# Tests of Financial Models in the Presence of Overlapping Observations

Matthew Richardson, Tom Smith

*The Review of Financial Studies*, Volume 4, Issue 2 (1991), 227-254.

Stable URL:
http://links.jstor.org/sici?sici=0893-9454%281991%294%3A2%3C227%3ATOFMIT%3E2.0.CO%3B2-E

# Tests of Financial Models in the Presence of Overlapping Observations

**Matthew Richardson**
University of Pennsylvania

**Tom Smith**
Duke University

***A general approach to testing serial dependence restrictions implied from financial models is developed. In particular, we discuss joint serial dependence restrictions imposed by random walk, market microstructure, and rational expectations models recently examined in the literature. This approach incorporates more information from the data by explicitly modeling dependencies induced by the use of overlapping observations. Because the estimation problem is sufficiently simple in this framework, the test statistics have simple representations in terms of only a few unknown parameters. As a result, relatively good size properties are attained in small samples. In addition, the benefit to overlapping observations and the advantage of examining multiperiod time series are explicitly quantified.***

Many financial asset pricing models impose serial dependence restrictions on time series. For example, the random walk theory of stock prices (RWTSP) implies that stock returns will be serially uncorrelated and the fundamental restriction from rational expectations models that forecast errors will be orthogonal

We would like to thank Michael Brennan, Stephen Brown, Michael Gibbons, David Hsieh, Allan Kleidon, Craig MacKinlay, Ken Singleton, and the referee for many helpful comments and suggestions. Address reprint requests to Tom Smith, Fuqua School of Business, Duke University, Durham, NC 27706.

*The Review of Financial Studies* 1991 Volume 4, number 2, pp. 227–254

to anything in the agent's information set also imposes some form of serial independence. In addition, many models of market microstructure generate testable serial dependence restrictions.

Using a procedure based on Hansen's (1982) generalized method of moments (GMM) procedure, we investigate new tests of serial dependence restrictions and relate them to the existing tests. We are concerned particularly with the use of multiperiod forecast errors to test asset pricing models. For example, Cochrane (1988) has suggested that low frequencies in the data (such as slow mean reversion in forecast errors) cannot be captured by short-run (e.g., single-period) correlation statistics. Traditionally, however, tests investigating multiperiod properties are severely limited by the number of observations. For example, for $T$ observations and a $j$-period sum, the econometrician has at most $T/j$ independent observations.

We propose a statistical procedure that incorporates more information in the data by explicitly modeling the dependencies induced by the use of overlapping observations. This is important because the use of overlapping observations can improve the properties of the test statistics. While other statistical procedures have been developed to account for overlapping data, the approach described here is of special interest since results are derived analytically, hence avoiding many problems associated with traditional sampling estimation techniques.

In particular, these analytical calculations provide a number of advantages over standard statistical procedures that adjust for overlapping observations. First, in many situations, the econometrician would like to employ long lags; and whereas the standard procedures would require estimation of many autocorrelations to take account of the overlap, the analytical calculations reduce the problem to estimation of only a few unknown parameters. For example, we can replace the Hansen–Hodrick (1980) standard errors adjustment for overlapping data in multiperiod regression tests with a very simple form that is independent of the data. This leads to more desirable size properties of the test statistics in small samples. Second, using a procedure developed by Geweke (1981), we can make explicit (i.e., quantifiable) asymptotic power comparisons between different test statistics. Since one of the purposes of asymptotic theory is its possible application in small samples and in developing intuition, this procedure has potential widespread use. For example, we compare two particular tests of recent interest to financial economists—namely, the variance ratio and multiperiod autocorrelation tests for no serial correlation—and find that there are substantial benefits to examining multiperiod properties of the data. Third, the benefit of using overlapping observations is quantified. For example, we show that using overlapping observations leads to a reduction in the variance of the multiperiod serial correlation estimator of about one third.

This article is organized as follows: The general approach to testing asset pricing models is introduced in Section 1. An important application in finance—the random walk theory of stock prices—is discussed in Section 2. The size and power properties of these analytical test statistics are investigated in Section 3. Some additional applications are discussed in Section 4. Section 5 is the conclusion.

## 1. GMM Test Procedure

In this section, a general procedure for testing serial dependence restrictions implied by financial models in the presence of overlapping observations, which is based on Hansen's (1982) GMM approach, is described. In order to test time-series implications from financial models, the econometrician requires sufficient serial dependence restrictions to identify the unknown parameters of the model. For example, consider testing whether a time series $\tilde{\epsilon}_t$ conforms to a particular set of serial dependence restrictions:

$$E[f_t(\tilde{\epsilon}_t, \ldots, \tilde{\epsilon}_{t-j}, \theta)] = 0, \tag{1}$$

where $\theta$ is an $M$-vector of parameters and $f_t$ is an $R$-vector.

With only weak assumptions on the $\tilde{\epsilon}_t$'s, it is possible to employ Hansen's GMM test procedure. Specifically, if the process $\tilde{\epsilon}_t$ is stationary and ergodic, and the second moments of $f_t(\cdot)$ exist and are finite, then with a long series the sample moments of $f_t(\cdot)$ should be close to its population moments; that is,

$$g_T(\theta) \equiv \frac{1}{T} \sum_{t=1}^{T} f_t(\tilde{\epsilon}_t, \ldots, \tilde{\epsilon}_{t-j}, \theta) \overset{T \to \infty}{\longrightarrow} 0. \tag{2}$$

The idea behind the GMM procedure is to choose values for $\theta$, such that the sample moment conditions, $g_T(\theta)$, equal zero. If the number of parameters in $\theta$ (denoted by $M$) is less than the number of restrictions in the vector $f_t(\cdot)$ (denoted by $R$), then the system is overidentified and no values of the parameters will set all the moment conditions equal to zero. It is possible, however, to find values of $\theta$ that set a linear combination of the $R$-vector $g_T(\theta)$ equal to zero:

$$A g_T(\theta) = 0, \tag{3}$$

where $A$ is an $(M \times R)$ matrix and $g_T(\theta)$ is an $(R \times 1)$ vector. Equation (3) suggests a particular class of estimators, namely those that can be derived from linear combinations of the moment conditions.

Note that the moment condition (3) involves overlapping data: for example, at $t$ and $t-1$, $f_t(\cdot)$ and $f_{t-1}(\cdot)$ have $j-1$ $\tilde{\epsilon}_t$'s in common. The problem then is to choose a weighting matrix $A$ that will lead to desirable properties for the estimator $\hat{\theta}$ in the presence of overlapping observations. A desirable property for the estimator $\hat{\theta}$ is that it have minimum variance–covariance matrix. Hansen (1982) shows that the optimal choice of $A$ (denoted by $A^*$) under this criterion is

$$A^* = D_0' S_0^{-1}, \tag{4}$$

where

$$D_0 = E\left[\frac{\partial g_0(\theta)}{\partial \theta}\right], \tag{5}$$

$$S_0 = \sum_{l=-\infty}^{+\infty} E[f_t(\cdot) f_{t-l}(\cdot)']. \tag{6}$$

Of special interest to this paper, under the assumptions stated above, Hansen (1982) proves the following results:

$$\sqrt{T}(\hat{\theta} - \theta) \overset{\text{asy}}{\approx} N(0, [D_0' S_0^{-1} D_0]^{-1}),$$

$$T g_T(\hat{\theta})' S_0^{-1} g_T(\hat{\theta}) \overset{\text{asy}}{\approx} \chi^2_{R-M}.$$

Moreover, the above results are valid even when $D_0$ and $S_0$ are replaced by their consistent estimators, denoted $D_T$ and $S_T$. For example, one popular way to calculate $S_T$ in practice is to employ the two-step procedure of Hansen and Singleton (1982): this yields an estimator $S_T(\bar{\theta})$, where $\bar{\theta}$ is a consistent estimate of $\theta$ and $S_T(\cdot)$ is the sample estimate of the variance–covariance matrix of the moment conditions.[1]

The analysis in this article is concerned with an alternative procedure for estimating $D_0$ and $S_0$ when the errors are serially uncorrelated and conditionally homoskedastic. Specifically, since this problem of estimation is sufficiently simple, the matrices $D_0$ and $S_0$ can be derived analytically. That is, a consistent estimate of $S_0$ can be provided by $S_0(\bar{\theta})$, where $\bar{\theta}$ is a consistent estimate of $\theta$ and $S_0(\cdot)$ is the analytical variance–covariance matrix of the moment conditions under the null hypothesis. The advantage here is that the asymptotic matrices will have simple representations in terms of a small number of unknown parameters. This leads to relatively desirable size properties and helps form intuition regarding the statistics and estimators.

[1] Depending on the model's assumptions, different estimators have been proposed. Some of these include estimators given in Hansen and Singleton (1982), Hansen and Hodrick (1980), Eichenbaum, Hansen, and Singleton (1988), and Newey and West (1987).

**Example: Variance restrictions.** As an illustration of the analytical approach advocated in this article, consider the following variance-ratio test [see Cochrane (1988), Lo and MacKinlay (1988a), Poterba and Summers (1988), and others, for applications of variance ratios]. Specifically, if $\tilde{\epsilon}_t$ is uncorrelated with past $\tilde{\epsilon}_{t-j}$, then the variance of the sum of the $\tilde{\epsilon}_t$'s should equal the sum of the individual variances of $\tilde{\epsilon}_t$. In the GMM representation, consider the following sample moment conditions to illustrate the variance-ratio restrictions:

$$g_T(m_1, m_2) = \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} \tilde{\epsilon}_t - m_1 \\ (\Sigma_{i=0}^{j-1} \tilde{\epsilon}_{t-i} - jm_1)^2 - jm_2 \\ (\Sigma_{i=0}^{k-1} \tilde{\epsilon}_{t-i} - km_1)^2 - km_2 \end{pmatrix} \qquad \forall j < k, \tag{7}$$

where

$$m_i = i\text{th central moment of } \epsilon_t.$$

From Equation (3),

$$Ag_T(m_1, m_2) = 0, \tag{8}$$

where $A$ is a $2 \times 3$ matrix and $g_T$ is a $3 \times 1$ vector. The procedure that has generally been followed for choosing $A$ is the two-step procedure of Hansen and Singleton (1982). Specifically, in the first step, minimize the sum of squares $g_T(\cdot)'g_T(\cdot)$ with respect to the two parameters, $m_1$ and $m_2$. In the second step, use these estimates $\bar{m}_1$ and $\bar{m}_2$ from the first step to construct a consistent estimator for $S_0$; for example, one choice for a consistent estimator is the sample estimator of $S_0$. The sample estimate $S_T(\bar{m}_1, \bar{m}_2)$ is then used for the optimal choice of $A$:

$$A_T^* = D_T(m_1, m_2)' S_T(\bar{m}_1, \bar{m}_2)^{-1}. \tag{9}$$

However, since the GMM test is derived under the null hypothesis, it is possible to replace the estimator $S_T(\bar{m}_1, \bar{m}_2)$ in (9) by its analytical counterpart $S_0$ under the following assumptions:

(i) $\epsilon_t$ is stationary and ergodic;
(ii) $E[\bar{\epsilon}_t^2 \bar{\epsilon}_{t-j} \bar{\epsilon}_{t-k}] = 0$;
(iii) $E[\bar{\epsilon}_t^2 \bar{\epsilon}_{t-j}] = m_2^2$;
(iv) $E[\bar{\epsilon}_t^3 \bar{\epsilon}_{t-j}] = 0 \qquad \forall t, \quad \forall j \neq k \neq 0$;
(v) all relevant moments $m_1 - m_4$ exist and are finite;

where $\bar{\epsilon}_t = \tilde{\epsilon}_t - m_1$, and $m_i = i$th central moment of $\epsilon_t$.

It is shown in Appendix A that under these assumptions $S_0$, the analytical variance–covariance of the moment conditions, is given by

$$S_0 = \begin{pmatrix} m_2 & jm_3 & km_3 \\ jm_3 & j^2m_4 + A(j)m_2^2 & jkm_4 + B(k,j)m_2^2 \\ km_3 & jkm_4 + B(k,j)m_2^2 & k^2m_4 + A(k)m_2^2 \end{pmatrix}, \quad (10)$$

where

$$A(i) = \left(\frac{(i-2)i(4i-1)}{3}\right),$$

$$B(k,j) = \left(\frac{3(2j^2-3j)(k-j+1) + (j-1)j(4j-11)}{3}\right).$$

This analytical derivation allows us to calculate the serial dependency in the moment conditions induced by the use of overlapping data, which is important because estimation of the sample variance–covariance matrix involves using the same observation numerous times in calculating the autocovariances. Therefore, for a given $T$, there can be a substantial dependence in the elements of the sample estimator, $S_T$. With the analytical derivation, this is not the case because the only unknown parameters are $m_2$, $m_3$, and $m_4$.

Given the derivation of the analytical variance–covariance matrix in (10), the optimal GMM estimators, $\hat{m}_1$ and $\hat{m}_2$, can be obtained by solving the system of equations

$$D_0' S_0^{-1} g_T(m_1, m_2) = 0, \quad (11)$$

where

$$D_0 = \begin{pmatrix} -1 & 0 \\ 0 & -j \\ 0 & -k \end{pmatrix}. \quad (12)$$

Specifically, inverting (10) and then substituting this matrix back into equation system (11) results in the following estimators:

$$\hat{m}_1 = \frac{1}{T}\sum_{t=1}^{T} [\tilde{\epsilon}_t], \quad (13)$$

$$\hat{m}_2 = \left\{(2k^2 - jk)\frac{1}{T}\sum_{t=1}^{T}\left[\sum_{i=0}^{j-1}\tilde{\epsilon}_{t-i} - j\hat{m}_1\right]^2 - (j^2-1)\frac{1}{T}\sum_{t=1}^{T}\left[\sum_{i=0}^{k-1}\tilde{\epsilon}_{t-i} - k\hat{m}_1\right]^2\right\} \times [2jk^2 + (1-2j^2)k]^{-1}. \quad (14)$$

For $j = 1$, the estimate of the variance reduces to the well-known sample variance of a population. If the $\tilde{\epsilon}_t$'s were normally distributed, then maximum likelihood procedures would give us the same estimate. Under alternative distributions, this will not necessarily be the case and, therefore, the estimator $\hat{m}_2$ may be inefficient relative to maximum likelihood estimators. Since in general the distribution of $\tilde{\epsilon}_t$ is unknown, maximum likelihood estimation will not, however, be applicable.

Using Hansen's (1982) results given earlier in this section, it is possible to calculate the variance–covariance matrix of the estimators:

$$\mathrm{VAR}(\hat{m}_1, \hat{m}_2) = \begin{pmatrix} m_2 & m_3 \\ m_3 & m_4 + C(j, k) m_2^2 \end{pmatrix}, \tag{15}$$

where

$$C(j, k) = [2(j - 2)(4j - 1)k^2 + (18j^2 - 10j^3 + 4j - 9)k + 2(j - 1)^2(j + 1)^2][3k(2jk - 2j^2 + 1)]^{-1}. \tag{16}$$

The benefits of the analytical derivation are not solely in parameter estimation. Under the null hypothesis, it is possible to test the overidentifying restriction that $\tilde{\epsilon}_t$ is uncorrelated with past values. In this particular case, there are three equations and two parameters ($m_1$ and $m_2$) and hence one overidentifying restriction. Using the optimal estimators given in Equations (13) and(14), it is possible to calculate the value of this $\chi^2$ statistic analytically:

$$Tg_T' S_0^{-1} g_T = 3\left\{ j \sum_{t=1}^{T} \left[ \sum_{i=0}^{k-1} \tilde{\epsilon}_{t-i} - k\hat{m}_1 \right]^2 - k \left[ \sum_{i=0}^{j-1} \tilde{\epsilon}_{t-i} - j\hat{m}_1 \right]^2 \right\}^2 \times [2jk(k - j)(2jk - 2j^2 + 1) T m_2^2]^{-1} \overset{\text{asy}}{\sim} \chi_1^2. \tag{17}$$

In order to relate this result to the previous literature, consider the variance ratio test examined by Cochrane (1988) and Lo and MacKinlay (1988a), among others. The variance ratio test is a special case of the restrictions in (7) with $j = 1$ and $k$ allowed to vary, and $m_2$ in (17) set equal to $\hat{m}_2$. To see this, in Section 1.1 of their article, Lo and MacKinlay show that, under the assumption that the $\tilde{\epsilon}_t$ are i.i.d. normal-random variables,

$$\sqrt{T}\left[\frac{(1/k)\ \Sigma_{t=1}^{T}\ [\Sigma_{i=0}^{k-1}\ \tilde{\epsilon}_{t-i} - k\hat{m}_1]^2}{\Sigma_{t=1}^{T}\ [\epsilon_t - \hat{m}_1]^2} - 1\right]$$

$$\overset{\text{asy}}{\approx} N\left(0, \frac{2(2k-1)(k-1)}{3k}\right). \tag{18}$$

Dividing (18) by its variance and squaring the resulting statistic gives the desired result.

With applications in which the GMM procedure is desirable, it has been standard to rely on numerical minimization procedures to estimate the parameters and form the overidentifying restrictions test. The problems with numerical optimization, such as failure to converge, are well-known. The analytical procedure overcomes these difficulties by providing closed-form solutions for the parameter estimators, their standard errors, and the $\chi^2$ statistic.

In addition, for the analytical GMM procedure, no parametric assumption like normality is necessary. Along with stationarity and ergodicity of $\tilde{\epsilon}_t$, all that we require for the derivation of (17) are assumptions (ii)–(v). Lo and MacKinlay, however, derive a more general, heteroskedasticity-consistent variance test. Their test relies on the calculation of standard errors of lagged serial correlation coefficients using White's (1980) heteroskedastic-consistent covariance matrix estimator. In contrast to the above derivation and their derivation of (18), however, this heteroskedasticity-consistent statistic involves the additional estimation of numerous autocorrelation parameters.[2]

## 2. Application: The Random Walk Theory of Stock Prices

The RWTSP has long been of interest to financial economists. Recently, Lo and MacKinlay (1988a) have presented evidence that, in the short term, stock prices do not follow a random walk. In addition, Fama and French (1988), Poterba and Summers (1988), and others have presented evidence that it is possible to forecast long-term movements in stock prices from past returns. The test strategy of these authors has been to calculate correlation statistics over different return

[2] Note that if the heteroskedasticity is specified, then this can be incorporated directly in the calculation of $S_0$. Otherwise, the GMM test can be adjusted in a similarly more general way. Specifically, an analogous heteroskedasticity-consistent estimator can be developed using the variance–covariance matrix estimators provided by, among others, Hansen and Singleton (1982); see, for example, Equation (9). These estimators do not require assumptions (ii)–(iv). Although the assumptions are generally weaker, the problem with these procedures is that for large $j$ they require the estimation of numerous autocovariances. This can lead to poor size properties in small samples.

horizons and to treat any of the statistics that are significant as evidence against the RWTSP. This ignores the joint implications of the RWTSP and the experimental design.

On the other hand, the approach described in Section 1 allows a joint test of the hypotheses:

$$\operatorname{cov}\left(\sum_{i=1}^{j} \tilde{R}_{t+i}, \sum_{i=1}^{j} \tilde{R}_{t+i-j}\right) = 0 \qquad \forall j \geq 1, \tag{19}$$

where $\tilde{R}_t$ is the return on a stock in period $t$.

Consider Fama and French's (1988) ordinary least-squares (OLS) regression test corresponding to condition (19). Let $\tilde{\epsilon}_t = \tilde{R}_t$. In the GMM framework, the corresponding moment conditions from OLS are the normal equations:

$$g_T(\alpha(j), \beta(j))$$

$$= \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} \Sigma_{i=1}^{j} \tilde{\epsilon}_{t+i} - \alpha(j) - \beta(j)(\Sigma_{i=1}^{j} \tilde{\epsilon}_{t-j+i}) \\ [\Sigma_{i=1}^{j} \tilde{\epsilon}_{t+i} - \alpha(j) - \beta(j)(\Sigma_{i=1}^{j} \tilde{\epsilon}_{t-j+i})][\Sigma_{i=1}^{j} \tilde{\epsilon}_{t-j+i}] \\ \vdots \\ \Sigma_{i=1}^{k} \tilde{\epsilon}_{t+i} - \alpha(k) - \beta(k)(\Sigma_{i=1}^{k} \tilde{\epsilon}_{t-k+i}) \\ [\Sigma_{i=1}^{k} \tilde{\epsilon}_{t+i} - \alpha(k) - \beta(k)(\Sigma_{i=1}^{k} \tilde{\epsilon}_{t-k+i})][\Sigma_{i=1}^{k} \tilde{\epsilon}_{t-k+i}] \end{pmatrix}. \tag{20}$$

The OLS estimators $\hat{\beta}$ have asymptotic normal distributions. Fama and French (1988) calculate the standard errors of the individual $\hat{\beta}_j$ estimators using the method of Hansen and Hodrick (1980). This introduces sampling error because of the additional nuisance parameters involved in estimating cross-equation correlations. Under the Hansen and Hodrick (1980) assumptions, the variance–covariance matrix of $V(\hat{\beta})$ has a representation independent of the data. The typical elements of $V(\hat{\beta})$ are given by (see Appendix B for derivation)[3]

$$V(\hat{\beta}) \equiv \operatorname{var}\begin{pmatrix}\hat{\beta}(j) \\ \hat{\beta}(k)\end{pmatrix} = \begin{pmatrix} \dfrac{2j^2+1}{3j} & \dfrac{s(j,k)+j^2}{jk} \\ \dfrac{s(j,k)+j^2}{jk} & \dfrac{2k^2+1}{3k} \end{pmatrix}, \tag{21}$$

[3] Note that these multiple-restriction results carry through to other settings. For example, consider the variance ratio statistic given in Equation (18) [denoted by $\hat{V}(j)$]. Using similar techniques, it is possible to show that for any lags $j$ and $k$ the variance–covariance matrix of $\hat{V}(j)$ and $\hat{V}(k)$ is

$$\begin{pmatrix} \dfrac{2(2j-1)(j-1)}{3j} & \dfrac{2(3k-j-1)(j-1)}{3k} \\ \dfrac{2(3k-j-1)(j-1)}{3k} & \dfrac{2(2k-1)(k-1)}{3k} \end{pmatrix}.$$

where

$$s(j, k) = 2 \sum_{l=1}^{j-1} [(j - l) \min(j, k - l)]. \tag{22}$$

Three important observations are in order. First, the asymptotic standard error given in (21) is independent of the data. Since the goal of the Hansen–Hodrick (1980) method under the null hypothesis is to estimate consistently the asymptotic standard errors given in (21), this formula greatly simplifies the estimation problem. Note also that when $j = 1$, we obtain the familiar result that the standard error of the autocorrelation coefficient is $1/\sqrt{T}$. Therefore, one way to view the result from (21) is that it extends the literature on test statistics for serial correlation of a random variable, $\tilde{\epsilon}_t$, to sums of that random variable, $\Sigma_{i=1}^{j} \tilde{\epsilon}_{t+1}$. To the extent that $\Sigma_{i=1}^{j} \tilde{\epsilon}_{t+i}$ can provide more information concerning the long-run properties of $\tilde{\epsilon}_t$, this extension has useful applications.

Second, the OLS multiperiod autocorrelation estimators are, in general, highly correlated. As $j$ approaches $k$, the correlation between $\hat{\beta}(j)$ and $\hat{\beta}(k)$ approaches unity. Note that even if the time horizons are far apart, these estimators may still be highly correlated. For example, even if $k > 2j$, the correlation between $\hat{\beta}(k)$ and $\hat{\beta}(j)$ can still be over 50 percent.

Third, it is straightforward to incorporate this joint dependency in designing test statistics. One such application is in testing the joint hypothesis that $\beta(j) = \beta(k) = \cdots = 0$. Let $\hat{\beta}$ equal the $M$-vector of different multiperiod correlation estimators. One popular statistic is the Wald statistic:

$$T\hat{\beta}'[V(\hat{\beta})]^{-1}\hat{\beta} \overset{\text{asy}}{\approx} \chi^2_M. \tag{23}$$

More generally, let $D$ be any $N \times M$ matrix. Then

$$T[D\hat{\beta}]'[DV(\hat{\beta})D']^{-1}[D\hat{\beta}] \overset{\text{asy}}{\approx} \chi^2_N. \tag{24}$$

The difference between a joint test of the restrictions (19) and a series of individual tests as proposed by the earlier authors is similar to the difference between reporting individual $t$-statistics versus an overall $F$-statistic [see, e.g., Gibbons, Ross, and Shanken (1989) for an application to multivariate tests of mean-variance efficiency]. While it is nontrivial to test formally restrictions of this type using standard techniques, the derivations follow quite naturally in our procedure. This distinction between a joint test and several individual tests is of practical importance since Richardson (1989) shows that Fama and French's (1988) rejection of the random walk on the basis of individual correlation statistics is not necessarily justified in a joint system.

**Multivariate series.** Many applications in financial economics deal with multivariate time series. For example, the results dealing with serial correlation in stock returns have been reported for many different groupings of assets and for a variety of countries. Using the framework in this article, it is possible to incorporate the joint nature of these different series. Consider testing for serial correlation in $\tilde{\epsilon}_{1t}, \tilde{\epsilon}_{2t}, \ldots, \tilde{\epsilon}_{Nt}$. Assume that the $\tilde{\epsilon}_{nt}$'s are contemporaneously correlated but otherwise satisfy assumptions similar to those of Hansen and Hodrick (1980):

(a) $E[(\tilde{\epsilon}_{nt} - \mu_{\tilde{\epsilon}_n})(\tilde{\epsilon}_{mt} - \mu_{\tilde{\epsilon}_m})] = \sigma_{\tilde{\epsilon}_n, \tilde{\epsilon}_m}$;
(b) $E[(\tilde{\epsilon}_{nt} - \mu_{\tilde{\epsilon}_n})(\tilde{\epsilon}_{mt-j} - \mu_{\tilde{\epsilon}_m})] = 0$;
(c) $E[(\tilde{\epsilon}_{nt} - \mu_{\tilde{\epsilon}_n})^2(\tilde{\epsilon}_{mt-j} - \mu_{\tilde{\epsilon}_m})(\tilde{\epsilon}_{mt-k} - \mu_{\tilde{\epsilon}_m})] = 0, \quad \forall j \neq k \neq 0, \forall n, m$;
(d) $E[(\tilde{\epsilon}_{nt} - \mu_{\tilde{\epsilon}_n})^2(\tilde{\epsilon}_{mt-j} - \mu_{\tilde{\epsilon}_m})^2] = \sigma^2_{\tilde{\epsilon}_n}\sigma^2_{\epsilon_m}$.

Stacking the $2 \times N$ normal equations from OLS for the $N \epsilon_{nt}$'s, we get the individual OLS serial correlation estimators of each series $\tilde{\epsilon}_{nt}$. Using the procedure in Appendixes A and B and Hansen's (1982) distributional results, the asymptotic distribution of the estimators is

$$\sqrt{T}\begin{pmatrix} \hat{\beta}_1(j) \\ \vdots \\ \hat{\beta}_N(j) \end{pmatrix} \overset{\text{asy}}{\approx} \sqrt{\frac{2j^2+1}{3j}} \times N\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{12}^2 & \cdots & \rho_{1N}^2 \\ \rho_{12}^2 & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1N}^2 & \cdots & \cdots & 1 \end{pmatrix}\right), \tag{25}$$

where $\rho_{ij}$ is the correlation coefficient between $\tilde{\epsilon}_{it}$ and $\tilde{\epsilon}_{jt}$. In practice, all we need do is replace the correlation coefficients $\rho_{ij}$ with consistent estimates and form test statistics analogous to those derived earlier.

## 3. Properties of the Statistics

In this section, some of the properties of the statistics discussed in Sections 1 and 2 are investigated: notably, the variance ratio and multiperiod autocorrelation regression tests.

### 3.1 Size

In this subsection, we perform Monte Carlo simulations relating to the size of the variance ratio and regression tests. For each of the

5000 replications in each simulation, 720 observations were independently drawn from a normally distributed random variable.[4]

**Variance ratio test.** We compare two particular ways of estimating the weighting matrix for the variance restrictions given in (7). The first estimator is the analytically derived estimator in Equation (10), denoted $S_0(\bar{\theta})$, where the vector $\bar{\theta}$ contains consistent estimates of the moments $m_1, \ldots, m_4$. The other estimator is the one suggested by Hansen and Singleton (1982). It is the sample variance–covariance matrix of the moment conditions associated with overlapping data given in (9), denoted $S_T(\bar{\theta})$.[5] For the moment conditions given by Equation (7), let $j = 1$ and $k = 40$.[6]

Average estimates for the elements of $S_0(\bar{\theta})$ and $S_T(\bar{\theta})$ are reported in Table 1. $S_0(\bar{\theta})$ is considerably closer to the true asymptotic variance–covariance matrix of the moment conditions, $S_0$, than $S_T(\bar{\theta})$ [note further that in about 5 percent of the simulations $S_T(\bar{\theta})$ was not positive definite]. For example, consider the (3, 3) element: $S_T(\bar{\theta})$ is only 85 percent of its true analytical counterpart $S_0$, while, in contrast, $S_0(\theta)$ estimates precisely. These results take on increasing importance for the size of the test as shown in Table 2. The overidentifying statistic $Tg_T'S^{-1}g_T$ using the Hansen–Singleton $S_T(\bar{\theta})$ estimator behaves more like a $\chi^2$ with two degrees of freedom than a $\chi^2$ one. For example, the 5 percent value is 5.11 versus a $\chi_1^2$'s table value of 3.84; its average (mean, variance) estimates are (1.34, 2.65) versus a $\chi_1^2$'s (1.0, 2.0);

[4] The parameters were chosen to match those of the equal-weighted index of stock returns ($m_1 = 0.01001$, $m_2 = 0.005685$). In order to avoid the normality assumption, we also calculated a bootstrap distribution of monthly returns on the index (1926–1985). By construction, the random variables will be i.i.d. with a discrete distribution that places equal probability on each of the returns. The idea is that the unconditional moments of this distribution will match the sample moments of the actual returns. The simulations were robust to the distributional changes.

[5] One problem with the Hansen and Singleton (1982) estimator is that it is not guaranteed to be positive definite. There are other variance–covariance matrix estimators, however, that are assured to be positive definite; for example, see Eichenbaum, Hansen, and Singleton (1988) and Newey and West (1987). These procedures are not particularly suited for the long lag structure we have here. For example, the Newey and West (1987) estimator fares relatively poorly because it imposes specific declining weights on the autocorrelations of the moment conditions—in this example, we can calculate the weights explicitly (see the example in Section 1) and, unlike the Hansen and Singleton (1982) method, these weights do not correspond to those of Newey and West (1987) for a given $j$ and $T$. The Newey and West (1987) procedure is a general method for taking account of a variety of econometric problems; its appropriateness for a particular situation is, therefore, at the discretion of the econometrician. Ignoring the cases that were not positive definite, $S_T(\bar{\theta})$ provides the "best" performance, so it is, therefore, the only one reported.

[6] These lags are chosen to coincide with the number of lags used elsewhere in the literature. For example, in most applications, the lags have varied between 2 and 120; hence, 40 lags represents a conservative choice. When $j = 1$, the various estimators yield similar estimates for the (1, 1), (1, 2), and (2, 2) elements of $S_0$, although $S_T(\bar{\theta})$ does pick up some spurious correlation. Hence, only the (2, 2), (2, 3), and (3, 3) elements are reported.

**Table 1**
**Estimates of the covariance matrix of moment conditions**

| Method | (1, 3) element | (2, 3) element | (3, 3) element |
|---|---|---|---|
| True $S_0$ | 0.000000 | 0.002568 | 2.758924 |
| Analytical $S_a(\bar{\theta})$ | −0.000020<br>(0.000002) | 0.002571<br>(0.000005) | 2.755560<br>(0.004090) |
| Hansen–Singleton $S_T(\bar{\theta})$ | −0.000395<br>(0.000706) | 0.001931<br>(0.000054) | 2.317565<br>(0.030715) |

Two ways of estimating the variance–covariance matrix of the GMM variance moment restrictions for tests of serial correlation are compared; namely, an analytical one-step procedure versus the more standard Hansen and Singleton (1982) two-step procedure. This estimator provides the weights for the GMM statistic given in Table 2. The standard error of the estimates is given in parentheses.

$$g_T(m_1, m_2) = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} \tilde{\epsilon}_t - m_1 \\ (\Sigma_{i=0}^{j-1}\tilde{\epsilon}_{t-i} - jm_1)^2 - jm_2 \\ (\Sigma_{i=0}^{k-1}\tilde{\epsilon}_{t-i} - km_1)^2 - km_2 \end{pmatrix}$$

$$H_0: \tilde{\epsilon}_t \sim \text{i.i.d.}, N(0.01001, 0.005685)$$

and its average $p$-value is .56. In contrast, the analytical-based overlapping statistic's behavior is closer to its $\chi_1^2$ asymptotic distribution.

Apparently, the reason for the success of the analytical estimator $S_0(\bar{\theta})$ is that it reduces sampling error. By calculating the variance–covariance matrix under the null hypothesis, the problem is reduced to estimating only $m_1, \ldots, m_4$. In contrast, calculating the sample variance–covariance matrix $S_T(\bar{\theta})$ involves estimating 240 autocovariances at various lags as well as $m_1, \ldots, m_4$. Intuitively, by the time the econometrician has calculated his 200th moment, the same data has been used numerous times in estimation. This induces sampling error into the estimation process.

**Table 2**
**Size of test**

Test statistic: $Tg_T'(\hat{m}_1, \hat{m}_2)S^{-1}g_T(\hat{m}_1, \hat{m}_2)$
$H_0$: $\tilde{\epsilon}_t \sim$ i.i.d., $N$ (0.01001, 0.005685)

| Method | Mean | 5% level | Average $p$-value |
|---|---|---|---|
| $\chi_1^2$ | 1.00 | 3.84 | |
| Analytical | 0.96<br>[1.77] | 3.28 | 0.51<br>(0.004) |
| Hansen–Singleton | 1.34<br>[2.65] | 5.11 | 0.56<br>(0.004) |

Details of the small sample distribution of the analytical and Hansen–Singleton statistics are reported. The asymptotic distribution using either estimator is $\chi^2$ with one degree of freedom. The statistics for the different methods were computed from 5000 replications of independently drawn normally distributed data. Standard errors are in parentheses; variances are in square brackets.

**Table 3**
**Standard errors of $\hat{\beta}$**

| Method | 12 lags | 60 lags | 120 lags |
|---|---|---|---|
| Analytical $SE_{\hat{\beta}}$ | 0.107 | 0.258 | 0.408 |
| Hansen–Hodrick $SE_{\hat{\beta}}$ | 0.105<br>(0.0001) | 0.231<br>(0.0005) | 0.294<br>(0.0013) |
| Hansen–Hodrick $SE_{\hat{\beta}}$ ($\hat{\beta} < -0.25$) | 0.092<br>(0.0001) | 0.190<br>(0.0008) | 0.257<br>(0.0024) |
| Hansen–Hodrick $SE_{\hat{\beta}}$ ($\hat{\beta} > 0.25$) | 0.109<br>(0.0001) | 0.252<br>(0.0009) | 0.323<br>(0.0031) |

Two ways of estimating the standard errors of the OLS serial correlation estimators for multiperiod time series are compared; namely, an analytical one-step procedure versus the more commonly used Hansen and Hodrick (1980) method. The statistics for the different methods were computed from 5000 replications of independently drawn normally distributed data. The standard error of the estimates is given in parentheses.

$$g_T(\alpha, \beta) = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} \sum_{i=1}^{j} \tilde{\epsilon}_{t+i} - \alpha - \beta\left(\sum_{i=1}^{j} \tilde{\epsilon}_{t-j+i}\right) \\ \left[\sum_{i=1}^{j} \tilde{\epsilon}_{t+i} - \alpha - \beta\left(\sum_{i=1}^{j} \tilde{\epsilon}_{t-j+i}\right)\right]\left[\sum_{i=1}^{j} \tilde{\epsilon}_{t-j+i}\right] \end{pmatrix}$$

$$H_0\colon \tilde{\epsilon}_t \sim \text{i.i.d.}, N(0.01001, 0.005685)$$

**Regression test**. With respect to the regression test of Section 2, we attain similar conclusions. Simulation evidence from 5000 repetitions is reported in Table 3. Specifically, the average standard error of $\hat{\beta}$ is given for lags 12, 60, and 120. The Hansen–Hodrick method understates the standard error (it is trying to estimate) for all lags chosen. Not surprisingly, this underestimation is more severe for longer lags as more moments need to be estimated (i.e., two for every additional lag). For example, the understatement is 2 percent, 11 percent, and 28 percent for lags 12, 60, and 120, respectively.[7] The fact that nuisance parameters affect the Hansen and Hodrick (1980) standard errors is best illustrated by comparing these standard errors for different values of $\hat{\beta}$ under the null hypothesis. When we stratify the sample conditional on extreme values of $\hat{\beta}$, Table 3 shows that the understatement is even more pronounced for low values of $\hat{\beta}$.

Table 4 reports the small sample distribution of the two statistics, $\hat{\beta}(j)/\sqrt{(2j^2 + 1)/3j}$ and $\hat{\beta}(j)/SE_{HH}$, for lags 12, 60, and 120. Note that the $\hat{\beta}(j)$'s have been adjusted for small sample "bias" using the method described in Fama and French (1988). Relative to normal table values, the analytical statistic of Equation (21), $\hat{\beta}(j)/\sqrt{(2j^2 + 1)/3j}$, is biased toward acceptance. For example, with 60 lags at the one-sided 10 percent level, the statistic represents the 14

[7] Part of this bias can be attributed to the well-known negative bias present in sample autocorrelations (of the $\tilde{\epsilon}_t$'s and OLS residuals in this case). Specifically, Kendall (1954) shows this bias is negative and that it increases with the lag. Huizinga (1983) extends this discussion to overlapping data.

**Table 4**
**Size of tests**

| Method (lags) | Mean | 0.01 | 0.05 | 0.10 | 0.25 | 0.50 | 0.75 | 0.90 | 0.95 | 0.99 |
|---|---|---|---|---|---|---|---|---|---|---|
| $N(0, 1)$ | 0.00 | −2.32 | −1.64 | −1.28 | −0.67 | 0.00 | 0.67 | 1.28 | 1.64 | 2.32 |
| Anal. (12) | −0.02 | −2.30 | −1.66 | −1.32 | −0.69 | −0.04 | 0.65 | 1.24 | 1.61 | 2.25 |
| Anal. (60) | 0.03 | −1.89 | −1.34 | −1.10 | −0.59 | 0.00 | 0.61 | 1.19 | 1.52 | 2.15 |
| Anal. (120) | 0.00 | −1.64 | −1.13 | −0.92 | −0.52 | −0.06 | 0.46 | 1.01 | 1.38 | 2.04 |
| H–H (12) | −0.05 | −2.57 | −1.78 | −1.41 | −0.71 | −0.04 | 0.65 | 1.22 | 1.59 | 2.23 |
| H–H (60) | −0.06 | −2.79 | −1.92 | −1.46 | −0.71 | 0.00 | 0.63 | 1.20 | 1.60 | 2.32 |
| H–H (120) | −0.08 | −3.42 | −2.00 | −1.53 | −0.80 | −0.08 | 0.59 | 1.37 | 1.95 | 3.38 |

The small sample distribution of the analytical and Hansen–Hodrick statistics is compared. The serial correlation coefficients for periods of length 12, 60, and 120 are then divided by their corresponding standard error estimator in order to provide an empirical distribution of the "*t*-statistic." Note that the OLS coefficients are adjusted for small sample "bias" using the method described in the appendix of Fama and French (1988). The asymptotic distribution using either estimator is standard normal. The statistics for the different methods were computed from 5000 replications of independently drawn normally distributed data.

$$\text{Test statistic: } \frac{\hat{\beta}}{\mathrm{SE}(\hat{\beta})}$$

$$\text{Anal. } \mathrm{SE}(\hat{\beta}_j) = \sqrt{\frac{2j^2 + 1}{3j(T - 2j)}}$$

$$\text{H–H } \mathrm{SE}(\hat{\beta}_j) = \mathrm{SE}_{\mathrm{HH}}(\hat{\beta}_j) \text{ Hansen–Hodrick}$$

percent table value. On the other hand, the statistic based on Hansen and Hodrick (1980) is biased toward rejection—at the 10 percent level, this statistic represents only the 7 percent table value. Which particular statistic is desirable depends upon the econometrician's own preference on acceptance/rejection biases. Nevertheless, the analytical statistic is not only computationally easier, but also has sampling variation from only one source—namely, the $\hat{\beta}(j)$ estimator. In contrast, $\hat{\beta}(j)/\mathrm{SE}_{\mathrm{HH}}$ also includes variation in the standard error estimator, $\mathrm{SE}_{\mathrm{HH}}$. This is because the Hansen–Hodrick method requires estimation of $2j$ autocorrelations, which introduces sampling error into their procedure.

## 3.2 Robust power calculations

Under fairly weak assumptions, the estimators and statistics in Sections 1 and 2 have well-known asymptotic distributions. Small sample considerations aside, the sizes of these tests are, therefore, straightforward to calculate. Under similar types of assumptions, what can we say about the power of these tests? Unfortunately, most of the specific power results available are based on Monte Carlo simulations—that is, they are specific to a particular, parameterized distribution. Monte Carlo analysis provides a useful gauge for evaluating aymptotic approximations. It does not on its own (i.e., without theoretical justification) provide an alternative to asymptotic theory.

Using a procedure developed by Bahadur (1960) and Geweke (1981), we perform large sample comparisons between tests using different multiperiods and different test designs. In particular, we compare the approximate slopes ("asymptotic power") of the different test statistics given in Sections 1 and 2. Intuitively, the approximate slope of the test is a measure of the rate at which the null hypothesis becomes more and more incredible as sample size increases. Geweke (1981) shows that if the test statistics have a limiting $\chi^2$ distribution under the null hypothesis (as they do in Sections 1 and 2), then asymptotically the approximate slope equals the probability limit (plim) of the statistic under the alternative hypothesis deflated by sample size. In addition, the ratio of these slopes between any two statistics, denoted $A$ and $B$, will be $T^*_B/T^*_A$, where $T^*$ is the minimum number of observations needed to reject the alternative for given power. Furthermore, in contrast to Monte Carlo analysis, the approximate slope results will be robust to parameterized distributional forms. For example, if the alternative is an AR(1) with uncorrelated innovations, then the approximate slope depends only on $j$, the length of the holding period, and $\rho$, the autoregressive parameter.

With respect to the GMM approach discussed in Sections 1 and 2, it is possible to calculate the approximate slopes of the regression and variance ratio tests analytically. This is especially convenient because we can quantify the benefit of exploring multiperiod properties of $\tilde{\epsilon}_t$. For example, consider the following alternative model to the one given in Sections 1 and 2:[8]

$$\tilde{\epsilon}_t = \tilde{\eta}_t + (\rho - 1) \sum_{i=1}^{\infty} \rho^{i-1} \tilde{\eta}_{t-i},$$

$$|\rho| < 1, \qquad E[\tilde{\eta}_t] = 0, \qquad E[\tilde{\eta}_t \tilde{\eta}_{t-i}] = 0. \tag{26}$$

In Tables 5 and 6, the optimal multiperiods are provided for the variance ratio test [denoted $(j^*,k^*)$] and regression test (denoted $j^*$) given in Sections 1 and 2. The periods are optimal in terms of maximizing the approximate slope for different values of $\rho$.[9] The optimal

[8] To see this alternative in light of previous studies, suppose $\tilde{\epsilon}_t$ were stock returns through time. This alternative is then equivalent to a first-order autoregressive process [AR(1)] on prices. We also considered other alternative processes; namely, an AR(1) in returns and a mixture of an AR(1) and a random walk in prices [see, e.g., Fama and French (1988) and Poterba and Summers (1988)]. The approximate slope results coincide with simulations given in Lo and MacKinlay (1988b) and Poterba and Summers (1988) for these alternatives, respectively. For purposes of space, we provide only the AR(1) in prices since it provides the "least successful" comparison. It, therefore, is a more conservative choice. The others are available upon request.

[9] Using Equations (17) and (21), it is possible to calculate a formula for the probability limit of the sum of squares (i.e., the approximate slope) for the variance ratio and regression test, respectively:

**Table 5**
**Approximate slope calculations for the variance-ratio test**

| $\rho$ | $c^{(j^*,k^*)}$ | $k^*$ | $j^*$ | $c^{(j^*,2)}/c^{(j^*,k^*)}$ | $c^{(j^*,k^*+50)}/c^{(j^*,k^*)}$ |
|---|---|---|---|---|---|
| .99 | 0.00122 | 214 | 1 | 0.021 | 0.986 |
| .98 | 0.00245 | 107 | 1 | 0.041 | 0.952 |
| .97 | 0.00369 | 71 | 1 | 0.061 | 0.910 |
| .96 | 0.00494 | 53 | 1 | 0.081 | 0.866 |
| .95 | 0.00621 | 42 | 1 | 0.101 | 0.822 |
| .94 | 0.00749 | 35 | 1 | 0.120 | 0.777 |
| .93 | 0.00878 | 30 | 1 | 0.140 | 0.735 |
| .92 | 0.01009 | 26 | 1 | 0.159 | 0.697 |
| .91 | 0.01141 | 23 | 1 | 0.178 | 0.661 |
| .90 | 0.01274 | 21 | 1 | 0.196 | 0.625 |
| .89 | 0.01409 | 19 | 1 | 0.215 | 0.594 |
| .88 | 0.01545 | 17 | 1 | 0.233 | 0.568 |
| .87 | 0.01683 | 16 | 1 | 0.251 | 0.539 |
| .86 | 0.01822 | 15 | 1 | 0.269 | 0.507 |
| .85 | 0.01963 | 14 | 1 | 0.287 | 0.491 |
| .84 | 0.02170 | 13 | 1 | 0.304 | 0.470 |
| .83 | 0.02249 | 12 | 1 | 0.321 | 0.452 |
| .82 | 0.02393 | 11 | 1 | 0.338 | 0.435 |
| .81 | 0.02541 | 11 | 1 | 0.355 | 0.414 |
| .80 | 0.02689 | 10 | 1 | 0.372 | 0.401 |
| .79 | 0.02839 | 10 | 1 | 0.388 | 0.383 |
| .78 | 0.02991 | 9 | 1 | 0.405 | 0.371 |
| .77 | 0.03145 | 9 | 1 | 0.421 | 0.356 |
| .76 | 0.03297 | 8 | 1 | 0.437 | 0.347 |
| .75 | 0.03458 | 8 | 1 | 0.452 | 0.333 |

Approximate slope comparisons of variance ratios using different period lengths are provided. The approximate slope $c^{(j,k)}$ is a measure of the rate at which the null hypothesis (i.e., the $\tilde{\epsilon}_t$'s are uncorrelated) becomes more and more incredible as sample size increases under a given alternative [i.e., if $\tilde{\epsilon}_t$ represents growth, then the alternative is an AR(1) in levels]. The period lengths denoted by an "*" are the optimal ones in that they maximize the approximate slope (i.e., relative asymptotic power). The final two columns show the sensitivity of the approximate slope calculations to departures from the optimal lag length, $k^*$.

$$\text{Alternative: } \tilde{\epsilon}_t = \tilde{\eta}_t + (\rho - 1) \sum_{i=1}^{\infty} \rho^{i-1} \tilde{\eta}_{t-i}; \qquad |\rho| < 1, \tilde{\eta}_t \text{ i.i.d.}$$

$$c^{(j,k)} \equiv \text{plim}(g_T' S_0^{-1} g_T) = \frac{3[j(1-\rho^k) - k(1-\rho^j)]^2}{2jk(k-j)(2jk - 2j^2 + 1)(1-\rho)^2}$$

$$j^*, k^* \in \text{argmax } c^{(j,k)}$$

multiperiod is an increasing function of $\rho$; this confirms the intuition that, as the mean-reversion alternative becomes closer to the random walk case, more power will come from examining multiperiod properties of the data. For example, consider $\rho = .95$ and the variance ratio test. Comparing the approximate slopes for $k = 2$ and $k = k^*$, the ratio is approximately 10 percent. This implies that about 10 times as many observations are needed to reject the null hypothesis when

$$\text{plim}(g_T' W_0 g_T) = \frac{3[j(1-\rho^k) - k(1-\rho^j)]^2}{2jk(k-j)(2jk - 2j^2 + 1)(1-\rho)^2},$$

$$\text{plim}\left(\frac{3j}{2j^2+1}\beta^2\right) = \frac{3j(1-\rho^j)^2}{4(2j^2+1)}.$$

**Table 6**
**Approximate slope calculations for the regression test**

| $\rho$ | $c^{j^*}$ | $j^*$ | $c^1/c^{j^*}$ | $c^{j^*+25}/c^{j^*}$ |
|---|---|---|---|---|
| .99 | 0.00153 | 125 | 0.016 | 0.987 |
| .98 | 0.00309 | 62 | 0.032 | 0.957 |
| .97 | 0.00465 | 41 | 0.048 | 0.916 |
| .96 | 0.00623 | 31 | 0.064 | 0.867 |
| .95 | 0.00783 | 25 | 0.080 | 0.816 |
| .94 | 0.00944 | 20 | 0.095 | 0.777 |
| .93 | 0.01106 | 17 | 0.111 | 0.732 |
| .92 | 0.01271 | 15 | 0.126 | 0.686 |
| .91 | 0.01436 | 13 | 0.141 | 0.649 |
| .90 | 0.01604 | 12 | 0.156 | 0.606 |
| .89 | 0.01772 | 11 | 0.171 | 0.570 |
| .88 | 0.01942 | 10 | 0.185 | 0.539 |
| .87 | 0.02114 | 9 | 0.200 | 0.512 |
| .86 | 0.02284 | 8 | 0.215 | 0.490 |
| .85 | 0.02462 | 8 | 0.228 | 0.457 |
| .84 | 0.02635 | 7 | 0.243 | 0.441 |
| .83 | 0.02815 | 7 | 0.257 | 0.414 |
| .82 | 0.02989 | 7 | 0.271 | 0.391 |
| .81 | 0.03174 | 6 | 0.284 | 0.380 |
| .80 | 0.03356 | 6 | 0.298 | 0.360 |
| .79 | 0.03532 | 6 | 0.312 | 0.342 |
| .78 | 0.03720 | 5 | 0.325 | 0.335 |
| .77 | 0.03911 | 5 | 0.338 | 0.319 |
| .76 | 0.04097 | 5 | 0.351 | 0.305 |
| .75 | 0.04277 | 5 | 0.365 | 0.292 |

Approximate slope comparisons of OLS serial correlation estimators using different period lengths are provided. The approximate slope $c^j$ is a measure of the rate at which the null hypothesis (i.e., the $\tilde{\epsilon}_t$'s are uncorrelated) becomes more and more incredible as sample size increases under a given alternative [i.e., if $\tilde{\epsilon}_t$ represents growth, then the alternative is an AR(1) in levels]. The period lengths denoted by an "*" are the optimal ones in that they maximize the approximate slope (i.e., relative asymptotic power). The final two columns show the sensitivity of the approximate slope calculations to departures from the optimal lag length, $k^*$.

$$\text{Alternative: } \tilde{\epsilon}_t = \tilde{\eta}_t + (\rho - 1) \sum_{i=1}^{\infty} \rho^{i-1}\tilde{\eta}_{t-i}; \qquad |\rho| < 1, \tilde{\eta}_t \text{ i.i.d.}$$

$$c^j \equiv \text{plim}\left(\frac{3j}{2j^2 + 1}\,\beta^2\right) = \frac{3j(1 - \rho^j)^2}{4(2j^2 + 1)}$$

$$j^* \in \text{argmax } c^j$$

$k$ is chosen to equal 2. This is especially interesting given that the case $k = 2$ reduces to the well-known one-period serial correlation estimator. Table 6 shows that a similar pattern holds for the regression test (i.e., the ratio of slopes for $j = 1$ and $j = j^*$ is 8 percent). Furthermore, since the asymptotic distribution of the regression and variance ratio test is $\chi_1^2$, it is possible to compare their approximate slopes as well. For all values of $\rho$ the regression test has higher relative power. For example, at $\rho = .95$, the ratio of the plim of the optimal regression test to the optimal variance test is 1.26. Therefore, the variance ratio test needs approximately 25 percent more observations to achieve the same power.

To check the robustness of these asymptotic power comparisons

**Table 7**
**Approximate slope calculations/small sample power**

| | Regression test | | Variance test | | |
|---|---|---|---|---|---|
| Lag ($j = k$) | $c^j$ | Power (5% size) | $c^{(1,k)}$ | Power (5% size) | $c^{(1,k)}/c^j$ |
| 12 | 0.00680 | 0.381 | 0.00389 | 0.257 | 0.592 |
| 24 | 0.00783 | 0.413 | 0.00560 | 0.497 | 0.715 |
| 36 | 0.00739 | 0.329 | 0.00615 | 0.574 | 0.833 |
| 48 | 0.00654 | 0.233 | 0.00618 | 0.733 | 0.945 |
| 60 | 0.00569 | 0.130 | 0.00596 | 0.727 | 1.05 |
| 72 | 0.00495 | 0.045 | 0.00566 | 0.684 | 1.14 |
| 84 | 0.00434 | 0.029 | 0.00532 | 0.678 | 1.22 |
| 96 | 0.00385 | 0.011 | 0.00499 | 0.688 | 1.30 |

Evidence is provided on the approximation of the asymptotic power procedure (i.e., the approximate slope comparisons) to small samples. The approximate slope is a measure of the rate at which the null hypothesis (i.e., the $\tilde{\epsilon}_t$'s are uncorrelated) becomes more and more incredible as sample size increases under a given alternative [i.e., if $\tilde{\epsilon}_t$ represents growth, then the alternative is an AR(1) in levels]. $c^j$ and $c^{(1,k)}$ represent the approximate slopes of the OLS serial correlation regression statistics and the variance ratio statistics, respectively, for different period lengths. The "Power" column represents the actual small sample power of these tests under this alternative. With respect to the simulation, the distributions of the innovations are normally distributed; 720 observations are drawn for each replication; and the simulation contains 5000 repetitions.

$$\text{Alternative: } \tilde{\epsilon}_t = \tilde{\eta}_t + (\rho - 1) \sum_{i=1}^{\infty} \rho^{i-1}\tilde{\eta}_{t-i}; \qquad \rho = .95,\ \tilde{\eta}_t \text{ i.i.d.}$$

$$c^{(1,k)} \equiv \operatorname{plim}(g_T' S_0^{-1} g_T) = \frac{3[j(1-\rho^k) - k(1-\rho^j)]^2}{2jk(k-j)(2jk - 2j^2 + 1)(1-\rho)^2}$$

$$c^j \equiv \operatorname{plim}\left(\frac{3j}{2j^2+1}\beta^2\right) = \frac{3j(1-\rho^j)^2}{4(2j^2+1)}$$

against small sample implications, Table 7 shows the small sample power of the statistics (from a Monte Carlo simulation) and the approximate slopes for $\rho = .95$ with lags 12,24,...,96. The approximate slope results suggest that the optimal multiperiod for the regression and variance ratio tests are lags 24 and 48, respectively. Similarly, the Monte Carlo small sample results imply maximum power is also achieved at these lags. Furthermore, this similarity between "asymptotic power" and small sample power seems to be present for other lags. For example, the value of the approximate slope and the magnitude of the small sample power have correlation coefficients across lags equal to .96 for the regression test and .73 for the variance ratio test.[10] The approximate slope calculations fare less well in direct comparisons of the regression and variance ratio test. As implied by the approximate slope theory, Table 7 shows that the small sample

[10] Note that the actual power of the regression test drops off more quickly for higher lags than would be predicted by the approximate slope procedure. This is because higher lags effectively reduce the sample size and, therefore, the asymptotic justification. The regression test is affected more by this problem since the number of observations equals $T - 2j$; whereas with the variance ratio test the number of observations is $T - j$. With respect to the variance ratio test, Lo and MacKinlay (1988b) formally discuss some of the limitations of large lags in small samples.

power of the regression test is greater than that of the variance ratio test for lag 12. With longer lags, however, the magnitude of the approximate slopes does not coincide with actual power. For example, the variance ratio test starts achieving higher power at 24 lags even though the ratio of the approximate slopes is only .715. Moreover, its 73 percent power at lag 48 exceeds the regression test's maximum power by about 30 percent. Nevertheless, all in all, the approximate slope comparisons do provide a fairly accurate account of the power properties of these statistics.

### 3.3 Overlapping observations

The analysis in Section 3.2 suggests there can be a real gain to choosing $j$ greater than unity. For example, given an AR(1) in prices and $\rho = .95$, the approximate slope was over 12 times that of the single-period serial correlation estimator. Suppose the alternative distribution is such that the optimal $j$ is greater than unity (denoted by $j^*$), and consider the regression test of Section 2. Should the econometrician use $j^*$ overlapping periods for a given sample of $T^*$ observations or use nonoverlapping periods for a sample of $T^*/j^*$ observations?

Hansen and Hodrick (1980) consider this question in their article and show that in general the econometrician is better off using overlapping periods. Because of the simple structure of the estimation problem in this article, it is possible to quantify this benefit to overlapping data. Specifically, the standard errors of the statistics for the overlapping and nonoverlapping case, respectively, will be given by

$$\sqrt{\frac{2j^{*2}+1}{3j^*(T^*-2j^*)}} \quad \text{and} \quad \sqrt{\frac{j^*}{T^*}}.$$

Taking the ratio of these standard errors for large $T^*$ results in approximately 22 percent higher standard errors in the nonoverlapping case. Note further that for fixed $j^*$, as $T^*$ increases, the square of these serial correlation statistics will both have asymptotic $\chi^2$ representations. In terms of approximate slope comparisons, the ratio of the slopes for the overlapping to nonoverlapping case is 1.5. Therefore, in order to achieve the same power, the nonoverlapping statistic requires about 50 percent more observations. It is in this sense that there is a definite gain to using overlapping observations.

## 4. Applications

This section discusses potential applications of the above procedure. The important point from these applications is that under fairly weak assumptions it is possible to test the joint serial dependence restrictions from financial models even in the presence of overlapping data.

### 4.1 Market microstructure models

The recent availability of transactions data, which has made possible a whole new series of empirical investigations, has also introduced a host of potential measurement problems related to market microstructure effects, such as the induced first-order negative serial correlation from the bid–ask spread or from nonsynchronous trading. The issue is how financial models that neglect market microstructure effects are to be tested using data that reflect these effects. This requires that we be able to differentiate these effects from more fundamental pricing effects.

The GMM procedure given in Section 1 offers a potential solution, since enough restrictions are placed on the market microstructure and the financial asset pricing models that it is possible to test jointly the restrictions imposed by the models while simultaneously estimating the parameters. For example, consider the bid–ask model posed by Blume and Stambaugh (1983), among others:

$$\ln(\tilde{p}_t) = \mu + \ln(\tilde{p}_{t-1}) + \tilde{\epsilon}_t, \qquad \hat{p}_t = (1 + \tilde{\delta}_t)\tilde{p}_t, \tag{27}$$

where $\hat{p}_t$ is the recorded price at time $t$, $\tilde{p}_t$ is the "true" price at time $t$, and $\tilde{\delta}_t$ is a mean zero i.i.d. random variable, independent of $p_t$.

The econometrician would like to test the validity of model (27) and simultaneously estimate parameters governing the fundamental process $\tilde{p}_t$ and the bid–ask spread $\tilde{\delta}_t$. Let $m_2$ and $\sigma_\delta^2$ be the variances of true log returns and the spread, respectively. Then, in the spirit of Section 1, one possible set of moment restrictions is

$$g_T(\mu, m_2, \sigma_\delta^2) = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} \ln(\hat{p}_t/\hat{p}_{t-1}) - \mu \\ [\ln(\hat{p}_t/\hat{p}_{t-j}) - j\mu]^2 - jm_2 - 2\sigma_\delta^2 \\ [\ln(\hat{p}_t/\hat{p}_{t-k}) - k\mu]^2 - km_2 - 2\sigma_\delta^2 \\ [\ln(\hat{p}_t/\hat{p}_{t-l}) - l\mu]^2 - lm_2 - 2\sigma_\delta^2 \\ \vdots \end{pmatrix} \quad \forall j,k,l. \tag{28}$$

It is possible to identify all the unknown parameters $\{\mu, m_2, \sigma_\delta^2\}$ (i.e., $D_0$ is full rank) because the market microstructure effect diminishes with holding period length. In addition, it is possible to test the overidentifying restrictions using the GMM procedure. (See section 1.) Smith (1989) investigates different market microstructure models and applies them to transactions data on the Dow Jones 30. In addition to formally testing these models, Smith (1989) is able to differentiate the market microstructure effects from more fundamental pricing effects. Since multiperiod variances are compared to single-period variances in (28), the use of overlapping observations is desirable as it increases the number of observations and also the precision of the estimators.

### 4.2 Rational expectations model

Consider the first-order condition from a specific consumption-based asset pricing model [see, e.g., Lucas (1978)]:

$$E_t\left[\beta\left(\frac{c_t + 1}{c_t}\right)^{-\gamma}(1 + \tilde{R}_{it+1}) - 1\right] = 0, \tag{29}$$

where $c_t$ is aggregate consumption at time $t$, $\tilde{R}_{it}$ is the return on asset $i$ from $t - 1$ to $t$, $\beta$ is the rate of time preference parameter, and $\gamma$ is the constant relative risk-aversion parameter.

There is a large literature concerned with tests of this model [see, e.g., Hansen and Singleton (1982)]. Define $\tilde{\epsilon}_{it+s} = [\beta(c_{t+s}/c_t)^{-\gamma}(1 + \tilde{R}_{it+s}) - 1]$, for all assets $i = 1, \ldots, N$. Exploring the properties of $\Sigma_{s=1}^{i} \tilde{\epsilon}_{it+s}$ has potential benefits over existing tests that use single-period returns because it is sometimes possible to obtain more information about a stochastic process from its multiperiod properties.

Using results in Section 2, it is possible to provide tests for serial correlation in the multiperiod forecast errors across different assets. For example, consider two multivariate forecast errors $\tilde{\epsilon}_{nt+1}$ and $\tilde{\epsilon}_{qt+1}$ from equilibrium condition (29). Using the mean restrictions (i.e., the mean forecast error is zero) and the $j$-period serial correlation restrictions for the two assets, it is possible to derive the asymptotic variance–covariance matrix for the serial correlation estimators $\hat{\beta}_n(j)$ and $\hat{\beta}_q(j)$ of the forecast errors $\tilde{\epsilon}_{nt}$ and $\tilde{\epsilon}_{qt}$, respectively:

$$\operatorname{var}\begin{pmatrix}\hat{\beta}_n(j)\\ \hat{\beta}_q\end{pmatrix}$$

$$= \frac{2j^2 + 1}{3j}\begin{pmatrix} 1 & \left(\dfrac{\sigma_{nq}}{\sigma_n\sigma_q}\right)^2 \\ \left(\dfrac{\sigma_{nq}}{\sigma_n\sigma_q}\right)^2 & 1 \end{pmatrix}$$

$$+ \left(\frac{\text{var}(\hat{\gamma})}{j^2}\right)\begin{pmatrix} \left[\frac{\text{cov}(j,\, n)}{\sigma_n^2}\right]^2 & \frac{\text{cov}(j,\, n)\text{cov}(j,\, q)}{\sigma_n^2\sigma_q^2} \\ \frac{\text{cov}(j,\, n)\text{cov}(j,\, q)}{\sigma_n^2\sigma_q^2} & \left[\frac{\text{cov}(j,\, q)}{\sigma_q^2}\right]^2 \end{pmatrix}, \tag{30}$$

where

$$\text{cov}(j,\, n) = \text{cov}\left(\sum_{s=1}^{j} \tilde{\epsilon}_{nt+s} \ln \frac{C_{t+s}}{C_{t+s-1}}, \quad \sum_{s=1}^{j} \tilde{\epsilon}_{nt+s-j}\right),$$

$$\text{cov}(j,\, q) = \text{cov}\left(\sum_{s=1}^{j} \tilde{\epsilon}_{qt+s} \ln \frac{C_{t+s}}{C_{t+s-1}}, \quad \sum_{s=1}^{j} \tilde{\epsilon}_{qt+s-j}\right),$$

$$\sigma_i^2 = \text{variance of } \tilde{\epsilon}_i \quad \forall i,$$

$$\sigma_{nq} = \text{covariance between } \tilde{\epsilon}_n \text{ and } \tilde{\epsilon}_q.$$

Note that the variance–covariance matrix of the $\hat{\beta}_i(j)$'s can be written as the sum of two terms. The first term reflects the variance–covariance matrix of the estimators $\hat{\beta}_i(j)$ given above. The second term reflects the portion of the standard errors caused by estimation of the parameter $\gamma$. Since the diagonal elements of this term are positive, the standard errors of $\hat{\beta}_n(j)$ and $\hat{\beta}_q(j)$ will in general be higher. Even if $\gamma$ cannot be measured precisely, the econometrician can potentially reduce this measurement problem by choosing assets to minimize $\text{cov}(j, i)$ $(i = n, q)$. This example serves to demonstrate the procedure's ability to tackle fairly complex models.

## 5. Conclusion

Financial models imply restrictions on the moments of economic variables. These restrictions can be tested directly using Hansen's (1982) GMM approach. This approach has the intuitively appealing property that only minimal distributional assumptions outside the model need to be made. While the application of GMM to asset pricing models is not new, the techniques introduced in this article are useful in tackling models that impose joint serial dependence restrictions; for example, the market microstructure models and multiyear autocorrelations discussed earlier. In addition, there are several technical contributions in this article.

(i) With few additional assumptions, we are able to derive simple expressions for resulting test statistics and estimators. Among other benefits, this results in less sampling variation, leading to better size properties.

(ii) Using the approximate slope of Geweke (1981), we make asymptotic power comparisons between different statistics. Specifically, we explore the power advantages of using multiperiod data.

(iii) The benefit to overlapping observations is explicitly quantified.

## Appendix A

Under the model example in Section 1, the demeaned terms of the moment conditions of Equation (7) can be written as

$$w_l = \begin{pmatrix} \epsilon_t \\ (\Sigma_{i=0}^{j-1} \epsilon_{t-i-l})^2 - jm_2 \\ (\Sigma_{i=0}^{k-1} \epsilon_{t-i-l})^2 - km_2 \end{pmatrix}. \tag{A1}$$

Now denote $R_w(l) = E[w_0 w_l']$. Hansen (1982) shows that the variance–covariance matrix of the moment conditions will in general be given by

$$S_0 = \sum_{l=-\infty}^{+\infty} R_w(l). \tag{A2}$$

For the model looked at in this article, $S_0$ takes an especially easy form since the autocovariance function is nonzero only up to the highest overlap. That is,

$$S_0 = \sum_{l=-k+1}^{k-1} R_w(l). \tag{A3}$$

Using the assumptions (i)–(v) given in Section 1, it is possible to calculate $E[w_0 w_l']$. Specifically, for a given $l$, each element has the following form:

$$E[\epsilon_t \epsilon_{t-l}] = \begin{cases} 0, & \text{if } l \neq 0, \\ m_2, & \text{if } l = 0; \end{cases}$$

$$E\left(\left[\left(\sum_{i=0}^{\kappa-1} \epsilon_{t-i-l}\right)^2 - \kappa m_2\right]\epsilon_t\right)$$

$$= \begin{cases} 0, & \text{if } l > 0, \\ m_3, & \text{if } l \leq 0, \end{cases} \qquad \forall \kappa = j \text{ or } k;$$

$$E\left(\left[\left(\sum_{i=0}^{\kappa-1}\epsilon_{t-i}\right)^2 - \kappa m_2\right]\left[\left(\sum_{i=0}^{\kappa-1}\epsilon_{t-i-l}\right)^2 - \kappa m_2\right]\right)$$

$$= ([2(\kappa - l)^2 - 3(\kappa - l)]m_2^2 + \kappa m_4) \qquad \forall l \leq \kappa, \quad \kappa = j \text{ or } k;$$

$$E\left(\left[\left(\sum_{i=0}^{j-1}\epsilon_{t-i}\right)^2 - jm_2\right]\left[\left(\sum_{i=0}^{k-1}\epsilon_{t-i-l}\right)^2 - km_2\right]\right)$$

$$= [2(\max(j - l, 0))^2 - 3\max(j - l, 0)]m_2^2$$

$$+ \max(j - l, 0)m_4 \qquad \forall l \leq k, j \leq k;$$

$$E\left(\left[\left(\sum_{i=0}^{k-1}\epsilon_{t-i}\right)^2 - km_2\right]\left[\left(\sum_{i=0}^{j-1}\epsilon_{t-i-l}\right)^2 - jm_2\right]\right)$$

$$= [2(\min(j, k - l, 0))^2 - 3\min(j, k - l)]m_2^2$$

$$+ \min(j, k - l)m_4 \qquad \forall l \leq k, j \leq k.$$

Substituting these covariances into (A3) above and performing the summation yields the desired result in Equation (10).

## Appendix B

From Hansen (1982), we know that the asymptotic variance of the estimators in moment condition (20) is given by

$$(D_0)^{-1}S_0(D_0')^{-1},$$

where

$$D_0 = E\left[\frac{\partial g_T}{\partial(\alpha, \beta)}\right],$$

$$S_0 = \sum_{l=-\infty}^{+\infty} E[w_0 w_l'],$$

$$w_l = \begin{pmatrix} \vdots \\ \Sigma_{i=1}^{j}\tilde{\epsilon}_{t+i-l} - \alpha_j - \beta_j(\Sigma_{i=1}^{j}\tilde{\epsilon}_{t-j+i-l}) \\ [\Sigma_{i=1}^{j}\tilde{\epsilon}_{t+i-l} - \alpha_j - \beta_j(\Sigma_{i=1}^{j}\tilde{\epsilon}_{t-j+i-l})][\Sigma_{i=1}^{j}\tilde{\epsilon}_{t-j+i-l}] \\ \vdots \\ \Sigma_{i=1}^{k}\tilde{\epsilon}_{t+i-l} - \alpha_k - \beta_k(\Sigma_{i=1}^{k}\tilde{\epsilon}_{t-j+i-l}) \\ [\Sigma_{i=1}^{k}\tilde{\epsilon}_{t+i-l} - \alpha_k - \beta_k(\Sigma_{i=1}^{k}\tilde{\epsilon}_{t-j+i-l})][\Sigma_{i=1}^{k}\tilde{\epsilon}_{t-j+i-l}] \\ \vdots \end{pmatrix}.$$

Under the Hansen and Hodrick (1980) assumptions, $S_0$ takes an especially easy form since the autocovariance function is nonnegative only up to the highest overlap. That is,

$$S_0 = \sum_{l=-k^*+1}^{k^*-1} E[w_0 w_l'], \tag{B1}$$

where $k^*$ is the largest holding period.

Without loss of generality, consider two different holding periods, $j$ and $k$. For a given $l$, denote the two moment conditions related to period $j$ as $w_{j,l}$ and $w_{jj,l}$, respectively. Using the assumptions, it is possible to then calculate for a given $l$ each element of $S_0$ analytically. Specifically, these elements take the form

$$E[w_{j,0} w_{j,l}] = \begin{cases} (j-l)m_2, & \text{if } j > l > 0, \\ (j+l)m_2, & \text{if } -j < l \leq 0; \end{cases}$$

$$E[w_{j,0} w_{jj,l}] = \begin{cases} j(j-l)m_1 m_2, & \text{if } j > l > 0, \\ j(j+l)m_1 m_2, & \text{if } -j < l \leq 0; \end{cases}$$

$$E[w_{jj,0} w_{jj,l}] = \begin{cases} (j-l)^2 m_2^2 + j^2(j-l)m_1^2 m_2, & \text{if } j > l > 0, \\ (j+l)^2 m_2^2 + j^2(j+l)m_1^2 m_2, & \text{if } -j < l \leq 0; \end{cases}$$

$$E[w_{j,0} w_{k,l}] = \begin{cases} \min(k-l, j)m_2, & \text{if } k > l > 0, \\ \max(j+l, 0)m_2, & \text{if } -k < l \leq 0; \end{cases}$$

$$E[w_{j,0} w_{kk,l}] = \begin{cases} \min(k-l, j)km_2 m_1, & \text{if } k > l > 0, \\ \max(j+l, 0)km_2 m_1, & \text{if } -k < l \leq 0; \end{cases}$$

$$E[w_{jj,0} w_{k,l}] = \begin{cases} \min(k-l, j)jm_2 m_1, & \text{if } k > l > 0, \\ \max(j+l, 0)jm_2 m_1, & \text{if } -k < l \leq 0; \end{cases}$$

$$E[w_{jj,0} w_{kk,l}] = \begin{cases} \min(j, k-l)\max(j-l, 0)m_2^2 \\ \quad + \min(j, k-l,)jkm_1^2 m_2, & \text{if } k > l > 0, \\ \min(j, k+l)\max(j+l, 0)m_2^2 \\ \quad + \max(j+l, 0)jkm_1^2 m_2, & \text{if } -k < l \leq 0; \end{cases}$$

Substituting these covariances and variances into (B1) above and performing the summation yields the following result:

$$S_0 = \begin{pmatrix} j^2 m_2 & j^3 m_2 m_1 & jkm_2 & jk^2 m_2 m_1 \\ j^3 m_2 m_1 & \frac{1}{3}[j(2j^2+1)m_2^2] + j^4 m_2 m_1^2 & j^2 k m_2 m_1 & (s(j,k)+j^2)m_2^2 + j^2 k^2 m_2 m_1^2 \\ jkm_2 m_1 & j^2 k m_2 m_1 & k^2 m_2 & k^3 m_1 m_2 \\ jk^2 m_2 m_1 & (s(j,k)+j^2)m_2^2 + j^2 k^2 m_2 m_1^2 & k^3 m_1 m_2 & \frac{1}{3}[k(2k^2+1)m_2^2] + k^4 m_2 m_1^2 \end{pmatrix}. \tag{B2}$$

Under the aforementioned assumptions, the analytical derivative matrix can also be calculated:

$$D_0 = \begin{pmatrix} -1 & -jm_1 & 0 & 0 \\ -jm_1 & j^2m_1^2 + jm_2 & 0 & 0 \\ 0 & 0 & 1 & km_1 \\ 0 & 0 & km_1 & k^2m_1^2 + km_2 \end{pmatrix}. \quad \text{(B3)}$$

Since the above matrices hold for all $j$ and $k$, it is possible to set up $2m \times 2m$ matrices $S_0$ and $D_0$, where $m$ is the number of holding period returns. Using (B2) and (B3), the matrix operations on $D_0^{-1} S_0 D_0^{-1}$ can be performed to get the desired result given in Equation (21).

**References**

Bahadur, R. R., 1960, "Stochastic Comparison of Tests," *Annals of Mathematical Statistics,* 31, 276–295.

Blume, M., and R. Stambaugh, 1983, "Biases in Computed Returns: An Application to the Size Effect," *Journal of Financial Economics,* 12, 387–404.

Cochrane, J., 1988, "How Big Is the Random Walk in GNP?" *Journal of Political Economy,* 96, 893–920.

Eichenbaum, M., L. Hansen, and K. Singleton, 1988, "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty," *Quarterly Journal of Economics,* 51–78.

Fama, E., and K. French, 1988, "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy,* 96, 246–273.

Geweke, J., 1981, "The Approximate Slope of Econometric Tests," *Econometrica,* 49, 1427–1442.

Gibbons, M. R., S. A. Ross, and J. Shanken, 1989, "A Test of the Efficiency of a Given Portfolio," *Econometrica,* 57, 1121–1152.

Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica,* 50, 1029–1054.

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.

Hansen, L., and K. Singleton, 1982, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica,* 50, 1269–1286.

Huizinga, J., 1983, "Small Sample Properties of Market Efficiency Tests in Models with Overlapping Data," University of Chicago and National Bureau of Economic Research.

Kendall, M., 1954, "Note on Bias in the Estimation of Autocorrelation," *Biometrica,* 41, 403–404.

Lo, A., and C. MacKinlay, 1988a, "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, 1988b, "The Size and Power of the Variance Ratio Test in Finite Samples: A Monte Carlo Investigation," *Journal of Econometrics,* 40, 203–238.

Lucas, R., 1978, "Asset Prices in an Exchange Economy," *Econometrica,* 46, 1426–1446.

Newey, W., and K. West, 1987, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica,* 55, 703–708.

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," working paper, Stanford University, Graduate School of Business.

Smith, T., 1989, "Financial Models and Market Microstructure," working paper, Fuqua School of Business, Duke University.

White, H., 1980, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica,* 48, 817–838.