NBER WORKING PAPER SERIES

DRAWING INFERENCES FROM STATISTICS BASED ON MULTI-YEAR ASSET RETURNS

Matthew Richardson

James H. Stock

Working Paper No. 3335

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 1990

The authors thank A. Lo, D. Quah, G.W. Schwert, M. Watson, and the participants in the NBER Summer Institute on Financial Markets and Monetary Economics for helpful discussions and suggestions. This paper was written while Stock was visiting the Graduate School of Business, Stanford University. Financial support received from the NSF (grants SES-86-18984 ad SES-89-10601) (Stock), and the University of Pennsylvania Research Foundation (Richardson) is gratefully acknowledged. This paper is part of NBER's research program in Financial Markets and Monetary Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #3335
April 1990

DRAWING INFERENCES FROM STATISTICS BASED ON MULTI-YEAR ASSET RETURNS

ABSTRACT

The possibility of mean reversion in stock prices recently has been examined using statistics based on multi-year returns. Previous researchers have noted difficulties in drawing inferences about these statistics because of poor performance of the usual approximating asymptotic distributions. We therefore develop an alternative asymptotic distribution theory for statistics involving multi-year returns. These distributions differ markedly from those implied by the conventional theory. This alternative theory provides substantially better approximations to the relevant finite-sample distributions. It also leads to empirical inferences much less at odds with the hypothesis of no mean reversion.

Matthew Richardson
Department of Finance
Wharton School of Business
University of Pennsylvania
Philadelphia, PA 19104

James H. Stock
Kennedy School of Government
Harvard University
Cambridge, MA 02138

1 Introduction

Starting with Fama and French (1988) and Poterba and Summers (1988), there has been a flurry of recent research which tests the random walk model of stock prices using multi-year returns. The intuitively appealing motivation for this work comes from the idea that it is easier to uncover low frequency correlations in the data, such as a slowly mean-reverting component of stock prices, if one examines the long term properties of returns. The conclusions reached in these articles is that there is indeed some statistically significant (albeit weak) evidence of long-run negative correlation in stock returns. The degree of significance has generally been assessed using conventional asymptotic theory, for example, by checking whether the estimates lie within two standard errors of the null. A number of recent authors, however, have concluded from Monte Carlo experiments that this approach provides a poor approximation to the sampling distribution of statistics involving multi-year returns, that is, for the sample sizes encountered in practice the normal distribution provides a poor approximation to the sampling distribution. The question, therefore, remains: is the mean reversion evidence due to a slow reverting component of stock prices or to the poor performance of the asymptotic theory in finite samples?

This paper adopts a different perspective on the problem of assessing the statistical significance of these statistics, primarily variance ratios and autocorrelations of multi-year returns. A commonly recognized feature of these statistics is that, even though the sample size might be large, the number of nonoverlapping observations can still be small. On an informal level, this suggests that there is not much independent information in a long time series of multi-year returns, which in turn suggests that conventional large-sample approximations to sampling distributions might perform poorly in practice. This paper formally implements this intuition. The key difference between our approach and the standard theory is that we treat the degree of overlap in the data (denoted by J) as tending to a fixed nonzero fraction (δ) of the sample size T, whereas the conventional theory treats J as fixed so that $\frac{J}{T}$ tends to zero.

Although the "fixed J" and the "$\frac{J}{T} \to \delta$" theories both provide large sample approximations that rely on versions of the central limit theorem, the two approaches yield sharply different qualititative results. For example, the fixed J theory implies that the more widely used multi-year correlation statistics, like variance ratios and sample autocorrelations, are consistent; in contrast, the $\frac{J}{T} \to \delta$ theory implies that these statistics will not be consistent. Under the fixed J theory, the statistics, when standardized by

$\sqrt{T}$, have asymptotic normal distributions; under the $\frac{J}{T} \rightarrow \delta$ theory these statistics have limiting distributions which are typically not normal, but rather have representations in terms of functionals of Brownian motion.

An attractive feature of the fixed J theory is that it is easy to use in practice. Similarly, even though the $\frac{J}{T} \rightarrow \delta$ theory implies that the limiting distributions of the statistics are not normal, we show that their representation under the null hypothesis that returns are unpredictable does not depend on any unknown parameters. Therefore, it is straightforward to compute their distributions by Monte Carlo methods. For example, to approximate the distribution of a particular statistic computed using say $J = 50$ and $T = 150$, the asymptotic results suggest performing a sequence of Monte Carlo simulations in which T increases and, for each T, J is set at $J = \frac{1}{3}T$.

Even though these two theories imply different distributions, they are both correct in their own regard. Whether or not the econometrician actually had a nonzero δ in mind when choosing a particular J is irrelevant. The major practical value of asymptotic theory is to provide a robust approximation to the small sample distribution of the statistics. Which theory is the most appropriate, therefore, is really an empirical question. Based on Monte Carlo simulations of the multi-year statistics, we find that the $\frac{J}{T} \rightarrow \delta$ theory provides a much better approximation to the sampling distributions than the conventional fixed J theory. For example, suppose that a χ^2-statistic is used to test the joint hypothesis that the J-period autocorrelations for $J = 2, 4, 6, 8, 10, 12, 16$, and 20 are all zero, with 120 observations (these are the ratios of $\frac{J}{T}$ used by Fama and French (1988)), and suppose that stock returns are standard normal random variates. Using the usual χ^2_8 asymptotic 5% critical value of 15.5, this statistic would in fact reject the null 18% of the time; using the $\frac{J}{T} \rightarrow \delta$ asymptotic 5% critical value of 25.4 would result in 4.8% rejections.

This paper is organized as follows. Section 2 examines the variance ratio statistics and Section 3 considers multi-year correlations. Section 4 presents a Monte Carlo study comparing the asymptotic approximations of the $\frac{J}{T} \rightarrow \delta$ theory to the fixed J theory. Section 5 applies the results to the question of whether stock prices exhibit mean reversion. Section 6 concludes.

2 Variance Ratio Statistics

As Cochrane (1988), Poterba and Summers (1988), and others have emphasized, variance ratio statistics provide an intuitively appealing way to search for mean reversion. If returns are serially uncorrelated, then the variance of the J-period return will increase linearly with J, but if there is mean reversion — so that returns are negatively correlated — then the variance of the J-period return will increase less than linearly.[1]

In this section we develop the alternative $\frac{J}{T} \to \delta$ approach to obtaining the asymptotic approximation to the null distribution of the variance ratio statistic. Throughout, it is assumed that returns R_t are unpredictable from their past values, except for a constant mean μ. That is,

$$R_t = \mu + \epsilon_t, \quad t = 1, \ldots, T \tag{1}$$

where $E(\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_1) = 0$. The returns are allowed to be conditionally heteroskedastic, as long as the average conditional heteroskedasticity tends to a constant variance. That is, let $h_t = E(\epsilon_t^2 | \epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_1)$; then $E\{\frac{1}{T}\sum_{t=1}^{T} h_t\} \to \sigma^2$ as $T \to \infty$ by assumption, where $\sigma^2 > 0$. Also, assume that $\sup_t E(\epsilon_t^4) < \infty$.

Two variance ratio statistics are considered in this section. The first, $\mathcal{J}_r(J)$, is computed using non-overlapping data; the second, $M_r(J)$, is computed using overlapping data. Lo and Mackinlay (1988,1989) analyzed the asymptotic properties of both when J is fixed and $T \to \infty$, and their notation is adopted here. Both variance ratio statistics are constructed using the J- period return,

$$x_t(J) = \sum_{i=0}^{J-1} R_{t-i}.$$

When constructed using the $\frac{T}{J}$ non-overlapping observations, the statistic is

$$\mathcal{J}_r(J) = \frac{\left(\frac{1}{T/J}\right)\left(\frac{1}{J}\right)\left[(X_J(J) - J\hat{\mu})^2 + (X_{2J}(J) - J\hat{\mu})^2 + \ldots + (X_T(J) - J\hat{\mu})^2\right]}{\hat{\sigma}_R^2} \tag{2}$$

where $\hat{\sigma}_R^2 = \frac{1}{T}\sum_{t=1}^{T}(R_t - \hat{\mu})^2$, $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t$, and where it is assumed in (2) that $\frac{T}{J}$ is an integer. The $M_r(J)$ statistic is constructed using overlapping observations on J-period returns:

$$M_r(J) = \frac{\frac{1}{JT}\sum_{t=J}^{T}(x_t(J) - J\hat{\mu})^2}{\hat{\sigma}_R^2}. \tag{3}$$

In the usual fixed J asymptotic treatment, under the null hypothesis the numerator and denominator converge in probability to $\text{var}(R_t)$, so that $\mathcal{J}_r(J)$ and $M_r(J)$ converge

[1]For additional applications, see Campbell and Mankiw (1987), Christiano and Eichenbaum (1989), Cochrane and Sbordone (1988), Faust (1989), Huizinga (1987), and Lo and Mackinlay (1988).

in probability to one. Lo and Mackinlay (1988,1989) have studied the fixed J asymptotic distribution of $\mathcal{J}_r(J)$ and $M_r(J)$. They assume that $\{R_t\}$ are i.i.d. with a normal distribution and show that:

$$\sqrt{T}(\mathcal{J}_r(J)-1) \xrightarrow{d} N(0, 2(J-1)) \tag{4}$$

$$\sqrt{T}(M_r(J)-1) \xrightarrow{d} N\left(0, \frac{2(2J-1)(J-1)}{3J}\right). \tag{5}$$

Lo and Mackinlay (1989) find that these approximations work well when J is small and T is large. They emphasize, however, that this large sample normal approximation is unsatisfactory for large J. For example, they show that for $J = \frac{1}{2}T$, the t-statistic constructed using the approximation (5) will never be less than -1.74 for all T. The poor performance of the fixed J asymptotic approximation when J is large relative to T indicates a need for an alternative approach in which $\frac{J}{T}$ is explicitly recognized to be large.

In the non-overlapping case (i.e. $\mathcal{J}_r(J)$), such an alternative approximation is in fact easy to develop. To simplify exposition, temporarily assume (unrealistically) that μ is known to equal zero so that $R_t = \epsilon_t$ and the terms involving $\hat{\mu}$ can be dropped from (2). In this case, when $\frac{T}{J}$ is an integer $\mathcal{J}_r(J)$ can be rewritten as

$$\mathcal{J}_r(J) = \frac{\frac{1}{T/J}\left[\left(\frac{1}{\sqrt{J}}\sum_{t=1}^{J}\epsilon_t\right)^2 + \left(\frac{1}{\sqrt{J}}\sum_{t=J+1}^{2J}\epsilon_t\right)^2 + \ldots + \left(\frac{1}{\sqrt{J}}\sum_{t=T-J+1}^{T}\epsilon_t\right)^2\right]}{\hat{\sigma}_R^2}. \tag{6}$$

If $\frac{J}{T} \to \delta$ as $T \to \infty$, the number of nonoverlapping observations used to construct $\mathcal{J}_r(J)$ remains fixed at $\frac{T}{J} = \frac{1}{\delta}$. (Note that as T increases $\frac{T}{J}$ will not remain an integer: for (6) formally to apply, take the limit along the sequence $T = \frac{J}{\delta}$, $J = 1, 2, 3, \ldots$, where $\frac{1}{\delta}$ is an integer.) This results in a simple limiting distribution for $\mathcal{J}_r(J)$. Because $J \to \infty$ as $T \to \infty$, each of the partial sums in (6) converge in distribution to independent $N(0, \sigma^2)$ variates. Upon dividing by $\hat{\sigma}_R^2$ (which is consistent for σ^2), one finds directly that this statistic has a limiting chi-squared distribution:

$$\mathcal{J}_r(J) \xrightarrow{d} \delta\chi^2_{\frac{1}{\delta}}. \tag{7}$$

This simple result highlights three key features that typically distinguish the $\frac{J}{T} \to \delta$ asymptotic approximations from the fixed J approximations. First, this variance ratio statistic is not consistent, but rather has a nondegenerate limiting distribution without first scaling by $\sqrt{T}$. Second, this limiting distribution is not the normal distribution in (4), but is non-normal — here, a $\chi^2_{\frac{1}{\delta}}$, divided by its degrees-of-freedom. Third, this result holds even if there is nonnormality and heteroskedasticity of the type permitted following (1).

The key observation used to obtain the alternative distribution (7) is that, when J is large relative to T, the multi-year return behaves more like a random walk than like a stationary process. To see this, it is useful to contrast the behavior of $x_t(J)$ in the fixed J and $\frac{J}{T} \to \delta$ cases. In the fixed J case, $x_t(J)$ is a random variable with its distribution on the real line and with variance $J\sigma^2$. Thus $\{x_t(J)\}$ is treated as a stationary stochastic process. In contrast, in the $\frac{J}{T} \to \delta > 0$ case the variance of $x_t(J)$ tends to infinity. In fact, $\frac{1}{\sqrt{T}}x_t(J)$ obeys a central limit theorem: $\frac{1}{\sqrt{T}}x_t(J) \xrightarrow{d} N(0, \delta\sigma^2)$. Because in this nesting the variance of $x_t(J)$ increases without bound, $\{x_t(J)\}$ cannot be treated as a stationary stochastic process. Hence the analogy between $x_t(J)$ and a random walk: as T gets large, the variance of each tends to infinity, but rescaled by $\frac{1}{\sqrt{T}}$ (or $\frac{1}{\sqrt{J}}$) they each obey a Central Limit Theorem.

The argument used to obtain the limiting distribution of $M_r(J)$ when $\frac{J}{T} \to \delta$ involves an additional technical difficulty. For now, continue to assume that μ is known to equal zero. Although the usual Central Limit Theorem applies to $\frac{1}{\sqrt{T}}x_t(J)$ for a fixed t, it does not apply to $\{\frac{1}{\sqrt{T}}x_t(J)\}$ treated as a stochastic process in t — i.e. it does not apply *simultaneously* for all t. Fortunately, however, there is an alternative limit theory that can be used to develop an asymptotic characterization of $x_t(J)$ in the $\frac{J}{T} \to \delta$ nesting. This theory — alternatively referred to as the Functional Central Limit Theorem (FCLT), Donsker's Theorem, or the invariance principle, and discussed in more detail in the Appendix — implies that, treated as a stochastic process, partial sums of returns converge to a Brownian motion process on the unit interval. This form of convergence is denoted by "$\Rightarrow$". To make this precise, let $[\cdot]$ denote the greatest lesser integer function. Then the FCLT states that, as $\frac{J}{T} \to \delta$,

$$\frac{1}{\sqrt{T}} \sum_{s=1}^{[T\lambda]} R_s \Rightarrow \sigma W(\lambda) \tag{8}$$

where $W(\lambda)$ is standard Brownian motion restricted to the unit interval.

The result (8) is the key ingredient in deriving the $\frac{J}{T} \to \delta$ asymptotic result. Write

$\frac{1}{\sqrt{T}}x_t(J)$ as

$$
\begin{aligned}
\frac{1}{\sqrt{T}}x_t(J) &= \frac{1}{\sqrt{T}}\sum_{s=t-J+1}^{t} R_s \\
&= \frac{1}{\sqrt{T}}\sum_{s=1}^{t} R_s - \frac{1}{\sqrt{T}}\sum_{s=1}^{t-J} R_s
\end{aligned}
\tag{9}
$$

If $\frac{t}{T} \to \lambda > 0$ and if $\frac{J}{T} \to \delta > 0$, then $\frac{t-J}{T} \to \lambda - \delta$ and, from (8) and (9),

$$
\frac{1}{\sqrt{T}}x_{[T\lambda]}(J) \Rightarrow \sigma(W(\lambda) - W(\lambda - \delta)).
\tag{10}
$$

The distribution of the variance ratio statistic $M_r(J)$ can now be obtained by applying (10). First consider the numerator:

$$
\begin{aligned}
\frac{1}{JT}\sum_{t=J}^{T} x_t^2(J) &= \left(\frac{T}{J}\right)\frac{1}{T}\sum_{t=J}^{T}\left(\frac{1}{\sqrt{T}}x_t(J)\right)^2 \\
&= \left(\frac{T}{J}\right)\int_{\frac{J}{T}}^{1}\left(\frac{1}{\sqrt{T}}x_{[T\lambda]}(J)\right)^2 d\lambda \\
&\Rightarrow \frac{1}{\delta}\int_{\delta}^{1}\sigma^2(W(\lambda) - W(\lambda-\delta))^2 d\lambda.
\end{aligned}
$$

The denominator in $M_r(J)$, which only involves the one-period return, converges in probability to σ^2. Thus, when μ is known to equal zero, $M_r(J)$ has the limiting representation,

$$
M_r(J) \Rightarrow \frac{1}{\delta}\int_{\delta}^{1}(W(\lambda) - W(\lambda-\delta))^2 d\lambda.
$$

In practice, μ is unknown and all the data are used to construct the variance ratio; thus the empirically relevant statistic to examine is $M_r(J)$ in (3) including the term in $\hat{\mu}$. Incorporating $\hat{\mu}$ is conceptually straightforward but complicates the algebra somewhat. The limiting representation of $M_r(J)$ when μ is estimated, which is derived in the Appendix, is

$$
M_r(J) \Rightarrow \frac{1}{\delta}\int_{\delta}^{1} Y_\delta(\lambda)^2 d\lambda
\tag{11}
$$

where $Y_\delta(\lambda) = W(\lambda) - W(\lambda - \delta) - \delta W(1)$.

These results warrant some remarks. First, while $M_r(J)$ is consistent for one under the fixed J treatment, in (11) it has a limiting distribution. An intuitive intrepretation of this result is that, in the $\frac{J}{T} \to \delta$ nesting, the amount of overlap in the data is so large that one is left with only a finite number $\frac{1}{\delta}$ of non-overlapping sets of J-period returns.

Second, the limiting distribution of $M_r(J)$ is not normal, but rather has a representation in terms of functionals of Brownian motion. Because this representation does not depend on any unknown parameters of the problem, however, it is straightforward to compute this distribution by Monte Carlo methods (this is done in Section 4). Finally, note that the details of the argument and the final representation change if the statistic is computed by deviating returns from their sample average. The latter procedure is of course appropriate in practice.

3 Multi-year Autocorrelation Statistics

Fama and French (1988) examine the autocorrelation of multi-year returns using the regression,

$$x_{t+J}(J) = \alpha(J) + \beta(J)x_t(J) + \epsilon_{t+J}(J), \tag{12}$$

where, as in Section 2, $x_t(J)$ denotes the J-period return. The multi-year correlation is estimated by the ordinary least squares estimator $\hat{\beta}(J)$.

The usual asymptotic approximation to the distribution of $\hat{\beta}(J)$ is obtained by holding J fixed and letting $T \to \infty$. Under the null hypothesis $\beta(J) = 0$, Richardson and Smith (1989) use this approach to show that

$$\sqrt{T-2J+1}\hat{\beta}(J) \xrightarrow{d} N\left(0, \frac{2J^2+1}{3J}\right)$$

where T-2J+1 is the number of observations used to compute $\hat{\beta}(J)$. Thus the distribution of $\sqrt{T-2J+1}\hat{\beta}(J)$ is approximated by a normal with a variance that increases with J.

When J is large, the reasoning leading to the $\frac{J}{T} \to \delta$ approach in Section 2 suggests a similar treatment of $\hat{\beta}(J)$. The method of calculation is similar to that used for the overlapping-data variance ratio statistic. The result, derived in the appendix, is that in the limit $\hat{\beta}(J)$ has the same distribution as a functional of Brownian motion:

$$\hat{\beta}([T\delta]) \Rightarrow \frac{\int_\delta^{1-\delta} X_\delta^0(\lambda)X_\delta^1(\lambda+\delta)d\lambda}{\int_\delta^{1-\delta} X_\delta^0(\lambda)^2 d\lambda} \equiv \beta_\delta^* \tag{13}$$

where $X_\delta^0(\lambda) = X_\delta(\lambda) - \frac{1}{1-2\delta}\int_\delta^{1-\delta} X_\delta(s)ds$, $X_\delta^1(\lambda) = X_\delta(\lambda) - \frac{1}{1-2\delta}\int_\delta^{1-\delta} X_\delta(s+\delta)ds$, and where $X_\delta(\lambda) = W(\lambda) - W(\lambda - \delta)$.

This limiting representation of $\hat{\beta}(J)$ depends only on δ and the standard Brownian motion process $W(\lambda)$, not on μ or any of the nuisance parameters describing the conditional or unconditional distributions of ϵ_t. Therefore, it is easy to compute its distribution by Monte Carlo methods.

There are several extensions of the result (13) that have practical value. In particular, it is reasonable to want to test the hypothesis that several multi-year correlations simultaneously equal zero using a F-test. Let $\hat{\beta} = (\hat{\beta}(J_1) \quad \hat{\beta}(J_2) \quad \cdots \quad \hat{\beta}(J_K))'$, where $J_1 < J_2 < \cdots < J_K$ are the K different periods over which the multiperiod returns are computed. Suppose that each of the elements of $\hat{\beta}$ is estimated using all the data available for that element, i.e. $T - 2J_i + 1$ observations. Let $V_T = V_T(J_1, \ldots, J_K)$ denote the fixed J asymptotic covariance matrix of the vector $\{\sqrt{T - 2J_i + 1}\hat{\beta}(J_i)\}$ under the null hypothesis provided by Richardson and Smith (1989).[2] Then a natural test statistic is the Wald statistic:

$$F_T = \hat{\beta}'(\Upsilon_T^{-1} V_T \Upsilon_T^{-1})^{-1} \hat{\beta} \tag{14}$$

where $\Upsilon_T = \mathrm{diag}(\sqrt{T - 2J_1 + 1}, \ldots, \sqrt{T - 2J_K + 1})$.

Although F_T does not have the usual large sample χ^2_K distribution if $\frac{J}{T} \to \delta$ ($\hat{\beta}$ is nonnormally distributed), F_T provides a simple way to check whether the restrictions on the multiperiod correlations apply simultaneously. If $V(J_1, \ldots, J_K)$ is evaluated for $\frac{J_i}{T} \to \delta_i$, $0 < \delta_1 < \delta_2 < \cdots < \delta_K < 1$, then $\Upsilon_T^{-1} V_T \Upsilon_T^{-1} \to V^*$.[3] With this observation in hand, the $\frac{J}{T} \to \delta$ asymptotic representation for F_T obtains directly from (13):

$$F_T \Rightarrow \beta^*(\delta_1, \ldots, \delta_k)' V^{*-1} \beta^*(\delta_1, \ldots, \delta_k)$$

where $\beta^*(\delta_1, \ldots, \delta_k)$ is a column vector with the i-th element given by $\beta^*_{\delta_i}$ as expressed in (13) with δ_i replacing δ.

A related statistic is the sum of the multi-year correlations. An heuristic motivation for considering this statistic is that, while any individual multi-year correlation might not differ significantly from zero, if there is mean reversion in stock prices over a long horizon then the sum of the correlations might be significantly negative. If there are a fixed number K of autocorrelations under consideration, then this statistic is $R_T = \sum_{i=1}^K \hat{\beta}(J_i)$. Adopt the same notation as in the case of the F-statistic. Then, as $\frac{J_i}{T} \to \delta_i$, $i = 1, \ldots, K$, $R_T \Rightarrow \sum_{i=1}^K \beta^*_{\delta_i}$. Like the individual estimators of the multi-year correlations,

[2]Richardson and Smith (1989) show that $V_{jj} = \frac{2J_j^2+1}{3J_j}$ and, for $J_k > J_j$, $V_{jk} = \frac{s(J_j,J_k)+J_j^2}{J_j J_k}$, where $s(J_j, J_k) = 2\sum_{m=1}^{J_j-1}(J_j - m)min(J_j, J_k - m)$.

[3]Specifically, $V^*_{jk} = 2\{\delta_j\delta_k\sqrt{(1-2\delta_j)(1-2\delta_k)}\}^{-1}\int_0^{\delta_j} x\min(\delta_j, \delta_k - \delta_j + x)dx$ for $0 < \delta_j < \delta_k$ and $V^*_{ii} = \frac{2\delta_i}{3(1-2\delta_i)}$.

R_T does not converge in probability to zero under the null but rather has a nonnormal limiting distribution. Alternative statistics with a similar motivation are the maximum correlation, $\max_i \hat{\beta}(J_i)$, the minimum correlation, or the maximum absolute correlation. These statistics are handled the same way as the sum R_T.

As a final note, Fama and French (1988) followed the conventional practice of computing standard errors for $\hat{\beta}(J)$ taking into account the moving average structure of the errors $\epsilon_t(J)$ in (12) induced by the use of overlapping data. The procedure they used was that propounded by Hansen and Hodrick (1980). Because this estimator is commonly used in related applications in finance with moving-average errors in regression models, it is of interest to consider the large-sample properties of the Hansen-Hodrick estimator when $\frac{J}{T} \to \delta$.

The differing behavior of $\hat{\beta}(J)$ under the fixed J and the $\frac{J}{T} \to \delta$ limits suggest the limiting behavior of the Hansen-Hodrick estimator might also differ in the two asymptotic nestings. In the fixed J treatment, the Hansen-Hodrick estimator $\hat{\omega}^2(J)$ of the variance of $\hat{\beta}(J)$ converges in probability to the true variance of the estimator, in this case, $\frac{2J^2+1}{3J}$. In contrast, with the assumption that $\frac{J}{T} \to \delta$, in the appendix we show that $\hat{\omega}^2(J)$ does not converge in probability to a constant but rather, after dividing by T, converges weakly to a functional of Brownian motion. Note that the Hansen-Hodrick estimator uses a specific set of weights on the sample autocovariances of the errors and regressors. The same qualitative results hold for estimators based on alternative weighting functions.

4 Monte Carlo Evidence

The purpose of this section is twofold. The first objective is to show how to calculate the asymptotic distribution of the statistics under the $\frac{J}{T} \to \delta$ theory. The asymptotic results of Sections 2 and 3 provide limiting representations of estimators and test statistics as functionals of Brownian motion. These random functionals typically do not have a standard distribution, so the usual approach of looking up previously tabulated critical values will not work. However, these limiting representations provide a clear recipe for obtaining asymptotic critical values using Monte Carlo simulations. The results imply that estimators generated from *any* null model (1) will have the same distribution as $\frac{J}{T} \to \delta$ and $T \to \infty$. Thus, to approximate the distribution of a particular statistic using (for example) $J = 60$ and $T = 360$, the asymptotic results suggest performing a sequence of Monte Carlo simulations in which T increases and, for each T, J is set at $J = \frac{1}{6}T$. Just how large

T must be in the Monte Carlo simulation to provide accurate numerical approximations to the limiting distribution will vary across statistics, but can be determined by checking when the percentiles from the Monte carlo distributions converge as T gets large.[4] The second, related objective is to compare the large-sample $\frac{J}{T} \to \delta$ approximation to the fixed J normal approximation. Specifically, in the controlled environment of Monte Carlo experiments, we wish to ascertain which theory (fixed J or $\frac{J}{T} \to \delta$) provides a better approximation to the finite-sample distribution of the multi-year statistics. In these Monte Carlo experiments, the one-period returns are generated according to model (1) where ϵ_t is i.i.d. $N(0, \sigma^2)$ and, without loss of generality, $\sigma = 1$ and $\mu = 0$.

4.1 The Variance Ratio Statistic

The first statistic considered is the variance ratio $M_r(J)$ with $\delta = \frac{1}{3}$.[5] Various percentiles of the distribution of $M_r\left(\left[\frac{T}{3}\right]\right)$ are given in Table 1. The first two rows present the asymptotic percentiles based on fixed J normal approximations. The first row presents the normal distribution (5), where the variance $\Omega^*(\delta) = \frac{4\delta}{3}$ is the limit of the sequence of variances $\frac{2(2J-1)(J-1)}{3JT}$ as $T \to \infty$ and $\frac{J}{T} \to \delta$. In deriving this result, Lo and Mackinlay (1988,1989) suggest an alternative, "bias-adjusted" approximation in which the normal distribution is scaled by the factor $b_T(J) = \frac{(T-J+1)(T-J)}{(T-1)T}$. As $T \to \infty$ and $\frac{J}{T} \to \delta$, this factor has the limit $b_T(J) \to (1-\delta)^2 \equiv b(\delta)$. This approximation is presented in the second row of Table 1. The remaining rows in Table 1 contain Monte Carlo results for the indicated sample sizes, with J evaluated at $\frac{T}{3}$.

Three aspects of these results are noteworthy. First, and most important, the convergence of the Monte Carlo percentiles appears rapid: there is little difference between the $T = 60$ and $T = 2880$ simulation results. Thus the $\frac{J}{T} \to \delta$ asymptotic distribution (the limit of the sequence of these Monte Carlo distributions) seems to yield a satisfactory approximation to the distribution of the variance ratio even for a relatively small value of T. Second, the percentiles based on the fixed J approximation differ substantially from

[4]This approach to the numerical evaluation of nonstandard asymptotic distributions (by a sequence of Monte Carlo simulations) is increasingly common in the empirical "unit roots" literature. See, for example, Perron (1989), Stock (1988) and Stock and Watson (1989).

[5]$\delta = \frac{1}{3}$ lies in the range of values found in the empirical literature. At one extreme, Huizinga (1987) reports this statistic for $J = 1, 2, \ldots, \frac{12}{13}T$. Poterba and Summers (1988) consider $\frac{J}{T}$ ranging from $\frac{1}{30}$ to $\frac{1}{7.5}$ for U.S. data, and as large as $\frac{1}{3.25}$ for foreign data. Cochrane and Sbordone (1988) consider $\frac{J}{T}$ up to $\frac{1}{2}$. In their Monte Carlo simulations, Lo and Mackinlay (1989) compute $M_r(J)$ for J ranging from 1 to $\frac{1}{2}T$.

the Monte Carlo percentiles. As noted by Lo and Mackinlay (1989) and Poterba and Summers (1988), the sampling distribution is skewed, with its mean well below one and its median below the mean. These authors' results indicate that this skewness persists for values of δ less than $\frac{1}{3}$; see for example Poterba and Summers (1988, Figure 1) for $\delta = .133$. Third, the first fixed J normal distribution provides a clearly unsatisfactory approximation to the percentiles of the variance ratio statistic. The second approximation is an improvement; the mean of the distribution $(1-\delta)^2$ can be shown to equal the mean of the limiting functional (11). This approximation, however, fails to capture the skewness of the Monte Carlo distributions and incorrectly puts substantial mass below zero (indeed, the lower 5% one-sided critical value is negative).

4.2 The Multi-year Correlation F-test

The next statistic considered is the Wald statistic (14) for tests of the hypothesis that $\beta(J_i) = 0$, $i = 1, \ldots, K$. The $\{J_i\}$ correspond to the multi-year correlations examined by Fama and French (1988): with 720 monthly observations, they examined 1,2,3,4,5,6,8 and 10-year returns, corresponding to $\delta = \left(\frac{1}{60}, \frac{1}{30}, \frac{1}{20}, \frac{1}{15}, \frac{1}{12}, \frac{1}{10}, \frac{1}{7.5}, \frac{1}{6}\right)$. The distributions are summarized in Table 2. Consistent with the large sample theory, the Monte Carlo percentiles converge, although the convergence is slower than in Table 1. The Monte Carlo 10% and 5% critical values differ substantially from the percentiles of the conventional χ^2_8 distribution: the $\frac{J}{T} \to \delta$ distribution places more weight in the right tail. There is little difference between the Monte Carlo distribution for $T = 120$ and $T = 2880$. Evidently, the $\frac{J}{T} \to \delta$ asymptotic distribution provides a better approximation to the distribution of the F-statistic than does the fixed J distribution. Inferences drawn using the χ^2_8 approximation would reject (1) far too often.

4.3 The Multi-year Correlation Sum Statistic

A third, related statistic is the sum of these multi-year correlation coefficients, $R_T = \sum_{i=1}^{K} \hat{\beta}(J_i)$, for the Fama-French choice of $\{\delta_i\}$. The fixed J normal approximation is based on $\Upsilon_T \hat{\beta} \xrightarrow{d} N(0, V_T)$, where $\hat{\beta}$, Υ_T and V_T are defined as in (14). Let ι be the K-vector, $\iota = (1 \cdots 1)'$. Then $R_T = \iota' \hat{\beta}$ is approximately (for fixed J) $N(0, \tau^2)$, where $\tau^2 = lim_{T\to\infty} \iota'(\Upsilon_T^{-1} V_T \Upsilon_T^{-1})\iota = \iota' V^* \iota$. The results, presented in Table 3, indicate that the $\frac{J}{T} \to \delta$ approximation is substantially better than the fixed J normal approximation, even for T as low as 60. Interestingly, the distribution of R_T is skewed and has a negative mean: 79% of the mass of the $\frac{J}{T} \to \delta$ asymptotic distribution falls below zero.

4.4 Implications

These simulations indicate substantial differences in the two competing asymptotic distributions for these statistics. These differences are largest in the tails. In each case, the $\frac{J}{T} \to \delta$ limiting distribution typically provides a better approximation to the sampling distributions for small T than does the fixed J approximation. The convergence to the asymptotic limit typically occurs quickly, arguably by $T = 120$. Although this last finding presumably is sensitive to the errors used to construct the pseudo-data, its implication is that Monte Carlo experiments with relatively few observations can be used to approximate numerically the $\frac{J}{T}$ asymptotic distributions.

These simulations suggest a different perspective on these multi-year return statistics. Under the fixed J asymptotics, $\hat{\beta}(120)$ (for example) would be considered an estimator of the 10-year autocorrelation. From the perspective of empirical inference, however, with 720 observations these results suggest that it is more fruitful to think of $\hat{\beta}(120)$ as the $\frac{1}{6}$-th sample autocorrelation. In particular, it is misleading to use the fixed J consistency to guide one's intuition when interpreting the point estimates.

In general, these statistics have nonnormal limits. It thus makes little sense to report standard errors, except perhaps as a measure of the spread of the distribution. However, because most readers of empirical research are in the habit of using the "two standard error" rule-of-thumb, reporting standard errors is likely to be misleading. A preferable procedure is to report asymptotic p-values. The asymptotic p-values are readily constructed by a sequence of Monte Carlo simulations. Moreover, although it might be tempting to use a fixed J approximation that entails a correction for small-sample bias (Lo and Mackinlay (1988,1989), Cochrane (1988)), these results suggest that in general doing so will not provide a satisfactory approximation.[6]

It is tempting to interpret this $\frac{J}{T} \to \delta$ nesting as representing what the researcher would do were he or she faced with a larger sample. Indeed, some articles contain explicit discussions of how J might be changed with T. However, this interpretation is tangential to this application of asymptotic theory, which is best understood simply as a device for approximating the sampling distribution for fixed J and T. The issue here is not what the

[6] Poterba and Summers (1988) perform what this theory suggests is the correct Monte Carlo simulation, and report selected Monte Carlo p-values in the text. In their tables, however, they focus on standard errors of bias-adjusted point estimates.

researcher would have done with different T. Rather, the point is to approximate the distribution of these statistics obtained by repeated samples of the same length T generated by the same null model. The asymptotic results show that these $\frac{J}{T} \to \delta$ approximations are valid even if the returns are non-normal and conditionally heteroskedastic, subject to the weak restrictions following (1).

5 Empirical Results

This section examines two key sets of empirical findings about mean reversion in U.S. stock prices: the multi-year correlations reported by Fama and French (1988) and the long-horizon variance ratio statistics reported by Poterba and Summers (1988). The analysis is based on point estimates reported in the respective articles.

5.1 Results

5.1.1 Correlations of Multi-Year Returns

Table 4 presents point estimates and corresponding $\frac{J}{T} \to \delta$ asymptotic p-values for autocorrelations of multi-year returns on value weighted, equal weighted and size portfolios, based on monthly returns on the New York Stock Exchange (NYSE) from 1926 to 1985. The return horizons, portfolios, and OLS estimates of $\beta(J)$ are taken from Fama and French (1988, Table 2). The values of the cumulative $\frac{J}{T} \to \delta$ distribution are given in parentheses; these are the asymptotic p-values for testing the null hypothesis (1) against the one-sided alternative that the true coefficient is negative. The asymptotic distribution was computed by Monte Carlo simulation with $T = 720$, although the results of Section 4 suggest that a numerically satisfactory approximation could have been achieved with a smaller T, say $T = 360$. It should be emphasized that although $T = 720$ corresponds to the sample size, the interpretation of this simulation is not to correct for small sample bias, but to evaluate the $\frac{J}{T} \to \delta$ asymptotic distribution.

As Fama and French (1988) emphasized, the point estimates are generally consistent with mean reversion. They assessed the statistical significance of these autocorrelations by (i) computing the Hansen-Hodrick standard errors; (ii) performing a Monte Carlo simulation of the uncorrelated null (1) with $T = 720$ to obtain mean "biases" of the OLS estimators; (iii) subtracting this "bias" from the point estimates; and (iv) checking whether the "bias-adjusted" point estimates are two (Hansen-Hodrick) standard errors away from zero. For the portfolios in Table 4, the Fama-French procedure led to 19 of the 96 bias-adjusted point estimates being more than 2 standard errors away from zero.

The $\frac{J}{T} \rightarrow \delta$ asymptotic p-values present a different picture. One way to perform a 5% two-sided hypothesis test of (1) is to reject if the point estimate falls in the extreme 2.5% of either tail; such point estimates are indicated by an asterisk in Table 4. Using the $\frac{J}{T} \rightarrow \delta$ asymptotic p-values, this results in 3 rejections at the 5% level rather than 19. Of course, many of the p-values remain small, so some evidence of mean reversion in the individual coefficients remains. Richardson and Smith (1989) and Richardson (1989) have argued, however, that, because the individual point estimates are highly correlated, it is inappropriate to emphasize any single $\hat{\beta}(J)$ or its t-statistic. We therefore consider the average of the point estimates for each portfolio as well; this statistic and its (one-sided) p-value are reported in the final column. (As discussed in Section 4, the motivation for considering this statistic is that, by preserving the signs of the point estimates, it might exhibit improved power against the mean reverting alternative.) For either the equal or value weighted portfolios, the hypothesis that the true value of this sum is zero cannot be rejected at the (two-sided) 40% significance level.

5.1.2 Ratios of Variances of Long Differences

Point estimates and one-sided p-values of variance ratios are presented in Table 5. The point estimates are based on those in Poterba and Summers (1988, Table 3).[7] The p-values were computed using the $\frac{J}{T} \rightarrow \delta$ asymptotic distribution, numerically approximated by Monte Carlo simulation.

The evidence about the mean-reversion hypothesis in Table 5 is at best mixed. Consistent with Poterba and Summers' (1988) interpretation, the "Average" statistic in Table 5(A) provides evidence of mean reversion in real returns (at the 10% level using a two-sided test). In both Table 5(A) and 5(B), however, the F_T statistic indicates rejection at the 10% level only for equal-weighted excess returns. The only 5% level two-sided

[7] Poterba and Summers (1988) report "bias-adjusted" ratios of the J-month variance over the 12-month variance. Their bias adjusted estimates were obtained by dividing the point estimates of the variance ratios by the mean from a Monte Carlo simulation of these statistics under an i.i.d. Gaussian null model. The point estimates in our Table 5(A) are "bias unadjusted": the estimates are the Poterba-Summers "bias adjusted" estimates, multiplied by the mean from an identical Monte Carlo simulation (which was also the Monte Carlo simulation used to compute p-values). For this table, sampling variances are scaled by $(T-J)^{-1}$ rather than T^{-1}. The estimates in Table 5(B) were obtained from the point estimates in Table 5(A). The first column in 5(B) is the inverse of the values in the first column of Table 5(A). For the next 7 columns, the Table 5(B) entry is the entry in the corresponding column in Table 5(A), divided by the Table 5(A), column 1 entry for that portfolio.

rejection in these summary statistics is for equal-weighted excess returns in Table 5(B). Here, the pattern is not so much one of mean reversion (all variance ratios below one), but of too-high ratios (positive autocorrelation) followed by too-low ratios.

5.2 Discussion

This analysis has emphasized the reporting of asymptotic p-values. In practice, these are just the fraction of times that pseudo-data generated under (1) produces point estimates or test statistics more adverse to the null than the one computed from the data. The $\frac{J}{T} \to \delta$ distribution theory therefore provides a formal justification for doing what many researchers might consider "natural" in these situations, evaluating statistical significance by Monte Carlo simulation. The conclusions drawn using these p-values are typically weaker than the inferences drawn using conventional standard errors. Poterba and Summers (1988) argue that it is appropriate to use a less conservative criterion than the 5% level used here. The choice of test level is an issue on which we do not take a stand; the point is that, for any level test, there are substantially fewer rejections using the $\frac{J}{T} \to \delta$ asymptotics than the fixed J asymptotics.

As Richardson (1989) and Richardson and Smith (1989) emphasize, the statistics considered here are far from independent. Thus, the combined results across portfolios, horizons, and statistics provides little new evidence beyond the results for individual statistics or individual portfolios. This suggests putting the greatest weight on overall summaries such as the sum statistic. In most cases, these statistics do not reject the null at the 10% level; the one exception to this is the 10%-level rejection of the Poterba-Summers equal-weighted real "average" statistic. Individually, each of these two arguments — the use of $\frac{J}{T} \to \delta$ distribution theory and the Richardson and Smith (1989) emphasis on overall summary statistics — weakens the case against the null hypothesis. Taken together, little evidence remains against the no-reversion null.

Recently, Kim, Nelson and Startz (1988) reached similar substantive conclusions using what at first might appear to be a quite different approach to approximating sampling distributions. They computed p-values by a technique closely related to the bootstrap, in which pseudo-data are computed by shuffling the original returns. Their motivation was that the underlying returns are arguably non-normally distributed, and they suggested that this might be one reason why conventional asymptotic theory delivers inadequate approximations. In view of the $\frac{J}{T} \to \delta$ asymptotic results, what Kim, Nelson and Startz

(1988) detected with their reshuffling scheme is simply the difference between the fixed J and the $\frac{J}{T} \to \delta$ limiting distributions. In fact, on a computational level their scheme and the one used here are closely related. The asymptotic p-values reported in Tables 4 and 5 were computed by Monte Carlo simulation with pseudo-random Gaussian returns. The "reshuffling" critical values reported by Kim, Nelson and Startz (1988) were also computed by Monte Carlo, except that the returns used to construct the pseudo-data are a shuffled version of the original returns. Because the asymptotic theory indicates that non-normality is unimportant for the ultimate asymptotic distribution, one would expect the two procedures to provide similar p-values, and indeed they do.[8]

6 Conclusion

This work has two main practical implications. First, the $\frac{J}{T} \to \delta$ approach to asymptotics is simple to implement. Although some of the limiting expressions might seem daunting, they typically are just the continuous-time counterparts of the moments computed using the discrete-time data, evaluated using the various limiting continuous-time stochastic processes. In practice, the primary role of these limiting expressions is simply to show that a $\frac{J}{T} \to \delta$ limiting distribution exists and that it can be approximated numerically by a sequence of Monte Carlo simulations with the appropriate $\frac{J}{T}$ nesting. Second, the difference in performance between the conventional and the $\frac{J}{T} \to \delta$ asymptotic approximations can be substantial; at least for the estimators considered here, the $\frac{J}{T} \to \delta$ approach represents a considerable improvement.

Although we have focused on multi-year correlations and variance ratio statistics, this approach can be extended to analyze other statistics based on multi-year returns. An example is the regression of one-month returns onto J-month returns as suggested by Jegadeesh (1989). The extension of this approach to multivariate statistics, such as the regression of multi-year returns on lagged dividend-price ratios, is a topic of ongoing research.

[8] As a methodological point, the reshuffling scheme of Kim, Nelson and Startz (1988) does not address the possibility that the usual approximation is unsatisfactory because of conditional heteroskedasticity (any conditional heteroskedasticity in returns is destroyed by the reshuffling). In contrast, the asymptotic results reported in Sections 2 and 3 handle this possibility and suggest that conditional heteroskedasticity is not the culprit in the performance of the conventional approximations. This statement applies to heteroskedasticity that "averages out" in the sense following (1). Much of the variance in returns comes from the Depression period, however, suggesting a nonstationarity of the conditional variances. This is reinforced by the findings of Pagan and Schwert (1989) for this period. An interesting, open question is to extend the results of Sections 2 and 3 to more persistent forms of heteroskedasticity.

These alternative limiting distributions present a revised picture of the evidence for mean reversion in the U.S. data. For example, based on the $\frac{J}{T} \rightarrow \delta$ asymptotic distribution, the number of multi-year correlations significant at the 5% level is reduced substantially from the number reported by Fama and French (1988). The conclusion from this analysis is that the case for mean reversion is less pronounced when the evidence is interpreted in light of the $\frac{J}{T} \rightarrow \delta$ rather than fixed J approximations. A more general conclusion from this analysis concerns alternatives that are detectable only by examining a statistic based on a horizon that is a fixed fraction of the sample size. For such alternatives, statistics such as those considered here are incapable of providing decisive evidence: the multi-year autocorrelation and variance ratio statistics are not consistent under this alternative nesting, but rather have nondegenerate limiting distributions. Thus the data ought not be expected to provide many insights about economic theories in which mean reversion at long horizons plays a central role.

Appendix

The two key results that will be used repeatedly below are the FCLT and the continuous mapping theorem. The FCLT is a central limit theorem for standardized partial sums of ϵ_t, treated as random elements of the space of functions on the unit interval that are right continuous with left limits, $D[0,1]$. Let

$$S_\tau = \sum_{t=1}^{\tau} \epsilon_t, \quad \tau = 1, 2, \ldots, T.$$

and set $S_0 = 0$. The FCLT states that the random function $\{\frac{1}{\sigma\sqrt{T}} S_{[T\lambda]}\}$, $\lambda \in [0,1]$, converges (weakly) to a standard Brownian motion process on the unit interval, $W(\lambda)$: as $T \to \infty$,

$$\frac{1}{\sigma\sqrt{T}} S_{[T\lambda]} \Rightarrow W(\lambda). \tag{15}$$

That is, as $T \to \infty$, the distribution of the random function $\frac{1}{\sigma\sqrt{T}} S_{[T\lambda]} \in D[0,1]$ converges to the distribution of $W(\lambda)$. The continuous mapping theorem states that, if g is a continuous function from $D[0,1]$ to $\Re^k$, then

$$g\left(\frac{1}{\sigma\sqrt{T}} S_{[T\lambda]}\right) \Rightarrow g(W(\lambda)). \tag{16}$$

More generally, (16) holds if g is measurable and has discontinuities on a set A_g such that $\Pr(W \in A_g) = 0$ (Hall and Heyde (1980, Theorem A.3)).[9]

Autocorrelation of Multi-Year Returns

The OLS estimator of the sample autocorrelation of J-period returns is

$$\hat{\beta}(J) = \frac{\sum_{t=J}^{T-J} x_t^0(J) x_{t+J}^1(J)}{\sum_{t=J}^{T-J} x_t^0(J)^2} \tag{17}$$

where $x_t^0(J) = x_t(J) - \frac{1}{T-2J+1}\sum_{t=J}^{T-J} x_t(J)$ and $x_{t+J}^1(J) = x_{t+J}(J) - \frac{1}{T-2J+1}\sum_{t=J}^{T-J} x_{t+J}(J)$. The $\frac{J}{T} \to \delta > 0$ nesting is handled by noting that, as in (8), $\frac{1}{\sqrt{T}} x_t^0(J)$ and $\frac{1}{\sqrt{T}} x_t^1(J)$ can be written in terms of partial sums of $\{\epsilon_t\}$ which obey the FCLT. Thus, for $\lambda \geq \delta$,

$$\begin{aligned} \frac{1}{\sqrt{T}} x_{[T\lambda]}^0([T\delta]) &\Rightarrow \sigma X_\delta^0(\lambda) \\ \frac{1}{\sqrt{T}} x_{[T\lambda]}^1([T\delta]) &\Rightarrow \sigma X_\delta^1(\lambda) \end{aligned} \tag{18}$$

[9]See Billingsley (1968) or Hall and Heyde (1980) for an introduction to functional central limit theory. Additional technical conditions on ϵ_t (beyond (1)) are available to ensure the convergence (15). For discussions of these and alternative conditions and of convergence on $D[0,1]$, see Herrndorff (1984,Corollary 1), Ethier and Kurtz (1986) or Phillips (1987). For applications of the FCLT and continuous mapping theorem to inference in the presence of a unit root, see (among others) Solo (1984), Phillips (1987), Chan and Wei (1988), and Sims, Stock and Watson (1990).

where $X_\delta^0(\lambda) = X_\delta(\lambda) - \frac{1}{1-2\delta}\int_\delta^{1-\delta} X_\delta(s)ds$, $X_\delta^1(\lambda) = X_\delta(\lambda) - \frac{1}{1-2\delta}\int_\delta^{1-\delta} X_\delta(s+\delta)ds$, and where $X_\delta(\lambda) = W(\lambda) - W(\lambda-\delta)$. Upon scaling by $\frac{1}{T^2}$, the denominator of (17) can be written,

$$\begin{aligned} \frac{1}{T^2}\sum_{t=J}^{T-J} x_t^0(J)^2 &= \frac{1}{T}\sum_{t=J}^{T-J}\left(\frac{1}{\sqrt{T}}x_t^0(J)\right)^2 \\ &= \int_\delta^{1-\delta}\left(\frac{1}{\sqrt{T}}x_{[T\lambda]}^0([T\delta])\right)^2 d\lambda. \end{aligned} \tag{19}$$

The argument $\frac{1}{\sqrt{T}}x_{[T\lambda]}^0([T\delta])$ converges by (18), and the integral in the final expression of (19) is a continuous mapping from $D[0,1]$ to $\Re$. Thus the FCLT and the continuous mapping theorem imply that, as $T \to \infty$ and $\frac{J}{T} \to \delta$,

$$\frac{1}{T^2}\sum_{t=J}^{T-J} x_t^0(J)^2 \Rightarrow \sigma^2 \int_\delta^{1-\delta} X_\delta^0(\lambda)^2 d\lambda.$$

A similar argument applies to the numerator, so

$$\hat{\beta}([T\delta]) \Rightarrow \frac{\int_\delta^{1-\delta} X_\delta^0(\lambda)X_\delta^1(\lambda+\delta)d\lambda}{\int_\delta^{1-\delta} X_\delta^0(\lambda)^2 d\lambda} \equiv \beta_\delta^*$$

which is the desired result (13). Note that β_δ^* has a simple interpretation. The random functions $X_\delta^1(\lambda)$ and $X_\delta^0(\lambda)$ are continuous-time stochastic processes on the unit interval. The expression β_δ^* shows that, in the limit, the estimator of the cross-covariance between the discrete-time processes x_t^1 and x_t^0 with lag J, is a cross-covariance between their corresponding limiting continuous-time processes, with lag δ on the transformed time scale.

Variance Ratios

Let $y_t(J) = x_t(J) - J\hat{\mu}$. Then straightforward algebra and the FCLT imply that, for $\lambda \geq \delta$,

$$\begin{aligned} \frac{1}{\sqrt{T}}y_{[T\lambda]}([T\lambda]) &= \frac{1}{\sqrt{T}}S_{[T\lambda]} - \frac{1}{\sqrt{T}}S_{[T(\lambda-\delta)]} - \left(\frac{J}{T}\right)\frac{1}{\sqrt{T}}S_T \\ &\Rightarrow \sigma Y_\delta(\lambda) \end{aligned} \tag{20}$$

where $Y_\delta(\lambda) = W(\lambda) - W(\lambda-\delta) - \delta W(1)$. Under (1), the denominator of $M_r(J)$ converges in probability to σ^2. By the continuous mapping theorem and (20), it follows that

$$M_r([T\delta]) \Rightarrow \frac{1}{\delta}\int_\delta^1 Y_\delta(\lambda)^2 d\lambda$$

which is the desired result (11).

Hansen-Hodrick Standard Errors

The Hansen-Hodrick estimator $\hat{\omega}(J)^2$ of the variance of $\hat{\beta}(J)$ is:

$$\begin{aligned}
\hat{\omega}(J)^2 &= \frac{\sum_{i=-(J-1)}^{J-1}\left(1-\frac{|i|}{T}\right)\hat{\gamma}_x(i)\hat{\gamma}_u(i)}{\hat{\gamma}_x(0)^2} \qquad (21)\\
\text{where } \hat{\gamma}_x(i) &= \frac{1}{T-2J+1}\sum_{t=J+|i|}^{T-J} x_t^0(J)x_{t-|i|}^0(J)\\
\hat{\gamma}_u(i) &= \frac{1}{T-2J+1}\sum_{t=J+|i|}^{T-J} \hat{u}_{t+J}(J)\hat{u}_{t+J-|i|}(J)\\
\hat{u}_{t+J}(J) &= x_{t+J}^1(J)-\hat{\beta}(J)x_t^0(J).
\end{aligned}$$

The results (13) and (18) imply that, for $\lambda \geq 2\delta$,

$$\frac{1}{\sqrt{T}}\hat{u}_{[T\lambda]}([T\delta]) \Rightarrow \sigma U_\delta(\lambda) \qquad (22)$$

where $U_\delta(\lambda) = X_\delta^1(\lambda) - \beta_\delta^* X_\delta^0(\lambda-\delta)$ for $\lambda \geq 2\delta$. For $\frac{i}{T} \to \rho$, (22) and the continuous mapping theorem therefore imply,

$$\frac{1}{T-2J+1}\hat{\gamma}_x(i) \Rightarrow \sigma^2\Gamma_x(\rho,\delta),\ \Gamma_x(\rho,\delta) = \frac{1}{(1-2\delta)^2}\int_{\delta+|\rho|}^{1-\delta} X_\delta^0(\lambda)X_\delta^0(\lambda-|\rho|)d\lambda \qquad (23)$$

$$\frac{1}{T-2J+1}\hat{\gamma}_u(i) \Rightarrow \sigma^2\Gamma_u(\rho,\delta),\ \Gamma_u(\rho,\delta) = \frac{1}{(1-2\delta)^2}\int_{\delta+|\rho|}^{1-\delta} U_\delta(\lambda+\delta)U_\delta(\lambda+\delta-|\rho|)d\lambda. \qquad (24)$$

It follows from (21), (23) and (24), and the continuous mapping theorem that

$$\frac{1}{T-2J+1}\hat{\omega}([T\delta])^2 \Rightarrow \frac{\int_{-\delta}^{\delta}(1-|\rho|)\Gamma_x(\rho,\delta)\Gamma_u(\rho,\delta)d\rho}{(1-2\delta)\Gamma_x(0,\delta)^2} \equiv \omega_\delta^{*2}.$$

References

Billingsley, P.. 1968, *Convergence of Probability Measure* (John Wiley and Sons, New York).

Campbell, J. and G. Mankiw, 1987, Are output fluctuations transitory?, *Quarterly Journal of Economics*, 102,857–880.

Chan, N. and C. Wei, 1988, Limiting distributions of least squares estimates of unstable autoregressive processes, *Annals of Statistics*, 16,367–401.

Christiano, L. and M. Eichenbaum. Unit roots in GNP: do we know and do we care? forthcoming, Carnegie-Rochester Conference on Public Policy, 1989.

Cochrane, J., 1988, How big is the random walk in GNP?, *Journal of Political Economy*, 96,893–920.

Cochrane, J. and A. Sbordone, 1988, Multivariate estimates of the permanent components of GNP and stock prices, *Journal of Economic Dynamics and Control*, 12,255–296.

Ethier, S. and T. Kurz, 1986, *Markov Processes: Characterization and Convergence* (John Wiley and Sons, New York).

Fama, E. and K. French, 1988, Permanent and temporary components of stock prices, *Journal of Political Economy*, 96,246–273.

Faust, J., 1989, A variance ratio test for mean reversion: statistical properties and implementation, Federal Reserve Bank of Kansas City.

Hall, P. and C. Heyde, 1980, *Martingale Limit Theory and Its Application* (Academic Press, New York).

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

Herrndorff, N.. 1984, A functional central limit theorem for weakly dependent sequences of random variables, *Annals of Probability*, 12,141–153.

Huizinga, J., 1987, An empirical investigation of the long-run behavior of real exchange rates, *Carnegie-Rochester Conference Series on Public Policy*, 27,149–214.

Jegadeesh, N., 1989, Seasonality in stock price mean reversion: evidence from the U.S. and U.K., Anderson Graduate School of Management, University of California at Los Angeles, January.

Kim, M., C. Nelson, and R. Startz, 1988, Mean reversion in stock prices? a reappraisal of the empirical evidence, Technical Report 2795, National Bureau of Economic Research.

Lo, A. and C. Mackinlay, 1988, Stock prices do not follow random walks: evidence from a simple specification test, *Review of Financial Studies*, 1,41–66.

Lo, A. and C. Mackinlay, 1989, The size and power of the variance ratio test in finite samples: a monte carlo investigation, *Journal of Econometrics*, 40,203–238.

Pagan, A. and G. W. Schwert, 1989, Alternative models for conditional stock volatility, Department of Economics, University of Rochester.

Perron, P., 1989, The great crash, the oil price shock, and the unit root hypothesis, *Econometrica*, 67,1361–1401.

Poterba, J. and L. Summers, 1988, Mean reversion in stock prices: evidence and implications, *Journal of Financial Economics*, 22,27–59.

Richardson, M., 1989, Temporary components of stock prices: a skeptic's view, Stanford University, Graduate School of Business, April.

Richardson, M. and T. Smith, 1989, Tests of financial models in the presence of overlapping observations, Stanford University, Graduate School of Business and Fuqua School of Business, Duke University, November.

Sims, C., J. Stock, and M. Watson, 1990, Inference in linear time series models with some unit roots, *Econometrica*, 58,113–144.

Solo, V., 1984, The order of differencing in ARIMA models, *Journal of the American Statistical Association*, 79,916–921.

Stock, J., 1988, A reexamination of Friedman's consumption puzzle, *Journal of Business and Economic Statistics*, 6,401–414.

Stock, J. and M. Watson, 1989, Interpreting the evidence on money-income causality, *Journal of Econometrics*, 40,161–182.

Table 1

Distribution of the variance ratio statistic $M_r([\delta T])$ for $\delta=1/3$: Conventional "Fixed J" asymptotics, "$J/T \to \delta$" asymptotics, and Monte Carlo results

	Mean	2.5%	5%	10%	50%	90%	95%	97.5%
		- - - - - - -	Percentile	- - - - - - - -				
Fixed J asymptotics:								
$N(1,\Omega^*(\delta))$	1.00	-0.31	-0.10	0.15	1.00	1.85	2.10	2.31
$N(b(\delta),b(\delta)^2\Omega^*(\delta))$	0.44	-0.14	-0.04	0.06	0.44	0.82	0.93	1.03
Monte Carlo results:								
$J/T \to \delta$, T = 60	0.46	0.10	0.12	0.15	0.36	0.89	1.10	1.32
120	0.45	0.09	0.11	0.14	0.36	0.88	1.09	1.32
180	0.45	0.09	0.11	0.14	0.36	0.87	1.08	1.32
360	0.44	0.09	0.11	0.14	0.35	0.86	1.09	1.29
720	0.45	0.09	0.11	0.14	0.35	0.88	1.11	1.35
1440	0.44	0.09	0.11	0.14	0.35	0.86	1.09	1.33
2880	0.45	0.09	0.11	0.14	0.35	0.88	1.09	1.33

Notes: The first two rows report conventional (i.e. fixed J, $T \to \infty$) large-sample approximations to the distribution of the variance ratio statistic $M_r(J) = Var(x_t(J))/(JVar(R_t))$, where $x_t(J)$ is the J-period cumulative return. For these two "fixed J" normal approximations, which are taken from Lo and MacKinlay (1988, 1989), $\Omega^*(\delta)=4\delta/3$ and $b(\delta)=(1-\delta)^2$. The remaining rows summarize the results of a sequence of Monte Carlo simulations, in which $M_r(J)$ was evaluated for $J=[\delta T]$, where [x] denotes the greatest integer less than x. Because $\delta=1/3$, for T=60, J=20; for T=120, J=40; etc. The Monte Carlo results were computed using i.i.d. N(0,1) returns with 8000 replications for each T. The "$J/T \to \delta$" asymptotic distribution is the limit (as $T\to\infty$, i.e. reading down the table) of this sequence of Monte Carlo distributions.

Table 2

Distribution of the multi-year correlation Wald statistic F_T: Conventional "Fixed J" asymptotics, "J/T → δ" asymptotics, and Monte Carlo results

	Mean	- - - - - - - Percentile - - - - - - - -						
		2.5%	5%	10%	50%	90%	95%	97.5%
Fixed J asymptotics:								
χ^2_8	8.00	2.18	2.73	3.49	7.34	13.4	15.5	17.5
Monte Carlo results:								
J/T →δ, T = 60	10.37	2.40	3.00	3.83	8.38	18.6	24.4	30.6
120	10.67	2.51	3.16	3.94	8.64	19.2	24.7	31.6
180	10.77	2.54	3.13	3.99	8.77	19.4	24.9	30.8
360	10.96	2.48	3.13	4.00	8.79	19.7	25.6	32.6
720	10.77	2.51	3.17	4.00	8.75	19.6	25.0	30.6
1440	11.12	2.49	3.17	4.08	8.98	20.2	25.7	32.3
2880	10.92	2.54	3.13	3.94	8.85	19.8	25.2	31.6

Notes: Each row reports a different approximation to the distribution of the Wald statistic F_T testing the joint hypothesis that each of 8 multiyear correlations ($\beta(J_i)$, i=1,...,8) are zero. Each J_i-month correlation is estimated by the ordinary least squares estimator $\hat{\beta}(J_i)$ in the regression $x_{t+J_i}(J_i) = \alpha(J_i) + \beta(J_i)x_t(J_i) + \epsilon_{t+J_i}(J_i)$; the F_T statistic is a quadratic form in these 8 individual correlations, weighted by the inverse of their ("fixed J" asymptotic) variance-covariance matrix. The first row reports the conventional (fixed J, T→∞) large-sample χ^2_8 approximation to the distribution of F_T. The remaining rows summarize Monte Carlo simulations in which F_T was evaluated for $(J_1,...,J_8) = [T\delta]$, where δ=(1/60, 1/30, 1/20, 1/15, 1/12, 1/10, 1/7.5, 1/6). The design of the Monte Carlo experiment is described in the note to Table 1. The "J/T → δ" asymptotic distribution is the limit (as T→∞) of this sequence of Monte Carlo distributions.

Table 3

Distribution of the sum R_T of multi-year correlations:
Conventional "Fixed J" asymptotics, "$J/T \to \delta$" asymptotics,
and Monte Carlo results

	Mean	2.5%	5%	10%	50%	90%	95%	97.5%
		- - - - - - -	Percentile	- - - - - - - -				
Fixed J asymptotics:								
$N(0,\tau^2)$	0.00	-3.08	-2.58	-2.01	0.00	2.01	2.58	3.08
Monte Carlo results:								
$J/T \to \delta$, T = 60	-1.00	-3.11	-2.87	-2.55	-1.13	0.72	1.25	1.69
120	-1.02	-3.12	-2.89	-2.59	-1.13	0.70	1.25	1.75
180	-1.01	-3.15	-2.91	-2.60	-1.11	0.71	1.27	1.76
360	-1.05	-3.13	-2.92	-2.61	-1.16	0.67	1.23	1.65
720	-1.01	-3.11	-2.89	-2.58	-1.14	0.76	1.28	1.75
1440	-1.04	-3.17	-2.94	-2.63	-1.15	0.70	1.27	1.79
2880	-1.01	-3.12	-2.91	-2.60	-1.12	0.76	1.30	1.77

Notes: Each row reports a different approximation to the distribution of the sum statistic $R_T = \sum_{i=1}^{8} \hat{\beta}([\delta_i T])$, where δ = (1/60, 1/30, 1/20, 1/15, 1/12, 1/10, 1/7.5, 1/6). This statistic tests the hypothesis that the sum of the 8 multiyear correlations $\beta(J_i)$ is zero, for $J_i = [T\delta_i]$. Each J_i-month correlation is estimated by the regression described in the note to Table 2. The first row reports the conventional (fixed J, $T\to\infty$) large-sample $N(0,\tau^2)$ approximation to the distribution of R_T, where τ^2 is given in Section 4.3 of the text. The remaining rows summarize the Monte Carlo simulations of R_T under the null hypothesis. The design of the Monte Carlo experiment is described in the note to Table 1. The "$J/T \to \delta$" asymptotic distribution is the limit (as $T\to\infty$) of this sequence of Monte Carlo distributions.

Table 4
Multi-year correlations of returns on
size and composite portfolios of New York Stock Exchange stocks:
point estimates and $J/T \rightarrow \delta$ asymptotic p-values, 1926-1985

	- - - - - - - - - - Return Horizon (months)						- - - - - - - - - -		
Portfolio	12	24	36	48	60	72	96	120	Average
Decile:									
1	-.01	-.18	-.30	-.46	-.45	-.21	.13	.27	-.151
	(.546)	(.186)	(.105)	(.034)	(.074)	(.426)	(.882)	(.949)	(.480)
2	-.01	-.16	-.32	-.51*	-.58*	-.42	-.24	-.20	-.305
	(.546)	(.227)	(.085)	(.017)	(.017)	(.145)	(.485)	(.628)	(.141)
3	-.06	-.20	-.34	-.46	-.48	-.35	-.30	-.33	-.315
	(.360)	(.154)	(.067)	(.034)	(.055)	(.222)	(.407)	(.463)	(.125)
4	-.04	-.23	-.37	-.48	-.52	-.39	-.30	-.24	-.321
	(.433)	(.113)	(.045)	(.026)	(.036)	(.176)	(.407)	(.578)	(.115)
5	-.08	-.27	-.37	-.42	-.46	-.32	-.23	-.16	-.289
	(.294)	(.069)	(.045)	(.056)	(.068)	(.264)	(.491)	(.675)	(.170)
6	-.07	-.25	-.38	-.41	-.41	-.26	-.18	-.16	-.265
	(.328)	(.089)	(.040)	(.061)	(.106)	(.350)	(.565)	(.675)	(.217)
7	-.09	-.32	-.42*	-.38	-.35	-.18	-.10	-.12	-.245
	(.263)	(.032)	(.023)	(.069)	(.167)	(.471)	(.667)	(.719)	(.262)
8	-.08	-.28	-.37	-.30	-.26	-.10	-.07	-.13	-.199
	(.294)	(.061)	(.045)	(.167)	(.285)	(.591)	(.701)	(.708)	(.366)
9	-.06	-.26	-.34	-.24	-.14	.05	.08	-.04	-.119
	(.360)	(.079)	(.067)	(.254)	(.483)	(.789)	(.844)	(.794)	(.556)
10	-.08	-.27	-.35	-.20	-.08	.09	.12	-.03	-.100
	(.294)	(.069)	(.059)	(.318)	(.584)	(.829)	(.875)	(.845)	(.598)
Value	-.05	-.24	-.32	-.19	-.07	.09	.10	-.08	-.095
	(.369)	(.101)	(.085)	(.334)	(.600)	(.829)	(.860)	(.757)	(.609)
Equal	-.07	-.26	-.39	-.46	-.47	-.29	-.14	-.06	-.268
	(.328)	(.079)	(.035)	(.034)	(.061)	(.306)	(.618)	(.775)	(.211)

Notes to Table 4:

The first eight columns report the sample correlations $\hat{\beta}(J)$ of the J-month return with its J-month lag, for values of J from 12 months to 120 months; these estimates, computed for the indicated decile and composite portfolios, are taken from Fama and French (1988, Table 2). Fama and French's data were monthly returns for all New York Stock Exchange (NYSE) stocks for 1926-1985 from the Center for Research in Security Prices (CRSP). The final column reports the average of these eight correlations, $(1/8)\sum_{i=1}^{8}\hat{\beta}(J_i)$. The p-values (reported in parentheses) were computed using the asymptotic "$J/T\to\delta$" approximation to the distribution of these statistics. The "$J/T\to\delta$" null distribution was computed by Monte Carlo: $\hat{\beta}(J_1),\ldots,\hat{\beta}(J_8)$ and their average were computed for T=720 and J=12, 24,..., 120 using i.i.d. N(0,1) monthly returns, with 25,000 replications. The p-value reports the fraction of these 25,000 Monte Carlo draws which were less than the observed value of the relevant statistic. Based on the results in Tables 2 and 3, these Monte Carlo distributions provide accurate approximations to the corresponding $J/T \to \delta$ asymptotic distributions. An asterisk indicates rejection at the (two-sided) 5% level.

Table 5

Variance ratio statistics for multi-year returns on composite portfolios of New York Stock Exchange stocks: point estimates and $J/T \to \delta$ asymptotic p-values, 1926-1985

A. $[\mathrm{Var}(x_t(J))/J]/[\mathrm{Var}(x_t(12))/12]$

Portfolio	Return Horizon (months): 1	24	36	48	60	72	84	96	Avg(1-96)	F_T
Value Weighted (real)	.826	.950	.836	.702	.615	.553	.502	.501	.686	7.72
	(.06)	(.40)	(.25)	(.14)	(.11)	(.10)	(.09)	(.12)	(.08)	(.58)
Value Weighted (excess)	.792	1.012	.947	.862	.789	.708	.613	.590	.789	10.50
	(.03)	(.63)	(.49)	(.39)	(.34)	(.34)	(.15)	(.22)	(.25)	(.35)
Equal Weighted (real)	.838	.941	.800	.700	.592	.473	.356*	.308*	.626	10.99
	(.08)	(.37)	(.19)	(.14)	(.09)	(.04)	(.01)	(.01)	(.04)	(.32)
Equal Weighted (excess)	.814	.987	.886	.825	.725	.588	.433	.370	.704	17.52
	(.05)	(.54)	(.35)	(.33)	(.24)	(.13)	(.04)	(.03)	(.11)	(.08)

B. $[\mathrm{Var}(x_t(J))/J]/\mathrm{Var}(R_t)$

Portfolio	Return Horizon (months): 12	24	36	48	60	72	84	96	Avg(12-96)	F_T
Value Weighted (real)	1.211	1.150	1.012	.850	.745	.669	.608	.607	.857	11.20
	(.93)	(.80)	(.62)	(.42)	(.33)	(.28)	(.24)	(.27)	(.44)	(.28)
Value Weighted (excess)	1.262	1.278	1.196	1.089	.996	.894	.774	.745	1.029	16.79
	(.96)	(.92)	(.82)	(.72)	(.64)	(.54)	(.45)	(.42)	(.68)	(.11)
Equal Weighted (real)	1.193	1.123	.955	.835	.706	.564	.425	.368	.771	15.51
	(.92)	(.77)	(.54)	(.44)	(.28)	(.16)	(.07)	(.05)	(.31)	(.13)
Equal Weighted (excess)	1.229	1.213	1.088	1.014	.891	.722	.532	.455	.893	26.49*
	(.95)	(.87)	(.71)	(.64)	(.51)	(.33)	(.16)	(.11)	(.48)	(.02)

Notes to Table 5:

The point estimates in panel A are from Poterba and Summers (1988, Table 2) after eliminating Poterba and Summers' "bias adjustment," in which they divided their point estimates by the mean variance ratio computed from a Monte Carlo simulation under the null hypothesis. Their data were monthly returns on the value-weighted and equal-weighted NYSE portfolios (CRSP). The variance ratios in panel A follow Poterba and Summers (1988) by normalizing by the 12-month variance. The ratios in panel B are the conventional variance ratio statistics $M_r(J)$, i.e. normalized by the one-month variance. The "Avg" column presents the average of the individual variance ratio statistics in that row. The final column reports the F-statistic testing the hypothesis that all the variance ratios in that row equal one. The p-values (in parentheses) were computed using a Monte Carlo simulation to approximate the large-sample "$J/T\to\delta$" distribution under the null hypothesis. The Monte Carlo simulation used N(0,1) monthly returns, T=720, and 10,000 draws. An asterisk indicates rejection of the null at the (two-sided) 5% level based on the $J/T\to\delta$ asymptotic distribution.