Journal of Empirical Finance 8 (2001) 493–535
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/econbase

The power and size of mean reversion tests

Kent Daniel *

Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208-2006, USA

Keywords: Mean reversion tests; Power; Stock prices

1. Introduction

The power of mean reversion tests has long been a tacit issue of the market efficiency literature. Early tests of market efficiency, as summarized in Fama (1970), found no economically significant evidence of serial correlation in stock returns. However, Summers (1986) later suggested that this was because these tests lacked power: Summers suggested a model of "fads" in which stock prices take long swings away from their fundamental values, and showed that even if a fads component such as this accounted for a large fraction of the variance of returns, the fads behavior might be difficult to detect by looking at short horizon autocorrelations of returns as these early tests had done.

The intuition behind Summers' reasoning was that if stock prices took large jumps away from their "fundamental" or full-information values, and then only reverted back towards the fundamental price over a period of years, the autocorrelations of monthly or daily returns would capture only a small fraction of this mean reversion.

Several attempts were made to develop tests that would have greater power against "fads" hypotheses such as Summers'. Fama and French (1988a) used a long horizon regression of multi-year returns on past multi-year returns, and Poterba and Summers (1988) used a variance ratio test to look for fads-type behavior in stock-index returns. In addition, variance ratio test are used by Cochrane (1988) and Lo and MacKinlay (1988) to investigate the time series properties of production and short horizon returns.

* Tel.: +1-847-491-4113.
E-mail address: kentd@nwu.edu (K. Daniel).

PII: S0927-5398(01)00038-X

Both Fama and French and Poterba and Summers develop intuition for why these long horizon tests should have more power to detect fads type behavior, and some effort has since been made to both verify and formalize this intuition. Lo and MacKinlay (1989) use Monte-Carlo methods to compare the power of the variance ratio, Box–Pierce Q, and the Dickey and Fuller (1979) τ-tests. Jegadeesh (1990) used the approximate slope method (Badahur, 1980; Geweke, 1981) to evaluate the power of a generalized long horizon regression, and Richardson and Smith (1991) use this method to evaluate the power of the variance ratio test and long horizon regression against specific alternatives. Hodrick (1992) and Campbell (1992) propose similar analyses for multivariate tests.

However, these papers are all comparisons of power across a discrete set of tests and for a specific mean reverting alternative; none presents a method for determining the most powerful test or suggests how far their tests might be from optimal for the specified alternative. Moreover, little or no intuition is provided as to the robustness of these results with respect to changes in the alternative hypothesis.

In this paper, we develop a methodology for determining asymptotic test power. This method allows us (1) to determine the most powerful test against a specified alternative; (2) to determine the distance of a test from the optimal test using an analytical measure of test power and (3) to determine the implicit alternative to any test.[1] Moreover, the straightforward geometric interpretation of test power we present facilitates consideration of test robustness issues.

It is important to note here that what we present a method for constructing an optimal test *once the alternative hypothesis has been determined*. We do not treat the problem of actually specifying the alternative hypothesis, which is very difficult problem, and is probably the reason that so many ad hoc tests have been used in the finance field. Nonetheless, in the debates over what type of test is appropriate, test power is often an issue that is ignored, or is addressed using Monte-Carlo methods that are not robust to small changes in the alternative hypothesis. The method we present here does allow us to address those questions.

The methodology we develop is applicable to all moment restriction tests where the instrument is a linear combination of past returns. This class encompasses the long horizon regression test, the variance ratio test, weighted spectral tests (Durlauf), and instrumental variable and generalized method of moments (GMM) tests involving past returns.

Since our analytical test-power results are valid only asymptotically and under local alternatives, we validate these results for small sample and nonlocal alternatives using Monte-Carlo experiments. We find that the asymptotic results extend

[1] The implicit alternative of a test is that alternative against which the test is the most powerful, which we shall discuss later.

well to small samples but also show that two asymptotically equivalent tests may have different small sample properties.

This test-power determination method extends naturally to the consideration of joint tests of moment restrictions. This issue is of importance in the finance literature: in attempting to characterize a time series of returns, a common approach in the finance literature is to run a set of tests in order to determine the time series properties of the returns series. For example, Fama and French run a set of eight long horizon regressions at return horizons of 1, 2, 3, 4, 5, 6, 8 and 10 years. Poterba and Summers (1988) perform variance ratio tests for similar horizons. Both find evidence of mean reversion at some horizons. However, as Richardson (1993) points out, the significance of these results must be based on the *joint* significance of all tests. Richardson and Smith (1991) suggest calculating the joint significance by forming a χ^2 statistic where the variance–covariance matrix of the sample regression coefficients is calculated under the null hypothesis. A similar approach is adopted by Jegadeesh (1990) and others.

However, we show that a χ^2 joint test of this form will have very low power, even if the individual tests are all powerful against the alternative.

One way of interpreting the Fama and French and Poterba and Summers tests is that a number of horizons were used because the researchers had the alternative that returns were mean reverting, but were unsure of the degree of persistence of the mean reverting component. They, therefore, studied a set regressions (or variance ratios) that bounded the range of mean reversion rates they expected to see, and estimated the rate of mean reversion by determining the return horizon at which the regression coefficient was most significant. This would have been a statistically correct procedure had they corrected for the fact that they had searched over a large number of regression coefficients. This would have been similar to a procedure in which the mean reversion coefficient was estimated from the data, and then, using this parameter estimate (for example in a GMM setting), a test was conducted of whether the variance of the mean reverting component was significantly different from zero. While a test such as this would have power that is independent of the number of regressions run, the power of the χ^2 test of the joint significance of the regression coefficient decreases as the number of regressions (variance ratios) increases. Thus, this method of testing is inherently statistically weak. This result is verified using Monte-Carlo Studies.

Finally, since our analytical test-power results are valid only asymptotically and under local alternatives, we also conduct Monte-Carlo experiments to investigate the robustness of these results for small sample sizes and for nonlocal alternatives. We find that the results are generally robust, but we also explore situations where the asymptotic theory will lead to incorrect conclusions. We extend the results of this section to show how small differences in the small sample properties of a test can lead to strikingly different statistical inferences. We show that the long horizon regression, which uses analytical standard errors as proposed by Richardson and Smith (1991), suffers from low power against simple mean reverting alternatives, and that this, not poor small properties, is the reason Richardson and Stock (1989) find no evidence of mean reversion using this test. We empirically calculate the small-sample corrected distribution for the Fama and French T-statistics, which are based on Hansen and Hodrick (1980) calculated standard errors, and show that there is still a good deal of evidence in favor of a mean reversion hypothesis. We show why a test based on the Hansen/Hodrick based T-statistic is more powerful even though the two test are asymptotically equivalent.

We proceed by showing that all of these are asymptotically equivalent to weighted autocorrelation tests, and develop the result that, for univariate tests, the most powerful test statistic is that which is a weighted sum of sample autocorrelations at different lags, for which the weights are proportional to the expected autocorrelation under the alternative hypothesis.

The intuition behind this method is straightforward, and is based on the fact that under the null hypothesis, the vector of sample autocorrelations at different lags is asymptotically mean zero, and is multivariate-normally distributed with a variance–covariance matrix $\mathbf{\Omega} = (1/T) \cdot \mathbf{I}$. In other words, sample autocorrelations at different lags have the same variance and are uncorrelated. If one changes the hypothesis from the null to the local alternative hypothesis (Davidson and MacKinnon (1987)) (i.e., if the serial correlation is small), the mean of the sample autocorrelation vector will shift in the direction of the alternative but the variance covariance matrix of sample autocorrelations will remain the same. Given these null and alternative autocorrelation distributions, we show that the most powerful test statistic is a linear combination of sample autocorrelations where the weighting vector is proportional to the vector of expected sample autocorrelations.

One of the virtues of writing these tests as weighted autocorrelation tests is that it leads to simple geometric interpretation of test power, which we provide in Section 2.3. We show in Section 2.5.1 that the weighted autocorrelation test can just as easily be written in the spectral domain as a weighted periodigram test, with an analogous result that the optimal test will have weights proportional to the expected periodigram under the alternative. This test has same optimality properties as the weighted autocorrelation test.

We also show another version of the optimal test is one tests the orthogonality of the current return to the optimal predictor of the current return, based on the alternative hypothesis.

Two other papers explore the topic of determining an optimal test. Faust (1992) presents a method for determining the optimal filtered variance ration test based on maximum likelihood methods. Perhaps most closely related to this paper is Richardson and Smith (1994), which develops a general method for determining the optimal test given a mean-reverting alternative. Using the approximate slope method as a measure of test power, Richardson and Smith reach conclusions on the optimal test statistics, which are similar to those we present in Sections 2.1 and 2.2. In addition, Richardson and Smith compare the power of their optimal test for the Summers' fads alternative to the long horizon regression test, to the variance ration test, and to the Jegadeesh (1990) regression both asymptotically (using the approximate slope measure) and in small-samples using Monte Carlo methods.

The paper is organized as follows. Section 2 develops the weighted autocorrelation test and proves its optimality, and extends this development to the spectral domain and to calculation of the optimal instrument. Section 3 demonstrates the equivalence of commonly used mean reversion tests to weighted autocorrelation tests, and investigates their optimality and the implicit alternatives of these tests. Section 4 extends the analysis to joint test of restrictions, and Section 3.4 presents Monte-Carlo results on the small sample power of the tests. Section 5 reexamines the Fama and French (1988b) long horizon test for mean reversion in light of this evidence. Section 6 concludes the paper.

2. The optimal univariate tests—the weighted autocorrelation test

In this section, we derive the asymptotic properties of the weighted autocorrelogram test and show that this test is asymptotically a uniformly most powerful test against a local alternative for which the return generating process can be described by an ARMA model. By uniformly most powerful, we mean that for any significance level (or probability of Type I error) selected by the econometrician, the probability of Type II error is minimized. We also provide a simple geometric illustration of the power of the test.

2.1. The local alternative hypothesis

We begin with a Pitman sequence of local data, or return, generating processes (DGPs):

$$\tilde{r}_t = \mu + \alpha T^{-\frac{1}{4}} \tilde{a}_t + \tilde{u}_t \tag{2.1}$$

where u_t and a_t are given by

$$\tilde{u}_t \sim \mathrm{IID}\left(0, \sigma_u^2\right) \qquad E\tilde{u}^4 = \eta_u \sigma_u^4 < \infty \tag{2.2}$$

$$\theta(L)\tilde{a}_t = \phi(L)\tilde{\epsilon}_t \tag{2.3}$$

$$\tilde{\epsilon}_t \sim \mathrm{IID}\left(0, \sigma_\epsilon^2\right) \qquad E\tilde{\epsilon}^4 = \eta_\epsilon \sigma_\epsilon^4 < \infty \tag{2.4}$$

$$E\left(\tilde{u}_t \tilde{\epsilon}_{t-\tau}\right) = 0 \, \forall \tau \tag{2.5}$$

$\phi(L)$ and $\theta(L)$ are finite-order lag polynomials, and $\phi(z)/[(1-z)\theta(z)]$ has roots outside the unit circle.

This return is seen to be composed of two components, the $\tilde{u}$ component, which is a differenced martingale, and the $\tilde{a}$ or "alternative" component, which has as ARMA representation. We assume that the correlation of ϵ_t and u_t is zero.[2]

[2] This assumption is not critical in that any ARMA process for returns can be decomposed in this way.

α is the parameter which determines how close the local alternative is to the null hypothesis of white noise returns. Notice that null hypothesis is nested within the alternative in the sense that when $\alpha = 0$ the null is true, and when α is any value other than zero, the alternative is true. Here, the return generating process under the null hypothesis is allowed to be nonnormal, but must have a finite fourth moment.[3]

Eq. (2.1) represents a *sequence* local DGPs: as the sample size of T increases, the variance of mean-reversion component grows smaller. The factor of $T^{-1/4}$ in the return generating process is chosen so that, given a fixed size, the power of the test will converge to some value in (0, 1) as $T \to \infty$.[4] In the interest of tractability, we must deal with asymptotic power, that is the power of the test $T \to \infty$. However, if we were to increase T without changing the importance of the mean-reverting component, the power would always go to one as $T \to \infty$. To allow asymptotic power analysis, it is necessary to modify the alternative hypothesis as T grows, to move it "closer" to the null so that the asymptotic probability of rejection under this local DGP is in (0, 1). As we show later, this type of convergence will occur only with an exponent of $-1/4$.

Given the definitions in Eqs. (2.1)–(2.5), the covariogram of the returns series r_t is given by:

$$c_\tau^a = E[\tilde{a}_t \tilde{a}_{t+\tau}] = \sigma_\epsilon^2 \int_\Gamma \frac{\phi(z)\phi(z^{-1})}{\theta(z)\theta(z^{-1})} z^\tau \mathrm{d}z \tag{2.6}$$

where Γ is the unit circle in the complex plane. We write the autocorrelation estimator for the r_t series as:

$$\hat{\rho}_\tau = \frac{\hat{c}_\tau}{\hat{c}_0} = \frac{\frac{1}{T}\Sigma r_t r_{t-\tau}}{\frac{1}{T}\Sigma r_t^2} \tag{2.7}$$

Expanding the numerator yields:

$$\hat{c}_\tau = \frac{1}{T}\sum_{t=\tau}^{T} r_t r_{t-\tau} = \frac{1}{T}\sum_{t=\tau}^{T}\left(\alpha T^{-\frac{1}{4}} \tilde{a}_t + \tilde{u}_t\right)\left(\alpha T^{-\frac{1}{4}} \tilde{a}_{t-\tau} + \tilde{u}_{t-\tau}\right)$$

---

[3] Richardson and Smith (1991) also show that their optimal test is robust to limited kinds of heteroskedasticity.

[4] In the work of Davidson and MacKinnon (1987) and others concerning local alternative hypotheses, this is usually a factor of $T^{-1/2}$. However, this is in a regression framework where only the DGP for the dependent variables varies with T. In our framework, where returns are both the dependent and independent variables, we want the product of these two to move towards the null at a rate of $T^{-1/2}$, so each part individually must move at the rate of $T^{-1/4}$.

$$\hat{c}_\tau = \underbrace{\frac{\alpha^2}{\sqrt{T}}\left(\frac{1}{T}\sum_t \tilde{a}_t \tilde{a}_{t-\tau}\right)}_{\overset{\text{asy}}{\sim} \mathcal{N}\left(\frac{a^2 c_\tau^a}{\sqrt{T}}, O[T^{-2}]\right)} + \underbrace{\frac{1}{T}\sum_t \tilde{u}_t \tilde{u}_{t-\tau}}_{\overset{\text{asy}}{\sim} \mathcal{N}\left(0, \sigma_u{}^4 T^{-1}\right)}$$

$$+ \underbrace{\alpha T^{-\frac{1}{4}}\left(\frac{1}{T}\sum_t a_t u_{t-\tau} + \frac{1}{T}\sum_t a_{t-\tau} u_t\right)}_{\overset{\text{asy}}{\sim} \mathcal{N}\left(0, O\left[T^{-\frac{3}{2}}\right]\right)} \tag{2.8}$$

By Eq. (A.3), the first term has an expected value of $\frac{\alpha^2 c_\tau^a}{\sqrt{T}}$ and a variance of $(\alpha^4/T^2)v_{\tau\tau} = O_p(T^{-2})$. By Eq. (A.4), the second term has a mean that is asymptotically zero and a variance of $\sigma_u{}^4 T^{-1}$. The expectation of the last term is zero since $\tilde{u}_t$ and $\tilde{a}_t$ are mean zero and independent. The asymptotic variance of this term is therefore:

$$= \alpha^2 T^{-\frac{5}{2}} E\left[\sum_t \left(a_t^2 u_{t-\tau}^2 + a_{t-\tau} u_t^2 + 2 a_{t+\tau} u_t^2 a_{t-\tau}\right)\right]$$

$$= \alpha^2 T^{-\frac{5}{2}}\left(2T\sigma_a^2 \sigma_u{}^2 + 2T\sigma_u{}^2 c_{2\tau}^a\right) = 2\alpha^2 \sigma_u{}^2 T^{-\frac{3}{2}}\left(\sigma_a{}^2 + c_{2\tau}^a\right) = O_p\left(T^{-\frac{3}{2}}\right) \tag{2.9}$$

The $\text{plim}_{T\to\infty}$ of the denominator is σ_u^2 while, based on the central limit theorem, the numerator tends to a sum of normally distributed random variables. Given this, the distribution of the sample autocorrelation is given by:

$$\hat{\rho}_\tau \overset{\text{asy}}{\sim} \mathcal{N}\left(\frac{\alpha^2 c_\tau^a}{\sigma_u^2 \sqrt{T}}, \frac{1}{T}\right) \tag{2.10}$$

The covariance of the autocorrelation estimator at different lags is obtained by performing term by term multiplication of the three in the expansion of the expansion of the covariance estimator $\hat{c}_\tau$ in Eq. (2.8). Denote these three terms by $\tilde{A}_\tau$, $\tilde{B}_\tau$, and $\tilde{C}_\tau$. We then have that

$$\begin{aligned}\mathrm{Cov}\left[\hat{c}_s,\hat{c}_\tau\right] &= E\left[\hat{c}_s\hat{c}_\tau\right] - E\left[\hat{c}_s\right]E\left[\hat{c}_\tau\right] \\ &= E\Big[\tilde{A}_s\cdot\tilde{A}_\tau + \tilde{B}_s\cdot\tilde{A}_\tau + \tilde{C}_s\cdot\tilde{A}_\tau \\ &\qquad + \tilde{A}_s\cdot\tilde{B}_\tau + \tilde{B}_s\cdot\tilde{B}_\tau + \tilde{C}_s\cdot\tilde{B}_\tau \\ &\qquad + \tilde{A}_s\cdot\tilde{C}_\tau + \tilde{B}_s\cdot\tilde{C}_\tau + \tilde{C}_s\cdot\tilde{C}_\tau\Big]\end{aligned}$$

Independence of $\tilde{u}$ and $\tilde{a}$ plus assumption (2.2) guarantees that the expectations of the $\tilde{A}\cdot\tilde{C}$, $\tilde{B}\cdot\tilde{C}$ and $\tilde{B}\cdot\tilde{B}\cdot$ terms are zero. Expansion of the $\tilde{C}_s\cdot\tilde{C}_\tau$ in a manner similar so that in Eq. (2.9) above yields:

$$\begin{aligned}&E\left[\tilde{C}_s\cdot\tilde{C}_\tau\right] \\ &\quad = \frac{2\alpha^2}{T^{3/2}}E\left[\tilde{a}_{t+s}\tilde{a}_{t+\tau}\tilde{u}_t^2 + \tilde{a}_{t-s}\tilde{a}_{t+\tau}\tilde{u}_t^2 + \tilde{a}_{t+s}\tilde{a}_{t-\tau}\tilde{u}_t^2 + \tilde{a}_{t-s}\tilde{a}_{t-\tau}\tilde{u}_t^2\right] \\ &\quad \overset{\text{asy}}{=} \frac{2\alpha^2\sigma_u^2}{T^{3/2}}\left(c_{s-\tau}^a + c_{s+\tau}^a\right)\end{aligned}$$

and from Eq. (A.3), the expectation of the remaining term, $\tilde{A}_s\cdot\tilde{A}_\tau - (\alpha^4/T)c_\tau^a c_s^a$ is asymptotically:

$$E\left[\tilde{A}_s\cdot\tilde{A}_\tau - \frac{\alpha^4}{T}c_\tau^a c_s^a\right] \overset{\text{asy}}{=} \frac{\alpha^4}{T^2}v_{s\tau}$$

and summing the last two terms gives:

$$\mathrm{Cov}\left(\hat{c}_\tau,\hat{c}_s\right) = E\left[\hat{c}_\tau\hat{c}_s\right] - E\left[\hat{c}_\tau\right]E\left[\hat{c}_s\right] \overset{\text{asy}}{=} \frac{2\alpha^2\sigma_u^2}{T^{\frac{3}{2}}}\left(c_{s-\tau}^a + c_{s+\tau}^a\right)$$

which is $O_p(T^{-3/4})$. Combining this result with the fact that $\mathrm{plim}_{T\to\infty}$ of the denominator of Eq. (2.7) is σ_u^2 and with Eq. (2.8) yields:

$$\sqrt{T}\,\hat{\boldsymbol{\rho}}_\tau \overset{\text{asy}}{\sim} \mathscr{N}\left(\frac{\alpha^2\mathbf{c}_\tau^a}{\sigma_u^2},\mathbf{I}\right) \tag{2.11}$$

2.2. *The weighted autocorrelogram test statistic*

We now proceed to find the most powerful test. We proceed by first deriving the optimal test among the class of tests that are *linear* functions of autocorrelations, and then showing in Section 2.4 that this linear test is optimal among *all* functions of the autocorrelations. Since the autocorrelations (plus the variance) summarize the properties of any series that has an ARMA representation, this test will be globally optimal.

We define the *weighted autocorrelogram test statistic* as:

$$\hat{A} = \left(\sum_{\tau} w_{\tau} \hat{\rho}_{\tau} \right) \tag{2.12}$$

where, without loss of generality, the length of the vector of weights is normalized to one:

$$\sum_{\tau} w_{\tau}^2 = 1 \tag{2.13}$$

From Eqs. (2.7) and (2.11) we have that:

$$\hat{A} \overset{\text{asy}}{\sim} \mathcal{N}\left(\frac{\alpha^2}{\sigma_u^2 T^{\frac{1}{2}}} \sum_{\tau} w_{\tau} c_{\tau}^a, \frac{1}{T} \sum_{\tau} w_{\tau}^2 \right)$$

$$\sqrt{T}\,\hat{A} \overset{\text{asy}}{\sim} \mathcal{N}\left(\frac{\alpha^2}{\sigma_u^2} \sum_{\tau} w_{\tau} c_{\tau}^a, 1 \right) \tag{2.14}$$

Notice that both the mean and the variance of the distribution of $\sqrt{T}\hat{A}$ are independent of T. This means that the probability of rejection as $T \to \infty$ is in (0, 1). Had we written the DGP in Eq. (2.1) with an exponent of $T^{-1/4}$, this would not have been the case.

If our alternative hypothesis does not suggest a sign for α, we will use the test statistic $T\hat{A}^2$, which based on Eq. (2.14), has a noncentral χ^2 distribution with one degree of freedom and with noncentrality parameter NCP.

$$\text{NCP} = \frac{\alpha^4}{\sigma_u^4} \left(\sum_{\tau} w_{\tau} c_{\tau}^a \right)^2 \tag{2.15}$$

Since under the null hypothesis ($\alpha = 0$) this statistic has a central χ_1^2 distribution, to maximize the power of the test under the local alternative represented by the $\tilde{a}_t$ DGP in Eq. (2.1), the weight ω_T must be chosen to maximize the noncentrality parameter (NCP), subject to normalization constraint that the sum of the squares of the weights equals 1.

The intuition behind this result is illustrated in Fig. 1, where χ_1^2 density functions with NCPs of 0, 2, and 4 are plotted. Since the test statistic $T\hat{A}^2$ for any set of weights satisfying Eq. (2.13) has a central χ^2 distribution under the null, a single critical value will give all tests the same size. For example, a critical value of $\chi^* = 3.84$ gives all tests a size of 5%. Maximizing the power of the test is then equivalent to choosing the test for which it is most likely that the test statistic will

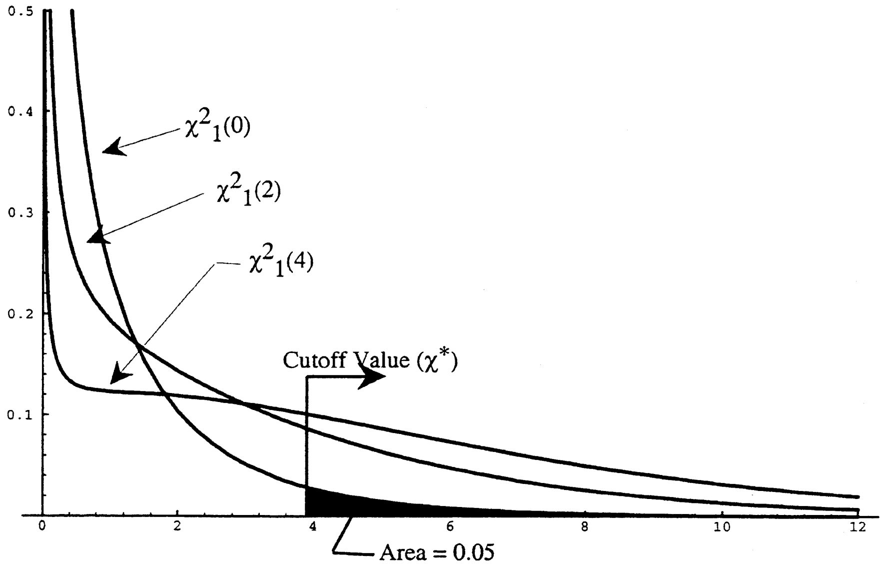

Fig. 1. Test power as a function of the noncentrality parameter.

exceed the critical value of 3.84 given the alternative is true. In other words, we need to find the value of the NCP which maximizes the integral

$$\int_{\chi^*}^{\infty} \chi_1^2(\mathrm{NCP})(x)\,\mathrm{d}x$$

Because a χ^2 distribution with a larger NCP first-order stochastically dominates a χ^2 with a lower NCP, the test which has the highest NCP will always maximize this integral, regardless of the size or critical value we choose.

To determine the set of weights which maximizes the NCP, we solve the Lagrangian:

$$\mathscr{L} = \sum_{\tau} w_\tau c_\tau^a - \lambda\left(\sum_{\tau} w_\tau^2 - 1\right)$$

Taking the first-order conditions gives the optimal weights:

$$\frac{\partial \mathscr{L}}{\partial w_\tau} = 0 \Rightarrow w_\tau^* = \frac{c_\tau^a}{2\lambda}$$

$$\frac{\partial \mathscr{L}}{\partial \lambda} = 0 \Rightarrow \sum_{\tau} w_\tau^{*\,2} = 1$$

or, simplifying:

$$w_\tau^* = \frac{c_\tau^a}{\sqrt{\sum_{\tau} c_\tau^{a2}}}$$

That is, the optimal weights are proportional to the autocorrelation expected under the alternative hypothesis.

Note also that, given a set of weights, we can recover the implicit alternative of a test, which is the alternative against which the test has the greatest possible power. This can be useful in providing some intuition as to what sort of alternatives a given test will have power against. In Sections 3.1 and 3.3, we will examine the implicit alternatives of variance ratio and long horizon regression test statistics.

Finally, the power of a weighted autocorrelogram test against a specified alternative can be summarized by the parameter

$$\cos^2\Psi = \frac{\left(\sum_\tau w_\tau c_\tau^a\right)^2}{\left(\sum_\tau w_\tau^2\right)\left(\sum_\tau c_\tau^{a2}\right)} \tag{2.16}$$

Using this parameter, the NCP as given in Eq. (2.15) can be written as:

$$\mathrm{NCP} = \frac{\alpha^4\left(\sum_\tau c_\tau^{a2}\right)}{\sigma_u^4}\cos^2\Psi$$

The geometric interpretation of this test statistic is explored in the next section (Section 2.3). From this equation it is clear that when the value of $\cos^2\Psi$ is 1, the test will be an optimal test, and when the value is zero, there will be no power against the alternative[5], as will be explained in more detail in Section 2.3.

2.3. A geometric interpretation of the weighted autocorrelogram test

Before we prove the general optimality of the autocorrelation test, it is useful to consider a simple geometric interpretation of the test power results from the previous section. First, note that the set of $\hat{\rho}$'s at different lags can be expressed as a vector in a p-dimensional space (where p is the number of nonzero weights in the test statistics). In this coordinate system, the component of the sample autocorrelation ($\hat{\boldsymbol{\rho}}$) vector would be $\hat{\boldsymbol{\rho}}' = (\hat{\rho}_1, \hat{\rho}_2 \ldots, \hat{\rho}_p)$. Under both the null and alternative hypotheses, the p-vector $\hat{\boldsymbol{\rho}}$ is distributed spherically, that is

$$E[(\hat{\boldsymbol{\rho}} - \mathbf{E}[\hat{\boldsymbol{\rho}}])(\hat{\boldsymbol{\rho}} - \mathbf{E}[\hat{\boldsymbol{\rho}}])'] = \frac{1}{T}\mathbf{I}$$

where $\mathbf{I}$ is the $p \times p$ identity matrix. However, under the null hypothesis, it is distributed about the origin and under the local alternative hypothesis $\hat{\boldsymbol{\rho}}$, it is

[5] Where by "no power", we mean that the test has no power to discriminate between the null and alternative, or alternatively that the distribution of the test statistic is the same under the null as under the alternative hypothesis.

centered at the point $T^{-1/2}\alpha\mathbf{c}^a$, where $\mathbf{c}^a$ is the p-vector of alternative autocorrelation, i.e., $\mathbf{c}^{a\prime} = (c_1^a, c_2^a \ldots, c_p^a)$.

The test statistic $T\hat{A}^2 = T(\Sigma_\tau w_\tau \hat{\rho})^2$ is therefore the square of the length of the projection of onto the vector of weights $\mathbf{w}$, where $\mathbf{w} = (w_1, w_2 \ldots, w_p)'$, based on our restriction that the length of $\mathbf{w}$ is 1. The length of this projection will be normally distributed as in Eq. (2.14), with mean $\frac{\alpha^2}{\sigma_u^2\sqrt{T}}\Sigma_\tau w_\tau c_\tau^a$. Rewriting this in terms of the vector we have defined. We have:

$$\sqrt{T}\,\hat{A} \overset{\text{asy}}{\sim} \mathcal{N}\left(\frac{\alpha^2|\mathbf{c}^a|}{\sigma_u^2}\cos\Psi, 1\right) \tag{2.17}$$

where Ψ is the angle between $\mathbf{w}$ and $\mathbf{c}^a$, as is illustrated in Fig. 2 and $|\mathbf{c}^a|$ denotes the length of the vector $\mathbf{c}^a$. Again, the test statistic $T\hat{A}^2$ will be noncentral χ_1^2 distributed with NCP $= ((\alpha^4|\mathbf{c}^a|^2)/\sigma_u^4)\cos^2\Psi$. Thus, to maximize the NCP, we want the vector of weights to point in the same direction as the vector of expected autocorrelations, as this results in a Ψ of zero and the maximum achievable value of $\cos^2\Psi$. On the other hand, if $\Psi = (\pi/2)$, then we are looking in a direction perpendicular to that in which we expect to see deviations, and the test will have no power.

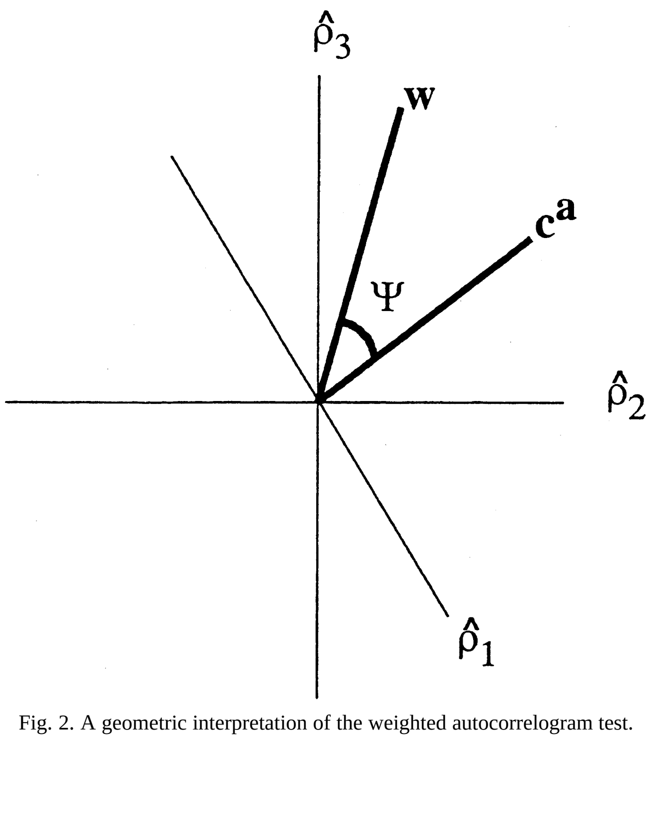

Fig. 2. A geometric interpretation of the weighted autocorrelogram test.

2.4. Proof of optimality for a general class of functions of autocorrelations

So far, we have only shown that this test is optimal among the class of test which are *linear* functions of the set of sample autocorrelations. We now show that this result holds for all functions of the p-vector of sample autocorrelations. First note any test $T(\hat{\boldsymbol{\rho}})$: $\mathbb{R}^p \rightarrow \{\text{accept } H_0, \text{ reject } H_0\}$ is a mapping from the vector of autocorrelation to a binary choice variable. Therefore, we can describe the test by the rejection region $\boldsymbol{\Omega} \subset \mathbb{R}^n$, which is the set of autocorrelation vectors $\hat{\boldsymbol{\rho}} \in \mathbb{R}^n$, which are mapped into *reject*.

Specifying the globally optimal test is equivalent to specifying the rejection region $\boldsymbol{\Omega}$ such that the probability of Type I error is minimized. Letting $f^{\{n,a\}}(\cdot)$: $\mathbb{R}^n \rightarrow \mathbb{R}$ denote the probability density functions under the null and the alternative and $\overline{\boldsymbol{\Omega}}$ denote the complement of $\boldsymbol{\Omega}$ or the acceptance region, this optimization problem can be written as:

$$\max_{\boldsymbol{\Omega}} \int_{\boldsymbol{\Omega}} f^a(\hat{\boldsymbol{\rho}}) \mathrm{d}\boldsymbol{\rho} \quad \text{such that} \quad \int_{\overline{\boldsymbol{\Omega}}} f^n(\hat{\boldsymbol{\rho}}) \mathrm{d}\boldsymbol{\rho} = \alpha$$

Differentiating the Lagrangian yields a first-order condition for maximization: that on the boundary of the region, which we denote by $\zeta \subset \mathbb{R}^{n-1}$, the ratio of the density functions under the alternative is a constant:

$$\left. \frac{f^a(\hat{\boldsymbol{\rho}})}{f^n(\hat{\boldsymbol{\rho}})} \right|_{\hat{\boldsymbol{\rho}} \in \zeta} = \lambda \tag{2.18}$$

To prove the optimality of the linear weighted autocorrelation test, we need to show that the manifold ζ is defined by $\mathbf{w}'\hat{\boldsymbol{\rho}} = \lambda \forall \hat{\boldsymbol{\rho}} \in \zeta$ for some $\mathbf{w}$.

To show this, we note that under the assumptions given in Eqs. (2.1)–(2.5), the autocorrelation vector $\hat{\boldsymbol{\rho}}$ is asymptotically distributed multivariate normal with a variance–covariance matrix equal to $\sigma^2\mathbf{I}$, and that, therefore, the distributions under the null and alternative are given by:

$$f^n(\hat{\boldsymbol{\rho}}) = \frac{1}{\sqrt{(2\pi)^n \sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \hat{\boldsymbol{\rho}}' \mathbf{I} \hat{\boldsymbol{\rho}} \right)$$

and

$$f^a(\hat{\boldsymbol{\rho}}) = \frac{1}{\sqrt{(2\pi)^n \sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}^a)' \mathbf{I} (\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}^a) \right)$$

where $\boldsymbol{\rho}^a$ is the vector of autocorrelations under the alternative hypothesis. Now define the idempotent matrix $\mathbf{M}$ as

$$\mathbf{M} = \left(\mathbf{I} - \frac{\boldsymbol{\rho}^a \boldsymbol{\rho}^{a'}}{(\boldsymbol{\rho}^{a'} \boldsymbol{\rho}^a)} \right)$$

With this we can write the log of the ratio of the alternative and null probability density functions as:

$$-2\sigma^2 \log\left(\frac{f^a(\hat{\boldsymbol{\rho}})}{f^n(\hat{\boldsymbol{\rho}})}\right) = (\hat{\boldsymbol{\rho}} - \rho^a)'\mathbf{M}'\mathbf{M}(\hat{\boldsymbol{\rho}} - \rho^a) + (\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}^a)'\left(\frac{\boldsymbol{\rho}^a\boldsymbol{\rho}^{a'}}{(\boldsymbol{\rho}^{a'}\boldsymbol{\rho}^a)}\right)(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}^a) + \hat{\boldsymbol{\rho}}'\mathbf{M}'\mathbf{M}\hat{\boldsymbol{\rho}} + \hat{\boldsymbol{\rho}}'\left(\frac{\boldsymbol{\rho}^a\boldsymbol{\rho}^{a'}}{(\boldsymbol{\rho}^{a'}\boldsymbol{\rho}^a)}\right)\hat{\boldsymbol{\rho}}$$

which after some simplification, becomes:

$$-2\sigma^2 \log\left(\frac{f^a(\hat{\boldsymbol{\rho}})}{f^n(\hat{\boldsymbol{\rho}})}\right) = 1 - 2\left(\frac{\hat{\boldsymbol{\rho}}'\boldsymbol{\rho}_\alpha}{\boldsymbol{\rho}'_\alpha\boldsymbol{\rho}_\alpha}\right)$$

We want to find the value $\hat{\boldsymbol{\rho}}$ of which makes this equal to λ. The value of $\hat{\boldsymbol{\rho}}$ that satisfies this restriction is:

$$\hat{\boldsymbol{\rho}} = \left(\frac{1-\lambda}{2}\right)\boldsymbol{\rho}_\alpha$$

Since this restriction is equivalent to the linear restriction derived earlier, this means that the linear restriction is optimal.

2.5. Other forms of the optimal test

2.5.1. The spectral domain: an optimal weighted periodigram test

Durlauf (1991) proposes a spectral based method of assessing whether a time series is a martingale. Basically, his method involves looking at the periodigram of the first differences of the series: under the null hypothesis that the series is a random walk, the expectation of the spectral density should be everywhere equal to $(\sigma_x(0))/2\pi$. Thus, asymptotically, the periodigram should be iid with mean $(\sigma_x(0))/2\pi$, and based on this the expectation of the function

$$\Gamma(\lambda) = \int_0^\lambda \left(I_T(\omega) - \frac{\sigma_x(0)}{2\pi}\right)\mathrm{d}\omega$$

is zero for all λ under the null hypothesis. $\Gamma(\lambda)$ is the "cumulated periodigram." By definition, it will be equal to zero at $\lambda = 0$ and at $\lambda = \pi$, and asymptotically, it obeys a Brownian bridge process on [0, 1] under the null hypothesis. Durlauf also suggests that if " . . . a researcher believes that the alternative to the martingale model is a long-run mean reversion, maximizing test power might dictate an examination of the low frequencies." In this section, we show how Durlauf's intuition can be formalized, and how an optimal test in the spectral domain can be constructed.

We show that since the periodigram can be thought of as just a representation of the autocorrelogram in another basis, the same intuition will apply here: the researcher should apply weights to the periodigram estimates which are proportional to the expected periodigram under the alternative.

The periodigram estimate of the spectral density is given by:

$$I_T(\omega) = \frac{1}{2\pi} \sum_{j=-(T-1)}^{T-1} \hat{\sigma}_x(j) e^{-ij\omega}$$

where $\hat{\sigma}_x(j)$ denotes the sample autocovariogram at lag j. Since $\hat{\sigma}_x(j) = {}_x(-j)$, this can be rewritten as:

$$I_T(\omega) = \frac{1}{2\pi} \left(\sum_{j=1}^{T-1} \hat{\sigma}_x(j)(e^{-ij\omega} + e^{ij\omega}) + \hat{\sigma}_x(0) \right)$$

Consider the following modified spectrum:

$$I'_T(\omega) = \frac{I_T(\omega)}{\hat{\sigma}_x(0)} - \frac{1}{2\pi} = \frac{1}{2\pi} \sum_{j=1}^{T-1} \hat{\rho}_x(j)(e^{-ij\omega} + e^{ij\omega})$$

If we define the quantity:

$$f(j,\omega) = (e^{ij\omega} + e^{-ij\omega}) = 2\cos(j\omega)$$

we see that the modified spectrum is given by:

$$I'_T(\omega) = \frac{1}{2\pi} \sum_{j=1}^{T-1} f(j,\omega)\hat{\rho}_x(j)$$

For $\omega_k = (2k\pi)/T,\ k \in \{1,\ T-1\}$, $f(\cdot)$ has the following properties:

$$E\left[\sum_{j=0}^{T-1} f(j,\omega_k) f(j,\omega_l)\right] = \begin{cases} 0 & k \neq l \\ 2T & k = l \end{cases}$$

Using this property and the fact that, asymptotically,

$$\sqrt{T}\,\hat{\boldsymbol{\rho}} \overset{\text{asy}}{\sim} \mathcal{N}(0,\mathbf{I})$$

we have that:

$$E[I'_T(\omega_k)] = 0 \forall k \geq 1$$

$$E[I'_T(\omega_k) I'_T(\omega_l)] = \begin{cases} 0 & k \neq l \\ \dfrac{1}{2\pi^2} & k = l \end{cases}$$

In other words, the modified periodigram at frequencies $\omega_k = (2k\pi)/T,\ k \in \{1,\ T-1\}$ is equivalent to the autocorrelogram in the sense that it is asymptotically mean zero and serially uncorrelated.

As an intuitive way of seeing this result, recall that, asymptotically, the vector of p autocorrelations is spherically distributed in p-dimensional space. Fourier transforming the sample autocorrelations to generate the spectrum is geometrically just transforming the vector of autocorrelations into another orthonormal basis; in this new basis, the vector must still be spherically distributed. Thus, we see that the basis of periodigram estimates has the same attractive properties as the autocorrelation basis and that, in fact, we can construct a weighted periodigram test which will have the same optimality properties as the weighted autocorrelogram test. Just as for the weighted autocorrelogram test, the weights of the optimal test should be proportional to the expected periodigram value under the alternative hypothesis.

2.5.2. An optimal instrumental variables test

We show in this subsection that another expression of the optimal test is a regression in which the dependent variable is a one-period return and the independent variable is the linear combination of past returns which is the optimal predictor of the dependent variable, given that the alternative hypothesis is true.[6]

Since the orthogonality condition is based on the characteristic that under the null hypothesis returns are not predictable using past returns, intuitively it seems that the most powerful instrumental variables test for a given alternative would be that for which the instrument was chosen to give the *greatest* possible predictive power under the alternative. That is, the optimal dependent variable should be $E[r_t \mid \mathbf{\Omega}_{t-1}]$, where $\mathbf{\Omega}_{t-1}$ is the set of all past returns. We now demonstrate that this intuition is correct. We do this by showing that an instrumental variables test using the $E[r_t \mid \mathbf{\Omega}_{t-1}]$ as the instrument is equivalent to the optimal weighted autocorrelation test.

The best forecast of r_t given the set of past returns $\mathbf{\Omega}_{t-1}$ will be given by the projection of $\mathbf{\Omega}_{t-1}$ onto r_t, which can be determined in a regression framework, that is

$$r_t = \beta \mathbf{x}_t + e_t$$

where

$$\mathbf{x}_t = \begin{pmatrix} r_{t-1} \\ r_{t-2} \\ \vdots \end{pmatrix}$$

[6] It has been noted by Hodrick (1992) that we can write any linear orthogonality condition involving returns in this way. The test of the above orthogonality condition is equivalent to either: (1) a test of whether a weighted average of future returns given by $\sum_{s=1}^{S} w_s r_{t+s}$ is predictable using the returns r_t; or to (2) a test of whether a weighted average of future returns $\sum_{s=0}^{S} w'_s r_{t+s}$ is predictable using the instrument $\sum_{r=1}^{R} w''_r r_{t-r}$, where the weights obey $\sum_{\tau=-\infty}^{\infty} w'_s w''_{\tau-r} = w_\tau$,and where the weights are defined in this equation so that $w'_s = 0$ for $s < 0$ and $s > S$, and $w''_r = 0$ for $r < 1$ and $r > R$. These tests are all precisely equivalent to the weighted autocorrelogram test if the sample moment variance is calculated under the null hypothesis and using only the single period variance. If the sample moment variance is calculated in some other way, then the tests will still be asymptotically equivalent.

We can use the OSL estimator of β here since under the local alternative the residuals will be uncorrelated. Therefore,

$$\hat{\beta} = (x'x)^{-1}(x'r) = \sum_t (x'x)^{-1} \begin{pmatrix} r_{t-1} r_t \\ r_{t-2} r_t \\ \vdots \end{pmatrix}$$

Given the local-alternative assumption, we have that $(x'x) \approx T\sigma_\epsilon{}^2\mathbf{I}$ and therefore that the project coefficients are

$$\beta = \frac{1}{\sigma_u{}^2} \begin{pmatrix} c_1^\alpha \\ c_2^\alpha \\ \vdots \end{pmatrix}$$

The regression of the single period return on the optimal predictor of this return under the alternative is therefore just a test of whether:

$$E(r_t \cdot \beta' \mathbf{x}_t) = \sum_\tau c_\tau^\alpha E(r_t \cdot r_{t-\tau}) = 0$$

is zero. This is of course the same as the optional weighted autocorrelation test.

3. The power of standard test for mean reversion

We now apply the method developed in the last section to analyze three standard mean reversion tests: the long horizon regression, the modified long horizon regression, and the variance ratio test. We show that these are asymptotically equivalent to weighted autocorrelation tests, calculated the vector of weight implicit in each test, and discuss the implicit alternative of each of the tests. In Section 3.4, we evaluate the power of these relative to an optimal test using Monte-Carlo methods.

3.1. The long horizon regression

Long horizon return regressions were used by Hansen and Hodrick (1980) to study forward rate predictions of exchange rate movements and later by Fama and French (1988a) to investigate autocorrection in stock returns. The intuition behind using a long horizon regression was that such a test could capture behavior such as the *long swings* proposed by Summers (1986) because, in aggregating returns, the price movements due to the "predictable" long swings would be aggregated, while the white noise components would be averaged out.

Consider the OLS regression coefficient $(\hat{\beta})$ for the regression

$$r(t,t+\tau) = \alpha(\tau) + \beta(\tau) r(t-\tau,t) + \epsilon(t,t+\tau)$$

where $r\ (t,t+\tau)$ represents the stock's return from t to $t+\tau$. The consistent OLS estimator of $\beta(\tau)$ is given by

$$\hat{\beta}=\frac{\hat{\text{cov}}(r(t,t+\tau),r(t-\tau,t))}{\hat{\sigma}^2(r(t-\tau,t))} \tag{3.1}$$

We can use the linearity of the covariance operator to write the OLS regression in Eq. (3.1) as:

$$\hat{\beta}=\frac{\sum_{s=0}^{2r}\min(s,2\tau-s)\hat{\text{cov}}(r_t,r_{t+s})}{\hat{\sigma}^2(r(t-\tau,t))}$$

Because overlapping observations are used, the residuals of the regressions will be correlated and the OLS standard error cannot be used. To compute the standard error of $\hat{\beta}$, Hansen and Hodrick (1980) propose estimating residual autocorrelations at all lags up to the return horizon (i.e., up to $\tau-1$ months), and then calculating the standard error using a weighted sum of these autocorrelations. Richardson and Smith (1991) propose calculating the variance–covariance matrix for the residuals assuming the null-hypothesis is true. Both methods result in consistent estimation of the residual variance–covariance matrix $\boldsymbol{\Omega}$ under the local alternative.[7] Given this, a consistent estimator of the variance of $(\hat{\beta}-\beta)$ will be:

$$E\left[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\right]=(X'X)^{-1}X'\hat{\boldsymbol{\Omega}}X(X'X)^{-1}$$

where $\hat{\boldsymbol{\Omega}}$ is the variance–covariance matrix of the residuals which, under the null hypothesis, is a block diagonal matrix where $\boldsymbol{\Omega}_{i,j}=0$ for $|i-j|\geq\tau$.

Under the local alternative we then have that:

$$\boldsymbol{\Omega}_{i,j}\overset{\text{asy}}{\rightarrow}\sigma_0^2\max(\tau-|i-j|,0)$$

$$(X'X)^{-1}X'\hat{\boldsymbol{\Omega}}X(X'X)^{-1}\overset{\text{asy}}{\rightarrow}\sigma_0^2\left(\sum_{s=0}^{2\tau}\min(s,2\tau-s)^2\right)$$

$$\hat{\sigma}^2(r(t-\tau,t))\overset{\text{asy}}{\rightarrow}\tau\sigma_0^2$$

$$t\overset{\text{asy}}{\rightarrow}\sum_{s=0}^{2\tau}\frac{\min(s,2\tau-s)}{\sqrt{\sum_{s=0}^{2\tau}\min(s,2\tau-s)^2}}\hat{\rho}\tau$$

[7] However, Richardson and Stock (1991) have pointed out that this estimator will have poor small sample properties when the sample size is not considerably larger than the aggregation interval.

where σ_0^2 is the single period return variance. Thus we see that the *t*-statistic is asymptotically equivalent to a weighted autocorrelation test, which has power against an MA process with lag polynomial weights as show in Fig. 3.

3.2. The modified long horizon regression (Jegadeesh, 1990)

Jegadeesh addresses the question of the power of the Fama and French regression against an AR(1) fads alternative such as that discussed in Section 3.4.1 of this paper. He looks at a generalized long horizon regression of the form:

$$R_{t,t+j} = \alpha(J,K) + \beta(J,K) R_{t-K,t} + \epsilon_t$$

and assesses the power of the test as a function of the parameters J and K, using the Geweke (1981) approximate slope coefficient as a measure of the test power. He finds that test power is maximized with $J=1$ However, he also finds that the optimal value of K is dependent on the parameterization of the fads alternative chosen in the process given in Eq. (3.3): the closer ϕ is to 1, the greater the optimal value of K.

The intuition for this result can be seen by referring to Fig. 6, which gives the autocorrelogram of returns generated by the AR(1) fads model. Under the fads alternative return, autocorrelation is negative at all lags, and is proportional to $\phi\tau$, where τ is the lag length. To maximize the power of the modified long horizon regression, we need to choose J and K such that the pattern of effective weights

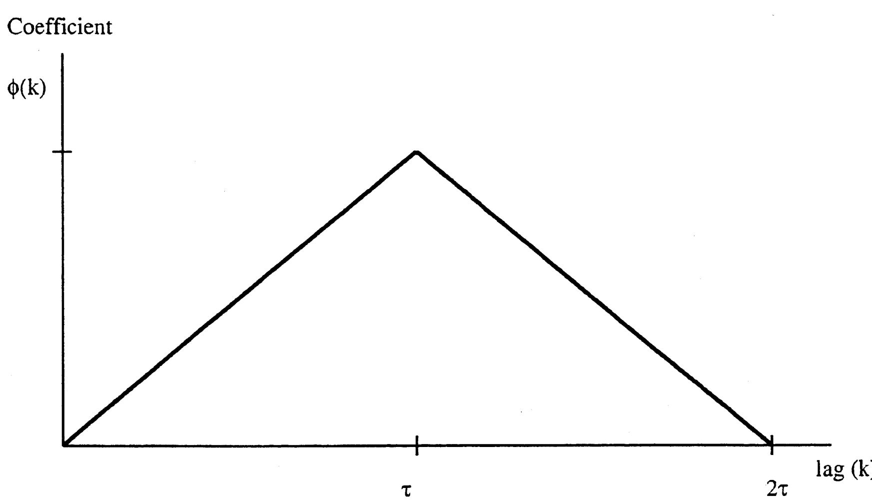

Fig. 3. Equivalent lag polynomial weights of long horizon regressions.

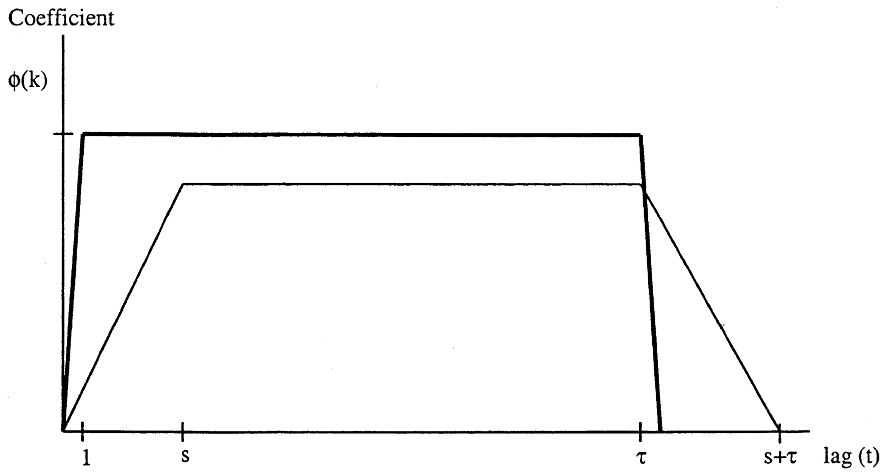

Fig. 4. Equivalent lag polynomial weights of modified long horizon regression.

will most closely resemble those in Fig. 6, that is choose w_τ to maximize $\sum_\tau w_\tau c_\tau^a$. The effective weights of the modified long horizon regression are given by:[8]

$$w_\tau = \max(0, \min(\tau, J+K-\tau, J, K))$$

Or imposing the requirement that the sine over the squares of the weights be 1, we have:

$$w_\tau = \frac{\max(0, \min(\tau, J+K-\tau, J, K))}{|J-K| \min(J,K)^2 + \frac{1}{3}\left(\min(J,K) + 2\min(J,K)^3\right)} \tag{3.2}$$

and a plot of the normalized weights for values of $J = 1$ and for $J > 1$ are given in Fig. 4.

We determine optimal weights for a set of ϕ's ranging from 0.95 to 0.99 and tabulate the results in Table 1. This is done by maximizing $\sum_\tau w_\tau \phi^\tau$ over J and K, where w_τ is taken from Eq. (3.2). Note that under the AR(1) fads alternative, c_τ^a is proportional to ϕ^τ, so this maximization will yield an asymptotically optimal test against the local AR(1) fads alternative. In addition to calculating these weights, we also calculate the optimal return horizon for a Fama and French (1988a) like regressions (where J is constrained to equal K), and calculate the value of $\sum_\tau w_\tau \phi^\tau$ for these two tests and for the optimal weighted autocorrelation test, where, for this alternative hypothesis, the optimal weights are given by

$$w_\tau = \sqrt{1-\phi^2}\, \phi^{\tau-1}$$

[8] Note that this is just a more general version of the equation for the long horizon regression weights.

Table 1
Optimal aggregation intervals and analytically calculated test power against a local alternative for optimal weighted autocorrelation test (WAC), modified long horizon regression (MLH regression) (Jegadeesh, 1990), and long horizon regression (LH regression)

ϕ	WAC test	MLH regression		LH regression	
	Power	Optimal (J, K)	Power	Optimal J	Power
0.99	7.017	(1, 125) or (125, 1)	6.334	55	5.360
0.98	4.925	(1, 62) or (62, 1)	4.445	27	3.779
0.97	3.990	(1, 41) or (41, 1)	3.601	18	3.077
0.96	3.428	(1, 31) or (31, 1)	3.095	14	2.655
0.95	3.042	(1, 25) or (25, 1)	2.746	11	2.367

There are several apparent differences between Jegadeesh's results and ours, which are explained by his use of the Geweke (1981) approximate slope coefficient as opposed to our use of a measure of local power. Note that while Jegadeesh finds that it is optimal to aggregate the independent variable and to use a single period return for the dependent variable, we find that either the dependent or independent variable may be aggregated, and the other variable should be a single period return. The reason for the differences in the results is that our test is optimal *under a local alternative*, while Jegadeesh determines optimality asymptotically, but using a *nonlocal* alternative. To make the Geweke approximate slope coefficient equivalent to a test of a local alternative, we need to calculate it in the limit as the variance of the temporary component of prices relative to the variance of the permanent component ($1/\phi$ in Jegadeesh's notation) goes to zero. When we recalculate the approximate slope coefficient given on page 5 of Jegadeesh (1990), we find both that the slope coefficient is the same whether the dependent or independent variable is aggregated, and that the optimal aggregation intervals are in agreement with those given in Table 1.

To see the reason why the Geweke approximate slope coefficient would dictate that the independent variable be aggregated, while under the local hypothesis there would be no difference, we can look at the regression coefficient for two possibilities: if the dependent variable is an n period return and the independent variable is a single period return, then the regression coefficient will be:

$$\hat{\beta} = (x'x)^{-1}(y'x) = \sum_{\tau=1}^{n} \hat{\rho}\tau$$

while the *independent* variable is an s period return and the *dependent* variable is a single period return then the regression coefficient will be:

$$\hat{\beta} = (x'x)^{-1}(y'x) = \sum_{\tau=1}^{n} \hat{\rho}_\tau \frac{\sum_t r^2(t,t+1)}{\sum_t r^2(t,t+n)}$$

Under the null hypothesis or a local alternative, the ration of variances should be $1/s$, but under a nonlocal alternative hypothesis, the ratio should be greater because of the negative autocorrelation of returns (just as the variance ratio should be less than 1 under the fads alternative). Thus, the T-statistic, which is just β divided by the standard error, is more likely to be significantly greater than zero if the independent variable is aggregated rather than the dependent.[9]

3.3. The variance ratio test

The variance ratio test has been used by Cochrane (1988) in testing for the presence of a permanent component in production data, and by Poterba and Summers (1988) and Lo and MacKinlay (1988) in testing for predictability in long and short horizon stock returns, respectively. Additionally, Lo and MacKinlay (1989) have investigated the size of the variance ratio test for both homoskedastic and heteroskedastic null hypothesis, and have calculated its power relative to the Dickey–Fuller τ-test and the Box–Pierce Q statistic for various alternatives involving simple fads processes.

The variance ratio statistic for a return horizon J is the ratio of the variance of J-period returns to J times the variance of one-period returns:

$$\mathrm{VR}(J) = \frac{\frac{1}{J}\sum_{t=1}^{T}\left(\sum_{j=1}^{J} r_{t+j} - J\bar{r}\right)^2}{\sum_{t=1}^{J}(r_t - \bar{r})^2}$$

The intuition behind the use of the variance ratio test is that if returns are uncorrelated, the variance of a return of a given horizon will be proportional to the horizon and this ratio will be 1. If, however, transitory movements in prices due to fads result in positive returns regularly being followed by negative returns, short horizon returns will exhibit a proportionally higher variance. Using a multi-layer variance ratio test, Poterba and Summers also find evidence of mean reversion at long horizons for real returns on common stocks over the 1926–1987 time period.

[9] Jegadeesh (1990) and Hodrick (1992) both point out the computational advantages of aggregating over the independent variable in that under the null hypothesis, the regression residuals will then be uncorrelated. However, as we have shown, it is straightforward to calculated standard error under the null hypothesis when a regression with an aggregated dependent variable is corrected into a weighted autocorrelation test.

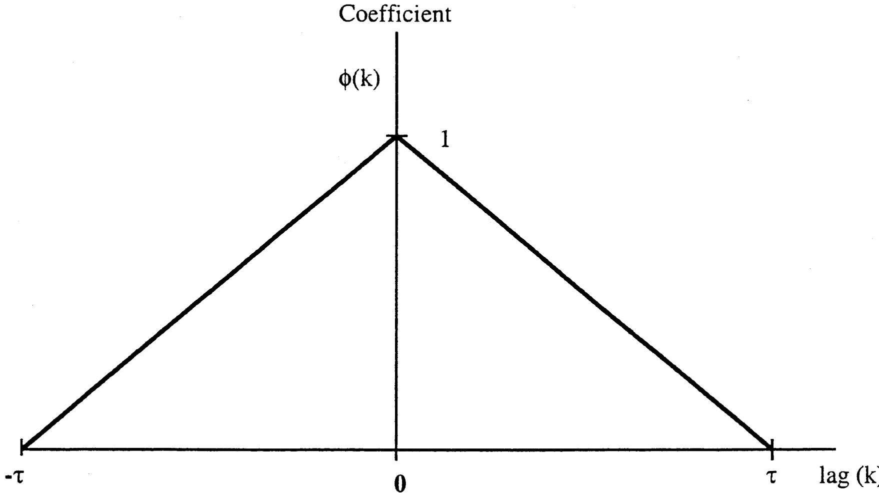

Fig. 5. Autocorrelation weights for variance-ratio equivalent tests.

As demonstrated by Cochrane (1988), the variance ratio statistic is equivalent to a test whether a weighted average of autocorrelations is equal to zero:

$$\mathrm{VR}(J) = \frac{\sum_{j=1}^{J-1}(J-j)r_t r_{t+j}}{\sum_{t=1}^{T} r_t^2} = \sum_{j=1-J}^{J-1}\left(\frac{J-|j|}{J}\right)\hat{\rho}_{\tau}$$

Thus, we see that the variance ratio test is precisely equivalent to a weighted autocorrelation test in which the pattern of weights forms an inverted triangle as in Fig. 5.

Note also that using the characteristics of the autocorrelation estimator given in Eqs. (2.7) and (2.11), one can easily show that[10]

$$\sqrt{T}\cdot \mathrm{VR}(J) \overset{\text{asy}}{\sim} \mathcal{N}\left(1, \frac{2(J-1)(2J-1)}{3}\right)$$

which is in agreement with Lo and MacKinlay's (1988) results in their Eqs. (14a) and (14b).

[10] Using the relation that $\sum_{i=1}^{j} i^2 = \frac{j(j+1)(2j+1)}{3}$.

3.4. Small sample properties of the tests

In this section, we perform a set of Monte-Carlo studies to study the small sample size and power properties of the weighted autocorrelation test. We use as an alternative hypothesis the AR(1) fads model of Summers (1986), which has been used extensively in the literature of mean-reversion test power. In Section 3.4.3, we present the power comparisons among the different test.

3.4.1. The AR(1) fads alternative hypothesis

The AR(1) fads model, suggested by Summers (1986), is important from a historical perspective in that has been widely cited and used as a basis for power comparisons (see, e.g., Fama and French, 1988a; Poterba and Summers, 1988; Lo and MacKinlay, 1989; Jegadeesh, 1990; Hodrick, 1992). Summers pointed out that if stock prices were equal to the fundamental value plus a “fads” component, this fads component might not be detected in low-order sample autocorrelations. To some extent, this observation prompted some of the long horizon regression and variance ratio tests, which were later carried out. However, the AR1 fads alternative is unsatisfying in that it is a model of overreaction rather than of rational variation in expected returns. Moreover, it implies that stock returns will be negatively autocorrelations at all lags, and the empirical evidence suggests that short horizon returns are positively autocorrelated.

The AR(1) fads model posits that observed stock prices (p_t) embody both a permanent component (p_t^*), assumed to follow a random walk a drift, and a stationary component (u_t), assumed to follow an autoregressive process of order one:

$$\begin{aligned} p_t &= p_t^* + u_t \\ p_t^* &= p_{t-1}^* + \mu + \epsilon_t, \quad \epsilon_t \sim \text{iid}\left(0, \sigma_\epsilon^2\right) \\ u_t &= \phi u_{t-1} + \nu_t, \quad 0 < \phi < 1, \nu_t \sim \text{iid}\left(0, \sigma_\nu^2\right) \end{aligned} \tag{3.3}$$

The persistence of the temporary component in determined by ϕ, while the share of the total variance due to the temporary component, γ, is defined by

$$\gamma \equiv \frac{2\sigma_\nu^2}{\sigma_\epsilon^2(1+\phi) + 2\sigma_\nu^2} \tag{3.4}$$

If ϵ_t and ν_t are independent, then the AR(1) fads model implies that demeaned returns $r_t \equiv \Delta p_t - \mu$, follow and ARMA(1,1) process $(1-\phi L)r_t = (1+\theta L)w_t$ where:

$$\theta = \frac{\left(-(1+\phi^2)\sigma_\epsilon^2 - 2\sigma_\nu^2 + (1-\phi)\left[4\frac{\sigma_\nu^2}{\sigma_\epsilon^2} + (1+\phi)^2\right]^{\frac{1}{2}}\sigma_\epsilon^2\right)}{2\left(\sigma_\nu^2 + \phi\sigma_\epsilon^2\right)}$$

and $\{w_t\}$ is an uncorrelated sequence of errors with $\sigma_w^2 = -(\phi + \sigma_\upsilon^2)/\theta$.

Under this hypothesis, the autocavariogram and autocorrelogram for the return generating process are given by

$$c_\tau = \begin{cases} \sigma_\nu^2\left(\dfrac{2}{1+\phi}\right) + \sigma_\epsilon^2 & \text{for } \tau = 0 \\ -\sigma_\nu^2\left(\dfrac{1-\phi}{1+\phi}\right)\phi^{\tau-1} & \text{for } \tau \geq 1 \end{cases}$$

$$\rho_\tau^\mu = \frac{c_\tau^\mu}{c_0^\mu} = -\left[\frac{\sigma_\nu^2(1-\phi)}{\sigma_\epsilon^2(1+\phi) + 2\sigma_\nu^2}\right]\phi^{\tau-1} \quad \text{for } \tau \geq 1 \tag{3.5}$$

Plots of the autocorrelation for several different values of ϕ are provided as Fig. 6. From this figure, it is seen that the value of ϕ controls the degree of persistence of the temporary shocks: a value of ϕ closer to 1 makes the shocks more persistent. Note that a value of 1 would make the shocks permanent.

3.4.2. Analytical calculation of test power

Table 1 presents the analytically calculated test power values against the AR(1) fads alternative for the weighted autocorrelation test, the modified long horizon regression (Jegadeesh, 1990), and the standard long horizon regression. This analytical comparison shows that, for all values of ϕ, the optimal test is most powerful, and that the modified long horizon regression is superior to the standard

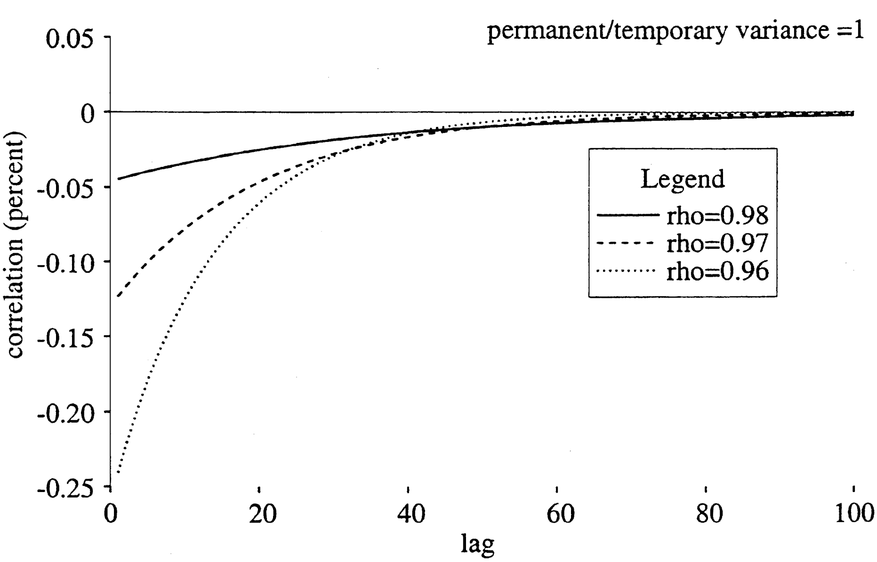

Fig. 6. Autocorrelogram of simulated AR(1) fads model.

long horizon regression. We next confirm these analytical results for small samples using a Monte-Carlo experiment.

3.4.3. Monte-Carlo results

We investigate the power of four tests against this alternative: the long horizon regression with Hansen and Hodrick (1980) calculated standard errors; the long horizon regression with standard errors calculated under the null hypothesis (Richardson and Stock, 1989; Richardson and Smith, 1991); the modified long horizon regression (Jegadeesh, 1990); and a weighted autocorrelation test with weights given by:

$$w_i = \frac{\sqrt{1-\phi^2}}{\sqrt{1-\phi^{360}}}\phi^{i-1} \quad \textit{for} \;\; i=1,\ldots,\ 180 \tag{3.6}$$

which is optimal against the AR(1) fads alternative. All of the test statistics are corrected for small sample bias using analytical corrections.

We use Monte-Carlo methods to calculate the size and power of the four tests. First, we calculate the empirical size of the tests. We simulate 60,000 returns series and compile the resulting test statistics to determine an empirical probability distribution of the test statistic under the null, and from this distribution determine the cutoff level for a size of 5%. We then simulate data under various parameterizations of the alternative hypothesis, and again compile the test statistics into an

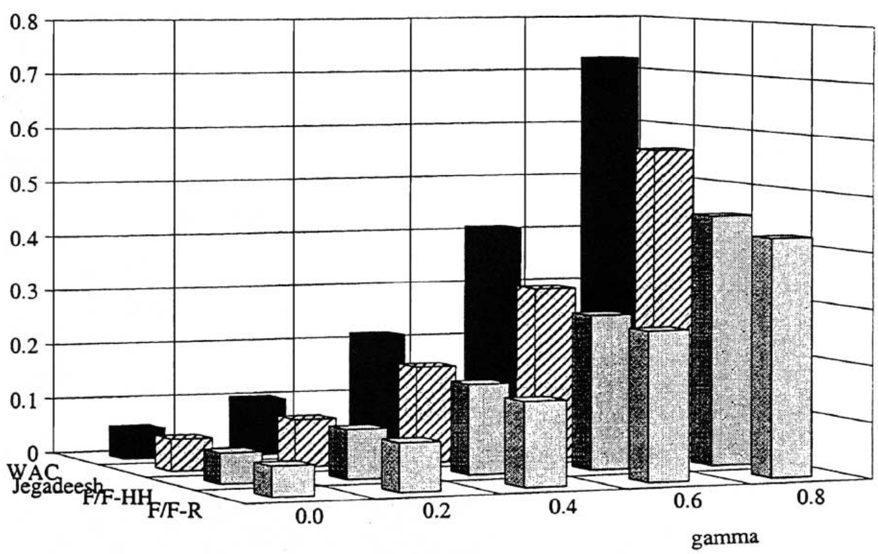

Fig. 7. Power comparison of weighted autocorrelation test, Jegadeesh regression, and Fama/French regression as a function of γ, for $\phi = 0.95$.

empirical distribution. The reported empirical power is the fraction of the test statistics which fall outside of the empirical cutoff value determined in the size analysis.

We report Monte-Carlo determined power levels for ϕ's of 0.95 and 0.98, and for γ's of 0.2, 0.4, 0.6 and 0.8. γ, which is given in Eq. (3.4), is the proportion of the variance due to the temporary component.

Figs. 7 and 8 give the power levels for the four tests, for the set of γ's, for a significance level of 0.05, for a persistence parameters ϕ of 0.95 and 0.98, respectively. All power levels presented here are calculated using 20,000 simulated returns series. In this and in Fig. 8, the return horizon used is that which gives the highest power against the particular alternative being evaluated.

In this figures, note that the power level for $\gamma = 0.0$ is approximately 0.05, which is to be expected since the alternative with $\gamma = 0.0$ is equivalent to the null, and we have set the critical value so that the null will be falsely rejected 5% of the time. As γ increases, the power increases for all tests, but more quickly for the weighted autocorrelation test. For these parameters, the weighted autocorrelation test is most powerful, followed by the modified long horizon regression (labeled "Jegadeesh"), followed by the long horizon regression using Hansen and Hodrick standard errors (labeled "FF-HH"), followed by the long horizon regression using analytical standard errors (labeled FF-R). Except for the relation between the FF-HH test and the FF-R test, about which our asymptotic theory makes no prediction, this is in agreement with the predictions as given in Table 1.

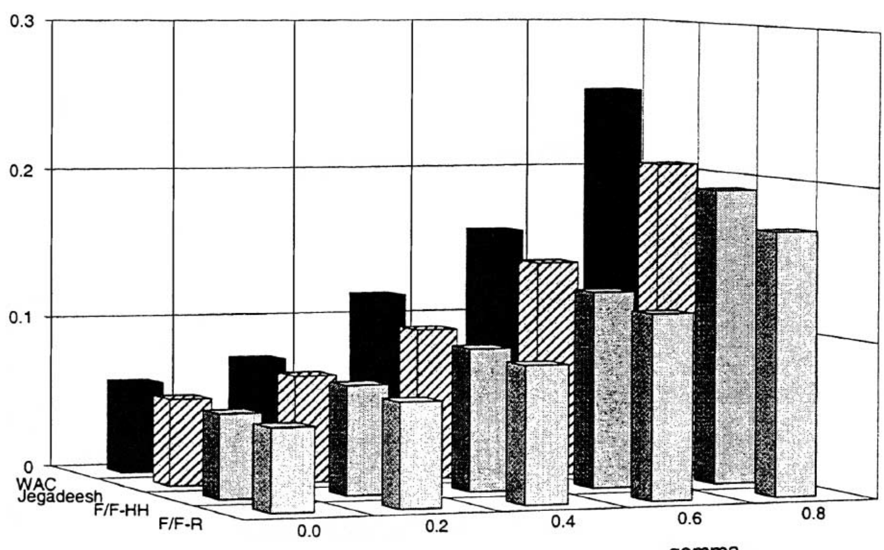

Fig. 8. Power comparison of weighted autocorrelation test, Jegadeesh regression, and Fama/French regression as a function of γ, for $\phi = 0.98$.

Thus, the asymptotic predictions appear to be generally verified, although the differences in the two long horizon regressions show that the asymptotic comparison is not a perfect predictor of test power in small samples. This issue is explored further in Section 5.

4. Joint tests of weighted sums of autocorrelations

A number of studies including Fama and French (1988b) and Poterba and Summers (1988) have looked at a set of mean-reversion test statistics, and calculated significance levels based on the most significant statistic. As Richardson (1993) points out, the statistical significance of the overall test should be evaluated by *jointly* testing whether all coefficients are equal to zero. Richardson and Smith (1991) propose a χ^2 joint test embedded in the GMM framework of Hansen (1982) and use this joint test to evaluate the significance of Fama and French's and Poterba and Summer's results. Jegadeesh (1990) uses the test to evaluate the joint significance of his modified long horizon regressions.

The intuition behind the results in this section is best expressed in terms of the geometric interpretation given in Section 2.3. There, we showed that a weighted autocorrelation test statistic can be interpreted as the length of the projection of the vector of autocorrelations onto a vector of weights. Here we show that a χ^2 test of whether a set of n weighted autocorrelation test statistics is zero is equivalent to a test of whether the projection of the autocorrelation vector onto the n-dimensional subspace spanned by the n weight vectors has a length zero.

An important implication of these results is that the χ^2 joint test may lack power against the very alternatives the econometrician is interested in. For example, if he wishes to look for mean reversion in stock price data, without having precise knowledge of the persistence of the mean reversion, he might elect to run a set of n long horizon regressions and then test their significance using a χ^2 test. Even if the individual regressions have considerable power against the alternative, the power of the joint test may be quite low. The reason for this is that joint test looks for deviations from the null in the entire n dimensional subspace, even if the alternative suggests deviations only in a particular direction within the subspace.

We perform the analysis in this section within a GMM framework. All of the tests we are concerned with are tests of whether returns are orthogonal to past returns, and GMM framework is a general way of analyzing this type of restriction. We establish the equivalence among the three tests: (1) a χ^2 or Wald joint test of M regression coefficients or variance ratios, as in Richardson and Smith (1991); (2) a GMM-test of a set of M overidentifying restrictions; and (3) a χ^2 joint test of a set of M weighted autocorrelogram tests. For expositional reasons, we show this equivalence in terms of long horizon regressions, though the method is applicable to any joint test of any regressions, moment restrictions, or variance ratio tests.

4.1. The long horizon regression in a GMM framework

A *single* long horizon regression can be written in a generalized instrumental variables framework by noting that the OLS estimates of $\alpha(\tau)$ and $\beta(\tau)$ are based on the moment restrictions that $E[\epsilon(t, t+\tau)]=0$ and $E[\epsilon(t, t+\tau)r(t, t+\tau)]=0$: that is, on the restriction that the regression residual $\epsilon(t, t+\tau)=r(t, t+\tau)-\alpha(\tau)-\beta(\tau)r(t-\tau, t)$ be orthogonal to the instruments of 1 and $r(t, t+\tau)$. These two orthogonality conditions are used to estimate the system of two unknowns. In the GMM framework, we can represent these restrictions in the following way:

$$\mathbf{g}_t(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} g_t(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} r(t,t+\tau)-\alpha(\tau)-\beta(\tau)r(t-\tau,t) \\ [r(t,t+\tau)-\alpha(\tau)-\beta(\tau)r(t-\tau,t)]r(t-\tau,t)\end{pmatrix} \tag{4.1}$$

The GMM estimator of $\boldsymbol{\theta}=(\alpha(\tau)\beta(\tau))'$ then minimizes the distance of the sample moment vector $\mathbf{g}_t(\boldsymbol{\theta})$ from zero. This is done by minimizing $\mathbf{g}_t(\boldsymbol{\theta})'\mathbf{W}_T\mathbf{g}_t(\boldsymbol{\theta})$, where $\mathbf{W}_T$ is some weighing matrix. For this just-identified system, the choice of weighing matrix is unimportant since for some choice of $\boldsymbol{\theta}$, every element of the $\mathbf{g}_t(\boldsymbol{\theta})$ vector will be equal to zero.

In general, there are more moment restrictions than variables to be estimated. Hansen (1982) shows that in this case, the optimal weighing matrix is the inverse of the variance–covariance matrix of $\mathbf{g}_T(\boldsymbol{\theta})$ evaluated at the true value of $\boldsymbol{\theta}$, $\boldsymbol{\theta}_0$:

$$\mathbf{W}^* = \mathbf{S}_0^{-1} = \left(\sum_{j=-\infty}^{\infty} E\left[\mathbf{g}_t(\boldsymbol{\theta}_0)\mathbf{g}_{t-j}(\boldsymbol{\theta}_0)'\right]\right)^{-1} \tag{4.2}$$

and that, given this weighing matrix, $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically normally distributed:

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\overset{\text{asy}}{\sim}\mathcal{N}\left(0,\left[\mathbf{D}_0'\mathbf{S}_0^{-1}\mathbf{D}_0\right]^{-1}\right) \tag{4.3}$$

where

$$\mathbf{D}_0 = E\left[\frac{\partial \mathbf{g}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right] \tag{4.4}$$

This representation can be extended to jointly estimate a set of regression coefficients. This results in the following set of just-identified moment equations:

$$\mathbf{g}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_t(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} [(r(t,t+j)-\beta(j)r(t-j,t)]r(t-j,t) \\ \vdots \\ [r(t,t+k)-\beta(k)r(t-k,t)]r(t-k,t)\end{pmatrix} \tag{4.5}$$

where now the returns are demeaned and the intercept terms and corresponding moment restrictions have been removed from the system.[11] Note that, again, the estimates of the β's will be identical to the OLS regression estimates since the system is just identified. The $\mathbf{D}_0$ and $\mathbf{S}_0$ matrices can be constructed under the null hypothesis, following Richardson and Smith (1991):

$$\mathbf{D}_0 = \begin{pmatrix} j\sigma^2 & 0 \\ 0 & k\sigma^2 \end{pmatrix} \tag{4.6}$$

$$\mathbf{S}_0 = \sigma^4 \begin{pmatrix} \dfrac{j(1+2j^2)}{3} & j^2 + s(j,k) \\ j^2 + s(j,k) & \dfrac{k(1+2k^2)}{3} \end{pmatrix} \tag{4.7}$$

where σ^2 is the variance of a single period return $(E[r_t^2])$ and where

$$s(j,k) = 2\sum_{l=1}^{j-1} (j-l)\min(j,k-l)$$

The variance–covariance matrix of the vector of β estimators then obtained by matrix manipulation from Eqs. (4.6) and (4.7).

$$V(\hat{\beta}) = \left(\mathbf{D}_0'\mathbf{S}_0^{-1}\mathbf{D}_0\right)^{-1} = \begin{pmatrix} \dfrac{2j^2+1}{3j} & \dfrac{j^2+s(j,k)}{jk} \\ \dfrac{j^2+s(j,k)}{jk} & \dfrac{2k^2+1}{3k} \end{pmatrix} \tag{4.8}$$

and the Wald test that the set of β's are equal to zero is given by

$$J = T\hat{\beta}'\left[V(\hat{\beta})\right]^{-1}\hat{\beta} \tag{4.9}$$

With this method, GMM is used to estimate the set of regression β's, and then a separate Wald test is performed to determine whether the β's are jointly significantly different from zero, using $V(\hat{\beta})$ as calculated analytically in the GMM

[11] The test statistic Richardson and Smith (1991) propose is not equivalent to this test in small samples because the estimated intercept terms (α's) obtained from estimating this overidentified system are slightly different. However, since $\bar{r}$ is a consistent estimator of $\alpha(t)$ under our assumptions, the tests will be identical asymptotically. We could set up an overidentified system in which the α's were estimated using the moment restriction that $E(\tilde{\epsilon}_t \cdot 1) = 0$, but this would complicate the analysis without adding additional insight.

framework. Alternatively, one could directly impose the moment restrictions which encompass this restriction (that $\beta(\tau) = 0 \forall \tau > 0$):

$$\mathbf{g}_T = \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}'_t(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} r(t,t+j)r(t-j,t) \\ \vdots \\ r(t,t+k)r(t-k,t) \end{pmatrix} \tag{4.10}$$

Again, from Hansen (1982), a test statistic that indicates the "distance" of this model from the data is given by

$$J' = T\mathbf{g}_t(\hat{\boldsymbol{\theta}})'\mathbf{S}_0^{-1}\mathbf{g}_t(\hat{\boldsymbol{\theta}}) \tag{4.11}$$

which is asymptotically χ^2 distributed with n degrees of freedom.

Just as is done above, we can construct $\mathbf{S}_0^{-1}$ under the null hypothesis, and from Eq. (4.2), we see that when calculated under the null, $\mathbf{S}_0$ will be identical to the $\mathbf{S}_0$ given in Eq. (4.7) because under the null hypothesis, β is zero for all return horizons. Now, we define $\mathbf{S}_0^* \equiv \mathbf{S}_0/\sigma^4$, where, from Eq. (4.7), $\mathbf{S}_0^*$ is now a function for j and k only. This means that we can construct an alternative $\mathbf{S}_0$, which we denote as $\mathbf{S}_0^\dagger$, in the following way:

$$\mathbf{S}_0^\dagger = \begin{pmatrix} \frac{\sigma(j)^2}{j} & 0 \\ 0 & \frac{\sigma(k)^2}{k} \end{pmatrix}\mathbf{S}_0^*\begin{pmatrix} \frac{\sigma(j)^2}{j} & 0 \\ 0 & \frac{\sigma(k)^2}{k} \end{pmatrix}$$

where

$$\sigma^2(j) = \frac{1}{T}\sum_{t=1}^{T} r(t-j,t)^2$$

Since under the null hypothesis, $(\sigma^2(j))/j$ is a consistent estimator of σ, the one-period return variance, $\mathbf{S}_0^\dagger$ is a consistent estimator of $\sigma^4\mathbf{S}_0^* = \mathbf{S}_0$.

When $(\mathbf{S}_0^\dagger)^{-1}$ is substituted into the definition of J' in Eq. (4.11), we obtain the following expression for the test statistic:

$$J' = T\mathbf{g}'_t\left(\mathbf{S}_0^\dagger\right)^{-1}\mathbf{g}_t = T\hat{\beta}'\underbrace{\begin{pmatrix} \frac{1}{j} & 0 \\ 0 & \frac{1}{k} \end{pmatrix}(\mathbf{S}_0^*)^{-1}\begin{pmatrix} \frac{1}{j} & 0 \\ 0 & \frac{1}{k} \end{pmatrix}}_{[V(\hat{\beta})]^{-1}}\hat{\beta}$$

Comparing this equation with Eqs. (4.6)–(4.8) confirms that the Wald statistic in Eq. (4.9) and the test statistic for the overidentified GMM system given here are identical.

4.2. Asymptotic equivalence to a set of weighted autocorrelation tests

We show now that the test of the moment restrictions in Eq. (4.10) is equivalent to the test of whether a set of weighted sums of autocorrelations are zero. Using these results, we present in Section 4.3 an intuitive geometric interpretation of the joint test.

First, note that the set of moment restrictions in Eq. (4.10) are equivalent to

$$\mathbf{g}_t = \frac{1}{t}\sum_{t=1}^{t}\mathbf{g}_t(\boldsymbol{\theta}) = \frac{1}{t}\sum_{t=1}^{t}\begin{pmatrix}\sum_{s=0}^{2j}\min(s,2j-s)\,r_t r_{t+s}\\ \vdots \\ \sum_{s=0}^{2k}\min(s,2k-s)\,r_t r_{t+s}\end{pmatrix} \tag{4.12}$$

Additionally, $\mathbf{S}_0$ will again be of the same from here as in Eq. (4.7). However, since we can choose *any* consistent estimator of the single period variance in Eq. (4.7), we now choose

$$\hat{\sigma}^2 = \sigma^2(1) = \frac{1}{T}\sum_{t=1}^{T} r_t^2,$$

which is the variance calculated using one-period returns. The test of the overidentifying restrictions can now be written

$$J'' = \underbrace{T\mathbf{g}'_T\frac{1}{\hat{\sigma}^2}}_{\mathbf{g}_T^{*\prime}}(\mathbf{S}_0)^{-1}\underbrace{\frac{1}{\hat{\sigma}^2}\mathbf{g}_T}_{\mathbf{g}_T^{*}}$$

and

$$\mathbf{g}_T^* = \frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_t^*(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\sum_{s=0}^{2j}\min(s,2j-s)\left(\frac{r_t r_{t+s}}{r_t^2}\right)\\ \vdots \\ \sum_{s=0}^{2k}\min(s,2k-s)\left(\frac{r_t r_{t+s}}{r_t^2}\right)\end{pmatrix}$$

$$= \begin{pmatrix}\sum_{s=0}^{2j}\min(s,2j-s)\,\hat{\rho}_s\\ \vdots \\ \sum_{s=0}^{2k}\min(s,2k-s)\,\hat{\rho}_s\end{pmatrix}$$

so we see that the moment conditions are equivalent to restrictions that weighted sums of autocorrelations be equal to zero. Using vector notation, we can rewrite the moment conditions in Eq. (4.13) as

$$\mathbf{g}_T^* = \begin{pmatrix} \mathbf{w}(j)' \hat{\boldsymbol{\rho}} \\ \vdots \\ \mathbf{w}(k)' \hat{\boldsymbol{\rho}} \end{pmatrix} \tag{4.13}$$

where $\mathbf{w}$(j) is a k-vector whose ith element is max(0, min(s, 2k − s)), and $\hat{\boldsymbol{\rho}}$ is a k-vector whose ith element is the autocorrelation at lag i.

Note that when $\mathbf{g}_t^*$ is expressed in this way, we can apply the asymptotic relationship

$$TE = [\hat{\rho}\hat{\rho}'] = \mathbf{I}$$

to show that that $\mathbf{S}_0^*$, the variance–covariance matrix of $\mathbf{g}_t^*$, will be given by:

$$\mathbf{S}_0^* = e[\mathbf{g}_T^{*\prime}\mathbf{g}_T^*] = \begin{pmatrix} \mathbf{w}(j)' \hat{\rho}'\hat{\rho}\mathbf{w}(j) & \cdots & \mathbf{w}(j)' \hat{\rho}'\hat{\rho}\mathbf{w}(k) \\ \vdots & \ddots & \vdots \\ \mathbf{w}(k)' \hat{\rho}'\hat{\rho}\mathbf{w}(j) & \cdots & \mathbf{w}(k)' \hat{\rho}'\hat{\rho}\mathbf{w}(k) \end{pmatrix}$$

$$= \frac{1}{T}\begin{pmatrix} \mathbf{w}(j)'\mathbf{w}(j) & \cdots & \mathbf{w}(j)'\mathbf{w}(k) \\ \vdots & \ddots & \vdots \\ \mathbf{w}(k)'\mathbf{w}(j) & \cdots & \mathbf{w}(k)'\mathbf{w}(k) \end{pmatrix}$$

which extensive algebraic manipulation reveals to be identical to the $\mathbf{S}_0^* = (1/\sigma^4)\mathbf{S}_0$ given in Eq. (4.7). Thus, when we express the regressions as weighted sums of correlations, we have a more straightforward way of deriving the Richardson and Smith (1991) variance–covariance matrix as given in Eq. (4.8). Moreover, writing the test in this way leads to a simple geometric interpretation of the test power, which we provide in the next section.

4.3. Geometric interpretation of the joint test

We can gain considerable intuition into the workings of the joint test by giving the joint test power issue a geometric interpretation analogous to what was done for the single weighted autocorrelation test in Section 2.3. Again, we consider a p-dimensional space in which the set of sample autocorrelations is expressed as a vector with elements $(\hat{\rho}_1, \hat{\rho}_2 \dots, \hat{\rho}_p)$. We showed in Section 2.3 that the test statistic $TA^2 = T\sum_\tau w_\tau \hat{\rho}_\tau$, was the square of the length of the projection onto the vector of weights $\mathbf{w}$, where $\mathbf{w} = (w_1, w_2 \dots, w_p)'$, and that this statistic was χ_1^2 distributed with NCP $((\alpha^4|\mathbf{c}^a|^2)/\sigma_u^4)\cos^2\Psi$ under the alternative hypothesis, where Ψ is the angle between the vector $\mathbf{c}^a$ and the vector of weights $\mathbf{w}$. Since the alternative hypothesis is noncentral χ_1^2 distributed and the null hypothesis is central χ_1^2 distributed, we showed that the most powerful test will maximize the NCP, which is done by packing weights so that Ψ is zero or π.

We can provide a similar interpretation of the joint test. Consider again the joint test of a set of M long horizon regressions, as expressed by the set of moment restrictions in Eq. (4.13). If the test statistic

$$J'' = T\mathbf{g}_T^{*\prime}(S_0^*)^{-1}\mathbf{g}_T^* \sim \chi_M^2 \quad (4.14)$$

is rewritten as

$$J'' = T\hat{\boldsymbol{\rho}}'\mathbf{W}'\left(\frac{1}{T}\mathbf{I}\right)\mathbf{W}\hat{\boldsymbol{\rho}}$$

where

$$\mathbf{W} = \begin{pmatrix} \mathbf{w}(j)' \\ \vdots \\ \mathbf{w}(k)' \end{pmatrix}$$

We see that J'' is the square of the length of the projection of $\hat{\boldsymbol{\rho}}$ onto the M-dimensional manifold (or subspace), which is spanned by the M eigenvectors $\mathbf{w}(j)\dots\mathbf{w}(k)$.

Additionally, just as in Section 2.3, the distribution of the joint test statistic J'' is χ_M^2 distributed with NCP $((\alpha^4|\mathbf{c}^a|^2)/\sigma_u^4)\cos^2\Psi$ where Ψ is now the angle between the vector of autocovariances $\mathbf{c}^a$ and the M-dimensional manifold containing the weight vectors.

For this setting, the power of the test will depend both on the angle Ψ and on the number of restrictions M. The most powerful test will both maximize $\cos(\Psi)$ and minimize M. However, when the alternative hypothesis is known precisely, there will be some tradeoff: increasing the dimensionality of the $\mathbf{W}$ matrix may decrease the expected value of $\cos(\Psi)$, thus increasing the expected NCP of the test statistic, but it will also increase the number of degrees of freedom of the test statistics χ^2 distribution. An extreme example of this is Box–Pierce Q test, which is the joint test of whether each of the p autocorrelations is zero: for this joint test, any process is an implicit alternative, but of course it will have a very little power against any specific alternative.

Fig. 9 illustrates the problem with the method of using a χ^2 joint test of the significance of a set of tests. Suppose the econometrician was using the first three autocorrelations to investigate mean reversion, and he believes that each of the three autocorrelations are likely to have roughly equal positive values. He might then run a joint test of three weighted autocorrelation tests using the weight vectors illustrated in Fig. 9. Each of these three tests would, individually, be powerful against the alternative. But, as our analysis shows, the joint test will have considerably lower power: while the econometrician wishes to place large weight on sample autocorrelations, vectors in a narrow region of $\mathbb{R}_+^3$ he is in fact putting equal weight on deviations in any direction in $\mathbb{R}^3$.

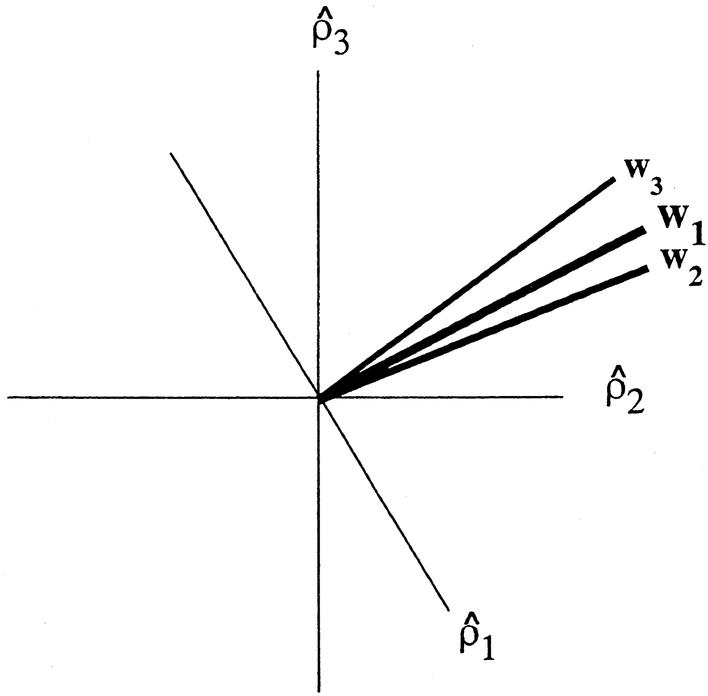

Fig. 9. A geometric illustration of the joint test problem.

Another way of stating this intuition is to say that when the number of restrictions being tested is increased, the "larger" the manifold of implicit alternatives becomes. The greater the dimensionality of this manifold, the greater the number of unreasonable alternatives the test is likely to have power against, and the lower the tests power against reasonable alternatives. We show in Section 5 that other types of joint tests can have greater power.

5. Small sample size and power of long horizon regression methods

Fama and French (1988a) perform the following long horizon regression on the CRSP, EW, VW and size decile portfolio real returns for 1926–1986.

$$R_{t,t+\tau} = \alpha(\tau) + \beta(\tau) R_{t-\tau,t} + \epsilon_t$$

The regressions are done for return horizons of 1, 2, 3, 4, 5, 6, 8 and 10 years, using monthly data, and the (consistent) OLS coefficients are calculated. A *T*-statistics is used as a test of statistical significance,

$$T(\tau) = \frac{\hat{\beta}(\tau)}{s\hat{e}_{\beta}} \tag{5.1}$$

where, because of the overlapping observations and the resulting correlated residuals, $s\hat{e}\beta$ is calculated using the Hansen and Hodrick (1980) method.

Fama and the French find that for the EW index and for size deciles 1–7, the slope coefficient for a return horizon of 4 years (48 months) is more than 2 standard errors from 0. Additionally, the slope coefficients for return horizons of 3 and 5 years are significantly different from 0 for a number of the portfolios. Fama and French conclude from this that there is evidence of mean reversion in the stock prices of small firms over the 1926–1986 period.

However, the Fama and French evidence is not unambiguous. First, Richardson (1993) challenges the statistical reliability of the Fama and French results on the ground that the test does not properly account for implicit multiple comparisons. That is, one cannot conclude from the statistical significance of the regression coefficient at a single return horizon that there is evidence of mean reversion; a joint test of significance of the coefficients at all eight return horizons must be conducted. Using this test, Richardson finds that the stationary random walk hypothesis cannot be rejected over the 1926–1986 sample period.

Second, Richardson and Stock (1989) suggest that the asymptotic standard errors used by Fama and French and by Richardson and Smith are flawed because of bad small sample properties. They suggest another asymptotic method of calculating the $\hat{\beta}$ standard errors, which is based on holding the return horizon at a constant fraction of the sample size as the length of the data series goes to infinity. They show that the J/T limiting distribution calculated under these assumptions has much better small sample properties than the conventional asymptotic distribution, and finally, they show that even the *individual* Fama and French regression coefficients ($\beta(\tau)$'s) are statistically insignificant when the significance is determined using the J/T asymptotics. They, therefore, concluded that statistical significance of the individual slope estimates that Fama and French finds is due to the poor small sample properties of the Hansen and Hodrick estimator.

We point out in this section that, while both of these critiques are well founded, their conclusions that the long horizon test statistics presented by Fama and French do not allow rejection of the null hypothesis are due to the use of a different test: Fama and French use Hansen/Hodrick calculated *T*-statistics while Richardson and Smith (1991) and Richardson and Stock use statistics calculated under the null hypothesis of no serial correlation. While these two statistics are asymptotically equivalent, their power differs in small samples, as was demonstrated in the Monte-Carlo results in Section 3.4.3.

We show in this section that if the small-sample corrected *p*-values of the Hansen/Hodrick $T(\tau)$ statistic are used instead of the *p*-values for the $\hat{\beta}(\tau)$ statistic, there is still evidence of mean reversion like. Therefore, the reversal of Fama and French's conclusion is not due to the small sample properties of the estimator, as claimed, but rather to the difference in power of the two tests.

In addition, we show that the poor small sample properties of the $\hat{\beta}(\tau)$ statistic can largely be corrected by adjusting the OLS regression coefficient, the weighting matrix and sample autocorrelations in the Hansen and Hodrick standard error calculation for small sample biases: Once these corrections are made, the small sample properties of the asymptotic statistics are greatly improved.

5.1. Calculation of the small sample distribution

We begin by deriving analytically the small-sample bias of the $\hat{\beta}(\tau)_{OLS}$, which is available on request from the author. Though tedious to derive, the intuition for the small sample bias is simply that a demeaning series induces negative serial correlation. As an extreme example, consider calculating the first-order serial correlation based on two observations: the calculated value will always be negative because once the observations are demeaned, one of them will be positive and the other negative. Next, using Monte-Carlo methods, we calculate the empirical distribution of the $\hat{\beta}(\tau)$OLS with and without the bias adjustment, for a sample of 720 points (the length of Fama and French's sample). This was done for return horizons of 12, 24, 36, 48, 60, 72, 96 and 120 periods, corresponding to the 1-, 2-, 3-, 4-, 5-, 6-, 8-, and 10-year horizons used by Fama and French and Richardson and Stock. The cumulative empirical distribution without the bias correction is plotted in Fig. 10.

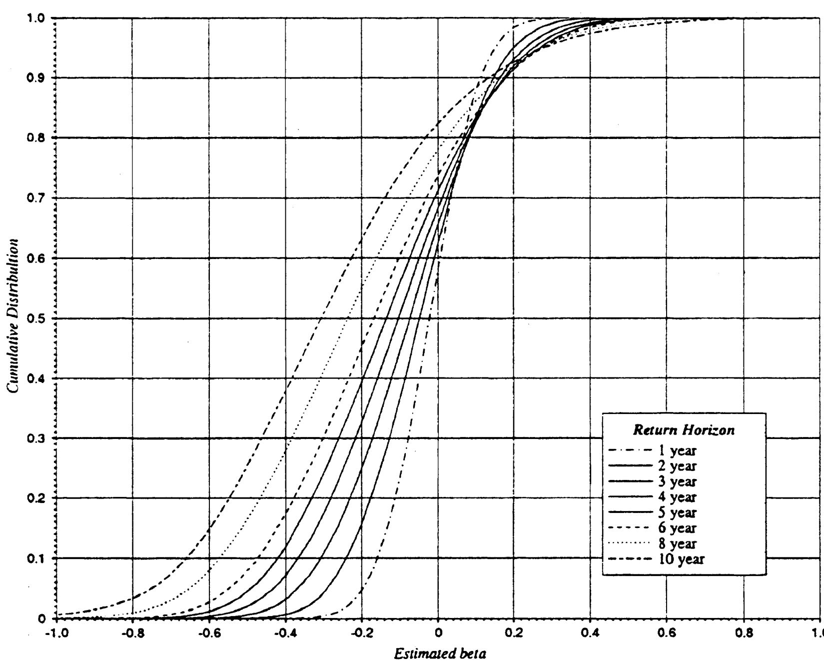

Fig. 10. Fama / French regression—cumulative distribution of Hansen / Hodrick T-statistic–Monte-Carlo results—no bias correction, 720 points, 20000 iterations.

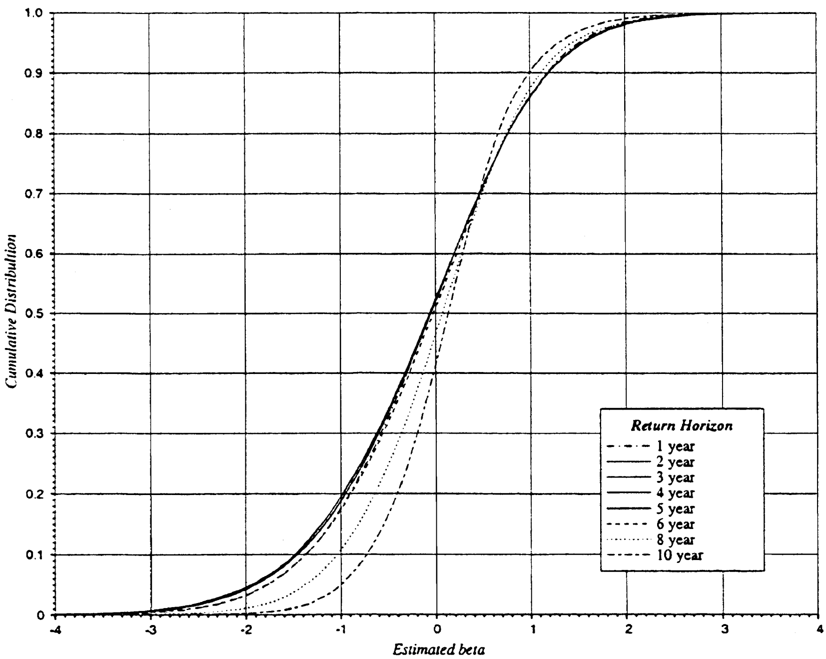

Fig. 11. Fama/French regression—cumulative distribution of Hansen/Hodrick *T*-statistic–Monte-Carlo results—bias correction, 720 points, 60000 iterations.

Fig. 10 shows that, as demonstrated by Richardson and Stock, for long return horizons, the OLS $\beta(\tau)$ estimator has a small sample distribution that is clearly not well represented by a mean-zero normal. Also, as they point out, the likelihood of negative values is quite high under the null hypothesis. However, Fig. 11 shows that the distributions of the bias-adjusted T statistics for return horizons of 1–6 years are almost identical to the asymptotic distribution, and that for the return of 8 and 10 years, the distribution is actually narrower: the probability of extreme negative values is lower than what is predicted by asymptotic theory.[12]

5.2. Empirical results

We performed the long horizon regression tests for real returns using the bias-adjusted Hansen/Hodrick T statistics. We present the regression $\hat{\beta}(\tau)$'s in

[12] The bias correction appears to be slightly too-small here, in that for all return horizons except 8 and 10 years, the mean of the distribution is somewhat negative.

Table 2
Fama and French regressions—real returns: 1926–1985—with significance levels based on small sample empirical distribution

Size decile portfolio	Bias adjusted $\hat{\beta}$'s and p-values							
	Return horizon							
	1	2	3	4	5	6	8	10
EW *	−0.05 (0.342)	−0.23 (0.077)	−0.33† (0.027)	−0.37* (0.011)	−0.35* (0.023)	−0.13 (0.254)	−0.12 (0.690)	0.31 (0.850)
1	0.02 (0.582)	−0.14 (0.197)	−0.26 (0.088)	−0.39* (0.024)	−0.34† (0.046)	−0.06 (0.397)	0.39 (0.951)	0.64 (964)
2 *	0.01 (0.558)	−0.12 (0.229)	−0.26 (0.075)	−0.42** (0.006)	−0.45** (0.003)	−0.27 (0.064)	0.01 (0.493)	0.17 (0.696)
3 *	−0.04 (0.394)	−0.17 (0.140)	−0.28 (0.051)	−0.37* (0.011)	−0.36* (0.017)	−0.18 (0.158)	−0.04 (0.386)	0.06 (0.539)
4 *	−0.02 (0.450)	−0.11 (0.251)	−0.21 (0.116)	−0.35* (0.010)	−0.38* (0.006)	−0.20 (0.111)	0.00 (0.473)	0.14 (0.696)
5	−0.05 (0.356)	−0.22 (0.084)	−0.29† (0.044)	−0.32* (0.025)	−0.32† (0.030)	−0.16 (0.205)	0.05 (0.570)	0.24 (0.790)
6	−0.05 (0.327)	−0.20 (0.099)	−0.32† (0.032)	−0.33* (0.018)	−0.29† (0.046)	−0.09 (0.310)	0.08 (0.622)	0.23 (0.763)
7	−0.06 (0.300)	−0.26† (0.048)	−0.33† (0.025)	−0.26 (0.058)	−0.20 (0.143)	−0.01 (0.482)	0.14 (0.704)	0.21 (0.722)
8	−0.06 (0.317)	−0.23 (0.071)	−0.31† (0.029)	−0.24† (0.049)	−0.18 (0.135)	0.03 (0.580)	0.17 (0.763)	0.24 (0.776)
9	−0.03 (0.403)	−0.22 (0.085)	−0.28 (0.053)	−0.15 (0.209)	−0.02 (0.488)	0.21 (0.842)	0.34 (0.878)	0.33 (0.799)
10	−0.06 (0.303)	−0.25† (0.050)	−0.29† (0.041)	−0.14 (0.230)	0.00 (0.532)	0.21 (0.848)	0.34 (0.876)	0.30 (0.790)
VW	−0.03 (0.399)	−0.21 (0.096)	−0.26 (0.072)	−0.10 (0.301)	0.06 (0.642)	0.26 (0.890)	0.36 (0.884)	0.30 (0.781)
Hansen / Hodrick T-statistics								
EW	(−0.46)	(−1.64)	(−2.29)	(−2.85)	(−2.35)	(−0.71)	(0.46)	(0.80)
1	(0.15)	(−0.98)	(−1.57)	(−2.39)	(−1.96)	(−0.30)	(1.46)	(1.45)
2	(0.08)	(−0.84)	(−1.68)	(−3.14)	(−3.42)	(−1.65)	(0.04)	(0.45)
3	(−0.32)	(−1.22)	(−1.92)	(−2.84)	(−2.51)	(−1.08)	(−0.17)	(0.18)
4	(−0.18)	(−0.76)	(−1.40)	(−2.89)	(−3.10)	(−1.31)	(0.01)	(0.44)
5	(−0.43)	(−1.59)	(−2.00)	(−2.37)	(−2.21)	(−0.88)	(0.21)	(0.64)
6	(−0.51)	(−1.47)	(−2.20)	(−2.56)	(−1.96)	(−0.53)	(0.32)	(0.58)
7	(−0.59)	(−1.93)	(−2.33)	(−1.86)	(−1.21)	(−0.08)	(0.49)	(0.49)
8	(−0.53)	(−1.70)	(−2.26)	(−1.97)	(−1.25)	(0.17)	(0.64)	(0.60)
9	(−0.30)	(−1.58)	(−1.88)	(−0.95)	(−0.08)	(0.93)	(1.02)	(0.66)
10	(−0.57)	(−1.90)	(−2.05)	(−0.88)	(0.02)	(0.94)	(1.00)	(0.63)
VW	(−0.32)	(−1.49)	(−1.69)	(−0.62)	(0.30)	(1.14)	(1.05)	(0.61)

†denotes a p-value < 0.05.
*denotes a p-value < 0.0025.
**denotes a p-value < 0.01.

the upper parts of Table 2. The Hansen/Hodrick T-statistics for each of these coefficients are given at the bottom of each table, and empirical p-values for these T-statistics are given under each coefficient. These p-values are calculated from the Monte-Carlo results plotted in Fig. 11 and are, therefore, correct in small samples.

The results here should be compared to those in Richardson and Stock's (1989) Table 4. They find that three slope coefficients are significant at the (two-sided) 5% level. In contrast, we find that 11 slope coefficients are statistically significant at the (two-sided) 5% level.

Based on the problems with the χ^2 joint test discussed above, we use a different statistic: we look at the most significant of the eight Fama French regression coefficients. Supposedly, the reason for performing a number of regressions in the first place is that since the optimal test return horizon is dependent on the parameters of the alternate hypothesis, regressions should be run for a range of return horizons corresponding to the range of parameter values in the prior distribution. The acceptance or rejection of the null hypothesis should then be based on a joint test of regression coefficients. However, we have shown that while the regression test may be powerful against one of the range of alternative hypotheses for a single regression, the power of the χ^2 joint test against the entire range of alternatives may be only peripherally related to the power of the individual regressions. Therefore, as a joint test, we use the most significant statistic as a measure of the overall significance but statistically correct for having selected this statistic from the set of regression coefficients.

In order to determine significance levels, we empirically calculate the distribution of the most negative of the statistics as described before. Based on this joint test, only the decile 2 returns exhibit evidence of mean-reversion at the 5% two-tailed level. However, based on the stated alternative of the AR(1) fads model, the one-tailed test is appropriate, and the EW and decile 2, 3 and 4 portfolios exhibit mean reversion at a 10% one-tailed level.

However, the economic significance of these results is still suspect based Jegadeesh's (1990) finding that all significant mean reversion appears to be due to high returns of small firms in January, and to the severe heteroskedasticity in the sample period.

6. Conclusions

We have developed a method that allows analytical calculation of the power of tests of mean reversion. This method allows us to calculate the power of any weighted autocorrelation tests. We have shown the equivalence of this test with the long horizon regression test, the variance ratio test, weighted spectral tests, and any instrumental variable or generalized method of moment (GMM) tests, which use linear functions of past returns as instruments.

This method has allowed us to make power comparisons tests of this class and to determine the implicit alternative the tests. In addition, we have shown how to determine the optimal test given a null hypothesis that returns are the sum of a differenced martingale and an alternative process with an ARMA(p, q) representation: in this setting, the optimal test will be a weighted sum of the sample autocorrelations at different lags, where the weights are proportional to the expected return autocorrelations under the alternative hypothesis. We have also provided a simple geometric analogy that gives the intuition for this result.

In the spectral domain, we have shown that the weighted autocorrelation test can just as easily be written as a weighted periodigram test, with an analogous result that the optimal weighted periodigram test will have weights proportional to the expected periodigram under the alternative. This test shares the optimality property of the weighted autocorrelogram test.

We have also addressed the issue of joint tests. We show that the results extend easily to the case of multiple autocorrelation or instrument restrictions as that in the simple geometric intuition developed in the first part of the paper. An important result of this section is that the power of a joint test of moment restrictions (or χ^2 joint test) of this sort may be only peripherally related to the power of the individual tests.

Since our analytical test-power results are valid only asymptotically and under local alternatives, we conducted Monte-Carlo experiments to investigate the robustness of these results for small sample sizes and for nonlocal alternatives found that the results were robust, at least for the limited set of alternatives we consider.

Finally, we have shown how small differences in the small sample properties of a test can lead to strikingly different statistical inferences. We show that the long horizon regression, which uses analytical standard errors, may have low power against simple mean reverting alternatives, and that this, not poor small sample properties, is the reason Richardson and Stock (1989) find no evidence of mean reversion using this test. We empirically calculate the small-sample corrected distribution for the Fama and French T-statistics, and find more evidence in favor of a mean reversion hypothesis.

Acknowledgements

I am indebted to Torben Anderson, John Cochrane, Burton Hollifield, Narasimhan Jegadeesh, Nicholas Polson, Sheridan Titman, Walter Torous and to the participants of the UBC, University of Chicago, USC, and Wharton finance seminars and the annual meetings of the Canadian Economics Association for their helpful discussions, comments and suggestions.

Appendix A. The statistical properties of the sample estimators under the local alternative

If a process is described by:

$$\theta(L)\,x_t = \phi(L)U_t \qquad \tilde{u} \sim IID\left(0,\sigma_u^2\right) \tag{A.1}$$

$$Eu^4 = \eta\sigma^4 < \infty \tag{A.2}$$

where $\phi(L)$ and $\theta(L)$ are finite-order lag polynomials, and where $\phi(z)/\theta(z)$ has roots outside the unit circle (for stationarity), then Brockwell and Davis (1991) show that if we define the vectors of covariances and correlations in the following way:

$$\hat{\mathbf{c}}^x \equiv \begin{bmatrix} \hat{c}_0^x \\ \vdots \\ \hat{c}_h^x \end{bmatrix} \qquad \hat{\boldsymbol{\rho}}^x \equiv \begin{bmatrix} \hat{\rho}_0^x \\ \vdots \\ \hat{\rho}_h^x \end{bmatrix}$$

where

$$\hat{c}_\tau^x \equiv \frac{1}{T}\sum_{t=0}^{T-\tau} x_t x_{t+\tau} \qquad \hat{\rho}_\tau^x \equiv \frac{\hat{c}_\tau^x}{\hat{c}_0^x}$$

then $\hat{\mathbf{c}}^x$ and $\hat{\boldsymbol{\rho}}^x$ will have the following asymptotic distributions:

$$\hat{\mathbf{c}}^x \sim \mathscr{N}(\mathbf{c}^x, T^{-1}\mathbf{V}) \tag{A.3}$$

$$\hat{\boldsymbol{\rho}}^x \sim \mathscr{N}(\rho^x, T^{-1}\mathbf{W}) \tag{A.4}$$

where the elements of the variance–covariance matrices **V** and **W** are given by:

$$v_{ij} = (\eta - 3)c_i c_j + \sum_{k=-\infty}^{\infty}\{c_k c_{k-i+j} + c_{k+j}c_{k-i}\}$$

$$w_{ij} = \sum_{k=-\infty}^{\infty}\left\{\rho_{k+i}\,\rho_{k+j} + \rho_{k-i}\,\rho_{k+j} + 2\rho_i\,\rho_j\,\rho_k^2 - 2\rho_i\,\rho_j\,\rho_{k+j}^2 - 2\rho_j\,\rho_k\,\rho_{k+i}^2\right\}$$

Under the assumption that x_t is generated by a stationary, finite-order ARMA process, each of these elements is finite.

References

Badahur, R.R., 1980. Rates of convergence of estimates and test statistics. Annals of Mathematical Statistics 38, 303–324.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods. 2nd edn. Springer Verlag, New York.

Campbell, J.Y., 1992. Why Long Horizons? Asymptotic Power Against Persistent Alternatives 8, 459–491.

Cochrane, J.H., 1988. How big is the random walk in GNP? Journal of Political Economy 96, 296–893.

Davidson, R., MacKinnon, J.G., 1987. Implicit alternatives and the local power of test statistics. Econometrica 55, 1305–1329.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive times series with a unit root. Journal of the American Statistical Association 79, 427–431.

Durlauf, S.N., 1991. Spectral based testing of the Martingale hypothesis. Journal of Econometrics 50, 355–376.

Fama, E.F., 1970. Efficient capital markets: a review of theory and empirical work. Journal of Finance 25, 383–417.

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

Fama, E.F., French, K., 1988b. Dividend yields and expected stock returns. Journal of Financial Economics 22, 3–25.

Faust, J., 1992. When are variance ratio tests for serial dependence optimal? Econometrica 60, 1215–1226.

Geweke, J., 1981. The approximate slopes of econometric tests. Econometrica 49, 1427–1442.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

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

Hodrick, R.J., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. Review of Financial Studies 5, 357–386.

Jegadeesh, N., 1990. Seasonality in stock price mean reversion; evidence from the U.S. and the U.K. Journal of Finance 45, 881–898.

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

Lo, A.W., MacKinlay, A.C., 1989. The size and power of the variance ratio test in finite samples. Journal of Econometrics 40, 203–238.

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

Richardson, M., 1993. Temporary components of stock prices: a skeptic's view. Journal of Business and Economic Statistics 11, 199–207.

Richardson, M., Smith, T., 1991. Tests of financial models in the presence of overlapping observations. Review of Financial Studies 4, 227–254.

Richardson, M., Smith, T., 1994. A Unified Approach to testing for serial correlation in stock returns. Journal of Business 67, 371–399.

Richardson, M., Stock, J.H., 1989. Drawing inferences from statistics based on multiyear asset returns. Journal of Financial Economics 25, 323–348.

Summers, L.H., 1986. Does the stock market rationally reflect fundamental values? Journal of Finance 41, 591–601.