Jackknife Estimation of Stationary Autoregressive Models

Marcus J. Chambers
University of Essex

January 2010

Abstract

This paper reports the results of an extensive investigation into the use of the jackknife as a method of estimation in stationary autoregressive models. In addition to providing some general theoretical results concerning jackknife methods it is shown that a method based on the use of non-overlapping sub-intervals is found to work particularly well and is capable of reducing bias and root mean squared error (RMSE) compared to ordinary least squares (OLS), subject to a suitable choice of the number of sub-samples, rules-of-thumb for which are provided. The jackknife estimators also outperform OLS when the distribution of the disturbances departs from normality and when it is subject to autoregressive conditional heteroskedasticity. Furthermore the jackknife estimators are much closer to being median-unbiased than their OLS counterparts.

Keywords. Jackknife; bias; autoregression.

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

Acknowledgements: I would like to thank Roy Bailey, Emma Iglesias, Maria Kyriacou, Joao Santos Silva and Michael Thornton for helpful comments on this piece of work. 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 above 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.

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. Although primarily motivated by the desire to achieve improved accuracy in finite samples, section 2 concludes with some results (contained in Theorem 4 and a Corollary) exploring the asymptotic properties of the jackknife statistic, which is particularly relevant for purposes of inference.

Section 3 explores jackknife methods of estimation in stationary autoregressive models. The main focus is on three variants of the first-order autoregressive, or AR(1), model, the variants corresponding to the presence or absence of intercept and trend. The motivation for employing jackknife estimators in this context is rooted in analytical work that provides expansions for the bias of the ordinary least squares (OLS) estimator of the AR parameter. The results from an extensive simulation exercise (involving 100,000 replications) are reported in an attempt to obtain evidence on: 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 (2010). The simulations also enable an optimal expansion rate for the bias-minimising number of sub-samples to be determined. Some results are also provided for the AR(2) and AR(4) models.

Additional considerations concerning the performance of the jackknife techniques are explored in section 4. Although designed to reduce bias other distributional aspects are important to the usefulness of an estimator, and so the mean squared error (MSE) is examined first. As in the case of estimator bias this is motivated by theoretical results concerning expansions of the MSE of the OLS estimator in the AR(1) model. Simulations reveal that it is possible to obtain an MSE less than the full-sample OLS estimator by using jackknife estimators, and optimal root MSE (RMSE) expansion rules for the number of sub-samples are determined. A feature of these is that a larger number of sub-samples is required to minimise RMSE than to minimise bias. Departures from normality are also explored, in particular disturbances generated by Student's t- and Gamma distributions, as well as autoregressive conditional heteroskedasticity (ARCH). Jackknife estimators continue to be characterised by a smaller bias than the OLS estimator in these scenarios. 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. All proofs are contained in the Appendix, and section 6 concludes.

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 term from its expectation. The statistic of interest 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 $S_n = S(y)$ denote a statistic of interest satisfying*

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

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 $S_i = S(Y_i)$ $(i = 1, \ldots, m)$ denote the corresponding sub-sample statistics. Then the jackknife statistic

$$S_J = \left(\frac{n}{n-\ell}\right) S_n - \left(\frac{\ell}{n-\ell}\right) \frac{1}{m} \sum_{i=1}^{m} S_i \tag{2}$$

satisfies $E(S_J) = S + O(n^{-2})$.

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. Some specific cases will now be considered and Theorem 1 will be employed to determine the appropriate weights to use in the construction of S_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 $Y_i = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)'$ $(i = 1, \ldots, n)$. Here, $m = n$ and the size of each sub-sample is $\ell = n - 1$. Hence, from Theorem 1, S_J takes the form (using the fact that $n - \ell = 1$)

$$S_J = nS_n - (n-1)\frac{1}{n} \sum_{i=1}^{n} S_i. \tag{3}$$

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 S_J 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

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

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:

$$S_J = \left(\frac{m}{m-1}\right) S_n - \left(\frac{1}{m-1}\right) \frac{1}{m} \sum_{i=1}^{m} S_i. \tag{4}$$

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

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

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

$$S_J = \left(\frac{n}{m-1}\right) S_n - \left(\frac{\ell}{m-1}\right) \frac{1}{m} \sum_{i=1}^{m} S_i. \tag{5}$$

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 sub-sampling 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

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

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

$$S_J = \left(\frac{m}{m-1}\right) S_n - \left(\frac{1}{m-1}\right) \frac{1}{2m-1} \sum_{i=1}^{2m-1} S_i. \tag{6}$$

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 S_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 $S_{1,i}$ $(i = 1, \ldots, m_1)$ and $S_{2,i}$ $(i = 1, \ldots, m_2)$ denote the corresponding sub-sample statistics. Then the jackknife statistic*

$$S_J = w_n S_n + w_{1n} \frac{1}{m_1} \sum_{i=1}^{m_1} S_{1,i} + w_{2n} \frac{1}{m_2} \sum_{i=1}^{m_2} S_{2,i} \tag{7}$$

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 $E(S_J) = S + O(n^{-3})$.

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 $S_{1,i}$ statistics are obtained from non-overlapping sub-samples and the $S_{2,i}$ 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

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

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 S_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 S_i $(i = 1, \ldots, m)$ denote the corresponding sub-sample statistics. Then the jackknife statistic*

$$S_J = k_n S_n + k_{1n} \frac{1}{m} \sum_{i=1}^{m} S_i, \tag{8}$$

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 $E(S_J) = S + O(n^{-2})$.

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.

2.6 Asymptotic properties

Although the jackknife methods are intended to provide bias reduction in small finite samples it is nevertheless important also to examine their asymptotic properties, not least because inferential methods are often based on such properties. In many cases of interest the asymptotic properties of the statistic S_n will already be known, and so it is natural to relate the limiting properties of S_J to those of S_n. The main result for the generic jackknife statistic S_J is given in Theorem 4.

Theorem 4. *(a) Under the conditions of Theorem 1, if $S_n \xrightarrow{p} S$ as $n \to \infty$ then $S_J \xrightarrow{p} S$ as $n \to \infty$ if:*
(i) m is fixed as $n \to \infty$; or
(ii) $m, \ell \to \infty$ and S_i satisfies a (weak) law of large numbers such that

$$\frac{1}{m} \sum_{i=1}^{m} (S_i - E(S_i)) \xrightarrow{p} 0 \quad \text{as } n \to \infty.$$

(b) If, in addition to the conditions of Theorem 1, $\sqrt{n}(S_n - S) \xrightarrow{d} Z$ as $n \to \infty$, where $Z = O_p(1)$, then

$$\sqrt{n}(S_J - S) \xrightarrow{d} \begin{cases} Z + O_p(1) & \text{if } m \text{ is fixed and } \ell \to \infty \text{ as } n \to \infty; \\ Z & \text{if } m \to \infty \text{ and } (1/\ell) + (\ell/n) \to 0 \text{ as } n \to \infty. \end{cases}$$

Theorem 4(a) establishes the consistency of the jackknife estimator. In addition to the fairly minimal conditions of Theorem 1, when m is fixed it is also assumed that the statistic S_n (and hence the $S_1, \ldots, S_m$ also) is a consistent estimator of S. Many statistics of interest will converge to a finite limit (at least) and such conditions are satisfied by the cases of interest in the context of stationary autoregressions. When m is allowed to grow with the sample size then it is also assumed that S_i satisfies a (weak) law of large numbers.

Although this is a high-level assumption it can usually be verified under certain conditions for the relevent application of interest. For example, in the cases of stationary autoregressions considered here one can appeal to the ergodic theorem to verify this requirement. Part (b) of Theorem 4 relates the asymptotic distribution of S_J to that of S_n and shows that the difference of the appropriately normalised and centered statistics is $O_p(1)$ when $\ell = O(n)$. In order for S_J to have the same limiting distribution as S_n requires, in addition, that ℓ increases more slowly than n. In the case of non-overlapping sub-samples this requires (recall that $\ell = n/m$) that

$$\frac{1}{m} + \frac{m}{n} \to 0 \quad \text{as} \quad n \to \infty,$$

and hence the number of sub-samples, m, must also grow with n but at a slower rate in order for S_J to share the same limiting distribution as S_n.

Corollary to Theorem 4. *(a) Assume that, as $n \to \infty$, $\ell = O(n)$,*

$$\sqrt{n}\,(S_n - S) \xrightarrow{d} Z \quad \text{and} \quad \sqrt{\ell}\,(S_i - S) \xrightarrow{d} Z_i, \quad i = 1, \ldots, m,$$

where Z and the Z_i $(i = 1, \ldots, m)$ are $O_p(1)$ random variables. Furthermore, let

$$K = \lim_{n\to\infty} \left(\frac{n}{n-\ell}\right), \quad \lambda = \lim_{n\to\infty} \left(\frac{\sqrt{n}}{\sqrt{\ell}}\right).$$

Then

$$\sqrt{n}\,(S_J - S) \xrightarrow{d} KZ - \frac{\lambda(K-1)}{m} \sum_{i=1}^{m} Z_i$$

as $n \to \infty$.
(b) Furthermore, in the case where $Z \sim N(0, \sigma^2)$ and $Z_i \sim N(0, \sigma^2)$ $(i = 1, \ldots, m)$, let $\sigma_{ij} = cov\,(Z_i, Z_j)$ $(i, j = 1, \ldots, m)$ and $\sigma_i = cov\,(Z_i, Z)$ $(i = 1, \ldots, m)$. Then, as $n \to \infty$,

$$\sqrt{n}\,(S_J - S) \xrightarrow{d} N(0, \sigma_J^2)$$

where

$$\sigma_J^2 = \sigma^2 \left[K^2 + \frac{\lambda^2(K-1)^2}{m}\right] + 2\lambda \left(\frac{K-1}{m}\right) \left[\lambda \left(\frac{K-1}{m}\right) \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \sigma_{ij} - K \sum_{i=1}^{m} \sigma_{in}\right].$$

The first part of the Corollary defines the limiting distribution of S_J in terms of the limits of the sub-sample estimators S_i and S_n when $\ell = O(n)$. In general the limiting distribution of S_J will be different to that of S_n. The second part of the Corollary deals with the case where the individual limits are Gaussian which results in the limiting distribution of S_J also being Gaussian albeit with a potentially different variance to S_n. The assumption of asymptotic normality for S_i $(i = 1, \ldots, m)$ is certainly consistent with the assumed expansion for $E(S_i)$ used in Theorems 1 and 2. Of course, if the limiting distributions of the sub-sample statistics were different from the limiting distribution of the full sample statistic S_n then the assumed expansions might also differ, in which case the jackknife statistic as defined would not succeed in eliminating the first-order finite sample bias. Such a situation arises when a unit root is present in an AR process; see Chambers and Kyriacou (2010) for details. Theorem 4 and its Corollary suggest that, in general, the limiting distribution of S_J will not necessarily be the same as the uncorrected statistic S_n unless the rate of increase of ℓ with n is suitably controlled. This issue is important if the limiting distribution of S_n, denoted Z, is used to conduct inference with S_J.

3. Bias reduction in stationary autoregressions

This section will be concerned with (special cases of) the general AR process of order p, or AR(p) process, given by

$$y_t = \phi' d_t + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \epsilon_t, \quad t = 1, \ldots, n, \tag{9}$$

where d_t denotes a vector of deterministic terms, for example an intercept and powers of t, ϵ_t is an i.i.d. disturbance, and the polynomial $\beta(z) = 1 - \beta_1 z - \ldots - \beta_p z^p$ has all its roots lying outside the unit circle. For convenience it will be assumed that $y_{-p+1}, \ldots, y_{-1}, y_0$ are observed. The unknown parameter vector of interest is $\theta = (\beta_1, \ldots, \beta_p, \phi')'$ enabling the regression model to be written in standard matrix form $y = X\theta + \epsilon$, where $y = (y_1, \ldots, y_n)'$ and X is the matrix with typical row $(y_{t-1}, \ldots, y_{t-p}, d_t')$. The OLS estimator of θ is given by $\hat{\theta} = (X'X)^{-1}X'y$ but $\hat{\theta}$ is not an unbiased estimator of θ in this model. Analytic expressions for the first-order bias term are provided by Shaman and Stine (1988) and Stine and Shaman (1989), and it is this term that the jackknife estimator aims to remove. From (2) the general form of the jackknife estimator of θ is given by

$$\hat{\theta}_J = \left(\frac{n}{n-\ell}\right)\hat{\theta} - \left(\frac{\ell}{n-\ell}\right)\frac{1}{m}\sum_{i=1}^{m}\hat{\theta}_i, \tag{10}$$

where $\hat{\theta}_i$ $(i = 1, \ldots, m)$ are the sub-sample OLS estimators and each sub-sample is of length l. Although $\hat{\theta} = \theta + O(n^{-1})$ the jackknife estimator satisfies $\hat{\theta}_J = \theta + O(n^{-2})$.

The main initial focus in this section is the application of jackknife methods to OLS estimation of the autoregressive parameter in the AR(1) model before moving on to more general AR(2)and AR(4) models. It is convenient to treat the pure AR(1) model, without intercept or trend, separately from the model containing deterministic terms.

3.1 The pure AR(1) model

The stationary pure AR(1) model is obtained from (9) by setting $p = 1$ and $d_t = 0$, resulting in

$$\text{Model A:} \;\; y_t = \beta y_{t-1} + \epsilon_t, \quad |\beta| < 1, \quad t = 1, \ldots, n, \tag{11}$$

where $\epsilon_t \sim$ i.i.d. $(0, \sigma^2)$ and y_0 will be assumed to be observed by the econometrician. The OLS estimator of β, given by

$$\hat{\beta} = \left(\sum_{t=1}^{n} y_{t-1}^2\right)^{-1} \sum_{t=1}^{n} y_t y_{t-1} = \beta + \left(\sum_{t=1}^{n} y_{t-1}^2\right)^{-1} \sum_{t=1}^{n} \epsilon_t y_{t-1}, \tag{12}$$

is biased but consistent in this model and the objective is to establish the extent to which jackknife techniques can produce an estimator with smaller bias. Shenton and Johnson (1965) demonstrated that, when y_0 is fixed and $\epsilon_t \sim$ i.i.d. $N(0,1)$, then

$$E\left(\hat{\beta} - \beta\right) = -\frac{2\beta}{n} + \frac{4\beta}{n^2} + O\left(n^{-3}\right). \tag{13}$$

More recently Bao (2007) relaxed the normality assumption underlying this expansion and demonstrated how the initial condition y_0 affects the $O(n^{-2})$ term. Concentrating on the effect of y_0 while maintaining normality, the results in Bao (2007) yield

$$E\left(\hat{\beta}-\beta\right) = -\frac{2\beta}{n} + \frac{1}{n^2}\left[4\beta + \frac{2\beta y_0^2}{\sigma^2}\right] + O\left(n^{-3}\right). \tag{14}$$

The simulations reported below take $y_0 = 0$ to correspond with the unconditional mean of the y_t process.

The simulations are based on a range of autoregressive parameter values that would appear to be most relevant in practice, so that $\beta \in \{0.1, 0.3, 0.5, 0.7, 0.9, 0.95, 0.99\}$, 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 applied, so that $m \in \{2, 3, 4, 6, 8, 12, 16, 24, 48\}$, although not all of these choices are possible for each sample size. The random shocks satisfy $\epsilon_t \sim$ i.i.d. $N(0,1)$, the estimator $\hat{\beta}$ being invariant to the standard error of ϵ_t when $y_0 = 0$.[2] The results reported below are based on a total of 100,000 replications of each combination of parameters and sample sizes.

Table 1 presents the bias of the OLS estimator and the jackknife estimator based on non-overlapping sub-samples for $m \in \{2, 3, 4, 6, 8\}$, the latter being defined by

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

where $\hat{\beta}_i$ $(i = 1, \ldots, m)$ denotes the OLS estimator based on sub-sample i. From (13) and (14) (with $y_0 = 0$) $\hat{\beta}_{J,m}$ satisfies

$$E\left(\hat{\beta}_{J,m}-\beta\right) = \frac{4\beta}{n^2} + O\left(n^{-3}\right). \tag{15}$$

It is clear from Table 1 that the jackknife estimators can achieve substantial bias reduction, compared to OLS, for all values of m, with the greatest reduction being obtained with $m = 2$ in all cases. For example, when $\beta = 0.5$ the jackknife achieves a 73% bias reduction for a sample size as small as $n = 24$ rising to 89% when $n = 192$. Although the percentage reduction falls as β increases the jackknife estimator with $m = 2$ still achieves a 50% reduction in bias for $n = 24$ when $\beta = 0.99$ which rises to 71% for $n = 192$.

Three further jackknife estimators were also considered, these being two moving-blocks jackknife estimators as well as the jackknife estimator that eliminates both first- and second-order biases. For the first moving-blocks-based estimator, in order to provide a point of reference with the results in Table 1, the sub-sample sizes, ℓ, are chosen to be the same size as for the non-overlapping case, and so $\ell = n/m$. The corresponding number of sub-samples is then $n - (n/m) + 1$ and so, using (5), the estimator is given by

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

For example, when $n = 24$ and $m = 2$, the sub-samples are of length $\ell = 12$, but the moving-blocks estimator uses 13 sub-samples compared to $m = 2$ in the non-overlapping case. The second moving-blocks jackknife estimator utilises a reduced number of $2m - 1$

[2]It can be seen from (14) that the ratio y_0/σ affects the $O(n^{-2})$ term in the expansion of bias.

moving blocks in which each block of length $\ell = n/m$ is shifted by $\ell/2$ observations, so that the estimator is of the form

$$\hat{\beta}_{J,m}^{\text{MB2}} = \left(\frac{m}{m-1}\right)\hat{\beta} - \left(\frac{1}{m-1}\right)\frac{1}{2m-1}\sum_{i=1}^{2m-1}\hat{\beta}_i;$$

see (6). The last jackknife estimator aims to eliminate both first- and second-order bias terms and is based upon combining estimators from two sets of non-overlapping sub-samples. From (7) and the ensuing discussion the estimator is

$$\begin{aligned}\hat{\beta}_{J,M} &= \left(\frac{m_1 m_2}{(m_1-1)(m_2-1)}\right)\hat{\beta} - \left(\frac{m_2}{m_1(m_1-1)(m_2-m_1)}\right)\frac{1}{m_1}\sum_{i=1}^{m_1}\hat{\beta}_{1,i} \\ &\quad + \left(\frac{m_1}{(m_2-1)(m_2-m_1)}\right)\frac{1}{m_2}\sum_{i=1}^{m_2}\hat{\beta}_{2,i},\end{aligned}$$

where m_1 and m_2 denote the numbers of sub-samples, $M = (m_1, m_2)$, and $\hat{\beta}_{1,i}$ and $\hat{\beta}_{2,i}$ denote the sub-sample estimators. Both moving-blocks jackknife estimators have bias of $O(n^{-2})$, in accordance with the bias expansion of $\hat{\beta}_{J,m}$ in (15), while the bias of $\hat{\beta}_{J,M}$ is $O(n^{-3})$.

The bias results for these three estimators are reported in Table 2, and relate to the value of m found to yield the *minimum* bias in Table 1, namely $m = 2$. For $\hat{\beta}_{J,M}$ the choice $M = (2, 3)$ was employed.[3] The entries for these estimators in Table 2 are the bias expressed as a ratio of the OLS bias. Also reported is the corresponding value for $\hat{\beta}_{J,2}$, based on the values reported in Table 1, as well as the actual bias of the OLS estimator from Table 1 to serve as a reference point. All of the jackknife estimators provide bias reduction compared to the OLS estimator. Both moving-blocks estimators are inferior to the non-overlapping sub-samples jackknife in all cases, but the jackknife estimator based on non-overlapping sub-samples that also eliminates second-order bias provides the most spectacular bias reductions uniformly across all parameter values and sample sizes. For example, when $\beta = 0.50$ and $n = 24$ this estimator produces a 92% reduction in bias and a 69% reduction even when $\beta = 0.99$ for the same (small) sample size. It is also evident from Table 2 that the proportion of OLS bias that is eliminated by the jackknife estimators is a decreasing function of the parameter β and an increasing function of sample size n for larger values of β although it appears to be U-shaped for smaller values of β.

3.2 The AR(1) model with constant and/or trend

Extending the pure AR(1) model to contain deterministic trend components is straightforward and two such extensions are considered. The first incorprates a constant in the regression; setting $d_t = 1$ in (9) the model becomes

$$\text{Model B:}\;\; y_t = \alpha + \beta y_{t-1} + \epsilon_t, \quad |\beta| < 1, \quad t = 1, \ldots, n, \tag{16}$$

[3] In the overwhelming majority of cases these values of m and M did yield the smallest bias. The exceptions are as follows: for $\hat{\beta}_{J,m}^{\text{MB}}$, when $\beta = 0.10$ bias is minimised at $m = 3$ for $n = 96$ and $m = 6$ for $n = 192$, while for $\beta = 0.30$ it is minimised at $m = 3$ for $n = 192$; for $\hat{\beta}_{J,2}^{\text{MB2}}$, when $\beta = 0.10$ bias is minimised at $m = 6$ for $n = 192$; and for $\hat{\beta}_{J,M}$, when $\beta = 0.10$ bias is minimised at $M = (6, 8)$ for $n = 96$ and at $M = (12, 16)$ for $n = 192$, when $\beta = 0.30$ the minimum bias occurs at $M = (4, 6)$ for $n = 96$ and at $M = (6, 8)$ for $n = 192$, and when $\beta = 0.50$ it is minimised at $M = (6, 8)$ for $n = 192$.

with ϵ_t and y_0 defined as before. In this model Sawa (1978) found the following moment expansion, originally due to Kendall (1954), to work well, despite its simplicity:

$$E\left(\hat{\beta}-\beta\right) = -\frac{1+3\beta}{n} + O\left(n^{-2}\right). \tag{17}$$

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

$$E\left(\hat{\beta}-\beta\right) = -\frac{1+3\beta}{n} + \frac{1}{n^2}\left[\frac{3\beta-9\beta^2-1}{1-\beta} + \frac{1+3\beta}{(1-\beta)^2}\left(\frac{g(\alpha,\beta,y_0)}{\sigma}\right)^2\right] + O\left(n^{-3}\right) \tag{18}$$

where $g(\alpha,\beta,y_0) = \alpha-(1-\beta)y_0$. In the simulations the intercept was chosen to satisfy $\alpha = (1-\beta)y_0$ so as to remove the dependence of the $O(n^{-2})$ term on y_0 and also to make $\hat{\beta}$ invariant to α and y_0. It then follows that

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

which is also satisfied by the two moving-blocks jackknife estimators. The bias of $\hat{\beta}_{J,M}$ is, of course, $O(n^{-3})$.

Incorporating a time trend into the model by setting $d_t = (1,t)'$ in (9) yields

$$\text{Model C:} \quad y_t = \alpha + \gamma t + \beta y_{t-1} + \epsilon_t, \quad |\beta| < 1, \quad t = 1,\ldots,n. \tag{19}$$

Theorem 1 of Kiviet and Phillips (1993) provides a general expression for the bias of $\hat{\beta}$ to $O(n^{-1})$ in a model that can contain a set of regressors in addition to a constant and trend. As an aid to simulation economy it can be shown that $\hat{\beta}$ is invariant to the values of α, γ and the variance of ϵ_t provided that

$$y_0 = \frac{1}{1-\beta}\left(\alpha - \frac{\beta\gamma}{1-\beta}\right),$$

a condition that was also imposed in the simulations. It can also be shown that $\hat{\gamma}$ is invariant to the values of α and the innovation variance under this initial value condition.

The simulation results for the estimators $\hat{\beta}_{J,2}$ and $\hat{\beta}_{J,(2,3)}$ in Models B and C are contained in Table 3. Results for the associated jackknife estimators of the intercept are not reported as this parameter is typically of secondary interest and the performance of the estimators relative to the OLS estimator was dependent on the particular parameter values considered. The estimators $\hat{\gamma}_{J,2}$ and $\hat{\gamma}_{J,(2,3)}$ of the trend parameter γ in Model C are, however, also reported (the true value of $\gamma = 0.1$). For both Models B and C it can be seen from Table 3 that substantial bias reduction can be obtained using these jackknife estimators, even for β approaching the upper boundary of the stationarity region and even for small sample sizes. In Model B the estimator $\hat{\beta}_{J,(2,3)}$ produces the smallest absolute bias in 20 out of the 28 combinations of β and n considered while in Model C this drops to 16 and to 13 for $\hat{\gamma}_{J,(2,3)}$. Nevertheless both estimators can be judged to perform well in terms of bias reduction.

Although impressive the bias results in Table 3 using $m = 2$ and $M = (2,3)$ do not necessarily correspond to the minimum bias possible across different choices of m and M. Theorem 4 implies that keeping these parameters fixed means that the limiting distribution of the jackknife estimators will be different to that of $\hat{\beta}$, in fact differing by an $O_p(1)$ random variable. Even if bias reduction is the sole objective it would seem possible to improve the performance of these estimators even further if m and M are allowed to increase with n and, indeed, this seems important if the limiting distribution of $\hat{\beta}$ is to be used for inference with the jackknife estimators. In the present case with stationary autoregressions the limiting distribution of $\sqrt{n}(\hat{\beta} - \beta)$ is Gaussian and in order for the jackknife estimators to share this property it is necessary for $(1/\ell) + (\ell/n) \to 0$ as $n \to 0$, which, in the case of $\hat{\beta}_{J,m}$, requires $(1/m) + (m/n) \to 0$ as $n \to 0$. This is certainly satisfied if $m = \delta_0 n^{\delta_1}$ with $\delta_0 > 0$ and $0 < \delta_1 < 1$ and so regressions were run using the bias-minimising values of m from the experiments reported in Table 3 to gain some insight as to whether an approximate rule-of-thumb can be derived that simultaneously minimises bias and satisfies the conditions in Theorem 4(b). In fact, the regressions also allowed m to depend on β and so were of the form

$$\ln m = \ln \delta_0 + \delta_1 \ln n + \delta_2 \ln \beta + u$$

where u denotes a random disturbance. The following results were obtained (where figures in parentheses denote standard errors and $\hat{\sigma}$ denotes the estimated standard error of u):

$$\text{Model B:} \quad \ln m = \underset{(0.3837)}{-0.6573} + \underset{(0.0884)}{0.3906} \ln n + \hat{u}, \quad \bar{R}^2 = 0.41, \quad \hat{\sigma} = 0.36.$$

The marginal probability value for including $\ln \beta$ in the regression was 0.46 and so this variable was omitted. This regression implies that $m \approx 0.5182 n^{0.3906}$ and the following rule-of-thumb was used: $m = 0.5 n^{0.4}$. For Model C the results were:

$$\text{Model C:} \quad \ln m = \underset{(0.6261)}{-1.1552} + \underset{(0.1442)}{0.6086} \ln n + \underset{(0.1445)}{0.3217} \ln \beta + \hat{u}, \quad \bar{R}^2 = 0.43, \quad \hat{\sigma} = 0.59.$$

This suggests that $m \approx 0.3150 n^{0.6086} \beta^{0.3217}$ or, roughly, $m = (1/3) n^{0.6} \beta^{1/3}$. Using these bias-minimising expansion paths for m resulted in the bias properties reported in Table 4. The precise values of m were chosen as $m = \hat{m}_K$, where $\hat{m}_K$ denotes the element of the set $K = \{2, 3, 4, 6, 8, 12, 16, 24, 48\}$ that is closest to the predicted value of m from the rule-of-thumb, while the values for M were taken as $M = (m, m^+)$ where m^+ denotes the next value from the set K. The rule-of-thumb appears to work well with substantial bias reductions for the jackknife estimators clearly evident.

3.3 Higher-order autoregression

In order to examine the performance of a jackknife estimator in higher-order autoregressions the following AR(p) model was considered for $p = 2$ and $p = 4$:

$$y_t = \alpha + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \epsilon_t, \quad t = 1, \ldots, n, \tag{20}$$

where $\alpha = \mu(1 - \eta)$, $\eta = \sum_{i=1}^{p} \beta_i$, $\mu = E(y_t)$ and $y_0 = y_{-1} = \ldots = y_{-p+1} = \mu$. The parameter values, taken from Patterson (2000), are as follows. In the AR(2), $\beta_1 = 1.25$ and $\beta_2 = -0.35$, while in the AR(4) model $\beta_1 = 1.20$, $\beta_2 = -0.55$, $\beta_3 = 0.40$ and $\beta_4 = -0.15$. 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 $\eta = 0.90$ 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. Defining $\beta = (\beta_1, \ldots, \beta_p)'$, Table 5 contains the bias results for the estimator based on non-overlapping sub-samples with $m = 2$ i.e. $\hat{\beta}_{J,2}$, expressed as a percentage of the OLS bias, the actual value of which is also reported for comparison. In all cases but one, which concerns the estimation of β_2 in the AR(2) model when n=24, the jackknife estimator succeeds in reducing the bias as compared to the OLS estimator. Of particular note is the accuracy with which the jackknife estimator is capable of estimating η, the sum of the autoregressive coefficients, even when the sample size is small. For example, in the AR(4) model with $n = 48$, the bias of the jackknife estimator of η is 6% that of the OLS estimator even though the percentage of bias of the jackknife estimator of the individual coefficients ranges (in absolute terms) from 30% to 86% that of the OLS estimator. Accurate estimation of η 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).

4. Other distributional considerations

4.1 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 case of Model A Shenton and Johnson (1965) provide an expansion for the mean square error (MSE) of $\hat{\beta}$, given by

$$E\left(\hat{\beta}-\beta\right)^2 = \frac{1-\beta^2}{n} - \frac{1-14\beta^2}{n^2} + O\left(n^{-3}\right), \tag{21}$$

while the results of Bao (2007), under normality, yield

$$E\left(\hat{\beta}-\beta\right)^2 = \frac{1-\beta^2}{n} + \frac{1}{n^2}\left[14\beta^2 - 1 - \frac{(1-\beta^2)y_0^2}{\sigma^2}\right] + O\left(n^{-3}\right). \tag{22}$$

When $y_0 = 0$ these two expressions are obviously equal. Furthermore in Model B the results of Bao (2007) provide (under normality)

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

When $\alpha = (1-\beta)y_0$, as in the simulations, the effect of y_0 in the $O(n^{-2})$ term is eliminated.

Analysis of the values of m that minimise the root mean square error (RMSE) for the estimator $\hat{\beta}_{J,m}$ in Models A, B and C reveals that m is an increasing function of sample size n and tends to vary inversely with β. As in the case of bias-minimisation it is useful to try to obtain a rule-of-thumb that can be used to determine m. The following results were obtained:

Model A: $\underset{}{\ln m} = \underset{(0.4371)}{-0.7499} + \underset{(0.1007)}{0.7038}\ln n - \underset{(0.1009)}{0.4523}\ln\beta + \hat{u}, \quad \bar{R}^2 = 0.71, \quad \hat{\sigma} = 0.41;$

Model B: $\ln m = \underset{(0.2631)}{-1.0228} + \underset{(0.0606)}{0.8345}\ln n - \underset{(0.0607)}{0.2051}\ln\beta + \hat{u}, \quad \bar{R}^2 = 0.88, \quad \hat{\sigma} = 0.25;$

Model C: $\ln m = \underset{(0.2458)}{-1.8055} + \underset{(0.0566)}{0.9647}\ln n - \underset{(0.0567)}{0.1558}\ln\beta + \hat{u}, \quad \bar{R}^2 = 0.92, \quad \hat{\sigma} = 0.23.$

The following rules-of-thumb for Models A, B and C, respectively, were used based on the above results: $m = 0.5n^{0.7}\beta^{-0.5}$; $m = 0.36n^{0.8}\beta^{-0.2}$; and $m = 0.16n^{0.96}\beta^{-0.15}$. The actual values of m and M were then derived in the same way as in the bias-minimising case.

Table 6 presents the RMSEs for the estimators $\hat{\beta}_{J,m}$ and $\hat{\beta}_{J,M}$ for Models A, B and C as a ratio of the OLS RMSE using the predicted values of m and M described above. For model C the performance of the estimators $\hat{\gamma}_{J,m}$ and $\hat{\gamma}_{J,M}$ is also reported. For all three models the RMSE of the jackknife estimators is a decreasing function of n and β and the RMSE of $\hat{\beta}_{J,M}$ is larger than that of $\hat{\beta}_{J,m}$ in the majority of cases. The jackknife RMSEs tend to be larger than the OLS RMSE for the smallest values of β but for larger values of β the ratio of the jackknife RMSE to the OLS RMSE falls to as low as 0.61 in Model C. The ratio for the estimator $\hat{\beta}_{J,M}$ is always larger than one for the smallest sample size ($n = 24$) in all three models. The fact that the jackknife estimators can achieve a smaller RMSE than the OLS estimator, even though they are intended mainly for bias reduction, is an interesting finding and may prove to have an impact on the properties of hypothesis tests using jackknife estimators. This topic warrants further exploration in future work.

4.2 Departures from normality

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 $x \sim \Gamma(a, b)$ 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 7 reports the bias of the estimators $\hat{\beta}_{J,m}$ and $\hat{\beta}_{J,M}$ in Model B using the bias-minimising rules-of-thumb for m and M described earlier. The actual bias of the OLS estimator $\hat{\beta}$ is typically smaller in absolute terms than under normality (compare the entries in Table 3) and, although the extent of bias reduction is comparable to the normal case when ϵ_t is a t-variate, the skewness introduced when ϵ_t is a Gamma variate typically has a small but noticeable negative impact on the extent of bias reduction. Nevertheless the jackknife estimators are capable of producing impressive reductions in bias, compared to OLS, even under non-normality.

4.3 ARCH effects

It is also of interest to examine the performance of the jackknife estimators in the AR(1) model subject to ARCH disturbances. In this case $\epsilon_t = h_t v_t$ where v_t is i.i.d. $N(0, 1)$ and

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

where σ^2 denotes the unconditional variance and ϕ is the ARCH parameter. Two values of

ϕ are considered, $\phi = 0.5$ and $\phi = 0.9$ 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 $\mu_4 = 3(1-\phi^2)/(1-3\phi^2)$, 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 biases of $\hat{\beta}_{J,m}$ and $\hat{\beta}_{J,M}$ relative to the OLS bias in Model B with ARCH disturbances are contained in Table 8. 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. Nevertheless the jackknife estimators do offer substantial improvements over OLS in terms of bias reduction.

4.3 The median

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

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

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 9 reports the percentage of the distribution of the estimators $\hat{\beta}$, $\hat{\beta}_{J,m}$ and $\hat{\beta}_{J,M}$ for which the bias is negative in Model B when subject to normal and gamma disturbances and for the values of m and M that minimise bias and RMSE using the rules-of-thumb described earlier. The OLS estimator suffers from substantial skewness especially for larger values of β but this is, to a large extent, eliminated by the jackknife estimators. The improvements are particularly striking for larger values of β for which the distribution of the OLS estimator $\hat{\beta}$ is severely skewed. The skewness of the distributions is also more severe when the disturbances are generated by the Gamma distribution than the normal.

5. Conclusions

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, rules-of-thumb for which are provided. The jackknife estimators also outperform OLS 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.

Other methods of bias reduction are, of course, possible in addition to the jackknife. An investigation of three such methods in the context of autoregressive models was carried out by Patterson (2007), the three methods being first-order bias correction (FOBC), based on the theoretical expansion of the OLS estimator, the bootstrap (BS), and recursive mean adjustment (RMA). In simulations with the AR(1) model he finds that the bootstrap achieves greatest bias reduction, recursive mean adjustment results in the smallest MSE, while first-order bias correction produces the most accurate confidence intervals. Although based on a different set of simulated random variables it is nevertheless of some interest to compare the results for bias and RMSE in Patterson (2007) with the results for the jackknife obtained here. Table 10 therefore presents the results from Patterson (2007, Table 2) for these three estimators and for the jackknife estimator $\hat{\beta}_{J,2}$, denoted JK(2), for the two common values of β, namely 0.90 and 0.99. The sample size in Patterson (2007) is $n = 100$, and so the jackknife results refer to $n = 96$, this being the nearest sample size used here. In terms of bias reduction the jackknife performs well, producing smaller bias then even the bootstrap in three of the four cases. But its larger sampling variability leads to the larger RMSE values reported in Table 10, although it must be recognised that the results refer to $m = 2$ which is not necessarily the value that minimises RMSE.

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 (2010), and further work (in progress) is examining the performance of the bias-reduced jackknife estimators in actually testing for unit roots.

Appendix

Proof of Theorem 1. Let S_J be of the generic form $S_J = k_{1n}S_n + k_{2n}(1/m)\sum_{i=1}^m S_i$ and note that

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

Then it follows from the above and (1) that

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

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 $k_{2n} = -\ell/(n-\ell)$. □

Proof of Theorem 2. First note that

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

It follows that

$$\begin{aligned} E(S_J) &= w_n\left(S + \frac{a_1}{n} + \frac{a_2}{n^2}\right) + w_{1n}\left(S + \frac{a_1}{\ell_1} + \frac{a_2}{\ell_1^2}\right) + w_{2n}\left(S + \frac{a_1}{\ell_2} + \frac{a_2}{\ell_2^2}\right) + r_n \\ &= (w_n + w_{1n} + w_{2n})S + a_1\left(\frac{w_n}{n} + \frac{w_{1n}}{\ell_1} + \frac{w_{2n}}{\ell_2}\right) + a_2\left(\frac{w_n}{n^2} + \frac{w_{1n}}{\ell_1^2} + \frac{w_{2n}}{\ell_2^2}\right) + r_n \end{aligned}$$

where $r_n = O(n^{-3})$. In order that $E(S_J) = S + O(n^{-3})$ it is therefore necessary for the following three conditions to be satisfied:

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

Solving these conditions yields the weights specified in the Theorem. □

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

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

To eliminate the first-order bias and to have $E(S_J) = S$ 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. □

Proof of Theorem 4. (a) From the definition of S_J in Theorem 1 it follows that

$$\begin{aligned} |S_J - S| &= \left|\left(\frac{n}{n-\ell}\right)(S_n - S) - \left(\frac{\ell}{n-\ell}\right)\frac{1}{m}\sum_{i=1}^m (S_i - S)\right| \\ &\leq \left(\frac{n}{n-\ell}\right)|S_n - S| + \left(\frac{\ell}{n-\ell}\right)\left|\frac{1}{m}\sum_{i=1}^m (S_i - S)\right|. \end{aligned} \tag{24}$$

With the assumption that $\ell = O(n)$ then

$$\frac{n}{n-\ell} = \left(1 - \frac{\ell}{n}\right)^{-1} = O(1), \quad \frac{\ell}{n-\ell} = \frac{\ell}{n}\left(1 - \frac{\ell}{n}\right)^{-1} = O(1).$$

(i) When m is fixed then $S_J - S = o_p(1)$ because $S_n - S = o_p(1)$ and $S_i - S = o_p(1)$ $(i = 1, \ldots, m)$ so that both components in (24) are $o_p(1)$.
(ii) When m increases with n then clearly the component involving S_n in (24) remains $o_p(1)$ while

$$\left|\frac{1}{m}\sum_{i=1}^{m}(S_i - S)\right| \leq \left|\frac{1}{m}\sum_{i=1}^{m}(S_i - E(S_i))\right| + \left|\frac{1}{m}\sum_{i=1}^{m}(E(S_i) - S)\right|.$$

The first component converges to zero in probability by assumption while the second is $O(n^{-1}) = o_p(1)$ because of (1) in Theorem 1 which applies to the sub-sample statistics as well as to S_n. Hence $S_J - S = o_p(1)$ as required.
(b) This follows using a result from Rao (1973, p.122, 2c.4.12) which shows that, for random variables X_n, Y_n and Y, if $|X_n - Y_n| \xrightarrow{p} 0$ and $Y_n \xrightarrow{d} Y$ then $X_n \xrightarrow{d} Y$. Here $X_n = \sqrt{n}(S_J - S)$ and $Y_n = \sqrt{n}(S_n - S)$ for which it is assumed that $Y_n \xrightarrow{d} Z$. The sub-sample statistics also satisfy $\sqrt{\ell}(S_i - S) = O_p(1)$ $(i = 1, \ldots, m)$. Note that $X_n - Y_n$ can be written

$$\begin{aligned} X_n - Y_n &= \left(\frac{n}{n-\ell}\right) Y_n - Y_n - \left(\frac{\ell}{n-\ell}\right)\frac{1}{m}\sum_{i=1}^{m}\sqrt{n}(S_i - S) \\ &= a_n Y_n - b_{n,m}\sum_{i=1}^{m}\sqrt{\ell}(S_i - S). \end{aligned} \tag{25}$$

where $a_n = \ell/(n-\ell)$ and

$$b_{n,m} = \left(\frac{\ell}{n-\ell}\right)\frac{\sqrt{n}}{m\sqrt{\ell}} = \frac{1}{m}\left(\frac{\ell}{n}\right)^{1/2}\left(1 - \frac{\ell}{n}\right)^{-1}.$$

When $\ell = O(n)$ it follows that $a_n = O(1)$, $b_{n,m} = O(1)$ and $X_n - Y_n = O_p(1)$ so that Rao's result does not apply, however the stated result holds because $a_n \to 1$ as $n \to \infty$ and the term involving the normalised sub-sample statistics is a finite sum and is $O_p(1)$. When $(1/\ell) + (\ell/n) \to 0$ as $n \to \infty$ then $a_n = o(1)$ and $b_{n,m} = o(m^{-1})$ so that

$$|X_n - Y_n| \leq a_n|Y_n| + b_{n,m}\left|\sum_{i=1}^{m}\sqrt{\ell}(S_i - S)\right| = o(1)O_p(1) + o(m^{-1})O_p(m) = o_p(1),$$

recognising that m may also increase with n. Application of Rao's result concludes the proof. □

Proof of Corollary to Theorem 4. (a) First note that

$$\sqrt{n}\,(S_J - S) = \left(\frac{n}{n-\ell}\right)\sqrt{n}\,(S_n - S) - \left(\frac{\ell}{n-\ell}\right)\frac{1}{m}\frac{\sqrt{n}}{\sqrt{\ell}}\sum_{i=1}^{m}\sqrt{\ell}\,(S_i - S)\,.$$

The result follows straightforwardly from this expression by taking the appropriate limits.
(b) Under normality the limit established in (a) is a linear combination of normal random variables with zero means and hence also has a mean of zero, so it remains to determine the variance which follows from the linear combination of random variables and the stated covariances. □

References

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

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.

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., 2010. Jackknife bias reduction in the presence of a unit root. In preparation.

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.

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.

Iglesias, E.M., Phillips, G.D.A., 2005. Bivariate ARCH models: Finite sample properties of QML estimators and an application to an LM-type test. Econometric Theory 21, 1058–1086.

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

Kiviet, J.F., Phillips, G.D.A., 1993. Alternative bias approximations in regressions with a lagged dependent variable. Econometric Theory 9, 62–80.

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.

Patterson, K.D., 2007. Bias reduction through first-order mean correction, bootstrapping and recursive mean adjustment. Journal of Applied Statistics, 34, 23–45.

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.

Rao, C.R., 1973. Linear Statistical Inference and Its Applications (2nd Edition). Wiley, New York.

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.

Shenton, L.R., Johnson, W.I., 1965. Moments of a serial correlation coefficient. Journal of the Royal Statistical Society Series B, 27, 308–320.

Stine, R.A., Shaman, P., 1989. A fixed point characterisation for bias of autoregressive estimators. Annals of Statistics 17, 1275–1284.

Sun, Y., Phillips, P.C.B., Jin, S., 2008. Optimal bandwidth selection in heteroskedasti-city-autocorrelation robust testing. Econometrica 76, 175–194.

Tanaka, K., 1983. Asymptotic expansions associated with the AR(1) model with unknown mean. Econometrica 51, 1221–1231.

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 in Model A

β	n	$\hat{\beta}$	$\hat{\beta}_{J,2}$	$\hat{\beta}_{J,3}$	$\hat{\beta}_{J,4}$	$\hat{\beta}_{J,6}$	$\hat{\beta}_{J,8}$
0.10	24	−0.0086	**−0.0029**	−0.0037	−0.0042	−0.0051	−0.0060
	48	−0.0045	**−0.0009**	−0.0011	−0.0014	−0.0018	−0.0021
	96	−0.0023	**−0.0004**	−0.0004	−0.0005	−0.0006	−0.0007
	192	−0.0014	**−0.0004**	−0.0004	−0.0004	−0.0004	−0.0005
0.30	24	−0.0238	**−0.0064**	−0.0086	−0.0104	−0.0134	−0.0160
	48	−0.0124	**−0.0018**	−0.0025	−0.0031	−0.0042	−0.0052
	96	−0.0064	**−0.0006**	−0.0008	−0.0010	−0.0013	−0.0016
	192	−0.0035	**−0.0005**	−0.0005	−0.0006	−0.0007	−0.0007
0.50	24	−0.0387	**−0.0104**	−0.0142	−0.0173	−0.0224	−0.0266
	48	−0.0202	**−0.0029**	−0.0041	−0.0051	−0.0070	−0.0086
	96	−0.0106	**−0.0010**	−0.0013	−0.0017	−0.0023	−0.0028
	192	−0.0056	**−0.0006**	−0.0007	−0.0008	−0.0009	−0.0011
0.70	24	−0.0533	**−0.0159**	−0.0214	−0.0260	−0.0329	−0.0384
	48	−0.0280	**−0.0046**	−0.0063	−0.0079	−0.0107	−0.0130
	96	−0.0147	**−0.0015**	−0.0021	−0.0026	−0.0035	−0.0043
	192	−0.0076	**−0.0007**	−0.0008	−0.0010	−0.0012	−0.0015
0.90	24	−0.0661	**−0.0271**	−0.0341	−0.0391	−0.0467	−0.0523
	48	−0.0353	**−0.0090**	−0.0119	−0.0141	−0.0177	−0.0204
	96	−0.0185	**−0.0029**	−0.0038	−0.0047	−0.0062	−0.0074
	192	−0.0095	**−0.0009**	−0.0012	−0.0015	−0.0020	−0.0024
0.95	24	−0.0677	**−0.0313**	−0.0382	−0.0430	−0.0502	−0.0554
	48	−0.0365	**−0.0122**	−0.0151	−0.0173	−0.0209	−0.0234
	96	−0.0192	**−0.0042**	−0.0054	−0.0065	−0.0081	−0.0093
	192	−0.0098	**−0.0012**	−0.0017	−0.0021	−0.0028	−0.0033
0.99	24	−0.0670	**−0.0338**	−0.0403	−0.0447	−0.0515	−0.0563
	48	−0.0358	**−0.0152**	−0.0179	−0.0200	−0.0231	−0.0252
	96	−0.0189	**−0.0068**	−0.0081	−0.0091	−0.0105	−0.0116
	192	−0.0098	**−0.0028**	−0.0035	−0.0039	−0.0046	−0.0052

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

Table 2
Bias of jackknife estimators as ratio of OLS bias in Model A

β	n	$\hat{\beta}$ Bias	$\hat{\beta}_{J,2}$	$\hat{\beta}_{J,2}^{\mathrm{MB}}$	$\hat{\beta}_{J,2}^{\mathrm{MB2}}$	$\hat{\beta}_{J,(2,3)}$
0.10	24	−0.0086	0.34	0.41	0.37	**0.16**
	48	−0.0045	0.20	0.29	0.25	**0.09**
	96	−0.0023	0.16	0.22	0.18	**0.14**
	192	−0.0014	0.29	0.41	0.38	**0.29**
0.30	24	−0.0238	0.27	0.33	0.30	**0.08**
	48	−0.0124	0.15	0.20	0.17	**0.04**
	96	−0.0064	0.10	0.14	0.12	**0.05**
	192	−0.0035	0.14	0.18	0.17	**0.13**
0.50	24	−0.0387	0.27	0.33	0.30	**0.08**
	48	−0.0202	0.15	0.19	0.16	**0.03**
	96	−0.0106	0.10	0.13	0.11	**0.04**
	192	−0.0056	0.11	0.13	0.12	**0.08**
0.70	24	−0.0533	0.30	0.37	0.33	**0.09**
	48	−0.0280	0.16	0.22	0.18	**0.04**
	96	−0.0147	0.11	0.14	0.12	**0.04**
	192	−0.0076	0.09	0.11	0.10	**0.06**
0.90	24	−0.0661	0.41	0.49	0.45	**0.20**
	48	−0.0353	0.26	0.34	0.29	**0.09**
	96	−0.0185	0.15	0.22	0.18	**0.05**
	192	−0.0095	0.09	0.14	0.11	**0.02**
0.95	24	−0.0677	0.46	0.54	0.50	**0.26**
	48	−0.0365	0.33	0.42	0.38	**0.17**
	96	−0.0192	0.22	0.30	0.25	**0.09**
	192	−0.0098	0.13	0.19	0.16	**0.03**
0.99	24	−0.0670	0.51	0.58	0.54	**0.31**
	48	−0.0358	0.42	0.51	0.47	**0.27**
	96	−0.0189	0.36	0.46	0.41	**0.23**
	192	−0.0098	0.29	0.39	0.33	**0.17**

Entries in bold denote the minimum ratio in each row.

Table 3
Bias of jackknife estimators as ratio of OLS bias in Models B and C

	Model B			Model C					
n	$\hat{\beta}$ Bias	$\hat{\beta}_{J,2}$	$\hat{\beta}_{J,(2,3)}$	$\hat{\beta}$ Bias	$\hat{\beta}_{J,2}$	$\hat{\beta}_{J,(2,3)}$	$\hat{\gamma}$ Bias	$\hat{\gamma}_{J,2}$	$\hat{\gamma}_{J,(2,3)}$
$\beta = 0.10$									
24	−0.0568	0.05	**0.03**	−0.1045	0.03	**0.01**	0.0117	**0.02**	−0.08
48	−0.0281	0.04	**0.03**	−0.0514	0.01	**0.00**	0.0057	0.01	**0.01**
96	−0.0139	0.02	**0.01**	−0.0255	**0.02**	0.03	0.0028	**0.01**	0.01
192	−0.0072	**0.06**	0.07	−0.0129	0.02	**0.02**	0.0014	0.02	**0.02**
$\beta = 0.30$									
24	−0.0820	0.08	**0.01**	−0.1401	0.05	**−0.01**	0.0201	**0.04**	−0.07
48	−0.0406	0.04	**0.01**	−0.0688	**0.01**	−0.01	0.0098	**0.00**	−0.01
96	−0.0202	0.02	**0.00**	−0.0342	**0.01**	0.02	0.0049	**0.01**	0.01
192	−0.0103	**0.05**	0.05	−0.0172	0.02	**0.01**	0.0024	0.02	**0.01**
$\beta = 0.50$									
24	−0.1091	0.09	**−0.01**	−0.1801	0.06	**−0.04**	0.0361	**0.05**	−0.07
48	−0.0537	0.03	**0.00**	−0.0877	**−0.01**	−0.03	0.0175	**−0.01**	−0.03
96	−0.0267	0.01	**0.00**	−0.0433	**0.00**	0.01	0.0087	**−0.01**	0.01
192	−0.0136	**0.04**	0.04	−0.0216	**0.00**	0.01	0.0043	**0.00**	0.01
$\beta = 0.70$									
24	−0.1409	0.11	**−0.03**	−0.2289	0.08	**−0.05**	0.0764	0.08	**−0.07**
48	−0.0686	**0.02**	−0.03	−0.1105	**−0.03**	−0.07	0.0368	**−0.03**	−0.07
96	−0.0338	**0.00**	−0.01	−0.0537	−0.04	**−0.01**	0.0179	−0.04	**−0.01**
192	−0.0169	**0.02**	0.03	−0.0263	−0.03	**0.00**	0.0088	−0.03	**−0.01**
$\beta = 0.90$									
24	−0.1856	0.21	**0.06**	−0.2995	0.14	**−0.02**	0.2994	0.14	**−0.03**
48	−0.0913	0.06	**−0.03**	−0.1480	**−0.01**	−0.09	0.1480	**−0.01**	−0.09
96	−0.0435	**−0.03**	−0.06	−0.0704	−0.08	**−0.08**	0.0704	−0.08	**−0.08**
192	−0.0209	0.03	**−0.01**	−0.0329	−0.11	**−0.06**	0.0329	−0.11	**−0.06**
$\beta = 0.95$									
24	−0.1976	0.23	**0.10**	−0.3265	0.18	**0.01**	0.6528	0.18	**0.01**
48	−0.1013	0.12	**0.04**	−0.1651	**0.03**	−0.08	0.3303	**0.03**	−0.08
96	−0.0487	**0.00**	−0.05	−0.0796	**−0.06**	−0.09	0.1593	**−0.07**	−0.09
192	−0.0229	−0.05	**−0.04**	−0.0369	−0.13	**−0.10**	0.0738	−0.13	**−0.10**
$\beta = 0.99$									
24	−0.2002	0.21	**0.07**	−0.3568	0.23	**0.07**	3.5677	0.23	**0.07**
48	−0.1069	0.14	**0.05**	−0.1880	0.10	**−0.02**	1.8803	0.10	**−0.01**
96	−0.0554	0.09	**0.05**	−0.0947	**0.01**	−0.06	0.9466	**0.01**	−0.06
192	−0.0275	0.05	**0.02**	−0.0459	**−0.06**	−0.08	0.4587	**−0.06**	−0.08

Entries in bold denote the minimum (absolute) ratio for each pair of estimators in each row.

Table 4
Bias of jackknife estimators as ratio of OLS bias in Models B and C using bias-minimising expansion rates for m and M

	Model B		Model C			
n	$\hat{\beta}_{J,m}$	$\hat{\beta}_{J,M}$	$\hat{\beta}_{J,m}$	$\hat{\beta}_{J,M}$	$\hat{\gamma}_{J,m}$	$\hat{\gamma}_{J,M}$
$\beta = 0.10$						
24	0.05	**0.03**	0.03	**0.01**	**0.02**	−0.08
48	0.04	**0.03**	0.01	**0.00**	0.01	**0.01**
96	0.02	**0.00**	**0.02**	0.03	**0.01**	0.01
192	0.06	**0.05**	0.03	**0.02**	0.03	**0.02**
$\beta = 0.30$						
24	0.08	**0.01**	0.05	**−0.01**	**0.04**	−0.07
48	0.04	**0.01**	**0.01**	−0.01	**0.00**	−0.01
96	0.03	**0.00**	**0.01**	0.02	**0.00**	0.01
192	0.05	**0.04**	0.02	**0.01**	0.01	**0.01**
$\beta = 0.50$						
24	0.09	**−0.01**	0.06	**−0.04**	**0.05**	−0.07
48	0.03	**0.00**	**−0.01**	−0.03	**−0.01**	−0.03
96	0.02	**−0.01**	**−0.01**	−0.02	**−0.01**	−0.02
192	0.04	**0.03**	−0.01	**0.00**	−0.01	**−0.01**
$\beta = 0.70$						
24	0.11	**−0.03**	0.08	**−0.05**	0.08	**−0.07**
48	**0.02**	−0.03	**−0.01**	−0.07	**−0.01**	−0.07
96	**0.00**	−0.03	**−0.05**	−0.07	**−0.05**	−0.07
192	0.01	**0.01**	−0.06	**−0.05**	−0.06	**−0.05**
$\beta = 0.90$						
24	0.21	**0.06**	0.14	**−0.02**	0.14	**−0.03**
48	0.06	**−0.03**	**0.04**	−0.09	**0.04**	−0.09
96	**−0.02**	−0.08	**−0.07**	−0.15	**−0.07**	−0.15
192	**−0.05**	−0.07	**−0.14**	−0.22	**−0.14**	−0.22
$\beta = 0.95$						
24	0.23	**0.10**	0.18	**0.01**	0.18	**0.01**
48	0.12	**0.04**	0.08	**−0.05**	0.08	**−0.05**
96	**0.03**	−0.05	**0.02**	−0.14	**0.02**	−0.13
192	**−0.05**	−0.09	**−0.12**	−0.22	**−0.12**	−0.22
$\beta = 0.99$						
24	0.21	**0.07**	0.23	**0.07**	0.23	**0.07**
48	0.14	**0.05**	0.16	**0.04**	0.16	**0.04**
96	0.11	**0.04**	0.14	**−0.02**	0.14	**−0.02**
192	0.07	**0.03**	**0.03**	−0.09	**0.03**	−0.09

Entries in bold denote the minimum (absolute) ratio for each pair of estimators in each row.

Table 5
Bias of jackknife estimator as ratio of OLS bias in AR(2) and AR(4) models

	AR(2)		AR(4)	
	$\hat{\beta}$ Bias	$\hat{\beta}_{J,2}$	$\hat{\beta}$ Bias	$\hat{\beta}_{J,2}$
$n = 24$				
β_1	−0.1126	−0.34	−0.1278	−0.33
β_2	−0.0295	1.32	0.0101	−0.71
β_3			−0.0279	−0.27
β_4			−0.0520	0.42
η	−0.1421	0.00	−0.1976	−0.11
$n = 48$				
β_1	−0.0482	−0.30	−0.0559	−0.30
β_2	−0.0156	0.60	0.0045	−0.86
β_3			−0.0116	−0.54
β_4			−0.0291	0.48
η	−0.0639	−0.08	−0.0921	−0.06
$n = 96$				
β_1	−0.0221	−0.19	−0.0251	−0.21
β_2	−0.0075	0.19	0.0020	−0.37
β_3			−0.0055	−0.24
β_4			−0.0135	0.13
η	−0.0295	−0.09	−0.0421	−0.10
$n = 192$				
β_1	−0.0107	−0.07	−0.0119	−0.10
β_2	−0.0033	0.01	0.0012	0.09
β_3			−0.0030	0.08
β_4			−0.0059	−0.06
η	−0.0141	−0.05	−0.0197	−0.07

Table 6
RMSE of jackknife estimators as ratio of OLS RMSE in Models A, B and C using RMSE-minimising expansion rates for m and M

	Model A		Model B		Model C			
n	$\hat{\beta}_{J,m}$	$\hat{\beta}_{J,M}$	$\hat{\beta}_{J,m}$	$\hat{\beta}_{J,M}$	$\hat{\beta}_{J,m}$	$\hat{\beta}_{J,M}$	$\hat{\gamma}_{J,m}$	$\hat{\gamma}_{J,M}$
$\beta = 0.10$								
24	**1.05**	1.29	**1.20**	2.32	**1.06**	1.65	**1.57**	6.33
48	**1.02**	1.07	**1.03**	1.09	**1.02**	1.17	**1.27**	3.89
96	**1.01**	1.02	**1.01**	1.03	**1.01**	1.05	**1.09**	2.26
192	**1.01**	1.01	**1.01**	1.01	**1.00**	1.01	**1.11**	1.16
$\beta = 0.30$								
24	**1.04**	1.29	**1.03**	1.31	**1.00**	1.63	**1.45**	5.60
48	**1.02**	1.05	**1.01**	1.09	**0.98**	1.16	**1.18**	3.33
96	**1.01**	1.02	**1.01**	1.03	**0.98**	1.04	**1.05**	1.96
192	**1.00**	1.01	**1.00**	1.01	**0.99**	1.00	**1.05**	1.10
$\beta = 0.50$								
24	**1.03**	1.16	**0.99**	1.30	**0.93**	1.60	**1.26**	4.48
48	**1.01**	1.04	**0.99**	1.16	**0.93**	1.14	**1.06**	2.64
96	**1.01**	1.02	**0.99**	1.04	**0.95**	1.02	**0.99**	1.64
192	**1.00**	1.01	**0.99**	1.00	**0.97**	0.99	**0.98**	1.05
$\beta = 0.70$								
24	**1.00**	1.15	**0.94**	1.26	**0.84**	1.53	**1.01**	3.00
48	**1.00**	1.07	**0.93**	1.17	**0.86**	1.11	**0.92**	1.59
96	**0.99**	1.01	**0.95**	1.03	**0.89**	0.98	**0.91**	1.16
192	**1.00**	1.00	**0.97**	0.99	**0.93**	0.96	**0.93**	1.02
$\beta = 0.90$								
24	**0.97**	1.13	**0.84**	1.14	**0.78**	1.38	**0.81**	1.60
48	**0.95**	0.97	**0.82**	0.93	**0.72**	0.98	**0.73**	1.08
96	**0.95**	0.98	**0.85**	0.89	**0.75**	0.86	**0.75**	0.91
192	0.97	**0.97**	**0.89**	0.90	**0.80**	0.84	**0.80**	0.84
$\beta = 0.95$								
24	**0.96**	1.12	**0.82**	1.11	**0.76**	1.32	**0.73**	1.38
48	**0.93**	0.95	**0.78**	0.87	**0.68**	0.91	**0.68**	0.93
96	**0.93**	0.94	**0.79**	0.83	**0.68**	0.79	**0.68**	0.80
192	0.95	**0.94**	**0.83**	0.84	**0.72**	0.75	**0.72**	0.75
$\beta = 0.99$								
24	**0.96**	1.13	**0.80**	1.10	**0.73**	1.25	**0.73**	1.25
48	**0.92**	0.93	**0.74**	0.84	**0.65**	0.84	**0.65**	0.84
96	0.91	**0.91**	**0.72**	0.75	**0.62**	0.69	**0.62**	0.69
192	0.90	**0.89**	0.72	**0.71**	0.62	**0.61**	0.62	**0.61**

Entries in bold denote the minimum (absolute) ratio for each pair of estimators in each row.

Table 7

Bias of jackknife estimators as ratio of OLS bias in Model B under non-normality

		$\epsilon_t \sim t_5$			$\epsilon_t \sim \Gamma\left(1, \sqrt{5/3}\right)$		
β	n	$\hat{\beta}$ Bias	$\hat{\beta}_{J,m}$	$\hat{\beta}_{J,M}$	$\hat{\beta}$ Bias	$\hat{\beta}_{J,m}$	$\hat{\beta}_{J,M}$
0.10	24	−0.0546	0.01	**−0.01**	−0.0554	**0.05**	0.06
	48	−0.0272	**0.01**	−0.01	−0.0277	0.04	**−0.02**
	96	−0.0136	0.02	**0.01**	−0.0140	0.06	**−0.04**
	192	−0.0068	0.01	**0.01**	−0.0070	0.05	**0.03**
0.30	24	−0.0790	0.05	**−0.01**	−0.0783	0.07	**0.04**
	48	−0.0393	0.02	**−0.01**	−0.0394	0.05	**0.02**
	96	−0.0198	0.03	**0.01**	−0.0201	0.07	**0.04**
	192	−0.0098	0.01	**0.00**	−0.0100	0.06	**0.03**
0.50	24	−0.1053	0.07	**−0.03**	−0.1031	0.08	**0.01**
	48	−0.0520	**0.02**	−0.02	−0.0516	0.05	**0.01**
	96	−0.0260	0.02	**−0.01**	−0.0262	0.06	**0.03**
	192	−0.0129	**0.01**	−0.01	−0.0131	0.05	**0.03**
0.70	24	−0.1365	0.09	**−0.04**	−0.1329	0.11	**0.00**
	48	−0.0666	**0.01**	−0.04	−0.0656	0.04	**−0.01**
	96	−0.0328	**0.00**	−0.04	−0.0327	0.03	**0.01**
	192	−0.0162	**−0.01**	−0.03	−0.0163	0.03	**0.02**
0.90	24	−0.1816	0.20	**0.06**	−0.1765	0.22	**0.09**
	48	−0.0893	0.05	**−0.04**	−0.0879	0.08	**−0.01**
	96	−0.0425	**−0.02**	−0.08	−0.0422	**0.00**	−0.07
	192	−0.0204	**−0.05**	−0.08	−0.0203	**−0.03**	−0.05
0.95	24	−0.1943	0.23	**0.09**	−0.1887	0.25	**0.12**
	48	−0.0996	0.12	**0.03**	−0.0980	0.14	**0.05**
	96	−0.0478	**0.03**	−0.04	−0.0474	0.05	**−0.04**
	192	−0.0226	**−0.05**	−0.09	−0.0224	**−0.03**	−0.08
0.99	24	−0.1975	0.22	**0.06**	−0.1924	0.24	**0.10**
	48	−0.1060	0.14	**0.06**	−0.1041	0.16	**0.06**
	96	−0.0551	0.13	**0.06**	−0.0544	0.14	**0.06**
	192	−0.0274	0.08	**0.03**	−0.0272	0.09	**0.03**

Entries in bold denote the minimum (absolute) ratio for each pair of estimators in each row.

Table 8

Bias of jackknife estimators as ratio of OLS bias in Model B with ARCH(1) disturbances

		$\phi = 0.50$			$\phi = 0.90$		
β	n	$\hat{\beta}$ Bias	$\hat{\beta}_{J,m}$	$\hat{\beta}_{J,M}$	$\hat{\beta}$ Bias	$\hat{\beta}_{J,m}$	$\hat{\beta}_{J,M}$
0.10	24	−0.0585	0.15	**0.11**	−0.0594	0.20	**0.18**
	48	−0.0314	0.17	**0.13**	−0.0336	0.27	**0.23**
	96	−0.0164	0.17	**0.15**	−0.0193	0.34	**0.29**
	192	−0.0086	0.19	**0.15**	−0.0118	0.44	**0.41**
0.30	24	−0.0907	0.22	**0.14**	−0.0967	0.31	**0.26**
	48	−0.0498	0.22	**0.15**	−0.0583	0.38	**0.32**
	96	−0.0269	0.23	**0.18**	−0.0365	0.47	**0.41**
	192	−0.0146	0.24	**0.20**	−0.0242	0.56	**0.52**
0.50	24	−0.1229	0.23	**0.13**	−0.1329	0.34	**0.26**
	48	−0.0670	0.21	**0.14**	−0.0803	0.39	**0.33**
	96	−0.0362	0.22	**0.15**	−0.0509	0.48	**0.43**
	192	−0.0198	0.24	**0.17**	−0.0342	0.58	**0.53**
0.70	24	−0.1567	0.22	**0.09**	−0.1690	0.32	**0.21**
	48	−0.0827	0.17	**0.09**	−0.0981	0.34	**0.29**
	96	−0.0435	0.18	**0.11**	−0.0602	0.44	**0.39**
	192	−0.0234	0.20	**0.14**	−0.0397	0.54	**0.50**
0.90	24	−0.2008	0.27	**0.11**	−0.2168	0.35	**0.20**
	48	−0.1014	0.12	**0.02**	−0.1148	0.25	**0.15**
	96	−0.0494	0.06	**−0.01**	−0.0621	0.27	**0.20**
	192	−0.0246	0.05	**0.01**	−0.0363	0.36	**0.32**
0.95	24	−0.2130	0.28	**0.13**	−0.2338	0.37	**0.23**
	48	−0.1100	0.15	**0.05**	−0.1234	0.25	**0.14**
	96	−0.0529	0.06	**−0.03**	−0.0631	0.21	**0.12**
	192	−0.0253	**−0.01**	−0.06	−0.0336	0.24	**0.18**
0.99	24	−0.2149	0.24	**0.08**	−0.2393	0.34	**0.20**
	48	−0.1144	0.13	**0.03**	−0.1308	0.24	**0.12**
	96	−0.0583	0.10	**0.03**	−0.0676	0.20	**0.11**
	192	−0.0288	0.05	**0.00**	−0.0340	0.16	**0.08**

Entries in bold denote the minimum (absolute) ratio for each pair of estimators in each row.

Table 9

Percentage of negative bias in Model B

	$\epsilon_t \sim N(0,1)$					$\epsilon_t \sim \Gamma(0.25,2)$				
n	$\hat{\beta}$	$\hat{\beta}_{J,m_b}$	$\hat{\beta}_{J,m_r}$	$\hat{\beta}_{J,M_b}$	$\hat{\beta}_{J,M_r}$	$\hat{\beta}$	$\hat{\beta}_{J,m_b}$	$\hat{\beta}_{J,m_r}$	$\hat{\beta}_{J,M_b}$	$\hat{\beta}_{J,M_r}$
$\beta = 0.10$										
24	60	51	51	54	51	68	59	50	57	55
48	57	50	50	52	50	65	59	54	58	56
96	55	50	50	50	50	61	58	53	57	56
192	54	50	50	50	50	59	56	53	56	57
$\beta = 0.30$										
24	64	51	51	55	51	71	57	53	57	56
48	60	50	51	52	49	66	57	53	58	56
96	57	49	50	49	49	63	56	55	56	56
192	55	49	50	49	49	60	55	53	55	56
$\beta = 0.50$										
24	69	51	52	55	50	74	56	53	57	54
48	63	49	50	53	49	68	56	54	57	53
96	59	48	49	49	48	64	54	54	54	54
192	57	49	49	49	49	61	54	54	54	54
$\beta = 0.70$										
24	76	52	52	55	50	79	55	54	56	53
48	69	49	50	54	48	72	53	52	56	51
96	64	47	49	48	46	66	52	52	52	51
192	60	48	49	48	47	63	52	52	52	51
$\beta = 0.90$										
24	90	58	58	56	52	90	59	60	56	55
48	83	52	53	55	48	84	54	56	55	51
96	75	46	50	48	45	76	49	53	49	50
192	68	45	47	47	44	69	47	51	48	48
$\beta = 0.95$										
24	93	58	60	56	53	92	59	62	56	55
48	91	55	57	56	50	91	57	61	56	53
96	83	47	53	48	46	84	50	57	49	52
192	76	44	48	46	43	76	46	52	47	48
$\beta = 0.99$										
24	95	56	60	55	52	92	57	62	56	55
48	95	55	60	55	51	94	56	65	55	54
96	94	51	57	50	51	94	53	62	52	58
192	92	47	54	49	47	92	49	60	50	54

m_b and M_b denote the bias-minimising values of m and M, respectively; m_r and M_r denote the RMSE-minimising values of m and M, respectively.

Table 10
Comparison of jackknife with other estimators in Patterson (2007) in Models B and C

	Patterson (2007)				Jackknife	
β	OLS	FOBC	BS	RMA	OLS	JK(2)
Bias						
Model B						
0.90	−0.039	−0.002	0.000	−0.006	−0.044	0.001
0.99	−0.050	−0.010	−0.013	−0.017	−0.055	−0.005
Model C						
0.90	−0.066	−0.010	−0.007	0.018	−0.070	0.006
0.99	−0.091	−0.032	−0.037	−0.013	−0.095	−0.001
RMSE						
Model B						
0.90	0.0685	0.0578	0.0592	0.0543	0.0738	0.0779
0.99	0.0668	0.0471	0.0502	0.0435	0.0725	0.0718
Model C						
0.90	0.0913	0.0667	0.0690	0.0633	0.0966	0.0961
0.99	0.1072	0.0674	0.0748	0.0525	0.1112	0.1026

Source: Patterson (2007, Table 2); Table 3; Table 6.