CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass

http://www.cass.city.ac.uk/cea/index.html

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

Jackknife Estimation of Stationary Autoregressive Models

Marcus J. Chambers

CEA@Cass Working Paper Series

WP–CEA–01-2012

Jackknife Estimation of Stationary Autoregressive Models

Marcus J. Chambers
University of Essex

August 2011

Abstract

This paper explores the properties of jackknife methods of estimation in stationary autoregressive models. Some general results concerning the correct weights for bias reduction under various sampling schemes are provided and the asymptotic properties of a jackknife estimator based on non-overlapping sub-samples are derived for the case of a stationary autoregression of order p when the number of sub-samples is either fixed or increases with the sample size at an appropriate rate. The results of a detailed investigation into the finite sample properties of various jackknife and alternative estimators are reported and it is found that the jackknife can deliver substantial reductions in bias in autoregressive models. This finding is robust to departures from normality, ARCH effects and misspecification. The median-unbiasedness and mean squared error properties are also investigated and compared with alternative methods as are the coverage rates of jackknife-based confidence inervals.

Keywords. Jackknife; bias; autoregression.

J.E.L. classification numbers. C12; C13; C22.

Acknowledgements: I am grateful to an anonymous Associate Editor and two referees for their helpful comments which have led to an improved and more focused piece of work. I would also like to thank Roy Bailey, Emma Iglesias, Maria Kyriacou, Joao Santos Silva and Michael Thornton for helpful comments on an earlier version of this paper. This research was funded by the Economic and Social Research Council under grant number RES-000-22-3082.

Address for Correspondence: Professor Marcus J. Chambers, Department of Economics, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, England.
Tel: +44 1206 872756; fax: +44 1206 872724; e-mail: mchamb@essex.ac.uk.

1. Introduction

Jackknife techniques have a long history in statistics. The jackknife method of bias reduction was originally proposed by Quenouille (1956) with Tukey (1958) subsequently demonstrating how the method could also be used to construct a nonparameteric estimator of variance. As a result it is often referred to as the Quenouille-Tukey jackknife; see, for example, Efron (1982, p.1). According to Miller (1964, p.1594) the procedure was named the jackknife by Tukey because "a boy scout's jackknife is symbolic of a rough-and-ready instrument capable of being utilized in all contingencies and emergencies." The applicability of the jackknife is certainly widespread but it has found fewer applications in econometrics than rival bootstrap methods. Indeed, Efron (1979) demonstrated that the jackknife is a linear approximation method for the bootstrap in the case of estimating the sampling distribution of a random variable based on a sample of i.i.d. (independently and identically distributed) data, a result that has perhaps been interpreted as favouring the bootstrap in a wider variety of situations than that to which this result relates. Moreover, as will be shown below, the standard formulation of the jackknife statistic is applicable only in the case of i.i.d. data, which may also help to explain its limited application in econometrics.

Notwithstanding the preceeding comments and the proliferation of bootstrap methods in econometrics, there has recently been a realisation that jackknife methods can be effective in reducing the bias of estimators in models of interest in econometrics. In models with more instruments than endogenous variables Angrist, Imbens and Krueger (1999) proposed the jackknife instrumental variables estimator and demonstrated its superior finite sample properties compared to the two-stage least squares estimator and its comparability to the limited information maximum likelihood estimator, although the performance of this estimator has subsequently been criticised by Davidson and MacKinnon (2006). Hahn, Kuersteiner and Newey (2003) considered both bootstrap and jackknife bias corrections to maximum likelihood estimators based on an i.i.d. sample while applications to panel data models (including nonlinear and dynamic models) have been considered by Hahn and Newey (2004), Hahn and Moon (2006) and Dhaene, Jochmans and Thuysbaert (2006). Jackknife methods have also been applied to maximum likelihood estimators of the parameters of continuous time models of the short-term interest rate by Phillips and Yu (2005) who also demonstrate the resulting gains that can be made by applying such techniques directly to the implied bond options prices. Based on the encouraging results obtained in the above situations this paper explores the properties of jackknife methods of estimation and inference in stationary autoregressive (AR) models. In the context of stationary time series Carlstein (1986) proposed an estimator of variance based on non-overlapping blocks while Künsch (1989) considered both jackknife and bootstrap methods of estimating standard errors by deleting whole blocks of observations. The focus here, however, is ultimately concerned more with issues of estimation of the parameters in AR models than it is with variance estimation, although the latter becomes important when using the jackknife estimator for inference.

Some general theoretical results on jackknife methods applied to a statistic of interest (such as an estimator of a parameter or a test statistic) are given in section 2. The first result (Theorem 1) shows how the full-sample and sub-sample statistics should be combined in order to eliminate the first-order bias in a general setting before considering specific sampling situations such as i.i.d. data as well as non-overlapping and moving-block sub-samples which are of particular relevance in time series settings. A further refinement (Theorem 2) shows how statistics from different sub-sampling methods, or from the same sub-sampling method with different numbers of sub-samples, can be combined to eliminate both first- and second-order bias from the statistic of interest. Specific cases of sub-sampling are also considered, and a further general result (Theorem 3) shows how the jackknife weights need to be modified in cases where sub-samples of unequal lengths are encountered, this being potentially important in empirical applications.

Section 3 explores jackknife methods of estimation in stationary autoregressive models, the focus being on the p'th order model with an intercept. The motivation for employing jackknife estimators in this context is rooted in analytical work that provides Nagar-type expansions for the bias of the ordinary least squares (OLS) estimator of the AR parameter vector. Theorem 4 derives the limiting distribution of the jackknife estimator based on non-overlapping sub-samples of the type utilised by Phillips and Yu (2005) and shows that it has the same form as the OLS estimator irrespective of whether the number of sub-samples is fixed or increases with the sample size at an appropriate rate. Hence the jackknife estimator has the potential to reduce the finite sample bias without any loss of asymptotic efficiency, although the effect on the finite sample mean squared error (MSE) is unknown (but is explored in simulations in section 5). Shao and Tu (1995, pp.66–67) provide examples where the jackknife statistic can have either a larger or smaller MSE than the underlying statistic, concluding that, in general, "the relative performance ... is indefinite and depends on the unknown population" (p.67) and, furthermore, "we should keep in mind that the jackknife estimator ... is designed to eliminate bias and, therefore, can be used when the bias is an important issue. We need to balance the advantage of unbiasedness against the drawback of a large mean squared error " (pp.67–68). The limiting distribution in Theorem 4 can be used as the basis for inference provided an appropriate estimator of the asymptotic variance matrix can be obtained, and two possibilities are provided in Theorem 5.

Section 4 reports the results of an extensive simulation exercise (involving 100,000 replications) using the AR(1) model in an attempt to obtain evidence on a number of issues, including: which sub-sampling method produces the greatest bias reduction; the optimal number of sub-samples to employ; how the optimal number of sub-samples varies with sample size; and the extent of additional bias reduction that can be achieved by eliminating the second-order bias. The results cover a range of sample sizes and a range of positive values for the AR parameter that approaches the boundary of the stationarity region, these being of greatest empirical relevance in economics and finance. The analysis of bias reduction using the jackknife when a unit root is present can be found in Chambers and Kyriacou (2011). Comparisons of the jackknife estimators are also made with respect to the exact median unbiased (MU) estimator of Andrews (1993) and a recursive-design wild boostrap estimator based on Gonçalves and Kilian (2004). The jackknife estimators are shown to result in the smallest bias in all cases considered. Section 4 also examines the robustness of the results to departures from normality, using Student's t- and Gamma distributions for the disturbances, as well as to autoregressive conditional heteroskedasticity (ARCH) and to higher-order and misspecified autoregressions.

Additional considerations concerning the performance of the jackknife (and other) techniques are explored in section 5. Although designed to reduce bias other distributional aspects are important to the usefulness of an estimator, and so the median-unbiasedness and mean squared error are examined first. Simulations reveal that it is possible to obtain an MSE less than the full-sample OLS estimator by using jackknife (and other) estimators, a feature of the jackknife estimators being that a larger number of sub-samples is required to minimise root MSE (RMSE) than to minimise bias. It is also shown that the distributions of the jackknife estimators are much closer to being median-unbiased than those of the OLS estimator, the latter being significantly negatively biased particularly for larger values of the AR parameter. Section 5 also looks at the coverage rates of jackknife confidence intervals based on the asymptotic distribution in Theorem 4 and compares them to those of OLS, MU and bootstrap methods. Proofs of all Theorems are contained in Appendix A, while Appendix B contains supplementary results that are used in the proofs and elsewhere. Section 6 concludes.

The following notation will be used throughout the paper. The symbol $\xrightarrow{p}$ denotes convergence in probability; $\xrightarrow{d}$ denotes convergence in distribution; and, for a $k \times 1$vector x, $\|x\| = (\sum_{i=1}^{k} x_i^2)^{1/2}$ denotes the Euclidean norm.

2. Jackknife methods: some general results

The idea behind the jackknife method of bias reduction is to combine a statistic based on a full sample of data with a set of statistics based on sub-samples in a way that eliminates the first-order bias. The statistic of interest, $\hat{\beta}_n$, is often an estimator of a parameter or parameter vector although functions of model parameters and test statistics, for example, can also be considered provided they satisfy (or are assumed to satisfy) certain properties. The following general result for the jackknife statistic will be used to deal with specific cases of interest.

Theorem 1. *Let $y = (y_1, \ldots, y_n)'$ be a sample of n observations on a random variable and let $\hat{\beta}_n = \beta(y)$ denote the statistic of interest satisfying*

$$E(\hat{\beta}_n) = \beta + \frac{a_1}{n} + \frac{a_2}{n^2} + O\left(n^{-3}\right), \tag{1}$$

where a_1 and a_2 are constants. Let Y_i $(i = 1, \ldots, m)$ denote a set of sub-samples of y, each of which has equal length $\ell = O(n)$, and let $\hat{\beta}_i = \beta(Y_i)$ $(i = 1, \ldots, m)$ denote the corresponding sub-sample statistics. Then the jackknife statistic

$$\hat{\beta}_J = \left(\frac{n}{n-\ell}\right)\hat{\beta}_n - \left(\frac{\ell}{n-\ell}\right)\frac{1}{m}\sum_{i=1}^{m}\hat{\beta}_i \tag{2}$$

satisfies $E(\hat{\beta}_J) = \beta + O(n^{-2})$.

Theorem 1 is a general result that holds for both i.i.d. samples as well as dependent samples of the type arising in time series. The expression for bias in (1) can usually be justified by a Nagar-type expansion; see, for example, Bao and Ullah (2007) for some results in the general time series setting and Bao (2007) for the AR(1) model under general error distributions. Some specific cases will now be considered and Theorem 1 will be employed to determine the appropriate weights to use in the construction of $\hat{\beta}_J$ based on different sub-sampling methods.

2.1 The i.i.d. case

In the case of a random sample of i.i.d. variables the sub-samples are usually obtained by deleting observation i from the full sample, so that the sub-samples are given by

$$Y_i = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)', \quad i = 1, \ldots, n.$$

Here, $m = n$ and the size of each sub-sample is $\ell = n - 1$. Hence, from Theorem 1, $\hat{\beta}_J$ takes the form (using the fact that $n - \ell = 1$)

$$\hat{\beta}_J^{\text{i.i.d.}} = n\hat{\beta}_n - (n-1)\frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_i. \tag{3}$$

This is sometimes known as the delete-1 jackknife because each sub-sample deletes one observation at a time, and its extension to the delete-d case was proposed by Wu (1986) although this extension will not be pursued here. The form of the jackknife statistic in (3) is the one commonly found in the literature[1] but, as demonstrated below, the weights involved in forming $\hat{\beta}_J^{\text{i.i.d.}}$ in (3) are not applicable when using different types of sub-sampling and/or with non-i.i.d. data.

2.2 Non-overlapping sub-samples

In time series settings the above jackknife method of deleting observations from the sample affects the correlation structure but the jackknife principle can still be applied subject to an appropriate sub-sampling scheme. The key requirement in constructing the sub-samples is that the dependence structure of the series is maintained. Phillips and Yu (2005) utilise non-overlapping sub-samples in applying the jackknife in an AR(1) model with intercept, the method working as follows.

Consider a set of m non-overlapping sub-samples, each of equal length ℓ, chosen so that $n = m \times \ell$. The number of sub-samples, m, will be treated as fixed and independent of n, so that the length of each sub-sample grows with n at the same rate; the assumption of fixed m will be relaxed later. Sub-sample i therefore contains the following observations:

$$Y_i = (y_{(i-1)\ell+1}, \ldots, y_{i\ell})', \quad i = 1, \ldots, m.$$

In this set-up we have $\ell = n/m$ and $n - \ell = (n/m)(m-1)$ and it follows that the weights in (2) become

$$\frac{n}{n-\ell} = \frac{m}{m-1} \quad \text{and} \quad \frac{\ell}{n-\ell} = \frac{1}{m-1},$$

resulting in the following jackknife statistic:

$$\hat{\beta}_{J,m} = \left(\frac{m}{m-1}\right)\hat{\beta}_n - \left(\frac{1}{m-1}\right)\frac{1}{m}\sum_{i=1}^{m}\hat{\beta}_i. \tag{4}$$

This expression corresponds to the form of jackknife estimator used by Phillips and Yu (2005).

2.3 Moving-block sub-samples

An alternative to non-overlapping sub-samples is to use a moving block of length ℓ. If each block is incremented by one observation the result is a set of $m = n - \ell + 1$ sub-samples of the form

$$Y_i = (y_i, \ldots, y_{i+\ell-1})', \quad i = 1, \ldots, m.$$

[1] See, for example, Quenouille (1956, p.354) or equation (2.8) of Efron (1982).

In this case $n-\ell = m-1$ and the jackknife statistic is easily seen to be

$$\hat{\beta}_{J,m}^{\mathrm{MB}} = \left(\frac{n}{m-1}\right)\hat{\beta}_n - \left(\frac{\ell}{m-1}\right)\frac{1}{m}\sum_{i=1}^{m}\hat{\beta}_i. \tag{5}$$

Note that in constructing the moving-block sub-samples it is the case that (some) observations are used more than once which is not the case with the non-overlapping blocks.

Another type of moving-block scheme is obtained by shifting the (non-overlapping) block of length $\ell = n/m$ by $\ell/2$ observations (assuming ℓ is even) rather than just one observation each time so that each block overlaps with just two others (except for the first and last blocks which overlap with just one other block). The result is a set of $2m-1$ moving blocks, each sub-sample being

$$Y_i = \left(y_{1+[(i-1)\ell/2]}, \ldots, y_{\ell+[(i-1)\ell/2]}\right)', \quad i = 1, \ldots, 2m-1,$$

where $[x]$ denotes the integer part of x. In this case $n-\ell = (n/m)(m-1)$ and hence

$$\hat{\beta}_{J,m}^{\mathrm{MB2}} = \left(\frac{m}{m-1}\right)\hat{\beta}_n - \left(\frac{1}{m-1}\right)\frac{1}{2m-1}\sum_{i=1}^{2m-1}\hat{\beta}_i. \tag{6}$$

Other types of moving-block sub-sampling can, of course, also be considered.

2.4 Higher-order bias reduction

The idea of re-applying the jackknife method in an attempt to reduce higher-order bias terms goes back to Quenouille (1956) and was further developed by Schucany, Gray and Owen (1971). As shown below it is not actually necessary to re-apply the jackknife itself because the ability to use different sub-sampling methods, or indeed to use statistics based on different numbers of a given type of sub-sample, enables higher-order bias corrections to be carried out directly. The result is presented in Theorem 2 below.

Theorem 2. *Let y and $\hat{\beta}_n$ be defined as in Theorem 1, and let $Y_{1,i}$ $(i = 1, \ldots, m_1)$ and $Y_{2,i}$ $(i = 1, \ldots, m_2)$ denote two differing sets of sub-samples of lengths ℓ_1 and ℓ_2 respectively, where $\ell_i = O(n)$ $(i = 1, 2)$. Let $\hat{\beta}_{1,i}$ $(i = 1, \ldots, m_1)$ and $\hat{\beta}_{2,i}$ $(i = 1, \ldots, m_2)$ denote the corresponding sub-sample statistics. Then the jackknife statistic*

$$\hat{\beta}_{J,(m_1,m_2)} = w_n\hat{\beta}_n + w_{1n}\frac{1}{m_1}\sum_{i=1}^{m_1}\hat{\beta}_{1,i} + w_{2n}\frac{1}{m_2}\sum_{i=1}^{m_2}\hat{\beta}_{2,i} \tag{7}$$

with weights given by

$$w_n = \frac{n^2}{(n-\ell_1)(n-\ell_2)}, \quad w_{1n} = -\frac{\ell_1^2}{(n-\ell_1)(\ell_1-\ell_2)}, \quad w_{2n} = \frac{\ell_2^2}{(n-\ell_2)(\ell_1-\ell_2)},$$

satisfies $E(\hat{\beta}_J) = \beta + O(n^{-3})$.

The sub-samples in Theorem 2 can be obtained either using different sub-sampling methods or using different numbers of sub-samples for a given method. As an example of the first type, consider the case where the $\hat{\beta}_{1,i}$ statistics are obtained from non-overlapping sub-samples and the $\hat{\beta}_{2,i}$ statistics are computed using moving blocks. In this case $\ell_1 = n/m_1$ and $\ell_2 = n-m_2+1$ so that $n - \ell_1 = (n/m_1)(m_1 - 1)$ and $n - \ell_2 = m_2 - 1$, resulting in the weights

$$w_n = \frac{nm_1}{(m_1 - 1)(m_2 - 1)}, \quad w_{1n} = -\frac{n}{m_1(m_1 - 1)(\ell_1 - \ell_2)}, \quad w_{2n} = \frac{\ell_2^2}{(m_2 - 1)(\ell_1 - \ell_2)}.$$

An example of the second type is where non-overlapping sub-samples are used for both methods, provided that $m_1 \neq m_2$. Then $\ell_1 = n/m_1$ and $\ell_2 = n/m_2$ so that $n - \ell_1 = (n/m_1)(m_1 - 1)$, $n - \ell_2 = (n/m_2)(m_2 - 1)$ and $\ell_1 - \ell_2 = (n/(m_1 m_2))(m_2 - m_1)$, yielding

$$w_n = \frac{m_1 m_2}{(m_1 - 1)(m_2 - 1)}, \quad w_{1n} = -\frac{m_2}{(m_1 - 1)(m_2 - m_1)}, \quad w_{2n} = \frac{m_1}{(m_2 - 1)(m_2 - m_1)}.$$

Other combinations of sub-sampling methods can, of course, be utilised.

2.5 Unequal sub-sample lengths

So far it has been assumed that the m sub-samples each have equal length ℓ, but in practical circumstances it is desirable to allow for situations in which this is not the case. For example, in the case of non-overlapping sub-samples, taking $m = 4$ with a sample of size $n = 50$ means that at least one sub-sample is of a different size to the others. In this example it would be possible to have three sub-samples of length $\ell = 12$ and one of length $\ell = 14$ or two of length $\ell = 12$ and two of length $\ell = 13$. Once different sub-sample lengths are employed the appropriate weights to use in constructing the jackknife estimator change from those presented previously. The following result extends Theorem 1 to deal with this situation.

Theorem 3. *Let y and $\hat{\beta}_n$ be defined as in Theorem 1, and let $Y_{1,i}$ $(i = 1, \ldots, m_1)$ and $Y_{2,i}$ $(i = m_1 + 1, \ldots, m_1 + m_2)$ denote two differing sets of sub-samples of lengths ℓ_1 and ℓ_2 respectively, where $\ell_i = O(n)$ $(i = 1, 2)$ and $m_1 + m_2 = m$. Let $\hat{\beta}_i$ $(i = 1, \ldots, m)$ denote the corresponding sub-sample statistics. Then the jackknife statistic*

$$\hat{\beta}_J^U = k_n \hat{\beta}_n + k_{1n} \frac{1}{m} \sum_{i=1}^{m} \hat{\beta}_i, \tag{8}$$

with weights given by

$$k_n = \frac{m_1 n(\ell_1 - \ell_2) - mn\ell_1}{m_1 n(\ell_1 - \ell_2) - m\ell_1(n - \ell_2)}, \quad k_{1n} = -\frac{m\ell_1\ell_2}{m_1 n(\ell_1 - \ell_2) - m\ell_1(n - \ell_2)},$$

satisfies $E(\hat{\beta}_J) = \beta + O(n^{-2})$.

The result in Theorem 3 assumes that there are two different sub-sample lengths in use which should be sufficient for most applications, although extending the result to more than two sub-sample lengths is straightforward. Alternatively, if $m\ell < n$, it would be possible to simply ignore the first $n - m\ell$ observations, although in relatively small samples such discarding of data may not be particularly desirable. Asymptotically, however, such discarding of data may be less important, as pointed out by Hall, Horowitz and Jing (1995) in the context of bootstrap blocking rules with dependent data.

3. Jackknife estimation in stationary autoregressions

The focus is the AR(p) process defined in Assumption 1.

Assumption 1. The observations $y_1, \ldots, y_n$ satisfy

$$y_t = \alpha + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \epsilon_t, \quad t = 1, \ldots, n, \tag{9}$$

where $\epsilon_t \sim$ i.i.d.$(0, \sigma^2)$ with finite fourth moment and the roots of the equation $\phi(z) = 0$, where $\phi(z) = 1 - \phi_1 z - \ldots \phi_p z^p$, lie outside the unit circle.

It is convenient to write the model in the form

$$y_t = x_t'\beta + \epsilon_t, \quad t = 1, \ldots, n, \tag{10}$$

where $x_t = (1, y_{t-1}, \ldots, y_{t-p})'$ and $\beta = (\alpha, \phi_1, \ldots, \phi_p)'$, and to assume that $y_{-p+1}, \ldots, y_0$ are also observed so that the ordinary least squares (OLS) estimator of β is

$$\hat{\beta}_n = \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \sum_{t=1}^{n} x_t y_t = \beta + \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \sum_{t=1}^{n} x_t \epsilon_t.$$

Writing $\beta = (\alpha, \phi')'$, where $\phi = (\phi_1, \ldots, \phi_p)'$ is the vector of autoregressive parameters, Shaman and Stine (1988) demonstrate that the expectation of the OLS estimator of ϕ satisfies a Nagar-type expansion of the form in (1) i.e. that

$$E(\hat{\phi}_n) = \phi + \frac{a_1}{n} + O(n^{-2}),$$

where $\hat{\phi}_n$ denotes the appropriate sub-vector of $\hat{\beta}_n$, i.e. $\hat{\beta}_n = (\hat{\alpha}_n, \hat{\phi}_n')'$. It seems reasonable to expect that the entire vector $\hat{\beta}_n$ also satisfies a similar expansion in view of the results of Bao and Ullah (2007) for bias expansions in general time series models. For $p = 1, 2, 4$ the first-order bias vector a_1 for $\hat{\phi}_n$ is given in Table 1 of Shaman and Stine (1988) by:

$$p = 1: \quad a_1 = (1 - 3\phi_1);$$

$$p = 2: \quad a_1 = (1 - \phi_1 - \phi_2, 2 - 4\phi_2)';$$

$$p = 4: \quad a_1 = (1 - \phi_1 - \phi_4, 2 + \phi_1 - 2\phi_2 - \phi_3 - 2\phi_4, 1 + 2\phi_1 - 5\phi_3 - \phi_4, 2 - 6\phi_4)'.$$

The proof of the existence of the first-order bias term a_1/n for $\hat{\phi}_n$ by Shaman and Stine (1988) suggests that the jackknife methods may well be successful in removing the first-order bias in finite samples in view of the results in section 2. For example, considering the jackknife estimator of ϕ based on non-overlapping sub-intervals as in (4), we would expect $E(\hat{\phi}_{J,m}) = \phi + O(n^{-2})$, and the ability of the jackknife estimators to reduce bias in this way is explored via the use of simulations in the next section. However, it is also of interest to examine the asymptotic properties of the jackknife estimators, especially if they are to be used for the purposes of inference about the autoregressive parameters.

Assumption 1 ensures that, as $n \to \infty$, $\hat{\beta}_n \xrightarrow{p} \beta$ and, furthermore, that

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1}),$$

where

$$Q = E(x_t x_t') = \begin{bmatrix} 1 & \mu & \mu & \dots & \mu \\ \mu & \gamma_0 + \mu^2 & \gamma_1 + \mu^2 & \dots & \gamma_{p-1} + \mu^2 \\ \mu & \gamma_1 + \mu^2 & \gamma_0 + \mu^2 & \dots & \gamma_{p-2} + \mu^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu & \gamma_{p-1} + \mu^2 & \gamma_{p-2} + \mu^2 & \dots & \gamma_0 + \mu^2 \end{bmatrix}, \tag{11}$$

$\mu = E(y_t) = \alpha/\phi(1)$ and $\gamma_j = E(y_t y_{t-|j|})$; such results are standard in the literature, an excellent textbook treatment being Hamilton (1994, p.216). The jackknife estimator based on non-overlapping sub-samples, of the type used by Phillips and Yu (2005), is defined by

$$\hat{\beta}_{J,m} = \frac{m}{m-1}\hat{\beta}_n - \left(\frac{1}{m-1}\right)\frac{1}{m}\sum_{j=1}^{m}\hat{\beta}_j, \tag{12}$$

where m denotes the number of sub-samples, each having length ℓ, where $n = m \times \ell$, and the sub-sample estimators are

$$\hat{\beta}_j = \left(\sum_{t\in\tau_j} x_t x_t'\right)^{-1} \sum_{t\in\tau_j} x_t y_t, \quad j = 1, \dots, m, \tag{13}$$

with $\tau_j = \{(j-1)\ell + 1, \dots, j\ell\}$ denoting the set of integers determining the observations in sub-sample j. Two scenarios concerning the number of sub-samples, m, shall be considered. The first is where m is fixed as $n \to \infty$, while the second allows m to increase with, but more slowly than, n. In both cases ℓ increases with n, an implication under Assumption 1 being that the sub-sample estimators satisfy $\hat{\beta}_j \xrightarrow{p} \beta$ as $\ell \to \infty$ and that

$$\sqrt{\ell}(\hat{\beta}_j - \beta) \xrightarrow{d} N(0, \sigma^2 Q_j^{-1}), \quad j = 1, \dots, m, \tag{14}$$

where $Q_j = \text{plim}_{\ell\to\infty}\ell^{-1}\sum_{t\in\tau_j} x_t x_t'$; clearly, $Q_j = Q$ and so the sub-sample estimators have the same limiting distribution as the full-sample estimator, $\hat{\beta}_n$. The limiting properties of $\hat{\beta}_{J,m}$ are given below for both scenarios concerning m.

Theorem 4. *Let $y = (y_1, \dots, y_n)'$ be a sample of n observations satisfying Assumption 1 and let $\hat{\beta}_{J,m}$ be defined as in (12). Then, as $n \to \infty$, $\hat{\beta}_{J,m} \xrightarrow{p} \beta$ and*

$$\sqrt{n}(\hat{\beta}_{J,m} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1}),$$

both when m is fixed and when $m^{-1} + mn^{-1} \to 0$.

Theorem 4 shows that $\hat{\beta}_{J,m}$ is consistent and asymptotically normally distributed and, furthermore, that there is no loss of asymptotic efficiency compared to $\hat{\beta}_n$. Combined with the bias reduction in finite samples this suggests that the jackknife method (at least the one based on non-overlapping sub-samples) may be a useful method of estimation in stationary autoregressive models. It is also interesting that this result holds irrespective of whether the number of sub-samples increases with n or is held fixed.

As far as asymptotic inference about β is concerned the distribution in Theorem 4 provides a suitable basis, and a consistent estimator of $V_J = \sigma^2 Q^{-1}$ is required. The following result demonstrates the consistency of an estimator of σ^2 based on the jackknife residuals as well as the consistency of two possible estimators of V_J.

Theorem 5. *Let $y = (y_1, \ldots, y_n)'$ be a sample of n observations satisfying Assumption 1 and let $\hat{\beta}_{J,m}$ be defined as in (12). Furthermore let $e_{t,J} = y_t - x_t'\hat{\beta}_{J,m}$ $(t = 1, \ldots, n)$ denote the residuals from the jackknife estimation, let*

$$\hat{\sigma}_J^2 = \frac{1}{n-p-1}\sum_{t=1}^{n} e_{t,J}^2 \tag{15}$$

denote an estimator of σ^2 based on these residuals, and define the following two estimators of V_J:

$$\hat{V}_J = \hat{\sigma}_J^2 \left(\frac{1}{n}\sum_{t=1}^{n} x_t x_t'\right)^{-1}, \tag{16}$$

$$\hat{V}_{J,m} = \frac{m(m-2)}{(m-1)^2}\hat{\sigma}_J^2 \left(\frac{1}{n}\sum_{t=1}^{n} x_t x_t'\right)^{-1} + \frac{1}{m(m-1)^2}\hat{\sigma}_J^2 \sum_{j=1}^{m}\left(\frac{1}{\ell}\sum_{t\in\tau_j} x_t x_t'\right)^{-1}. \tag{17}$$

Then, as $n \to \infty$, $\hat{\sigma}_J^2 \xrightarrow{p} \sigma^2$, $\hat{V}_J \xrightarrow{p} V_J$ and $\hat{V}_{J,m} \xrightarrow{p} V_J$, both for fixed m and when $m^{-1} + mn^{-1} \to 0$.

The consistency of the estimator $\hat{\sigma}_J^2$ follows by standard arguments, using the consistency of $\hat{\beta}_{J,m}$. The estimator $\hat{V}_J$ is, in many ways, the 'obvious' estimator of V_J, although it does not take into account sub-sample information in the way that the jackknife estimator $\hat{\beta}_{J,m}$ does. Given that Q is also consistently estimated by the sub-sample moment matrices suggests that it may be advantageous to incorporate these sub-sample quantities in a single expression in much the same way as the jackknife estimator itself utilises the sub-sample estimators of β. Lemma 2 in Appendix B provides an expression for the variance matrix of $\hat{\beta}_{J,m}$ based on OLS regression formulae which motivates the definition of $\hat{V}_{J,m}$ in (17).

4. Finite sample bias reduction

In this section the finite sample bias reduction properties of the jackknife estimators of autoregressive parameters are explored using simulations. The initial focus is bias reduction in the AR(1) model, including an assessment of the effects of non-normal and ARCH disturbances, before higher-order and misspecified autoregressive models are considered. Comparisons with alternative estimators are also provided, while other distributional considerations and the issue of inference are explored in section 5.

4.1 AR(1) model with i.i.d. Normal disturbances

The model under consideration is defined by

$$y_t = \alpha + \phi y_{t-1} + \epsilon_t, \quad |\phi| < 1, \quad \epsilon_t \sim \text{i.i.d. } N(0, \sigma^2), \quad t = 1, \ldots, n, \tag{18}$$

which satisfies Assumption 1 with $p = 1$. The full-sample OLS estimator of $\beta = (\alpha, \phi)'$ is

$$\hat{\beta}_n = \begin{pmatrix} \hat{\alpha}_n \\ \hat{\phi}_n \end{pmatrix} = \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1}\sum_{t=1}^{n} x_t y_t,$$

where $x_t = (1, y_{t-1})'$ and y_0 is assume to be observed. In this model Sawa (1978) found the following moment expansion, originally due to Kendall (1954), to work well, despite its simplicity:

$$E\left(\hat{\phi}_n - \phi\right) = -\frac{1+3\phi}{n} + O\left(n^{-2}\right). \tag{19}$$

Using results in Bao (2007) while maintaining normality shows how y_0 affects the $O(n^{-2})$ term:

$$E\left(\hat{\phi}_n - \phi\right) = -\frac{1+3\phi}{n} + \frac{1}{n^2}\left[\frac{3\phi - 9\phi^2 - 1}{1-\phi} + \frac{1+3\phi}{(1-\phi)^2}\left(\frac{g(\alpha,\phi,y_0)}{\sigma}\right)^2\right] + O\left(n^{-3}\right) \tag{20}$$

where $g(\alpha, \phi, y_0) = \alpha - (1-\phi)y_0$.

In the simulations the intercept was chosen to satisfy $\alpha = (1-\phi)y_0$ so as to remove the dependence of the $O(n^{-2})$ term on y_0 and also to make $\hat{\phi}_n$ invariant to α and y_0. A total of 100,000 replications of each experiment were carried out with $\sigma^2 = 1$ and $y_0 = 0$ (without loss of generality). The simulations are based on a range of autoregressive parameter values that would appear to be most relevant in practice, so that $\phi_1 \in \{0.1, 0.3, 0.5, 0.7, 0.9, 0.95, 0.99\}$, and the sample sizes are $n \in \{24, 48, 96, 192\}$. These particular sample sizes enable a range of values of m, the number of non-overlapping sub-samples used in constructing the jackknife estimator, to be considereded.

The first jackknife estimator considered was the estimator based on non-overlapping sub-samples defined in (4). In view of the above expansions this estimator of ϕ satisfies

$$E\left(\hat{\phi}_{J,m} - \phi\right) = \frac{1}{n^2}\left(\frac{3\phi - 9\phi^2 - 1}{1-\phi}\right) + O(n^{-3}),$$

and a key issue to address is how the remaining finite sample bias is affected by the choice of m, the number of sub-samples. Table 1 reports the bias of the OLS estimator $\hat{\phi}_n$ as well as the jackknife estimators $\hat{\phi}_{J,m}$ for $m \in \{2, 3, 4, 6, 8\}$. It is clear that, for all sample sizes and parameter values, each of the jackknife estimators produces substantial bias reduction compared to the OLS estimator, with the bias increasing in absolute value monotonically with m. The minimum bias is obtained using just $m = 2$ sub-samples. For $\phi = 0.1$ and $n = 24$ the resulting bias reduction is of the order of 94% while for $\phi = 0.99$ and $n = 24$ it is 78% rising to 98% when $n = 192$.

The performance of other jackknife estimators was also examined. Table 2 contains the bias of the OLS estimator $\hat{\phi}_n$, the jackknife estimator $\hat{\phi}_{J,2}$ from Table 1, the moving-block jackknife estimators $\hat{\phi}_{J,2}^{\mathrm{MB}}$ and $\hat{\phi}_{J,2}^{\mathrm{MB2}}$ defined in (5) and (6) respectively, and the second-order jackknife estimator $\hat{\phi}_{J,(2,3)}$ defined in Theorem 2 and based on non-overlapping sub-samples. In eighteen of the twenty-eight combinations of ϕ and n reported in Table 2 it is the estimator $\hat{\phi}_{J,(2,3)}$ that produces the smallest absolute bias, improving even on the performance of the estimator $\hat{\phi}_{J,2}$ in Table 1. For example, when $\phi = 0.1$ and $n = 24$ the resulting bias reduction is of the order of 96% while for $\phi = 0.99$ and $n = 24$ it is 92% rising to 98% when $n = 192$. In only two cases does the estimator $\hat{\phi}_{J,2}^{\mathrm{MB}}$ produce the minimum bias while $\hat{\phi}_{J,2}^{\mathrm{MB2}}$ does so in none of the cases, although both estimators do reduce the bias substantially compared to the OLS estimator.

It is also of interest and of some importance to compare the performance of the jackknife estimators not only with the OLS estimator but also with other rival estimators that have been developed to improve on the properties of $\hat{\phi}_n$. For this purpose two further estimators of ϕ were considered, these being the exact median unbiased estimator (MUE) of Andrews (1993) and a recursive-design wild bootstrap of the type proposed by Gonçalves and Kilian (2004) for autoregressions with non-i.i.d. disturbances (with a view to relaxing the i.i.d. assumption in the next section). Other estimators are, of course, also possible, including generalised method of moments estimators, based on unconditional moments, as well as estimators that exploit conditional moment restrictions. Examples of the latter estimators can be found in Gospodinov and Otsu (2009) who extend previous work of, for example, Kitamura, Tripathi and Ahn (2004) and Smith (2007) to the case of dependent data, in particular AR models. These particular alternatives are not pursued here, however, with the focus of the comparisons of the jackknife estimators being with the exact MUE and the bootstrap in addition to the OLS estimator.

The exact MUE estimator, denoted $\hat{\phi}_{\text{MU}}$, is defined by

$$\hat{\phi}_{\text{MU}} = \begin{cases} 1 & \text{if } \hat{\phi}_n > \text{med}(1); \\ \text{med}^{-1}(\hat{\phi}_n) & \text{if } \text{med}(-1) < \hat{\phi}_n \leq \text{med}(1); \\ -1 & \text{if } \hat{\phi}_n \leq \text{med}(-1), \end{cases}$$

where $\text{med}(\phi)$ is the median function of $\hat{\phi}_n$. The function $\text{med}(\phi)$ was calculated as in Andrews (1993) at a grid of points in $(-1, 1]$ for each of the four sample sizes and then linear interpolation was used to determine $\hat{\phi}_{\text{MU}}$. The bootstrap estimator, denoted $\hat{\phi}_{\text{BS}}$, is defined as follows. Let $e_t = y_t - x_t'\hat{\beta}_n$ $(t = 1, \ldots, n)$ denote the full-sample OLS residuals. Then, with $y_0 = 0$, a total of $B = 1000$ bootstrap samples are computed recursively according to

$$y_t^* = x_t^*\hat{\beta}_n + \epsilon_t^*, \quad t = 1, \ldots, n,$$

where $x_t^* = [1, y_{t-1}^*]'$ and $\epsilon_t^* = e_t\eta_t$ with η_t an i.i.d. sequence with mean zero and unit variance, which was taken to be $N(0, 1)$. For each bootstrap sample the OLS vector

$$\hat{\beta}_{n,b}^* = \left(\sum_{t=1}^{n} x_t^* x_t^{*\prime}\right)^{-1} \sum_{t=1}^{n} x_t^* y_t^*, \quad b = 1, \ldots, B,$$

is obtained and then the bias-corrected estimator

$$\hat{\phi}_{\text{BS}} = 2\hat{\beta}_n - \frac{1}{B}\sum_{b=1}^{B} \hat{\beta}_{n,b}^*$$

is computed. This bootstrap process is repeated for each of the 100,000 replications.

The results of the comparison are contained in Table 3, the entries in which are the simulated biases of $\hat{\phi}_n$, $\hat{\phi}_{J,2}$, $\hat{\phi}_{J,(2,3)}$, $\hat{\phi}_{\text{MU}}$ and $\hat{\phi}_{\text{BS}}$. The new estimators under consideration, $\hat{\phi}_{\text{MU}}$ and $\hat{\phi}_{\text{BS}}$, both have considerably reduced bias compared to $\hat{\phi}_n$, although it is the second-order bias reduced jackknife estimator $\hat{\phi}_{J,(2,3)}$ that produces the smallest absolute bias in nineteen of the twenty-eight cases, the other jackknife estimator $\hat{\phi}_{J,2}$ showing the minimum bias in the remaining nine cases.

4.2 Non-normal and ARCH disturbances

All results reported so far have been based on $\{\epsilon_t\}_{t=1}^n$ being an i.i.d. Normal sequence. Indeed most theoretical results concerning the moments of the OLS estimator in the AR(1) model are based on such an assumption, although Bao (1997) has derived analagous expansions for bias and mean square error that allow for non-normality. Two departures from normality are considered here; it is convenient to let μ_3 and μ_4 denote the skewness and kurtosis coefficients, respectively.

The first generates the ϵ_t from a Student's t-distribution with five degrees of freedom, in which case $E(\epsilon_t) = 0$, $var(\epsilon_t) = 5/3$, $\mu_3 = 0$ and $\mu_4 = 9$. The second generates ϵ_t as a sequence of (mean-corrected) gamma variates. If a random variable $x \sim \Gamma(a, b)$ then $E(x) = ab$, $var(x) = ab^2$, $\mu_3 = 2/\sqrt{a}$ and $\mu_4 = 3 + (6/a)$. Setting $a = 1$ ensures that the Gamma variate has the same kurtosis as the t_5 variate (i.e. $\mu_4 = 9$), such a choice yielding a skewness coefficient of $\mu_3 = 2$. It is also possible, by appropriate choice of b, to ensure that the variance is equal to that of the t_5, namely 5/3; this requires $b = \sqrt{5/3}$, the resulting mean being $\sqrt{5/3}$. Here $\epsilon_t \sim \Gamma(1, \sqrt{5/3})$ so that $(\epsilon_t - \sqrt{5/3})$ has zero mean, variance and kurtosis equal to the t_5 variate, but has skewness equal to 2. The t-variates therefore introduce kurtosis relative to the normal while the gamma variates additionally introduce skewness.

Table 4 reports the bias of the estimators $\hat{\phi}_n$, $\hat{\phi}_{J,2}$, $\hat{\phi}_{J,(2,3)}$, $\hat{\phi}_{\text{MU}}$ and $\hat{\phi}_{\text{BS}}$ under Student's t and Gamma disturbances as outlined above for values of $\phi \in \{0.5, 0.7, 0.9, 0.95, 0.99\}$. The actual bias of the OLS estimator $\hat{\phi}_n$ is typically smaller in absolute terms than under normality (compare the entries in Table 3) but the jackknife and bootstrap estimators are still able to reduce the bias significantly. The estimator that is adversely affected by the non-normality is $\hat{\phi}_{\text{MU}}$ where, in some cases, the bias is larger than that of $\hat{\phi}_n$, although perhaps this should not be too surprising as the median function used in the determination of $\hat{\phi}_{\text{MU}}$ is based explicitly on the assumption of normality. Under the Student's t-distribution the estimator $\hat{\phi}_{J,2}$ produces the minimum (absolute) bias in seven of the twenty cases with $\hat{\phi}_{J,(2,3)}$ producing the minimum value in the remaining thirteen, while under the Gamma distribution the numbers are six and fourteen, respectively.

In the case of ARCH disturbances $\epsilon_t = h_t v_t$ where v_t is i.i.d. $N(0, 1)$,

$$h_t^2 = \sigma^2(1 - \lambda) + \lambda\epsilon_{t-1}^2,$$

σ^2 denotes the unconditional variance and λ is the ARCH parameter. Two values of λ are considered, $\lambda = 0.5$ and $\lambda = 0.9$ which (with $\sigma^2 = 1$) produce time-varying conditional variances of $0.5 + 0.5\epsilon_{t-1}^2$ and $0.1 + 0.9\epsilon_{t-1}^2$ respectively. In the former case the kurtosis, given by $\mu_4 = 3(1 - \lambda^2)/(1 - 3\lambda^2)$, is equal to 9 and matches the kurtosis of the t_5 and Gamma variates used previously, while in the second case the kurtosis is infinite. The simulated biases of the five estimators considered above are reported in Table 5 for values of $\phi \in \{0.5, 0.7, 0.9, 0.95, 0.99\}$. The OLS bias tends to be larger in magnitude than the corresponding values in Table 3 and, although both jackknife estimators are capable of substantial bias reduction, it is not as great as when the disturbances are i.i.d. (normal or non-normal) and tends to be smaller for the larger value of the ARCH parameter. Although the two jackknife estimators produce the smallest (absolute) bias in nineteen of the twenty cases when $\lambda = 0.5$ this dominance falls when λ rises to 0.9, with $\hat{\phi}_{J,(2,3)}$ producing the minimum value in half of the cases and the bootstrap estimator producing the minimum in eight cases. The improvement of the bootstrap estimator relative to the jackknife methods is perhaps due to the recursive-design wild bootstrap method having been proposed by Gonçalves and Kilian (2004) precisely for autoregressions with conditional heteroskedasticity.

4.3 Higher-order and misspecified autoregressions

In order to examine the preformance of the jackknife and alternative estimators in higher-order autoregressions simulations were carried out for the AR(p) model with i.i.d. Normal disturbances with $p = 2$ and $p = 4$, the analysis of the effects of misspecification being conducted by estimating the over- or under-parameterised model. Instead of reporting the results for each individual coefficient the attention focuses instead on the persistence measure ρ which is equal to the sum of the autoregressive coefficients. The model is defined (for general p) by

$$y_t = \alpha + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \epsilon_t, \quad t = 1, \ldots, n, \tag{21}$$

where $\alpha = \mu(1-\rho)$, $\rho = \sum_{i=1}^{p} \phi_i$, $\mu = E(y_t)$ and $y_0 = y_{-1} = \ldots = y_{-p+1} = \mu$. The parameter values, taken from Patterson (2000), are as follows. In the AR(2) model, $\phi_1 = 1.25$ and $\phi_2 = -0.35$, while in the AR(4) model $\phi_1 = 1.20$, $\phi_2 = -0.55$, $\phi_3 = 0.40$ and $\phi_4 = -0.15$. The corresponding roots are 2.3616 and 1.2098 in the AR(2) model and -0.2428 ± 1.6834, 2 and 1.1523 in the AR(4) model. In both cases the persistence measure $\rho = 0.90$ and the unconditional mean of y_t is taken to be $\mu = 0.1$. The parameters were chosen to be consistent with the high persistence found in many macroeconomic time series.

The results of the simulations are reported in Table 6 which contains the bias of $\hat{\rho}_n$, $\hat{\rho}_{J,2}$, $\hat{\rho}_{J,(2,3)}$ and $\hat{\rho}_{\mathrm{BS}}$. Not surprisingly the biases of all estimators have a tendency to be higher when the model is misspecified, but the smallest (absolute) biases are obtained with the jackknife estimators, particularly $\hat{\rho}_{J,(2,3)}$ which produces the minimum bias in twelve out of the sixteen cases considered in Table 6 with $\hat{\rho}_{J,2}$ producing the smallest bias in the four remaining cases. Accurate estimation of ρ is important in determining long-run multiplier effects based on stationary autoregressions as well as in the estimation of spectral densities at the origin; see, for example, Berk (1974). The estimation of such quantities is also important in the construction of the modified unit root statistics of Ng and Perron (2001).

5. Other finite sample properties and inference

Although finite sample bias reduction is an important and attractive feature of the jackknife estimators considered in the previous section, there are also other aspects of these estimators' performance that are important to consider. In this section three further considerations are addressed, namely median-unbiasedness, mean squared error, and coverage rates of confidence intervals.

5.1 Median-unbiasedness

An estimator $\tilde{\beta}$ of β is said to be median-unbiased if

$$\Pr(\tilde{\beta} \geq \beta) \geq 0.5 \quad \text{and} \quad \Pr(\tilde{\beta} \leq \beta) \geq 0.5.$$

This concept can be of more relevance than mean-unbiasedness in situations where the distribution of the estimator is asymmetric or skewed, as in the case of the OLS estimator of the autoregressive parameter. Although the jackknife estimators are not designed to be median-unbiased it is nevertheless of interest to assess whether they also provide advantages over OLS in this regard. Table 7 reports the percentage of the distribution of the estimators $\hat{\phi}_n$, $\hat{\phi}_{J,2}$, $\hat{\phi}_{J,(2,3)}$, $\hat{\phi}_{\mathrm{MU}}$ and $\hat{\phi}_{\mathrm{BS}}$ for which the bias is negative in the AR(1) model when subject to Normal and Gamma disturbances. The OLS estimator suffers from substantial skewness especially for larger values of ϕ but this is, to a large extent, eliminated by all of the estimators under Normal distrubances, the MU estimator working particularly well in this setting, as it is designed to do. The improvements are particularly striking for larger values of ϕ for which the distribution of the OLS estimator $\hat{\phi}_n$ is severely skewed.

The estimators also show improvements over OLS when the error distribution is Gamma instead of Normal, although the MU estimator loses its advantages due to an incorrect (Normal-based) median function being used. The jackknife estimators appear to enjoy good median-unbiased properties even when ϕ is large and also for small sample sizes.

5.2 Mean squared error

Although the jackknife method is designed to eliminate the first-order bias of a statistic, MSE considerations are also of interest to explore. In the AR(1) model with intercept the results of Bao (2007) provide (under normality)

$$E\left(\hat{\phi}_n - \phi\right)^2 = \frac{1-\phi^2}{n} + \frac{1}{n^2}\left[23\phi^2 + 10\phi - \frac{1+\phi}{1-\phi}\left(\frac{g(\alpha,\phi,y_0)}{\sigma}\right)^2\right] + O\left(n^{-3}\right). \tag{22}$$

When $\alpha = (1-\phi)y_0$, as in the simulations, the effect of y_0 (and σ^2) in the $O(n^{-2})$ term is eliminated. Table 8 reports the root mean squared error (RMSE) of the estimators $\hat{\phi}_n$, $\hat{\phi}_{J,2}$, $\hat{\phi}_{J,(2,3)}$, $\hat{\phi}_{\mathrm{MU}}$ and $\hat{\phi}_{\mathrm{BS}}$, along with two additional jackknife estimators, denoted $\hat{\phi}_{J,m_r}$ and $\hat{\phi}_{J,(m_1,m_2)_r}$. The former is the jackknife estimator $\hat{\phi}_{J,m}$ obtained using the RMSE-minimising value of m, denoted m_r, while the latter is the estimator $\hat{\phi}_{J,(m_1,m_2)}$ obtained using the RMSE-minimising values of m_1 and m_2, denoted $(m_1,m_2)_r$. The values m_r and $(m_1,m_2)_r$ were determined by computing the estimators for a range of values of $m \in \{2,3,4,6,8,12,16,24,48\}$ (subject to sample size) and combinations of m_1 and m_2 from this set of values, and choosing those values resulting in the smallest simulated RMSE. The values of m_r and $(m_1,m_2)_r$ are reported in Table 8 in superscripts next to the appropriate entry. The estimators $\hat{\phi}_{J,2}$, $\hat{\phi}_{J,(2,3)}$ are, of course, the jackknife estimators that minimise bias, and so it is useful to see not only how these estimators based on the bias-minimising values of m perform in an RMSE sense but also to ascertain how these bias-minimising values compare with the RMSE-minimising values.

For the smallest values of ϕ considered, namely $\phi = 0.1$ and $\phi = 0.3$, the OLS estimator $\hat{\phi}_n$ results in the smallest RMSE, although $\hat{\phi}_{J,m_r}$ and $\hat{\phi}_{\mathrm{MU}}$ are on a par for the largest sample size when $\phi = 0.3$. For larger values of ϕ it is the estimator $\hat{\phi}_{\mathrm{MU}}$ that dominates, producing the smallest RMSE value in fifteen of the remaining twenty cases. This is presumably due to the way in which the estimator never exceeds the value of one by definition, a truncation that is not a feature of the other estimators and which would tend to reduce the variance for larger values of ϕ. The estimators that provide the smallest RMSE in the other cases are $\hat{\phi}_{J,m_r}$ (in seven cases, three jointly with $\hat{\phi}_{\mathrm{MU}}$) and $\hat{\phi}_{J,(m_1,m_2)}$ (in one case).

In terms of the jackknife estimators, it appears in most cases in Table 8 that the greatest bias reduction (as demonstrated by $\hat{\phi}_{J,2}$ and $\hat{\phi}_{J,(2,3)}$) comes at the cost of increased variance which also increases the RMSE as compared to $\hat{\phi}_n$. This increased variance is particularly acute in the case of $\hat{\phi}_{J,(2,3)}$. However, Table 8 also shows that the RMSE of the jackknife estimators can be reduced by suitable choice of m and (m_1,m_2), and these RMSE-minimising values are shown to be an increasing function of sample size and to be larger than the bias-minimising values.

5.3 Confidence intervals

Although bias and RMSE reduction may be important aspects of an estimator, the issue of statistical inference is also important, and so the coverage rates of (nominal) 90% confidence intervals of the jackknife estimators were compared with those of rival methods, namely OLS, the MUE and the bootstrap. The benchmark case was taken to be the OLS estimator, $\hat{\phi}_n$, of the parameter ϕ in the AR(1) model (18) with normally distributed innovations. The limiting distribution of $\sqrt{n}(\hat{\phi}_n - \phi)$ is $N(0, \sigma^2 Q_{22}^{-1})$,$Q_{22}^{-1}$ being the second diagonal element of the matrix Q^{-1} where $Q = \text{plim}_{n\to\infty} n^{-1} \sum_{t=1}^n x_t x_t'$ and $x_t = (1, y_{t-1})'$. A two-sided asymptotic 90% confidence interval can therefore be constructed as

$$\hat{\phi}_n \pm 1.645 \frac{\hat{\sigma}\sqrt{\hat{Q}_{22}^{-1}}}{\sqrt{n}},$$

where $\hat{\sigma}^2$ is the OLS error variance estimator and $\hat{Q}_{22}^{-1}$ is the second diagonal element of the matrix $(n^{-1} \sum_{t=1}^n x_t x_t)^{-1}$. From Theorems 4 and 5 a similar two-sided asymptotic 90% confidence interval based on the jackknife estimator $\hat{\phi}_{J,m}$ can be constructed as

$$\hat{\phi}_{J,m} \pm 1.645 \frac{\sqrt{\hat{V}_{J,m,22}}}{\sqrt{n}},$$

where $\hat{V}_{J,m,22}$ denotes the second diagonal element of the matrix $\hat{V}_{J,m}$ defined in Theorem 5. A similar procedure can be applied for the second-order jackknife estimator $\hat{\phi}_{J,(m_1,m_2)}$ resulting in

$$\hat{\phi}_{J,(m_1,m_2)} \pm 1.645 \frac{\sqrt{\hat{V}_{J,(m_1,m_2),22}}}{\sqrt{n}},$$

where $\hat{V}_{J,(m_1,m_2),22}$ denotes the second diagonal element of an estimator of the asymptotic covariance matrix of $\hat{\phi}_{J,(m_1,m_2)}$ which can be derived from the expression for the variance matrix given in Lemma 3 in Appendix B. For the MUE the technique of obtaining (exact) confidence intervals is outlined in Andrews (1993, pp. 152–153) for which linear interpolation between a grid of values for the sample sizes used here is employed. Finally, the bootstrap is used to determine appropriate critical values for the confidence interval based on $\hat{\phi}_n$ by simulating the distribution using $B = 1000$ replications, yielding upper and lower values, c_U and c_L, each of which puts 5% of the simulated distribution into the relevant tail. The resulting two-sided 90% bootstrap confidence interval is

$$\left[\hat{\phi}_n - c_U \frac{\hat{\sigma}\sqrt{\hat{Q}_{22}^{-1}}}{\sqrt{n}}, \hat{\phi}_n - c_L \frac{\hat{\sigma}\sqrt{\hat{Q}_{22}^{-1}}}{\sqrt{n}}\right].$$

The results of 100,000 simulations for $\phi \in \{0.5, 0.7, 0.9, 0.95, 0.99\})$ are reported in Table 9, the entries being the estimated type I error measured as the propertion of replications in which the true parameter value was outside the confidence interval. Two variants of each jackknife estimator were considered, these corresponding to the bias- and RMSE-minimising values of m and (m_1, m_2). For $\phi = 0.5$ all of the entries in Table 9 are relatively close to the nominal value of 0.1, the largest discrepancies occuring with the bootsrap, particularly for small samples. The most accurate confidence intervals are based on the MUE. As ϕ approaches unity the OLS-based type I errors increase substantially and all of the alternative methods are able to improve the coverage of the confidence intervals. Again the MUE-based confidence intervals are the most accurate, although this performance is based on explicit knowledge of normality and would likely disappear under non-normal disturbances.

The coverage rates of the OLS-based confidence intervals reported in Table 9 are reasonably close to the nominal 10% level but actually mask large discrepancies in each tail. This feature is revealed more fully in Table 10 which reports, for each method, the separate upper and lower tail probabilities, denoted P_U and P_L respectively, of these two-sided confidence intervals. They can equivalently be regarded as the type I errors associated with one-sided 95% confidence intervals. It is clear from Table 10 that the OLS-based confidence intervals have much larger type I errors in the upper tail than in the lower tail, the divergence becoming increasingly severe as ϕ tends towards unity. The other methods are better able to equalise these upper and lower tail probabilities, particularly as the sample sizes increase. The MUE-based confidence intervals achieve the best equalisation overall.

6. Concluding comments

This paper, in addition to providing some general theoretical results concerning jackknife methods, has conducted an extensive investigation into the use of the jackknife as a method of estimation and inference in stationary autoregressive models. A method based on the use of non-overlapping sub-intervals is found to work particularly well and is capable of reducing bias and RMSE compared to OLS, subject to a suitable choice of the number of sub-samples. The jackknife estimators also outperform OLS and other estimators when the distribution of the disturbances departs from normality and when it is subject to ARCH effects, and is much closer to being median-unbiased. The performance of jackknife-based confidence intervals is also competitive with rival methods and is able to improve on the OLS-based confidence intervals, although choosing the number of sub-intervals is an important consideration.

The results obtained in this paper are encouraging for the use of jackknife methods in time series models, and many further avenues present themselves for exploration. A natural extension of potential importance is to consider jackknife methods of bias reduction applied to AR models containing a unit root in which the OLS estimator is known to be severely negatively biased. An analysis of bias reduction in the unit root case can be found in Chambers and Kyriacou (2011), and further work (in progress) is examining the performance of the bias-reduced jackknife estimators in actually testing for unit roots. In addition data-based methods for the determination of the optimal number of sub-samples for bias- and RMSE-minimisation would appear to be an important avenue of investigation.

Appendix A: Proofs of Theorems

Proof of Theorem 1. Let $\hat{\beta}_J$ be of the generic form $\hat{\beta}_J = k_{1n}\hat{\beta}_n + k_{2n}(1/m)\sum_{i=1}^{m}\hat{\beta}_i$ and note that

$$E(\hat{\beta}_i) = \beta + \frac{a_1}{\ell} + \frac{a_2}{\ell^2} + O\left(n^{-3}\right), \quad i = 1, \ldots, m.$$

Then it follows from the above and (1) that

$$\begin{aligned} E(\hat{\beta}_J) &= k_{1n}\left(\beta + \frac{a_1}{n}\right) + k_{2n}\left(\beta + \frac{a_1}{\ell}\right) + O\left(n^{-2}\right) \\ &= (k_{1n} + k_{2n})\beta + a_1\left(\frac{k_{1n}}{n} + \frac{k_{2n}}{\ell}\right) + O\left(n^{-2}\right). \end{aligned}$$

The result in the Theorem then holds if $k_{1n} + k_{2n} = 1$ and $(k_{1n}/n) + (k_{2n}/\ell) = 0$; these conditions are easily solved to give $k_{1n} = n/(n-\ell)$ and $k_{2n} = -\ell/(n-\ell)$. $\square$

Proof of Theorem 2. First note that

$$E(\hat{\beta}_{j,i}) = \beta + \frac{a_1}{\ell_j} + \frac{a_2}{\ell_j^2} + O\left(n^{-3}\right), \quad j = 1, 2.$$

It follows that

$$\begin{aligned} E(\hat{\beta}_{J,(m_1,m_2)}) &= w_n\left(\beta + \frac{a_1}{n} + \frac{a_2}{n^2}\right) + w_{1n}\left(\beta + \frac{a_1}{\ell_1} + \frac{a_2}{\ell_1^2}\right) + w_{2n}\left(\beta + \frac{a_1}{\ell_2} + \frac{a_2}{\ell_2^2}\right) + r_n \\ &= (w_n + w_{1n} + w_{2n})\beta + a_1\left(\frac{w_n}{n} + \frac{w_{1n}}{\ell_1} + \frac{w_{2n}}{\ell_2}\right) + a_2\left(\frac{w_n}{n^2} + \frac{w_{1n}}{\ell_1^2} + \frac{w_{2n}}{\ell_2^2}\right) + r_n \end{aligned}$$

where $r_n = O(n^{-3})$. In order that $E(\hat{\beta}_{J,(m_1,m_2)}) = \beta + O(n^{-3})$ it is therefore necessary for the following three conditions to be satisfied:

$$\text{(i) } w_n + w_{1n} + w_{2n} = 1; \quad \text{(ii) } \frac{w_n}{n} + \frac{w_{1n}}{\ell_1} + \frac{w_{2n}}{\ell_2} = 0; \quad \text{and (iii) } \frac{w_n}{n^2} + \frac{w_{1n}}{\ell_1^2} + \frac{w_{2n}}{\ell_2^2} = 0.$$

Solving these conditions yields the weights specified in the Theorem. $\square$

Proof of Theorem 3. Following the proof of Theorem 1 it is convenient to let the jackknife estimator have the generic form $\hat{\beta}_J^{\text{U}} = k_{1n}\hat{\beta}_n + k_{2n}(1/m)\sum_{i=1}^{m}\hat{\beta}_i$ so that the objective is to determine k_{1n} and k_{2n}. Under the conditions of the Theorem it follows that

$$\begin{aligned} E(\hat{\beta}_J^{\text{U}}) &= k_{1n}\left(\beta + \frac{a_1}{n}\right) + k_{2n}\left(\beta + \frac{a_1 m_1}{m\ell_1} + \frac{a_1 m_2}{m\ell_2}\right) + O\left(n^{-2}\right) \\ &= (k_{1n} + k_{2n})\beta + a_1\left(\frac{k_{1n}}{n} + \frac{k_{2n}m_1}{m\ell_1} + \frac{k_{2n}m_2}{m\ell_2}\right) + O\left(n^{-2}\right). \end{aligned}$$

To eliminate the first-order bias and to have $E(\hat{\beta}_J^{\text{U}}) = \beta + O(n^{-2})$ it is necessary that $k_{1n} + k_{2n} = 1$ and $(k_{1n}/n) + (k_{2n}m_1/m\ell_1) + (k_{2n}m_2/m\ell_2) = 0$; solving these equations yields the values for k_{1n} and k_{2n} in the Theorem. $\square$

Proof of Theorem 4. The proof of consistency when m is fixed follows straightforwardly by considering

$$(\hat{\beta}_{J,m} - \beta) = \frac{m}{m-1}(\hat{\beta}_n - \beta) - \frac{1}{m(m-1)}\sum_{j=1}^{m}(\hat{\beta}_j - \beta) \xrightarrow{p} 0 \qquad (23)$$

by virtue of the consistency of $\hat{\beta}_n$ and the $\hat{\beta}_j$ $(j = 1, \ldots, m)$. When $m^{-1} + mn^{-1} \to 0$ we first show that the second term in (23) is $o_p(1)$ by considering, for any $\epsilon > 0$,

$$\begin{aligned} \Pr\left(\frac{1}{m(m-1)}\left\|\sum_{j=1}^{m}(\hat{\beta}_j - \beta)\right\| > \epsilon\right) &\leq \Pr\left(\sum_{j=1}^{m}\left\|\hat{\beta}_j - \beta\right\| > \epsilon m(m-1)\right) \\ &\leq \frac{E\left(\sum_{j=1}^{m}\left\|\hat{\beta}_j - \beta\right\|\right)^2}{\epsilon^2 m^2 (m-1)^2} \end{aligned} \tag{24}$$

by use of Chebyshev's inequality. But $E\left(\sum_{j=1}^{m}\left\|\hat{\beta}_j - \beta\right\|\right)^2 \leq m\sum_{j=1}^{m} E\left\|\hat{\beta}_j - \beta\right\|^2$ by the Cauchy-Schwarz inequality, while

$$E\left\|\hat{\beta}_j - \beta\right\|^2 = \sum_{k=1}^{p+1} E(\hat{\beta}_{jk} - \beta_k)^2 = O(\ell^{-1});$$

see, for example, equation (A.2) of Bao and Ullah (2007) for the MSE of the OLS estimator in the general case and Bao (2007) for the AR(1) model. Hence the right-hand side of (24) is bounded by an $O(\ell^{-1}m^{-2})$ quantity which is also $o(1)$ under the assumption on m. It follows that

$$(\hat{\beta}_{J,m} - \beta) = \frac{m}{m-1}(\hat{\beta}_n - \beta) + o_p(1) \xrightarrow{p} 0$$

by the consistency of $\hat{\beta}_n$.

The proof of asumptotic normality considers the joint convergence of $\hat{\beta}_n$ and the sub-sample estimators $\hat{\beta}_j$ $(j = 1, \ldots, m)$ and utilises the convergence result in Lemma 1 of Appendix B. First, note that the normalised jackknife estimator can be written

$$\sqrt{n}(\hat{\beta}_{J,m} - \beta) = a_m\sqrt{n}(\hat{\beta}_n - \beta) - b_m\sum_{j=1}^{m}\sqrt{\ell}(\hat{\beta}_j - \beta), \tag{25}$$

where $a_m = m/(m-1)$ and $b_m = 1/(\sqrt{m}(m-1))$. Clearly

$$\sqrt{n}(\hat{\beta}_n - \beta) = \left(\frac{1}{n}\sum_{t=1}^{n}x_t x_t'\right)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}x_t\epsilon_t,$$

$$\sqrt{\ell}(\hat{\beta}_j - \beta) = \left(\frac{1}{\ell}\sum_{t\in\tau_j}x_t x_t'\right)^{-1}\frac{1}{\sqrt{\ell}}\sum_{t\in\tau_j}x_t\epsilon_t, \quad (j = 1, \ldots, m).$$

It is also convenient to define the $(p+1) \times (m+1)(p+1)$ matrix B_n and the $(m+1)(p+1) \times 1$ vector b_n as follows:

$$B_n = \left[a_m\left(\frac{1}{n}\sum_{t=1}^{n}x_t x_t'\right)^{-1}, -b_m\left(\frac{1}{\ell}\sum_{t\in\tau_1}x_t x_t'\right)^{-1}, \ldots, -b_m\left(\frac{1}{\ell}\sum_{t\in\tau_m}x_t x_t'\right)^{-1}\right],$$

$$b_n = \left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t' \epsilon_t, \frac{1}{\sqrt{\ell}} \sum_{t \in \tau_1} x_t' \epsilon_t, \ldots, \frac{1}{\sqrt{\ell}} \sum_{t \in \tau_m} x_t' \epsilon_t \right]';$$

this enables the normalised jackknife estimator to be written in the form

$$\sqrt{n}(\hat{\beta}_{J,m} - \beta) = B_n b_n.$$

When m is fixed as $n \to \infty$ the results in Lemma 1 are relevant. In the notation of Lemma 1 the sub-blocks of B_n are the inverses of the appropriate diagonal blocks of $D_n^{-1} \sum_{t=1}^{n} z_t z_t' D_n^{-1}$ while $b_n = D_n^{-1} \sum_{t=1}^{n} z_t \epsilon_t$, and so immediately it follows that

$$B_n \xrightarrow{p} \bar{B}_m = \left[a_m Q^{-1}, -b_m Q^{-1}, \ldots, -b_m Q^{-1} \right], \quad b_n \xrightarrow{d} N(0, \sigma^2 \bar{Q}_m).$$

Hence

$$\sqrt{n}(\hat{\beta}_{J,m} - \beta) \xrightarrow{d} N(0, \sigma^2 \bar{B}_m \bar{Q}_m \bar{B}_m').$$

It can further be shown that

$$\bar{B}_m \bar{Q}_m \bar{B}_m' = a_m(a_m - b_n \sqrt{m}) Q^{-1} - m b_m \left(\frac{a_m}{\sqrt{m}} - b_m \right) Q^{-1} = Q^{-1}$$

upon evaluation of the coefficients, which proves the result for fixed m. When $m^{-1} + mn^{-1} \to 0$ we demonstrate in (25) that

$$b_m \sum_{j=1}^{m} \sqrt{\ell}(\hat{\beta}_j - \beta) = \frac{1}{\sqrt{m}(m-1)} \sum_{j=1}^{m} \sqrt{\ell}(\hat{\beta}_j - \beta) = o_p(1)$$

by showing that, for all $\epsilon > 0$,

$$\begin{aligned} \Pr\left(\frac{\sqrt{\ell}}{\sqrt{m}(m-1)} \left\| \sum_{j=1}^{m} (\hat{\beta}_j - \beta) \right\| > \epsilon \right) &\leq \Pr\left(\sum_{j=1}^{m} \left\| \hat{\beta}_j - \beta \right\| > \frac{\epsilon \sqrt{m}(m-1)}{\sqrt{\ell}} \right) \\ &\leq \frac{\ell E \left(\sum_{j=1}^{m} \left\| \hat{\beta}_j - \beta \right\| \right)^2}{\epsilon^2 m(m-1)^2} \leq \frac{\ell m \sum_{j=1}^{m} E \left\| \hat{\beta}_j - \beta \right\|^2}{\epsilon^2 m(m-1)^2}, \end{aligned} \quad (26)$$

where use is made of the Chebyshev and Cauchy-Schwarz inequalities, respectively. But, as shown in the consistency proof, the last expectation in (26) is $O(\ell^{-1})$ and it follows that the right-hand side of (26) is $O(m^{-1})$ and hence $o(1)$, implying that the second term on the right-hand side of (25) is $o_p(1)$ as claimed. Therefore

$$\sqrt{n}(\hat{\beta}_{J,m} - \beta) = \frac{m}{m-1} \sqrt{n}(\hat{\beta}_n - \beta) + o_p(1) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$$

as $n \to \infty$, as required. □

Proof of Theorem 5. The consistency of $\hat{\sigma}_J^2$ follows in the usual way by first writing

$$e_{t,J} = y_t - x_t' \hat{\beta}_{J,m} = \epsilon_t - x_t'(\hat{\beta}_{J,m} - \beta).$$

Defining $n^* = n - p - 1$ we then obtain

$$\begin{aligned}\hat{\sigma}_J^2 &= \frac{1}{n^*}\sum_{t=1}^n \epsilon_t^2 - \frac{2}{n^*}\sum_{t=1}^n \epsilon_t x_t'(\hat{\beta}_{J,m} - \beta) + (\hat{\beta}_{J,m} - \beta)'\frac{1}{n^*}\sum_{t=1}^n x_t x_t'(\hat{\beta}_{J,m} - \beta) \\ &= \frac{1}{n^*}\sum_{t=1}^n \epsilon_t^2 + o_p(1) \xrightarrow{p} \sigma^2\end{aligned}$$

as $n \to \infty$, which follows from the consistency of $\hat{\beta}_{J,m}$ (for fixed m and m tending to ∞) and the properties of the moment matrices. The consistency of $\hat{V}_J$ is straightforward while for $\hat{V}_{J,m}$, for fixed m,

$$\begin{aligned}\hat{V}_{J,m} &= \frac{m(m-2)}{(m-1)^2}\hat{\sigma}_J^2\left(\frac{1}{n}\sum_{t=1}^n x_t x_t'\right)^{-1} + \frac{1}{m(m-1)^2}\hat{\sigma}_J^2\sum_{j=1}^m\left(\frac{1}{\ell}\sum_{t\in\tau_j} x_t x_t'\right)^{-1} \\ &\xrightarrow{p} \frac{m(m-2)}{(m-1)^2}\sigma^2 Q^{-1} + \frac{1}{m(m-1)^2}\sigma^2 m Q^{-1} = \sigma^2 Q^{-1}\end{aligned}$$

because $m(m-2)+1 = (m-1)^2$. When $m^{-1} + mn^{-1} \to 0$ as $n \to \infty$ we have

$$\hat{V}_{J,m} = \frac{m(m-2)}{(m-1)^2}\hat{\sigma}_J^2\left(\frac{1}{n}\sum_{t=1}^n x_t x_t'\right)^{-1} + o_p(1) \xrightarrow{p} \sigma^2 Q^{-1},$$

which completes the proof. □

Appendix B: Supplementary results

Lemma 1. *Let y_t satisfy Assumption 1, let x_t be defined as in (10) and let*

$$z_t = (x_t', x_t'1(t\in\tau_1), \ldots, x_t'1(t\in\tau_m))', \quad D_n = \begin{pmatrix} n^{1/2}I_{p+1} & 0 \\ 0 & \ell^{1/2}I_{m(p+1)}\end{pmatrix},$$

where z_t is of dimension $(m+1)(p+1)\times 1$ and $1(x)$ denotes the indicator function taking the value 1 if the event x is true and 0 otherwise. Furthermore, let

$$\bar{Q}_m = \begin{pmatrix} Q & m^{-1/2}Q & m^{-1/2}Q & \ldots & m^{-1/2}Q \\ m^{-1/2}Q & Q & 0 & \ldots & 0 \\ m^{-1/2}Q & 0 & Q & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m^{-1/2}Q & 0 & 0 & \ldots & Q\end{pmatrix},$$

where $Q = E(x_t x_t')$ is defined in (11). Then, as $n \to \infty$,

$$D_n^{-1}\sum_{t=1}^n z_t z_t' D_n^{-1} \xrightarrow{p} \bar{Q}_m \quad \textit{and} \quad D_n^{-1}\sum_{t=1}^n z_t\epsilon_t \xrightarrow{d} N(0, \sigma^2\bar{Q}_m).$$

Proof. From the definition of z_t we have

$$D_n^{-1}\sum_{t=1}^{n} z_t z_t' D_n^{-1} = \begin{pmatrix} \frac{1}{n}\sum_{t=1}^{n} x_t x_t' & \frac{1}{\sqrt{n\ell}}\sum_{t\in\tau_1} x_t x_t' & \frac{1}{\sqrt{n\ell}}\sum_{t\in\tau_2} x_t x_t' & \dots & \frac{1}{\sqrt{n\ell}}\sum_{t\in\tau_m} x_t x_t' \\ \frac{1}{\sqrt{n\ell}}\sum_{t\in\tau_1} x_t x_t' & \frac{1}{\ell}\sum_{t\in\tau_1} x_t x_t' & 0 & \dots & 0 \\ \frac{1}{\sqrt{n\ell}}\sum_{t\in\tau_2} x_t x_t' & 0 & \frac{1}{\ell}\sum_{t\in\tau_2} x_t x_t' & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{n\ell}}\sum_{t\in\tau_m} x_t x_t' & 0 & 0 & \dots & \frac{1}{\ell}\sum_{t\in\tau_m} x_t x_t' \end{pmatrix},$$

the $(p+1)\times(p+1)$ blocks of zeros involving cross-products of the form

$$\sum_{t=1}^{n} x_t 1(t\in\tau_j) x_t' 1(t\in\tau_k) = 0 \quad (j\neq k)$$

which are zero due to the non-overlapping sub-samples. When m is fixed the diagonal blocks each converge in probability to Q while the non-zero off-diagonal blocks converge in probability to $m^{-1/2}Q$ by noting that $\sqrt{nl} = \ell\sqrt{m}$, thereby establishing the convergence of $D_n^{-1}\sum_{t=1}^{n} z_t z_t' D_n^{-1}$. Next, note that $z_t\epsilon_t$ is a vector martingale difference sequence (mds). The proof will proceed by verifying that the conditions of Proposition 7.9 of Hamilton (1994) are satisfied, which yields the stated result. The requirements are that: (i) $E(z_t z_t' \epsilon_t^2) = \Omega_t$, a positive definite matrix with $D_n^{-1}\sum_{t=1}^{n}\Omega_t D_n^{-1} \to \Omega$, a positive definite matrix, as $n\to\infty$; (ii) $E(z_{it}z_{jt}z_{rt}z_{st}\epsilon_t^4) < \infty$ for all t,i,j,r,s; and (iii) $D_n^{-1}\sum_{t=1}^{n} z_t z_t' \epsilon_t^2 D_n^{-1} \xrightarrow{p} \Omega$ as $n\to\infty$. To demonstrate that (i) is satisfied, note that $E(z_t z_t'\epsilon_t^2) = E(\epsilon_t^2)E(z_t z_t') = \sigma^2 Q_t$, where

$$Q_t = E(z_t z_t') = \begin{pmatrix} Q & Q1(t\in\tau_1) & Q1(t\in\tau_2) & \dots & Q1(t\in\tau_m) \\ Q1(t\in\tau_1) & Q1(t\in\tau_1) & 0 & \dots & 0 \\ Q1(t\in\tau_2) & 0 & Q1(t\in\tau_2) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Q1(t\in\tau_m) & 0 & 0 & \dots & Q1(t\in\tau_m) \end{pmatrix},$$

a positive definite matrix. Noting that $\sum_{t=1}^{n} Q1(t\in\tau_j) = \ell Q$ $(j=1,\dots,m)$ we find that

$$D_n^{-1}\sum_{t=1}^{n} Q_t D_n^{-1} = \begin{pmatrix} Q & (\ell/n)^{1/2}Q & (\ell/n)^{1/2}Q & \dots & (\ell/n)^{1/2}Q \\ (\ell/n)^{1/2}Q & Q & 0 & \dots & 0 \\ (\ell/n)^{1/2}Q & 0 & Q & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (\ell/n)^{1/2}Q & 0 & 0 & \dots & Q \end{pmatrix} \to \bar{Q}_m.$$

Hence (i) holds with $\Omega_t = \sigma^2 Q_t$ and $\Omega = \sigma^2\bar{Q}_m$. In order to verify property (ii) first note that $E(z_{it}z_{jt}z_{rt}z_{st}\epsilon_t^4) = E(\epsilon_t^4)E(z_{it}z_{jt}z_{rt}z_{st})$. Now $E(\epsilon_t^4) < \infty$ under Assumption 1 and $E(z_{it}z_{jt}z_{rt}z_{st}) < \infty$ for all t,i,j,r,s by Proposition 10.1 of Hamilton (1994); hence (ii) is satisfied. As for (iii), the matrix of interest is given by

$$D_n^{-1}\sum_{t=1}^{n} z_t z_t' \epsilon_t^2 D_n^{-1} =$$

$$\begin{pmatrix} \frac{1}{n}\sum_{t=1}^{n} x_t x_t' \epsilon_t^2 & \frac{1}{\sqrt{n\ell}}\sum_{t\in\tau_1} x_t x_t' \epsilon_t^2 & \ldots & \frac{1}{\sqrt{n\ell}}\sum_{t\in\tau_m} x_t x_t' \epsilon_t^2 \\ \frac{1}{\sqrt{n\ell}}\sum_{t\in\tau_1} x_t x_t' \epsilon_t^2 & \frac{1}{\ell}\sum_{t\in\tau_1} x_t x_t' \epsilon_t^2 & \ldots & 0 \\ \frac{1}{\sqrt{n\ell}}\sum_{t\in\tau_2} x_t x_t' \epsilon_t^2 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{n\ell}}\sum_{t\in\tau_m} x_t x_t' \epsilon_t^2 & 0 & \ldots & \frac{1}{\ell}\sum_{t\in\tau_1} x_t x_t' \epsilon_t^2 \end{pmatrix}.$$

Note that

$$\frac{1}{n}\sum_{t=1}^{n} x_t x_t' \epsilon_t^2 = \frac{1}{n}\sum_{t=1}^{n} (\epsilon_t^2 - \sigma^2) x_t x_t' \epsilon_t^2 + \frac{1}{n}\sigma^2 \sum_{t=1}^{n} x_t x_t'.$$

But the elements of $(\epsilon_t^2 - \sigma^2)x_t x_t'$ are each an mds and so $n^{-1}\sum_{t=1}^{n}(\epsilon_t^2 - \sigma^2)x_t x_t' \xrightarrow{p} 0$ as $n \to \infty$ so that $n^{-1}\sum_{t=1}^{n} x_t x_t' \epsilon_t^2 \xrightarrow{p} \sigma^2 Q$. A similar argument applied to the sub-samples results in $\ell^{-1}\sum_{t=1}^{n} x_t x_t' \epsilon_t^2 \xrightarrow{p} \sigma^2 Q$ while

$$\frac{1}{\sqrt{n\ell}}\sum_{t\in\tau_j} x_t x_t' \epsilon_t^2 = \frac{1}{m^{1/2}}\frac{1}{\ell}\sum_{t\in\tau_j} x_t x_t' \epsilon_t^2 \xrightarrow{p} m^{-1/2}\sigma^2 Q.$$

Hence (iii) holds with Ω defined as in part (a). □

Lemma 2. *Let the $n \times 1$ random vector y satisfy $y = X\beta + \epsilon$ where X is an $n \times k$ matrix of non-random regressors, β is an unknown $k \times 1$ parameter vector and ϵ is an $n \times 1$ vector of random disturbances satisfying $E(\epsilon) = 0$ and $E(\epsilon\epsilon') = \sigma^2 I_n$, I_n being the $n \times n$ identity matrix. Furthermore partition y and X into m non-overlapping sub-samples as follows:*

$$y = [y_1', y_2', \ldots, y_m']', \quad X = [X_1', X_2', \ldots, X_m']',$$

where $y_1, \ldots, y_m$ are $\ell \times 1$ and $X_1, \ldots, X_m$ are $\ell \times k$, where $n = m \times \ell$. Let $\hat{\beta} = (X'X)^{-1}X'y$, $\hat{\beta}_j = (X_j'X_j)^{-1}X_j'y_j$ $(j = 1, \ldots, m)$ and $\hat{\beta}_{J,m} = (m/(m-1))\hat{\beta} - (1/(m(m-1)))\sum_{j=1}^{m}\hat{\beta}_j$. Then the variance matrix of $\hat{\beta}_{J,m}$ is given by

$$var(\hat{\beta}_{J,m}) = \frac{m(m-2)}{(m-1)^2}\sigma^2(X'X)^{-1} + \frac{1}{m^2(m-1)^2}\sigma^2\sum_{j=1}^{m}(X_j'X_j)^{-1}.$$

Proof. Let $a_0 = m/(m-1)$ and $a_1 = 1/[m(m-1)]$ so that $\hat{\beta}_{J,m} = a_0\hat{\beta}_n - a_1\sum_{j=1}^{m}\hat{\beta}_j$. Under the assumptions of the model $E(\hat{\beta}_n) = E(\hat{\beta}_j) = \beta$ and so $E(\hat{\beta}_{J,m}) = \beta$ also, implying that $var(\hat{\beta}_{J,m}) = E(\hat{\beta}_{J,m} - \beta)(\hat{\beta}_{J,m} - \beta)'$. Then

$$\begin{aligned} var(\hat{\beta}_{J,m}) &= a_0^2 var(\hat{\beta}_n) - a_0 a_1\left[cov\left(\hat{\beta}_n, \sum_{j=1}^{m}\hat{\beta}_j\right) + cov\left(\sum_{j=1}^{m}\hat{\beta}_j, \hat{\beta}_n\right)\right] + a_1^2 var\left(\sum_{j=1}^{m}\hat{\beta}_j\right) \\ &= a_0^2 var(\hat{\beta}_n) - a_0 a_1\sum_{j=1}^{m}\left[cov(\hat{\beta}_n, \hat{\beta}_j) + cov(\hat{\beta}_j, \hat{\beta}_n)\right] \\ &\quad + a_1^2\left[\sum_{j=1}^{m} var(\hat{\beta}_j) + \sum_{j=1}^{m-1}\sum_{k=j+1}^{m}\left(cov(\hat{\beta}_j, \hat{\beta}_k) + cov(\hat{\beta}_k, \hat{\beta}_j)\right)\right], \end{aligned}$$

where, for example, $cov(\hat{\beta}_n, \hat{\beta}_j) = E(\hat{\beta}_n - \beta)(\hat{\beta}_j - \beta)'$ and other covariances are similarly defined. The variance matrices are $var(\hat{\beta}_n) = \sigma^2(X'X)^{-1}$ and $var(\hat{\beta}_j) = \sigma^2(X_j'X_j)^{-1}$ so it remains to determine the covariances. From the usual OLS formulae

$$cov(\hat{\beta}_n, \hat{\beta}_j) = E\left[(X'X)^{-1}X'\epsilon\epsilon_j'X_j(X_j'X_j)^{-1}\right] = (X'X)^{-1}X'E(\epsilon\epsilon_j')X_j(X_j'X_j)^{-1},$$

where ϵ_j denotes the appropriate subvector of $\epsilon = (\epsilon_1', \ldots, \epsilon_m')'$. Note that

$$E(\epsilon\epsilon_j') = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \sigma^2 I_\ell \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sigma^2 K_j$$

where K_j is the null matrix with $\sigma^2 I_\ell$ in the j'th block. It then follows that

$$cov(\hat{\beta}_n, \hat{\beta}_j) = (X'X)^{-1}X' \cdot \sigma^2 K_j \cdot X_j(X_j'X_j)^{-1} = \sigma^2(X'X)^{-1}$$

because $X'K_j = X_j$. Also, noting that $E(\epsilon_j\epsilon_k') = 0$ for $j \neq k$, it follows that

$$cov(\hat{\beta}_j, \hat{\beta}_k) = E\left[(X_j'X_j)^{-1}X_j'\epsilon_j\epsilon_k'X_k(X_k'X_k)^{-1}\right] = 0.$$

Combining these results yields

$$var(\hat{\beta}_{J,m}) = a_0^2\sigma^2(X'X)^{-1} - 2a_0a_1m\sigma^2(X'X)^{-1} + a_1^2\sigma^2\sum_{j=1}^{m}(X_j'X_j)^{-1},$$

from which the expression in the Lemma follows by substituting for a_0 and a_1 and simplifying. □

Lemma 3. *Let the $n \times 1$ random vector y satisfy $y = X\beta + \epsilon$ where X is an $n \times k$ matrix of non-random regressors, β is an unknown $k \times 1$ parameter vector and ϵ is an $n \times 1$ vector of random disturbances satisfying $E(\epsilon) = 0$ and $E(\epsilon\epsilon') = \sigma^2 I_n$, I_n being the $n \times n$ identity matrix. Furthermore let y and X be partitioned into two sets of non-overlapping sub-samples as follows:*

$$y = [y_{11}', y_{12}', \ldots, y_{1m_1}']' = [y_{21}', y_{22}', \ldots, y_{2m_2}']',$$

$$X = [X_{11}', X_{12}', \ldots, X_{1m_1}']' = [X_{21}', X_{22}', \ldots, X_{2m_2}']',$$

where the $y_{j\cdot}$ are $\ell_j \times 1$ and the $X_{j\cdot}$ are $\ell_j \times k$ $(j = 1, 2)$, where $n = m_1 \times \ell_1 = m_2 \times \ell_2$. Let $\hat{\beta} = (X'X)^{-1}X'y$, $\hat{\beta}_{j\cdot} = (X_{j\cdot}'X_{j\cdot})^{-1}X_{j\cdot}'y_{j\cdot}$ $(j = 1, 2)$ and

$$\hat{\beta}_{J,(m_1,m_2)} = w\hat{\beta} + w_1\frac{1}{m_1}\sum_{j=1}^{m_1}\hat{\beta}_{1j} + w_2\frac{1}{m_2}\sum_{j=1}^{m_2}\hat{\beta}_{2j},$$

where

$$w = \frac{m_1m_2}{(m_1-1)(m_2-1)}, \quad w_1 = \frac{m_2}{(m_1-1)(m_1-m_2)}, \quad w_2 = \frac{m_1}{(m_2-1)(m_2-m_1)}.$$

Then the variance matrix of $\hat{\beta}_{J,(m_1,m_2)}$ *is given by*

$$\begin{aligned} var(\hat{\beta}_{J,(m_1,m_2)}) &= \sigma^2(X'X)^{-1}\left[w^2+2w(w_1+w_2)\right] \\ &+\frac{w_1^2}{m_1^2}\sigma^2\sum_{j=1}^{m_1}(X_{1j}'X_{1j})^{-1}+\frac{w_2^2}{m_2^2}\sigma^2\sum_{j=1}^{m_2}(X_{2j}'X_{2j})^{-1} \\ &+\frac{w_1w_2}{m_1m_2}\sigma^2\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\left[(X_{1i}'X_{1i})^{-1}X_{1i}'K_{1i}K_{2j}'X_{2j}(X_{2j}'X_{2j})^{-1}\right. \\ &\left.+\,(X_{2j}'X_{2j})^{-1}X_{2j}'K_{2j}K_{1i}'X_{1i}(X_{1i}'X_{1i})^{-1}\right], \end{aligned}$$

where K_{1i} *is a null matrix of dimension* $\ell_1 \times n$ *with* I_{ℓ_1} *in the i'th block and* K_{2j} *is a null matrix of dimension* $\ell_2 \times n$ *with* I_{ℓ_2} *in the j'th block.*

Proof. First note that, because $w+w_1+w_2=1$, we may write

$$\hat{\beta}_{J,(m_1,m_2)}-\beta = w(\hat{\beta}-\beta)+w_1\frac{1}{m_1}\sum_{j=1}^{m_1}(\hat{\beta}_{1j}-\beta)+w_2\frac{1}{m_2}\sum_{j=1}^{m_2}(\hat{\beta}_{2j}-\beta),$$

and hence

$$\begin{aligned} var(\hat{\beta}_{J,(m_1,m_2)}) &= E(\hat{\beta}_{J,(m_1,m_2)}-\beta)(\hat{\beta}_{J,(m_1,m_2)}-\beta)' \\ &= w^2E(\hat{\beta}-\beta)(\hat{\beta}-\beta)' \\ &+\frac{w_1^2}{m_1^2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_1}E(\hat{\beta}_{1i}-\beta)(\hat{\beta}_{1j}-\beta)'+\frac{w_2^2}{m_2^2}\sum_{i=1}^{m_2}\sum_{j=1}^{m_2}E(\hat{\beta}_{2i}-\beta)(\hat{\beta}_{2j}-\beta)' \\ &+\frac{ww_1}{m_1}\sum_{i=1}^{m_1}\left[E(\hat{\beta}-\beta)(\hat{\beta}_{1i}-\beta)'+E(\hat{\beta}_{1i}-\beta)(\hat{\beta}-\beta)'\right] \\ &+\frac{ww_2}{m_2}\sum_{i=1}^{m_2}\left[E(\hat{\beta}-\beta)(\hat{\beta}_{2i}-\beta)'+E(\hat{\beta}_{2i}-\beta)(\hat{\beta}-\beta)'\right] \\ &+\frac{w_1w_2}{m_1m_2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\left[E(\hat{\beta}_{1i}-\beta)(\hat{\beta}_{2j}-\beta)'+E(\hat{\beta}_{2j}-\beta)(\hat{\beta}_{1i}-\beta)'\right]. \end{aligned}$$

Using results from regression theory and the proof of Lemma 2 we have that $E(\hat{\beta}-\beta)(\hat{\beta}-\beta)' = \sigma^2(X'X)^{-1}$, $E(\hat{\beta}-\beta)(\hat{\beta}_{1i}-\beta)' = E(\hat{\beta}-\beta)(\hat{\beta}_{1j}-\beta)' = \sigma^2(X'X)^{-1}$ $(i=1,\ldots,m_1; j=1,\ldots,m_2)$,

$$E(\hat{\beta}_{1i}-\beta)(\hat{\beta}_{1j}-\beta)' = \begin{cases} \sigma^2(X_{1i}'X_{1i})^{-1}, & i=j, \\ 0, & i\neq j, \end{cases}$$

$$E(\hat{\beta}_{2i}-\beta)(\hat{\beta}_{2j}-\beta)' = \begin{cases} \sigma^2(X_{2i}'X_{2i})^{-1}, & i=j, \\ 0, & i\neq j. \end{cases}$$

For the remaining cross-product terms we have

$$E(\hat{\beta}_{1i}-\beta)(\hat{\beta}_{2j}-\beta)' = (X_{1i}'X_{1i})^{-1}X_{1i}'E(\epsilon_{1i}\epsilon_{2j}')X_{2j}(X_{2j}'X_{2j})^{-1}.$$

It is possible to write $\epsilon_{1i} = K_{1i}\epsilon$ and $\epsilon_{2j} = K_{2j}\epsilon$ where the selection matrics K_{1i} and K_{2j} are defined in the Lemma. Hence $E(\epsilon_{1i}\epsilon_{2j}') = E(K_{1i}\epsilon\epsilon'K_{2j}' = \sigma^2K_{1i}K_{2j}'$. Combining all the relevant terms yields the required expression. □

References

Angrist, J.D., Imbens, G.W., Krueger, A.B., 1999. Jackknife instrumental variables estimation. Journal of Applied Econometrics 14, 57–67.

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

Bao, Y., 2007. The approximate moments of the least squares estimator for the stationary autoregressive model under a general error distribution. Econometric Theory 23, 1013–1021.

Bao, Y., Ullah, A., 2007. The second-order bias and mean squared error of estimators in time-series models. Journal of Econometrics 140, 650–669.

Berk, K.N., 1974. Consistent autoregressive spectral estimates. Annals of Statistics 2, 489–502.

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.

Chambers, M.J., Kyriacou, M., 2011. Jackknife bias reduction in autoregressive models with a unit root. Submitted.

Davidson, R., MacKinnon, J.G., 2006. The case against JIVE. Journal of Applied Econometrics 21, 827–833.

Dhaene, G., Jochmans, K., Thuysbaert, B., 2006. Jackknife bias reduction for nonlinear dynamic panel data models with fixed effects. Working paper, KU Leuven.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Annals of Statistics 7, 1–26.

Efron, B., 1982. The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia.

Gonçalves, S., Kilian, L., 2004. Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. Journal of Econometrics 123, 89–120.

Gospodinov, N., Otsu, T., 2009. Local GMM estimation of time series models with conditional moment restrictions. Concordia University Working Paper 08-010.

Hahn, J., Kuersteiner, G., Newey, W., 2003. Higher order properties of bootstrap and jackknife bias corrected maximum likelihood estimators. Working paper, MIT.

Hahn, J., Moon, H.R., 2006. Reducing bias of MLE in a dynamic panel model. Econometric Theory 22, 499–512.

Hahn, J., Newey, W., 2004. Jackknife and analytical bias reduction for nonlinear panel models. Econometrica 72, 1295–1319.

Hall, P., Horowitz, J.L., Jing, B.-Y., 1995. On blocking rules for the bootstrap with dependent data. Biometrika 82, 561–574.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton.

Kendall, M.G., 1954. Notes on bias in the estimation of autocorrelation. Biometrika 41, 403–404.

Kitamura, Y., Tripathi, G., Ahn, H., 2004. Empirical l;ikelihood-based inference in conditional momenmt restriction models. Econometrica 72, 1667–1714.

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

Miller, R.G., 1964. A trustworthy jackknife. Annals of Mathematical Statistics 35, 1594–1605.

Ng, S., Perron, P., 2001. Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519–1554.

Patterson, K., 2000. Finite sample bias of the least squares estimator in an AR(p) model: estimation, inference, simulation and examples. Applied Economics 32, 1993–2005.

Phillips, P.C.B., Yu, J., 2005. Jackknifing bond option prices. Review of Financial Studies 18, 707–742.

Quenouille, M.H., 1956. Notes on bias in estimation. Biometrika 43, 353–360.

Sawa, T., 1978. The exact moments of the least squares estimator for the autoregressive model. Journal of Econometrics 8, 159–172.

Schucany, W.R., Gray, H.L., Owen, D.B., 1971. On bias reduction in estimation. Journal of the American Statistical Association 66, 524–533.

Shaman, P., Stine, R.A., 1988. The bias of autoregressive coefficient estimators. Journal of the American Statistical Association 83, 842–848.

Shao, J., Tu, D., 1995. The Jackknife and Bootstrap. Springer-Verlag, New York.

Smith, R.J., 2007. Efficient information theoretic inference for conditional moment restrictions. Journal of Econometrics 138, 430–460.

Tukey, J.W., 1958. Bias and confidence in not-quite large samples. Annals of Mathematical Statistics 29, 614.

Wu, C.F.J., 1986. Jackknife, bootstrap and other resampling methods in regression analysis. Annals of Statistics 14, 1261–1295.

Table 1
Bias of OLS and non-overlapping jackknife estimators

ϕ	n	$\hat{\phi}_n$	$\hat{\phi}_{J,2}$	$\hat{\phi}_{J,3}$	$\hat{\phi}_{J,4}$	$\hat{\phi}_{J,6}$	$\hat{\phi}_{J,8}$
0.10	24	−0.0568	**−0.0031**	−0.0037	−0.0044	−0.0059	−0.0070
	48	−0.0281	**−0.0011**	−0.0012	−0.0015	−0.0018	−0.0023
	96	−0.0139	**−0.0003**	−0.0003	−0.0005	−0.0005	−0.0007
	192	−0.0072	**−0.0004**	−0.0004	−0.0004	−0.0004	−0.0005
0.30	24	−0.0820	**−0.0063**	−0.0090	−0.0117	−0.0168	−0.0216
	48	−0.0406	**−0.0016**	−0.0022	−0.0029	−0.0041	−0.0055
	96	−0.0202	**−0.0004**	−0.0005	−0.0008	−0.0010	−0.0014
	192	−0.0103	**−0.0005**	−0.0005	−0.0005	−0.0006	−0.0007
0.50	24	−0.1091	**−0.0095**	−0.0149	−0.0200	−0.0295	−0.0372
	48	−0.0537	**−0.0017**	−0.0027	−0.0038	−0.0061	−0.0087
	96	−0.0267	**−0.0004**	−0.0006	−0.0008	−0.0013	−0.0018
	192	−0.0136	**−0.0005**	−0.0005	−0.0005	−0.0006	−0.0007
0.70	24	−0.1409	**−0.0152**	−0.0247	−0.0333	−0.0475	−0.0588
	48	−0.0686	**−0.0013**	−0.0030	−0.0050	−0.0091	−0.0132
	96	−0.0338	**−0.0001**	−0.0001	−0.0004	−0.0011	−0.0020
	192	−0.0169	**−0.0003**	−0.0003	−0.0003	−0.0003	−0.0003
0.90	24	−0.1856	**−0.0382**	−0.0515	−0.0627	−0.0807	−0.0954
	48	−0.0913	**−0.0054**	−0.0096	−0.0133	−0.0200	−0.0260
	96	−0.0435	**−0.0001**	−0.0008	−0.0014	−0.0019	−0.0038
	192	−0.0209	**−0.0007**	−0.0009	−0.0010	−0.0011	−0.0015
0.95	24	−0.1976	**−0.0460**	−0.0594	−0.0711	−0.0900	−0.1052
	48	−0.1013	**−0.0123**	−0.0167	−0.0207	−0.0278	−0.0340
	96	−0.0487	**−0.0001**	−0.0013	−0.0026	−0.0049	−0.0070
	192	−0.0229	**−0.0002**	−0.0007	−0.0012	−0.0012	−0.0013
0.99	24	−0.2002	**−0.0429**	−0.0574	−0.0699	−0.0898	−0.1055
	48	−0.1069	**−0.0145**	−0.0193	−0.0236	−0.0313	−0.0378
	96	−0.0554	**−0.0052**	−0.0063	−0.0077	−0.0102	−0.0125
	192	−0.0275	**−0.0013**	−0.0016	−0.0020	−0.0026	−0.0033

Entries in bold denote the minimum absolute bias in each row.

Table 2
Bias of OLS and various jackknife estimators

ϕ	n	$\hat{\phi}_n$	$\hat{\phi}_{J,2}$	$\hat{\phi}_{J,2}^{\text{MB}}$	$\hat{\phi}_{J,2}^{\text{MB2}}$	$\hat{\phi}_{J,(2,3)}$
0.10	24	−0.0568	−0.0031	−0.0072	−0.0047	**−0.0019**
	48	−0.0281	−0.0011	−0.0024	−0.0017	**−0.0008**
	96	−0.0139	−0.0003	−0.0007	−0.0005	**−0.0001**
	192	−0.0072	**−0.0004**	−0.0007	−0.0006	−0.0005
0.30	24	−0.0820	−0.0063	−0.0120	−0.0087	**−0.0010**
	48	−0.0406	−0.0016	−0.0035	−0.0024	**−0.0006**
	96	−0.0202	−0.0004	−0.0010	−0.0007	**−0.0001**
	192	−0.0103	**−0.0005**	−0.0007	−0.0006	−0.0006
0.50	24	−0.1091	−0.0095	−0.0173	−0.0129	**0.0012**
	48	−0.0537	−0.0017	−0.0044	−0.0028	**0.0002**
	96	−0.0267	−0.0004	−0.0012	−0.0008	**0.0000**
	192	−0.0136	**−0.0005**	−0.0007	−0.0006	−0.0006
0.70	24	−0.1409	−0.0152	−0.0252	−0.0199	**0.0038**
	48	−0.0686	**−0.0013**	−0.0054	−0.0029	0.0021
	96	−0.0338	**−0.0001**	−0.0012	−0.0005	0.0004
	192	−0.0169	**−0.0003**	−0.0006	−0.0004	−0.0005
0.90	24	−0.1856	−0.0382	−0.0436	−0.0411	**−0.0117**
	48	−0.0913	−0.0054	−0.0106	−0.0078	**0.0030**
	96	−0.0435	**−0.0001**	−0.0014	0.0002	0.0027
	192	−0.0209	−0.0007	**−0.0001**	0.0004	0.0002
0.95	24	−0.1976	−0.0460	−0.0474	−0.0468	**−0.0192**
	48	−0.1013	−0.0123	−0.0151	−0.0137	**−0.0036**
	96	−0.0487	**−0.0001**	−0.0029	−0.0014	0.0025
	192	−0.0229	−0.0002	**−0.0001**	0.0006	0.0010
0.99	24	−0.2002	−0.0429	−0.0426	−0.0428	**−0.0141**
	48	−0.1069	−0.0145	−0.0144	−0.0144	**−0.0048**
	96	−0.0554	−0.0052	−0.0053	−0.0052	**−0.0028**
	192	−0.0275	−0.0013	−0.0018	−0.0016	**−0.0005**

Entries in bold denote the minimum absolute bias in each row.

Table 3
Bias of OLS, jackknife, MU and bootstrap estimators

ϕ	n	$\hat{\phi}_n$	$\hat{\phi}_{J,2}$	$\hat{\phi}_{J,(2,3)}$	$\hat{\phi}_{\mathrm{MU}}$	$\hat{\phi}_{\mathrm{BS}}$
0.10	24	−0.0568	−0.0031	**−0.0019**	−0.0057	−0.0081
	48	−0.0281	−0.0011	**−0.0008**	−0.0029	−0.0026
	96	−0.0139	−0.0003	**−0.0001**	−0.0013	−0.0007
	192	−0.0072	**−0.0004**	−0.0005	−0.0009	−0.0006
0.30	24	−0.0820	−0.0063	**−0.0010**	−0.0141	−0.0136
	48	−0.0406	−0.0016	**−0.0006**	−0.0070	−0.0041
	96	−0.0202	−0.0004	**−0.0001**	−0.0035	−0.0012
	192	−0.0103	**−0.0005**	−0.0006	−0.0020	−0.0007
0.50	24	−0.1091	−0.0095	**0.0012**	−0.0227	−0.0200
	48	−0.0537	−0.0017	**0.0002**	−0.0113	−0.0056
	96	−0.0267	−0.0004	**0.0000**	−0.0057	−0.0017
	192	−0.0136	**−0.0005**	−0.0006	−0.0031	−0.0008
0.70	24	−0.1409	−0.0152	**0.0038**	−0.0334	−0.0298
	48	−0.0686	**−0.0013**	0.0021	−0.0161	−0.0077
	96	−0.0338	**−0.0001**	0.0004	−0.0081	−0.0023
	192	−0.0169	**−0.0003**	−0.0005	−0.0042	−0.0010
0.90	24	−0.1856	−0.0382	**−0.0117**	−0.0668	−0.0596
	48	−0.0913	−0.0054	**0.0030**	−0.0258	−0.0162
	96	−0.0435	**−0.0001**	0.0027	−0.0111	−0.0038
	192	−0.0209	−0.0007	**0.0002**	−0.0055	−0.0010
0.95	24	−0.1976	−0.0460	**−0.0192**	−0.0817	−0.0756
	48	−0.1013	−0.0123	**−0.0036**	−0.0364	−0.0256
	96	−0.0487	**−0.0001**	0.0025	−0.0141	−0.0066
	192	−0.0229	**−0.0002**	0.0010	−0.0060	−0.0015
0.99	24	−0.2002	−0.0429	**−0.0141**	−0.0917	−0.0877
	48	−0.1069	−0.0145	**−0.0048**	−0.0471	−0.0384
	96	−0.0554	−0.0052	**−0.0028**	−0.0233	−0.0162
	192	−0.0275	−0.0013	**−0.0005**	−0.0104	−0.0059

Entries in bold denote the minimum absolute bias in each row.

Table 4
Bias of OLS, jackknife, MU and bootstrap estimators under non-normal disturbances

ϕ	n	$\hat{\phi}_n$	$\hat{\phi}_{J,2}$	$\hat{\phi}_{J,(2,3)}$	$\hat{\phi}_{\mathrm{MU}}$	$\hat{\phi}_{\mathrm{BS}}$
$\epsilon \sim t_5$						
0.50	24	−0.1053	−0.0072	**0.0027**	0.0540	−0.0171
	48	−0.0520	**−0.0010**	0.0011	0.0505	−0.0046
	96	−0.0260	−0.0003	**0.0002**	−0.1159	−0.0014
	192	−0.0129	**0.0000**	0.0002	−0.0611	−0.0003
0.70	24	−0.1365	−0.0129	**0.0050**	−0.1066	−0.0288
	48	−0.0666	**−0.0008**	0.0026	−0.0807	−0.0081
	96	−0.0328	**0.0002**	0.0007	−0.1170	−0.0021
	192	−0.0162	**0.0002**	0.0002	−0.0519	−0.0007
0.90	24	−0.1816	−0.0367	**−0.0104**	−0.2789	−0.0574
	48	−0.0893	−0.0049	**0.0032**	−0.2019	−0.0159
	96	−0.0425	**0.0014**	0.0028	−0.1142	−0.0036
	192	−0.0204	0.0008	**0.0004**	−0.0808	−0.0010
0.95	24	−0.1943	−0.0453	**−0.0184**	−0.3035	−0.0733
	48	−0.0996	−0.0119	**−0.0032**	−0.2044	−0.0255
	96	−0.0478	**−0.0001**	0.0027	−0.1266	−0.0068
	192	−0.0226	0.0011	**0.0010**	−0.0882	−0.0016
0.99	24	−0.1975	−0.0426	**−0.0124**	−0.3112	−0.0853
	48	−0.1060	−0.0152	**−0.0061**	−0.1977	−0.0390
	96	−0.0551	−0.0056	**−0.0031**	−0.1049	−0.0172
	192	−0.0274	−0.0014	**−0.0005**	−0.0259	−0.0060
$\epsilon \sim \Gamma\left(1, \sqrt{5/3}\right) - \sqrt{5/3}$						
0.50	24	−0.1031	−0.0083	**−0.0015**	0.0616	−0.0140
	48	−0.0516	−0.0025	**−0.0008**	0.0505	−0.0030
	96	−0.0262	−0.0014	**−0.0009**	−0.1120	−0.0015
	192	−0.0131	−0.0005	**−0.0002**	−0.0542	−0.0008
0.70	24	−0.1329	−0.0141	**0.0006**	−0.0993	−0.0227
	48	−0.0656	**−0.0023**	0.0006	−0.0738	−0.0045
	96	−0.0327	**−0.0009**	−0.0005	−0.1097	−0.0019
	192	−0.0163	**−0.0004**	−0.0002	−0.0472	−0.0010
0.90	24	−0.1765	−0.0386	**−0.0156**	−0.2796	−0.0522
	48	−0.0879	−0.0069	**0.0009**	−0.2017	−0.0131
	96	−0.0422	**0.0006**	0.0020	−0.1111	−0.0037
	192	−0.0203	0.0005	**0.0003**	−0.0789	−0.0015
0.95	24	−0.1887	−0.0470	**−0.0222**	−0.3036	−0.0692
	48	−0.0980	−0.0137	**−0.0049**	−0.2037	−0.0229
	96	−0.0474	**−0.0009**	0.0019	−0.1267	−0.0068
	192	−0.0224	**0.0010**	0.0011	−0.0894	−0.0020
0.99	24	−0.1924	−0.0462	**−0.0187**	−0.3069	−0.0829
	48	−0.1041	−0.0165	**−0.0062**	−0.1938	−0.0365
	96	−0.0544	−0.0060	**−0.0031**	−0.1044	−0.0163
	192	−0.0272	−0.0016	**−0.0009**	−0.0256	−0.0063

Entries in bold denote the minimum absolute bias in each row.

Table 5
Bias of OLS, jackknife, MU and bootstrap estimators under ARCH disturbances

ϕ	n	$\hat{\phi}_n$	$\hat{\phi}_{J,2}$	$\hat{\phi}_{J,(2,3)}$	$\hat{\phi}_{\text{MU}}$	$\hat{\phi}_{\text{BS}}$
$\lambda = 0.5$						
0.50	24	−0.1229	−0.0280	−0.0156	**0.0134**	−0.0312
	48	−0.0670	−0.0141	**−0.0093**	0.0331	−0.0144
	96	−0.0362	−0.0071	**−0.0051**	−0.1257	−0.0071
	192	−0.0198	−0.0038	**−0.0025**	−0.0830	−0.0038
0.70	24	−0.1567	−0.0349	**−0.0142**	−0.1232	−0.0428
	48	−0.0827	−0.0139	**−0.0077**	−0.0928	−0.0165
	96	−0.0435	−0.0066	**−0.0044**	−0.1382	−0.0071
	192	−0.0234	−0.0037	**−0.0026**	−0.0727	−0.0038
0.90	24	−0.2008	−0.0532	**−0.0230**	−0.2772	−0.0736
	48	−0.1014	−0.0124	**−0.0021**	−0.2036	−0.0218
	96	−0.0494	−0.0018	**0.0006**	−0.1260	−0.0056
	192	−0.0246	−0.0010	**−0.0008**	−0.0827	−0.0022
0.95	24	−0.2130	−0.0589	**−0.0274**	−0.3052	−0.0907
	48	−0.1100	−0.0164	**−0.0057**	−0.2093	−0.0307
	96	−0.0529	**−0.0015**	0.0018	−0.1303	−0.0076
	192	−0.0253	**0.0004**	0.0007	−0.0872	−0.0020
0.99	24	−0.2149	−0.0520	**−0.0170**	−0.3140	−0.1025
	48	−0.1144	−0.0154	**−0.0033**	−0.2033	−0.0437
	96	−0.0583	−0.0045	**−0.0016**	−0.1083	−0.0172
	192	−0.0288	−0.0011	**−0.0005**	−0.0294	−0.0061
$\lambda = 0.9$						
0.50	24	−0.1329	−0.0446	−0.0342	**−0.0101**	−0.0390
	48	−0.0803	−0.0312	−0.0268	**0.0164**	−0.0238
	96	−0.0509	−0.0235	−0.0215	−0.1342	**−0.0145**
	192	−0.0342	−0.0182	−0.0160	−0.0983	**−0.0102**
0.70	24	−0.1690	−0.0541	**−0.0352**	−0.1373	−0.0524
	48	−0.0981	−0.0337	−0.0281	−0.1035	**−0.0265**
	96	−0.0602	−0.0251	−0.0227	−0.1543	**−0.0150**
	192	−0.0397	−0.0200	−0.0178	−0.0937	**−0.0102**
0.90	24	−0.2168	−0.0752	**−0.0425**	−0.2815	−0.0885
	48	−0.1148	−0.0282	**−0.0171**	−0.2072	−0.0294
	96	−0.0621	−0.0153	−0.0121	−0.1429	**−0.0099**
	192	−0.0363	−0.0122	−0.0111	−0.0925	**−0.0052**
0.95	24	−0.2338	−0.0869	**−0.0542**	−0.3105	−0.1122
	48	−0.1234	−0.0305	**−0.0172**	−0.2166	−0.0396
	96	−0.0631	−0.0112	**−0.0067**	−0.1380	−0.0105
	192	−0.0336	−0.0071	−0.0060	−0.0858	**−0.0033**
0.99	24	−0.2393	−0.0820	**−0.0469**	−0.3216	−0.1294
	48	−0.1308	−0.0309	**−0.0162**	−0.2145	−0.0579
	96	−0.0676	−0.0110	**−0.0059**	−0.1185	−0.0219
	192	−0.0340	−0.0040	**−0.0024**	−0.0397	−0.0070

Entries in bold denote the minimum absolute bias in each row.

Table 6
Bias of OLS, jackknife and bootstrap estimators of persistence measure (ρ) in correctly specified and mis-specified AR(2) and AR(4) models

	Estimation model is AR(2)				Estimation model is AR(4)			
n	$\hat{\rho}_n$	$\hat{\rho}_{J,2}$	$\hat{\rho}_{J,(2,3)}$	$\hat{\rho}_{\text{BS}}$	$\hat{\rho}_n$	$\hat{\rho}_{J,2}$	$\hat{\rho}_{J,(2,3)}$	$\hat{\rho}_{\text{BS}}$
DGP is AR(2)								
24	−0.1421	**−0.0006**	0.0078	−0.0423	−0.1803	0.0322	**−0.0263**	−0.0601
48	−0.0639	0.0054	**0.0050**	−0.0098	−0.0767	0.0125	**0.0048**	−0.0128
96	−0.0295	0.0027	**0.0014**	−0.0032	−0.0342	0.0047	**0.0017**	−0.0036
192	−0.0141	0.0007	**−0.0001**	−0.0010	−0.0160	0.0012	**−0.0000**	−0.0009
DGP is AR(4)								
24	−0.2136	−0.0317	**−0.0164**	−0.1082	−0.2133	**0.0168**	−0.0342	−0.0779
48	−0.1057	−0.0101	**−0.0071**	−0.0453	−0.0936	0.0103	**0.0058**	−0.0181
96	−0.0555	**−0.0093**	−0.0103	−0.0250	−0.0420	0.0052	**0.0026**	−0.0050
192	−0.0327	**−0.0114**	−0.0124	−0.0172	−0.0195	0.0015	**0.0001**	−0.0011

Entries in bold denote the minimum bias in each row of four columns.

Table 7
Percentage of negative bias in AR(1) model

	$\epsilon_t \sim N(0,1)$					$\epsilon_t \sim \Gamma(1, \sqrt{5/3})$				
n	$\hat{\phi}_n$	$\hat{\phi}_{J,2}$	$\hat{\phi}_{J,(2,3)}$	$\hat{\phi}_{\mathrm{MU}}$	$\hat{\phi}_{\mathrm{BS}}$	$\hat{\phi}_n$	$\hat{\phi}_{J,2}$	$\hat{\phi}_{J,(2,3)}$	$\hat{\phi}_{\mathrm{MU}}$	$\hat{\phi}_{\mathrm{BS}}$
$\phi = 0.10$										
24	60	51	54	51	51	63	54	55	38	53
48	57	50	52	50	50	60	53	55	13	53
96	55	50	51	50	50	57	53	54	56	52
192	54	50	50	50	50	55	52	52	51	51
$\phi = 0.30$										
24	64	51	55	51	51	66	53	56	33	52
48	60	50	52	50	50	63	53	55	20	52
96	57	49	51	50	49	58	52	54	58	51
192	55	50	50	50	50	56	51	52	50	51
$\phi = 0.50$										
24	69	51	55	51	51	70	53	55	32	51
48	63	49	53	51	49	65	52	55	24	50
96	59	49	51	50	48	61	51	54	60	50
192	57	49	50	50	49	58	51	52	43	50
$\phi = 0.70$										
24	76	52	55	52	51	77	53	55	42	52
48	69	49	54	51	48	70	51	55	53	48
96	64	48	53	50	48	65	50	54	52	49
192	60	48	51	50	48	61	49	52	41	49
$\phi = 0.90$										
24	90	58	56	54	57	90	58	56	95	57
48	83	52	55	53	49	84	53	56	93	49
96	75	48	55	51	46	76	49	55	95	47
192	68	47	54	51	46	69	48	54	95	47
$\phi = 0.95$										
24	93	58	56	54	61	93	58	56	92	62
48	91	55	56	54	53	91	56	56	85	54
96	83	51	55	53	48	84	51	55	71	49
192	76	48	55	51	45	76	48	55	48	46
$\phi = 0.99$										
24	95	56	55	52	66	94	57	55	85	66
48	95	55	55	53	61	95	55	55	71	61
96	94	54	56	54	56	95	55	56	41	57
192	92	54	56	54	51	93	54	56	10	52

Table 8
RMSE of OLS, jackknife, MU and bootstrap estimators

ϕ	n	$\hat{\phi}_n$	$\hat{\phi}_{J,2}$	$\hat{\phi}_{J,m_r}$	$\hat{\phi}_{J,(2,3)}$	$\hat{\phi}_{J,(m_1,m_2)_r}$	$\hat{\phi}_{\mathrm{MU}}$	$\hat{\phi}_{\mathrm{BS}}$
0.10	24	**0.2089**	0.2411	0.2222^{6}	0.3298	$0.2592^{2,8}$	0.2198	0.2235
	48	**0.1456**	0.1572	0.1492^{12}	0.1902	$0.1588^{8,12}$	0.1492	0.1512
	96	**0.1021**	0.1066	0.1033^{24}	0.1195	$0.1055^{16,24}$	0.1033	0.1042
	192	**0.0721**	0.0734	0.0724^{48}	0.0785	$0.0729^{24,48}$	0.0725	0.0728
0.30	24	**0.2156**	0.2460	0.2225^{6}	0.3553	$0.2651^{2,8}$	0.2196	0.2231
	48	**0.1455**	0.1565	0.1466^{12}	0.1991	$0.1579^{8,12}$	0.1463	0.1483
	96	**0.1002**	0.1044	0.1003^{24}	0.1213	$0.1029^{16,24}$	0.1003	0.1012
	192	**0.0700**	0.0714	**0.0700**^{48}	0.0779	$0.0705^{24,48}$	**0.0700**	0.0703
0.50	24	0.2222	0.2492	0.2191^{6}	0.3881	$0.2730^{2,8}$	**0.2164**	0.2189
	48	0.1425	0.1527	0.1393^{12}	0.2102	$0.1527^{8,12}$	**0.1389**	0.1407
	96	0.0949	0.0988	0.0934^{24}	0.1223	$0.0963^{16,24}$	**0.0933**	0.0941
	192	0.0650	0.0663	**0.0644**^{48}	0.0755	$0.0650^{24,48}$	**0.0644**	0.0646
0.70	24	0.2313	0.2519	0.2128^{6}	0.4304	$0.2813^{2,8}$	**0.2097**	0.2119
	48	0.1373	0.1466	**0.1270**^{12}	0.2278	$0.1422^{8,12}$	**0.1270**	0.1280
	96	0.0860	0.0896	0.0816^{24}	0.1241	$0.0847^{16,24}$	**0.0815**	0.0820
	192	0.0565	0.0576	**0.0547**^{48}	0.0716	$0.0553^{24,48}$	**0.0547**	0.0548
0.90	24	0.2500	0.2528	0.2081^{6}	0.4667	$0.2847^{4,6}$	**0.1897**	0.2021
	48	0.1352	0.1406	0.1109^{8}	0.2577	$0.1254^{8,12}$	**0.1098**	0.1111
	96	0.0738	0.0780	**0.0627**^{16}	0.1352	$0.0657^{16,24}$	0.0636	0.0640
	192	0.0427	0.0443	**0.0381**^{48}	0.0694	$0.0386^{24,48}$	0.0383	0.0388
0.95	24	0.2552	0.2519	0.2068^{6}	0.4678	$0.2829^{4,6}$	**0.1825**	0.1983
	48	0.1377	0.1382	0.1069^{8}	0.2613	$0.1204^{8,12}$	**0.1021**	0.1061
	96	0.0721	0.0752	**0.0570**^{12}	0.1408	$0.0596^{16,24}$	0.0574	0.0575
	192	0.0386	0.0406	**0.0320**^{24}	0.0719	$0.0324^{24,48}$	0.0328	0.0330
0.99	24	0.2543	0.2470	0.2028^{6}	0.4612	$0.2790^{4,6}$	**0.1744**	0.1927
	48	0.1374	0.1342	0.1017^{8}	0.2584	$0.1149^{8,12}$	**0.0938**	0.0998
	96	0.0725	0.0718	0.0521^{12}	0.1402	$0.0541^{16,24}$	**0.0503**	0.0520
	192	0.0372	0.0376	0.0266^{16}	0.0741	**0.0265**^{24,48}	0.0267	0.0271

Entries in bold denote the minimum RMSE in each row and superscripts denote the relevant values of m_r and $(m_1, m_2)_r$.

Table 9
AR(1) 90% confidence interval type I errors

ϕ	Estimator	$n = 24$	$n = 48$	$n = 96$	$n = 192$
0.50	$\hat{\phi}_n$	0.126	0.113	0.107	0.104
	$\hat{\phi}_{J,2}$	0.121	0.121	0.115	0.110
	$\hat{\phi}_{J,m_r}$	0.112	0.112	0.107	0.104
	$\hat{\phi}_{J,(2,3)}$	0.073	0.113	0.120	0.117
	$\hat{\phi}_{J,(m_1,m_2)_r}$	0.113	0.113	0.111	0.108
	$\hat{\phi}_{\text{MU}}$	0.108	0.104	0.101	0.101
	$\hat{\phi}_{\text{BS}}$	0.183	0.147	0.125	0.114
0.70	$\hat{\phi}_n$	0.156	0.130	0.117	0.109
	$\hat{\phi}_{J,2}$	0.129	0.133	0.127	0.116
	$\hat{\phi}_{J,m_r}$	0.115	0.117	0.112	0.107
	$\hat{\phi}_{J,(2,3)}$	0.084	0.121	0.135	0.129
	$\hat{\phi}_{J,(m_1,m_2)_r}$	0.115	0.115	0.118	0.114
	$\hat{\phi}_{\text{MU}}$	0.109	0.106	0.103	0.102
	$\hat{\phi}_{\text{BS}}$	0.193	0.154	0.129	0.116
0.90	$\hat{\phi}_n$	0.263	0.201	0.158	0.131
	$\hat{\phi}_{J,2}$	0.145	0.162	0.158	0.143
	$\hat{\phi}_{J,m_r}$	0.130	0.136	0.133	0.116
	$\hat{\phi}_{J,(2,3)}$	0.115	0.142	0.182	0.173
	$\hat{\phi}_{J,(m_1,m_2)_r}$	0.081	0.106	0.129	0.136
	$\hat{\phi}_{\text{MU}}$	0.101	0.105	0.107	0.104
	$\hat{\phi}_{\text{BS}}$	0.209	0.180	0.154	0.130
0.95	$\hat{\phi}_n$	0.335	0.273	0.209	0.161
	$\hat{\phi}_{J,2}$	0.158	0.175	0.185	0.166
	$\hat{\phi}_{J,m_r}$	0.140	0.139	0.149	0.140
	$\hat{\phi}_{J,(2,3)}$	0.123	0.161	0.202	0.217
	$\hat{\phi}_{J,(m_1,m_2)_r}$	0.084	0.099	0.129	0.154
	$\hat{\phi}_{\text{MU}}$	0.097	0.100	0.107	0.105
	$\hat{\phi}_{\text{BS}}$	0.220	0.186	0.171	0.147
0.99	$\hat{\phi}_n$	0.422	0.408	0.368	0.304
	$\hat{\phi}_{J,2}$	0.170	0.190	0.219	0.221
	$\hat{\phi}_{J,m_r}$	0.149	0.155	0.164	0.166
	$\hat{\phi}_{J,(2,3)}$	0.118	0.164	0.206	0.264
	$\hat{\phi}_{J,(m_1,m_2)_r}$	0.080	0.094	0.125	0.167
	$\hat{\phi}_{\text{MU}}$	0.097	0.094	0.097	0.100
	$\hat{\phi}_{\text{BS}}$	0.245	0.216	0.197	0.178

Table 10
AR(1) 90% two-sided confidence interval type I errors

		$n = 24$		$n = 48$		$n = 96$		$n = 192$	
ϕ	Estimator	P_L	P_U	P_L	P_U	P_L	P_U	P_L	P_U
0.50	$\hat{\phi}_n$	0.014	0.112	0.022	0.091	0.028	0.079	0.034	0.071
	$\hat{\phi}_{J,2}$	0.052	0.070	0.062	0.059	0.062	0.054	0.057	0.052
	$\hat{\phi}_{J,m_r}$	0.045	0.067	0.052	0.060	0.052	0.055	0.051	0.054
	$\hat{\phi}_{J,(2,3)}$	0.007	0.066	0.044	0.068	0.056	0.064	0.058	0.059
	$\hat{\phi}_{J,(m_1,m_2)_r}$	0.044	0.069	0.060	0.052	0.061	0.051	0.059	0.050
	$\hat{\phi}_{\text{MU}}$	0.051	0.057	0.051	0.054	0.050	0.052	0.050	0.052
	$\hat{\phi}_{\text{BS}}$	0.084	0.099	0.072	0.075	0.062	0.063	0.057	0.057
0.70	$\hat{\phi}_n$	0.007	0.149	0.014	0.116	0.022	0.096	0.028	0.081
	$\hat{\phi}_{J,2}$	0.047	0.082	0.069	0.064	0.071	0.056	0.064	0.052
	$\hat{\phi}_{J,m_r}$	0.036	0.080	0.051	0.065	0.052	0.060	0.052	0.055
	$\hat{\phi}_{J,(2,3)}$	0.002	0.082	0.036	0.085	0.059	0.076	0.062	0.067
	$\hat{\phi}_{J,(m_1,m_2)_r}$	0.035	0.080	0.062	0.053	0.069	0.049	0.066	0.048
	$\hat{\phi}_{\text{MU}}$	0.050	0.059	0.051	0.055	0.049	0.053	0.050	0.052
	$\hat{\phi}_{\text{BS}}$	0.082	0.111	0.074	0.080	0.063	0.066	0.058	0.058
0.90	$\hat{\phi}_n$	0.002	0.261	0.003	0.197	0.009	0.149	0.016	0.115
	$\hat{\phi}_{J,2}$	0.026	0.118	0.075	0.087	0.091	0.067	0.085	0.058
	$\hat{\phi}_{J,m_r}$	0.018	0.112	0.054	0.082	0.067	0.066	0.054	0.062
	$\hat{\phi}_{J,(2,3)}$	0.000	0.114	0.009	0.134	0.061	0.121	0.073	0.100
	$\hat{\phi}_{J,(m_1,m_2)_r}$	0.003	0.078	0.044	0.062	0.078	0.051	0.090	0.046
	$\hat{\phi}_{\text{MU}}$	0.040	0.061	0.046	0.060	0.050	0.057	0.050	0.054
	$\hat{\phi}_{\text{BS}}$	0.052	0.157	0.071	0.109	0.072	0.082	0.065	0.066
0.95	$\hat{\phi}_n$	0.001	0.334	0.001	0.272	0.003	0.206	0.008	0.153
	$\hat{\phi}_{J,2}$	0.015	0.143	0.063	0.111	0.101	0.084	0.100	0.066
	$\hat{\phi}_{J,m_r}$	0.014	0.126	0.042	0.097	0.076	0.074	0.079	0.060
	$\hat{\phi}_{J,(2,3)}$	0.000	0.123	0.004	0.157	0.047	0.155	0.083	0.134
	$\hat{\phi}_{J,(m_1,m_2)_r}$	0.002	0.082	0.030	0.069	0.071	0.058	0.105	0.049
	$\hat{\phi}_{\text{MU}}$	0.039	0.058	0.039	0.061	0.047	0.060	0.049	0.056
	$\hat{\phi}_{\text{BS}}$	0.031	0.189	0.049	0.137	0.069	0.102	0.069	0.078
0.99	$\hat{\phi}_n$	0.001	0.421	0.000	0.408	0.001	0.368	0.001	0.303
	$\hat{\phi}_{J,2}$	0.009	0.161	0.040	0.150	0.085	0.134	0.112	0.109
	$\hat{\phi}_{J,m_r}$	0.012	0.137	0.038	0.117	0.063	0.101	0.087	0.079
	$\hat{\phi}_{J,(2,3)}$	0.000	0.118	0.002	0.162	0.017	0.190	0.067	0.197
	$\hat{\phi}_{J,(m_1,m_2)_r}$	0.001	0.079	0.020	0.074	0.051	0.074	0.100	0.067
	$\hat{\phi}_{\text{MU}}$	0.046	0.051	0.041	0.053	0.040	0.058	0.040	0.060
	$\hat{\phi}_{\text{BS}}$	0.016	0.229	0.019	0.198	0.029	0.168	0.044	0.134