ISSN 1032-3813
ISBN 0 7326 1029 X

MONASH UNIVERSITY

ANCORA · IMPARO

AUSTRALIA

ANALYTIC SMALL SAMPLE BIAS AND STANDARD ERROR CALCULATIONS FOR TESTS OF SERIAL CORRELATION IN MARKET RETURNS

Michael Smith
Narayan Y. Naik

Working Paper 6/97
May 1997

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

Analytic Small-Sample Bias and Standard Error Calculations for Tests of Serial Correlation in Market Returns

MICHAEL SMITH and NARAYAN Y. NAIK [1]

May 23rd 1997

[1]Michael Smith is from the Department of Econometrics and Business Statistics, Monash University, Australia and Narayan Naik is from the London Business School, United Kingdom. Michael Smith is grateful for support from a Monash University Faculty of Business and Economics research grant. Narayan Naik is grateful for support from the research fund of the Institute of Finance and Accounting of London Business School.

Analytic Small Sample Bias and Standard Error Calculations for Tests of Serial Correlation in Market Returns

ABSTRACT

This article derives analytic finite sample approximations to the bias and standard error of a class of statistics which test the hypothesis of no serial correlation in market returns. They offer an alternative to both the widely used Monte Carlo approach for calculating the bias, as well as asymptotic standard error calculations. These approximations are calculated under the assumption that returns are spherically symmetrically distributed (such as Gaussian) and also under the weaker assumption that returns follow any arbitrary continuous distribution. The class of statistics examined here includes many of those employed in the finance and macroeconomics literature to test for the existence of random walk, including the variance ratio and the multi-period return regression on past returns. The accuracy of the approximations is benchmarked using simulated data, where arbitrarily tight estimates of the bias and standard error can be calculated. The approximations are then applied to adjust the statistics calculated using returns on the NYSE from 1926-1991.

There is a large literature in finance which documents the serial correlation properties of asset returns (see Bollerslev & Hodrick (1992) for a survey). A common feature of this literature is that the autocorrelation-based statistics are estimated over a shorter sampling interval than the horizon of the statistics. For example, consider the case where the researcher is interested in autocorrelation of annual stock returns but, due to the availability of data, is able to sample yearly returns every month. From an efficiency standpoint, there are good reasons for employing fine sampling intervals. Although annual returns sampled monthly are serially correlated due to the presence of overlapping observations, the estimators based on more frequent sampling are known to be asymptotically more efficient (see Hansen & Hodrick (1980)). With this enhanced efficiency, however, small sample issues become important, especially the bias and standard error of these estimators.

The common approach to determining the bias under the null hypothesis of no correlation has been through Monte Carlo simulation (see Huizinga (1987), Fama & French (1988), and Lo & MacKinlay (1989)). For example, Fama & French (1988) estimate J-period autocorrelations via a simulation and then tabulate, for given T and J, the small sample bias of these estimators. These tables are then used to adjust autocorrelations estimated from the actual data. This approach requires simulation for every value of T and J used in practice, which in turn requires pre-specification of the distribution of returns– something that is impossible to know. For calculating the standard errors, it is common to use asymptotic values (Richardson & Smith, 1991), which are potentially very different from the actual small sample standard errors. The problems arising from not carefully accounting for small sample issues have been examined by Nelson & Kim (1993) in the case where returns are regressed upon lagged dividend yields. This paper examines these issues for the long-horizon autocorrelation based statistics that do not employ an additional independent variable.

There exists a large literature in statistics on the small sample moments of autocorrelation estimators (see Mariott & Pope (1954) and Dufour & Roy (1985), among others). For example, the sample autocorrelation of a random series is well known to have a negative bias of the order T^{-1}, where T is the number of observations used in the estimation. However, the results in this literature do not generally carry through to the frequently sampled long-horizon autocorrelation estimators because these estimators are nonlinear functions of different order autocorrelations (Richardson & Smith, 1994). Thus, analytically, little has been known about the bias and standard error of these estimators for an uncorrelated series. However, as shown in section I, it is possible to employ Taylor series expansions and derive analytical formula for both the finite sample bias and standard error as functions of known moments in the autocorrelation estimators. Although these formulae are approximate (due to higher-order Taylor series terms), we show in section II that they work well in small samples. To the extent that long-horizon statistics are used frequently in the literature, the results in this paper have application in practice.

I. The Bias and Standard Error of Autocorrelation-Based Estimators

The finance literature has especially focused on long-horizon statistics with frequent sampling intervals. The main reason for this focus is a general belief that many interesting phenomena occur at low frequencies of the data, leading to the choice of long-horizon estimators for increased statistical power (see Cochrane (1988), Poterba & Summers (1988) and Richardson & Smith (1991)). Given the choice of long-horizon statistics and the availability of higher frequency data, it is natural to use all the information available in the estimation. Since much of this literature in finance considers testing the null hypothesis of serially uncorrelated stock returns, we consider deriving the small sample bias and standard error of long-horizon autocorrelation based estimators under this null hypothesis.

A. Long-Horizon Autocorrelation-Based Estimators

Consider a time series $\{R_t\}_{t=1}^{T}$, such as continuously compounded dividend adjusted real returns $R_t = \log(P_t + D_t) - \log(P_{t-1})$, where P_t is the real price of the underlying asset at time t and D_t is the dividend paid in the period $(t-1, t]$. A number of tests have been suggested in the literature to decide whether there is evidence that such a series is correlated. Typically, these tests fall into two categories: those which regress future returns on past returns and those which compute variance ratios over different intervals.[1] Although at first glance the regression tests and variance ratio

[1]For example, Fama & French (1988), Huizinga (1987) and Jegadeesh (1991) use the regression approach while Campbell & Mankiw (1987), Cochrane (1988), French & Roll (1986) and Lo & MacKinlay

tests appear to be taking rather different routes, both these approaches essentially test for the random walk in a very similar way. Both these approaches propose test statistics which are nonlinear combinations of consistent estimators of autocorrelations of different order obtained from the observed time series. These approaches differ from each other only in terms of the weights they attach to the autocorrelations of different order while forming the test statistic.

Richardson & Smith (1994) show that these long-horizon statistics employed in the current literature can be written as a linear combination of different order autocovariances of R_t weighted by the sample J-period variance estimator of R_t. That is,

$$F_D(\hat{\rho}(I,J)) = \frac{\Sigma_{i=1}^{I} D_i \hat{\text{cov}}(R_t, R_{t-i})}{\hat{\text{var}}(\Sigma_{j=1}^{J} R_{t-j})/J} \tag{1}$$

where $\hat{\text{cov}}(R_t, R_{t-i})$ is the i-th order autocovariance estimator, D_i are the weights on these autocovariances, and $\hat{\text{var}}(\Sigma_{j=1}^{J} R_{t-j})$ is the J-period variance estimator. Most of the existing long-horizon estimators fall within the class of estimators described by equation (1). For example,

- the J-period autocorrelation estimator used by Fama & French (1988) implies the weights $D_i = \min(i, 2J-i)/J$ where $I = 2J-1$ in equation (1),
- the I-period variance ratio estimator used by Cochrane (1988), Lo & MacKinlay (1988) and Faust (1992) implies the weights $D_i = 2(I-i)/I$ where $J = 1$ in equation (1), and
- the One-period on J-period regression estimator used by Jegadeesh (1991) implies the weights $D_i = 1/J$ where $I = J$ in equation (1).

Similarly, various other autocorrelation-based estimators in the presence of overlapping observations described in Richardson & Smith (1991) also fit in this framework.

(1988) employ the variance ratio approach.

Note that equation (1) can be written as a nonlinear function of estimators of the i-th order sample autocorrelations of R_t (denoted by $\hat{\rho}_i$). To see this, divide both the numerator and the denominator of equation (1) by the sample variance, $\hat{\text{var}}(R_t)$, so that[2]

$$\begin{aligned} F_D(\hat{\rho}(I,J)) &= \left(\frac{\Sigma_{i=1}^{I} D_i \hat{\text{cov}}(R_t, R_{t-i})}{\hat{\text{var}}(R_t)}\right)\left(\frac{\hat{\text{var}}(\Sigma_{j=1}^{J} R_{t-j})}{J \times \hat{\text{var}}(R_t)}\right)^{-1} \\ &= \frac{\Sigma_{i=1}^{I} D_i \hat{\rho}_i}{1 + 2\Sigma_{j=1}^{J-1} \frac{J-j}{J} \hat{\rho}_j} \end{aligned} \tag{2}$$

Under the null hypothesis of no correlation $F_D(\hat{\rho}(I,J)) = 0$ because $\rho_i = 0$, for all i. Deviations from the null are represented by values of F_D that differ significantly from zero. Therefore, the small sample bias and standard error of F_D require calculation. The complication in analytically obtaining these comes from the fact that the numerator and denominator in equation (2) are not independent. Consistent with the literature in this area, we approximate the bias and standard error using Taylor series expansions. The rest of this paper is devoted to deriving, justifying (via simulation) and applying these approximations.

B. Expression for the Bias

Various estimators of autocorrelations, ρ_i, are considered in the literature. For the random walk case, in which $\rho_i = 0$, there is practically no difference between choosing one estimator over another one (see Moran (1948), Mariott & Pope (1954)). In order to calculate the bias, we consider a particular estimator, although it should be noted that the results carry through (albeit slightly differently) to other choices as well.

Specifically, following Moran (1948), we calculate the i-th order sample autocorre-

[2]See Cochrane (1988, Appendix A) for the derivation of denominator in equation (2). The denominator is simply the ratio of variance of J-period returns to the variance of One-period return scaled by J. Under the null of random walk this ratio equals one; however, to the extent the data generating process departs from random walk, this ratio differs from one.

lation, $\hat{\rho}_i$, as

$$\hat{\rho}_i = \frac{\frac{1}{T-i}\Sigma_{t=1}^{T-i}(R_t - \bar{R})(R_{t+i} - \bar{R})}{\frac{1}{T}\Sigma_{t=1}^{T}(R_t - \bar{R})^2} \tag{3}$$

where the sample mean $\bar{R} = \frac{1}{T}\Sigma_{t=1}^{T} R_t$. If we substitute these $\hat{\rho}_i$'s into equation (2), then we can obtain an estimate of the long-horizon autocorrelation based estimator $F_D(\hat{\rho}(I,J))$.

Note that the expectation of $F_D(\hat{\rho}(I,J))$ can be decomposed into the weighted sum of the individual expectations of $\hat{\rho}_i/(1 + 2\Sigma_{j=1}^{J-1}\frac{J-j}{J}\hat{\rho}_j)$. Thus, for our purposes it suffices to derive the expectation for any i. For the purpose of exposition, we denote the variance ratio in the denominator of equation (2) as $VR_J \equiv (1 + 2\Sigma_{j=1}^{J-1}\frac{J-j}{J}\hat{\rho}_j)$. Following standard methodology in this area, we consider a Taylor series expansion of $\hat{\rho}_i/VR_J$ around $E[\hat{\rho}_i]$ and $E[VR_J]$. We carry out the expansion to the second order because for $\rho_i = 0$, the higher order terms are of $O(T^{-2})$ and below. Specifically, we have

$$\begin{aligned}\frac{\hat{\rho}_i}{VR_J} \approx\ & \frac{E[\hat{\rho}_i]}{E[VR_J]} + \frac{(\hat{\rho}_i - E[\hat{\rho}_i])}{E[VR_J]} - \frac{E[\hat{\rho}_i](VR_J - E[VR_J])}{E[VR_J]^2} \\ & + \frac{E[\hat{\rho}_i](VR_J - E[VR_J])^2}{E[VR_J]^3} - \frac{(\hat{\rho}_i - E[\hat{\rho}_i])(VR_J - E[VR_J])}{E[VR_J]^2}\end{aligned} \tag{4}$$

Taking the expectation of equation (4), the second and the third term immediately drop out yielding the following expression, up to $O(T^{-2})$ terms, for the bias.

$$E\left[\frac{\hat{\rho}_i}{VR_J}\right] \approx \frac{E[\hat{\rho}_i]}{E[VR_J]} + \frac{E[\hat{\rho}_i]\mathrm{var}(VR_J)}{E[VR_J]^3} - \frac{\mathrm{cov}(\hat{\rho}_i, VR_J)}{E[VR_J]^2} \tag{5}$$

The expansion at equation (5) suggests that the bias relies on the moments of the sample autocorrelation coefficient, $\hat{\rho}_i$, and variance ratio, VR_J. Since the variance ratio is itself a linear function of sample autocorrelation coefficients, it is possible to use results for the moments of $\hat{\rho}_i$ to calculate the unknown moments in equation (5).

B.1. Strong Assumptions

Under the null hypothesis that $\{R_t\}_{t=1}^{T}$ are independent and identically distributed with a spherically symmetrical distribution (such as the normal distribution) the following moments can be calculated using the results in the appendix.

$$
\begin{aligned}
E[VR_J] &= \frac{T-J}{T-1} \\
\operatorname{var}(VR_J) &= \frac{4T^4}{(T+1)(T-1)^2J^2}\sum_{j=1}^{J-1}\frac{(J-j)^2}{(T-j)^2} + O\left(T^{-2}\right) \\
\operatorname{cov}(VR_J,\hat{\rho}_i) &= \frac{2T^4\max(0,J-i)}{J(T-i)^2(T+1)(T-1)^2} + O\left(T^{-2}\right)
\end{aligned}
$$

Substituting these values into equation (5), it is possible to express the bias as

$$
E\left[\frac{\hat{\rho}_i}{VR_J}\right] = \frac{-1}{T-J}\left[1+\frac{2T^4\max(0,J-i)}{J(T-i)^2(T-J)(T+1)}\right] + O\left(T^{-2}\right) \tag{6}
$$

where the second term in equation (5) disappears as it is of $O(T^{-2})$. Taking expectations of (2) and substituting in (6), the bias of the long-horizon autocorrelation-based estimator $F_D(\hat{\rho}(I,J)))$ can be written, up to $O(T^{-2})$, as

$$
E[F_D(\hat{\rho}(I,J))] = \sum_{i=1}^{I} D_i E\left[\frac{\hat{\rho}_i}{VR_J}\right] \approx -\sum_{i=1}^{I}\frac{D_i}{T-J}\left[1+\frac{2T^4\max(0,J-i)}{J(T-i)^2(T-J)(T+1)}\right] \tag{7}
$$

The formula for the bias in (7) holds for all estimators that fall within the class of long-horizon estimators described by equation (2). As pointed out in section I.A., this class contains most of the long-horizon methods currently used in practice. As examples, consider substituting the weights D_i for the J-period autocorrelation, the I-period variance-ratio and the 1-on-J regression into equation (7). It is then possible to show that, up to $O(T^{-2})$, the resulting bias of these estimators are then:

$$
\begin{aligned}
J\text{-period autocorrelation} &: -\left\{\frac{J}{T-J}+\frac{2T^4}{(T-J)^2J^2(T+1)}\sum_{i=1}^{J-1}\frac{i(J-i)}{(T-i)^2}\right\} \\
I\text{-period variance-ratio} &: \frac{-(I-1)}{(T-1)} \\
1\text{-on-}J\text{ regression} &: \frac{-1}{T-J}\left[1+\frac{2T^4}{(T-J)(T+1)J^2}\sum_{i=1}^{J-1}\frac{(J-i)}{(T-i)^2}\right]
\end{aligned}
$$

B.2. Weak Assumptions

If the null hypothesis is weakened, so that $\{R_t\}_{t=1}^T$ remain independent and identically distributed, but with an arbitrary continuous distribution, then the results in the appendix allow the construction of the following bounds.

$$E[VR_J] = \frac{T-J}{T-1}$$

$$\text{var}(VR_J) \leq \frac{4T^5}{J^2(T-1)^2(T-2)(T-3)} \sum_{j=1}^{J-1} \frac{(J-j)^2}{(T-j)^2} + O(T^{-2})$$

$$O(T^{-2}) \leq \text{cov}(VR_J, \hat{\rho}_i) \leq \frac{2T^5 \max(0, J-i)}{J(T-i)^2(T-1)^2(T-2)(T-3)} + O(T^{-2})$$

Substituting these into equation (5) allows the determination of upper and lower bounds for the bias, so that

$$\frac{-1}{T-J}\left[1 + \frac{2T^5 \max(0, J-i)}{J(T-2)(T-3)(T-i)^2(T-J)}\right] + O(T^{-2}) \leq E\left[\frac{\hat{\rho}_i}{VR_J}\right] \leq \frac{-1}{T-J} + O(T^{-2})$$

Because $D_i \geq 0$, approximate bounds, up to $O(T^{-2})$, for the bias of the long-horizon autocorrelation based estimators are therefore

$$\sum_{i=1}^{I} \frac{-D_i}{T-J}\left[1 + \frac{2T^5 \max(0, J-i)}{J(T-2)(T-3)(T-i)^2(T-J)}\right] \leq E[F_D(\hat{\rho}(I,J))] \leq \sum_{i=1}^{I} \frac{-D_i}{T-J}$$

By substituting in the weights for the three estimators considered in section I.A., this results in the following approximate bounds on the bias for the J-period autocorrelation, I-period variance-ratio and 1-on-J regression estimators, respectively.

$$-\left\{\frac{J}{T-J} + \frac{2T^5}{J^2(T-2)(T-3)(T-J)^2} \sum_{i=1}^{J-1} \frac{i(J-i)}{(T-i)^2}\right\} \leq E[F_D] \leq \frac{-J}{T-J}$$

$$E[F_D] = \frac{-(I-1)}{T-1}$$

$$\frac{-1}{T-J}\left\{1 + \frac{2T^5}{J^2(T-J)(T-3)(T-2)} \sum_{i=1}^{J-1} \frac{J-i}{(T-i)^2}\right\} \leq E[F_D] \leq \frac{-1}{T-J}$$

C. Expression for the Variance

Using a similar procedure it is possible to develop an approximate expression for the finite sample variance of such autocorrelation based estimators. That is,

$$\begin{aligned}\operatorname{var}(F_D) &= \sum_{i=1}^{I}\sum_{k=1}^{I} D_i D_k \operatorname{cov}\left(\frac{\hat\rho_i}{VR_J}, \frac{\hat\rho_k}{VR_J}\right) \\ &= \sum_{i=1}^{I}\sum_{k=1}^{I} D_i D_k \left\{ E\left[\frac{\hat\rho_i\hat\rho_k}{(VR_J)^2}\right] - \underbrace{E\left[\frac{\hat\rho_i}{VR_J}\right] E\left[\frac{\hat\rho_k}{VR_J}\right]}_{\text{from previous section}} \right\} \qquad (8)\end{aligned}$$

In a similar manner as in section I.B., taking the expected value of a second order Taylor series expansion of $\frac{\hat\rho_i\hat\rho_k}{(VR_J)^2}$ around $E[\hat\rho_i], E[\hat\rho_k]$ and $E[VR_J]$ gives, up to $O(T^{-3})$,

$$\begin{aligned}E\left[\frac{\hat\rho_i\hat\rho_k}{(VR_J)^2}\right] &\approx \frac{E[\hat\rho_i]E[\hat\rho_k]}{E[VR_J]^2} + \frac{3\operatorname{var}(VR_J)E[\hat\rho_i]E[\hat\rho_k]}{E[VR_J]^4} + \frac{\operatorname{cov}(\hat\rho_i, \hat\rho_k)}{E[VR_J]^2} \\ &- \frac{2\operatorname{cov}(\hat\rho_i, VR_J)E[\hat\rho_k]}{E[VR_J]^3} - \frac{2\operatorname{cov}(\hat\rho_k, VR_J)E[\hat\rho_i]}{E[VR_J]^3}\end{aligned}$$

Again, it is possible to use the results on sample autocorrelations found in section I.B. and the appendix to calculate approximations for the variance.

For example, under the strong assumptions (specified by the null hypothesis found in section I.B.1.), the results on the moments of sample autocorrelations can be used to deduce that

$$\begin{aligned}E\left[\frac{\hat\rho_i\hat\rho_k}{(VR_J)^2}\right] &= \frac{1}{(T-J)^2} + \frac{4T^4}{J(T+1)(T-J)^3}\left\{\frac{\max(0,J-i)}{(T-i)^2} + \frac{\max(0,J-k)}{(T-k)^2}\right\} \\ &+ \frac{T^4 + T^3(i+3)}{(T-i)^2(T+1)(T-J)^2}\mathcal{I}_{(i=k)} + O(T^{-3})\end{aligned}$$

Here, $\mathcal{I}_{(i=k)}$ is an indicator variable equal to one when $i = k$ and zero otherwise. Therefore, substituting this and the expansion for $E\left[\frac{\hat\rho_i}{VR_J}\right]$ found at equation (6) into equation (8), the variance of the long-horizon autocorrelation-based estimator $F_D(\hat\rho(I,J))$ under the strong assumptions can be expressed, up to $O(T^{-3})$, as

$$\begin{aligned}\operatorname{var}(F_D) &\approx \sum_{i=1}^{I}\sum_{k=1}^{I} D_i D_k \left\{\frac{2T^4}{J(T+1)(T-J)^3}\left[\frac{\max(0,J-i)}{(T-i)^2} + \frac{\max(0,J-k)}{(T-k)^2}\right]\right. \\ &- \left.\frac{4T^8\max(0,J-i)\max(0,J-k)}{J^2(T-J)^4(T-i)^2(T-k)^2(T+1)^2}\right\} + \sum_{i=1}^{I}\frac{D_i^2(T^4 + T^3(i+3))}{(T-i)^2(T+1)(T-J)^2} \qquad (9)\end{aligned}$$

Substituting in the weights for the popular J-period autocorrelation estimator, the resulting estimate of the variance, up to $O(T^{-3})$, is

$$\begin{aligned}\operatorname{var}(F_D) &\approx \frac{4T^4}{J(T+1)(T-J)^3}\left\{\sum_{i=1}^{J-1}\frac{i(J-i)}{(T-i)^2}\right\} - \frac{4T^8}{(T-J)^4(T+1)^2J^4}\left\{\sum_{i=1}^{J-1}\frac{i(J-i)}{(T-i)^2}\right\}^2 \\ &+ \frac{T^3}{J^2(T+1)(T-J)^2}\sum_{i=1}^{J-1}\left\{\left(\frac{i^2}{(T-i)^2}+\frac{(J-i)^2}{(T-J-i)^2}\right)(T+i+3)\right\}\end{aligned}$$

Other estimators can be evaluated by substituting their respective weights into equation (9). In addition, approximate bounds for the variance (and hence standard errors) of these estimators can also be calculated under the weak assumptions discussed in section I.B.2. It is also important to stress that, unlike the asymptotic variance used in much of the literature, (see Richardson & Smith, 1994) this is a finite sample approximate expression for the variance.

II. Simulated and Real Data Examples

A. Example 1: The Random Walk

The analytic approximations developed in section I for the bias of autocorrelation-based estimators are only exact up to $O(T^{-2})$ terms. The accuracy of these approximations, for a given underlying distribution and values of T and J, can be verified using tight Monte Carlo estimates for the bias based on many simulated iterations. This section does so using returns, $\{R_t\}_{t=1}^{T}$, which were generated as independent $N(0,1)$ variates, so that the logarithm of the underlying is assumed to follow a random walk without drift. As the normal is a spherically symmetrical distribution, the bias and variance approximations derived under the strong conditions are applicable.

—figure 1 about here—

A range of values for T and J is taken and the analytic estimates of the bias calculated and plotted in figure 1 for the three common estimators discussed in section I.A. The panel columns correspond to three typical values of T, namely (from left to right)

T= 360, 720 and 1440. In each panel, the approximate bias for that particular estimator and value of T is plotted as a bold line for $12 < J < 120$, while the dotted line gives the Monte Carlo estimate of the bias. For the J-period autocorrelation and J-period variance-ratio estimator these are obtained using a Monte Carlo simulation based on 150,000 iterations. However, this did not prove sufficient to get tight Monte Carlo values for the 1-on-J regression estimator, where 1,000,000 iterations were required. In line with the theme of the paper the estimators were calculated as sums of sample autocorrelations of the form given at equation (3).

For the J-period variance-ratio estimator the approximations appear to be excellent for all values of J and T examined here. The approximation for the J-period autocorrelation and 1-on-J regression estimators are more accurate for larger values of T and smaller values of J. Finite sample estimates of the variance of each estimator, up to $O(T^{-3})$, can also be calculated by substituting in the relevant weights into equation (9).

B. Example 2: Stratification by Market Value and Industry

We applied the popular J-period autocorrelation estimator to monthly returns for companies on the NYSE from January 1926 to December 1991 (so that $T = 792$), where the data have been sorted into deciles according to their market value. Table I details the calculated statistic, along with the finite sample bias and standard error approximations calculated under the strong assumptions. As the underlying distribution of real returns are not known, the finite sample approximate upper and lower bounds for the bias under the weak assumptions are also given. None of the statistics have a bias-adjusted value greater that two standard errors away from zero using the approximations derived under the strong assumptions. However, it should be noted that while these statistics are asymptotically normally distributed (Richardson & Smith, 1994), they may not be in small samples, so that treating such tests as simple t-tests can be misleading.[3]

This data was also stratified by industry (minus the observation for January 1926 which could not be stratified this way) and table II provides the resulting values of the J-period autocorrelation estimator. Only the four year bias-adjusted statistic for capital goods is more than two standard errors away from zero. Overall, using these finite sample bias adjustments and standard error approximations, there appears to be little evidence of serial correlation in this returns data, whether the data is stratified by market value or industry size.

—tables I and II about here—

C. Example 3: Asymmetrically Distributed Returns

The returns for the first decile (the smaller companies) appear far from normally distributed. We mimicked the observed distribution of these returns using a 'mixture of normals', where $R_t \sim N(0.01, 0.047)$ with probability 0.6, $R_t \sim N(0.01, 0.1)$ with probability 0.37 and $R_t \sim N(0.3, 0.3)$ with probability 0.03. Here, $N(\mu, \sigma)$ is the normal distribution with mean μ and standard deviation σ. Figure 2 plots a histogram of the first decile returns, showing that the density of this normal mixture compares very closely to the empirical density of the actual data.

—figure 2 about here—

We therefore calculated a tight Monte Carlo estimate of the bias of the three estimators using returns, $\{R_t\}_{t=1}^{T}$, generated from the above mixture of normals. A range of values for J ($12 < J < 120$) were considered, while $T = 792$ (which corresponded to the actual number of monthly observations in example 2) and 150,000 iterations were used to obtain a tight estimate of the bias. The $N(0.3, 0.3)$ distribution, which mimics the heavy positive tail of the observed distribution of the first decile returns, renders

[3] For example, if the true finite sample distribution of the J-period autocorrelation was heavy tailed on the left, a deviation of even more than two standard deviations from zero would be required to reject the null hypothesis of no serial correlation with 95% certainty.

the mixture of normals asymmetric. Hence, the strong conditions no longer apply, though the finite sample approximations based on the weak conditions still hold. Figure 3 plots the approximate upper and lower bounds for the bias for each of the three estimators, along with the corresponding tight Monte Carlo estimate of the bias. They reveal that the approximate bounds appear quite accurate for all three estimators for this value of T.

—figure 3 about here—

III. Conclusion

This paper has provided a set of finite sample approximations of the bias and standard error of a class of statistics often used in the financial literature to determine whether, or not, series of returns are correlated. They can be calculated under either the assumption that returns are spherically symmetrically distributed, or under the weaker condition that returns arise from an arbitrary continuous distribution. The bias calculation improves on previous finite sample Monte Carlo derived adjustments (such as in Fama & French, 1988) as it is both analytic and does not require specification of a particular distribution (see example 3). The estimates of the standard error improve upon the popular asymptotic standard errors (such as in Richardson & Smith, 1994), as they are finite sample approximations. Moreover, as the bias and standard error approximations are exact up to $O(T^{-2})$ and $O(T^{-3/2})$ terms, respectively, they improve with larger sample sizes.

Appendix: Moments of Sample Autocorrelations

Much work in the statistical literature has gone into calculating the moments of sample autocorrelation coefficients. Dufour & Roy (1985) give an excellent summary and this appendix is based mainly on their survey. Here, we list moments that arise from two different null hypotheses which differ in the assumption that the process $\{R_t\}_{t=1}^{T}$

follows a spherically symmetric (eg normal) or an arbitrary continuous distribution. We label these 'strong' and 'weak' assumptions, respectively.

Moments under strong assumptions

Under the null hypothesis that $\{R_t\}_{t=1}^T$ are independent and identically distributed with a spherically symmetrical distribution (such as the normal distribution) the following hold.

$$
\begin{aligned}
E[\hat{\rho}_i] &= \frac{-1}{T-1} \\
\operatorname{var}(\hat{\rho}_i) &= \frac{T^4-(i+3)T^3+3iT^2+2i(i+1)T-4i^2}{(T-i)^2(T+1)(T-1)^2} \\
\operatorname{cov}(\hat{\rho}_i,\hat{\rho}_j) &= \frac{2\{ij(T-1)-(T-j)(T^2-i)\}}{(T+1)(T-1)^2(T-i)(T-j)} \quad \text{for} \quad 1 \leq i < j \leq T-1
\end{aligned}
$$

Moments under weak assumptions

Under the null hypothesis that $\{R_t\}_{t=1}^T$ are independent and identically distributed, but with an arbitrary continuous distribution, the following results on sample autocorrelation moments hold.

$$E[\hat{\rho}_i] = \frac{-1}{T-1}$$

$$\operatorname{var}(\hat{\rho}_i) \leq \frac{T\{T^4-(i+7)T^3+(7i+16)T^2+2(i^2-9i-6)T-4i(i-4)\}}{(T-i)^2(T-1)^2(T-2)(T-3)}$$

$$\frac{-2(T-j+3)}{T(T-i)(T-j)}+O\left(T^{-4}\right) \leq \operatorname{cov}(\hat{\rho}_i,\hat{\rho}_j) \leq \frac{2(i+2)}{T(T-i)(T-j)}+O\left(T^{-4}\right) \quad \text{for } 1 \leq i < j \leq T-1$$

REFERENCES

Bollerslev, T. and Hodrick, R., 1992, 'Financial Market Efficiency Tests', NBER working paper, no. 4108

Campbell, J. and Mankiw G., 1987, 'Are Output Fluctuations Transitory?,' *Quarterly Journal of Economics*, vol.102, 857-880

Cochrane, J., 1988, 'How Big is the Random Walk in GNP?' *Journal of Political Economy*, 96, 247-273

Dufour, J. and Roy, R., 1985, 'Some robust exact results on sample autocorrelations and tests of randomness', *Journal of Econometrics*, vol.29, 257-273

Fama, E., and French, K., 1988, 'Permanent and Temporary Components of Stock Prices', *Journal of Political Economy*, vol.96, no.2, 246-272

Faust, J., 1992, 'When are Variance Ratio Tests for Serial Dependence Optimal?', *Econometrica*, vol.60, 1215-1226

French, K. and Roll, R., 1986, 'Stock Return Variances: The Arrival of Information and the Reaction of Traders', *Journal of Financial Economics*, 17, 5-26

Hansen, L. and Hodrick, R., 1980, 'Forward exchange rates as optimal predictors of future spot rates: An econometric analysis', *Journal of Political Economy*, vol.88, 829-853

Huizinga, J., 1987, 'An Empirical Investigation of the Long-Run Behaviour of Real Exchange Rates,' *Carnegie-Rochester Conference Series on Public Policy*, 27

Jagadeesh, N., 1991, 'Seasonality in Stock Price Mean Reversion: Evidence from the US and UK', *The Journal of Finance*, vol.46, no. 4, 1427-1444

Lo, A. and MacKinlay, A., 1988, 'Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test', *The Review of Financial Studies*, vol.1, no.1, 41-66

Marriot, F. and Pope, J., 1954, 'Bias in the Estimation of Autocorrelations', *Biometrika*, vol.41, 390-402

Moran, P., 1948, 'Bias in Serial Correlations of Random Series,' *Biometrika*, vol.35, Nelson, C. and Kim, M., 1993, 'Predictable Stock Returns: The Role of Small Sample Bias', *Journal of Finance*, no.2, 641-661

Poterba, J. and Summers, L., 1988, 'Temporary components of stock prices: Evidence and Implications', *Journal of Financial Economics*, vol.22, 27-59

Richardson, M. and Smith, T., 1991, 'Tests of financial models in the presence of overlapping observations', *Review of Financial Studies*, vol.4, 227-254

Richardson, M. and Smith, T., 1994, 'A Unified Approach to Testing for Serial Correlation in Stock Returns,' *Journal of Business*, vol. 67, no.3, 371-399

	Unadjusted J-Period Autocorrelation Statistics							
Year ($J/12$)	1	2	3	4	5	6	7	8
1st Decile	0.008	-0.144	-0.328	-0.503	-0.493	-0.316	-0.166	-0.143
2nd Decile	-0.015	-0.147	-0.301	-0.482	-0.527	-0.443	-0.364	-0.431
3rd Decile	-0.059	-0.190	-0.341	-0.478	-0.487	-0.394	-0.325	-0.413
4th Decile	-0.068	-0.184	-0.337	-0.497	-0.522	-0.423	-0.356	-0.438
5th Decile	-0.082	-0.244	-0.365	-0.468	-0.498	-0.401	-0.314	-0.399
6th Decile	-0.098	-0.254	-0.416	-0.519	-0.506	-0.352	-0.250	-0.329
7th Decile	-0.106	-0.312	-0.438	-0.477	-0.461	-0.320	-0.230	-0.341
8th Decile	-0.077	-0.280	-0.395	-0.421	-0.377	-0.202	-0.096	-0.174
9th Decile	-0.068	-0.270	-0.399	-0.407	-0.310	-0.092	-0.020	-0.140
10th Decile	-0.059	-0.231	-0.329	-0.319	-0.205	-0.014	0.039	-0.077
	Bias and Standard Error Estimates ($T = 792$)							
Bias (Strong)	-0.021	-0.042	-0.065	-0.089	-0.114	-0.140	-0.168	-0.197
St. Error (Strong)	0.098	0.150	0.193	0.234	0.273	0.313	0.353	0.394
Upper Bias (Weak)	-0.015	-0.031	-0.048	-0.065	-0.082	-0.100	-0.119	-0.138
Lower Bias (Weak)	-0.021	-0.042	-0.065	-0.089	-0.114	-0.141	-0.168	-0.198

Table I: J-Period Autocorrelation Statistics for Returns Stratified by Size. The top half of the table provides the unadjusted values of the J-period autocorrelation statistic for monthly returns on the NYSE between January 1926 and December 1991. These have been provided for ten portfolios, each corresponding to a decile of the companies sorted by market value. Here, the 1st decile represents the smallest 10% of companies and the 10th decile the largest 10% of companies. The second half of the table provides the finite sample bias and standard error approximations, calculated under the strong assumptions, and bounds on the bias calculated under the weak assumptions. No bias-adjusted statistic is greater than two standard errors away from zero using the adjustments derived under the strong assumptions.

	Unadjusted J-Period Autocorrelation Statistics							
Year ($J/12$)	1	2	3	4	5	6	7	8
Petroleum	-0.058	-0.306	-0.268	-0.317	-0.263	-0.164	-0.029	-0.051
Finance/Real Estate	-0.033	-0.208	-0.397	-0.454	-0.347	-0.116	-0.024	-0.141
Consumer Durables	-0.077	-0.254	-0.418	-0.456	-0.389	-0.273	-0.233	-0.364
Basic Industries	-0.050	-0.275	-0.435	-0.460	-0.384	-0.227	-0.154	-0.208
Food/Tobacco	-0.015	-0.069	-0.014	0.081	0.135	0.229	0.193	0.106
Construction	-0.064	-0.189	-0.263	-0.376	-0.364	-0.269	-0.167	-0.188
Capital Goods	-0.014	-0.241	-0.462*	-0.506	-0.423	-0.238	-0.124	-0.189
Transportation	-0.191	-0.305	-0.315	-0.347	-0.338	-0.222	-0.187	-0.308
Utilities	0.006	-0.151	-0.240	-0.204	-0.122	0.071	0.026	-0.126
Textiles/Trade	-0.061	-0.209	-0.263	-0.248	-0.219	-0.146	-0.152	-0.266
Services	0.057	0.074	0.033	-0.138	-0.247	-0.256	-0.243	-0.265
Leisure	0.012	-0.146	-0.361	-0.477	-0.402	-0.227	-0.184	-0.261
	Bias and Standard Error Estimates ($T = 791$)							
Bias (Strong)	-0.021	-0.042	-0.065	-0.089	-0.114	-0.140	-0.168	-0.198
St. Error (Strong)	0.098	0.150	0.193	0.234	0.273	0.313	0.353	0.395
Upper Bias (Weak)	-0.015	-0.031	-0.048	-0.065	-0.082	-0.100	-0.119	-0.138
Lower Bias (Weak)	-0.021	-0.042	-0.065	-0.089	-0.114	-0.141	-0.169	-0.198

Table II: J-Period Autocorrelation Statistics for Returns Stratified by Industry. The top half of the table provides the unadjusted values of the J-period autocorrelation statistic for monthly returns on the NYSE between Febuary 1926 and December 1991. These have been provided for 12 portfolios corresponding to a stratification by industry. The second half of the table provides the finite sample bias and standard error approximations, calculated under the strong assumptions, and bounds on the bias calculated under the weak assumptions. Bias-adjusted statistics that are greater than two standard errors away from zero using the adjustments derived under the strong assumptions are marked with a star (*).

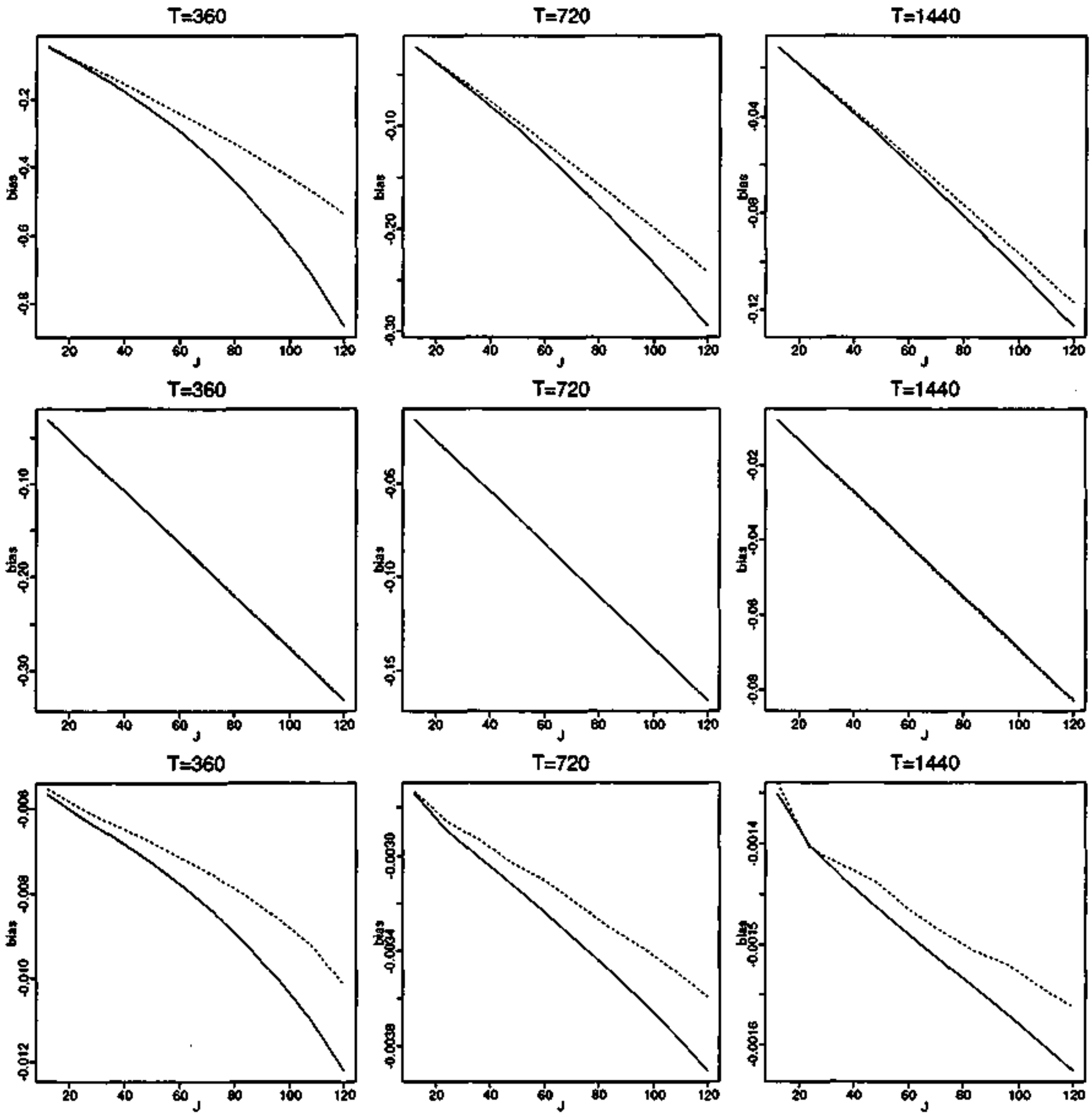

Figure 1: A Comparison of Tight Monte Carlo and Finite Sample Bias Approximations for a Random Walk. Each panel plots the bias on the y-axis against J on the x-axis. The bold line gives the analytic approximation for the bias, while the dotted line gives the tight Monte Carlo based estimate. The panels on the three rows (from top to bottom) correspond to the J-period autocorrelation, J-period variance-ratio and the 1-on-J regression estimators. The three columns correspond (from left to right) to T=360, 720 and 1440.

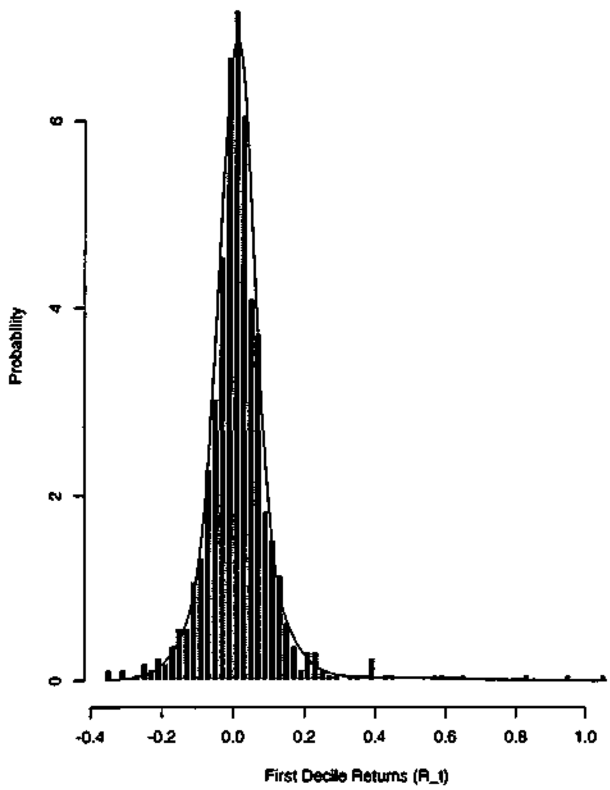

Figure 2: A Mixture of Normals Estimate of the Distribution of Returns from Small Market Value Companies. The histogram (normalized to have the area sum to one) is of the first decile of monthly returns on the NYSE from Janurary 1926 to December 1991 stratified by market size. The bold line is the density arising from the mixture of normals introduced in example 3. This mimics the observed behavior of returns quite well, including the heavy right hand tail.

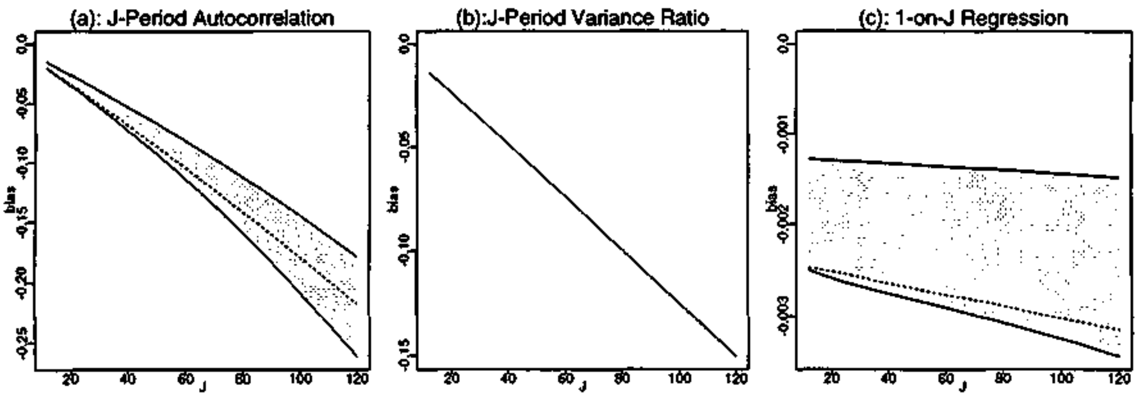

Figure 3: A Comparison of Tight Monte Carlo and Finite Sample Bias Approximations for Assymetric Returns. Each panel plots the analytic bias approximations (bold lines) from the asymmetric simulated returns in example 3 for one of the three estimators. In panels (a) and (c) the analytic approximations are of the upper and lower bounds only, so these are plotted as bold lines with the region between being shaded. The tight Monte Carlo based estimate of the bias is given as a dotted line. The x-axis provides a variety of values for J, while T is fixed to 792 throughout.