Tests of Conditional Asset Pricing Models: A New Approach

Kevin Wang

Faculty of Management
McGill University
1001 Sherbrooke Street West
Montreal, PQ H3A 1G5, Canada

Current Draft: November 1998

¤Comments welcome. E-mail: qingwang@management.mcgill.ca; Phone: (514) 398-4054. I am grateful to Yacine Aït-Sahalia, Xiaohong Chen, John Cochrane, George Constantinides, and Robert McCulloch for advice. I also thank Michael Brandt, Philip Dybvig, Lars Hansen, David Hsieh, Ravi Jagannathan, Raymond Kan, Ekaterini Kyriazidou, Kai Li, Ruey Tsay, Guofu Zou, and seminar participants at McGill University, the University of Chicago, the University of North Carolina, the University of Toronto, the University of Waterloo, and Washington University in St Louis for helpful comments and suggestions. This paper has also benefited from a presentation at the Northern Finance Association 1998 meetings in Toronto.

Tests of Conditional Asset Pricing Models: A New Approach

Abstract

Using a nonparametric discount factor, we construct a regression approach to testing the conditional CAPM. A large sample theory, an extension to testing multifactor models, and simulation results are provided. We find that the test performs well for simulated monthly data of the postwar period. In application we find that the conditional CAPM exhibits volatile pricing errors that are positively correlated with the stock market. In addition, the size and book-to-market factors of Fama and French (1993) and the labor income risk factor of Jagannathan and Wang (1996) do not explain the CAPM deviations.

1 Introduction

A problem arises when testing the conditional version of the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965). While the conditional CAPM imposes cross-sectional restrictions on moments of asset returns, the model does not describe exactly how these moments vary over time. The conditional CAPM states that conditional expected excess returns on securities at any given point in time are linear in their conditional covariances with the market, but it says nothing about functional forms of time-varying moments of asset returns. To conduct empirical tests, however, we have to cope with this functional form uncertainty.

This paper explores a new testing approach. Our approach is based on a simple regression testing idea. We utilize the fact that the conditional CAPM implies a stochastic discount factor which is determined by the first two conditional moments of the market return. Using standard kernel regression estimates of these two moments, we obtain a nonparametric estimate of the discount factor. Then we use a regression approach to test if discounted excess returns are predictable. This design circumvents discount factor misspecification and provides a simple way to look at pricing errors.

For conditional multifactor models, we construct a testing approach by combining the nonparametric regression test with Hansen's (1982) generalized method of moments (GMM). In addition, we provide a nonparametric regression test to aid selection of state variables. We have also considered two alternative testing approaches, since the proposed regression test does not have any optimal properties.

We conduct a simulation study to address two questions. First, does the test have power in finite samples? In general nonparametric tests can avoid effects of misspecification, but they may not have sufficient power in finite sample applications. Secondly, how sensitive are test results to state variable selection? This issue is interesting since in practice it is difficult to completely avoid errors in selecting state variables. We design simulation experiments using a postwar monthly data set from the Center for Research in Security Prices (CRSP). The simulation evidence is encouraging. The regression test performs well in the simulations.

In the applications, we test the conditional Sharpe-Lintner CAPM, the Fama and French (1993) three factor model, and the model of Jagannathan and Wang (1996), using monthly data on U.S. stock returns. We find that the conditional CAPM exhibits volatile pricing errors which are positively correlated with the stock market, though it performs significantly better than the unconditional CAPM. These positive correlations indicate that when the market goes up (down), investors tend to set expected returns higher (lower) than the rate of return required by the CAPM, and hence seem to suggest that investors tend to overreact to market movements. On the other hand, the size and book-to-market factors of Fama and French and the labor income risk factor of Jagannathan and Wang have rather limited abilities to explain the CAPM pricing errors.[1]

On implementing the regression testing idea, this paper is related to the works of Powell, Stock, and Stoker (1989), Robinson (1989), Yoshihara (1990), Lee (1992), and Khashimov (1993). The regression estimator underlying our testing approach is similar to the estimators of Powell et al and Lee. The limiting distribution of this estimator is derived using an extension of classical U-statistic theorems to β-mixing processes. Similar extensions have been obtained by Robinson, Yoshihara, and Khashimov.[2]

The target, methodology, and results of this paper can be contrasted with those of Bansal and Viswanathan (1993), Bansal, Hsieh, and Viswanathan (1993). These two papers propose a series expansion GMM approach to testing nonlinear APT models. In this approach, one first approximates the stochastic discount factor with a series expansion (such as a long polynomial) and then employs GMM for estimation and testing. Chapman (1997) has used this approach to study consumption-based models. However, large sample and finite sample properties of this series expansion GMM approach to testing asset pricing models are largely unknown.

There is a substantial and growing literature on testing the conditional CAPM and conditional multifactor models. Existing tests rely upon auxiliary functional form assumptions about moments of asset returns and/or certain functions of these moments. For example, Gibbons and Ferson (1985), and Ferson, Kandel, and Stambaugh (1987) assume that conditional covariances of asset returns are constant over time. Bollerslev, Engle, and Wooldridge (1988) use an ARCH approach

[1] These results are consistent with recent findings of MacKinlay (1995), He, Kan, Ng, and Zhang (1996), and Kan and Zhang (1997).

[2] Powell et al proposed an instrumental variable estimator for limited dependent variables models. Lee designed a density-weighted least squares estimator for heteroskedasticity testing. Robinson provided a class of unconditional moments tests.

to model time variation in conditional covariances. Harvey (1989) assumes a parametric structure on the time-varying first moments and specifies the reward-to-risk ratio. Shanken (1990) uses parametric models in which the conditional betas are linear functions of state variables. Carhart et al. (1994), Cochrane (1996), Jagannathan and Wang (1996) employ auxiliary models for stochastic discount factors, which can also be viewed as functional form assumptions on some function of conditional moments of security returns.

Our testing approach does not rely on functional form assumptions about the discount factor. To motivate this new method, we show with a GMM approach that there exist strong effects of discount factor misspecification on inference and pricing error estimation. The problem that incorrect auxiliary functional forms can bias estimation and test results is well known in this literature.[3] Recently Ghysels (1997) provides a demonstration about effects of beta misspecification. Ghysels shows that in many cases where betas are misspecified, pricing errors of the conditional CAPM are even larger than those of the unconditional CAPM.

This paper is organized as follows. Section 2 details the regression testing idea, constructs a weighted least squares estimator and the test statistic, and presents their asymptotic properties. Section 2 also describes the extension for testing conditional multifactor models and the state variable selection test. Section 3 briefly outlines two closely related alternative approaches. Section 4 conducts simulation experiments to study finite sample performance of the nonparametric regression test and to demonstrate effects of discount factor misspecification. Section 5 presents empirical applications. Section 6 concludes.

[3] For example, see Harvey (1991).

2 The Regression Testing Approach

2.1 The Basic Idea

Using a nonparametric discount factor, we carry out a regression testing idea through a weighted least squares estimator. This section details the idea.

We consider a framework in which there is a conditionally riskless asset. Let $r_{p,t+1}$ be the return on a benchmark portfolio p in excess of the riskless rate, and $r_{i,t+1}$ be the excess return on the i-th test asset for $i = 1, \cdots, n$. Let x_t be a $k \times 1$ vector of state variables such that

$$E(r_{p,t+1} | I_t) = E(r_{p,t+1} | x_t) \quad (1)$$

$$E(r^2_{p,t+1} | I_t) = E(r^2_{p,t+1} | x_t) \quad (2)$$

where I_t is the time-t information set of investors.[4] The excess returns and the state variables are assumed to be strictly stationary.

We present our approach as a test of conditional mean-variance efficiency of a given portfolio. If the benchmark portfolio p is conditionally mean-variance efficient, then

$$E(r_{i,t+1} | I_t) = E(r_{p,t+1} | I_t) \frac{cov(r_{i,t+1}, r_{p,t+1} | I_t)}{var(r_{p,t+1} | I_t)}, \quad (3)$$

or equivalently

$$E(r_{i,t+1} | I_t) = E(r_{p,t+1} | I_t) \frac{E(r_{i,t+1} r_{p,t+1} | I_t)}{E(r^2_{p,t+1} | I_t)}, \quad (4)$$

for $i = 1, \cdots, n$. The covariance representation (3) is the familiar beta-pricing equation. The regression test aims at equation (4), the 'cross moment' representation.[5]

Let $g_p(x_t) = E(r_{p,t+1} | x_t)$, $g_{pp}(x_t) = E(r^2_{p,t+1} | x_t)$, and $b(x_t) = g_p(x_t)/g_{pp}(x_t)$. Given that (1) and (2) hold, conditional expected return errors from (4) can be expressed as

$$E(r_{i,t+1} | I_t) - E(r_{p,t+1} | I_t) \frac{E(r_{i,t+1} r_{p,t+1} | I_t)}{E(r^2_{p,t+1} | I_t)} = E(m_{t+1} r_{i,t+1} | I_t),$$

where $m_{t+1} = 1 - b(x_t) r_{p,t+1}$. Thus (4) is equivalent to

$$E(m_{t+1} r_{i,t+1} | I_t) = 0. \quad (5)$$

---

[4] Note that (1) and (2) are only for the benchmark portfolio p. They are not requiring x_t to be a full characterization of the information set I_t, but they are sufficient for developing the nonparametric test.

[5] Conditional expected return errors from (4) are smaller in absolute value than those from (3). The ratio of the errors from (3) and (4) is one plus squared conditional Sharpe ratio. A nonparametric regression test that targets (3) is proposed in Section 3.

Let $e_{i,t+1} = m_{t+1} r_{i,t+1}$ (discounted excess returns) and z_t be a $q \times 1$ vector of observed stationary variables in I_t. If we could observe the discount factor m_{t+1}, a natural way to test the condition $E(e_{i,t+1} | I_t) = 0$ would be a regression approach: regress $e_{i,t+1}$ on z_t and test if the regression coefficients are zero. This is because the following regression equations

$$e_{i,t+1} = z_t' \delta_i + u_{i,t+1}, \quad (6)$$

where $E(u_{i,t+1} | I_t) = 0$, for $i = 1, \cdots, n$, are always consistent with (5). Obviously, the moment condition (5) implies that the regression equations in (6) hold, with the coefficients satisfying $\delta = 0$ where $\delta = (\delta_1' \delta_2' \cdots \delta_n')'$.

To implement this simple idea, we replace m_{t+1} with a nonparametric discount factor $\hat{m}_{t+1}$, and estimate the parameter vector δ_i by

$$\hat{\delta}_i = \left(\frac{1}{N} \sum_{t=1} \hat{w}_t z_t z_t' \right)^{-1} \left(\frac{1}{N} \sum_{t=1} \hat{w}_t z_t \hat{e}_{i,t+1} \right), \quad (7)$$

for $i = 1, \cdots, n$, where $\hat{e}_{i,t+1} = \hat{m}_{t+1} r_{i,t+1}$ and

$$\hat{m}_{t+1} = 1 - \hat{b}(x_t) r_{p,t+1}$$

with $\hat{b}(x) = \hat{g}_p(x)/\hat{g}_{pp}(x)$. The weighting function is set to be $\hat{w}_t = \hat{f}(x_t)\hat{g}_{pp}(x_t)$. Here $\hat{f}$, $\hat{g}_p$, and $\hat{g}_{pp}$ are kernel estimators defined below

$$\hat{f}(x) = N^{-1} h^{-k} \sum_{s=1} K\left(\frac{x - x_s}{h}\right);$$

$$\hat{g}_p(x) = N^{-1} h^{-k} \hat{f}(x)^{-1} \sum_{s=1} K\left(\frac{x - x_s}{h}\right) r_{p,s+1};$$

$$\hat{g}_{pp}(x) = N^{-1} h^{-k} \hat{f}(x)^{-1} \sum_{s=1} K\left(\frac{x - x_s}{h}\right) r_{p,s+1}^2. \quad (8)$$

The nonparametric estimators in (8) are standard. $\hat{f}(x)$ is the Rosenblatt-Parzen kernel density estimator, with kernel function $K(\cdot)$ and bandwidth parameter h. $\hat{g}_p(x)$ and $\hat{g}_{pp}(x)$ are the Nadaraya-Watson kernel regression function estimators.

The weighting function is chosen purely for a technical constraint. Because of this choice, both $N^{-1} \sum_{t=1}^{N} \hat{w}_t z_t z_t'$ and $N^{-1} \sum_{t=1}^{N} \hat{w}_t z_t \hat{e}_{i,t+1}$ can be expressed as second order generalized U-statistics, making it straightforward to analyze large sample properties of $\hat{\delta}_i$.[6] In contrast, rather complex technical problems arise in developing distribution theory if we use weighting functions that do not give rise to simple U-statistic structures.[7] The weighting function chosen above is the simplest among those that yield U-statistic structures. In a similar spirit, Powell et al. (1989) employed density-weighting to obtain an instrumental variable estimator for limited dependent variables models, Robinson (1989) provided a class of unconditional moments tests for general econometric applications, and Lee (1992) designed a density-weighted least squares estimator for heteroskedasticity testing.

The test that we propose is based on the weighted least squares estimator $\hat{\delta}_N$, where

$$\hat{\delta}_N = (\hat{\delta}_1', \hat{\delta}_2', \cdots, \hat{\delta}_n')'.$$

Intuitively, $\hat{\delta}_N$ converges to zero if the benchmark p is conditionally mean-variance efficient. Otherwise, the estimator converges to a nonzero limit in general (unless $e_{i,t+1}$ is orthogonal to all the components of z_t for $i = 1, \cdots, n$). Thus we can conduct a test by checking how far $\hat{\delta}_N$ is away from zero, using asymptotic distribution theory to account for sampling errors. In addition, the regression testing approach provides a simple way to look into pricing errors. In applications we may view (6) as a model for pricing errors. That is, $z_t'\delta_i$ may serve as an approximation for $E(e_{i,t+1} | I_t)$, the conditional expected return errors from (4). We impose no restrictions on choice of z_t to establish the large sample theory.

2.2 Asymptotics and the Test Statistic

Applying some basic results from the U-statistic literature, we show that the weighted least squares estimator $\hat{\delta}_N$ has a limiting multivariate normal distribution. Then we propose the test statistic, using a simple estimator for the covariance matrix of $\hat{\delta}_N$. The regularity conditions and the proofs are given in Appendix A and Appendix B.

The following notations are used to present the results. Let r_{t+1} be a vector of scaled excess returns $(r_{1,t+1}, \cdots, r_{n,t+1})' \otimes z_t$ and $y_{t+1} = (x_t' \otimes z_t', r_{p,t+1}, r_{t+1}')'$, where '$\otimes$' is the Kronecker operator.

[6] There is a large literature on U-statistics. See Arcones (1995), Khashimov (1993), Powell et al (1989), Robinson (1989), and Yoshihara (1990) for some recent contributions.

[7] In particular, the choice of setting $\hat{w}_t = 1$ does not yield U-statistic structures. This is why it is not used here.

Denote $w_t = f(x_t) g_{pp}(x_t)$, $A = I_n \otimes E[w_t z_t z_t']$, and $\hat{A}_N = I_n \otimes N^{-1} \sum_{t=1}^{N} \hat{w}_t z_t z_t'$, where I_n is the $n \times n$ identity matrix. Let $\delta = (\delta_1' \cdots \delta_n')'$ with $\delta_i = [E(w_t z_t z_t')]^{-1} E[w_t z_t e_{i,t+1}]$.

Define

$$\gamma(y_{t+1}) = \eta(y_{t+1}) - [I_n \otimes a(y_{t+1})]\delta; \tag{9}$$

$$\begin{aligned} \eta(y_{t+1}) = {} & f(x_t)[g_{pp}(x_t) r_{t+1} - g_p(x_t) r_{p,t+1} r_{t+1} \\ & + g_r(x_t) r_{p,t+1}^2 - g_{pr}(x_t) r_{p,t+1}]; \end{aligned} \tag{10}$$

$$a(y_{t+1}) = f(x_t)[g_{pp}(x_t) z_t z_t' + r_{p,t+1}^2 g_{zz}(x_t)]; \tag{11}$$

where $g_r(x_t) = E(r_{t+1}|x_t)$, $g_{pr}(x_t) = E(r_{p,t+1} r_{t+1}|x_t)$, and $g_{zz}(x_t) = E(z_t z_t'|x_t)$.

In Appendix B, we show that (i) $\hat{A}_N$ converges in probability to A, (ii) the limiting distribution of $\sqrt{N}\hat{A}_N(\hat{\delta}_N - \delta)$ is identical to that of $N^{-1/2}\sum_{t=1}^{N} \gamma(y_{t+1})$, and (iii) $E\gamma(y_{t+1}) = 0$. Because $N^{-1}\sum_{t=1}^{N} \gamma(y_{t+1})$ is a simple average of stationary random vectors, application of a central limit theorem thus gives the following result.

Theorem 1: Given Assumptions A1–A6 stated in Appendix A, if $h \to 0$, $Nh^{2k} \to \infty$, and $Nh^{2k+2} \to 0$, then the weighted least squares estimator $\hat{\delta}_N$ is such that $\sqrt{N}(\hat{\delta}_N - \delta)$ has a limiting multivariate normal distribution with mean 0 and variance-covariance matrix Ω, where $\Omega = A^{-1}\Gamma A^{-1}$, $\Gamma = \sum_{j=-\infty}^{\infty} \Gamma_j$, and $\Gamma_j = E[\gamma(y_{t+1})\gamma(y_{t+j+1})']$.

Theorem 1 shows that the weighted least squares estimator $\hat{\delta}_N$ has the standard limiting properties, $\sqrt{N}$-consistency and asymptotic normality, of parametric estimators. This result does not rely upon (1) and (2), nor does it require that the regression equations in (6) are correctly specified. Note that the bandwidth conditions are different from those for pointwise kernel estimators. The conditions $Nh^{2k} \to \infty$ and $Nh^{2k+2} \to 0$ place upper and lower bounds on the rate that the bandwidth h converges to 0, for $\hat{\delta}_N$ to exhibit the desired asymptotic behavior. The condition $Nh^{2k+2} \to 0$ is due to the use of a kernel of order $k+1$ (Assumption A4), and hence the admissible range for the rate can be relaxed using a kernel of order higher than $k+1$.

To construct a test using the distribution of $\hat{\delta}_N$, we need to have an estimator for the covariance matrix Ω. For this purpose, consider first estimation of $\gamma(y_{t+1})$. Replacing the functions $f(x)$, $g_p(x)$,

$g_{pp}(x)$, $g_r(x)$, $g_{pr}(x)$, and $g_{zz}(x)$ in (10) and (11) by standard kernel estimators,[8] and replacing δ in (9) by $\hat{\delta}_N$, we obtain a natural approximation for $\gamma(y_{t+1})$:

$$\hat{\gamma}_N(y_{t+1}) = \hat{\psi}_N(y_{t+1}) - [\iota_h \otimes \hat{a}_N(y_{t+1})]\hat{\delta}_N; \tag{12}$$

$$\begin{aligned} \hat{\psi}_N(y_{t+1}) &= \hat{f}(x_t)[\hat{g}_{pp}(x_t) r_{t+1} - \hat{g}_p(x_t) r_{p,t+1} r_{t+1} \\ &\quad + \hat{g}_r(x_t) r^2_{p,t+1} - \hat{g}_{pr}(x_t) r_{p,t+1}]; \end{aligned} \tag{13}$$

$$\hat{a}_N(y_{t+1}) = \hat{f}(x_t)[\hat{g}_{pp}(x_t) z_t z_t' + r^2_{p,t+1}\hat{g}_{zz}(x_t)]; \tag{14}$$

where $\hat{f}$, $\hat{g}_p$, and $\hat{g}_{pp}$ are defined as in Section 2.1, and

$$\hat{g}_r(x) = N^{-1} h^{-k} \hat{f}(x)^{-1} \sum_{s=1}^{N} K\left(\frac{x - x_s}{h}\right) r_{s+1};$$

$$\hat{g}_{pr}(x) = N^{-1} h^{-k} \hat{f}(x)^{-1} \sum_{s=1}^{N} K\left(\frac{x - x_s}{h}\right) r_{p,s+1} r_{s+1};$$

$$\hat{g}_{zz}(x) = N^{-1} h^{-k} \hat{f}(x)^{-1} \sum_{s=1}^{N} K\left(\frac{x - x_s}{h}\right) z_s z_s'.$$

We show in Appendix B that the estimator

$$\hat{\Gamma}_j = N^{-1} \sum_{t=1}^{N-j} \hat{\gamma}_N(y_{t+1}) \hat{\gamma}_N(y_{t-j+1})'$$

is consistent for Γ_j. It is also shown that given (1) and (2), $\Gamma_j = 0$ for any $j \neq 0$ when the regression equations in (6) hold.

Thus we propose to use the test statistic

$$\hat{T}_\delta = N \hat{\delta}_N' \hat{\Omega}_N^{-1} \hat{\delta}_N; \tag{15}$$

where $\hat{\Omega}_N = \hat{A}_N^{-1} \hat{\Gamma}_0 \hat{A}_N^{-1}$, for testing conditional mean-variance efficiency of the benchmark. The following theorem gives the limiting distribution of the test statistic.

Theorem 2: Let the conditions of Theorem 1 hold.

(i) Given (1) and (2), if the portfolio p is conditionally mean-variance efficient, then the test statistic $\hat{T}_\delta$ has a limiting chi-squared distribution with $q \times n$ degrees of freedom.

---

[8] Note that $g_{zz}(x_t) = z_t z_t'$ when z_t is a fixed transformation of x_t, for example, when $z_t = (1\ x_t')'$. In such circumstances there is no need to use the kernel estimator $\hat{g}_{zz}(x_t)$ in (14). Instead one can simply replace $g_{zz}(x_t)$ by $z_t z_t'$, which gives $\hat{a}_N(y_{t+1}) = \hat{f}(x_t)[\hat{g}_{pp}(x_t) + r^2_{p,t+1}] z_t z_t'$.

(ii) $\hat{\gamma}_{ij}$ is a consistent estimator of γ_{ij} for any fixed j.

The regression test is constructed using a simple covariance matrix estimator. Note that Theorem 2 has also provided necessary inputs ($\hat{\gamma}_{ij}$) to obtain covariance matrix estimators that are consistent under general circumstances. Whether there exists any advantage to use such estimators in applications remains to be studied.[9]

A few points are worth noticed. First, the parametric convergence rate of $\hat{\delta}_N$ is pleasant, as it provides a reason to expect good finite sample performance. However, it should be noted that there is certain cost due to nonparametric estimation, which is embedded in the covariance matrix of $\hat{\delta}_N$. Secondly, it is easy to compute the test statistic and the underlying estimators, given the kernel function and the bandwidth parameter. Finally, note that the test will not have power if we choose some vector z_t that is orthogonal to $e_{i,t+1}$ (i.e., $E(z_t e_{i,t+1}|x_t) = 0$, $i = 1, \cdots, n$).[10] However, the test will have power as long as one component of z_t can significantly forecast $e_{i,t+1}$, no matter whether the regression model (6) is correctly specified or not.

2.3 Selection of the State Variables

In general we can not observe the entire information set of investors. Even if we can, a test which is constructed using a large number of state variables is likely to have poor power. In practice, we can only use a subset of the information set, typically a small number of state variables. Although it is invalid to ignore the conditioning information set (Hansen and Richard 1987), it is possible to use a much smaller subset to develop a valid test. The result below, implied in Dybvig and Ross (1985), shows that security-specific information may be ignored. One can proceed with a subset as long as it contains variables that characterize the first two moments of the market.[11]

Theorem 3: Let x_t satisfy (1) and (2), and let I_t^* be such that $x_t \in I_t^* \subseteq I_t$. Define

$$\epsilon_{i,t} \equiv E(r_{i,t+1}|I_t) - E(r_{p,t+1}|I_t)\frac{E(r_{i,t+1}r_{p,t+1}|I_t)}{E(r_{p,t+1}^2|I_t)};$$

[9] In application we have attempted to use a Newey-West covariance matrix estimator with a lag length of six, twelve, and eighteen. However, the estimates are (or so close to be) singular such that the computer cannot invert them. On the other hand, the regression residuals look quite like martingale sequences, supplying no incentive to pursue a more complex approach for estimation of the covariance matrix.

[10] In statistical term, the test is not consistent, for a given vector z_t, against certain processes that generate $e_{i,t+1}$.

[11] The proof of Theorem 3 is straightforward and hence omitted.

$$\varepsilon^*_{i,t} \equiv E(r_{i,t+1}|I^*_t) - E(r_{p,t+1}|I^*_t)\frac{E(r_{i,t+1}r_{p,t+1}|I^*_t)}{E(r^2_{p,t+1}|I^*_t)}.$$

Then $\varepsilon^*_{i,t} = E(\varepsilon_{i,t}|I^*_t)$. Thus it follows that (i) if $\varepsilon_{i,t} = 0$, then $\varepsilon^*_{i,t} = 0$; (ii) $E(\varepsilon^*_{i,t}) = E(\varepsilon_{i,t})$; (iii) $var(\varepsilon^*_{i,t}) \leq var(\varepsilon_{i,t})$; (iv) if $z_t \in I^*_t$, then $cov(z_t, \varepsilon^*_{i,t}) = cov(z_t, \varepsilon_{i,t})$.

Theorem 3 shows that as long as I^*_t contains x_t, a test based on I^*_t is indeed a test of implications of the model. Yet there are consequences for using a subset of the information set. First, the volatility of pricing errors based on I^*_t is not equal to that based on I_t. Instead, it is a lower bound. Second, any test based on I^*_t is likely to be inconsistent, since $\varepsilon^*_{i,t}$ can be zero while $\varepsilon_{i,t}$ is not.

For power consideration, a sensible approach is to use a conditioning information set that just consists of x_t. A simple way to select such variables is to use linear regression methods. Below, we outline a nonparametric test for state variable selection. That is, a test of (1) and (2) in Section 2.1, given a candidate vector x_t. This approach is more robust to possible nonlinear relations.

Let z_t be a $q_\mu \times 1$ vector in the information set I_t, where components of z_t are different from those of x_t. The idea of the selection test is to check if z_t can forecast the residuals $r_{p,t+1} - g_p(x_t)$ and $r^2_{p,t+1} - g_{pp}(x_t)$.

If (1) holds, i.e., if $E(r_{p,t+1}|I_t) = g_p(x_t)$, then

$$r_{p,t+1} - g_p(x_t) = z_t'\mu + \epsilon_{t+1},$$

with $E(\epsilon_{t+1}|I_t) = 0$, and $\mu = 0$. Under certain regularity conditions, if $Nh^{2k} \to \infty$ and $Nh^{2k+2} \to 0$, then one can show that the following estimator of μ

$$\hat\mu = \left(\frac{1}{N}\sum_{t=1}^{N}\hat f(x_t)z_tz_t'\right)^{-1}\left(\frac{1}{N}\sum_{t=1}^{N}\hat f(x_t)z_t[r_{p,t+1} - \hat g_p(x_t)]\right)$$

is such that $\sqrt{N}(\hat\mu - \mu) \xrightarrow{d} N(0, \Omega_\mu)$.

One may construct a covariance matrix estimator as in Theorem 2. Specifically, one can use $\hat\Omega_\mu = \hat A_\mu^{-1}\hat\Gamma_\mu\hat A_\mu^{-1}$, where $\hat A_\mu = N^{-1}\sum_{t=1}^{N}\hat f(x_t)z_tz_t'$, and $\hat\Gamma_\mu = N^{-1}\sum_{t=1}^{N}\hat\gamma_{\mu,t+1}\hat\gamma'_{\mu,t+1}$. The input $\hat\gamma_{\mu,t+1}$ is

$$\hat\gamma_{\mu,t+1} = \hat f(x_t)[z_tr_{p,t+1} - z_t\hat g_p(x_t) - r_{p,t+1}\hat g_z(x_t) + \hat g_{zp}(x_t) - \hat g_{zz}(x_t)\hat\mu - z_tz_t'\hat\mu]$$

with $\hat{g}_{zz}(x)$ defined as in Section 2.2 and

$$\hat{g}_z(x) = N^{-1}h^{-k}\hat{f}(x)^{-1}\sum_{s=1}^{N} K\left(\frac{x - x_s}{h}\right) z_s;$$

$$\hat{g}_{zp}(x) = N^{-1}h^{-k}\hat{f}(x)^{-1}\sum_{s=1}^{N} K\left(\frac{x - x_s}{h}\right) z_s r_{p,s+1}.$$

These results yield a selection test statistic $N\hat{\Delta}'\hat{\Omega}_1^{-1}\hat{\Delta}$, which has a limiting $\chi^2(q_1)$ distribution under (1). A test of (2) can be constructed similarly, just replacing $r_{p,t+1}$ and $r_{p,s+1}$ with $r^2_{p,t+1}$ and $r^2_{p,s+1}$, respectively.

2.4 An Extension for Testing Multifactor Models

We extend the regression testing approach to allow a parametric structure for excess returns on the benchmark portfolio p. That is, excess returns $r_{p,t+1}(\theta)$ are assumed to be a function of a $l \times 1$ parameter vector θ. The hypothesis is that the benchmark is conditionally mean-variance efficient for some parameter value θ_0. The methodology developed in this section aims at multifactor asset pricing predictions that a portfolio of several factor mimicking portfolios is on the conditional mean-variance efficiency frontier.

A conditional version of the Fama and French (1993) three factor model, for example, is such that a benchmark portfolio with time-$(t+1)$ excess return

$$r_{p,t+1}(\theta) = RM_{t+1} + \theta_1 SMB_{t+1} + \theta_2 HML_{t+1}$$

is conditionally mean-variance efficient, where RM_{t+1} is the excess return on the Fama and French market portfolio, SMB_{t+1} and HML_{t+1} are returns on their mimicking portfolios for the size and the book-to-market factors.

Testing one factor models may also involve parametric benchmark returns, when the factor is latent or unobservable. For instance, in their study of the conditional CAPM, Jagannathan and Wang (1996) use a parametric proxy for return on the market portfolio. They assume that return on the market portfolio is a linear function of return on a stock market index and labor income growth rate. Alternatively, their specification may be viewed as a conditional two factor model.

To test such parametric hypotheses, we propose an approach that combines the nonparametric regression test with the conventional GMM procedure.

For any given value of θ, one can obtain an estimator $\hat{\delta}_N(\theta)$ from (7) and (8) of Section 2.1, with $r_{p;t+1}(\theta)$ replacing $r_{p;t+1}$. Let $\Omega(\theta)$ and $\hat{\Omega}_N(\theta)$ denote the asymptotic covariance matrix of $\hat{\delta}_N(\theta)$ and the estimator for this matrix, respectively, defined as in Section 2.2 with $r_{p;t+1}(\theta)$ replacing $r_{p;t+1}$. Let $\hat{\theta}_N$ be the parameter value that minimizes

$$\hat{\delta}_N(\theta)'\hat{W}_N\hat{\delta}_N(\theta);$$

where the weighting matrix $\hat{W}_N$ is regarded as fixed with respect to θ, $\hat{W}_N \xrightarrow{p} \Omega_0^{-1}$, and $\Omega_0 \equiv \Omega(\theta_0)$. Let $D_N(\theta)$ denote the $l \times q$ matrix of partial derivatives of $\hat{\delta}_N(\theta)$ with respect to the parameter vector θ: $D_N(\theta) \equiv \partial\hat{\delta}_N(\theta)'/\partial\theta$. The estimator $\hat{\theta}_N$ is assumed to satisfy the following first-order condition

$$D_N(\hat{\theta}_N)\hat{W}_N\hat{\delta}_N(\hat{\theta}_N) = 0: \qquad (16)$$

We propose to use the test statistic $N\hat{\delta}_N(\hat{\theta}_N)'\hat{W}_N\hat{\delta}_N(\hat{\theta}_N)$. The following theorem gives the limiting distributions of the estimator $\hat{\theta}_N$ and the test statistic under the hypothesis that the benchmark p is conditionally mean-variance efficient.[12]

Theorem 4: Let the conditions of Theorems 1 and 2 hold with $r_{p;t+1}(\theta_0)$ replacing $r_{p;t+1}$. Let $\hat{\delta}_N(\theta)$ be differentiable in θ and $\hat{\theta}_N$ be the estimator satisfying (16) with $l < q$. Let $\{\hat{W}_N\}_{N=1}^{\infty}$ be a sequence of positive definite matrices such that $\hat{W}_N \xrightarrow{p} \Omega_0^{-1}$. Suppose further that (i) $\hat{\theta}_N \xrightarrow{p} \theta_0$, and (ii) for any sequence $\{\theta_N^*\}_{N=1}^{\infty}$ satisfying $\theta_N^* \xrightarrow{p} \theta_0$, $\text{plim} D_N(\theta_N^*) = \text{plim} D_N(\theta_0) \equiv D_0$, with the l rows of D_0 linearly independent. Then

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{d} N(0;(D_0\Omega_0^{-1}D_0')^{-1});$$

and the conditional efficiency test statistic $N\hat{\delta}_N(\hat{\theta}_N)'\hat{W}_N\hat{\delta}_N(\hat{\theta}_N)$ has a limiting chi-squared distribution with $qn - l$ degrees of freedom.

In applications we set the weighting matrix $\hat{W}_N$ to be the inverse of the covariance matrix estimate $\hat{\Omega}_N(\theta)$ and update value of θ through iteration. The weighting matrix is evaluated at zero ($\theta = 0$) for the first stage estimation. Then the first stage estimate of θ is used to update the weighting matrix at the second stage. And so on.

[12] Proof of this theorem is straightforward and hence omitted.

3 Two Alternative Tests

Like most nonparametric methods, the proposed regression testing approach is not associated with any optimal properties. On the other hand, there are two closely related approaches to testing the conditional mean-variance efficiency hypothesis. This section presents a brief outline of the two alternatives.

The test proposed above aims at the cross moment representation (4). Alternatively, we can target the beta-pricing equation (3) to construct a regression test. It turns out that we just need to redefine $e_{i,t+1}$ and $\hat{w}_t$ and then repeat the above procedure. Given (1) and (2), conditional expected return errors from (3) can be expressed as

$$E(r_{i,t+1}|I_t) - E(r_{p,t+1}|I_t)\frac{\mathrm{cov}(r_{i,t+1}, r_{p,t+1}|I_t)}{\mathrm{var}(r_{p,t+1}|I_t)} = E(e_{i,t+1}|I_t);$$

where

$$e_{i,t+1} = \left(\frac{g_{pp}(x_t) - g_p(x_t)r_{p,t+1}}{\sigma_p^2(x_t)}\right) r_{i,t+1};$$

and $\sigma_p^2(x_t) = g_{pp}(x_t) - g_p^2(x_t)$.

Thus (3) is equivalent to $E(e_{i,t+1}|I_t) = 0$, which implies

$$e_{i,t+1} = z_t'\alpha_i + u_{i,t+1}$$

with $E(u_{i,t+1}|I_t) = 0$ for $i = 1, \cdots, n$, and $\alpha = 0$, where $\alpha = (\alpha_1' \cdots \alpha_n')'$.

We estimate the parameter vector α_i by

$$\hat{\alpha}_i = \left(\frac{1}{N}\sum_{t=1} \tilde{w}_t z_t z_t'\right)^{-1}\left(\frac{1}{N}\sum_{t=1} \tilde{w}_t z_t \hat{e}_{i,t+1}\right);$$

where $\tilde{w}_t = \hat{f}^2(x_t)\hat{\sigma}_p^2(x_t)$, $\hat{\sigma}_p^2(x_t) = \hat{g}_{pp}(x_t) - \hat{g}_p^2(x_t)$, and

$$\hat{e}_{i,t+1} = \left(\frac{\hat{g}_{pp}(x_t) - \hat{g}_p(x_t)r_{p,t+1}}{\hat{\sigma}_p^2(x_t)}\right) r_{i,t+1}.$$

With the weight $\tilde{w}_t$, both $N^{-1}\sum_{t=1}^N \tilde{w}_t z_t z_t'$ and $N^{-1}\sum_{t=1}^N \tilde{w}_t z_t \hat{e}_{i,t+1}$ can be expressed as third order generalized U-statistics. Under certain regularity conditions, if $N h^{4k} \to \infty$ and $N h^{4k+2} \to 0$, then the weighted least squares estimator $\hat{\alpha}_N = (\hat{\alpha}_1' \cdots \hat{\alpha}_n')'$ has a limiting centered normal distribution.

Similarly as in Theorem 2, we can show that given (1) and (2), the test statistic

$$\hat{T}_{\alpha} = N \hat{\alpha}_N' \hat{\Sigma}_N^{-1} \hat{\alpha}_N \tag{17}$$

has a limiting $\chi^2(qn)$ distribution if the benchmark portfolio p is conditionally mean-variance efficient.[13]

The other alternative is an unconditional moments test. Given (1) and (2), the beta-pricing equation (3) implies a moment condition $\lambda = 0$, where

$$\lambda = E\{f(x_t)[g_{pp}(x_t) - g_p(x_t) r_{p,t+1}] r_{t+1}\}.$$

We estimate the vector λ of unconditional moments by

$$\hat{\lambda}_N = \frac{1}{N}\sum_{t=1}^{N} \hat{f}(x_t)[\hat{g}_{pp}(x_t) - \hat{g}_p(x_t) r_{p,t+1}] r_{t+1}.$$

This estimator can be expressed as a generalized U-statistic of second order. In addition to certain regularity conditions, if $N h^{2k} \to \infty$ and $N h^{2k+2} \to 0$, then $\hat{\lambda}_N$ has a limiting normal distribution. This leads to the following test statistic

$$\hat{T}_{\lambda} = N \hat{\lambda}_N' \hat{\Omega}_N^{-1} \hat{\lambda}_N. \tag{18}$$

Given (1) and (2), we can show that the test statistic $\hat{T}_{\lambda}$ has a limiting $\chi^2(qn)$ distribution if the benchmark is conditionally efficient.[14] In the econometrics literature, Robinson (1989) has studied a class of unconditional moments tests. The test given in (18) fits within his framework.

The proposed regression test and these two alternative tests are not equivalent. Analytical comparison of their performances is rather difficult. The central concern of this paper is whether these asymptotically justified methods have practical values for the asset pricing applications at hand. In simulation experiments, we find that the regression test of Section 2 performs significantly better than these two alternatives.[15]

[13] Details of the covariance matrix estimator $\hat{\Sigma}_N$ are omitted for brevity.

[14] Details of the covariance matrix estimator $\hat{\Omega}_N$ are omitted for brevity.

[15] In this version of the paper, we have omitted results using these two alternatives. These results as well as details of the covariance matrix estimators in (17) and (18) are available upon request.

4 Simulation Experiments

4.1 The Data

Postwar monthly data on six excess returns and five forecasting variables are obtained from CRSP to design the simulation experiments. We use the value-weighted portfolio of NYSE stocks as the benchmark, and the value-weighted NYSE size decile 1, 3, 5, 7, and 9 portfolios as test assets. Excess returns on these portfolios are denoted as VWR, SZ1, SZ3, SZ5, SZ7, and SZ9, respectively. The five forecasting variables are the dividend/price ratio (DPR), the default premium (DEF), the one-month Treasury bill rate (RTB), excess return on the NYSE equally-weighted portfolio (EWR), and the term premium (TERM). Table 1 provides definitions of the variables and summary statistics of the data.

The forecasting variables are selected out of a larger set of ten variables including industry growth rate, inflation rate, short-end term structure slope, January dummy, and excess return on the NYSE value-weighted index. The ten variables are in one-month lag to the excess return VWR.[16] We proceed in two steps to select the state variables. We apply first the ordinary least squares regression method and then the nonparametric selection test of Section 2.3. The OLS regressions indicate that joint use of the three popular forecasters DPR, DEF, and RTB drives out other variables in predicting VWR, and EWR shows up strongly in predicting VWR squared. The nonparametric selection tests do not reject that the four variables DPR, DEF, RTB, and EWR are sufficient to characterize the first two conditional moments of VWR, but produce rejections against dropping any one of the four.[17] We keep TERM as a redundant state variable in simulations, just to check sensitivity of test results to state variable selection.

4.2 Simulated Economies

We construct four simulated economies, or four ways to generate excess returns on the assets, to study finite sample performances of the nonparametric tests. The structures of these economies are identical except for conditional expected returns on the five size portfolios.

The forecasting variables DPR, DEF, RTB, EWR, and TERM are generated through a VAR(1)

[16] More distant lags have also been considered.

[17] The nonparametric selection tests use the samples from 2/47 to 12/95 and from 1/60 to 12/95.

model. Let $y_1 = \ln(DPR)$, $y_2 = \ln(DEF)$, $y_3 = \ln(RTB)$, $y_4 = EWR$, and $y_5 = TERM$. The 5×1 vector y_t is assumed to follow the process

$$y_t - \bar{y} = \Phi(y_{t-1} - \bar{y}) + \epsilon_t;$$

where $\bar{y}$ is the mean vector of y_t, Φ is the matrix of coefficients, and the residual ϵ_t is normally distributed. The model is estimated from the historical data. Parameter estimates, reported in part (A) of Table 2, are used as the 'true' parameter values in simulations.

The conditional distribution of the six excess returns in these economies is multivariate normal. (This does not imply that the unconditional distribution is normal.) The conditional correlation matrix is assumed to be constant and equal to the sample correlation matrix given in Table 1. For each asset, time-t conditional standard deviations of the time-(t+1) returns are of the form

$$\sigma_t = |a_0 + a_1 DPR_t + a_2 DEF_t + a_3 RTB_t + a_4 EWR_t|:$$

The parameter values are determined by a two-step procedure of Davidian and Carroll (1987). In the first step, we regress asset returns on the four state variables DPR, DEF, RTB, and EWR. Then we regress the absolute residual from the first step on the four predictors. With a scaling constant adjustment, the estimates at the second step become the parameter values in simulations.[18]

Time-t conditional mean of the excess return VWR takes the form:

$$E_t(VWR_{t+1}) = b_0 + b_1 DPR_t + b_2 DEF_t + b_3 RTB_t + b_4 EWR_t:$$

In the first economy, H_0, the NYSE value-weighted portfolio is conditionally mean-variance efficient. The conditional means of SZ1 through SZ9 are determined by the beta-pricing equation (3). In the other three economies H_1, H_2, and H_3, the benchmark is not conditionally mean-variance efficient. We generate the conditional means of SZ1 through SZ9 by (i) a linear model under H_1, (ii) a log-linear model under H_2, and (iii) a quadratic model (similar to an ARCH-m model) under H_3. Details of these models are given in Table 2. In every case we let regression estimates from

[18] Let return r_{t+1} be generated as $r_{t+1} = \mu(x_t) + \sigma(x_t)\epsilon_{t+1}$, where the shock ϵ_{t+1} is i.i.d. and independent from x_t, with mean 0 and variance 1. Then $E_t|r_{t+1} - \mu(x_t)| = \phi(x_t)$, where $\phi(x_t) = \sigma(x_t)=c$ for some constant c. Let $\mu(x_t)$ and $\sigma(x_t)$ be linear functions. First, compute least squares regression fit $\hat{\mu}(x_t)$ for $\mu(x_t)$. Then regress the absolute residual $|r_{t+1} - \hat{\mu}(x_t)|$ on x_t, producing regression fit $\hat{\phi}(x_t)$ for $\phi(x_t)$. The scaling constant c is estimated by the square root of $N^{-1} \sum_{t=1}^{N} |r_{t+1} - \hat{\mu}(x_t)|^2 = |\hat{\phi}(x_t)|^2$. At only one point in the sample period, $\hat{\phi}(x_t)$ is below 0.5(%). So nonsmoothness of σ_t is not an issue here.

the historical data, all reported in part (B) of Table 2, to be the 'true' parameter values in the simulation study.

Conditional expected return errors in the economies H_1, H_2, and H_3 are all nonlinear functions of the four state variables. Summary measures of the pricing errors are provided in Table 2 (B). In terms of these measures, the pricing errors in the three economies share some common features. In every case, the average pricing errors are small, but the pricing errors are substantial in terms of volatility. The standard deviations have a clearly negative relation to size, while the first auto-correlations of the pricing errors have a positive size pattern. Pleasantly, as we will see in Section 5, the pricing errors estimated from the historical data also exhibit these characteristics.[19]

4.3 The Kernel and the Bandwidth

Two kernel functions are used in the simulation study of the nonparametric regression test. The first kernel is an independent multivariate normal density function

$$K(u) = \prod_{i=1}^{k} \phi_i(u_i)$$

where ϕ_i is the density of a univariate normal with mean zero and variance σ_i^2, and σ_i is the standard deviation of the i-th state variable. In computation, σ_i is replaced by the sample standard deviation estimate. An independent multivariate normal kernel is a popular choice in kernel estimation methods (e.g. Boudoukh et al. 1995). This kernel is referred to as the normal kernel in the following discussion.

The second one is a "bias-corrected" kernel, which is constructed from the normal kernel as follows (see Schucany and Sommers 1977 or Powell et al. 1989)

$$K^*(u) = \frac{K(u) - \sum_{j=1}^{k} a_j b_j^{-k} K(u/b_j)}{1 - \sum_{j=1}^{k} a_j}$$

where $a = B^{-1}e$. Here $a = (a_1; \ldots; a_k)'$, B is a $k \times k$ matrix with the $(i;j)$-th component $B_{ij} = b_j^{i}$, e is a $k \times 1$ vector of ones, and $b_j = k + j$ for $j = 1; \ldots; k$. It is easy to verify that this is a kernel of order $k+1$ satisfying Assumption A4 in Appendix.

[19] For every of the four simulated economies, we have computed means, standard deviations, first autocorrelations, and crosscorrelations of asset returns. The summary measures from the simulated data are very similar to each other across the simulated economies, and to those from the historical data. For brevity, such double checking results are not reported.

Optimal bandwidth selection is an unresolved issue. There is no theoretical guidance in this context. For a practical choice, we set the bandwidth parameter to be

$$h = N^{-\frac{1}{2k+1}}.$$

Obviously, this takes into account the bandwidth convergence rate conditions for the limiting distribution result of Theorem 1. Note that scale adjustment of the state variables is already made in the above kernels through the standard deviations σ_i. Such a simple bandwidth rule has been regarded as an objective starting point and adopted by many authors (e.g., Silverman (1986), Pagan and Schwert (1990), Harvey (1991), among others).[20]

4.4 The Simulation Results

Table 3 presents the simulation results on finite sample performances of the regression test that we have proposed in Section 2. In constructing the test, the state variable vector is

$$x_t = (DPR_t, DEF_t, RTB_t, EWR_t)'$$

except for panel (C) where we add the redundant variable TERM. The instrument vector is $z_t = (1, x_t')'$ throughout the simulations. See Table 3 for other details.

Panel (A) records % rejections of the test against conditional mean-variance efficiency of the NYSE benchmark, under the null and the three alternative economies at 5% and 10% levels. First, look at the results based on the normal kernel. At $N = 300$, the test has reasonable performance given that four state variables are being used. Most impressive is its performance at sample size $N = 600$. Under H_0, it produces rejection probabilities very close to the corresponding significance levels. Meanwhile, the power of the test is nearly perfect under H_1, H_2, and H_3.

The results based on the higher order kernel are similar but weaker. As one would expect, the rejection probabilities with the higher order kernel are always lower. This is consistent with the fact that higher order kernels tend to produce more variable estimates (e.g., Härdle 1990). Thus we focus on the normal kernel for the rest of the experiments.[21]

[20] We tried a cross-validation method to determine constant c in $h = cN^{-\frac{1}{2k+1}}$. For the postwar monthly dataset, $c = 1.04$.

[21] In terms of bias reduction, the higher order kernel is not effective in this finite sample context.

Panel (B) presents a snapshot of finite sample distribution of the test statistic (i.e., $\hat{T}_{\delta}$ in (15)) under H_0, with sample size $N = 600$. It shows that the distribution of the test statistic is surprisingly close to the predicted limiting distribution.

Panel (C) presents the case in which the redundant variable $TERM$ is added to the state variable set. That is, x_t is now composed of the five variables DPR, DEF, RTB, EWR, and $TERM$. As expected, adding this redundant variable has clearly affected powers of the tests. Overall, the rejection probabilities drop, with an amount between 5% and 15% in most cases. However, the test still has impressive power at $N = 600$, with rejection probabilities all above 85% against the three alternatives.

Along with panel (C), panel (D) presents evidence on sensitivity of the test to selection of state variables. In each case, one of the four state variables is assumed to have been missed.[22] For the first four cases, the variable $TERM$ is selected and used with the other three state variables in constructing the tests. Surprisingly, the test is still well behaved. The rejection probabilities under H_0 are all fairly close to the corresponding significance levels, and the powers under the three alternatives remain strong.

Panel (E) reports the regression estimates. Under H_0, the WLS regression estimates are well behaved in the sense that most of the estimates are within one standard deviation from zero (all within 2 standard deviations). Under H_1, the slopes on EWR are negatively related to size. The EWR slopes in the $SZ1$, $SZ3$, and $SZ5$ regressions have averages that are at least two standard errors above zero. The means of other regression estimates are all within two standard deviations from zero. These suggest that EWR is the major contributor to the power of the regression test. Regression results under H_2 and H_3 are similar (with larger EWR slopes) and hence not reported.

Panel (F) presents results on testing a conditional two factor model. The hypothesis is that a portfolio of the value-weighted index and the first decile portfolio, with excess return $(1 - \mu)VWR + \mu SZ1$, is conditionally mean-variance efficient. By construction of the simulated economies, this hypothesis holds with $\mu = 0$ under H_0, but does not hold under H_1, H_2, and H_3.

The results show that the regression-GMM test of Section 2.4 performs well. At $N = 600$, performance of the test is impressive, in terms of both size and power. The rejection probabilities

[22] In practice, this could happen either due to lack of data or an error of a selection procedure.

are similar to those reported in (A). At $N = 300$, the power is significantly lower. The rejection probabilities at 10% level range between 50% and 80%.

The iterated procedure produces slightly better powers than the two-stage procedure. However, the two-stage procedure produces better estimates of μ under H_0, in terms of both bias and standard deviation. Under the three alternatives, there is no prediction what the estimates of μ should be like. The estimates are all positive on average, but only the estimates in the economy H_2 have means above two standard errors from zero.

In sum, the regression testing approach performs well in the simulated economies.

4.5 Effects of Discount Factor Misspecification

We use a set of parametric discount factors (or pricing kernels) listed in Table 4 to demonstrate effects of discount factor misspecification. In every case we implement the GMM procedure to test conditional efficiency of the NYSE benchmark. Note that finite sample behavior of a GMM test with correct specification is not considered here.[23] Instead we aim at consequences of discount factor misspecification.

All the specifications are of linear form as recommended by Carhart et al. (1994) and Cochrane (1996). The first one (SP1) implies unconditional mean-variance efficiency of the benchmark, if it is free of misspecification. This specification is included here to serve as a special example of the Hansen and Richard (1987) critique. The specifications SP2, SP3, and SP4 (or their variants) have been employed by Carhart et al. (1994), Cochrane (1996), Jagannathan and Wang (1996), in their tests of conditional asset pricing models.[24] SP5 and SP6 are two generalizations of the above choices. SP5 includes SP1 through SP4 as special cases. SP6 is even more general, which does not omit any of the four state variables.

The simulation results in Table 4 show that arbitrary functional form assumptions can easily create erroneous rejections. Most notably, SP1 produces rejection probabilities above or near 90% under H_0 when the sample size $N = 600$. This shows that strong rejections can be a consequence of assuming constant conditional moments, so that a rejection of unconditional mean-variance

[23] This has been studied by Ferson and Foerster (1994) among others.

[24] Cochrane studies a two-factor model. Jagannathan and Wang consider a framework without riskless rate. We ignore such details when drawing similarities in specifications.

efficiency is not necessarily evidence against conditional mean-variance efficiency. Apparently, incorrect models for time-varying moments can also cause the problem. As the results show, SP2, SP3, and SP4 have also generated tests that reject much too often under H_0, even though these specifications do permit time-variation in conditional moments of asset returns. At 10% level, the rejection probabilities of these tests range from 40% to 80% under H_0. As expected, the problems are more serious in larger samples. At $N = 600$, the three 10% (5%) tests all reject with probability above 50% (40%) under H_0.

Meanwhile, the simulation results show that serious power problems can easily arise as well in finite samples, as a result of some specification choices. The tests based on SP5 and SP6 are poor in power, as reflected by the rejection frequencies of these tests under H_1, H_2, and H_3. At $N = 600$, there is only one case in which the power of a 10% test exceeds 60%, but in four out of the six cases the power is less than 40%. At $N = 300$, only in one case does the power of a 10% test exceed 40%.

The test results from the historical data also indicate strong effects of discount factor specification. The tests based on SP1 through SP3 give strong rejections with p-values topped by 3%. The specification SP4 yields an 11.9% p-value. On the other hand, the tests based on SP5 and SP6 do not lead to rejection and produce large p-values, 42.5% and 68.1%, respectively. The existence of the test problems implies that some specifications can produce substantially biased pricing error estimates while some can give rise to very noisy estimates. Consistent with this implication, panel (B) shows that pricing error estimates are also sensitive to specification choice. The estimated expected return errors vary wildly across the specifications, suggesting that either significant biases or substantial noises exist in the pricing error estimation.

5 Empirical Results

5.1 Set-up

We use two sets of stock portfolios as test assets. In the first set are the five NYSE size portfolios, the same as for the simulation experiments. The data are monthly excess returns from February 1947 to December 1995. The second set contains the Fama and French (1993) twenty-five size and book-to-market (BE/ME) portfolios of common stocks listed on NYSE, Amex, and Nasdaq. The data of monthly excess returns on these portfolios are from July 1963 to December 1995. For the benchmark portfolio or components of the benchmark, we use the NYSE value-weighted portfolio and the three factor mimicking portfolios of Fama and French. Descriptions and summary statistics of all the variables involved in the testing are given in Table 1.

The conditioning information set in the applications consists of four state variables: the dividend/price ratio (DPR), the default premium (DEF), the one-month Treasury bill rate (RTB), and excess return on the NYSE equally-weighted index (EWR). These four variables are selected out of a ten-variable set according to the procedure outlined in Section 4.1. The instrument vector for constructing the nonparametric tests is $z_t = (1\ x_t')'$, where x_t is the state variable vector $(DPR_t\ DEF_t\ RTB_t\ EWR_t)'$. The kernel and bandwidth inputs are also the same as in the simulation study. We use the two kernel functions and the bandwidth given in Section 4.3.

5.2 Tests of the Conditional CAPM

First, we test the conditional Sharpe-Lintner CAPM using the NYSE value-weighted portfolio as proxy for the market. Test assets are the five NYSE size portfolios. The results are reported in panel (A) of Table 5. Next we use the Fama-French market portfolio as the benchmark and the twenty-five size-BE/ME portfolios as test assets. Panel (B) of Table 5 present the test results, including six subset cases. In each of the cases, five out of the twenty-five size-BE/ME portfolios in a single size or BE/ME quintile are taken as test assets. The six quintiles are three size quintiles (small, medium, large) and three BE/ME quintiles (low, medium, high).

The regression tests reject that the NYSE market proxy is conditionally mean-variance efficient, at the p-values below 2% for the full sample period. For the subsample period from February 1947 to December 1970, the tests fail to reject at 10% level, but the p-values are fairly low (11.1% using the normal kernel). For the other subsample case, the tests reject at 10% level. In panel (B), the tests strongly reject that the Fama-French market proxy is conditionally mean-variance efficient. The p-values using the 25 size-BE/ME portfolios are 0.0%. Moreover, the tests reject at 10% level for five out of the six subset cases. The results from the subset cases suggest that portfolios of smaller size or higher BE/ME value are associated with statistically more significant pricing errors. In particular, the five portfolios in the small size quintile produce the largest test statistics with 0.0% p-values.[25]

The regression testing approach provides a simple way to look into conditional expected return errors or pricing errors. Panel (C) of Table 5 presents two cases of the regression results. In the first case (C1), the benchmark is the NYSE value-weighted portfolio, and test assets are the five NYSE size portfolios. In the second case (C2), the Fama-French market proxy is the benchmark, while the test assets are the twenty-five size-BE/ME portfolios. In these regressions, the four nonconstant regressors are in deviation form (deviation from mean),[26] so that the intercepts are just average pricing errors or biases of the conditional CAPM.

The results in (C1) show that the conditional CAPM does not have difficulty pricing average excess returns on the NYSE size portfolios. The intercepts or average pricing errors are small, only between -0.02% and 0.08%, and are statistically insignificant. However, the pricing errors are substantial in terms of volatility. In the SZ1, SZ3, and SZ5 regressions,[27] the slopes on EWR are salient, about three or four standard errors above zero. The joint significance test for slopes on EWR rejects with 0.0% p-value. These slopes are also large in economic terms, noting that the pricing error components 0.21EWR, 0.11EWR, and 0.09EWR in the three regressions have standard deviation of 1.02%, 0.53%, and 0.44%, respectively. The slopes on EWR are clearly related to size. The slope falls strictly with size from 0.21 to -0.01. To much extent, this determines a negative size pattern in the pricing error volatilities. Indeed, the pricing error standard deviation decreases

[25] In sensitivity analyses, we have considered small changes in the bandwidth around the selected values. We have also checked results using a larger set of state variables: $x_t = (DPR_t\ DEF_t\ RTB_t\ EWR_t\ TERM_t)'$. The test results are qualitatively robust. For brevity, these robustness checking results are omitted.

[26] This transformation does not affect estimation and inference for slopes on the nonconstant regressors. Nor does it alter the joint statistic and pricing error estimates. For inference about the intercepts (which is affected), we ignore estimation noise in the sample means of the regressors. This is equivalent to assuming that the regressors are observed in deviation form.

[27] For a WLS regression with dependent variable $\hat{m}_t r_{i,t+1}$, we refer to it as 'r_i regression' for short.

with size from 1 08% to 0 31%.

Panel (C2) presents a look of the regression results using the twenty-five Size-BE/ME portfolios. The joint significance test for the regression intercepts gives a strong rejection. Compared to those obtained from the NYSE size portfolios, the intercepts from the size-BE/ME portfolios are no longer trivial, ranging from -0 29% to 0 45%. The intercepts are obviously related to book-to-market equity. In every size quintile, the intercepts tend to increase with BE/ME. On average, the intercepts increase strictly from -0 16% in the lowest BE/ME quintile to 0 33% in the highest. There is, however, no clear size pattern in the intercepts. Similarly as in (C1), the joint tests for individual regressors indicate that the slopes on EWR is significant and that EWR is the only statistically significant contributor to volatilities of the pricing errors.

Does the conditional CAPM perform better than the unconditional CAPM? Panel (D) gives an answer. For the overall sample period, the small firm effect does not show up in the average pricing errors of the conditional CAPM, while this effect is prominent in those of the unconditional model. On the other hand, the conditional CAPM performs uniformly better in terms of RMSE, especially for the small to medium size portfolios.

Panel (E) presents means and standard deviations of the pricing errors estimated from the 25 size-BE/ME portfolios, with the Fama-French market proxy as the benchmark. Both the WLS regression method and the pointwise kernel method are used to estimate the pricing errors. Details of the two methods are given in Appendix C. As in (C2), the average pricing errors estimated by the pointwise kernel method have a strong positive relation to BE/ME, but no clear size pattern. On the other hand, the standard deviations based upon either of the two methods exhibit a negative relation to size, but no clear BE/ME pattern.

The positive correlations between the pricing errors and EWR seem to suggest that investors tend to overreact to stock market fluctuations. When the stock market is up (down), investors tend to set expected returns higher (lower) than the rate of return predicted by the CAPM.[28] However, it is always possible that a multifactor model may explain the CAPM pricing errors. Next we test the Jagannathan and Wang (1996) two factor model and the Fama and French (1993) three factor

[28] When we replace the state variable EWR with VWR (the value-weighted NYSE index), the results are very similar.

model.

5.3 The Labor Income Risk Factor

Jagannathan and Wang (1996) find that after including a labor income risk factor, their specification of the conditional CAPM performs substantially better in explaining cross-sectional variation in average stock returns.[29] Their labor income risk factor is motivated as a proxy for return on human capital. Alternatively, their model may be viewed as a conditional two factor model.

We test the two factor model using two data sets. Specifically, we assume that the benchmark portfolio has return of the form $(1 - \mu)RSM + \mu LBR$, where LBR is the labor income risk factor, measured as the per capita labor income growth rate. RSM is return on a stock market index. We set this index to be the NYSE value-weighted index in tests using the five NYSE size portfolios. For the second data set of excess returns on the 25 size-BE/ME portfolios, we let RSM be return on the Fama-French market portfolio.

The regression-GMM tests strongly reject the two factor model. The test results are presented in panel (A) of Table 6. The p-values obtained from both data sets are below 1%. Estimates of the parameter μ show no support for significance of the labor income risk factor. For the first data set, the estimates are trivial, only around 0.04 and less than a quarter of the standard errors. The estimates of μ from the second data set are about 0.3, close to but less than 2 standard errors.[30] The regression tests for fixed μ from panel (B) also give strong rejections, with maximums of the p-values below one percent.

Panel (B) of Table 6 presents minimums of three summary measures of the pricing errors that can be achieved with the free parameter μ. The measures are average absolute bias (AAB), average standard deviation (ASD), average root mean squared error (ARMSE). Detailed definitions are given in Appendix C. Figure 1 plots the measures as functions of μ, using the second data set.

Performance of the conditional CAPM does not improve much when including the labor income risk factor. The ARMSE cannot achieve any significant reduction with the free parameter μ. The

[29] Using Japanese stock return data but an unconditional four factor model, Jagannathan, Kubota and Takehara (1997) also find that the labor income risk factor has a significant role in explaining the cross-sectional variation in average stock returns.

[30] In contrast, Jagannathan and Wang (1996) have encountered disappearance of their stock index when including the labor income growth rate, i.e., the fraction parameter μ is almost equal to 1.0.

minimums of ARMSE in the two cases are 0.59% and 0.91%, respectively, compared to 0.61% and 0.92% for the conditional CAPM without the labor factor. For the first data set, the minimizer of AAB is close to 0 and the minimum AAB is about the same as the bias at $\mu = 0$. In contrast, the AAB from the second data set has minimum at $\mu = 0.93$, and the minimum AAB is 0.14%, which is a thirty percent reduction of the bias 0.20% at $\mu = 0$. However, Figure 1 shows that the bias reduction comes at the cost of more volatile pricing errors. As the AAB decreases over (0.80, 0.93), the ASD rises sharply.

While it does not help much to include the labor income risk factor, it does not hurt either, at least for a reasonable range of μ values.[31] For μ between 0.0 and 0.8, the three pricing error measures stay practically flat, indicating that effects of including the labor income growth rate are neutral.[32]

5.4 The Size and Book-to-Market Factors

Fama and French (1993, 1996) find that two factors, size and book-to-market, do a very good job in characterizing failure of the static CAPM in explaining cross-section of average stock returns. Can these two factors explain pricing errors of the conditional CAPM?

We test a conditional version of the Fama and French (1993) three factor model. The hypothesis is that a benchmark portfolio with excess return $FFVW + \mu_1 SMB + \mu_2 HML$ is conditionally mean-variance efficient. FFVW is excess return on the Fama-French market portfolio. SMB and HML are returns on the Fama-French mimicking portfolios for the size factor and the book-to-market factor, respectively. The tests use excess returns on the 25 size-BE/ME portfolios. The results are presented in Table 7.

The regression-GMM tests strongly reject the conditional three factor model. In panel (A), both the two-stage and the iterated procedures produce 0.0% p-values. The regression test for fixed μ in panel (B) also gives a strong rejection; the p-values that can be achieved with the two free parameters are topped by 0.1%. Estimates of μ_1 and μ_2 in panel (A), however, provide some support for statistical significance of the two factors. The parameter estimates are all more than

[31] Campbell (1993) suggested that the fraction parameter for the market portfolio may be somewhere around two-thirds. If so, [0.5; 0.8] seems to be a reasonable range for μ.

[32] A Hansen and Jagannathan (1997) measure of pricing error biases has been computed. It is also very flat.

six standard errors above zero.

Panel (B) of Table 7 reports minimums of the three summary measures of pricing errors that can be achieved with the free parameters μ_1 and μ_2. The cases with constraint $\mu_1 = 0$ or $\mu_2 = 0$ are also included. Figure 2 plots the three measures against μ_1 while letting $\mu_2 = 0$. Figure 3 presents the case in which μ_2 is free but μ_1 is fixed at zero.[33]

Used together, the size and book-to-market factors can significantly reduce either the bias measure (AAB) or the volatility measure (ASD) of the pricing errors. The minimum AAB is about a fifty percent reduction while the minimum ASD is about a twenty-five percent reduction, compared to the measures from the conditional CAPM. However, the minimizers are very different. The pair (0.40, 2.40) minimizing the AAB produces an ASD of 1.32%, much larger than the 0.89% ASD from the conditional CAPM. The minimizer (0.94, -0.06) for the ASD yields an AAB of 0.23%, worse than the CAPM bias 0.20%. What is going on here? A separate look at each of the two factors provides the answer.

The size factor can significantly reduce the pricing error volatility, but it is not effective at all for bias reduction. This is evident from (iii) of panel (B) and Figure 2. Used alone, the size factor can achieve an ASD of 0.68%, the same as the minimum from joint use of both factors. But Figure 2 shows that the factor tends to increase the AAB. On the other hand, the book-to-market factor is effective on bias reduction. Used alone, it gives minimum AAB of 0.09%, the same as that from joint use of the two factors. However, Figure 3 shows that the book-to-market factor inflates the ASD, so that the bias reduction comes at the expense of more volatile pricing errors.

The two factors jointly produce minimum ARMSE of 0.72%, about a twenty percent reduction of the ARMSE (0.92%) from the conditional CAPM. Nearly all of the drop comes from volatility reduction. In terms of this joint measure for pricing errors, the book-to-market factor is of no importance. The size factor is significant, but it does not appear to be a cure.[34]

[33] The pricing error measures in Table 6 and Table 7 and Figures 1-3 are based on the WLS regression fit. In every case, the measures based on the pointwise kernel method give rise to the same conclusions.

[34] The book-to-market factor (and the labor income risk factor) can not reduce the correlations between the CAPM pricing errors and the stock market. The size factor reduces the EWR slopes for small size portfolios but increases magnitude of the EWR slopes for large size portfolios.

6 Conclusion

In this paper we have proposed a new approach to testing conditional asset pricing models. The approach carries out a simple regression testing idea through a nonparametric discount factor. In the simulation experiments, we find that the regression test performs well for simulated monthly data of the postwar period, and the performance is reasonably robust to selection of state variables. In the empirical applications we find that the conditional CAPM exhibits volatile pricing errors that are positively correlated with the stock market. The positive correlations seem to suggest that investors tend to overreact to stock market fluctuation when they set expected returns. On the other hand, our results do not support the Fama and French (1993) three factor model and the model of Jagannathan and Wang (1996). We find that the size and book-to-market factors of Fama and French, and the labor income risk factor of Jagannathan and Wang do not explain the correlations between the CAPM pricing errors and the stock market.

Appendix A: Assumptions

There is a large statistical literature on the U-statistics introduced by Hoeffding (1948). Numerous asymptotic results for U-statistics or generalized U-statistics have been established under various statistical setups. In particular, Robinson (1989), Yoshihara (1990), and Khashimov (1993) have provided conditions and established central limit theorems of second order generalized U-statistics for β-mixing processes.

We present a set of conditions oriented to the asset pricing applications. In particular, simplicity of conditions on the bandwidth and the kernel is pursued to facilitate practical choices. In addition, restrictions on higher order moment existence are kept at a minimal level, due to concerns about possibly fat-tailed distributions of asset returns. Moreover, the β-mixing condition below requires no change for higher order extensions (e.g., to deal with third order generalized U-statistics in the test defined by (17) of Section 3). Under this set of assumptions, we provide a large sample justification for the regression testing approach.

Notations:

For a matrix of random variables $X \equiv (x_{ij})$ and any positive real ρ, define $\|X\|_\rho \equiv (\|x_{ij}\|_\rho)$ and $\|x_{ij}\|_\rho \equiv [E\,|x_{ij}|^\rho]^{1/\rho}$. The matrix $\|X\|_\rho$ is referred to as the ρ-norm of X. Let $|X|^\rho$ denote $(|x_{ij}|^\rho)$. These notations are also used for vectors. Let $\|x\|$ stand for the Euclidean norm of a vector x. For convenience, we use an inequality sign between a matrix (vector) and a scalar or between two matrices (vectors) to denote component-by-component inequalities. As usual, the information set I_t stands for a σ-field which contains in particular the σ-field generated by the data sequence $\{y_t, y_{t-1}, \cdots\}$. In addition, define

$$W_1(t,s) \equiv (r^2_{p,s+1} - r_{p,s+1} r_{p,t+1}) r_{t+1} + (r^2_{p,t+1} - r_{p,t+1} r_{p,s+1}) r_{s+1};$$
$$W_2(t,s) \equiv r^2_{p,s+1} z_t z_t' + r^2_{p,t+1} z_s z_s'.$$

A function $\phi(x)$ is said to satisfy local Lipschitz condition for some function $m(x)$ if

$$|\phi(x+\delta) - \phi(x)| < m(x)\|\delta\|.$$

Let $\nabla_{l_1,\cdots,l_j}\phi(x)$ stand for the partial derivative $\partial^j \phi(x)/\partial x_{l_1}\cdots\partial x_{l_j}$ where x_l is the l-th element of x.

Assumptions:

Assumption A1: The data sequence $\{y_{t+1}\}$ is a strictly stationary β-mixing process, and the subvector x_t has absolutely continuous distribution with density $f(x_t)$. For some $\rho > 2$, the mixing numbers β_n, $n = 1, 2, \ldots,$ satisfy $\sum_{n=1}^{\infty} n \beta_n^{(\rho-2)/\rho} < \infty$.

Assumption A2: (i) $r_{p;t+1}^2$, r_{t+1}, and $r_{p;t+1} r_{t+1}$ have finite first moment; (ii) $\|\mathcal{W}_1(t;s)\|_\rho < \infty$ and $\|\mathcal{W}_2(t;s)\|_\rho < \infty$ for all $t < s$; (iii) $\|\delta(y_{t+1})\|_\rho < \infty$ and $\|a(y_{t+1})\|_\rho < \infty$.

Assumption A3: fg_p, fg_{pp}, fg_r, fg_{pr}, and fg_{zz} satisfy the local Lipschitz condition for some $m(x)$, where $m(x_t) r_{t+1}$, $m(x_t) r_{p;t+1} r_{t+1}$, $m(x_t) r_{p;t+1}$, $m(x_t) r_{p;t+1}^2$, and $m(x_t) z_t z_t'$ have finite ρ-norm.

Assumption A4: The kernel $K(u)$ is a bounded symmetric function satisfying

$$\int K(u)du = 1;$$

$$\int |u|^j |K(u)| du < \infty \quad \text{if } 0 \le j \le k+1;$$

$$\int u_1^{l_1} \cdots u_k^{l_k} K(u) du = 0 \quad \text{if } 0 < l_1 + \cdots + l_k < k+1;$$

where u_j is the j-th element of vector u. That is, the kernel $K(u)$ is of order $k+1$.

Assumption A5: (i) The j-th partial derivatives of fg_p, fg_{pp}, fg_r, fg_{pr}, and fg_{zz} exist for all $j \le k+1$; (ii) The expectations $E[g_{pr} \nabla_{l_1;\cdots;l_j}(fg_p)]$, $E[g_r \nabla_{l_1;\cdots;l_j}(fg_{pp})]$, $E[g_{pp} \nabla_{l_1;\cdots;l_j}(fg_r)]$, $E[g_p \nabla_{l_1;\cdots;l_j}(fg_{pr})]$, $E[g_{pp} \nabla_{l_1;\cdots;l_j}(fg_{zz})]$ exist for all $j \le k+1$, where the functions and the partial derivatives are evaluated at x_t.

Assumption A6: The matrices A and Γ_0 are nonsingular.

The mixing condition in Assumption A1 restricts the amount of dependence allowed in the data sequence, permitting among other things a central limit theorem to be applied. Conditions that require β_n to vanish as a power of n do not seem to be restrictive for most financial data processes and such assumptions are commonly used in the literature. For instance, a β-mixing condition of this type is adopted by Aït-Sahalia (1996) for continuous time models of interest rates. Note that Assumption A2 is related to A1. As the moment restrictions become stronger (larger ρ), more dependence is allowable. Such a trade-off is not uncommon in establishing asymptotic results for serially correlated data (e.g., White and Domowitz 1984). Lipschitz conditions and higher order kernel assumptions, similar to Assumptions A3 and A4, have been employed by many authors (e.g., Härdle and Stoker 1989). Assumption A5 is a regular condition for asymptotic bias correction through use of a higher order kernel (see Powell et al. 1989).

For the moment conditions in A2, note first that ρ can be set arbitrarily close to 2 if β_n decays exponentially. Next, it is straightforward to verify that Assumption A2 holds for all $\rho > 2$ if the joint distribution of all the variables is normal, or lognormal, or mixture normal. For a more interesting example, consider the case in which the return vector is generated as follows:

$$\hat{r}_{t+1} = \mu(z_t) + \sigma(z_t)\varepsilon_{t+1};$$

where $\hat{r}_{t+1} = (r_{p,t+1}, r_{t+1}')'$. z_t has finite second moment and it contains x_t as a subvector. Being independent from z_t, the shock ε_{t+1} has zero mean and finite moment of any order. Then Assumption A2 holds for all $\rho > 2$ if $\mu(\cdot)$, $\sigma(\cdot)$, and $f(\cdot)$ are bounded functions. Of course, the conditional mean and standard deviation do not have to be bounded to obtain A2. For instance, $\mu(\cdot)$ and $\sigma(\cdot)$ can be polynomials of any order when z_t has finite moments of any order, which is enough to deliver A2 in the example above. These examples suggest that the moment conditions in A2 are not as restrictive as they may seem.

Appendix B: Proofs of the Theorems

We use three basic lemmas for the proofs of the theorems. Proofs of these lemmas are omitted but available upon request. The first lemma is a modified version of Yoshihara's (1976) lemma 1 which has been an indispensable tool to analyze U-statistics for β-mixing processes.

Lemma 1 (Yoshihara's Fundamental Lemma): Let $\{y_t\}$ be a strictly stationary β-mixing process with the mixing numbers β_n, $n = 1, 2, \ldots$. For any given j, $1 \le j \le m - 1$, and $t_1 < \cdots < t_m$, let ξ_{j+1}, ..., ξ_m be $m - j$ random vectors which are identical in joint distribution to $y_{t_{j+1}}, \ldots, y_{t_m}$, but independent from $y_{t_1}, \ldots, y_{t_j}$. Let $\phi(y_{t_1}, \ldots, y_{t_m})$ be a function such that

$$E[\phi(y_{t_1}, \ldots, y_{t_j}, \xi_{j+1}, \ldots, \xi_m)] = 0;$$

and

$$\sup_{1 \le t_1 < \cdots < t_m < \infty} \|\phi(y_{t_1}, \ldots, y_{t_m})\|_{\rho_1} \le M$$

for some $\rho_1 > 1$ and $M > 0$. Then for $n_j = t_{j+1} - t_j$,

$$|E\phi(y_{t_1}, \ldots, y_{t_m})| \le 4M\beta_{n_j}^{(\rho_1 - 1)/\rho_1}.$$

Define two generalized U-statistics of second order:

$$U_{1N} \equiv \frac{1}{N(N-1)} \sum_{t=1}^{N-1} \sum_{s=t+1}^{N} q_N(y_{t+1}, y_{s+1});$$

$$U_{2N} \equiv \frac{1}{N(N-1)} \sum_{t=1}^{N-1} \sum_{s=t+1}^{N} \phi_N(y_{t+1}, y_{s+1});$$

where

$$q_N(y_{t+1}, y_{s+1}) \equiv h^{-k} K\left(\frac{x_t - x_s}{h}\right) \{W_1(t,s) - [\mu_h \cdot W_2(t,s)]\};$$
$$\phi_N(y_{t+1}, y_{s+1}) \equiv h^{-k} K\left(\frac{x_t - x_s}{h}\right) W_2(t,s).$$

Note that q_N and ϕ_N are symmetric and they vary with N through h.

For application of the Hoeffding projection technique, define

$$\gamma_N(y) \equiv \int q_N(y, y_{s+1}) dF(y_{s+1});$$
$$\hat{H}_{1N} \equiv \frac{1}{2} E\gamma_N(y_{t+1}) + \frac{1}{N} \sum_{t=1}^{N} [\gamma_N(y_{t+1}) - E\gamma_N(y_{t+1})];$$
$$a_N(y) \equiv \int \phi_N(y, y_{s+1}) dF(y_{s+1});$$
$$\hat{H}_{2N} \equiv \frac{1}{2} E a_N(y_{t+1}) + \frac{1}{N} \sum_{t=1}^{N} [a_N(y_{t+1}) - E a_N(y_{t+1})];$$

where $F(y_{s+1})$ is the distribution function of y_{s+1}.

Lemma 2 below is a version of the Hoeffding projection equivalence result. This result has been obtained under a variety of regularity conditions (see Yoshihara (1976, 90), Powell et al. (1989), Robinson (1989), Khashimov (1993), and Arcones (1995) among others), producing various central limit theorems or asymptotic normality results.

Lemma 2 (Hoeffding Decomposition): Let $\{y_t\}$ be a strictly stationary β-mixing process with the mixing numbers β_n, $n = 1, 2, \ldots$, satisfying

$$\sum_{n=1}^{\infty} n \beta_n^{(\rho-2)/\rho} < \infty$$

for some $\rho > 2$. If there exists $B_N = o(\sqrt{N})$ such that $\|\gamma_N(y_{t+1})\|_\rho \le B_N$ and

$$\sup_{1 \le t < s < \infty} \|q_N(y_{t+1}, y_{s+1})\|_\rho \le B_N;$$

then

$$N E[(U_{1N} - H_{1N})(U_{1N} - H_{1N})'] = o(1).$$

Lemma 2 provides general conditions for the equivalence of generalized U-statistics to their Hoeffding projection in time-series contexts. Note that no specific contents of U_{1N}, q_N, and γ_N

are required here, i.e., the lemma holds if we use U_{2N}, κ_N, and a_N instead. Under Assumptions A1 through A5, Lemma 3 below establishes (i) equivalence of $\gamma_N(z_{t-1})$ to $\gamma(z_{t-1})$ and $a_N(z_{t-1})$ to $a(z_{t-1})$, and (ii) the limiting distribution of $\sqrt{N}H_{1N}$ is centered at zero.

Lemma 3: Let Assumptions A1 through A5 hold. Let

$$\varepsilon_{1N}(y_{t-1}) \equiv \gamma_N(y_{t-1}) - \gamma(y_{t-1});$$

$$\varepsilon_{2N}(y_{t-1}) \equiv a_N(y_{t-1}) - a(y_{t-1}).$$

(i) If $h \to 0$, then $\|\varepsilon_{1N}(y_{t-1})\|_\rho \le b_N$ for some $b_N = o(1)$ and

$$\frac{1}{\sqrt{N}}\sum_{t=1}[\varepsilon_{1N}(y_{t-1}) - E\,\varepsilon_{1N}(y_{t-1})] = o_p(1).$$

The same result also holds for $\varepsilon_{2N}(y_{t-1})$.
(ii) If $Nh^{2k+2} \to 0$, then $\sqrt{N}E(H_{1N}) = o(1)$.

Proof of Theorem 1:

From the definition of the WLS estimator $\hat{\delta}_N$, it follows that

$$\chi_N(\hat{\delta}_N - \delta) = (1 - \frac{1}{N})U_{1N};$$

and

$$\chi_N = (1 - \frac{1}{N})\iota_h \otimes U_{2N}:$$

By Assumptions A2 and A4, there exists an upper bound M such that

$$E\,|K(\frac{x_t - x_s}{h})W(t;s)|^\rho \le M$$

where $W(t;s) = W_1(t;s) - [\iota_h \otimes W_2(t;s)]\delta$. Hence

$$E\,|q_N(y_{t-1};y_{s+1})|^\rho = h^{-\rho k}E\,|K(\frac{x_t - x_s}{h})W(t;s)|^\rho \le h^{-\rho k}M:$$

This gives

$$\|q_N(y_{t-1};y_{s+1})\|_\rho^2 \le h^{-2k}M^{2/\rho} = O(h^{-2k}) = O(\frac{N}{Nh^{2k}}) = o(N);$$

since $Nh^{2k} \to \infty$.

From Assumption A2, Lemma 3, and the triangular inequality

$$\|\gamma_N(y_{t-1})\|_\rho \le \|\gamma(y_{t-1})\|_\rho + \|e_N(y_{t-1})\|_\rho;$$

it follows that $\|\gamma_N(y_{t-1})\|_\rho$ has a constant upper bound.

Thus the conditions of Lemma 2 are satisfied. Lemma 2 provides

$$\sqrt{N}\,(U_{1N} - EH_{1N}) = \sqrt{N}\,(H_{1N} - EH_{1N}) + o_p(1). \tag{19}$$

Repeating the above arguments for κ_N and a_N to apply Lemma 2 to U_{2N}, we have

$$U_{2N} = H_{2N} + o_p(1).$$

By Lemma 3 (i),

$$\begin{aligned} H_{2N} &= \frac{1}{2}E\,a(y_{t+1}) + \frac{1}{N}\sum_{t=1}^{N}[a(y_{t+1}) - E\,a(y_{t+1})] + o_p(1) \\ &= \frac{1}{2}E\,a(y_{t+1}) + o_p(1). \end{aligned}$$

Since $E\,a(y_{t+1}) = 2E(w_t z_t z_t')$, it then follows that $\hat{A}_N \xrightarrow{p} A$.

Lemma 3 (i) also gives

$$\begin{aligned} \sqrt{N}\,(H_{1N} - EH_{1N}) &\equiv \frac{1}{\sqrt{N}}\sum_{t=1}^{N}[\gamma_N(y_{t+1}) - E\gamma_N(y_{t+1})] \\ &= \frac{1}{\sqrt{N}}\sum_{t=1}^{N}[\gamma(y_{t+1}) - E\gamma(y_{t+1})] + o_p(1). \end{aligned} \tag{20}$$

By (19), (20), and Lemma 3 (ii),

$$\sqrt{N}\,U_{1N} = \frac{1}{\sqrt{N}}\sum_{t=1}^{N}[\gamma(z_{t+1}) - E\gamma(z_{t+1})] + o_p(1).$$

Note next that $E\gamma(y_{t+1}) = 0$, since it follows from the Law of Iterated Expectations that

$$\begin{aligned} E\eta(y_{t+1}) &= 2E(w_t e_{t+1}) \\ E\,a(y_{t+1}) &= 2E(w_t z_t z_t') \end{aligned}$$

where $e_{t+1} = (e_{1,t+1} \cdots e_{n,t+1})' - z_t$. Thus $E\eta(y_{t+1}) - [\iota_h \otimes E\,a(y_{t+1})]\delta = 0$.

By the central limit theorem of Doukhan et al. (1994)[35],

$$\frac{1}{\sqrt{N}}\sum_{t=1}^{N}\gamma(y_{t+1}) \xrightarrow{d} N(0, \Gamma)$$

where $\Gamma = \sum_{j=-\infty}^{\infty} E\gamma(y_t)\gamma(y_{t-j})'$. This completes the proof.

[35] The authors have improved the classical central limit theorems (e.g., theorem 18.5.3) of Ibragimov and Linnik (1971) and they proved that if $2 < p < \infty$, $E|X|^p < \infty$, and $\sum_{n=1}^{\infty} \alpha_n n^{2/(p-2)} < \infty$, then $n^{-1/2}\sum_{i=1}^{n}(X_i - EX_i)$ converges to a centered normal random vector. Note that $\sum_{n=1}^{\infty} n\,\beta_n^{(\rho-2)/\rho} < \infty$ implies that $\sum_{n=1}^{\infty} \beta_n n^{\rho/(\rho-2)} < \infty$. Thus it is obvious that Assumptions A1 and A2 imply the conditions for the central limit theorem; they are also sufficient for Γ to be finite.

Proof of Theorem 2: Proof of part (ii): Define

$$\tilde\gamma_N(y_{t+1}) = \hat\gamma_N(y_{t+1}) - [\nabla h \cdot \hat a_N(y_{t+1})]'\delta.$$

Since $\hat\delta_N \to^p \delta$, it suffices for part (ii) to show that

$$\frac{1}{N}\sum_{1}^{N-j} \tilde\gamma_N(y_{t+1})\tilde\gamma_N(y_{t+j+1})' \to^p \Gamma_j.$$

Note first that

$$\tilde\gamma_N(y_{t+1}) = \frac{1}{N}\sum_{s=1}^{N} q_N(y_{t+1}; y_{s+1}).$$

Thus

$$\tilde\gamma_N(y_{t+1}) - \gamma_N(y_{t+1}) = N^{-1}\sum_{s=1}^{N} J_N(t;s)$$

with $J_N(t;s) \equiv q_N(y_{t+1}; y_{s+1}) - \gamma_N(y_{t+1})$.

As shown above, $\|q_N(y_{t+1}; y_{s+1})\|_\rho$ has an upper bound of order $o(\sqrt{N})$ and $\|\gamma_N(y_{t+1})\|_\rho$ has a constant upper bound. Thus by the Minkowski's inequality, $\|J_N(t;s)\|_\rho$ has an upper bound of order $o(\sqrt{N})$. So there exists $M_N = o(N)$ such that

$$\|J_N(t;s_1)J_N(t;s_2)'\|_{\rho/2} \le \|J_N(t;s_1)\|_\rho \|J_N(t;s_2)'\|_\rho \le M_N.$$

Moreover, note that for $s_1 < s_2$

$$\int J_N(t;s_1)J_N(t;s_2)'dF(z_{j+1}) = 0;$$

where $j = s_1$ if $s_1 < t$, and $j = s_2$ otherwise.

Applying Lemma 1, we have

$$\begin{aligned}
&N^2 E\{[\tilde\gamma_N(y_{t+1}) - \gamma_N(y_{t+1})][\tilde\gamma_N(y_{t+1}) - \gamma_N(y_{t+1})]'\} \\
&\le \sum_{s_1=1}^{N} |E J_N(t;s_1)J_N(t;s_1)'| + 2\sum_{1\le s_1<s_2\le N} |E J_N(t;s_1)J_N(t;s_2)'| \\
&\le N M_N + 8N M_N \sum_{n=1}^{N-1} \alpha_n^{(\rho-2)/\rho} = o(N^2).
\end{aligned}$$

So there exists $b_{1,N} = o(1)$ such that $E|\tilde\gamma_N(y_{t+1}) - \gamma_N(y_{t+1})|^2 \le b_{1,N}$. Combined with Lemma 3, this gives $E|\tilde\gamma_N(y_{t+1}) - \gamma(y_{t+1})|^2 \le b_{2,N}$ for some $b_{2,N} = o(1)$.

Since $E|\tilde\gamma_N(y_{t+1})|^2$ and $E|\gamma(y_{t+1})|^2$ have constant upper bounds, it follows that for any fixed j there exists $b_{3,N} = o(1)$ such that

$$E|\tilde\gamma_N(y_{t+1})\tilde\gamma_N(y_{t+j+1})' - \gamma(y_{t+1})\gamma(y_{t+j+1})'| \le b_{3,N}.$$

Chebyshev's inequality thus yields

$$\frac{1}{N-j}\sum_{t=1}^{N-j}\hat{\gamma}_N(y_{t+1})\hat{\gamma}_N(y_{t+j+1})' - \frac{1}{N}\sum_{t=1}^{N-j}\gamma(y_{t+1})\gamma(y_{t+j+1})' = o_p(1).$$

Now note that $\beta_n = o(n^{-2\rho/(\rho-2)})$, since

$$\sum_{j=n}^{2n} j\beta_j^{(\rho-2)/\rho} \geq \beta_{2n}^{(\rho-2)/\rho}\sum_{j=n}^{2n} j = \beta_{2n}^{(\rho-2)/\rho}O(n^2)$$

and $\sum_{j=n}^{2n} j\beta_j^{(\rho-2)/\rho} = o(1)$ by Assumption A1. Thus the strong mixing coefficients of the data process satisfy $\alpha_n = o(n^{-r/(r-2)})$ for some $2 < r < \rho$. Note further that from Assumption A2, $\|\gamma(y_{t+1})\gamma(y_{t+j+1})'\|_{\rho/2} < \infty$. Thus by Lemma 2.1 and Theorem 2.3 of White and Domowitz (1984) (or McLeish's (1975) Theorem 2.10),

$$\frac{1}{N}\sum_{t=1}^{N-j}\gamma(y_{t+1})\gamma(y_{t+j+1})' = E[\gamma(y_{t+1})\gamma(y_{t+j+1})'] + o_p(1).$$

This establishes part (ii) of the theorem that $\hat{\Gamma}_j$ is consistent for Γ_j.

Proof of part (i): Given (1) and (2) of Section 2,

$$\begin{aligned} E[\eta(y_{t+1})|I_t] &= E(w_t e_{t+1}|I_t) + E(w_t e_{t+1}|x_t) \\ E[a(y_{t+1})|I_t] &= w_t z_t z_t' + w_t g_{zz}(x_t) \end{aligned}$$

where $e_{t+1} = (e_{1,t+1} \cdots e_{n,t+1})' \otimes z_t$. On the other hand, (6) of Section 2 implies

$$\begin{aligned} E(w_t e_{t+1}|I_t) - [\iota_h \otimes (w_t z_t z_t')]\delta = 0; \\ E(w_t e_{t+1}|x_t) - [\iota_h \otimes w_t g_{zz}(x_t)]\delta = 0. \end{aligned}$$

Thus $E[\gamma(y_{t+1})|I_t] = 0$ if (6) holds. It follows that $\Gamma_j = 0$ for $j \neq 0$.

Given (1) and (2), the null hypothesis of conditional efficiency implies (6). So $\hat{\Omega}_N \xrightarrow{p} \Omega$. Then with $\delta = 0$, Theorem 1 delivers the result: $N\hat{\delta}_N'\hat{\Omega}_N^{-1}\hat{\delta}_N \xrightarrow{d} \chi^2(qn)$.

Appendix C: Pricing Error Measures and Two Estimation Methods

Three summary measures of conditional expected return errors or pricing errors are used in applications. They are average absolute bias (AAB), average standard deviation (ASD), and average root mean squared error (ARMSE). Let $\varepsilon_{i,t}$ be time-t conditional expected return error associated with $E_t r_{i,t+1}$, for $t = 1, \cdots, N$, and for $i = 1, \cdots, n$. For every i, let B_i, SD_i, and $RMSE_i$ be the sample mean, the sample standard deviation, and the sample root mean squared error of the series

$\varepsilon_{i,1}, \varepsilon_{i,2}, \cdots, \varepsilon_{i,N}$. Then the three summary measures for cross-section of the n pricing error series are:

$$
\begin{aligned}
AAB &= \frac{1}{n}\sum_{i=1}^{n} |B_i| \\
ASD &= \frac{1}{n}\sum_{i=1}^{n} SD_i \\
ARMSE &= \frac{1}{n}\sum_{i=1}^{n} RMSE_i.
\end{aligned}
$$

AAB is a measure for average pricing error or bias of the model. ASD is for pricing error volatility. $ARMSE$ is a joint measure for both bias and volatility of the pricing errors.

These summary measures are plain and simple. Yet we do not directly observe $\varepsilon_{i,t}$. In applications, we use two methods to estimate the pricing errors. The first approach estimates $\varepsilon_{i,t}$ by

$$\hat{\varepsilon}_{i,t} = z_t' \hat{\delta}_i$$

where $z_1 = (1 \;\; x_t')'$, and $\hat{\delta}_i$ is defined by (7) of Section 2. This approach is referred to as the WLS regression method. The second approach uses the following estimator

$$\hat{\varepsilon}_{i,t} = \hat{g}_i(x_t) - \hat{g}_p(x_t)\hat{g}_{ip}(x_t)/\hat{g}_{pp}(x_t)$$

where $\hat{g}_p(x)$ and $\hat{g}_{pp}(x)$ are defined as in Section 2, and

$$
\begin{aligned}
\hat{g}_i(x) &= N^{-1}h^{-k}\hat{f}(x)^{-1}\sum_{s=1}^{N} K\left(\frac{x - x_s}{h}\right) r_{i,s+1}; \\
\hat{g}_{ip}(x) &= N^{-1}h^{-k}\hat{f}(x)^{-1}\sum_{s=1}^{N} K\left(\frac{x - x_s}{h}\right) r_{p,s+1}\, r_{i,s+1}.
\end{aligned}
$$

For this method, we use the normal kernel of Section 4.3, with bandwidth $h = N^{-\frac{1}{k+4}}$. This approach is referred to as the pointwise kernel method.

References

Aït-Sahalia, Yacine, 1996, "Nonparametric Pricing of Interest Rate Derivative Securities", Econometrica 64, 527-560.

Arcones, Miguel A., 1995, "On the Central Limit Theorem for U-statistics Under Absolute Regularity", Statistics & Probability Letters 24, 245-249.

Bansal, Ravi, David A. Hsieh, and S. Viswanathan, 1993, "A New Approach to International Arbitrage Pricing", Journal of Finance 48, 1719-1747.

Bansal, Ravi, and S. Viswanathan, 1993, "No Arbitrage and Arbitrage Pricing: A New Approach", Journal of Finance 48, 1231-1262.

Bollerslev, T., R.F. Engle and J.M. Wooldridge, 1988, "A Capital Asset Pricing Model with Time-varying Covariances", Journal of Political Economy, 96, 116-131.

Campbell, John Y., 1993, "Understanding Risk and Return", NBER Working Paper, No. 4554.

Carhart, M.M., R.J. Krail, R.L. Steven, and K.D. Welch, 1994, "Testing the Conditional CAPM", Working Paper, Graduate School of Business, University of Chicago.

Chapman, David A., 1997, "Approximating the Asset Pricing Kernel", Journal of Finance 63, 1383-1410.

Cochrane, John H., 1996, "A Cross-Sectional Test of an Investment-Based Asset Pricing Model", Journal of Political Economy 104, 572-621.

Davidian M., and R J. Carroll, 1987, "Variance Function Estimation", Journal of American Statistical Association 82, 1079-1091.

Doukhan, P., P. Massart and E. Rio, 1994, "The Functional Central Limit Theorem for Strongly Mixing Processes", Ann. Inst. H. Poincare Probab. Statist. 30, 63-82.

Dybvig, P.H., and S.A. Ross, 19855, "Differential Information and Performance Measurement Using a Security Market Line", Journal of Finance 40, 383-400.

Fama, Eugene F. and Kenneth R. French, 1993, "Common Risk Factors in the Returns on Bonds and Stocks", Journal of Financial Economics 33, 3-56.

Fama, Eugene F. and Kenneth R. French, 1996, "Multifactor Explanations of Asset Pricing Anomalies", Journal of Finance 51, 55-84.

Ferson, Wayne E., 1995, "Theory and Empirical Testing of Asset Pricing Models", in R. Jarrow, W. Ziemba, and V. Maksimovic, eds., The Finance Handbook, North Holland.

Ferson, Wayne E., Shmuel Kandel, and Robert F. Stambaugh, 1987, "Tests of Asset Pricing with Time-varying Expected Risk Premiums and Market Betas", Journal of Finance 42, 201-220.

Ferson, Wayne E. and Stephen R. Foerster, 1994, "Finite Sample Properties of the Generalized Method of Moments in Tests of Conditional Asset Pricing Models", Journal of Financial Economics 36, 29-55.

Ghysels, Eric, 1997, "On Stable Factor Structures in the Pricing of Risk: Do Time Varying Betas Help or Hurt?", forthcoming on Journal of Finance.

Gibbons, Michael R., and Wayne E. Ferson, 1985, "Testing Asset Pricing Models with Changing Expectations and An Unobservable Market Portfolio", Journal of Financial Economics 14, 217-236.

Hansen, Lars Peter, 1982, "Large Sample Properties of Generalized Method of Moments Estimators", Econometrica 50, 1029-1054.

Hansen, Lars Peter and Ravi Jagannathan, 1997, "Assessing Specification Errors in Stochastic Discount Factor Models", Journal of Finance 62, 557-590.

Hansen, Lars Peter, and Scott F. Richard, 1987, "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models", Econometrica 55, 587-613.

Härdle, W., 1990, Applied Nonparametric Regression, Cambridge University Press.

Härdle, Wolfgang, and Thomas M. Stoker, 1989, "Investigating Smooth Multiple Regression by the Method of Average Derivatives", Journal of the American Statistical Association 84, 986-995.

Harvey, Campbell R., 1989, "Time-Varying Conditional Covariances in Tests of Asset Pricing Models", Journal of Financial Economics 24, 289-317.

Harvey, Campbell R., 1991, "The Specification of Conditional Expectations", Working Paper, Fuqua School of Business, Duke University.

He, J., R. Kan, L. Ng, and C. Zhang, 1996, "Tests of the Relations Among Marketwide Factors, Firm-Specific Variables, and Stock Returns Using a Conditional Asset Pricing Model", Journal of Finance 60, 1891-1908.

Hoeffding, W., 1948, "A Class of Statistics with Asymptotically Normal Distribution", Annals of Mathematical Statistics 19, 293-325.

Ibragimov, I.A. and Yu. V. Linnik, 1971, Independent and Stationary Sequences of Random Variables (Wolters-Noordhoff, Groningen, The Netherlands).

Jagannathan, Ravi and Zhenyu Wang, 1996, "The Conditional CAPM and the Cross-Section of Expected Returns", Journal of Finance 51, 3-53.

Jagannathan, Ravi, Keiichi Kubota, and Hitoshi Takehara, 1997, "Relationship Between Labor-Income Risk and Average Return: Empirical Evidence From the Japanese Stock Market", working paper.

Kan, Raymond, and Chu Zhang, 1997, "Two-pass Tests of Asset Pricing Models with Useless Factors", forthcoming on Journal of Finance.

Khashimov, Sh. A., 1993, "The Central Limit Theorem for Generalized U-Statistics for Weakly Dependent Vectors", Theory of Probability and Applications 38, 456-469.

Lee, Byung-Joo, 1992, "A Heteroskedasticity Test Robust to Conditional Mean Misspecification", Econometrica 60, 159-171.

Lintner, John, 1965, "The Valuation of Risk Assets and the Selection of Risky Investments in Stock portfolios and Capital Budgets", Review of Economics and Statistics 47, 13-37.

MacKinlay, A. Craig, 1995, "Multifactor Models Do Not Explain Deviations from the CAPM", Journal of Financial Economics 38, 3-28.

McLeish, D.L., 1975, "A Maximal Inequality and Dependent Strong Laws", Annals of Probability 3, 829-839.

Pagan, A., and G.W. Schwert, 1990, "Alternative Models for Conditional Stock Volatility", Journal of Econometrics 45, 267-290.

Powell, James L., James H. Stock, and Thomas M. Stoker, 1989, "Semiparametric Estimation of Index Coefficients", Econometrica 57, 1403-1430.

Robinson, P.M., 1989, \Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series", Review of Economic Studies 56, 511-534.

Schucany, W.R. and J.P. Summers, 1977, \Improvement of Kernel-Type Density Estimators", Journal of the American Statistical Association 72, 420-423.

Shanken, Jay, 1990, \Intertemporal Asset Pricing: An Empirical Investigation", Journal of Econometrics 45, 99-120.

Sharpe, William F., 1964, \Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk", Journal of Finance 19, 425-442.

Silverman, B.W., 1986, Density Estimation for Statistics and Data Analysis, London: Chapman and Hall.

White, Halbert and Ian Domowitz, 1984, \Nonlinear Regression with Dependent Observations", Econometrica 52, 143-161.

Yoshihara, Ken-ichi, 1976, \Limiting Behavior of U-Statistics for Stationary, Absolutely Regular Processes", Z. Wahrsch. Verw. Gebiete 35, 237-252.

Yoshihara, Ken-ichi, 1990, \Limiting Behavior of Generalized Quadratic Forms Generated by Absolutely Regular Sequences", Yokohama Mathematical Journal 37, 109-123.

Table 1: Summary Statistics for the Data

Monthly observations from January 1947 to December 1995 for the following asset returns and instrument variables are obtained from CRSP. The returns are arithmetic nominal rates of return in excess of the one-month Treasury bill rate, measured in percent (multiplied by 100). VWR denotes the excess return on the value-weighted portfolio of NYSE common stocks. SZN refers to the excess return on the value-weighted portfolio of the N th size decile NYSE stocks. LBR refers to the growth rate (in percent) on per capita personal labor income, which is constructed as $1 + LBR_t = (L_{t-1} + L_{t-2})/(L_{t-2} + L_{t-3})$, where L_{t-1} denotes the per capita labor income for month $t-1$.

Variable	Mean	Std. Dev.	First Auto.	Cross Correlations					
SZ1	1.01	6.44	0.12						
SZ3	0.82	5.39	0.14	0.94					
SZ5	0.77	4.94	0.13	0.89	0.97				
SZ7	0.75	4.66	0.12	0.84	0.94	0.97			
SZ9	0.71	4.33	0.07	0.77	0.88	0.92	0.97		
VWR	0.65	4.06	0.04	0.73	0.85	0.90	0.95	0.97	
LBR	0.49	0.52	0.49	0.00	-0.02	-0.03	-0.05	-0.08	-0.07
DPR	4.03	1.06	0.98						
DEF	0.92	0.43	0.97	0.19					
RTB	4.64	2.99	0.97	0.00	0.64				
EWR	0.77	4.84	0.14	-0.04	0.12	-0.11			
TERM	2.22	1.47	0.91	-0.12	0.39	-0.10	0.11		

DPR is the dividend yield (in percent) on the NYSE value-weighted index, measured as the sum of previous 12 months' dividend payments divided by the level of the index. DEF is the Baa-rated corporate bond yield minus that of the Aaa-rated bond. RTB is the 1-month T-bill yield. EWR is the excess return on the NYSE equally-weighted index. TERM is the Aaa-rated corporate bond yield minus the 1-month T-bill yield. All the bond or bill yields are annualized and measured in percent.

Presented below are means and standard deviations of excess returns on the twenty-five size and book-to-market portfolios of Fama and French (1993). The sample period is from July 1963 to December 1995. SZ1 through SZ5 stand for the five size quintiles (from small to large), while BM1 through BM5 stand for the five book-to-market equity quintiles (from low to high).

	Means					Standard Deviations				
	BM1	BM2	BM3	BM4	BM5	BM1	BM2	BM3	BM4	BM5
SZ1	0.29	0.72	0.80	0.95	1.08	7.48	6.58	5.98	5.69	5.99
SZ2	0.51	0.71	0.89	0.93	1.06	6.98	6.09	5.57	5.12	5.82
SZ3	0.46	0.65	0.74	0.86	1.06	6.37	5.38	4.99	4.68	5.37
SZ4	0.53	0.42	0.64	0.80	1.03	5.74	5.15	4.86	4.71	5.52
SZ5	0.42	0.43	0.41	0.59	0.74	4.73	4.52	4.20	4.12	4.81

Variable	Mean	Std. Dev.	First Auto.	Correlations	
FFVW	0.48	4.32	0.05		
SMB	0.25	2.84	0.17	0.31	
HML	0.42	2.53	0.20	-0.37	-0.10

The lower panel is for the Fama-French mimicking portfolios for three stock market factors. FFVW is excess return on a value-weighted portfolio of stocks listed on NYSE, AMEX, and NASDAQ that Fama and French use to proxy for the market portfolio. SMB is return on a zero-investment portfolio constructed to mimic the size factor. HML is return on a zero-investment portfolio constructed to mimic the book-to-market factor. For more details on the twenty-five size-BE/ME portfolios and the factor mimicking portfolios, see Fama and French (1993).

Table 2: The Simulated Economies

The forecasting variables DPR, DEF, RTB, EWR, and TERM are generated through a VAR(1) model. Let $y_1 = \ln(DPR)$, $y_2 = \ln(DEF)$, $y_3 = \ln(RTB)$, $y_4 = EWR$, and $y_5 = TERM$. The 5×1 vector y_t follows the process

$$y_t - \mu = \Phi(y_{t-1} - \mu) + \epsilon_t;$$

where μ is the mean vector of y_t and ϵ_t is normally distributed. The process is estimated from the historical data and parameter estimates become the `true' parameter values in simulations.

The conditional distribution of the excess returns SZ1 through SZ9 and VWR is multivariate normal in these economies. The conditional correlation matrix is assumed to be constant and equal to the sample correlation matrix of the returns given in Table 1. The time-t conditional standard deviations of the time-(t+1) returns take the form:

$$\sigma_{i,t} = |a_{i,0} + a_{i,1}DPR_t + a_{i,2}DEF_t + a_{i,3}RTB_t + a_{i,4}EWR_t|.$$

The parameter values are set to be the estimates obtained by a two step procedure described in Section 4.2.

The conditional expected return of the benchmark takes the form:

$$E_t(VWR_{t+1}) = b_0 + b_1 DPR_t + b_2 DEF_t + b_3 RTB_t + b_4 EWR_t.$$

Conditional expected excess returns, $E_t r_{i,t+1}$, on the five test portfolios take the following forms under H_1, H_2, and H_3, respectively.

H_1: $b_{i,0} + b_{i,1}DPR_t + b_{i,2}DEF_t + b_{i,3}RTB_t + b_{i,4}EWR_t$.

H_2: $b_{i,0} + b_{i,1}\ln(DPR_t) + b_{i,2}\ln(DEF_t) + b_{i,3}\ln(RTB_t) + b_{i,4}EWR_t$.

H_3: $b_{i,0} + b_{i,1}DPR_t + b_{i,2}DEF_t + b_{i,3}RTB_t + b_{i,4}EWR_t + b_{i,5}\sigma^2_{i,t}$.

In each case, regression estimates become the `true' parameter values in simulation. All the estimates are obtained with the data from January 1947 to December 1995.

The summary measures of conditional alphas (i.e., conditional expected return errors from the beta-pricing equation) under H_1, H_2, and H_3 given in panel (B2) are calculated from a simulated sequence of 100,000 data points.

(A) The VAR(1) Process of the Forecasting Variables

The mean vector μ	1.36	-0.17	1.29	0.77	2.22
The Φ matrix of coefficients	0.99	-0.01	0.01	0.00	0.00
	0.06	0.95	0.03	0.00	0.00
	-0.02	-0.03	0.98	0.00	0.01
	1.76	0.86	-0.67	0.11	0.35
	-0.08	0.31	-0.03	-0.02	0.88
The standard deviations of ε	0.04	0.07	0.15	4.71	0.61
The correlation matrix of ε	1.00				
	-0.05	1.00			
	0.04	-0.07	1.00		
	-0.87	0.05	-0.03	1.00	
	0.02	0.07	-0.71	-0.02	1.00

(B) Conditional Moments of the Excess Returns

Panel (B1)

	Conditional Standard Deviations of the Excess Returns						Conditional Mean
Regressors	SZ1	SZ3	SZ5	SZ7	SZ9	VWR	VWR
in.	2.91	2.80	2.95	2.91	2.46	2.24	-0.79
DPR	0.30	0.31	0.19	0.09	0.15	0.21	0.29
DEF	1.72	0.50	0.46	0.61	0.67	0.63	2.21
RTB	0.08	0.15	0.16	0.14	0.12	0.07	-0.38
EWR	-0.10	-0.17	-0.18	-0.19	-0.20	-0.17	0.00

Panel (B2)

	Conditional Expected Returns of the Size Portfolios under H_1, H_2, and H_3						Conditional Alphas		
	int.	DPR	DEF	RTB	EWR	¾^2_i	Mean	Std. Dev.	First Auto.
H_1:									
SZ1	-1 05	0 26	3 26	-0 46	0 16		0 16	0 .72	0 23
SZ3	-1 45	0 34	3 06	-0 44	0 11		0 06	0 58	0 36
SZ5	-1 30	0 32	2 85	-0 41	0 09		0 09	0 49	0 36
SZ7	-1 20	0 31	2 99	-0 45	0 04		0 11	0 35	0 69
SZ9	-1 07	0 36	2 37	-0 40	0 00		0 08	0 13	0 85
H_2:									
		(ln)	(ln)	(ln)					
SZ1	1 57	0 .72	2 04	-1 04	0 19		0 06	1 04	0 44
SZ3	0 .75	1 16	1 89	-1 00	0 14		-0 01	0 89	0 52
SZ5	0 86	1 03	1 83	-0 99	0 11		-0 04	0 81	0 60
SZ7	1 40	0 85	2 03	-1 17	0 07		-0 01	0 67	0 .70
SZ9	0 91	1 03	1 58	-1 05	0 03		-0 03	0 59	0 85
H_3:									
SZ1	0 22	-1 06	-4 46	-0 84	0 59	0 338	0 10	1 00	0 32
SZ3	-1 62	-0 19	2 20	-0 .71	0 40	0 155	0 12	0 .73	0 28
SZ5	-1 51	0 20	2 60	-0 51	0 18	0 053	0 04	0 50	0 34
SZ7	-1 32	0 28	2 86	-0 48	0 08	0 023	0 10	0 36	0 65
SZ9	-1 05	0 37	2 43	-0 39	-0 01	-0 009	0 07	0 13	0 80

Table 3: The Simulation Results

This table presents simulation results on performances of the nonparametric tests. The state variable vector is $x_t = (DPR_t, DEF_t, RTB_t, EWR_t)'$ except for panel (C). The instrument vector is $z_t = (1, x_t')'$. Both the normal kernel (n.k.) and the higher order kernel (h.k.) are used for panels (A) and (B). For the rest panels, the results are based on the normal kernel. Details of the kernel functions and the bandwidth are given in Section 4.3. All the simulation results are based on 1,000 replications.

Panel (A) records the % rejections of conditional efficiency of the value-weighted index in the nonparametric regression testing.

Panel (B) records the % rejections of the three tests for sample size $N = 600$, under H_0 at 5%, 10%, 15%, 20%, 25%, 50%, and 75% significance levels.

Panel (C) presents the tests using a larger set of state variables. The redundant variable TERM is added to the set. The test procedures are otherwise identical to those in (A).

Panel (D) presents the regression tests subject to misselection of state variables. In the cases (1) through (4), one of the true state variables is replaced by the variable TERM. In the cases (5) through (8), one of the state variables is omitted in constructing the tests.

Panel (E) contains means and standard deviations of components of the nonparametric weighted least squares estimator $\hat{\delta}_N$ under the null H_0 and the alternative H_1. Here the nonconstant regressors are in deviation form, i.e., deviation from mean.

Panel (F) tests a two-factor hypothesis that a linear combination of the value-weighted index and the first decile portfolio, with excess return $(1 - \theta)VWR + \theta SZ1$, is conditionally mean-variance efficient. The regression-GMM approach of Section 2.4 is implemented. The weighting matrix is the inverse of $\hat{\Omega}_N(\theta)$ that is being updated through iteration. The matrix is evaluated at zero ($\theta = 0$) for the first stage estimation, and evaluated at the first stage estimate of θ for the second stage estimation, and so on.

(A) % Rejections in the Nonparametric Testing

Kernel Function	H_0 10%	H_0 5%	H_1 10%	H_1 5%	H_2 10%	H_2 5%	H_3 10%	H_3 5%
			Sample Size N = 300					
n.k.	5.6	1.8	66.3	53.3	70.2	57.0	80.5	70.0
h.k.	3.6	1.6	55.4	40.4	65.9	54.4	70.7	59.0
			Sample Size N = 600					
n.k.	10.5	4.7	97.8	95.5	97.2	94.1	99.4	98.7
h.k.	4.5	2.1	92.1	83.6	95.1	90.7	96.6	93.9

(B) % Rejections under H_0

(The test with n.k. and N = 600)

5%	10%	15%	20%	25%	50%	75%
4.7	10.5	16.2	21.9	27.1	54.6	80.2

(C) Tests Using a Larger State Variable Set

$$x_t = (DPR_t, DEF_t, RTB_t, EWR_t, TERM_t)'$$

Sample Size	H_0 10%	H_0 5%	H_1 10%	H_1 5%	H_2 10%	H_2 5%	H_3 10%	H_3 5%
300	4.2	2.4	52.7	40.1	57.5	45.2	67.1	56.0
600	6.1	1.9	93.5	88.0	93.0	88.3	98.0	95.3

(D) Sensitivity to State Variable Selection

% Rejections (with N = 600)

State Variables	H_0 10%	H_0 5%	H_1 10%	H_1 5%	H_2 10%	H_2 5%	H_3 10%	H_3 5%
(1) DEF,RTB,EWR,TERM	8 9	3 6	96 0	92 3	96 1	92 8	98 .7	97 2
(2) DPR,RTB,EWR,TERM	7 5	3 0	92 3	88 6	95 0	91 .4	96 8	93 9
(3) DPR,DEF,EWR,TERM	6 5	3 0	91 .7	85 .7	90 2	83 .4	96 .7	93 8
(4) DPR,DEF,RTB,TERM	10 .4	4 5	91 0	85 .4	90 .4	83 5	96 1	93 5
(5) DEF,RTB,EWR	13 6	7 .7	93 .4	88 5	92 0	87 .7	97 5	95 2
(6) DPR,RTB,EWR	11 .4	4 9	84 6	76 0	91 1	84 3	91 9	86 6
(7) DPR,DEF,EWR	8 9	4 5	86 6	77 8	82 .4	73 5	93 1	88 9
(8) DPR,DEF,RTB	14 .7	8 .4	96 5	92 5	95 1	89 2	98 8	97 1

(E) The WLS Regression Estimation

Mean and Standard Deviation of Regression Coefficient Estimates under H_0 with Sample Size $N = 600$

	Mean of $\hat{\beta}_N$					Std. Dev. of $\hat{\beta}_N$				
	Int.	DPR	DEF	RTB	EWR	Int.	DPR	DEF	RTB	EWR
SZ1	0 01	-0 02	0 36	-0 07	0 01	0 17	0 28	0 84	0 11	0 04
SZ3	-0 01	-0 01	0 36	-0 07	0 01	0 12	0 21	0 62	0 08	0 03
SZ5	-0 01	-0 02	0 36	-0 07	0 01	0 09	0 18	0 53	0 07	0 03
SZ7	-0 01	-0 02	0 36	-0 07	0 01	0 06	0 14	0 44	0 05	0 02
SZ9	-0 01	-0 02	0 35	-0 07	0 01	0 05	0 12	0 38	0 05	0 02

Mean and Standard Deviation of Regression Coefficient Estimates under H_1 with Sample Size $N = 600$

	Mean of $\hat{\beta}_N$					Std. Dev. of $\hat{\beta}_N$				
	Int.	DPR	DEF	RTB	EWR	Int.	DPR	DEF	RTB	EWR
SZ1	0 15	-0 06	0 .75	-0 09	0 12	0 18	0 29	0 86	0 11	0 04
SZ3	0 02	0 00	0 83	-0 08	0 09	0 13	0 21	0 63	0 08	0 03
SZ5	0 04	0 00	0 .70	-0 07	0 07	0 11	0 18	0 54	0 07	0 03
SZ7	0 05	0 01	0 81	-0 10	0 04	0 08	0 14	0 44	0 06	0 02
SZ9	0 04	0 04	0 41	-0 08	0 01	0 06	0 12	0 38	0 05	0 02

(F) The Regression-GMM Test

	Sample Size N = 300				Sample Size N = 600			
	% Rejections		Estimate of μ		% Rejections		Estimate of μ	
	10%	5%	Mean	S.D.	10%	5%	Mean	S.D.
H_0 :								
2-stage	5 5	2 3	0 09	0 22	10 0	4 6	0 07	0 16
iterated	7 5	3 5	0 15	0 35	11 9	6 2	0 10	0 22
H_1 :								
2-stage	59.7	46.4	0 21	0 23	97 0	93.7	0 22	0 16
iterated	63 9	49.4	0 35	0.40	97.4	94 8	0 36	0 29
H_2 :								
2-stage	53 6	38 6	0 38	0 37	91 0	84.4	0.48	0 23
iterated	58 2	44 6	0 61	0 62	94 6	90 1	0 80	0 32
H_3 :								
2-stage	77 1	65 9	0 15	0 32	99 2	98 1	0 15	0 19
iterated	78.7	67 6	0 27	0 61	99 5	98.7	0 28	0.40

Table 4: Effects of Discount Factor Misspecification

The following functional form specifications of a discount factor m_t are used to test the conditional efficiency of the NYSE value-weighted index. In each case, the moment condition $E\, m_t r_t = 0$ is tested, where r_t is a vector of twenty-five scaled excess returns: the five excess asset returns SZ1, SZ3, SZ5, SZ7, SZ9, multiplied by the constant one and the four 1-month lagged instruments DPR, DEF, RTB, and EWR, respectively.

	Specifications of the Discount Factor m_t
SP1:	$1 - b_0 VWR_t$
SP2:	$1 + a_1 DEF_{t-1} - b_0 VWR_t$
SP3:	$1 - (b_0 + b_1 RTB_{t-1}) VWR_t$
SP4:	$1 - (b_0 + b_1 DPR_{t-1} + b_2 DEF_{t-1}) VWR_t$
SP5:	$1 + a_1 DPR_{t-1} + a_2 DEF_{t-1} + a_3 RTB_{t-1}$ $- (b_0 + b_1 DPR_{t-1} + b_2 DEF_{t-1} + b_3 RTB_{t-1}) VWR_t$
SP6:	$1 + a_1 DPR_{t-1} + a_2 DEF_{t-1} + a_3 RTB_{t-1} + a_4 EWR_{t-1}$ $- (b_0 + b_1 DPR_{t-1} + b_2 DEF_{t-1} + b_3 RTB_{t-1} + b_4 EWR_{t-1}) VWR_t$

An iterated GMM testing procedure is implemented. The identity weighting matrix is used at the first stage. For the second stage and so on, the weighting matrix is set to be the inverse of the standard sample covariance matrix of $m_t r_t$ evaluated at the parameter estimate from the previous stage. The simulation results are based on 1,000 replications. Panel (A) records percentage rejections in the GMM tests at 10% and 5% significance levels under H_0, H_1, H_2, and H_3 (defined as in Table 2).

Panel (B) reports test results from the CRSP data. The test statistics are computed using the iterated procedure if the GMM criterion function decreases through iteration beyond the second stage. Otherwise, reported GMM test statistics are from the second stage. The p-value is the probability that a draw from the chi-squared distribution exceeds the test statistic. The expected return errors, i.e., $E(m_t r_t)/E\, m_t$, are computed using one stage GMM with the weighting matrix being the inverse of $N^{-1} \sum_{t=1}^{N} r_t r_t'$.

(A) GMM Testing Under Misspecification

Sample Size $N = 300$

	H_0		H_1		H_2		H_3	
Specification	10%	5%	10%	5%	10%	5%	10%	5%
SP1	71.7	60.3	99.3	98.5	98.4	97.0	100	100
SP2	41.2	28.4	79.8	69.0	66.0	53.2	91.3	85.5
SP3	52.8	37.4	98.0	97.1	97.6	94.6	99.6	98.9
SP4	56.9	42.0	98.6	97.3	97.4	94.4	100	99.7
SP5	6.1	2.4	21.3	11.7	19.9	11.1	40.6	29.5
SP6	3.2	1.1	14.6	7.8	14.3	7.6	27.4	17.3

Sample Size $N = 600$

	H_0		H_1		H_2		H_3	
Specification	10%	5%	10%	5%	10%	5%	10%	5%
SP1	95.0	89.8	100	100	100	100	100	100
SP2	59.6	48.2	96.5	93.5	91.0	84.3	99.7	99.3
SP3	81.4	69.7	100	100	100	100	100	100
SP4	81.6	72.2	100	100	100	100	100	100
SP5	6.1	2.3	37.2	24.5	30.9	19.5	62.7	50.4
SP6	3.4	1.3	28.1	16.7	26.2	15.1	44.4	32.6

(B) Test Results from the CRSP Data

$\hat{A}^2$-statistics, % p-values, and Expected Return Errors

	DPR	DEF	RTB	EWR	ONE	DPR	DEF	RTB	EWR	ONE
			SP1					SP2		
	$\hat{A}^2$-stat: 73 1		% p-value: 0 0			$\hat{A}^2$-stat: 37.7		% p-value: 2 .7		
SZ1	1 12	0 36	-1 48	71 7	0 19	-0 51	-0 98	-4 65	4 61	-0 09
SZ3	0 59	0 19	-2 14	5 61	0 02	-0 .74	-0 96	-4 92	2 62	-0 20
SZ5	0 44	0 14	-2 20	4 87	-0 01	-0 59	-0 85	-4 45	2 33	-0 18
SZ7	0 42	0 12	-2 37	3 97	-0 02	-0 44	-0 .78	-4 40	1 50	-0 16
SZ9	0 40	0 04	-2 34	2 .76	-0 02	0 35	-0 52	-3 26	1 04	0 01
			SP3					SP4		
	$\hat{A}^2$-stat: 51 3		% p-value: 0 1			$\hat{A}^2$-stat: 30 0		% p-value: 11 9		
SZ1	1 22	0 56	1 01	6 28	0 21	0 51	0 28	-1 61	7 46	0 17
SZ3	0 .70	0 40	0 36	4 .76	0 04	0 07	0 13	-2 23	5 .76	0 03
SZ5	0 56	0 35	0 21	4 05	0 01	-0 01	0 09	-2 24	4 95	0 01
SZ7	0 54	0 33	0 00	3 20	0 00	-0 02	0 07	-2 41	4 08	0 00
SZ9	0 49	0 24	-0 08	2 01	-0 01	-0 06	-0 01	-2 42	2 91	-0 01
			SP5					SP6		
	$\hat{A}^2$-stat: 18 5		% p-value: 42 5			$\hat{A}^2$-stat: 12 9		% p-value: 68 1		
SZ1	1 40	-0 04	-2 40	23 06	0 96	-0 .74	-0 24	-3 55	14 91	0 18
SZ3	-0 99	-0 65	-4 58	14 67	0 05	-2 20	-0 .72	-4 80	6 41	-0 36
SZ5	-0 84	-0 45	-3 14	11 89	-0 08	-1 .75	-0 50	-3 40	4 .70	-0 36
SZ7	0 37	-0 14	-1 33	10 40	0 14	-0 05	-0 14	-1 32	3 13	0 00
SZ9	0 04	0 08	0 21	7 30	0 07	0 09	0 09	0 18	1 36	0 06

Table 5: Testing the Conditional CAPM

The state variable vector is $x_t = (DPR_t\ DEF_t\ RTB_t\ EWR_t)'$ and the instrument vector is $z_t = (1\ x_t')'$ for constructing the tests. Both the normal kernel (n.k.) and the higher order kernel (h.k.) are used for panels (A) and (B). For the rest panels, the results are based on the normal kernel. Details of the kernel functions and the bandwidth are in Section 4.3.

Panel (A) presents test results using the five NYSE size portfolios. The NYSE value-weighted portfolio is the benchmark, or proxy for the market. The full sample period is from February 1947 to December 1995. The p-value is the probability that a draw from the chi-squared distribution exceeds the test statistic.

Panel (B) records test results using the 25 size-BE/ME portfolios. The benchmark is the Fama-French value-weighted portfolio. FF25 refers to the set of the 25 size-BE/ME portfolios. BM-Low (SZ-Small) stands for subset of the five stock portfolios in the BM1 (SZ1) quintile, BM-Medium (SZ-Medium) for the five in the BM3 (SZ3) quintile, and BM-High (SZ-Large) for the five in the BM5 (SZ5) quintile. The sample period is from July 1963 to December 1995.

Panel (C) presents two cases of the WLS regression results. The nonconstant regressors are in deviation form (deviation from mean). The tests of joint significance (across equations) of each regressor are constructed as follows. To test $H\delta = 0$, where H is a $n \times qn$ matrix, the test statistic is $N(H\hat{\delta}_N)'(H\hat{\Omega}_N H')^{-1}(H\hat{\delta}_N)$, which has a limiting $\chi^2(n)$ distribution.

Panel (D) presents means and standard deviations of estimated pricing errors of the conditional CAPM and the unconditional CAPM, where the benchmark is the NYSE value-weighted portfolio and the test assets are the five NYSE size portfolios. The errors of the conditional CAPM is estimated by the WLS regression method. See Appendix C for details. For the errors of the unconditional CAPM, the same method is applied (with the same regressors, but without any state variables).

Panel (E) continues the case (C3) in panel (C), where the benchmark is the Fama and French market portfolio and test assets are the 25 size-BE/ME portfolios. It presents means and standard deviations of pricing errors estimated by both the WLS regression method and the pointwise kernel method. Details of the two methods are in Appendix C.

(A) Tests Using the NYSE Size Portfolios

Kernel Function	2/47-12/95 Â^2-stat (25)	p-value (%)	2/47-12/70 Â^2-stat (25)	p-value (%)	1/71-12/95 Â^2-stat (25)	p-value (%)
n.k.	45 0	0 8	33 9	11 1	37 2	5 5
h.k.	42 4	1 6	31 4	17 5	37 1	5 6

(B) Tests Using the Size-BE/ME Portfolios

	FF25 Â^2-stat (125)	%p	BM-Low Â^2-stat (25)	%p	BM-Medium Â^2-stat (25)	%p	BM-High Â^2-stat (25)	%p
n.k.	220 2	0 0	36 0	7 1	45 5	0 .7	52 6	0 1
h.k.	213 9	0 0	34 4	9 9	41 9	1 9	49 3	0 3

	SZ-Small Â^2-stat (25)	%p	SZ-Medium Â^2-stat (25)	%p	SZ-Large Â^2-stat (25)	%p
n.k.	79 5	0 0	69 2	0 0	18 .7	81 2
h.k.	72 5	0 0	65 1	0 0	18 2	83 4

(C) T he W eighted L east S quares R egression R esults

(C1)

P roxy for the M arket: N Y S E-V W ; T est A ssets: the N Y S E Size P ortfolios
S ample P eriod: 2 /47 ¡12 /95 .

	Regression Estimates					Standard Errors				
	Int.	DPR	DEF	RTB	EWR	Int.	DPR	DEF	RTB	EWR
SZ1	0 08	-0 18	-0 05	-0 04	0 21	0 19	0 16	0 .75	0 14	0 05
SZ3	-0 02	-0 08	0 42	-0 10	0 11	0 13	0 11	0 54	0 10	0 03
SZ5	0 00	-0 03	0 60	-0 09	0 09	0 11	0 09	0 47	0 09	0 03
SZ7	0 00	-0 01	1 01	-0 16	0 04	0 08	0 08	0 42	0 07	0 03
SZ9	0 03	0 03	0 .77	-0 13	-0 01	0 06	0 07	0 39	0 06	0 02

Joint Signi¯cance T esting for Individual R egressors

	Int.	DPR	DEF	RTB	EWR
Â² (5)-stat	2 4	2 6	11 1	8 6	29 6
% p-value	79 3	75 8	4 9	12 .7	0 0

Summary S tatistics of P ricing E rrors

	Mean	Std. Dev.	1st Auto.	Cross Correlations			
SZ1	0 08	1 08	0 18				
SZ3	-0 02	0 64	0 32	0 96			
SZ5	0 00	0 55	0 38	0 91	0 98		
SZ7	0 00	0 48	0 .75	0 63	0 82	0 89	
SZ9	0 03	0 31	0 88	0 04	0 32	0 43	0 .79

(C2)

Proxy for the Market: FF-VW; Test Assets: the 25 size-BE/ME Portfolios
Sample Period: 7/63 – 12/95

Joint Significance Testing for Individual Regressors

	Int.	DPR	DEF	RTB	EWR
χ^2(25)-stat	672	170	220	275	390
% p-value	0 0	88 3	63 8	33.4	3.7

The Intercepts and the EWR Slopes

	BM 1	BM 2	BM 3	BM 4	BM 5	BM 1	BM 2	BM 3	BM 4	BM 5
	Intercepts					Standard Errors				
SZ 1	-0 29	0 09	0 16	0 38	0 33	0 27	0 23	0 22	0 21	0 23
SZ 2	-0 20	0 05	0 34	0 33	0 31	0 21	0 20	0 17	0 17	0 19
SZ 3	-0 20	0 02	0 17	0 36	0.42	0 18	0 15	0 15	0 14	0 17
SZ 4	-0 09	-0 09	0 06	0 24	0.45	0 15	0 13	0 12	0 13	0 16
SZ 5	-0 02	0 05	0 03	0 12	0 16	0 11	0 10	0 12	0 11	0 14
	Slopes on EWR					Standard Errors				
SZ 1	0 24	0 21	0 21	0 22	0 24	0 07	0 06	0 05	0 06	0 06
SZ 2	0 11	0 15	0 13	0 11	0 13	0 06	0 05	0 04	0 04	0 06
SZ 3	0 08	0 09	0 09	0 08	0 06	0 06	0 05	0 04	0 04	0 05
SZ 4	-0 01	0 05	0 02	0 00	0 03	0 05	0 04	0 04	0 04	0 05
SZ 5	-0 05	-0 04	-0 06	-0 04	-0 03	0 04	0 03	0 03	0 03	0 05

(D) Conditional Expected Return Errors
Estimated from the Size Portfolios

	Conditional CAPM $m_{t+1} = 1 ¡ b(x_t) r_{p;t+1}$					Unconditional CAPM $m_{t+1} = 1 ¡ b_0 r_{p;t+1}$				
	(Benchmark return: VWR)									
	SZ 1	SZ 3	SZ 5	SZ 7	SZ 9	SZ 1	SZ 3	SZ 5	SZ 7	SZ 9
	Sample Period: 2 /47 ¡ 12 /95									
Mean	0 08	-0 02	0 00	0 00	0 03	0 25	0 09	0 05	0 05	0 04
RMSE	1 09	0 64	0 55	0 48	0 31	1 .78	1 57	1 42	1 36	1 14
	Subperiod 1 : 2 /47 ¡ 12 /70									
Mean	0 23	0 05	-0 04	0 00	0 01	0 20	0 06	-0 05	0 02	0 06
RMSE	0 95	0 59	0 48	0 31	0 20	1 38	1 25	1 18	1 06	0 91
	Subperiod 2 : 1 /71 ¡ 12 /95									
Mean	-0 04	-0 10	0 00	-0 04	-0 01	0 31	0 11	0 15	0 08	0 03
RMSE	1 41	0 99	0 90	0 .78	0 59	2 50	2 22	1 99	1 89	1 52

(E) Summary Measures of Conditional Expected Return Errors Estimated from the Size-BE/ME Portfolios

	The WLS regression method estimates					The pointwise kernel method estimates				
	BM1	BM2	BM3	BM4	BM5	BM1	BM2	BM3	BM4	BM5
Means										
SZ1	-0.29	0.09	0.16	0.38	0.33	-0.34	0.07	0.18	0.36	0.42
SZ2	-0.20	0.05	0.34	0.33	0.31	-0.17	0.05	0.32	0.33	0.37
SZ3	-0.20	0.02	0.17	0.36	0.42	-0.18	0.05	0.18	0.32	0.45
SZ4	-0.09	-0.09	0.06	0.24	0.45	-0.08	-0.11	0.09	0.25	0.40
SZ5	-0.02	0.05	0.03	0.12	0.16	-0.06	-0.01	0.01	0.13	0.26
Standard Deviations										
SZ1	1.45	1.31	1.34	1.32	1.42	1.29	1.15	1.05	1.03	1.19
SZ2	1.07	1.15	0.97	0.93	0.92	0.85	0.88	0.69	0.66	0.77
SZ3	0.93	0.99	0.89	0.84	0.72	0.62	0.59	0.59	0.52	0.59
SZ4	0.63	0.85	0.67	0.66	0.63	0.40	0.44	0.45	0.48	0.52
SZ5	0.43	0.49	0.61	0.48	0.57	0.41	0.37	0.48	0.43	0.64

Table 6: The Labor Income Risk Factor

This table presents results on testing the conditional mean-variance efficiency of a benchmark portfolio which has return of the form

$$(1-\theta)RSM + \theta LBR;$$

where LBR is the per capita labor income growth rate. RSM is return on a stock market index. The index is the NYSE value-weighted index in tests using the five NYSE size portfolios. The sample period for this data set is from February 1947 to December 1995. In the second data set, the test assets are the 25 size-BE/ME portfolios, and RSM is excess return on the Fama-French market portfolio. The sample period is from July 1963 to December 1995.

The state variable vector is $x_t = (DPR_t\ DEF_t\ RTB_t\ EWR_t)'$ and the instrument vector is $z_t = (1\ x_t')'$ for constructing the tests. All the results are based on the normal kernel. See Section 4.3 for details of the kernel function and the bandwidth parameter.

Panel (A) reports test results using the regression-GMM approach of Section 2.4. The weighting matrix is the inverse of $\hat{\Omega}_N(\theta)$ that is being updated through iteration. The matrix is evaluated at zero ($\theta = 0$) for the first stage estimation, and evaluated at the first stage estimate of θ for the second stage estimation, and so on. The p-value is the probability that a draw from the chi-squared distribution exceeds the test statistic.

Panel (B) reports three summary measures of pricing errors. The measures are average absolute bias (AAB), average standard deviation (ASD), and average root mean squared error (ARMSE). The pricing errors are estimated by the WLS regression method. See Appendix C for details about the measures and the estimation method. It also includes the minimum of the test statistic $\hat{T}_\delta$, defined by (15) of Section 2, as a function of the free parameter θ.

(A) T he R egression-G M M Test R esults

D ata Set		Â^2 -statistic	% p-value	estimate of µ	std . err.
Set 1	2-stage	45 0	0 6	0 04	0 28
	iterated	45 0	0 6	0 04	0 27
Set 2	2-stage	219 2	0 0	0 24	0 18
	iterated	219 0	0 0	0 30	0 16

(B) S ummary M easures of P ricing E rrors

	(i) D ata Set 1				(ii) D ata Set 2			
	A A B	A SD	A R M SE	$\hat{T}_{\pm}$	A A B	A SD	A R M SE	$\hat{T}_{\pm}$
µ = 0 0	0 03	0 61	0 61	45 0	0 20	0 89	0 92	220 2
M in	0 03	0 56	0 59	44 9	0 14	0 88	0 91	219 0
µ¤	0 01	0 .77	0 52	0 08	0 93	0 45	0 52	0 27

Table 7: The Size and Book-to-Market Factors

This table presents results on testing a conditional Fama and French (1993) three factor model. The hypothesis is that a benchmark portfolio with excess return

$$FFVW + \mu_1 SMB + \mu_2 HML$$

is conditionally mean-variance efficient. FFVW is excess return on the Fama-French market portfolio. SMB and HML are returns on the Fama-French mimicking portfolios for the size factor and the book-to-market factor, respectively. The test assets are the 25 size-BE/ME portfolios. The sample period is from July 1963 to December 1995.

The state variable vector is $x_t = (DPR_t\ DEF_t\ RTB_t\ EWR_t)'$ and the instrument vector is $z_t = (1\ x_t')'$ for constructing the tests. All the results are based on the normal kernel. See Section 4.3 for details of the kernel function and the bandwidth parameter.

Panel (A) reports test results using the regression-GMM approach of Section 2.4. The weighting matrix is the inverse of $\hat{\Omega}_N(\mu)$ that is being updated through iteration. The matrix is evaluated at zero ($\mu = 0$) for the first stage estimation, and evaluated at the first stage estimate of μ for the second stage estimation, and so on. The p-value is the probability that a draw from the chi-squared distribution exceeds the test statistic.

Panel (B) reports three summary measures of pricing errors. The measures are average absolute bias (AAB), average standard deviation (ASD), and average root mean squared error (ARMSE). The pricing errors are estimated by the WLS regression method. See Appendix C for details about the measures and the estimation method. It also includes the minimum of the test statistic $\hat{T}_{\pm}$, defined by (15) of Section 2, as a function of the free parameter(s).

(A) The Regression-GMM Test Results

	Â^2-statistic	% p-value	Parameter	Estimate	Std. Err.
2-stage	181 6	0 0	μ_1	1 99	0 25
			μ_2	1 67	0 23
iterated	184 4	0 0	μ_1	4 19	0 57
			μ_2	2 28	0 36

(B) Summary Measures of Pricing Errors

	AAB	ASD	ARMSE	$\hat{T}_{\pm}$	AAB	ASD	ARMSE	$\hat{T}_{\pm}$
	(i) Three Factor Model				(ii) Conditional CAPM			
Min	0 09	0 68	0 72	178 8	0 20	0 89	0 92	220 2
$\mu_1^{¤}$	0 40	0 94	0 92	2 45	(μ_1 = 0, μ_2 = 0)			
$\mu_2^{¤}$	2 40	-0 06	0 19	1 11				
	(iii) Size E®ects				(iv) BE/ME E®ects			
Min	0 19	0 68	0 73	201 2	0 09	0 87	0 92	220 0
$\mu^{¤}$	0 32	0 94	0 86	2 30	1 94	-0 38	-0 10	0 06
	(μ_1 free; μ_2 = 0)				(μ_2 free; μ_1 = 0)			