and some small-scale simulations showed encouraging results. However, it would be of real interest to directly examine the performance of bootstrap methods utilizing this data-dependent block size choice. Unfortunately, this is computationally too demanding at the present time. Perhaps a faster method to pick the block size can be found to alleviate the computational issue. Second, studentized bootstrap methods improve upon normal theory. At least in principle, further improvements are available by a second round of bootstrapping using either bootstrap calibration (Loh, 1987, 1988, 1991) or bootstrap prepivoting (Beran, 1987). So far, not much is known about the finite sample performance of these methods given their great compuational cost. However, in a few years computers should be fast enough to fill this void and, in particular, the application of these methods to time series regressions could be examined.

References

Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–858.

Andrews, D.W.K. and Monahan, J.C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* **60**, 953–966.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74**, 457–68.

Davidson, R. and MacKinnon, J.G. (1981). Efficient estimation of tail-area probabilities in sampling experiments. *Economic Letters* **8**, 73–77.

Davison, A.C and Hall, P. (1993). On studentizing and blocking methods for implementing the bootstrap with dependent data. *Australian Journal of Statistics* **35**, 215–224.

Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and their Application*. Cambridge University Press.

Fitzenberger, B. (1997). The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *Journal of Econometrics* **82**, 235–287.

Götze, F. and Künsch, H.R. (1996). Second order correctness of the blockwise bootstrap for stationary observations. *Annals of Statistics* **24**, 1914–1933.

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

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

Hall, P., Horowitz J.L., and Jing, B.-Y. (1996). On blocking rules for the bootstrap with dependent data. *Biometrika* **50**, 561–574.

Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**, 357–384.

Hannan, E.J. (1970). *Multiple Time Series*. John Wiley, New York.

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

Liu, R.Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap*, 225–248. Edited by LePage, R. and Billard, L., John Wiley, New York.

Loh, W.Y. (1987). Calibrating confidence coefficients. *Journal of the American Statistical Association* **82**, 155–162.

Loh, W.Y. (1988). Discussion of "Theoretical comparison of bootstrap confidence intervals" by P. Hall. *Annals of Statistics* **16**, 972–976.

Loh, W.Y. (1991). Bootstrap calibration for confidence interval construction and selection. *Statistica Sinica* **1**, 479–495.

Newey, W.K. and West, K.D. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* **61**, 631–653.

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

Politis, D.N. Romano, J.P., and Wolf, M. (1999). *Subsampling.* Springer, New York.

Press, H. and Tukey, J. W. (1956). Power spectral methods of analysis and their application to problems in airplane dynamics. Bell Systems Monograph No. 2606.

Priestley, M.B. (1981). *Spectral Analysis and Time Series.* Academic Press, New York.

White, H. (1984). *Asymptotic Theory for Econometricians.* Academic Press, Orlando.

Wu, C.F. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics* **14**, 1261–1343.

7 Tables

Table 1: Estimated coverage probabilities (EC) of various confidence intervals with nominal levels 0.95 and 0.90. The specification for the residual process is model (9). The first two columns correspond to the smaller block size always.

MA(1) residuals, $n = 50$, $b = 3$, and $b = 5$				
Interval	EC-0.95	EC-0.90	EC-0.95	EC-0.90
NT	0.87	0.81	0.87	0.81
NT-PW	0.91	0.86	0.91	0.86
BO-ET	0.87	0.80	0.87	0.82
BO-SD-ET	0.92	0.85	0.90	0.83
BO-SD-SYM	0.92	0.86	0.91	0.85
BO-SD-PW	0.93	0.88	0.93	0.88
MA(1) residuals, $n = 100$, $b = 5$ and $b = 10$				
Interval	EC-0.95	EC-0.90	EC-0.95	EC-0.90
NT	0.91	0.85	0.91	0.85
NT-PW	0.94	0.90	0.94	0.90
BO-ET	0.88	0.82	0.87	0.81
BO-SD-ET	0.91	0.86	0.90	0.85
BO-SD-SYM	0.92	0.87	0.90	0.85
BO-SD-PW	0.95	0.90	0.93	0.88
MA(1) residuals, $n = 200$, $b = 5$ and $b = 10$				
Interval	EC-0.95	EC-0.90	EC-0.95	EC-0.90
NT	0.92	0.87	0.92	0.87
NT-PW	0.97	0.93	0.97	0.93
BO-ET	0.92	0.87	0.90	0.86
BO-SD-ET	0.94	0.89	0.92	0.87
BO-SD-SYM	0.94	0.89	0.92	0.87
BO-SD-PW	0.95	0.91	0.94	0.89

Table 2: Estimated coverage probabilities (EC) of various confidence intervals with nominal levels 0.95 and 0.90. The specification for the residual process is model (10). The first two columns correspond to the smaller block size always.

EXP residuals, $n = 50$, $b = 3$, and $b = 5$				
Interval	EC-0.95	EC-0.90	EC-0.95	EC-0.90
NT	0.91	0.85	0.91	0.85
NT-PW	0.92	0.86	0.92	0.86
BO-ET	0.93	0.87	0.91	0.85
BO-SD-ET	0.91	0.85	0.90	0.84
BO-SD-SYM	0.93	0.87	0.92	0.86
BO-SD-PW	0.94	0.90	0.94	0.89
EXP residuals, $n = 100$, $b = 5$ and $b = 10$				
Interval	EC-0.95	EC-0.90	EC-0.95	EC-0.90
NT	0.92	0.86	0.92	0.86
NT-PW	0.94	0.88	0.94	0.88
BO-ET	0.93	0.86	0.90	0.85
BO-SD-ET	0.91	0.86	0.89	0.83
BO-SD-SYM	0.93	0.88	0.91	0.86
BO-SD-PW	0.95	0.89	0.93	0.87
EXP residuals, $n = 200$, $b = 5$ and $b = 10$				
Interval	EC-0.95	EC-0.90	EC-0.95	EC-0.90
NT	0.94	0.88	0.94	0.88
NT-PW	0.95	0.90	0.95	0.90
BO-ET	0.94	0.88	0.94	0.88
BO-SD-ET	0.92	0.85	0.93	0.87
BO-SD-SYM	0.95	0.89	0.94	0.90
BO-SD-PW	0.95	0.90	0.95	0.90

Table 3: Estimated coverage probabilities (EC) of various confidence intervals with nominal levels 0.95 and 0.90. The specification for the residual process is model (11). The first two columns correspond to the smaller block size always.

MARKOV residuals, $n = 50$, $b = 3$, and $b = 5$				
Interval	EC-0.95	EC-0.90	EC-0.95	EC-0.90
NT	0.85	0.79	0.85	0.79
NT-PW	0.87	0.82	0.87	0.82
BO-ET	0.86	0.80	0.83	0.77
BO-SD-ET	0.92	0.86	0.90	0.83
BO-SD-SYM	0.93	0.88	0.90	0.84
BO-SD-PW	0.93	0.87	0.92	0.86
MARKOV residuals, $n = 100$, $b = 5$ and $b = 10$				
Interval	EC-0.95	EC-0.90	EC-0.95	EC-0.90
NT	0.89	0.82	0.89	0.82
NT-PW	0.91	0.85	0.91	0.85
BO-ET	0.88	0.81	0.88	0.80
BO-SD-ET	0.93	0.87	0.90	0.84
BO-SD-SYM	0.94	0.87	0.91	0.85
BO-SD-PW	0.94	0.88	0.92	0.86
MARKOV residuals, $n = 200$, $b = 5$ and $b = 10$				
Interval	EC-0.95	EC-0.90	EC-0.95	EC-0.90
NT	0.90	0.82	0.90	0.82
NT-PW	0.92	0.87	0.92	0.87
BO-ET	0.88	0.81	0.90	0.84
BO-SD-ET	0.92	0.86	0.93	0.87
BO-SD-SYM	0.92	0.87	0.93	0.88
BO-SD-PW	0.93	0.88	0.94	0.90

Table 4: Estimated coverage probabilities (EC) of various confidence intervals with nominal levels 0.95 and 0.90. The specification for the residual process is model (12). The first two columns correspond to the smaller block size always.

AR(1) residuals, $n = 50$, $b = 3$, and $b = 5$				
Interval	EC-0.95	EC-0.90	EC-0.95	EC-0.90
NT	0.82	0.74	0.82	0.74
NT-PW	0.87	0.81	0.87	0.81
BO-ET	0.81	0.73	0.80	0.72
BO-SD-ET	0.90	0.82	0.88	0.81
BO-SD-SYM	0.91	0.84	0.90	0.83
BO-SD-PW	0.92	0.86	0.91	0.84
AR(1) residuals, $n = 100$, $b = 5$ and $b = 10$				
Interval	EC-0.95	EC-0.90	EC-0.95	EC-0.90
NT	0.87	0.81	0.87	0.81
NT-PW	0.92	0.88	0.92	0.88
BO-ET	0.86	0.80	0.85	0.78
BO-SD-ET	0.91	0.87	0.89	0.83
BO-SD-SYM	0.92	0.87	0.90	0.85
BO-SD-PW	0.94	0.89	0.92	0.87
AR(1) residuals, $n = 200$, $b = 5$ and $b = 10$				
Interval	EC-0.95	EC-0.90	EC-0.95	EC-0.90
NT	0.90	0.85	0.90	0.85
NT-PW	0.94	0.90	0.94	0.90
BO-ET	0.88	0.83	0.90	0.83
BO-SD-ET	0.92	0.87	0.93	0.88
BO-SD-SYM	0.93	0.88	0.94	0.89
BO-SD-PW	0.94	0.89	0.94	0.80

Table 5: Estimated coverage probabilities (EC) of various confidence intervals with nominal levels 0.95 and 0.90. The specification for the residual process is model (13). The first two columns correspond to the smaller block size always.

AR(1)-HET residuals, $n = 50$, $b = 3$, and $b = 5$				
Interval	EC-0.95	EC-0.90	EC-0.95	EC-0.90
NT	0.79	0.70	0.79	0.70
NT-PW	0.84	0.76	0.84	0.76
BO-ET	0.75	0.67	0.74	0.67
BO-SD-ET	0.89	0.82	0.88	0.81
BO-SD-SYM	0.92	0.85	0.91	0.85
BO-SD-PW	0.92	0.86	0.92	0.85
AR(1)-HET residuals, $n = 100$, $b = 5$ and $b = 10$				
Interval	EC-0.95	EC-0.90	EC-0.95	EC-0.90
NT	0.86	0.78	0.86	0.78
NT-PW	0.91	0.85	0.91	0.85
BO-ET	0.82	0.76	0.81	0.73
BO-SD-ET	0.91	0.85	0.90	0.83
BO-SD-SYM	0.93	0.88	0.92	0.86
BO-SD-PW	0.94	0.90	0.93	0.88
AR(1)-HET residuals, $n = 200$, $b = 5$ and $b = 10$				
Interval	EC-0.95	EC-0.90	EC-0.95	EC-0.90
NT	0.89	0.82	0.89	0.82
NT-PW	0.93	0.88	0.93	0.88
BO-ET	0.86	0.79	0.89	0.82
BO-SD-ET	0.91	0.86	0.92	0.87
BO-SD-SYM	0.93	0.88	0.94	0.89
BO-SD-PW	0.94	0.90	0.95	0.90

Table 6: Frequencies of block choices chosen by Algorithm 4.1 for BO-ET and BO-SD-PW confidence intervals with nominal levels 0.95 and 0.90. The first two columns correspond to the intervals with nominal level 0.95 always.

MARKOV residuals, $n = 50$				
Interval	$b = 3$	$b = 5$	$b = 3$	$b = 5$
BO-ET	0.80	0.20	0.72	0.28
BO-SD-PW	0.78	0.18	0.86	0.14
MARKOV residuals, $n = 100$				
Interval	$b = 5$	$b = 10$	$b = 5$	$b = 10$
BO-ET	0.81	0.19	0.73	0.27
BO-SD-PW	0.96	0.04	1.00	0.00
AR(1) residuals, $n = 50$				
Interval	$b = 3$	$b = 5$	$b = 3$	$b = 5$
BO-ET	0.66	0.34	0.63	0.37
BO-SD-PW	0.84	0.16	0.85	0.15
AR(1) residuals, $n = 100$				
Interval	$b = 5$	$b = 10$	$b = 5$	$b = 10$
BO-ET	0.77	0.23	0.66	0.34
BO-SD-PW	0.96	0.04	0.97	0.03