Jackknife Estimation in Autoregressive Models with a Unit Root*

Maria Kyriacou
Department of Economics
University of Essex

November 9, 2009

Abstract

This paper examines the performance of the jackknife resampling method in unit root autoregression. Two different sub-sampling methods used for deriving jackknife estimators are considered: the first one employs non-overlapping subgroups, while the second one uses moving blocks of observations. In the case of a pure random walk, from the second subgroup onwards, the associated initial conditions become explosive. A simple alteration to the data generation process of these subgroups allows to overcome this issue. The performance of the proposed jackknife estimators compared to their full-sample, least squares counterparts is examined in a Monte Carlo study. The simulations reveal that both jackknife estimators provide a substantive reduction to the bias in every case, especially in moderate sample sizes. Specifically, at a fixed subgroup length, the non-overlapping jackknife outperforms the moving-blocks one. For any given sample size, the bias of the non-overlapping jackknife estimator is minimised when only two subgroups are used to derive the estimator. However, subject to the appropriate choice of subgroups, the moving-blocks jackknife is able to reduce the mean square error of the original estimator. In the final part of the paper, the robustness of the results under deviations from Gaussian innovations, zero initial conditions and the presence of deterministic components (constant and/or linear trend) in the regression is also studied.

Keywords. Jackknife; bias; autoregression; unit root; resampling methods.

J.E.L. classification numbers. C01; C13; C14; C22.

Address for Correspondence: Maria Kyriacou, Department of Economics, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, Essex, England. Email: mkyriaw@essex.ac.uk; Tel: +44 1206 874234.

*Acknowledgments: I am indebted to Prof. Marcus J. Chambers for guidance and comments on this paper. I would also like to thank Prof. Karim Abadir, Dr Emma Iglesias, Prof. Joao Santos Silva, the participants of the Research Strategy Seminar at the University of Essex in 2008 and INFER 2009 in Stirling, Scotland for their comments and suggestions on preliminary versions of this paper. All remaining errors are mine.

1 Introduction

Resampling methods like the jackknife and the bootstrap are becoming increasingly popular in econometrics. They serve as useful tools for making inference in finite sample situations where the large sample asymptotics fail to work and the population distribution is unknown. With the rapid development of computers, these numerically intensive methods have become highly accessible. The last two decades have witnessed a vast amount of literature focusing on theoretical and applied work of these methods. However, applications of the jackknife method in a time-series context are still in their infancy.

This paper investigates the performance of the jackknife in a non-stationary dynamic environment. Specifically, the ability of the jackknife to provide reductions to the bias of the correlation coefficient which emerges from an autoregressive process with a unit root is examined. In doing so, two different methods of obtaining the jackknife estimator are illustrated, the performance of which is assessed in a Monte Carlo study.

The jackknife was introduced by Quenouille (1956) as a non-parametric method for bias reduction. The basic idea is to subsequently delete points from a data set (of an unknown distribution) and recompute the emerging averages. The jackknife bias-reduced estimator is established by eliminating the leading term of a power series expansion. It is both simple as a method, and most importantly it requires no *a priori* assumptions about the underlying population distribution. It was not named the jackknife until Tukey (1958) used Quenouille's idea in confidence interval estimation. The jackknife is a valuable and general statistical tool as it allows for many estimation methods to be used. There is a vast amount of statistics literature on variations and applications of the jackknife method[1] . However, applications of resampling methods like the jackknife in time series requires further attention, for implementing them in a way so that the dependence structure is not affected.

The primary motivation for this paper arises from the recent use of the jackknife in stationary time series situations. Phillips & Yu (2005) apply the jackknife method in a bond options market application to handle bias estimation issues. In fact, the authors examine two methods of applying the jackknife technique: either by applying the jackknife to the bond options function directly or by applying it to each of the variables that constitute the bond option function. Additionally, Hahn & Newey (2006) provide a jackknife based bias application on non-linear panel data. Also, Phillips & Yu (2007) apply the jackknife in a continuous time financial application. Chambers (2009) provides an exhaustive investigation on the use of the jackknife in stationary autoregressive models under different specifications which include, among others, the presence of an intercept and/or a trend, non-Gaussian innovations and heteroskedasticity. As these papers focus exclusively on stationary settings, an obvious extension from these is to assess the jackknife's ability to provide reductions to the finite sample bias under non-stationarity. The case of non-stationary autoregression constitutes a particularly interesting setting, especially in economics, for various reasons that we shall discuss next.

Furthermore, another resampling method,the bootstrap, has been recently applied in non-stationary environments mainly for the purposes of constructing bootstrap-based unit root tests. Since the jackknife and the bootstrap are closely related methods[2], we can be confident that an appropriate way of successfully applying the jackknife method in a non-stationary environment can also be found. The bootstrap is a widely used tool in financial applications; however it is a highly computationally intensive method as it requires for numerous "boot-

[1]Interested readers may refer to Miller (1974), Efron (1982), Shao & Tu (1995) for comprehensive reviews and extensions for the jackknife technique.

[2]See Efron (1979) for an exhaustive analysis of the jackknife and the bootstrap and their equivalence.

strap replications" to be generated. The jackknife however, appears to be a much simpler and computationally less demanding resampling method. For example, Phillips & Yu (2005) in their jackknife application find that, for the majority of the cases that they examine, the bias is minimised when only two subgroups are used.

The main objective of this paper is to apply the jackknife and assess its performance as a bias reduction tool in a non-stationary environment. For this purpose, one of the simplest, yet widely used models in time series analysis is used: the first order autoregressive model $AR(1)$. The least squares estimator of the correlation coefficient which links the present value of a variable with its lagged value is known to be severely biased in an $AR(1)$ model with a unit root (an exhaustive study of this can be found in Evans & Savin (1981)). This constitutes a very interesting case in which the finite sample bias of the full-sample least squares estimator requires a better estimation method. Although in many occasions reductions to the bias can be linked with increases in the mean square error, in this setting bias corrections are particularly desirable as they can lead to improvements in unit root testing.

The primary reason why this has not been attempted before is because the Edgeworth expansions which are typically used to describe the bias in such cases do not hold under non-stationarity. The Edgeworth expansions derived by Kendall (1954) and White (1961) become explosive as the autoregressive coefficient reaches unity. However, Phillips (1987) and Abadir (1993, 1995) derive appropriate expansions used to describe the bias of the least-squares estimator of an $AR(1)$ model with a unit root. As Phillips & Yu (2005) correctly point out, the jackknife bias reduction does require a functional form for the bias, although its performance does not explicitly rely on the functional form of the expansion. Two different methods of sub-sampling are used for obtaining the jackknife estimator are discussed and assessed. These are the non-overlapping jackknife (also used in Phillips & Yu (2005)) and the moving-blocks jackknife (introduced by Kuensch (1989)). As we proceed, issues which are not immediately evident are addressed, such as the effect of initial conditions on the performance of the estimators, the optimal subgroup choice and the appropriate way of comparing these two jackknife estimators. A Monte Carlo study is carried out to evaluate the performance of the jackknife as opposed to the full-sample least squares estimator and to give insights on these issues. Although the main analysis primarily relies on the case of Gaussian random walk with no deterministic components, the robustness of the results under the presence of deterministics, deviations from normality and non-zero initial conditions are also considered.

The findings of the Monte Carlo study reveal that the jackknife provides a substantial reduction to the bias of the least squares estimator in every case examined. The non-overlapping jackknife outperforms the moving-blocks one, and the bias is minimised when only two non-overlapping subgroups are used to derive the jackknife estimator. The moving blocks jackknife is able to reduce the mean square error of the original estimator.

The paper is organised as follows. Section 2 discusses in detail the data generation process and the main framework analysed in this paper, while in Section 3 the non-overlapping and moving-blocks jackknife methods are formally defined. Section 4 presents a Monte Carlo study for examining the performance of the two methods of deriving the jackknife estimator, and also to compare the jackknife estimators with its least squares counterparts. Finally, sections 5 and 6 consider the cases of non-zero initial values, non-Gaussian innovations and the presence of deterministic components in the regression model respectively. The paper's concluding remarks are shown in section 7, while Proofs and Figures are left in the appendix.

2 Bias in Autoregression

The first-order autoregression model (known as the $AR(1)$ model) satisfies the first order equation in which it relates the present value of a variable to its first lag. Let the realisation $\{y_1,\ y_2, \ldots,\ y_T\}$ emerge from the process defined as in (1) below.

$$y_t = A^{'}x_t + \beta y_{t-1} + \epsilon_t, \quad y_0 = 0 \quad \epsilon_t \sim i.i.d.(0, \sigma^2) \tag{1}$$

where the index t is the time indicator, β the autoregressive (or correlation) coefficient, $A^{'}x_t$ denotes a linear-in-parameters deterministic components where A is a vector of unknown parameters and x_t is a known vector of deterministic terms. Finally, ϵ_t denotes the (unobserved) error/innovation term.

Since the 1950s, the bias[3] of the least squares estimator which emerges from the $AR(1)$ model has been a subject of interest in the statistics and econometrics literature, with Hurwitz (1950) being the first to investigate the small sample bias of this model. The bias of the estimated least squares coefficient of β is substantial in small samples for stationary[4] models, with the bias being in fact proportional to the true value of β: the bias vanishes as β tends towards zero and increases as the parameter reaches unity (in absolute terms).

The stationary case of the $AR(1)$ process seems appealing as it involves many desirable characteristics regarding its moments[5]. However, assuming stationarity when dealing with economic or financial data might not appear to be very realistic, especially since most economic variables (for example population sized variables such as employment, output) tend to be integrated to the first order $I(1)$[6]. For this reason, after the the 1960s, the literature turned its focus on the non-stationary case. The non-stationary case can be well approximated by an autoregressive (AR) process with a unit root (the true value of the parameter β being equal to one) and with the root of the equation $\phi(z) = 1 - \beta z = 0$ lying on the unit circle.

This paper aims to investigate the ability of the jackknife to provide improvements regarding bias in a non-stationary environment. For this purpose, the results will primarily rely on the simplest form of a non-stationary autoregressive process: the $AR(1)$ process with a unit root and no deterministic components, known as the pure random walk case. This is obtained by setting the deterministic compoments to equal zero $A^{'}x_t = 0$, while the value of the autoregressive parameter is equal to unity: $\beta = 1$.

To fix ideas, let $y_1, y_2, \ldots y_T$ be a sequence of observations obtained from the following process:

$$y_t = \beta y_{t-1} + \epsilon_t \quad \text{with} \quad \beta = 1, \quad y_0 = 0 \quad \& \quad \epsilon_t \sim i.i.d(0, \sigma^2) \tag{2}$$

where $t = 1, \ldots, T$ is the time indicator, y_0 is the initial value and ϵ_t is the innovation term which is identically independently distributed ($i.i.d$) with mean zero and variance σ^2. It is common to assume that the stationary component of the random walk process (i.e. the error term) follows a Gaussian distribution $\epsilon_t \sim N(0, \sigma^2)$.

[3]The bias of an estimator $\hat{\theta}$ is defined as the expected value of the difference between the estimated value $\hat{\theta}$ and the real value θ: $BIAS(\hat{\theta}) = E(\hat{\theta} - \theta)$

[4]By stationarity we mean covariance/weak stationarity rather than strict stationarity, where under the former, the moments of the process are independent of the time lags. The main requirement for attaining covariance stationarity in the case of the $AR(1)$ model is $|\beta| < 1$. Under the non-stationary case (or unit root case) the value of the correlation coefficient is one $\beta = 1$.

[5]In stationary environments, the first moment of y_t, μ, is given by $\mu = \frac{\alpha}{(1-\beta)}$, where α is the constant term (or drift)-see section 6 on a detailed discussion of models with constant terms; while the variance and autocovariance (γ_0 and γ_j respectively) are $\gamma_0 = \frac{\sigma^2}{(1-\beta^2)}$ and $\gamma_j = [\beta^j/(1-\beta^2)]\sigma^2$. These values are not valid at $\beta = 1$ as the denominator of these becomes zero.

[6]A non-stationary series is said to be integrated of order d, $I(d)$, if it becomes stationary after being first differenced d times.

There are two reasons why this is the main data generating process for the purposes of this analysis. First, because of the simplicity in algebraic and programming requirements of this process, as having a constant or a trend makes the model more complicated. By using backwards substitution, the random walk data generating process can be expressed in terms of the sum of the innovation terms as:

$$y_t = \sum_{j=1}^{t} \epsilon_j \tag{3}$$

which implies that the variance of the dependent variable $Var(y_t) = t\sigma^2$ reaches infinity as $t \to \infty$, so the distribution degenerates and therefore, in order to make any inference, different transformations need to be used.

Secondly, the resulting finite sample bias of the least squares estimator in the random walk case is highly evident, as the estimates are biased downwards not only for very small samples (< 50) but also for larger samples (> 200). The driftless random walk case might not appear as very realistic with economic data (economic variables tend to have a trend rather than being pure random walks[7]), and random walk cases can only be found in some extreme cases of stock prices examples. Therefore, further extensions of this model will also be considered in order to make this analysis more coherent with economic applications. These extension include the cases of non-zero initial values ($\frac{y_0}{\sigma} \neq 0$), non-Gaussian innovations and the presence of deterministic components in the regression model.

The behaviour of the Ordinary Least Squares (OLS) estimator of β, denoted as $\hat{\beta}$, has been excessively studied in the literature (e.g. Evans & Savin (1981)). Simulation results reveal that $\hat{\beta}$ is severely biased downwards with the bias being particularly evident in moderate samples. Specifically, when the underlying process incorporates a unit root, as in (2), the estimated coefficient becomes:

$$\hat{\beta} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} = 1 + \frac{\sum_{t=1}^{T} y_{t-1}\epsilon_t}{\sum_{t=1}^{T} y_{t-1}^2} \tag{4}$$

The bias of $\hat{\beta}$ becomes obvious after taking the expectations from both sides of (4):

$$E(\hat{\beta}) = 1 + E\left(\frac{\sum_{t=1}^{T} y_{t-1}\epsilon_t}{\sum_{t=1}^{T} y_{t-1}^2}\right) \neq 1 + \frac{E(\sum_{t=1}^{T} y_{t-1}\epsilon_t)}{E(\sum_{t=1}^{T} y_{t-1}^2)} \tag{5}$$

Under a unit root, $\hat{\beta}$ is biased downwards; however it converges to its true value at rate T and therefore is said to be super-consistent or T-consistent[8]. The normalised estimated coefficient $T(\hat{\beta} - 1)$ has a limiting distribution which is expressible in terms of the Wiener process[9] W(.) as:

$$T(\hat{\beta} - 1) \Rightarrow \frac{1}{2}\frac{\left[(W(1)^2 - 1\right]}{\int_0^1 W(r)^2 dr} \tag{6}$$

where " $\Rightarrow$ " denotes weak convergence of the associated probability measures. These results were first derived by White (1958) while also extensions of these for the case of serially correlated errors can be found in Phillips (1987a, 1987b). From the definition of the Wiener process, $W(1)$ is a standard normal variable, $W(1) \sim N(0, 1)$, which means that its squared

[7]Except Hall's Random walk which is a theoretical model that assumes that consumption follows a random walk.

[8]In the stationary case the consistency rate is reached at a slower rate $\sqrt{T}$.

[9]A Wiener process (or Brownian motion) $W(t)$ is a continuous function which has the following properties:(a) $W(0) = 0$, (b) For any $0 \leq t_1 < t_2 < \ldots < t_k < 1$, the increments $W(t_2) - W(t_1), \ldots, W(t_k) - W(t_{k-1})$ are independent multivariate Gaussian random variables with $W(t) - W(s) \sim N(0, t - s)$ for $s < t$, (N indicates the Normal/Gaussian Distribution) and (c) $W(t)$ is continuous in t with probability 1 (See Hamilton (1994), pp.478).

term, which appears in the numerator of (6), follows a chi-square distribution with one degree of freedom: $W(1)^2 \sim \chi^2_1$. The probability that $W(1)^2$ is below one is 0.68; this along with the fact that the denominator of (6) is always positive, imply that the ratio must be negative at two third of the time. These features of the distribution suggest that when the true value of the parameter β is equal to one, the least squares estimator $\hat{\beta}$ tends to underestimate the true value of the parameter even asymptotically in 68% of the time.

2.1 Application of the Jackknife in Time Series

Having to deal with dependent data means that we now deviate from the $i.i.d$ (independent identically distributed) case on which the initial jackknife methods were based. In a time-series context one needs to be careful when trying to implement resampling methods like the jackknife so that no distortion is caused to the structure of the variance/covariance matrix. Therefore, using the usual delete-1/delete-d jackknife methods means that we have to delete random points from the data set, which can only be applicable in $i.i.d$ cases but not with dependent data.

Applications of the jackknife in a time series environment are rather recent in the literature and focus exclusively on stationary cases (for example Phillips & Yu (2005)). The relevant literature encompasses two methods of obtaining the jackknife estimator whilst maintaining the dependence structure of the data. These are namely the non-overlapping blocks and moving-blocks jackknife. Their distinction arises from the different way by which the subgroups are generated, with the former using consecutive subgroups of non-overlapping observations and the latter using subgroups of overlapping observations (moving blocks).

Using blocks of observations or sub-series to improve estimation emerges from the idea that these sub-series act as replicates of the (unknown) population distribution, and therefore result in better estimates in the sense that these are closer to the parameter's true value. This does not have to imply that we need a large number of subgroups in order to improve our estimates. In fact, Phillips & Yu (2005) use the non-overlapping version of the jackknife in order to estimate a complicated bonds option model to find that the Mean Square Error (MSE hereafter)[10] minimising results are obtained when two subgroups are used. Carlstein (1986) suggests the idea of employing adjacent non-overlapping sub-series to be used in estimation by employing methods like the jackknife and the bootstrap. The moving-blocks jackknife (and also the widely used moving-blocks bootstrap) were introduced by Kuensch (1989) to be used in stationary time-series.

The main reason why the jackknife has only been used in stationary cases up to now is because in order to achieve bias reduction, a power-series expansion for the bias is required. As many papers address the issue of bias in estimation for the $AR(1)/AR(p)$ models, by deriving asymptotic expansions for the moments (such as Kendall (1954), White (1961), Tanaka (1983)), these Edgeworth-type expansions are derived for stationary environments and cannot be applicable to the unit root case. Even though the jackknife requires a functional form for the bias, it does not rely explicitly on the functional form of it (the power-series in this case). In fact, Phillips (1987) (p.p 274) provides the leading term of the asymptotic expansion of the distribution of the normalised estimated coefficient $T(\hat{\beta}-1)$, whose true value is one.

[10]The mean square error (MSE) of an estimator $\hat{\theta}$ of a parameter θ is defined to be: $MSE(\hat{\theta}) = E(\hat{\theta}-\theta)^2$. The MSE measures the trade off between changes in bias and the variance and it can also be expressed as: $MSE(\hat{\theta}) = BIAS(\hat{\theta})^2 + Var(\hat{\theta})$.

His expansion is expressed in terms of a Wiener process $W(.)$ and a term ξ which follows a standard Gaussian distribution $N(0,1)$ as:

$$T(\hat{\beta}-1) \equiv \frac{\frac{1}{2}\left[W(1)^2-1\right]}{\int_0^1 W(r)^2 dr} - \frac{\frac{1}{\sqrt{2T}}\xi}{\int_0^1 W(r)^2 dr} + O_p(T^{-1}) \tag{7}$$

By taking expectations from each side of (7), the $\frac{1}{\sqrt{T}}$ term disappears due to presence of the standard normal term ξ. Thus, an appropriate expression for the bias of $\hat{\beta}$ under a unit root is found by then multiplying each side with $\frac{1}{T}$.

$$E(\hat{\beta}) = 1 + \frac{c}{T} + O_p(T^{-2}) \quad \text{with} \quad c = E\left\{\frac{\frac{1}{2}\left[W(1)^2-1\right]}{\int_0^1 W(r)^2 dr}\right\} \tag{8}$$

The expansion for the bias of $\hat{\beta}$ described by (8) above, resembles a power series expansion. In the framework examined here, the jackknife will eliminate the $\frac{c}{T}$ term.

3 Methods of Obtaining a Jackknife Estimator

3.1 Non-overlapping Jackknife

The first and simplest way of obtaining a jackknife estimator is by employing consecutive, non-overlapping subgroups. Suppose that we have a sample of size T, and we use m consequent non-overlapping subgroups of observations (where $m \geq 2$), each of length n_m, so that $T = m \times n_m$. Employing non-overlapping subgroups indicates that each observation is only used once, since each subgroup i contains observations $(i-1)n_m + 1$ until in_m. Therefore, along with the full-sample estimator $\hat{\beta}$ which is based on a sample of size T, we also need to obtain the least squares estimators for each subgroup i (where $i \in \{1, 2, \ldots, m\}$) denoted by $\hat{\beta}_i$ [11]. It is important at this point to stress the distinction between the non-overlapping jackknife which employs m non-overlapping subgroups in order to maintain the dependence structure of the time series and the delete-d jackknife. The latter is established by deleting blocks of random observations from a sample, while the former uses subgroups that are part of a time series. The reason of using the delete-d jackknife, instead of the delete-1 jackknife in $i.i.d$ situations, is mainly for computational saving purposes.

The function for the proposed jackknife estimator is given by the weighted difference between the full-sample estimator $\hat{\beta}$ and the average of all the subgroup estimators $\hat{\beta}_i$s where $i = 1, \ldots, m$. The objective of the jackknife method is to reduce the order[12] of the bias from $O(T^{-1})$ to the second order $O(T^{-2})$, so the weights are chosen in a manner so that the first order of the bias (that is, the leading term of the power-series expansion) is eliminated. The non-overlapping jackknife estimator has the following general form:

$$\hat{\beta}_J = \gamma\hat{\beta} - \delta\frac{1}{m}\sum_{i=1}^{m}\hat{\beta}_i \tag{9}$$

[11] This means that obtaining the non-overlapping jackknife requires for $(m+1)$ estimators to be calculated: the full sample estimator and the estimator for each one of the m subgroups.
[12] A function f is of order ϕ: $O(\phi)$, if and only if there exists a positive real number A so that $|f(x)| < A\phi(x)$.

where $\hat{\beta}$ denotes the least squares estimator based on the entire sample T, while for each subgroup i the corresponding estimator is given by $\hat{\beta}_i$ is calculated. These are defined as (10) and (11) below.

$$\hat{\beta} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} \tag{10}$$

$$\hat{\beta}_i = \frac{\sum_{t=(i-1)n_m+1}^{in_m} y_t y_{t-1}}{\sum_{t=(i-1)n_m+1}^{in_m} y_{t-1}^2}, \quad i \quad \in 1, 2, .., m \tag{11}$$

Proposition 1 below explains how analytical solutions for the jackknife weights γ and δ are found (these are merely functions of m which is assumed to be fixed in this paper). The results of Proposition 1 rely on the assumption that the expansion which describes the bias of the original estimator $\hat{\beta}$ also describes the bias of the subgroup estimators $\hat{\beta}_i$ (with $i = 1, 2, \ldots, m$).

Proposition 1 (Non-overlapping weights):
The non-overlapping groups jackknife estimator $\hat{\beta}_J$: $\hat{\beta}_J = \gamma\hat{\beta} - \delta\frac{1}{m}\sum_{i=1}^{m}\hat{\beta}_i$ is biased to the second order ($O(T^{-2})$) if and only if the weights are $\gamma = \frac{m}{m-1}$ and $\delta = \frac{1}{m-1}$ (where m is an integer number with $m \geq 2$).

The jackknife estimator $\hat{\beta}_J$, prompted by Proposition 1 is therefore be defined as:

$$\hat{\beta}_J = \frac{m}{m-1}\hat{\beta} - \frac{1}{(m-1)}\frac{1}{m}\sum_{i=1}^{m}\hat{\beta}_i \tag{12}$$

The weights $\gamma = \frac{m}{m-1} > 1 \ \forall \ \ m > 1$, while $\delta = \frac{1}{m-1} \in \ \ [0, 1]$. The functional form described in (12) coincides with the one used in Phillips & Yu (2005) for their bond options application.
For the unit root case examined here, we want the resulting jackknife estimates to be (on average) as close to unity as possible and for a given subgroup size m, this is driven by the trade off between the increase in the first term $\frac{m}{m-1}E(\hat{\beta})$ (which will be below $\frac{m}{m-1}$ since the original estimator $\hat{\beta}$ is underestimated from its true value of one) and the reduction of the weighted average term of the subgroups estimators $\frac{1}{(m-1)}\frac{1}{m}E\sum_{i=1}^{m}E(\hat{\beta}_i)$.

3.1.1 Subgroups' Initial Conditions: Proposition 1 Revisited

The weights illustrated in Proposition 1 work ideally in stationary situations (for example Phillips & Yu (2005) employ the estimator defined by Proposition 1 and (12) in a stationary time-series application) as these manage to eliminate the first order of the original estimator's bias. However, when the jackknife estimator described by (12) is used within the the pure random walk framework described by (2), this requires further attention. Specifically, the problem arises at the second subsample and onwards: the initial values of subgroup(s) i (where $i \geq 2$) are no longer fixed but are, in fact, equal to the sum of all previous observations' error terms.
As implied by (2), for every subgroup i, with $i \geq 2$, the associated data generating process is defined as:

$$y_t = y_{(i-1)n_m} + \sum_{k=(i-1)n_m+1}^{t} \epsilon_k; \quad \forall \ \ t \in \{(i-1)n_m + 1, \ldots, in_m\} \ \text{ and } \ i \geq 2 \tag{13}$$

The initial conditions which correspond to each subgroup i are found at time point $t = (i - 1)n_m$, as shown by (14) below.

$$y_{(i-1)n_m} = y_0 + \sum_{k=1}^{(i-1)n_m} \epsilon_k \tag{14}$$

As the equation above indicates, for each one of these subgroups $i \geq 2$, the corresponding initial values $y_{(i-1)n_m}$ are not fixed, but instead are equal to the sum of y_0, which is $O_p(1)$, and the accumulated errors which are not even bounded random variables $O_p(1)$, but are instead explosive[13] $O_p(T^{1/2})$.

This issue has implications not only on the limiting distribution but also on the finite-sample properties of $\hat{\beta}_J$ as defined up to this point. In terms of bias reduction, it implies that the weights of the jackknife estimator $\hat{\beta}_J$ as illustrated in Proposition 1 do not completely eliminate the first order of the bias.

In Proposition 1, the bias of the full-sample estimator $\hat{\beta}$ is assumed to follow a power-series expansion in $\frac{1}{T}$ as illustrated in (48). The same applies for the first subgroup for $\hat{\beta}_1$ whose initial condition, y_0, is also fixed:

$$E(\hat{\beta}_1) = \beta + \frac{\alpha_1}{n_m} + O(n_m^{-2}) \tag{15}$$

However, the difference appears in the subsequent subgroup(s) i, where $i \geq 2$, as the value of the α_1 term is now going to vary for each one of these subgroups. This occurs because now, the α_1 term for these subsamples is depended on its position on the sample, since (14) implies that α_1 is going to vary and depend on the sample size T (Proposition 1 assumes that α_1 is fixed and does not depend on t). Therefore, we have that:

$$E(\hat{\beta}_i) = \beta + \frac{\alpha_{1,i}}{n_m} + O(n_m^{-2}) \ \ \forall\, i \in \ \{2, \ldots, m\} \tag{16}$$

This modification implies that the expectation of the jackknife estimator $\hat{\beta}_J$ now becomes:

$$E(\hat{\beta}_J) = \gamma \left[\beta + \frac{\alpha_1}{T} + O(T^{-2})\right] - \frac{\delta}{m}\left[\beta + \frac{\alpha_1}{n_m} + O(n_m^{-2})\right] - \frac{\delta}{m}\sum_{i=2}^{m} E(\hat{\beta}_i)$$

$$\Rightarrow E(\hat{\beta}_J) = \gamma \left[\beta + \frac{\alpha_1}{T}\right] - \frac{\delta}{m}\left[\beta + \frac{\alpha_1 m}{T}\right] - \frac{\delta}{m}\sum_{i=2}^{m}\left[\beta + \frac{\alpha_{1,i} m}{T}\right] + O(T^{-2})$$

$$\Rightarrow E(\hat{\beta}_J) = [\gamma - \delta]\,\beta + \left[\gamma\alpha_1 - \delta\alpha_1 - \delta\sum_{i=2}^{m}\alpha_{1,i}\right]\frac{1}{T} + O(T^{-2}) \tag{17}$$

Equation (17) implies that $\hat{\beta}_J$ will be unbiased of the first order if the following (now revised) conditions $(i)'$ and $(ii)'$ hold:

$(i)'$ $\gamma - \delta = 1$ (same as condition (i) from the proof of Proposition 1)

$(ii)'$ $\gamma\alpha_1 - \delta\alpha_1 - \delta\sum_{i=2}^{m}\alpha_{1,i} = 0$

Condition $(ii)'$ as shown above is different from condition (ii) of Proposition 1. This indicates that the weights as chosen in Proposition 1 do not completely eliminate the first order of the bias as they indent to.

The first order of the bias has not been removed since there is still a term of $O(T^{-1})$ (due to the $\delta\sum_{i=2}^{m}\alpha_{1,i}$ term in (17)) which still remains.

In other words, as shown from (17) and condition $(ii)'$, Proposition 1 does not work as it intends to as it neglects the $\delta\sum_{i=2}^{m}\alpha_{1,i}$ term which is responsible for the remaining $O(T^{-1})$

[13]From the Functional Central Limit Theorem (see Hamilton (1994) pp. 479-486) we know that: $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\epsilon_t \xrightarrow{L} \sigma W(1)$. Since $(i-1)n_m$ is $O(T)$ we can verify that also $\sum_{k=1}^{(i-1)n_m}\epsilon_k$ is $O_p(T^{1/2})$.

term. The magnitude of this remaining term is going to be examined in the simulation study which follows in Section 4.

A straightforward way of overcoming this problem is by performing a simple alteration to the data generating process (dgp) of those subgroups for which the non-fixed initial values occur. This change will only affect the data generating process of subgroup(s) two and beyond. The form of the jackknife estimator remains as it was defined by (9) and, most importantly, no change is required to the weights shown by Proposition 1 to be $\gamma = \frac{m}{m-1}$ and $\delta = \frac{1}{m-1}$. Specifically, for those subgroups, the data generating process needs to be re-defined, by subtracting the initial values $y_{(i-1)n_m}$ from y_t and adding the initial value y_0. The new variable $\tilde{y}_t$ has the following form:

$$\tilde{y}_t = y_t - y_{(i-1)n_m} + y_0, \quad t \in \{(i-1)n_m + 1, \ldots, in_m\} \;\; \& \;\; i \geq \; 2 \tag{18}$$

The associated data generating process which links $\tilde{y}_t$ with $\tilde{y}_{t-1}$ now has a fixed initial value and is given as the random walk process described in (19):

$$\tilde{y}_t = \tilde{y}_{t-1} + \epsilon_t, \quad t \; \in \{(i-1)n_m + 1, \ldots, m\} \;\; \& \;\; i \geq 2 \tag{19}$$

The formula for the (initial-values-corrected) jackknife estimator, denoted by $\tilde{\beta}_J$, is a function of the modified dependent variable defined in (18), while the weights for γ and δ remain unchanged.

$$\tilde{\beta}_J = \frac{m}{m-1}\hat{\beta} - \frac{1}{m(m-1)}\sum_{i=1}^{m}\tilde{\beta}_i \tag{20}$$

where the least squares estimator remains unchanged, thus as previously: $\hat{\beta} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}$ while for each subgroup i (where i=1, 2,. . . ,m) we have that:

$$\tilde{\beta}_i = \begin{cases} \frac{\sum_{t=1}^{n_m} y_t y_{t-1}}{\sum_{t=1}^{n_m} y_{t-1}^2} & \text{if} \;\; i = 1 \\ \frac{\sum_{t=(i-1)n_m+1}^{in_m} \tilde{y}_t \tilde{y}_{t-1}}{\sum_{t=(i-1)n_m+1}^{in_m} \tilde{y}_{t-1}^2} & \text{if} \quad i \geq 2 \end{cases} \tag{21}$$

For subgroup 1 there is no adjustment required for the initial values, since the initial value y_0 is fixed. Hence, since the first subgroup comprises observations 1 until n_m we have that the subgroup estimator $\tilde{\beta}_1 = \hat{\beta}_1$ emerges from the random walk process in (2) For subgroups i with $i = 2, \ldots, m$ an adjustment for the initial values is required, so $\hat{\beta}_i$ arise from (18) and (19).

To sum up this section, Proposition 1^1 is a revised version of Proposition 1 which takes into account the findings of this subsection to define a correct non-overlapping estimator $\tilde{\beta}_J$ which is unbiased to the first order.

Proposition $1'$ *(A correct non-overlapping jackknife estimator):*

The (initial values corrected) jackknife estimator $\tilde{\beta}_J$ defined by (20) to be a function of the full-sample least squares estimator $\hat{\beta}$ and the the subgroup estimators defined by (21) is biased to the second order $O(T^{-2})$.

3.2 Moving-Blocks Jackknife

Alternatively, a jackknife estimator can be derived by utilizing overlapping (moving) blocks of observations. The moving-blocks jackknife estimator proposed in this section is obtained by moving a length of observations (now denoted by n_M) across a time series by one observation each time and calculate the corresponding subgroup estimators, until the last observation is reached. This way, the moving blocks sub-sampling scheme manages to maintain the dependence structure of a time series sample. This means that each of the subgroups is composed of overlapping observations, with the first observation of subgroup $k+1$ being the second observation in subgroup k as Figure I graphically illustrates.

In contrast with the non-overlapping case, the moving-blocks jackknife uses almost all observations more than once. In a sample of T observations, defining the moving blocks jackknife estimator (for a fixed length n_M), requires $T - n_M + 1$ sub-samples where each moving block/subgroup i includes observation i until observation $n_M + i - 1$.
While the first and the last observations (observation 1 and T respectively) are only used once (the former is used in sub-sample 1 and the latter in sub-sample M), observations 2 until $T-1$ are, in fact, repeated more than once. Indeed, the frequency with which each observation is repeated depends on its position in the time-series, which can be described by the following pattern[14]:

$$\text{Frequency of repetition of j-th observation} \begin{cases} 1 & \text{for j=1} \\ j & \text{for } 2 < j \leq n_M \\ n_M & \text{for } n_M < j \leq T - n_M \\ (T - j + 1) & \text{for } T - n_M < j \leq T - 1 \\ 1 & \text{for } j = T \end{cases}$$

An important implication of the fact that the same observations appear repeatedly in several subgroups is that we anticipate this method will result in smaller standard errors (since the covariance of the overlapping subgroups i and l (for subgroup i and l, with $i \neq l$) is higher).

As in the non-overlapping case, the proposed moving blocks jackknife estimator, $\hat{\beta}_{J,MB}$, is given by the (weighted) difference between the full-sample estimator and the average of the (moving-block) subgroup estimators. Again, the jackknife weights (now denoted by γ' and δ') are chosen in a way so that the first order of the bias is eliminated (as explained in Proposition 2 below).

The proposed moving-blocks jackknife estimator, $\hat{\beta}_{J,MB}$, obtained using M subgroups of overlapping observations each of length n_M, has the following form:

$$\hat{\beta}_{J,MB} = \gamma' \hat{\beta} - \delta' \frac{1}{M} \sum_{i=1}^{M} \hat{\beta}_{i,mb} \tag{22}$$

Again $\hat{\beta}$ denotes the full-sample least squares estimator given in (4), while for each block i, the overlapping subgroup estimator $\hat{\beta}_{i,mb}$ is obtained using observations i until $n_m + i - 1$:

$$\hat{\beta}_i = \frac{\sum_{t=i}^{n_m+i-1} y_t y_{t-1}}{\sum_{t=i}^{n_m+i-1} y_{t-1}^2} \; \forall \; i \;\; \in \{1, 2, ..., T - n_m + 1\} \tag{23}$$

The weights γ' and δ' need not be identical with the non-overlapping weights γ and δ since each moving block has a different composition of observations. The values of the non-overlapping weights which eliminate the leading term of the full-sample estimator's bias are found to be functions of both the subgroups M and the sample size T: $\gamma' = \frac{T}{M-1}$ and $\delta' = \frac{T-M+1}{M-1}$.

Proposition 2 (Moving-Blocks Weights):
The moving-blocks jackknife estimator $\hat{\beta}_{J,MB}$: $\hat{\beta}_{J,MB} = \gamma' \hat{\beta} - \delta' \frac{1}{M} \sum_{i=1}^{M} \hat{\beta}_{i,mb}$ is biased to the second order ($O(T^{-2})$) if and only if the weights γ' and δ' are: $\gamma' = \frac{T}{M-1}$ and $\delta' = \frac{T-M+1}{M-1}$ (where the number of overlapping subgroups denoted by M is an integer number, with $M \geq 2$).

[14] Where j indicates the j-th observation

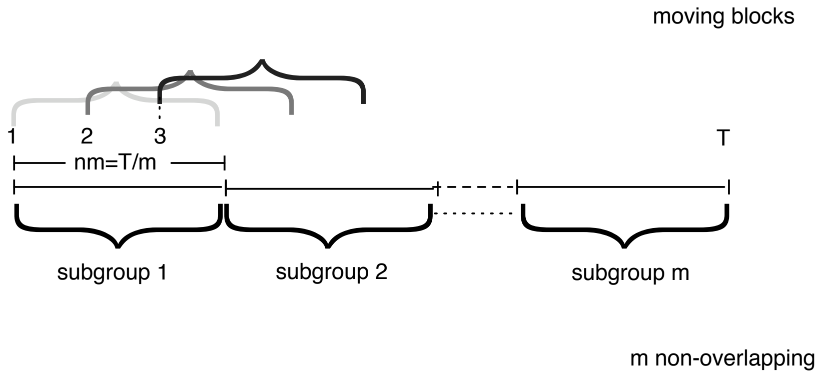

Figure I: Non-overlapping & Moving-Blocks jackknife at a given subgroup length $n_m = n_M$**.**

Therefore, the proposed moving-blocks jackknife estimator can be formally defined in (24) as:

$$\hat{\beta}_{J,MB} = \frac{T}{M-1}\hat{\beta} - \frac{(T-M+1)}{(M-1)}\frac{1}{M}\sum_{i=1}^{M}\hat{\beta}_{i,mb} \tag{24}$$

where $\hat{\beta}$ and $\hat{\beta}_{i,mb}$ are defined in (4) and (23) respectively. (The subgroup length n_M in this case is given by: $n_M = T - M + 1$)
For a given subgroup length $n_m = n_M$, the moving-blocks jackknife requires more subgroups than the non-overlapping groups jackknife (since the former uses almost all observations more than once). Specifically, for a given subgroup length n_m the non-overlapping jackknife can be calculated by using m subgroups (with $m = \frac{T}{n_m}$), whilst the corresponding moving-blocks jackknife requires $M = T - n_m + 1$ (with each subgroup containing more than one observation: $n_m > 1$). It can be easily shown that $M > m$ since $M = T - n_m + 1 = mn_m - n_m + 1 = n_m(m-1)+1$ so that $M > m$ holds $\forall\, n_m > 1$. This can be graphically illustrated by the Figure I. As it will be analysed in a following section, this does not mean that the moving blocks version of the jackknife is computationally more intensive, since when comparing the performance of the row methods we do not have to keep the subgroup length constant.

4 Jackknife Bias Reduction: A Monte Carlo Investigation

In this section, a Monte Carlo Simulation study carried out to assess the performance of the jackknife estimators presented in the previous section as bias reduction tools when the data generating process incorporates a unit root.

The behaviour of the least squares estimator of the correlation coefficient which emerges from non-stationary autoregression has been studied in the literature. Indeed, in the excessive simulation study of Evans & Savin (1981), it has been shown that the least squares estimator is severely biased downwards from its true value of one.

The jackknife (Monte Carlo) expected values will outperform their least-squares counterparts if their expected values exceed the full-sample ones: $E(\tilde{\beta}_J) > E(\hat{\beta})$, which in turn implies that for the non-overlapping jackknife case we wish that:

$$\frac{m}{m-1}E(\tilde{\beta}) - \frac{E(\sum_{i=1}^{m}\tilde{\beta}_i)}{m(m-1)} > E(\hat{\beta})$$

$$\Rightarrow E(\hat{\beta}) > \frac{1}{m}\sum_{i=1}^{m} E(\tilde{\beta}_i) \tag{25}$$

So the jackknife will perform better in the simulations in comparison to its least squares counterparts if the (Monte Carlo) expectation of the full-sample estimator $E(\hat{\beta})$ exceeds the corresponding average expectation of the subgroups: $\frac{1}{m}\sum_{i=1}^{m} E(\hat{\beta}_i)$.
Likewise, for the moving blocks jackknife it applies that $E(\hat{\beta}_{J,MB}) > E(\hat{\beta})$ if:
$E(\hat{\beta}) > \frac{1}{M}\sum_{i=1}^{M} E(\hat{\beta}_{i,mb})$. The latter imply that an alternative way of assessing the performance of the jackknife is, instead of comparing the simulated expectations of the estimators $E(\hat{\beta})$ and $E(\hat{\beta}_J)$, to base simulations on comparing the expectation of the full-sample estimator and the average sum of the expectation of the subgroups, as expression (25) suggests.

4.1 Comparison between the non-overlapping and moving-blocks jackknife estimators: Same Subgroup length

An obvious question arising at this stage is whether the two methods of obtaining the jackknife estimators should be compared in terms of the same subgroup length (and hence allowing the number of subgroups to vary between the two estimators) or in terms of the same subgroup number (therefore allowing for the subgroup length to differ for the two methods). While the former means that we are saving some computational time as the number of subgroups used is set to be identical in both the non-overlapping and moving blocks jackknife), it might however cause the moving blocks estimator to under-perform in comparison with its non-overlapping counterpart. For instance, if one decides to compare the non-overlapping and moving blocks jackknife estimators at a fixed number of subgroups (say two subgroups: $m = M = 2$), this will work for the benefit of the non-overlapping jackknife. In this case, each one of the subgroups used to derive the moving- blocks jackknife will be almost identical, varying only by two observations. As a result, the moving-blocks jackknife estimator $\hat{\beta}_{J,MB}$ will differ very little compared to the original full-sample least squares estimator $\hat{\beta}$. For this reason, it is more appropriate to compare the two methods used to obtain the jackknife estimators (at a given sample size T) using each time identical subgroup length n_m and letting the subgroups used to differ in both number and observations used. This is in line with the literature concerned with optimal sub-series use (for example Carlstein (1986)) who compare their results in terms of same subgroup length.

Once again, let $\tilde{\beta}_J$ and $\hat{\beta}_{J,MB}$ denote the non-overlapping and moving-blocks jackknife estimators respectively which have been defined in section 3.
By fixing the subgroup length to be the same in both cases, we impose the condition shown in (26) below.

$$n_m = \frac{T}{m} = n_M \tag{26}$$

For a fixed subgroup length, the number of subgroups required for the moving blocks estimator will exceed the corresponding subgroups required for the non-overlapping groups jackknife. This can be easily shown from: $M = T - n_m + 1 = mn_m - n_m + 1 = n_m(m-1) + 1$ which, in turn, implies that $M > m$ means that $n_m(m-1) + 1 > m$ which holds for all subgroups which comprise more than one observation, i.e. when $n_m > 1$.

Lemma 1: Let $\tilde{\beta}_J$ and $\hat{\beta}_{J,MB}$ be defined as in (20) and (24) respectively. If the condition in (26) holds, it follows that the jackknife weights for the two estimators (as shown in Propositions 1 & 2) will be equal, that is:
(i). $\frac{m-1}{m} = \frac{T}{M-1}$
and
(ii). $\frac{1}{m-1} = \frac{T-M+1}{M-1}$

Therefore, from Lemma 1 we can conclude that the difference in the values of the jackknife estimators described by equation (20) and (24) lies exclusively on the averaged subgroup sums of the two estimators. The two values, for a given subgroup length n_m, will differ in both the number of subgroups used: m and M, but also in the observations included in each of these subgroups.
The magnitude of this difference will depend on the choices of both m and T. At this stage, the only speculations that we can make about the performance of the two methods of obtaining the jackknife estimators is that $E(\tilde{\beta}_J) > E(\hat{\beta}_{J,MB})$ indicates that:

$$\frac{m}{m-1}E(\hat{\beta}) - \frac{\sum_{i=1}^{m} E(\tilde{\beta}_i)}{m(m-1)} > \frac{T}{M-1}E(\hat{\beta}) - \frac{ME(\sum_{i=1}^{M} \hat{\beta}_{i,mb})}{(M-1)(MT-T+M)} \tag{27}$$

The first terms of each hand side and the weights of the second terms cancel out and therefore the expression simplifies into:

$$\frac{1}{m}E(\sum_{i=1}^{m} \tilde{\beta}_i) < \frac{1}{M}E(\sum_{i=1}^{M} \hat{\beta}_{i,mb}) \tag{28}$$

The analysis above is important for the purposes of the simulation work, as it indicates that differences in the Monte Carlo bias for the two methods of obtaining the jackknife estimator, when we pre-fix the subgroup length to be identical, are merely due to differences on the averaged sum of the subgroup expectations. Instead of basing simulations on the expectation of the jackknife estimators, one can directly conduct a Monte Carlo study directly on the average sums of subgroups which mean that the full-sample estimator will not be required.

4.2 Monte Carlo Results

The Monte Carlo expected values (and hence the corresponding values for the bias) of the jackknife and least squares estimators are calculated using 10^4 replications. The underlying data generating process is a simple linear function so employing as many as 10^4 replications ensures the accuracy of the results without making this experiment time consuming. Initially, we concentrate on the pure random walk process with zero initial values for the simulations, and as the analysis proceeds, some of the assumptions of the original framework are relaxed (e.g zero initial values and Gaussian error terms).

Correcting the bias can be advantageous is various econometric settings. For example, Iglesias & Phillips (2005) show that correcting for bias can lead to finite sample improvements in the context of LM tests in the context of multivariate ARCH models. In addition, Linton (1997) p.p. 475, lists a number of occasions in which bias corrections can be desirable: namely in unit root testing, in cases with large numbers of nuisance parameters and for improving convergence rates in non-parametric and semi-parametric estimation.

The range of the sample sizes examined was carefully chosen to start at a moderate value[15] and gradually increase to higher values. Therefore, starting at a moderate sample of $T = 24$, by doubling the sample size ($T \in \{24,\ 50,\ 100,\ 200,\ 400\}$) we can examine each time this increase in the sample size affects the bias. As discussed earlier, in each case, the subgroup length is restricted to be same for the non-overlapping and moving-blocks jackknife, which allows us to use the virtues of Lemma 1. Phillips and Yu (2005) find that, for the non-overlapping jackknife, in most cases the bias is minimized when only two or three subgroups are used. Based on this, for each sample size T, subject to the rule that $m = \frac{T}{n_m}$, the choices of $m = 2,\ 3,\ 4,\ 5,\ 8$ and 10 are used (for example at $T = 50$ the only suitable choices of subgroups are $m = 2,\ 5$ and 10). Using these choices of non-overlapping subgroups m, the number of subgroups required for the moving-blocks jackknife, M is worked out so that we end up with the same subgroup lengths for the two cases. At a fixed subgroup length, for each sample size T, more subgroups M are required for the moving-blocks jackknife than for the non-overlapping one[16], so still the moving-blocks jackknife will be computationally more demanding (although computational time here is not really an issue in the frameworks studied here).

The simulation results are summarised in Table 1 below. The resulting Monte Carlo values for the bias of the (initial-values corrected) non-overlapping jackknife, the moving-blocks and the full-sample least squares estimator, are shown under the columns for $\tilde{b}_J$, $b_{J,MB}$ and b_T respectively. The high accuracy of the simulations can be verified by observing the values for the least squares bias; these are similar (up to 4 decimal places) to the exact values from Evans & Savin (1981) (Table III p.p. 769).

For each case, the bias decreases as the sample size increases; however even for sample sizes as large as 200 or 400, the bias is still evident. The simulations reveal that the jackknife estimators provide an obvious and substantive reduction to the bias in all cases, compared to their full-sample least squares counterparts. In particular, the bias reduction proves to be especially remarkable in the case of the non-overlapping estimator $\tilde{\beta}_J$. The jackknife bias reduction is particularly evident for small/ moderate samples like 24 and 50, but is also non negligible in larger samples $T = 200$ or $T = 400$ where the original estimator's bias is still present. For example, at $T = 24$ by employing only two subgroups ($m = 2$) the non-overlapping jackknife $\tilde{\beta}_J$ which is defined in equation (20) provides a prominent bias reduction of 79.3% as opposed to the full-sample estimator $\hat{\beta}$. The magnitude of the bias reduction of $\tilde{\beta}_J$ can be better illustrated by the bias ratio $\frac{\tilde{b}_J}{b_T}$ shown in the penultimate column of Table 1. Indeed these ratios are shown to be significantly lower than one which means that the non-overlapping jackknife reduces a substantial amount of the original estimators' bias. For example, when $T = 200$ and $m = 2$, the ratio is merely 0.045 which indicates that only 4% of the original estimators' bias remains in $\tilde{\beta}_J$.

Moreover, the Monte Carlo expectations for the jackknife estimator $\hat{\beta}_J$ which is not adjusted for the initial values (described by (12)) are also shown in the fourth column of Table 1. The interesting aspect of these results arises from the fact that $\hat{\beta}_J$ also provides an important reduction to the initial estimator's bias even though, as shown in the previous section, this estimator is still biased of $O(T^{-1})$. As expected, the bias reduction of this estimator is not as substantial as of that of the initial-values-corrected non-overlapping estimator $\tilde{\beta}_J$. This occurs because the latter is biased of the second order, while in the former there still remains a component which makes the estimator biased to the first order.

[15]The choice of $T = 24$ was preferred to $T = 25$ as the former provides a convenience in the choice of non-overlapping subgroups: at $T = 24$ the values $m = 2,\ 3,\ 4,$ and 8 are examined.

[16]At a given subgroup length $n_m = n_M$, for a given sample size T: $T - M + 1 = \frac{T}{m} \Rightarrow M = \frac{T(m-1)+m}{m}$.

Table 1: Monte Carlo Simulations of the bias of the non-overlapping and moving-blocks jackknife estimators; fixed subgroup length n_m (using 10^4 replications).

T	m	M	n_m [a]	$-100 \times b_J$	$-100 \times \tilde{b}_J$ [b]	$-100 \times b_{J,MB}$	$-100 \times b_T$	$\tilde{b}_J/b_T$	$b_{J,MB}/b_T$
24	2	13	12	3.361	1.346	3.867		0.207	0.593
	3	17	8	3.999	1.973	4.266	6.517	0.303	0.655
	4	19	6	4.428	2.462	4.606		0.378	0.707
	8	22	3	5.513	4.726	5.514		0.725	0.846
50	2	26	25	1.387	0.303	1.672		0.092	0.505
	5	41	10	2.041	0.726	2.139	3.311	0.219	0.646
	10	46	5	2.497	1.414	2.542		0.427	0.768
100	2	51	50	0.744	0.137	0.875		0.079	0.502
	4	76	25	0.914	0.190	0.997	1.743	0.109	0.572
	5	81	20	0.987	0.214	1.050		0.123	0.602
	10	91	10	1.198	0.419	1.229		0.240	0.705
200	2	101	100	0.341	0.040	0.431		0.045	0.485
	4	151	50	0.445	0.071	0.490	0.889	0.080	0.551
	5	161	40	0.473	0.073	0.513		0.082	0.577
	10	181	20	0.570	0.124	0.591		0.139	0.665
400	2	201	200	0.187	0.028	0.222		0.061	0.484
	4	301	100	0.227	0.033	0.252	0.459	0.072	0.549
	5	321	80	0.245	0.034	0.263		0.074	0.573
	10	361	40	0.288	0.046	0.300		0.100	0.654

[a] Here, in order to make the two methods of obtaining the jackknife estimator comparable, the subgroup length n_m is restricted to be the same for the two methods of obtaining the jackknife estimator: $n_m = n_M = \frac{T}{m}$.

[b] The bias terms of the non-overlapping jackknife, the moving-blocks jackknife and the full-sample estimator are defined as $\tilde{b}_J = E(\tilde{\beta}_J) - 1$, $b_{J,MB} = E(\hat{\beta}_{J,MB}) - 1$ and $b_T = E(\hat{\beta}) - 1$ respectively.

At a given subgroup length, even though both methods of sub-sampling provide reductions to the full-sample estimator's bias the non-overlapping jackknife provides us with noticeably better results in terms of bias reduction in every case compared to their moving-blocks counterparts. The difference between the two methods of obtaining the jackknife estimator is noticeable at $m = 2$ for a given sample size.

In fact, at a given sample size, the optimal results for the non-overlapping jackknife are established at $m = 2$ and hence at the largest possible subgroup length n_m. This finding is of considerable importance, as it indicates that by employing only two subgroups, the non-overlapping jackknife provides a computationally simple way of minimising the bias of the original estimator ($\hat{\beta}$ in this case). These results also imply that employing more subgroups does not necessarily provide any improvements in estimation. Additionally, bearing in mind Lemma 1, the simulation study suggests that at $m = 2$, for a given sample size T the average of the expected values of the subgroups' estimates is minimised, which is what drives the bias minimising result to be achieved at two subgroups:

$$\frac{1}{2}\sum_{i=1}^{2} E(\tilde{\beta}_i) < \frac{1}{m'}\sum_{j=1}^{m'} E(\tilde{\beta}_j) \quad \forall \quad \text{integer} \quad m' \geq 3 \tag{29}$$

The best bias reductive results regarding the moving-blocks jackknife are also obtained at the highest subgroup length n_m, which is the same subgroup length at which the best results for the non-overlapping jackknife are obtained (since $n_m = \frac{T}{m}$). The moving-blocks jackknife always provides a reduction to the full-sample bias but its results are not as good as the non-overlapping ones.

At this point, it would be interesting to examine how the performance of the moving blocks jackknife estimator changes with the overlapping subgroups, since generally for a sample size T, we can use every integer M between two and $T - 1$ (unlike the non-overlapping case where we have the restriction of $m = \frac{T}{n_m}$). Therefore, the Monte Carlo bias terms for the moving-blocks jackknife are calculated for every possible subgroup choice M. These results are illustrated in Figures 1-5 in the appendix, with each graph showing the relationship between overlapping subgroups and jackknife bias at a given sample size T.

These figures reveal a that a clear convex relationship exists (particularly clear for small sample sizes $T = 24$ or $T = 50$), between subgroups and moving-blocks jackknife bias, as there exists an optimal subgroup size M where the bias is minimised. Unlike the non-overlapping case where the bias is always minimised at $m = 2$, the best choice of subgroups is not fixed, but it changes for different sample sizes. The reason behind the non-constant optimal subgroup choice M for the moving-blocks jackknife lies within the weights chosen for eliminating the first-order bias. As shown by Proposition 2, these weights for the moving-blocks case are found to be functions of both T and M, whereas the corresponding weights for the non-overlapping jackknife are merely a function of the non-overlapping subgroups m but not T.

4.2.1 Mean Square Error considerations

The jackknife has been designed for bias reduction, as it eliminates the leading term of an expansion for the bias. In many occasions, bias correction is associated with an increase in the mean square error. This subsection examines the performance of the jackknife estimators in terms of another measure of accuracy, namely the mean square error. The mean square error of the estimator $\hat{\beta}$ is defined as its expected square deviation from the true value β shown in (30) below.

$$MSE(\hat{\beta}) = E[(\hat{\beta} - \beta)^2] = Var(\hat{\beta}) + (b_T)^2 \tag{30}$$

Table 2: Mean Square errors of the non-overlapping jackknife, moving blocks jackknife and least square estimators $MSE(\tilde{\beta}_J)$, $MSE(\hat{\beta}_{J,MB})$ and $MSE(\hat{\beta})$ respectively.(using 10^4 replications).

T	m	M	n_m	$MSE(\tilde{\beta}_J)$	$MSE(\hat{\beta}_{J,MB})$	$MSE(\hat{\beta})$	$\frac{MSE(\tilde{\beta}_J)}{MSE(\hat{\beta})}$	$\frac{MSE(\hat{\beta}_{J,MB})}{MSE(\hat{\beta})}$
24	2	13	12	0.0385	0.0214	0.0184	2.09	1.16
	3	17	8	0.0263	0.0173		1.43	0.94
	4	19	6	0.0229	0.0165		1.24	0.90
	8	22	3	0.0265	0.0170		1.44	0.92
50	2	26	25	0.0009	0.0051	0.0046	1.96	1.11
	5	41	10	0.0051	0.0038		1.11	0.83
	10	46	5	0.0044	0.0039		1.13	0.85
100	2	51	50	0.0026	0.0014	0.0013	2.00	1.08
	4	76	25	0.0016	0.0011		1.23	0.85
	5	81	20	0.0014	0.0010		1.08	0.77
	10	91	10	0.0012	0.0011		0.92	0.85
200	2	101	100	0.0007	0.0004	0.0003	2.33	1.33
	4	151	50	0.0004	0.0003		1.33	1.00
	5	161	40	0.0004	0.0003		1.33	1.00
	10	181	20	0.0003	0.0003		1.00	1.00
400	2	201	200	0.0002	9.07×10^{-5}	8.5×10^{-5}	2.35	1.07
	4	301	100	0.0001	6.73×10^{-5}		1.18	0.79
	5	321	80	9.2×10^{-5}	6.60×10^{-5}		1.08	0.78
	10	361	40	7.2×10^{-5}	6.62×10^{-5}		0.85	0.78

From (30) it becomes clear that the MSE captures the trade-off between (possible) reduction to the bias and increase in the variance. The MSE properties of the $\hat{\beta}$ which emerges from a pure random walk setting as in (2) has been previously studied by Abadir (1995) in which the author derives heuristic expansions for both the bias and the MSE of $\hat{\beta}$.
The MSE of the the both the non-overlapping and the moving blocks jackknife, $MSE(\tilde{\beta}_J)$ and $MSE(\hat{\beta}_{J,MB})$, have been calculated and is shown in Table 2, along with the values of the MSE of the full sample estimator $\hat{\beta}$. The simulations reveal that the moving blocks jackknife, is in fact able to reduce the mean square error of the original estimator, subject to an appropriate choice in the subgroups M. Specifically, at a given sample size T, the moving blocks jackknife reduces the MSE of $\hat{\beta}$ as the number of subgroups increases. Unlike the bias case, the optimal subgroups number is not fixed but changes for the range of sample sizes T that are considered here.

5 Deviations from Zero Initial Conditions and Normality

In the previous subsection, a Monte Carlo study was carried out in order to assess the performance of the jackknife estimator. As the underlying process of a random walk with zero initial values might appear to be restrictive, in this section some of the underlying assumptions of the process in (2) are relaxed. The reason of doing this is to ensure that the desirable results of the jackknife do not explicitly rely on any of the properties of the random walk process with zero initial values. Specifically, the robustness of the results is examined for two scenarios: the cases of non-zero initial values and that of non-Gaussian error terms.

5.1 Non-zero Initial Conditions

In this case, we deviate form the random walk process with zero initial values to allow for non-zero initial values: $y_0 \neq 0$. In economic applications, assuming a zero initial value appears to be unrealistic as it implies that the sample used starts at the beginning of the series. Indeed, the ratio $\frac{y_0}{\sigma}$ is highly likely to be non-zero; in the unit root case the y_0 component emerges from the non-stationary series, while the σ component comes from the well-behaved stationary error term. Here we still hold on the assumption that the error term is drawn from the standard normal distribution $\epsilon_t \sim N(0,1)$, so the variance is normalised at one: $\sigma^2 = 1$. Specifically, the cases of $\frac{y_0}{\sigma} = -10,\ -1,\ 1,\ 10$ are considered.

A fixed non-zero initial value provides some information about the process y_t, compared to the zero initial value case; for this reason it is anticipated non-zero initial values will cause the least squares bias to decrease. Also, Evans & Savin (1981) find that as $\frac{y_0}{\sigma}$ increases, the full-sample normalised bias measure $\mu_T = \frac{T}{\sqrt{2}}\left(E(\hat{\beta}-1)\right)$, reaches its limiting distribution value μ_∞ as the initial value increases.

The results are illustrated in Table 3 above. For every case considered, the least squares bias is smaller than the corresponding bias for the zero initial value. The bias is particularly small in the cases where the initial values are $\frac{y_0}{\sigma} = -10,\ 10$ (hence the values for the jackknife in these cases are multiplied by 10^{-3} instead of 10^{-2}). It is important to notice that the values for -1 with 1 and -10 with 10 are not identical: this would only be true if we also replaced the correlation coefficient's true value to be at -1 in the cases for negative initial values. The jackknife provides reduction to the bias in each case, with the magnitude of the reduction being especially evident at $\frac{y_0}{\sigma} = -1$ and $\frac{y_0}{\sigma} = 1$.

Table 3: Non-zero Initial values: $\frac{y_0}{\sigma} = -10,\ -1,\ 1,\ 10$ (Based on 10^4 Replications)

Sample Size	Subgroups	$\frac{y_0}{\sigma} = -10$		$\frac{y_0}{\sigma} = -1$		$\frac{y_0}{\sigma} = 1$		$\frac{y_0}{\sigma} = 10$	
T	m	$-100 \times \tilde{b}_J$	$-100 \times b_T$	$-100 \times \tilde{b}_J$	$-100 \times b_T$	$-100 \times \tilde{b}_J$	$-100 \times b_T$	$-100 \times \tilde{b}_J$	$-100 \times b_T$
24	2	1.169		1.262		1.780		1.088	
	3	1.162	1.083	1.750	6.098	2.423	6.435	1.082	1.015
	4	1.152		2.398		2.790		1.077	
	8	1.138		3.662		4.132		1.068	
50	2	1.019		0.454		0.448		0.922	
	5	1.070	1.050	0.887	3.245	0.965	3.288	0.980	0.968
	10	1.074		1.486		1.557		0.987	
100	2	0.666		0.140		0.150		0.644	
	4	0.776	0.846	0.244	1.721	0.241	1.724	0.749	0.817
	5	0.797		0.279		0.288		0.767	
	10	0.834		0.502		0.512		0.804	
200	2	0.324		0.060		0.0388		0.307	
	4	0.435	0.582	0.079	0.883	0.084	0.883	0.421	0.567
	5	0.465		0.092		0.087		0.452	
	10	0.531		0.159		0.154		0.516	
400	2	0.132		0.030		0.029		0.129	
	4	0.204	0.361	0.041	0.458	0.033	0.456	0.194	0.353
	5	0.225		0.040		0.037		0.218	
	10	0.286		0.058		0.053		0.277	

For negative initial values, the least squares bias is higher than in positive cases; still the bias is smaller than the zero initial value. Finally, the jackknife estimators provide improvements to the bias in all cases; the magnitude of the bias reduction however, is not as significant as in the zero initial value case.

5.2 Non-Gaussian Innovations

The random walk process described in (2) assumes that the error term ϵ_t is normally distributed with mean zero and standard error σ. For the simulation purposes, the error terms were generated from a standard normal distribution as $\epsilon_t \sim N(0,1)$, so $\sigma = 1$. Here, we consider two alternative distributions for the error term, namely the Gamma distribution and the Student-t distribution. The former is a left-skewed (asymmetric) distribution while the latter exhibits fatter tails (i.e difference in the kurtosis) compared to the normal distribution. By examining these alternatives, we can assess how sensitive the performance of the jackknife (and also the full-sample least squares) is to these changes. The initial value (y_0) is kept at zero.

Case 1: $\epsilon_t \sim Gam(\alpha, \beta)$

The assumption of normal error terms is replaced by assuming instead that the error term follows a Gamma distribution: $\epsilon_t \sim Gam(\alpha, \beta)$, where α denotes the shape parameter and β the scale parameter ($\alpha > 0$ and $\beta > 0$)[17]. The expectation and variance are given by[18]:

$$E(\epsilon_t) = \alpha\beta \;\;\&\;\; Var(\epsilon_t) = \alpha\beta^2 \tag{31}$$

The Gamma distribution is widely used in time-series applications for modeling financial or weather situations. The Gamma distribution is used here to examine how (and whether) its asymmetry (left-skewed) and thicker tails will affect the performance of the jackknife method. Specifically, the skewness and kurtosis are functions of the shape parameter α and the corresponding coefficients α_3 and α_4 are given by $\alpha_3 = 2\alpha^{-1/2}$ and $\alpha_4 = 3 + \frac{6}{\alpha}$ respectively. The choice of the scale parameter β in this case is not really crucial as it only affects the first two moments; the shape parameter however determines the asymmetry (how left-skewed the distribution will be) and the thickness of the tails (measured by the kurtosis). In order to make the results comparable with the normally distributed case, the Gamma-distributed error term as been centralised at $E(\epsilon_t) = 0$, by subtracting the expectation term $\alpha\beta$, while the parameters were chosen in a way so that the resulting variance will be equal to one (i.e identical to the standard normal case $\epsilon_t \sim N(0,1)$).

For the simulations' purposes the following cases were examined: $\epsilon_t \sim Gam(1,1)$, $\epsilon_t \sim Gam(\frac{1}{4}, 2)$ and $\epsilon_t \sim Gam(\frac{1}{16}, 4)$, as these choices of parameters give variances equal to one ($\sigma^2 = 1$) in every case. The Monte Carlo results are summarised in Table 4 below. The full-sample estimator's bias b_T in this case is marginally different under Gamma distributed error terms; however it is still severe. As the skewness and kurtosis coefficients α_3 and α_4 increase (hence deviating further for the normal distribution characteristics) the bias b_T decreases. Again, the jackknife estimator provides reduction to the bias of the full-sample estimator for all the choices of α and β which are examined (for every sample size and subgroups). Again, jackknife minimises the bias at two subgroups ($m = 2$). These suggest that replacing the normally distributed assumption, to a distribution which is both skewed and with thicker tails does not affect the performance of the jackknife in terms of bias reduction.

[17] The range of a Gamma distributed variable, say X, is $0 \leq X \leq \infty$

[18] The numerical characteristics of the Gamma Distribution as defined here are taken from Spanos (1999) pp.140.

Table 4: Gamma Distributed Error Terms: $\epsilon_t \sim Gam(1,1)$, $\epsilon_t \sim Gam(\frac{1}{4},2)$ and $\epsilon_t \sim Gam(\frac{1}{16},4)$.

Sample Size	Subgroups	$\epsilon_t \sim Gam(1,1)$		$\epsilon_t \sim Gam(1\frac{1}{4},2)$		$\epsilon_t \sim Gam(\frac{1}{16},4)$	
T	m	$-100 \times \tilde{b}_J$	$-100 \times b_T$	$-100 \times \tilde{b}_J$	$-100 \times b_T$	$-100 \times \tilde{b}_J$	$-100 \times b_T$
24	m=2	1.585		2.073		3.417	
	m=3	2.287	6.094	2.795	4.985	3.968	3.258
	m=4	2.763		3.214		4.283	
	m=8	4.754		4.625		4.540	
50	m=2	0.501		0.774		1.342	
	m=5	1.009	3.204	1.374	2.810	2.146	2.239
	m=10	1.628		1.983		2.673	
100	m=2	0.124		0.318		0.467	
	m=4	0.241	1.671	0.451	1.596	0.754	1.328
	m=5	0.286		0.518		0.859	
	m=10	0.529		0.799		1.203	
200	m=2	0.061		0.0935		0.153	
	m=4	0.088	0.878	0.160	0.851	0.262	0.737
	m=5	0.114		0.191		0.314	
	m=10	0.182		0.290		0.466	
400	m=2	0.019		0.0101		0.047	
	m=4	0.031	0.451	0.0343	0.422	0.087	0.395
	m=5	0.042		0.0396		0.105	
	m=10	0.061		0.083		0.167	
Numerical characteristics:		$\alpha_3 = 2,\ \alpha_4 = 9$		$\alpha_3 = 4,\ \alpha_4 = 27$		$\alpha_3 = 8,\ \alpha_4 = 99$	

Case 2: Student t Distribution: $\epsilon_t \sim Stt(n)$

The student-t distribution resembles the normal distribution in terms of shape (bell-shaped); however compared to the normal distribution, it exhibits fatter tails (this is due to the difference in the 4th moments i.e. the kurtosis). The thickness of the tails depends on the shape parameter of the t-distributed random variable; the shape parameter n is the degrees of freedom. As n increases, the fatness of the tail decreases as the t-distribution is getting better approximated by the normal distribution. Additionally, the expected value and variance are given as follows:

$$E(\epsilon_t) = 0 \;\; and \;\; Var(\epsilon_t) = \frac{n}{n-2} \; \forall \; n > 2 \tag{32}$$

(where n denotes the degrees of freedom) while the skewness and kurtosis are given by α_3 and α_4 respectively.

$$\alpha_3 = 0 \;\; \& \;\; \alpha_4 = \frac{3n-6}{n-4} \tag{33}$$

The Monte Carlo bias of the (non-overlapping) jackknife and full-sample least squares estimators were calculated for the case where the underlying process is an $AR(1)$ process with a unit root and t-distributed random errors (and zero initial values). Three different choies of degrees of freedom were considered: $\epsilon_t \sim Stt(5)$, $\epsilon_t \sim Stt(10)$ and $\epsilon_t \sim Stt(20)$.

The simulation results are summarised in Table 5 below. For thicker tails (i.e. small degrees of freedom $n = 5$ or $n = 10$) the least squares bias b_T is smaller compared to the normally distributed case. As the degrees of freedom increase, the t-distribution is getting better approximated by the standard normal distribution (the tail thickening decreases) and hence the least squares bias increases. For all cases, the full-sample estimator appears to be underestimated, and the jackknife provides a reduction to the bias compared to the corresponding original estimator's counterparts. Again, the bias-minimising non-overlapping subgroup size is found at two subgroups ($m = 2$).

Table 5: Student-t Distributed error terms: $\epsilon_t \sim Stt(n)$ with $n = 5,\ 10,\ 20$ (n denotes the degrees of freedom).

Sample Size	Subgroups	$\epsilon_t \sim Stt(5)$		$\epsilon_t \sim Stt(10)$		$\epsilon_t \sim Stt(20)$	
T	m	$-100 \times \tilde{b}_J$	$-100 \times b_T$	$-100 \times \tilde{b}_J$	$-100 \times b_T$	$-100 \times \tilde{b}_J$	$-100 \times b_T$
24	2	1.541		1.290		1.398	
	3	2.203	6.441	1.851	6.459	2.033	6.540
	4	2.672		2.506		2.551	
	8	4.663		4.518		4.665	
50	2	0.332		0.289		0.347	
	5	0.869	3.272	0.814	3.306	0.808	3.366
	10	1.513		1.468		1.467	
100	2	0.113		0.123		0.125	
	4	0.228	1.715	0.215	1.731	0.238	1.760
	5	0.286		0.257		0.275	
	10	0.470		0.456		0.453	
200	2	0.046		0.039		0.042	
	4	0.062	0.874	0.052	0.885	0.071	0.889
	5	0.081		0.066		0.075	
	10	0.138		0.127		0.126	
400	2	0.019		.0129		0.023	
	4	0.022	0.441	0.016	0.443	0.024	0.449
	5	0.024		0.022		0.019	
	10	0.040		0.032		0.039	
Numerical Characteristics:		$\alpha_3 = 0,\ \alpha_4 = 9$		$\alpha_3 = 0,\ \alpha_4 = 4$		$\alpha_3 = 0,\ \alpha_4 = 3.4$	

6 The AR(1) model with deterministic components

This section complements the analysis by allowing for the presence of deterministic components in the regression model. This indicates that, unlike in (2), the vector of deterministic components $A^{'}x_t$ in (1) is now non-zero. These settings correspond to more realistic situations in which a constant term and/or a linear trend are incorporated in the regression model. Specifically, for the case where the process includes a constant term but no linear trend, it follows that A is equal to a scalar α and $x_t = 1$, while for the case where both a constant term and a linear time trend are included we have that $A = [\alpha, \delta]^{'}$ and $x_t = [1, t]^{'}$. It has been shown in the literature, for example Andrews (1993), that the presence of a constant and/or a linear trend in non-stationary autoregression accentuates the bias of the least squares estimator of β. An additional virtue of including such deterministic terms in the regression model, as this section shows, is that the associated least squares estimators of β become invariant with respect to the initial conditions. Therefore, the implementation of the jackknife in such settings is even more straightforward as no adjustments to the initial conditions are required.

6.1 True model is a pure random walk; Regression model includes a constant term.

In the first case considered, the true process remains that of a pure random walk shown in (2); however, now, a constant term is also included in the regression model as shown in (34) below.

$$y_t = \alpha + \beta y_{t-1} + \epsilon_t \tag{34}$$

At first sight, incorporating a constant in the regression model seems to be merely a minor specification error. However, the least squares coefficient of β which emerges from this framework has a different behaviour not only asymptotically, but also in finite samples, compared to its counterpart which comes from the setting examined until now.

6.1.1 Least Squares estimation of $(\alpha, \beta)^{'}$:

Using ordinary least squares, the model in (34) is estimated to obtain the least squares estimators of both α and β, namely $(\hat{\alpha}, \hat{\beta})^{'}$. This is given by (35) below; these are identical to the case in which both the true and regression models incorporate a constant (Case 2 in Hamilton (1994), p.p 490-495).

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^{T} y_{t-1} \\ \sum_{t=1}^{T} y_{t-1} & \sum_{t=1}^{T} y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} y_t \\ \sum_{t=1}^{T} y_t y_{t-1} \end{bmatrix} \tag{35}$$

However now, when the estimation error is defined, i.e the deviation of the OLS estimators from their true value, the true model under which $\alpha = 0$ and $\beta = 1$, needs to be taken into account. Therefore, for this case it follows that:

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} - 1 \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^{T} y_{t-1} \\ \sum_{t=1}^{T} y_{t-1} & \sum_{t=1}^{T} y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} \epsilon_t \\ \sum_{t=1}^{T} y_{t-1} \epsilon_t \end{bmatrix} \tag{36}$$

In the present framework, $\hat{\alpha}$ and $\hat{\beta}$ converge at different rates, namely the former at rate $T^{1/2}$ and the latter at rate T.

$$\left[\begin{array}{c} T^{1/2}\hat{\alpha} \\ T(\hat{\beta}-1) \end{array}\right] \xrightarrow{L} \frac{1}{\Delta}\left[\begin{array}{c} \sigma\left\{W(1)\int_0^1 W(r)^2 dr - \frac{1}{2}\left\{W(1)^2-1\right\}\int_0^1 W(r)dr\right\} \\ \frac{1}{2}\left\{W(1)^2-1\right\} - W(1)\int_0^1 W(r)dr \end{array}\right] \tag{37}$$

where Δ denotes the determinant $\Delta = \int_0^1 [W(r)]^2\, dr - \left[\int_0^1 W(r)dr\right]^2$

In line with the pure random walk case examined in the previous sections [19], the estimated coefficient of the lagged dependent variable $\hat{\beta}$ converges to its true value at rate T; however the limiting distribution (shown below) is now consistent with a demeaned Wiener process. This, in turn, implies that under this framework different critical values are required.

$$T(\hat{\beta}-1) \xrightarrow{L} \frac{\frac{1}{2}\left\{W(1)^2-1\right\} - W(1)\left[\int_0^1 W(r)dr\right]}{\int_0^1 [W(r)]^2\, dr - \left[\int_0^1 W(r)dr\right]^2} \tag{38}$$

The proposed jackknife estimator which employs m non-overlapping subgroups (each of which contain n_m observations), B_J, is a 2×1 vector defined in (39).

$$B_J = \left[\begin{array}{c} \hat{\alpha}_J \\ \hat{\beta}_J \end{array}\right] = \frac{m}{m-1}\left[\begin{array}{c} \hat{\alpha} \\ \hat{\beta} \end{array}\right] - \frac{1}{(m-1)}\frac{1}{m}\left[\begin{array}{c} \sum_{i=1}^m \hat{\alpha}_i \\ \sum_{i=1}^m \hat{\beta}_i \end{array}\right] \tag{39}$$

where $(\hat{\alpha},\hat{\beta})'$ is defined in (35) and, in the same fashion, for every subgroup i the corresponding least squares estimators are given in (40) below.

$$\left[\begin{array}{c} \hat{\alpha}_i \\ \hat{\beta}_i \end{array}\right] = \left[\begin{array}{cc} n_m & \sum_{t=(i-1)n_m+1}^{in_m} y_{t-1} \\ \sum_{t=(i-1)n_m+1}^{in_m} y_{t-1} & \sum_{t=(i-1)n_m+1}^{in_m} y_{t-1}^2 \end{array}\right]^{-1} \left[\begin{array}{c} \sum_{t=(i-1)n_m+1}^{in_m} y_t \\ \sum_{t=(i-1)n_m+1}^{in_m} y_t y_{t-1} \end{array}\right] \tag{40}$$

An interesting feature which occurs in the present case is that the least squares estimators $\hat{\beta}$ and subgroup estimators $\hat{\beta}_i$s (where $i = 1, 2, \ldots, m$), are invariant with respect to the initial conditions y_0 and $y_{(i-1)n_m}$ respectively. This indicates that also the corresponding limiting distributions are not affected by these initial conditions. This, in turn, implies that when a constant is included in the (regression) model no adjustment to the subgroups' data generation process is required. The same result applies for the case where both the regression and true model include a constant term (see Kiviet and Phillips (2005) for a thorough examination of this case).

This result is also of practical importance as it suggests that the jackknife estimator defined in (39) can be directly used for empirical work. These findings are summarised by Propositions 3 and $3'$ which follow.

Proposition 3:

The least squares estimator of β, $\hat{\beta}$, which emerges from the framework described by (2) as the true process and the regression model in (34) is invariant of the initial condition(s) y_0 (where y_0 is assumed to be fixed).

Proposition $3'$:

The jackknife estimator $\hat{\beta}_J$ shown in (39) which arises from a regression model with a constant term shown in (34) requires no adjustments to the initial-values.

[19]The pure random walk case is consistent with Case 1 in Hamilton (1994) p.p. 487-490, while the case examined in this subsection is Case 2 in Hamilton (1994) p.p. 490-495.

6.2 Constant term and linear trend in the regression model; True process is a random walk with or without a drift.

An alternative case to the one discussed in the last subsection, is the case where the regression model, in addition to the constant term, also incorporates a linear trend. This occurs when one wishes to model a variable which seems to have a gradual movement in time. The Gross national product (GNP) of developed economies is a prime example of such case as it is shown to exhibit a linear trend across time. In this setting, the true model is a stationary process with a possible drift term (depending on whether the mean of the series is zero or not). Equations (41) and (42) refer to the real and regression models respectively.

$$y_t = \alpha + y_{t-1} + \epsilon_t \tag{41}$$

$$y_t = \alpha + \beta y_{t-1} + \delta t + \epsilon_t \tag{42}$$

If $\alpha \neq 0$, then y_{t-1} is asymptotically equivalent to a time trend (which is already separately included as an explanatory variable in the regression model). A transformation on the regression model helps to avoid the model to become collinear. The regression model can be re-written as follows:

$$y_t = \alpha^* + \beta^* \xi_{t-1} + \delta^* t + \epsilon_t \tag{43}$$

where $\alpha^* = (1-\beta)\alpha$, $\beta^* = \beta$, $\delta^* = (\delta + \beta\alpha)$ and $\xi_t = y_t - \alpha t$.
Under the null hypothesis (i.e the true process) it follows that $\beta = 1$ and $\delta = 0$. Hence, in the transformed regression model, the null hypothesis implies that $\beta^* = 1$, $\alpha^* = 0$ and $\delta^* = \alpha$.

6.2.1 Least squares estimation of $(\alpha^*, \beta^*, \delta^*)'$

$$\begin{bmatrix} \hat{\alpha}^* \\ \hat{\beta}^* - 1 \\ \hat{\delta}^* - \alpha \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^T \xi_{t-1} & \sum_{t=1}^T t \\ \sum_{t=1}^T \xi_{t-1} & \sum_{t=1}^T \xi_{t-1}^2 & \sum_{t=1}^T \xi_{t-1} t \\ \sum_{t=1}^T t & \sum_{t=1}^T t\xi_{t-1} & \sum_{t=1}^T t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^T \epsilon_t \\ \sum_{t=1}^T \xi_{t-1}\epsilon_{t-1} \\ \sum_{t=1}^T t\epsilon_t \end{bmatrix} \tag{44}$$

Each element of the vector $(\alpha^*, \beta^*, \delta)'$ converges at different rates, namely $T^{1/2}$, T and $T^{3/2}$ respectively. This indicates that, also in this case the lagged dependent variable's coefficient converges to its true value of one at rate T. The limiting distribution is expressed in terms of a standard Brownian motion $W(.)$ as shown by (45) below.

$$\begin{bmatrix} T^{1/2}\hat{\alpha}^* \\ T(\hat{\beta} - 1) \\ T^{3/2}(\hat{\delta} - \alpha) \end{bmatrix} \rightarrow_L \begin{bmatrix} \sigma & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sigma \end{bmatrix} \begin{bmatrix} 1 & A & 1/2 \\ A & C & B \\ 1/2 & B & 1/3 \end{bmatrix}^{-1} \times \begin{bmatrix} D \\ \frac{1}{2}\left\{D^2 - 1\right\} \\ D - A \end{bmatrix} \tag{45}$$

where $A = \int W(r)dr$, $B = \int rW(r)dr$, $C = \int [W(r)]^2 dr$ and $D = W(1)$.
The limiting distribution of $T(\hat{\beta} - 1)$ is independent of the innovation term's variance σ^2. Also, the limiting distribution of all three elements of $(\alpha^*, \beta^*, \delta)'$ is invariant with respect to the value of the constant term α.

6.2.2 Jackknife estimation

The jackknife estimator of the parameter vector $(\alpha^*, \beta^*, \delta^*)'$, which employs m adjacent subgroups, is the 3×1 vector B_J given by (46) below.

$$B_J = \begin{bmatrix} \hat{\alpha}_J^* \\ \hat{\beta}_J^* = \hat{\beta}_J \\ \hat{\delta}_J^* \end{bmatrix} = \frac{m}{m-1} \begin{bmatrix} \hat{\alpha}^* \\ \hat{\beta}^* = \hat{\beta} \\ \hat{\delta}^* \end{bmatrix} - \frac{1}{m(m-1)} \begin{bmatrix} \sum_{i=1}^m \hat{\alpha}_i^* \\ \sum_{i=1}^m \hat{\beta}_i^* \\ \sum_{i=1}^m \hat{\delta}_i^* \end{bmatrix} \tag{46}$$

where for every subgroup i, ($i = 1, 2, \ldots, m$), it holds that:

$$\begin{bmatrix} \hat{\alpha}_i^* \\ \hat{\beta}_i^* \\ \hat{\delta}_i^* \end{bmatrix} = \begin{bmatrix} n_m & \sum_{t=(i-1)n_m+1}^{in_m} \xi_{t-1} & \sum_{t=(i-1)n_m+1}^{in_m} t \\ \sum_{t=(i-1)n_m+1}^{in_m} \xi_{t-1} & \sum_{t=(i-1)n_m+1}^{in_m} \xi_{t-1}^2 & \sum_{t=(i-1)n_m+1}^{in_m} \xi_{t-1} t \\ \sum_{t=(i-1)n_m+1}^{in_m} t & \sum_{t=(i-1)n_m+1}^{in_m} t\xi_{t-1} & \sum_{t=(i-1)n_m+1}^{in_m} t^2 \end{bmatrix}^{-1} \times \tag{47}$$

$$\times \begin{bmatrix} \sum_{t=(i-1)n_m+1}^{in_m} y_t \\ \sum_{t=(i-1)n_m+1}^{in_m} \xi_{t-1} y_t \\ \sum_{t=(i-1)n_m+1}^{in_m} t y_t \end{bmatrix}$$

The Monte Carlo simulation results regarding the bias of the estimated correlation coefficient of β are summarised in Table 6 below. The first interesting feature is that for both cases examined here, the least squares bias, b_T is noticeably higher as opposed to the pure random walk case shown in Table 1. In particular, the presence of both a constant and a trend in the regression model is shown to shoot up the least squares bias, especially in samples of moderate size such as $T = 24$ of $T = 50$. These results can be also verified by looking at Tables II and III of Andrews (1993) which illustrate the quantiles of the least squares estimator in autoregression with deterministic components. The jackknife values for the bias are calculated for the same choices of sample size T and subgroups m as in the simulation work done for the pure random walk case. These results confirm the ability of the (non-overlapping) jackknife to reduce bias in the presence of deterministic components in the regression model. Indeed, the jackknife manages to provide substantial reductions to the bias in these settings. The magnitude of the bias reduction is captured by the bias ratio $\frac{b_J}{b_T}$, which is found to be even lower than its counterparts from Table 1. To sum up, the results of Table 6 indicate that the jackknife is robust to the presence of a constant term and/or a linear trend in the regression model.

Table 6: Deterministic Components in the Regression model (based on 10^4 repl.)

		Constant term in regression			Constant term & Linear trend in regression		
T	m	$-100 \times b_J$	$-100 \times b_T$	b_J/b_T	$-100 \times b_J$	$-100 \times b_T$	b_J/b_T
24	2	3.710	19.670	0.19	9.439	36.748	0.26
	3	5.294		0.27	12.49		0.34
	4	6.546		0.33	14.73		0.40
	8	10.12		0.51	14.55		0.40
50	2	0.942	10.054	0.09	2.384	18.835	0.13
	5	2.259		0.22	5.273		0.28
	10	3.807		0.38	8.242		0.44
100	2	0.315	5.190	0.06	0.696	9.786	0.07
	4	0.531		0.10	1.284		0.13
	5	0.640		0.12	1.546		0.16
	10	1.878		0.12	2.739		0.28
200	2	0.126	2.657	0.05	0.117	4.966	0.02
	4	0.175		0.07	0.314		0.06
	5	0.194		0.07	0.403		0.08
	10	0.352		0.13	0.772		0.16
400	2	0.044	1.354	0.03	0.099	2.539	0.04
	4	0.067		0.05	0.119		0.05
	5	0.074		0.05	0.129		0.05
	10	0.114		0.08	0.239		0.09

7 Concluding Remarks

This paper examined the performance of the jackknife resampling technique in $AR(1)$ models with a unit root. Two methods of sub-sampling for obtaining the jackknife estimator are examined, namely the non-overlapping jackknife and the moving-blocks jackknife. The jackknife bias reduction is obtained by eliminating the leading term of a power-series expansion. This paper showed that the same non-overlapping estimator used by Phillips & Yu (2005) can be also used in non-stationary situations, only in the latter case within the pure random walk framework, an alteration in the subgroups' data generating process is required in order to prevent the initial values from becoming explosive. Another way of overcoming this issue is by including a constant and/or a linear trend into the regression model, as shown in section 6 and by propositions 3 and $3^{'}$.

Once the appropriate weights (and hence functions) for the non-overlapping and moving-blocks jackknife are derived, the next step was to assess the performance of the two methods of deriving the jackknife estimator with their full-sample counterparts. The main findings of the Monte Carlo study of Section 4 reveal that the jackknife provides a substantive reduction to the bias of the original estimator. Most importantly, the non-overlapping jackknife, which performs better than the moving-blocks one, minimises the bias (at a given sample size) when only two subgroups are used in deriving the corresponding jackknife estimator. This result is in line with the results of Phillips & Yu (2005), who in their complicated stationary time series application, find that the bias is minimised at two or three subgroups. This implies that the non-overlapping jackknife is a computationally simple bias reduction method to be used in unit root cases. What is also of interest to notice from the simulation results is that the non-overlapping jackknife $\tilde{\beta}_J$ before the initial values correction might still comprise a $O(T^{-1})$ term, however it performs better in terms of bias reduction as opposed to the moving-blocks jackknife. However, the latter performs better in terms of the mean square error (MSE), rather than the bias, as it manages to reduce the MSE of the original estimator. As the analysis proceeded, some of the assumptions of the pure random walk case in (2) were relaxed. Indeed, the jackknife was found to be robust to non-zero initial values, non-Gaussian innovations and to the presence of deterministic components. A general recommendation which arises from the results of this paper is that in situations where the optimality condition requires bias corrections, the non-overlapping jackknife (with two subgroups) should be used, while in cases where the MSE needs to be reduced the moving blocks jackknife can be used.

A possible shortcoming of the non-overlapping jackknife as opposed to the moving-blocks one, is that the former (as defined in section 3) lacks some flexibility on the subgroups choice as the restriction that $m = \frac{T}{n_m}$ means that at a given sample size T, it might be difficult to find many subgroup choices. Also, for some sample sizes, it might not be possible to split the sample in half, e.g if $T = 2 \times x - 1$ and x is an integer number. What one can do in situations like this, is to adjust the subsamples to the given sample size, possibly by ending up with uneven (in size) subsamples. In fact, theorem 3 in Chambers (2009) provides a solution to this problem as the author proposes a jackknife estimator which utilises sub-samples of different length.

To conclude, there exist countless extensions that one can pursue from here. The results of these paper open new avenues of research in which the jackknife can be employed as a simple method to improve estimation in non-stationary dynamic situations at very little computational cost. A fairly straightforward extension would be to examine how these gains in bias can be used for inferential purposes, in unit root testing.

8 References

1. Abadir, K. M. (1993) OLS Bias in a Nonstationary Autoregression. *Econometric Theory* Vol.9, No. 1, 81-93.

2. Abadir, K. M. (1995) Unbiased estimation as a solution to testing for random walks. *Economics Letters* Vol. 47, 263-268.

3. Andrews, D. W. K. (1993) Exacly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica* Vol. 61, No. 1, pp. 139-165.

4. Carlstein, E. (1986) The Use of Subseries Values for Estimating the Variance of a General Statistic from a Stationary Sequence. *The Annals of Statistics* Vol. 14, No. 13. 1171-1179.

5. Chambers, M. J. (2009) Jackknife Estimation and Inference in Stationary Autoregressive Models. *Working Paper* .

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

7. Efron, B. (1982) The Jackknife, the Bootstrap and Other Resampling Plans. *Society for Industrial and Applied Mathematics.*

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

9. Hahn, J. & Newey, W. (2004) Jackknife and Analytical Bias Reduction for Nonlinear Panel Models. *Econometrica* Vol. 72, No. 4, 1295-1319.

10. Hamilton, J. D. (1994) Time Series Analysis. *Princeton University Press.*

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

12. Iglesias, E. M. and G. D. A. Phillips (2005) Bivariate ARCH Models: Finite-Sample properties of QML estimators and an application to an LM-type test. *Econometric Theory* Vol. 21, 1058-1086.

13. Kendal, M. G. (1954) Note on Bias in the Estimation of Autocorrelation. *Biometrika* Vol. 41, No. 3/4, 403-404.

14. Kiviet, J. F and G. D. A. Phillips (2005) Moment approximation for least-squares estimators in dynamic regression models with a unit root. *Econometrics Journal* Vol. 8, pp. 115-142.

15. Kuensch, H. R. (1989) The Jackknife and the Bootstrap for General Stationary Observations *The Annals of Statistics* Vol. 17, No. 13, 1217-1241.

16. Linton, O. (1997) An asymptotic expansion in the Garch(1,1) Model. *Econometric Theory* Vol. 13, No. 4, 558-581.

17. Miller, R. G. (1974) The Jackknife–A Review. *Biometrika* Vol. 61, No. 1. 1-15.

18. Phillips, P. C. B. (1987a) Time Series Regression with a Unit Root. *Econometrica* Vol. 55, No. 2, 277-301.

19. Phillips, P. C. B. (1987b) Asymptotic Expansions in Nonstationary Vector Autoregressions. *Econometric Theory* Vol.3, 44-68.

20. Phillips, P. C. B. and Yu, J. (2005) Jackknifing Bond Option Prices. *The Review of Financial Studies* Vol. 18, No. 2, 707-742.

21. Phillips, P. C. B. and Yu, J. (2007) Maximum Likelihood and Gaussian Estimation of Continuous Time Models in Finance. *Cowles Foundation Discussion Paper No. 1597.*

22. Shao, J. & Tu, D. (1995) The Jackknife and the Bootstrap. *Springer Series in Statistics.*

23. Spanos, A. (1999) Probability Theory and Statistical Inference: econometric modeling with Observational Data. *Cambridge University Press.*

24. Quenouille, M. H. (1956) Notes on Bias in Estimation. *Biometrika* Vol. 43, 353-360.

25. Tanaka, K. (1983) Asymptotic Expansions Associated with the AR(1) Model with Unknown Mean. *Econometrica* Vol. 51, No. 4, 1221-1232.

26. Tukey, J. W. (1958) Bias in Confidence in Not-Quite Large Samples. (abstract)*Ann. Math. Statist.*Vol. 29, 614.

27. White, J. S. (1958) The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case. *Annals of Mathematical Statistics* Vol. 26, 1188-1197.

28. White, J. S. (1961) Asymptotic Expansions for the Mean and Variance of the Serial Correlation Coefficient. *Biometrika* Vol. 48, No. 1/2, 85-94.

A Proofs

Proof of Proposition 1:

By assumption, the bias of the full-sample estimator $\hat{\beta}$ follows a power-series expansion in $\frac{1}{T}$ as shown in (48) below[20].

$$E(\hat{\beta}) = \beta + \frac{\alpha_1}{T} + O(T^{-2}) \tag{48}$$

Likewise, the bias of each of the (non-overlapping) subgroup estimators $\hat{\beta}_i$ follows a power-series expansion in $\frac{1}{n_m}$, (where $n_m = \frac{T}{m}$) which has the following form:

$$E(\hat{\beta}_i) = \beta + \frac{\alpha_1}{n_m} + O(n_m^{-2})$$

Therefore, the expectation of the proposed jackknife estimator $\hat{\beta}_J$, denoted as $E(\hat{\beta}_J)$, is as follows:

$$E(\hat{\beta}_J) = \gamma(\beta + \frac{\alpha_1}{T}) - \delta\left(\beta + \frac{\alpha_1 m}{T}\right) + O(T^{-2})$$

By rearranging the terms we get:

$$\Rightarrow E(\hat{\beta}_J) = (\gamma - \delta)\beta + \alpha_1\left(\frac{\gamma}{T} - \frac{\delta m}{T}\right) + O(T^{-2})$$

Then, $\hat{\beta}_J$ will unbiased to the first order if the following two conditions hold:
(i) $\gamma - \delta = 1$

[20] $E(.)$ denotes the expectations operator, whilst $O(.)$ refers to the "big- O" notation and T is the sample size. Also, the parameter α_1 does not depend on the sample size T.

(ii) $\frac{\gamma}{T} - \frac{\delta m}{T} = 0$
From (i), it follows that $\gamma = \delta + 1$ By substituting this into (ii) we obtain that $\frac{\delta+1}{T} - \frac{\delta m}{T} = 0$ $\rightarrow \quad \delta + 1 - \delta m = 0 \quad \rightarrow \delta = \frac{1}{m-1}$.

Finally, the value for γ in terms of m can be derived by substituting the value for δ into (i). This gives us $\gamma = \frac{m}{m-1}$.//

Proof of Proposition $1'$:

This proposition can be proofed in a very straightforward way by looking at the modified variable $\tilde{y}_t$ defined by (18). This modification is implemented to fix the initial conditions of subgroup(s) $i \geq 2, \ldots, m$ at y_0. From (18), the associated initial conditions at time point $t = (i-1)n_m$ are:

$$\tilde{y}_{(i-1)n_m} = y_{(i-1)n_m} - y_{(i-1)n_m} + y_0$$

This leaves that:

$$\tilde{y}_{(i-1)n_m} = y_0$$

Since the initial conditions are now fixed (at zero in the main framework), it follows that for every non-overlapping subgroup estimator $\tilde{\beta}_i$ with $i = 1, 2, \ldots, m$, the power series expansions in $\frac{1}{n_m}$ used to prove proposition 1 are also restored, that is:

$$E(\tilde{\beta}_i)\beta + \frac{\alpha_1}{n_m} + O(n_m^{-2})$$

The bias of the full sample estimator $\hat{\beta}$ follows a power-series expansion in $\frac{1}{T}$, so the jackknife weights illustrated in proposition 1 can, now, be successfully employed.

Proof of Proposition 2:

Since the bias of the full-sample estimator $\hat{\beta}$ follows a power-series expansion in $\frac{1}{T}$, this implies that, in the same manner, the bias of the sub-group estimator $\hat{\beta}_{i,mb}$ follows a power-series expansion in $\frac{1}{n_M}$, which has the following form:

$$E(\hat{\beta}_{i,mb}) = \beta + \frac{\alpha_1}{n_M} + O(n_M^{-2}) \tag{49}$$

From (22) and (49), it follows that the expected value of $\hat{\beta}_{J,MB}$, $E(\hat{\beta}_{J,MB})$, is as follows:

$$E(\hat{\beta}_{J,MB}) = \gamma'(\beta + \frac{\alpha_1}{T}) - \delta' \left[\beta + \frac{\alpha_1}{n_M}\right] + O(T^{-2})$$

By rearranging the terms we obtain that:

$$\Rightarrow E(\hat{\beta}_{J,MB}) = (\gamma' - \delta')\beta + \alpha_1 \left(\frac{\gamma'}{T} - \frac{\delta'}{n_M}\right) + O(T^{-2})$$

Then, $\hat{\beta}_{J,MB}$ will be unbiased to the first order if the following two conditions hold:
(iii) $\gamma' - \delta' = 1$
(iv) $\frac{\gamma'}{T} - \frac{\delta'}{n_M} = 0$
Substituting (iii) into (iv) we obtain: $\frac{\delta'+1}{T} - \frac{\delta'}{T-M+1} = 0 \Rightarrow \quad (\delta'+1)(T-M+1) = \delta' T \ \Rightarrow \delta' = \frac{T-M+1}{M-1}$
Finally, the value for γ' in terms of M can be derived by substituting the value for δ' into (iii). This gives us that $\gamma' = \frac{T-M+1}{M-1} + 1 \Rightarrow \gamma' = \frac{T}{M-1}$.//

Proof of Lemma 1:

We can notice that the first terms of equations (20) and (24) are weighted values of the full-sample estimator $\hat{\beta}$. Even though at first glance the weights do not seem to be identical for a given n_m , this can be easily shown by substituting (26) into the weight of the first term of equation (20) as:

$$\frac{m}{m-1} = \frac{\frac{T}{n_m}}{\frac{T}{n_m}-1} = \frac{T}{T-n_m} = \frac{T}{T-(T-M+1)} = \frac{T}{M-1} \tag{50}$$

This implies that, for a given subgroup length n_m, the difference in the estimators (in expectation terms) described by equations (20) and (24) will rely exclusively on the extent by which the second terms of the equations differ. The second terms are in turn, the weighted averaged sums of the subgroup estimators: $\frac{1}{m}\sum_{i=1}^{m}\tilde{\beta}_i$ and $\frac{1}{M}\sum_{i=1}^{M}\hat{\beta}_{i,mb}$
Additionally, the weights of these averaged subgroup sums are in fact equal, as we can see by substituting $M = T - n_m + 1$ and equation (26) into the moving blocks subgroups weight $\frac{T-M+1}{M-1}$:

$$\frac{T-M+1}{M-1} = \frac{T-(T-n_m+1)+1}{T-n_m+1-1} = \frac{n_m}{T-n_m} = \frac{\frac{T}{m}}{T-\frac{T}{m}} = \frac{1}{m-1} \tag{51}$$

Proof of Proposition 3 (and $3^{'}$):

Under the framework described in section 6.1, the least squares estimator of the correlation coefficient, $\hat{\beta}$ is defined as:

$$\hat{\beta} = \frac{T\sum_{t=1}^{T} y_t y_{t-1} - \left(\sum_{t=1}^{T} y_t\right)\left(\sum_{t=1}^{T} y_{t-1}\right)}{T\sum_{t=1}^{T} y_{t-1}^2 - \left(\sum_{t=1}^{T} y_{t-1}\right)^2} \tag{52}$$

Since the true process is the pure random walk process described by (2), it follows that the lagged term y_{t-1} is given by equation(53) below.

$$y_{t-1} = y_0 + S_{t-1} \tag{53}$$

with $S_t = \sum_{j=1}^{t}\epsilon_j$
This implies that the sum of the sequence $\{y_{t-1}\}$ can be expressed as in (54).

$$\sum_{t=1}^{T} y_{t-1} = \sum_{t=1}^{T} S_{t-1} + Ty_0 \tag{54}$$

- For the terms in the numerator of (52), by using (53), it follows that:

  - ◇ $\sum_{t=1}^{T} y_t y_{t-1} = \sum_{t=1}^{T} S_t S_{t-1} + y_0\sum_{t=1}^{T} S_t + y_0\sum_{t=1}^{T} S_{t-1} + Ty_0^2$
  - ◇ $\sum_{t=1}^{T} y_t = \sum_{t=1}^{T} S_t + Ty_0$
  - ◇ $\sum_{t=1}^{T} y_{t-1} = \sum_{t=1}^{T} S_{t-1} + Ty_0$

Therefore, the numerator $T\sum_{t=1}^{T} y_t y_{t-1} - \left(\sum_{t=1}^{T} y_t\right)\left(\sum_{t=1}^{T} y_{t-1}\right)$ can be expressed in terms of the sum of innovations S_t (and its lagged counterpart S_{t-1}(and the initial value y_0:

$$T\sum_{t=1}^{T} y_t y_{t-1} - \left(\sum_{t=1}^{T} y_t\right)\left(\sum_{t=1}^{T} y_{t-1}\right) = T\sum_{t=1}^{T} S_t S_{t-1} + y_0 T\sum_{t=1}^{T} S_t + y_0 T\sum_{t=1}^{T} S_{t-1} +$$
$$+T^2 y_0^2 - \sum_{t=1}^{T} S_t \sum_{t=1}^{T} S_{t-1} - y_0 T\sum_{t=1}^{T} S_t - y_0 T\sum_{t=1}^{T} S_{t-1} - T^2 y_0^2 \tag{55}$$

As it turns out, all the terms which include y_0 cancel out, and therefore we obtain that:

$$T\sum_{t=1}^{T} y_t y_{t-1} - \left(\sum_{t=1}^{T} y_t\right)\left(\sum_{t=1}^{T} y_{t-1}\right) = T\sum_{t=1}^{T} S_t S_{t-1} - \sum_{t=1}^{T} S_t \sum_{t=1}^{T} S_{t-1}$$

Hence, the numerator is invariant of the initial value y_0.

- For the Denominator $T\sum_{t=1}^{T} y_{t-1}^2 - \left(\sum_{t=1}^{T} y_{t-1}\right)^2$ we have:
  - ◇ $\sum_{t=1}^{T} y_{t-1}^2 = \sum_{t=1}^{T} S_{t-1}^2 + 2y_0 \sum_{t=1}^{T} S_{t-1} + Ty_0^2$
  - ◇ $\left(\sum_{t=1}^{T} y_{t-1}\right)^2 = \left(\sum_{t=1}^{T} S_{t-1}\right)^2 + 2Ty_0 \sum_{t=1}^{T} S_{t-1} + T^2 y_0^2$

$$\Rightarrow T\sum_{t=1}^{T} y_{t-1}^2 - \left(\sum_{t=1}^{T} y_{t-1}\right)^2 = T\sum_{t=1}^{T} S_{t-1}^2 + 2y_0 T\sum_{t=1}^{T} S_{t-1} + T^2 y_0^2 +$$
$$-\left(\sum_{t=1}^{T} S_{t-1}\right)^2 - 2Ty_0 \sum_{t=1}^{T} S_{t-1} - T^2 y_0^2 \tag{56}$$

  Again, all the terms which incorporate y_0 cancel out and we have that for the denominator, it holds that:

$$T\sum_{t=1}^{T} y_{t-1}^2 - \left(\sum_{t=1}^{T} y_{t-1}\right)^2 = T\sum_{t=1}^{T} S_{t-1}^2 - \left(\sum_{t=1}^{T} S_{t-1}\right)^2$$

Therefore the estimated coefficient $\hat{\beta}$ can be expressed in terms of the sums of innovation terms as: $\hat{\beta} = \frac{T\sum_{t=1}^{T} S_t S_{t-1} - \sum_{t=1}^{T} S_t \sum_{t=1}^{T} S_{t-1}}{T\sum_{t=1}^{T} S_{t-1}^2 - \left(\sum_{t=1}^{T} S_{t-1}\right)^2}$ and is invariant to y_0.

B Figures

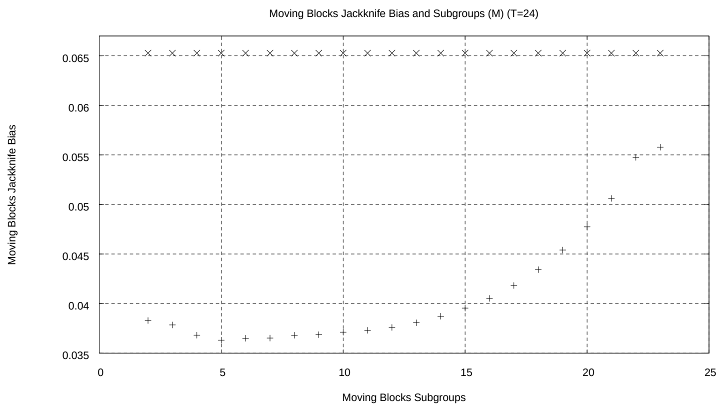

Figure 1: Moving-Blocks Jackknife bias and number of subgroups (M) at $T = 24$. The horizontal line denotes the corresponding full-sample least squares bias b_T. (10^4 replications)

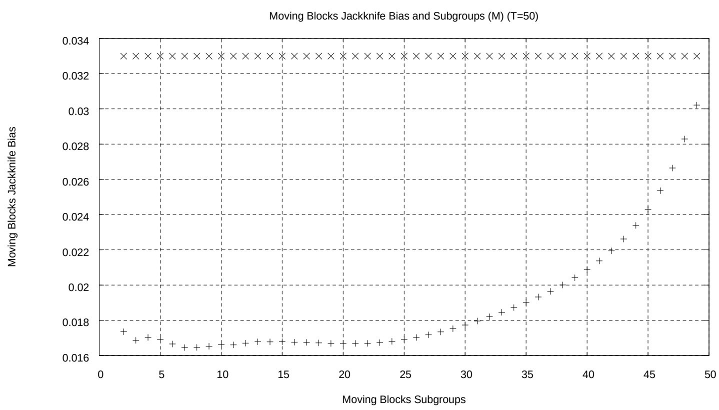

Figure 2: Moving-Blocks Jackknife bias and number of subgroups (M) at $T = 50$. The horizontal line denotes the corresponding full-sample least squares bias b_T . (10^4 replications)

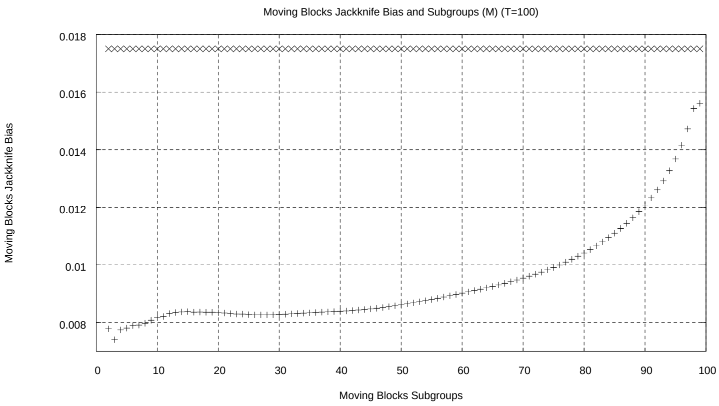

Figure 3: Moving-Blocks Jackknife bias and number of subgroups (M) at $T = 100$. The horizontal line denotes the corresponding full-sample least squares bias b_T. (10^4 replications)

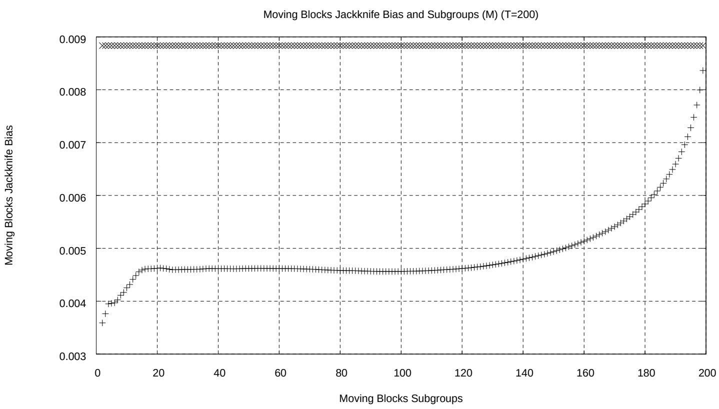

Figure 4: Moving-Blocks Jackknife bias and number of subgroups (M) at $T = 200$. The horizontal line denotes the corresponding full-sample least squares bias b_T. (10^4 replications)

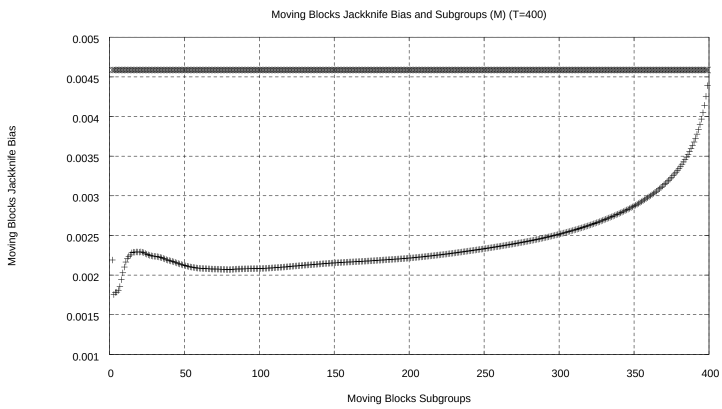

Figure 5: Moving-Blocks Jackknife bias and number of subgroups (M) at $T = 400$. The horizontal line denotes the corresponding full-sample least squares bias b_T.(10^4 replications)