Economics Division

University of Southampton

Southampton SO17 1BJ, UK

Discussion Papers in
Economics and Econometrics

Title Overlapping Sub-sampling and invariance to
initial conditions

By Maria Kyriacou

No. 1203

Overlapping Sub-sampling and Invariance to Initial Conditions

Maria Kyriacou[1]

University of Southampton

April 2012

Abstract

This paper studies the use of the overlapping blocking scheme in unit root autoregression. When the underlying process is that of a random walk, the blocks' initial conditions are not fixed, but are equal to the sum of all the previous observations' error terms. When non-overlapping subsamples are used, as first shown by Chambers and Kyriacou (2010), these initial conditions do not disappear asymptotically. In this paper we show that a simple way of overcoming this issue is to use overlapping blocks. By doing so, the effect of these initial conditions vanishes asymptotically. An application of these findings to jackknife estimators indicates that an estimator based on moving-blocks is able to provide obvious reductions to the mean square error.

Keywords. moving-blocks, bias, autoregression.

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

---

[1]Economics Division, School of Social Sciences, University of Southampton, Highfield, Southampton, SO17 1BJ. Email: M.Kyriacou@soton.ac.uk, Tel: +44(0)23 8059 3645.

1 Introduction

It is well known that the original jackknife and bootstrap methods, proposed by Quenouille (1949) and Efron (1978) respectively, are not applicable to non-*iid* situations. The main approach for making such methods operational within dependent data is to block arguments and then resample from these blocks, instead from individual data. These blocks can be formed either as consecutive blocks of non-overlapping observations, as first proposed by Carlstein (1986), or as overlapping observations (moving blocks) as first used by Kuensch (1989) for stationary time series. While the primary focus of the aforementioned papers has been variance estimation, these blocking techniques can be also employed to maintain the dependence structure when the main issue is parameter estimation.

In many econometric situations of interest, non-overlapping blocks were recently found to work well in dynamic settings in which the primary objective is bias reduction. Phillips and Yu (2005) show that a jackknife estimator based on non-overlapping blocks is capable of effectively reducing the bias in a stationary bond option price application. Meanwhile, Chambers (2010) shows that in stationary autoregression, the same estimator is able to provide important reductions to the bias of the estimator of the autoregressive parameter. Much of the appeal of the non-overlapping scheme stems from the fact that in most cases it minimises the bias by using only two non-overlapping blocks. Under this sub-sampling scheme however, bias reduction can come at a cost. At the bias–minimizing number of subgroups, the estimator's variance, and evidently the mean square error, may increase. On the other hand, using overlapping blocks has the advantage that, subject to the appropriate choice of the number of blocks, it is able to balance the trade-off between bias reduction and increase

in variance.

In the presence of a unit root, the use of non-overlapping blocks is not as straightforward as in the stationary case. As Chambers and Kyriacou (2010) first illustrate, (CK hereafter), each block of length ℓ, has an initial value which is explosive $O_p(\ell^{1/2})$, the effect of which does not vanish asymptotically. As a result, in this case a jackknife estimator based on non-overlapping blocks will not work as intended.

In this paper we tackle this issue simply by constructing subgroups as overlapping blocks of observations. The contribution of the paper is twofold. First, we show that if the block's length grows at the same rate as the sample size, the effect of initial conditions vanishes asymptotically. This implies that for every block of observations, the same expansion can be used to describe the bias of the associated least squares estimator of the correlation coefficient. Second, these results motivate the use of the blocking technique for constructing jackknife estimators in unit root settings. This jackknife estimator can be, in fact, constructed in the exact same manner as in the stationary case. The moving-blocks jackknife estimator is able to provide obvious reductions to the mean square error. The results presented in the paper are based on a first order unit root autoregression but these can also apply to higher order non-stationary autoregression.

The remainder of the paper is structured as follows. Section 2 introduces the main framework and the moving blocks scheme, while Section 3 addresses the issue of the blocks' initial conditions. Section 4 summarises the asymptotic results of the moving blocks' subgroup estimators and Section 5 presents an application of these findings to jackknife estimators. Finally, Section 6 concludes. Proofs and Figures are left in the Appendix.

Throughout the paper, the symbol $\equiv$ denotes equivalence in distribution, $\rightarrow_d$ denotes convergence in distribution, $\rightarrow_p$ denotes convergence in probability, $\Rightarrow$ denotes weak convergence of the relevant probability measures and $W(r)$ denotes a Wiener process on $C[0, 1]$, the space of continuous real-valued functions on the unit interval and $\square$ denotes the end of a proof.

2 Subsampling schemes and initial conditions

Consider that the data $\{y_1, \ldots, y_n\}$ are generated by the random walk process described in (1) below.

$$y_t = \beta y_{t-1} + \epsilon_t; \quad \beta = 1, \quad \epsilon_t \sim iid\ (0, \sigma^2), \quad t = 1, \ldots, n \tag{1}$$

where y_0 can be any (observed) $O_p(1)$ random variable.

The behaviour of the Ordinary Least Squares (OLS) estimator of the autoregressive parameter, denoted by $\hat{\beta}$, has been extensively studied in the literature, with Hurwitz (1950) being the first to investigate its small sample properties. As shown in the simulation results of Evans and Savin (1981), the bias of the estimated least squares coefficient of β is evidently large for small samples. In fact, the bias is proportional to the true value of β: it vanishes as β tends towards zero and increases as the parameter reaches unity (in absolute terms). In the presence of a unit root, the bias is especially severe. As shown from the entries in Table III of Evans and Savin (1981); not only in small samples but also persists within larger samples.

As first shown by White (1958) and later extended by Phillips (1987), the limiting distribution

of $\hat{\beta}$, shown in (2), is non-standard, skewed to left and in finite sample situations, the estimator appears to be negatively biased.

$$n(\hat{\beta} - 1) \Rightarrow \frac{\frac{1}{2}[W(1)^2 - 1]}{\int_0^1 W(r)^2 dr} \tag{2}$$

In this setting, Phillips (2012) also provides the leading term of an expansion to describe the behaviour of β, under Gaussian innovations:

$$n(\hat{\beta} - 1) \equiv \frac{\frac{1}{2}(W(1)^2 - 1) - \frac{1}{\sqrt{2n}}\xi}{\int_0^1 W(r)^2 dr} + O_p(n^{-1}) \tag{3}$$

where $\xi \sim N(0, 1)$ and is independent of $W(r)$. If one takes expectations from each side of (3), an appropriate expansion for the bias of $\hat{\beta}$ is obtained, shown in (4) below.

$$E(\hat{\beta} - 1) = \frac{\mu}{n} + O(n^{-2}) \tag{4}$$

where $\mu = E\left(\frac{\frac{1}{2}(W(1)^2 - 1)}{\int_0^1 W(r)^2 dr}\right) \approx -1.781$

When dealing with dynamic settings such as in (1), sub-samples used from these realisations need to be constructed in a way so that they maintain the dependence structure of the original process. The main approach of doing so is to group observations into either consecutive non-overlapping blocks or as adjacent blocks of overlapping observations. The former method employs consecutive subgroups of non-overlapping observations while the latter uses subgroups of overlapping observations (moving blocks). Carlstein (1986) first explores the idea of employing adjacent non-overlapping sub-series to be used within resampling methods like the jackknife and the bootstrap. Under this scheme, the original sample of size n, is

grouped into a fixed number of blocks, m, each of the same length ℓ so that $n = m \times \ell$. By doing so, each observation is only used once, since each subgroup i contains observations from $(i-1)\ell + 1$ until $i \times \ell$. The use of non-overlapping blocks has recently proved to be particularly useful for constructing jackknife estimators used for bias reduction purposes. Specifically, Phillips and Yu (2005), Chambers (2010), Chambers and Kyriacou (2010) and Chen and Yu (2011) show that the bias of the OLS or ML estimators in various situations of interest can be effectively reduced when only two non-overlapping groups are employed to construct the corresponding jackknife estimators.

The moving-blocks jackknife and moving-blocks bootstrap were introduced by Kuensch (1989) for variance estimation in stationary time-series. The moving blocks technique is obtained by moving the length of observations, ℓ, across a time series by one observation each time until the last observation is reached. At block length ℓ, where $\ell = \frac{n}{m}$, M overlapping blocks are used so that $n = M + \ell - 1$. In contrast to the non-overlapping case, the moving-block technique uses almost all observations more than once. The advantage of doing so, as we shall explore more later on, is that it manages to reduce the variance of the jackknife estimators something which its non-overlapping counterpart does not achieve.

Finally, to illustrate the two sub-sampling schemes described in this section, let τ_i and κ_i denote the set of integers incorporated in each non-overlapping and overlapping block respectively:

$$\tau_i = \{(i-1)\ell + 1, \ldots, i\ell\} \quad i = 1, \ldots, m$$

$$\kappa_i = \{i, \ldots, i + \ell - 1\} \quad i = 1, \ldots, M$$

2.1 Subgroup least squares estimation

To fix ideas, let $\tilde{\beta}_i$ and $\hat{\beta}_i$ denote the least squares estimators of β which emerges from (1) using the non-overlapping and overlapping schemes respectively for a given subgroup length ℓ, with $\ell = \frac{n}{m}$:

$$\ell(\tilde{\beta}_i - 1) = \frac{\sum_{t=(i-1)\ell+1}^{i\ell} y_{t-1}\epsilon_t}{\sum_{t=(i-1)\ell+1}^{i\ell} y_{t-1}^2} \quad i = 1, \ldots, m \tag{5}$$

where $m = \frac{n}{\ell}$

$$\ell(\hat{\beta}_i - 1) = \frac{\sum_{t=i}^{i+\ell-1} y_{t-1}\epsilon_t}{\sum_{t=i}^{i+\ell-1} y_{t-1}^2} \quad i = 1, \ldots, M \tag{6}$$

where $M = n - \ell + 1$

As illustrated in Chambers and Kyriacou (2010), if the number of subgroups, m is kept fixed, then if $\ell \to \infty$ as $n \to \infty$ the limiting distribution of the non-overlapping subgroup least squares estimator $\tilde{\beta}_i$, is shown to be:

$$\ell(\tilde{\beta}_i - 1) \Rightarrow \frac{1}{m} \frac{\frac{1}{2}\left\{W(\frac{i}{m})^2 - W(\frac{i-1}{m})^2 - \frac{1}{m}\right\}}{\int_{\frac{i-1}{m}}^{\frac{i}{m}} W^2} \tag{7}$$

From (7) it becomes obvious that the limiting distribution of every subgroup i depends on both the subgroup indicator i and the number of subgroups m. Clearly, each subgroup estimator $\tilde{\beta}_i$ does not inherit the same asymptotic properties as the full-sample estimator. Of course, this only holds for the second subgroup onwards, as one can easily see that when

6

$i = 1$ the limiting distribution of $\tilde{\beta}_1$ coincides with that of the full sample estimator, shown in (2). In fact, as it will be discussed in more detail in the next section, the driving force of this issue is the sub-sample initial condition $y_{(i-1)\ell}$. For each subgroup $i \geq 2$, the initial condition is $O_p(\ell^{1/2})$ and its effect does not vanish asymptotically. Evidently, the initial condition for the fist subgroup and full sample is assumed to be $O_p(1)$.

This result also follows when deriving an asymptotic expansion for $\ell(\tilde{\beta}_i - 1)$, which is required to verify the applicability of the jackknife in this setting. Indeed jackknife estimators are built on the requirement that the same expansions can be used to describe the bias of both the full sample estimator and each subgroup estimator. Clearly, in the case of a unit root, this is not the case as it becomes obvious from (8) below.

$$\ell(\tilde{\beta}_i - 1) \equiv \frac{\int_{\frac{i-1}{m}}^{\frac{i}{m}} W \, dW}{m \int_{\frac{i-1}{m}}^{\frac{i}{m}} W^2} - \frac{\eta_i}{m^3 \sqrt{2\ell} \int_{\frac{i-1}{m}}^{\frac{i}{m}} W^2} + O_p(\ell^{-1}) \qquad i = 1, \ldots, m \tag{8}$$

where $\eta_i \sim iidN(0, \ 1)$ and independent of $W(r)$

By taking expectations from each side of (8), the $1/\sqrt{\ell}$ term disappears due to the presence of the standard normal term η_i. Thus, an appropriate expression for the bias of $\tilde{\beta}_i$ under a unit root is found by then multiplying each side of (8) with $1/\ell$.

An expansion of the bias for each non-overlapping sub-group i is given in (9) below.

$$E(\tilde{\beta}_i - 1) = \frac{\mu_i}{\ell} + O(\ell^{-2}) \tag{9}$$

i	μ_i
1	-1.781430
2	-1.138209
3	-0.931929
4	-0.814330
5	-0.734818
6	-0.676084
7	-0.630246
8	-0.593099
9	-0.562154
10	-0.535827
11	-0.513053
12	-0.493085

Table 1: Expectations for each non-overlapping subgroup i (Entries taken from Table 1 in Chambers and Kyriacou (2010))

where $\mu_i = E \left(\dfrac{\int_{\frac{i-1}{m}}^{\frac{i}{m}} W dW}{m \int_{\frac{i-1}{m}}^{\frac{i}{m}} W^2} \right)$

The expression in (9) verifies that the leading term for the bias expansion varies for each $\tilde{\beta}_i$.

Table 1 shows the values for the μ_i for each subgroup $i = 1, \ldots, m$. The entries in Table 1 confirm that these values vary substantially within different subgroups, while subgroup 1 shares the same leading term with the full sample estimator, that is $\mu = \mu_1$. These results indicate that the weights used to derive jackknife estimators using non-overlapping sub-groups in stationary autoregressive situations cannot be applied in the unit root case. Evidently, because of this issue, a jackknife estimator which utilises such subgroups will not eliminate the leading term of the original estimator's expansion as it intends to. In this paper we show that when moving blocks are used, the initial conditions issue vanishes asymptotically and the associated subgroup estimators in (6).

3 Block initial conditions

The sub-samples initial (or pre-sample) value are equal to the accumulated sum of all previous innovations and, as shown by CK, is not eliminated in the asymptotics in the non-overlapping case.

It is convenient to re-write the data generation process in (1) in terms of the initial value y_0 and the sum of innovations $\sum_{j=1}^{n} \epsilon_j$:

$$y_t = y_0 + \sum_{j=1}^{n} \epsilon_j$$

The associated data generation processes for each non-overlapping and moving block i are given by (10) and (11) respectively:

$$y_t = y_{(i-1)\ell} + \sum_{j=(i-1)\ell+1}^{t} \epsilon_j \ \ \forall \ \ t\,\{(i-1)\ell+1,\ldots,i\ell\} \tag{10}$$

$$y_t = y_{i-1} + \sum_{j=i}^{t} \epsilon_j \ \ \forall \ \ t\,\{i,\ldots,i+\ell-1\} \tag{11}$$

For each block i, the initial (or pre-sample) values are found at $t = (i-1)\ell$ and $t = i-1$ for the non-overlapping and overlapping scheme respectively:

$$y_{(i-1)\ell} = y_0 + \sum_{j=1}^{(i-1)\ell} \epsilon_j \tag{12}$$

$$y_{i-1} = y_0 + \sum_{j=1}^{i-1} \epsilon_j \tag{13}$$

As indicated by both (12) and (13), the blocks' initial values, with the mere exception of

the first subgroup, are not fixed but are of the same order as the partial sum of all previous blocks' innovation terms, which, in turn are $O_p(\ell^{-1/2})$.

4   Moving Blocks subgroup estimators: Asymptotic results

This section collectively presents the results for the asymptotic behaviour of the subgroup estimators based on the moving blocks scheme. Indeed, we show that the issue of initial conditions outlined in the previous section vanishes asymptotically as the subgroup length ℓ grows at the same rate with the sample size n. This assumption is not restrictive as this is generally used within this subsampling technique, see for example Politis, Romano and Wolf (1997). In fact, in dynamic models such as the one examined here, the blocks' length needs to be large enough to capture the dependence of the original process. The issue of the optimal block length is explored in more detail in the next section within an application of moving blocks in a jackknife estimator.

Lemma 1 and Theorem 1 show that the moving blocks estimators inherit the same distribution as their full sample counterpart, something that does not occur in the non-overlapping case.

LEMMA 1. *Let* $y_1, \ldots, y_n$ *be generated by the random walk in (1), with* y_0 *being any* $O_p(1)$ *random variable. If* M *is kept fixed, then if* $\ell \to \infty$ *as* $n \to \infty$, :

$$(a)\ \ell^{-3/2} \sum_{t=i}^{i+\ell-1} y_{t-1} \Rightarrow \sigma \int_0^1 W(r) dr \tag{14}$$

$$(b)\ \ell^{-2} \sum_{t=i}^{i+\ell-1} y_{t-1}^2 \Rightarrow \sigma^2 \int_0^1 W(r)^2 dr \tag{15}$$

$$(c) \ \ell^{-1} \sum_{t=i}^{i+\ell-1} y_{t-1} \Rightarrow \frac{\sigma^2}{2}[W(1)^2 - 1] \tag{16}$$

THEOREM 1: *Let* $y_1, \ldots, y_n$ *be generated by the random walk process in (1) and* y_0 *being any* $O_p(1)$ *random variable. If M is kept fixed and $\ell \to \infty$ as $n \to \infty$, then for every $j = 1, \ldots, M$ it follows that:*

$$\ell(\hat{\beta}_j - 1) \Rightarrow \frac{\frac{1}{2}[W(1)^2 - 1]}{\int_0^1 W(r)^2 dr} \tag{17}$$

THEOREM 2: *Let* $y_1, \ldots, y_n$ *be generated by the random walk process in (1) and that the initial condition is set at zero, $y_0 = 0$ and, in addition, that the error term ϵ_t is iid $N(0, \sigma^2)$. Then, if M is kept fixed, while $\ell \to \infty$ as $n \to \infty$:*

$$\ell(\hat{\beta}_i - 1) \equiv \frac{\frac{1}{2}(W(1)^2 - 1) - \frac{1}{\sqrt{2l}}\xi_i}{\int_0^1 W(r)^2 dr} + O_p(\ell^{-1})$$

where $W(r)$ is a standard Wiener process and for every sub-sample i, $\xi_i \sim N(0,1)$ and it is independent of $W(r)$.

The expression derived in Theorem 2 is identical to the associated expansion for the full sample estimator, shown in (3). The expression derived in Theorem 2 provides the leading term of the asymptotic expansion of the distribution of $\ell(\hat{\beta}_i - 1)$. In the same line as in the full-sample case, if one takes expectations from both sides from the expression shown

from Theorem 2 , the same expression can be used to describe each moving-block subgroup estimator's bias:

$$E(\hat{\beta}_i - 1) = \frac{\mu}{\ell} + O(\ell^{-2}) \qquad i = 1, \ldots, M \tag{18}$$

This result is, in fact, of substantial importance when trying to derive a jackknife estimator in the unit root case. Equation (18) indicates that since the same expansion can be used to describe both the full sample and the sub-sample estimators. Hence, the jackknife estimator derived in Chambers (2010) for stationary autoregressive models can be employed in unit root situations. In fact, as we shall see in the next section, this moving blocks jackknife estimators not only can be invariant to initial conditions asymptotically, but can outperform the non-overlapping one and reduce the mean square error.

5   An application to jackknife estimators

An appropriate jackknife estimator is defined as a weighted, linear combination between the full sample estimator and a number of subgroup estimators. In the case of the first order autoregression studied here, this is defined as in (19) below.

$$\hat{\beta}_J = \gamma \hat{\beta} + \delta \frac{1}{M} \sum_{i=1}^{m,M} \hat{\beta}_i \tag{19}$$

Where γ and δ correspond to the weights which depend on the sub-sampling scheme used to describe the subgroup estimators. As explained in detail in Chambers (2010), these weights are found to be $\gamma = \frac{m}{m-1}$ and $\delta = -\frac{1}{m-1}$ for the non-overlapping scheme, and as $\gamma = \frac{n}{M-1}$

and $\delta = \frac{\ell}{M-1}$ for the moving blocks one.

These weights are based upon the requirement that the expressions used to describe the full sample and sub-sample estimators' bias are the same. Clearly, as (9) suggests, this requirement is not met in the case when non-overlapping subgroups in the unit root case. This, in turn, implies that the usual weights will not manage to fully eliminate the leading term of the bias in the unit root case.

One way of overcoming this issue is to simply employ moving blocks when deriving the jackknife estimator. As suggested by Theorem 1, each moving block subgroup estimator inherits the same asymptotic distribution as its full sample counterpart. In addition, Theorem 2 proves the validity of an asymptotic expansion for each moving blocks estimator. This expansion is shown to be identical with the one derived by Phillips (1987) for the full–sample estimator $\hat{\beta}$. These findings motivate the use of the jackknife estimator in the presence of a unit root. The jackknife estimator under (1) can employ the same weights the ones in the stationary case and can effectively eliminate the leading term of the bias expansion. This estimator, $\hat{\beta}_{J,MB}$ is defined in (20) below.

$$\hat{\beta}_{J,MB} = \frac{n}{M-1}\hat{\beta} - \frac{\ell}{M-1}\frac{1}{M}\sum_{i=1}^{M}\hat{\beta}_i \tag{20}$$

where M denotes the number of moving blocks, and $\hat{\beta}_i$ is defined in (6).

THEOREM 3. *Let $\hat{\beta}$ and $\hat{\beta}_i$ denote the full-sample and moving blocks sub-sample estima-*

tors of β which emerges from (1). Then, the moving-blocks jackknife estimator $\hat{\beta}_{J,MB}$ defined in (20) is biased to the to the second order $O(T^{-2})$, where M is an integer satisfying $2 \leq M \leq n - 1$.

Table 2 summarises the bias and mean square error (MSE) of the non-overlapping and moving blocks jackknife estimators which emerge from (1). These simulations reveal that both estimators roughly have the same performance in terms of bias reduction. Most importantly, the entries in Table 2 indicate that the moving blocks jackknife provides clear reductions to the MSE of the original estimator at each sample size examined. On the other hand, for each case the non-overlapping jackknife's bias reduction comes at a massive increase to the MSE.

In addition, the entries of Table 2 were set at fixed sub-group lengths for the two estimators. This may have worked at the benefit of the non-overlapping case as this utilised smaller subgroup lengths than the optimal ones for the moving blocks one. The optimal block size therefore requires further investigation. The figures in the appendix graph the moving block jackknife's bias for all possible choices of M, at a given sample size n. Two features become obvious from these graphs: (1) The bias minimising M is found to be relatively small and (2) the optimal M depends on the sample size.

n	m	M	ℓ	$\frac{b_J}{b_{OLS}}$	$\frac{b_{J,MB}}{b_{OLS}}$	$\frac{MSE_J}{MSE_{OLS}}$	$\frac{MSE_{J,MB}}{MSE_{OLS}}$
24	2	13	12	0.58	0.60	2.09	1.16
	3	17	8	0.60	0.65	1.43	0.94
	4	19	6	0.70	0.71	1.24	0.90
	8	22	3	0.84	0.84	1.44	0.92
50	2	26	25	0.35	0.50	1.96	1.11
	5	41	10	0.64	0.65	1.11	0.83
	10	46	5	0.76	0.77	1.13	0.85
100	2	51	50	0.48	0.50	2.00	1.08
	4	76	25	0.55	0.57	1.23	0.85
	5	81	20	0.59	0.60	1.08	0.77
	10	91	10	0.70	0.70	0.92	0.85
200	2	101	100	0.40	0.49	2.33	1.33
	4	151	50	0.53	0.55	2.33	1.00
	5	161	40	0.55	0.58	1.33	1.00
	10	181	20	0.64	0.66	1.00	1.00
400	2	201	200	0.46	0.48	2.35	1.07
	4	301	100	0.53	0.55	1.18	0.79
	5	321	80	0.56	0.57	1.08	0.78
	10	361	40	0.63	0.65	0.85	0.78

Table 2: Bias and Mean square error results of the non-overlapping and moving-blocks jackknife at fixed subgroup length ℓ (10^4 repl.)
Note: The subscripts OLS, J and J, MB denote the full-sample, non-overlapping jackknife and moving blocks jackknife estimators respectively.

6   Concluding Remarks

In this paper, we show that in the presence of a unit root, subgroup estimators of the autoregressive parameter based on the moving blocks scheme, inherit the same asymptotic distribution as their full-sample counterparts. When subgroups which emerge from a process as in (1) are formed either as non-overlapping or moving blocks of observations, their initial conditions are $O_p(\ell^{-1/2})$. Unlike the non-overlapping case, in the moving blocks case studied here, the effect of these initial conditions vanishes asymptotically. This result motivates

the use of moving blocks in deriving a jackknife estimator in the presence of a unit root. This estimator uses the same weights as in the one derived for stationary autoregression in Chambers (2010). Simulation findings summarised in Table 2 suggest that the proposed moving blocks jackknife estimator is able to provide obvious reductions to the MSE, unlike its non-overlapping counterpart. The performance of the estimator is dependent upon the choice of subgroup length, something that is currently under investigation by the author. Finally, these results easily apply to higher order autoregressions and continuous time models.

7   References

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. (2010) Jackknife estimation and inference in stationary autoregressive models. University of Essex Discussion paper 684.

Chambers, M. J. and M. Kyriacou (2010) Jackknife bias reduction in the presence of a unit root. University of Essex Discussion paper 685.

Chen, Y. and J. Yu (2011) Optimal Jackknife for Discrete Time and continuous Time Unit Root Models. Preprint.

Efron, B. (1979) Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics Vol. 7, No. 1, 1-26.

Evans, G. B. A and N. E. Savin (1981) Testing for Unit Roots: I. Econometrica Vol. 49, No. 3, 753-779.

Hurwitz, L.(1950) Least Squares Bias in Time Series. In T. C. Koopmans (ed.) Statistical Inference in Econometric Models (New York: Wiley).

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

Phillips, P. C. B. (2012) Folklore Theorems, Implicit Maps, and Indirect Inference. Econometrica 80, 425–454.

Phillips, P. C. B. (1987) Time Series Regression with a Unit Root. Econometrica 55, 277–301.

Phillips, P. C. B.,Yu, J. (2005) Jackknifing Bond Option Prices. Review of Financial Studies. 18, 707–742.

Politis, D.N., Romano, J.P., and M. Wolf (1997). Subsampling for heteroskedastic time series. Journal of Econometrics 81, 281-317.

Quenouille, M. H. (1949) Approximate tests of correlation in time series. Journal of the Royal Statistical Society, Series B Vol. 11, pp. 68-84.

8   Appendix

Proof of LEMMA 1. Define the partial sum $S_{[nr]} = \sum_{j=1}^{[nr]} \epsilon_j$. From the FCLT it follows that as $n \to \infty$:

$$n^{-1/2} S_{[nr]} \Rightarrow \sigma W(r)$$

The random walk process shown in (1) can be rewritten as:

$$y_t = y_0 + S_t$$

$$S_{t-1} = S_{t-1} \int_{(t-1)/n}^{t/n} dr = n \int_{(t-1)/n}^{t/n} S_{[nr]} dr \tag{21}$$

(a) From (21), it follows that for the moving blocks case we have that:

$$\sum_{t=i}^{i+\ell-1} S_{t-1} = n \sum_{t=i}^{i+\ell-1} \int_{(t-1)/n}^{t/n} S_{[nr]} dr = n \int_{\frac{i-1}{n}}^{\frac{i+\ell-1}{n}} S_{[nr]} dr$$

In view of the fact that the block's length ℓ and the sample size n grow at the same rate as $n \to \infty$, it holds that in the limits of the integral:

$$\lim_{n \to \infty} \frac{i-1}{n} = 0 \text{ and } \lim_{n \to \infty} \frac{i+\ell-1}{n} = 1$$

so that

$$\sum_{t=i}^{i+\ell-1} S_{t-1} = n \int_0^1 S_{[nr]} dr$$

which, in turn, verifies that:

$$\sum_{t=i}^{i+\ell-1} S_{t-1} \Rightarrow \sigma \int_0^1 W(r) dr \tag{22}$$

18

In the same lines, it follows that:

$$\sum_{t=i}^{i+\ell-1} S_{t-1}^2 = n \int_0^1 S_{[nr]}^2 dr \qquad (23)$$

Finally, from (22) It follows that:

$$\ell^{-3/2} \sum_{t=i}^{i+\ell-1} y_{t-1} = \ell^{-1/2} y_0 + \ell^{-3/2} \sum_{t=i}^{i+\ell-1} S_{t-1} \Rightarrow \sigma \int_0^1 W(r) dr$$

(b) Since $\frac{n}{\ell} \to 1$ as $n \to \infty$ and by taking into account (23) we have that:

$$\ell^{-2} \sum_{t=1}^{i+\ell-1} S_{t-1}^2 = \frac{n}{\ell} \int_0^1 \left(n^{-1/2} S_{[nr]}\right)^2 \Rightarrow \sigma^2 \int_0^1 W(r)^2 dr$$

Since $y_{t-1}^2 = y_0^2 + 2y_0 S_{t-1} + S_{t-1}^2$ it follows that for the denominator of (6) we have that:

$$l^{-2} \sum_{t=i}^{i+\ell-1} y_{t-1}^2 = \ell^{-2} y_0^2 + 2\ell^{-1/2} y_0 \ell^{-3/2} \sum_{t=i}^{i+\ell-1} S_{t-1} + \ell^{-2} \sum_{t=i}^{i+\ell-1} S_{t-1}^{-2} \Rightarrow \sigma^2 \int_0^1 W(r)^2 dr$$

(c) The numerator of (6) can be expressed in the following way:

$$\sum_{t=1}^{i+\ell-1} y_{t-1} \epsilon_t = \frac{1}{2} \left(y_{i+\ell-1}^2 - y_{i-1}^2 - \sum_{t=1}^{i+\ell-1} \epsilon_t^2 \right) \qquad (24)$$

From the LLN we have that $\frac{1}{\ell} \sum_{t=i}^{i+\ell-1} \epsilon_t^2 \to_p \sigma^2$

$$y_{i+\ell-1} = y_0 + S_{i+\ell-1} = o_p(1) + S_{[n\frac{i+\ell-1}{n}]}$$

19

$$y_{i-1} = y_0 + S_{i-1} = o_p(1) + S_{[n\frac{i-1}{n}]}$$

$$\ell^{-1/2} y_{i-1} =^d o_p(1) + S_{[n\frac{i-1}{n}]} \Rightarrow \sigma W(0)$$

$$\ell^{-1/2} y_{i+\ell-1} =^d o_p(1) + S_{[n\frac{i+\ell-1}{n}]} \Rightarrow \sigma W(1)$$

The above imply that the first two terms of (24) have the following behaviour asymptotically:

$$\ell^{-1} y_{i-1} \Rightarrow \sigma^2 W(0)^2 \tag{25}$$

$$\ell^{-1} y_{i+\ell-1} \Rightarrow \sigma^2 W(1)^2 \tag{26}$$

Thus, the numerator of (6) becomes:

$$\ell^{-1} \sum_{t=i}^{i+\ell-1} y_{t-1}\epsilon_t \Rightarrow \frac{\sigma^2}{2} \left[W(1)^2 - 1\right]$$

$\square$

Proof of THEOREM 1. This result follows directly from parts (b) and (c) of Lemma 1, as these correspond to the numerator and denominator of the subgroup normalised bias measure, for every $i = 1, \ldots, M$:

$$l(\hat{\beta}_i - 1) = \frac{\ell^{-1} \sum_{t=i}^{i+\ell-1} y_{t-1}\epsilon_t}{\ell^2 \sum_{t=i}^{i+\ell-1} y_{t-1}^2}$$

$\square$

Proof of THEOREM 2: The proof for this theorem follows from section 7.1 in Phillips (1987), by using the following representation:

$$X_n(r) \equiv W(r) \left\{ 1 - \frac{1}{2} \frac{nr - [nr]}{nr} \right\} + O_p(n^{-2})$$

We need to derive an asymptotic expansion for $l(\hat{\beta}_i - 1)$:

$$l(\hat{\beta}_i - 1) = \frac{\sum_{t=i}^{i+\ell-1} y_{t-1} \epsilon_t}{\sum_{t=i}^{i+\ell-1} y_{t-1}^2} \tag{27}$$

For the numerator of (27) can be re-written as:

$$\frac{1}{\ell} \sum_{t=i}^{i+\ell-1} y_{t-\epsilon_t} = \frac{1}{2} \left\{ \frac{1}{\ell} y_{i+\ell-1}^2 - \frac{1}{\ell} y_{i-1}^2 - \frac{1}{\ell} \sum_{t=i}^{i+\ell-1} (\epsilon_t^2 - \sigma^2) - \sigma^2 \right\}$$

It is straightforward that $l^{-1/2} y_{i+\ell-1} = X_n(1)$ and $\ell^{-1/2} y_{i-1} = X_n(0)$ from which it follows that:

$$\frac{1}{\ell} y_{i+\ell-1}^2 = X_n(1)^2 \equiv \sigma^2 W(1)^2 + O_p(\ell^{-1})$$

$$\frac{1}{\ell} y_{i-1}^2 = X_n(0)^2 \equiv \sigma^2 W(0)^2 + O_p(\ell^{-1})$$

In addition, we know that $n^{-1/2}(\sum_{t=1}^{n} (\epsilon_t^2 - \sigma^2) \to_d \sqrt{2}\sigma^2 \xi$, where $\xi \sim N(0, 1)$. Since in the current framework the block's length ℓ grows at the same rate with n, as $n \to \infty$, it follows that:

$$\ell^{-1/2} \sum_{t=i}^{i+\ell-1} (\epsilon_t^2 - \sigma^2) \to_d \sqrt{2}\sigma^2 \xi_i$$

This implies that:

$$\frac{1}{2\ell} \sum_{t=i}^{i+\ell-1} (\epsilon_t^2 - \sigma^2) \to_d \frac{1}{\sqrt{2\ell}} \sigma^2 \xi_i$$

Therefore, for the numerator of (27) it follows that:

$$\frac{1}{\ell} \sum_{t=i}^{i+\ell-1} y_{t-1}\epsilon_t \equiv \frac{\sigma^2}{2} \{X_n(1)^2 - 1\} - \frac{1}{\sqrt{2l}}\sigma^2 \xi_i + O_p(\ell^{-1}) \tag{28}$$

$$\frac{1}{\ell} \sum_{t=i}^{i+\ell-1} y_{t-1}\epsilon_t \equiv \frac{\sigma^2}{2} \{W(1)^2 - 1\} - \frac{1}{\sqrt{2\ell}}\sigma^2 \xi_i + O_p(\ell^{-1}) \tag{29}$$

For the denominator of (27) we have that:

$$\frac{1}{\ell^2} \sum_{t=i}^{i+\ell-1} y_{t-1}^2 = \frac{1}{\ell^2} \sum_{t=i}^{i+\ell-1} S_{t-1}^2$$

$$\frac{1}{\ell^2} \sum_{t=i}^{i+\ell-1} y_{t-1}^2 \equiv \sigma^2 \int_0^1 W(r)^2 dr + O_p(\ell^{-1})$$

$\square$

Proof of THEOREM 3. By taking expectations from each side of (20), we get that:

$$E(\hat{\beta}_{J,MB}) = \frac{n}{M-1} E(\hat{\beta}) - \frac{\ell}{M-1} E(\hat{\beta}_i) \tag{30}$$

By substituting (4) and (18) into the expression above, we get:

$$E(\hat{\beta}_{J,MB}) = \frac{n}{M-1}\left(\beta + \frac{\mu}{n} + O(n^{-2})\right) - \frac{\ell}{M-1}\left(\beta + \frac{\mu}{\ell} + O(\ell^{-2})\right)$$

22

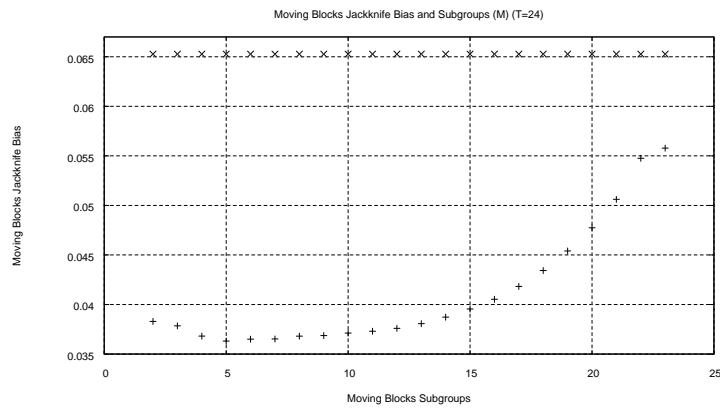$$\Rightarrow \ E(\hat{\beta}_{J,MB}) = \beta + O(n^{-2})$$

$\square$

8.1 Figures

Figure 1: Moving-Blocks Jackknife bias and number of subgroups (M) at $n = 24$. The horizontal line denotes the corresponding full-sample least squares bias. (10^4 replications used)
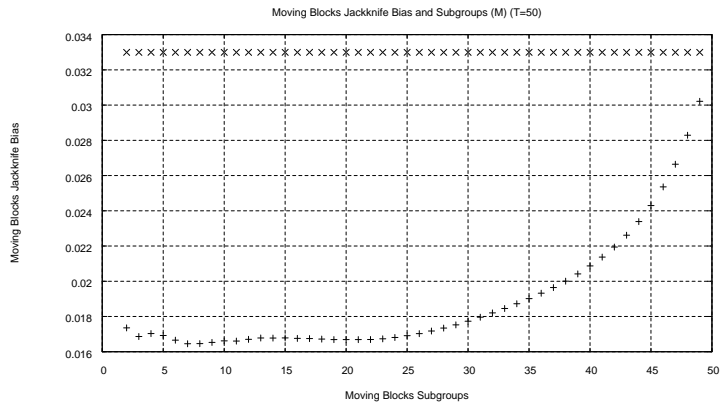
Figure 2: Moving-Blocks Jackknife bias and number of subgroups (M) at $n = 50$. The horizontal line denotes the corresponding full-sample least squares bias. $(10^4$ replications)
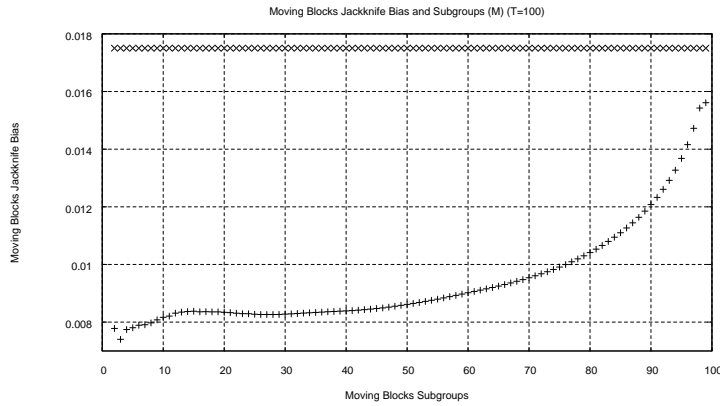

Figure 3: Moving-Blocks Jackknife bias and number of subgroups (M) at $n = 100$. The horizontal line denotes the corresponding full-sample least squares bias. $(10^4$ replications)
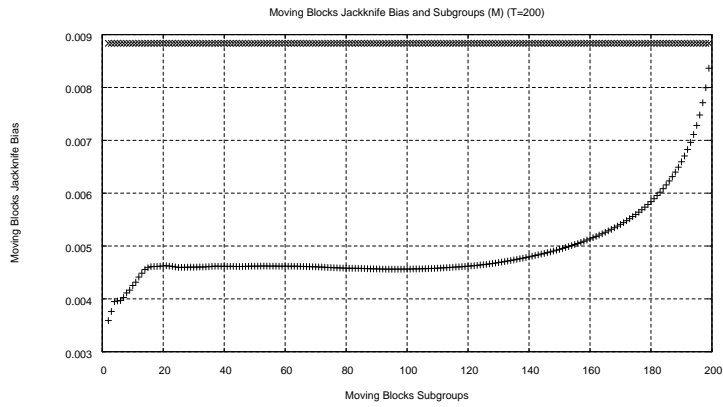
Figure 4: Moving-Blocks Jackknife bias and number of subgroups (M) at $n = 200$. The horizontal line denotes the corresponding full-sample least squares bias (10^4 replications)
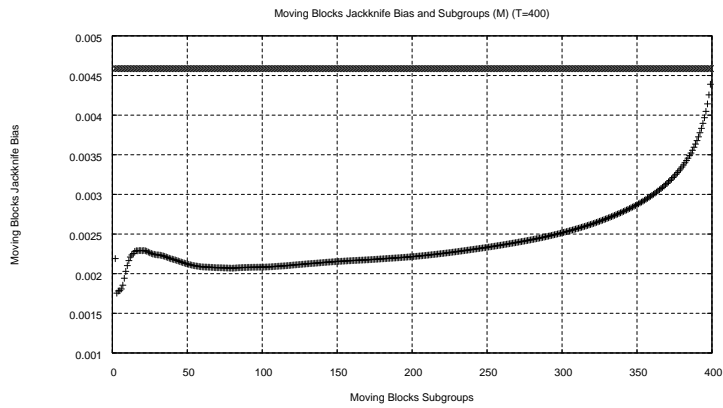

Figure 5: Moving-Blocks Jackknife bias and number of subgroups (M) at $n = 400$. The horizontal line denotes the corresponding full-sample least squares bias. (10^4 replications)