Journal of Empirical Finance 9 (2002) 133–169
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/econbase

Nonparametric tests of conditional mean-variance efficiency of a benchmark portfolio

Kevin Q. Wang *

Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario, Canada

Accepted 17 October 2001

Abstract

In this paper, we propose three nonparametric methods for testing conditional mean-variance efficiency of a benchmark portfolio. These approaches avoid functional form misspecification and share a pleasant feature that the test statistics are based on estimators that converge at the fast parametric rate. We derive limiting distributions of the test statistics and compare the three methods in a simulation study. We also present an application of the methods for testing the conditional CAPM. © 2002 Elsevier Science B.V. All rights reserved.

JEL classification: C12; C14; G11; G12
Keywords: Mean-variance efficiency; Nonparametric tests; Conditional CAPM; Portfolio choice; Nonparametric discount factor

1. Introduction

Mean-variance efficiency is a central concept in asset allocation, asset pricing, and performance evaluation. How to test portfolio mean-variance efficiency is one of the earliest and most enduring questions in financial economics. Earlier studies have focused on testing the unconditional version of the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965), or equivalently, the unconditional mean-variance efficiency of the market portfolio. In recent years, the role of

* Tel.: +1-416-946-5059; fax: +1-416-971-3048.
E-mail address: kwang@mgmt.utoronto.ca. (K.Q. Wang).

PII: S0927-5398(01)00049-4

conditioning information has received considerable attention, motivated by mounting evidence of time-variation in expected returns, return volatilities, and betas of financial assets. Numerous researchers have incorporated conditioning information in various tests and applications.[1]

A challenging issue arises when testing conditional mean-variance efficiency of a benchmark portfolio. That is, there is no theoretical guidance on how conditional moments of the portfolio and betas vary with variables that represent conditioning information. On evaluation of conditional asset pricing models, Ghysels (1998) argues that misspecification of functional forms (of betas and so on) is of first-order importance. Using several time-varying beta models that have been proposed in the literature, he shows that these time-varying beta models are seriously misspecified such that they give rise to even more serious pricing errors than constant beta models in many cases. Harvey (1991) and He et al. (1996), among others, have also noted the problem. More recently, Brandt (1999) has stressed the importance of this issue in the estimation of portfolio and consumption choice.

In this paper, we explore a solution to this problem. We propose a new testing approach that completely avoids specification of time-varying betas. We focus on the pricing kernel (i.e., stochastic discount factor) and use a flexible nonparametric approach to incorporate conditioning information into the testing. This method utilizes the fact that the conditional mean-variance efficiency implies a pricing kernel which is determined by the first two conditional moments of the return on the benchmark portfolio. We use standard kernel regression estimates of the two moments to obtain a nonparametric estimate of the pricing kernel. Next, we use a regression approach to test if excess returns discounted by the nonparametric pricing kernel are predictable. In addition, we provide two alternative tests that are also built on pricing kernels estimated by the standard kernel regression method.

These testing approaches have an important feature. We show that although nonparametric discount factor estimation is involved, the test statistics are built on estimators that converge at the standard parametric rate ($\sqrt{N}$-convergence), regardless of the number of conditioning variables being used.[2] This fast convergence

[1] There are a growing number of tests for conditional mean-variance efficiency which are proposed in the context of testing the conditional CAPM (i.e., conditional mean-variance efficiency of some stock market) or other conditional asset pricing models. See Ferson et al. (1987), Bollerslev et al. (1988), Harvey (1989), Shanken (1990), Carhart et al. (1995), Engel et al. (1995), Cochrane (1996), He et al. (1996), Jagannathan and Wang (1996), Clare et al. (1997), Ferson and Siegel (1998), and Ferson and Harvey (1999), among others. See Ferson and Schadt (1996), and Christopherson et al. (1998) for examples on conditional performance evaluation.

[2] On implementing the regression testing idea, this paper is related to Powell et al. (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. We derive the limiting distribution of this estimator using an extension of classical U-statistic theorems to β-mixing processes. Similar extensions have been obtained by Robinson, Yoshihara, and Khashimov.

property is pleasant as it provides hope that the tests may, to some extent, escape from the "curse of dimensionality" that plagues applications of nonparametric methods.[3] We compare the three methods in a simulation study. We find that the nonparametric tests have better power than a standard discount factor GMM approach.

To illustrate the practical value of the new methodology, we test the conditional Sharpe–Lintner CAPM with the nonparametric methods. We find that the conditional CAPM performs substantially better than the static version, but still it is statistically rejected. There is a strong size pattern in the volatilities but not in the averages of the pricing errors. The pricing errors for small to medium size portfolios are positively correlated with the stock market. In other words, one can generate abnormal returns using simple strategies that increase holdings of small stocks and reduce holdings of large stocks when the stock market goes up, and vice versa.

This paper is organized as follows. Section 2 details a regression testing approach and presents analytical results for the test statistic and the underlying estimator. In Section 2, we also propose the two alternative tests. Section 3 provides a simulation study. Section 4 presents an empirical application. Section 5 concludes. The Appendix contains technical assumptions and proofs.

2. Nonparametric tests

2.1. A regression testing approach

We first present a regression testing idea, which we carry out through a weighted least squares estimator. We use a nonparametric discount factor to incorporate conditioning information into the testing of portfolio conditional mean-variance efficiency. This section describes the basic idea in details.

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 ith test asset for $i=1,\dots,n$. Let x_t be a $k\times 1$ vector of conditioning variables (or state variables) such that

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

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

[3] It is well known that nonparametric estimates typically converge at slower rates than parametric estimates, and the problem is more serious for higher dimensional applications. This is known as the curse of dimensionality (e.g., Silverman, 1986).

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

If the benchmark portfolio p is conditionally mean-variance efficient, then

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

which implies

$$E(r_{i,t+1} \mid I_t) = E(r_{p,t+1} \mid I_t) \frac{E(r_{i,t+1} r_{p,t+1} \mid I_t)}{E(r_{p,t+1}^2 \mid I_t)}, \tag{4}$$

for $i = 1, \ldots, n$. The covariance representation (3) is the familiar beta pricing equation. Our focus is on Eq. (4), the cross moment representation.[5]

Let $g_p(x_t) = E(r_{p,t+1} \mid x_t)$, $g_{pp}(x_t) = E(r_{p,t+1}^2 \mid x_t)$, and $b(x_t) = g_p(x_t)/g_{pp}(x_t)$. Given that Eqs. (1) and (2) hold, conditional expected return errors (or Jensen's alphas) from Eq. (4) can be expressed as

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

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

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

Let $e_{i,t+1} = m_{t+1} r_{i,t+1}$ (i.e., discounted excess returns) and z_t be a $q \times 1$ vector of observed variables in I_t. If we could observe the discount factor m_{t+1}, a natural way to test the condition $E(e_{i,t+1} \mid I_t) = 0$ 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

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

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

[4] Note that Eqs. (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 tests.

[5] Eq. (3) implies that $E(r_{i,t+1} \mid I_t)\mathrm{var}(r_{p,t+1} \mid I_t) = E(r_{p,t+1} \mid I_t)\mathrm{cov}(r_{i,t+1}, r_{p,t+1} \mid I_t)$, which is equivalent to Eq. (4). Conversely, Eq. (4) also implies Eq. (3).

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

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

for $i = 1, \dots, n$, 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

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

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

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

The nonparametric estimators in Eq. (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 $\hat{\delta}_i$. 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.[6] 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.

[6] In particular, the choice of setting $\hat{w}_t = 1$ does not yield U-statistic structures. This is why it is not used here. Note that it is possible to avoid the limiting distribution theory by using a time series bootstrap method. A rigorous analysis of potential gains and problems of such an alternative approach is beyond the scope of this paper.

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

$$\hat{\hat{\delta}}_N = \left(\hat{\hat{\delta}}_1' \hat{\hat{\delta}}_2' \dots \hat{\hat{\delta}}_n'\right)'.$$

Intuitively, $\hat{\hat{\delta}}_N$ 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 $i = 1, \dots, n$). Thus, we can conduct a test by checking how far $\hat{\hat{\delta}}_N$ is away from zero, using asymptotic distribution theory to account for sampling errors. Note that this regression testing approach provides a simple way to look into pricing errors. By design, our test aims at detecting time-variation in Jensen's alphas. A significant component of $\hat{\hat{\delta}}_N$ indicates that the expected return errors are correlated with the corresponding state variable. Thus, potentially the test may lead to a simple characterization of time-variation in the pricing errors. Moreover, in applications we may view Eq. (6) as a model for pricing errors. That is, $z_t' \delta_i$ may serve as an approximation for $E(e_{i,t+1} \mid I_t)$, the time-varying Jensen's alphas from Eq. (4).

2.2. Asymptotics and the test statistic

This section presents an important property of the test. Although we use nonparametric kernel density and regression estimates in construction, the estimator $\hat{\hat{\delta}}_N$ behaves like a parametric estimator. It has a standard limiting multivariate normal distribution and in particular it converges at the fast parametric convergence rate.[7] This property makes it straightforward to construct and implement the testing approach. The fast convergence rate is especially pleasant, as it provides a theoretical reason to expect that the test may have good power in applications.

We use the following notations to present the results. Let r_{t+1} be a vector of scaled excess returns $(r_{1,t+1} \dots r_{n,t+1})' \otimes z_t$ and $y_{t+1} = (x_t' z_t' r_{p,t+1} r_{t+1}')'$, where '$\otimes$' is the Kronecker operator. Denote $w_t = f(x_t) g_{pp}(x_t)$, $A = \iota_n \otimes E[w_t z_t z_t']$, and $\hat{A}_N = \iota_n \otimes N^{-1} \sum_{t=1}^N \hat{w}_t z_t z_t'$, where ι_n is the $n \times n$ identity matrix. Let $\delta = (\delta_1' \dots \delta_n')'$ with $\delta_i = [E(w_t z_t z_t')]^{-1} E[w_t z_t e_{i,t+1}]$. Define

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

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

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

[7] This result arises from the fact that the average of some functions of slow-converging nonparametric kernel estimates can converge at the parametric rate ($\sqrt{N}$). See Hardle and Stoker (1989), Powell et al. (1989), Robinson (1989), and Lee (1992) for examples about this fact.

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

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.

Proposition 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 $\Omega=A^{-1}\Gamma A^{-1}$, $\Gamma=\sum_{-\infty}^{\infty}\Gamma_j$, and $\Gamma_j=E[\gamma(y_{t+1})\gamma(y_{t+j+1})']$.*

Proposition 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 Eqs. (1) and (2), nor does it require that the regression equations in Eq. (6) are correctly specified. Note that the bandwidth conditions are different from those for the kernel density and regression function 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 estimate the covariance matrix Ω, we first consider the 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 Eqs. (10) and (11) by standard kernel estimators,[8] and replacing δ in Eq. (9) by $\hat{\delta}_N$, we obtain a natural approximation for $\gamma(y_{t+1})$:

$$\hat{\gamma}_N(y_{t+1})=\hat{\eta}_N(y_{t+1})-\left[\iota_n\otimes\hat{a}_N(y_{t+1})\right]\hat{\delta}_N, \tag{12}$$

$$\hat{\eta}_N(y_{t+1})=\hat{f}(x_t)\Big[\hat{g}_{pp}(x_t)r_{t+1}-\hat{g}_p(x_t)r_{p,t+1}r_{t+1}+\hat{g}_r(x_t)r_{p,t+1}^2 \\ -\hat{g}_{pr}(x_t)r_{p,t+1}\Big], \tag{13}$$

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

[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 Eq. (14). Instead, one can simply replace $g_{zz}(x_t)$ by $z_t z_t'$, which gives $\hat{a}_N(y_{t+1})=\hat{f}(x_t)[\hat{g}_{pp}(x_t)+r_{p,t+1}^2]z_t z_t'$.

where $\hat{f}$, $\hat{g}_p$, and $\hat{g}_{pp}$ are defined as in Eq. (8), 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 Eqs. (1) and (2), $\Gamma_j = 0$ for any $j \neq 0$ when the regression equations in Eq. (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 proposition gives the limiting distribution of the test statistic.

Proposition 2 Let the conditions of Proposition 1 hold. *(i) Given Eqs. (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. (ii) $\hat{\Gamma}_j$ is a consistent estimator of Γ_j for any fixed j.*

The regression test is constructed using a simple covariance matrix estimator. Note that Proposition 2 has also provided the necessary inputs $(\hat{\Gamma}_j)$ 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 remarks are in order. First, the limiting distribution of $\hat{\delta}_N$ does not depend on the scaling constant c when we set the bandwidth as $h = cN^{-1/(2k+1)}$. This property suggests that unlike kernel density and kernel regression estimates, the estimate $\hat{\delta}_N$ will become less sensitive to choice of c as the sample size

[9] In application, we have attempted to use a Newey–West covariance matrix estimator with a lag length of 6, 12, and 18. 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.

increases. In other words, the bandwidth choice in our approach is not as important as in kernel density and kernel regression estimation.[10] The trade-off is that we cannot justify cross-validation procedures. In simulations and applications, we adopt a simple rule, and we find that results are not very sensitive to small changes in the bandwidth. Secondly, it looks like a "free lunch" that $\hat{\hat{\delta}}_N$ has standard limiting distribution and the parametric convergence rate, no matter how many conditioning variables we use. However, it should be noted that there is certain cost due to nonparametric estimation, which is embedded in the covariance matrix of $\hat{\hat{\delta}}_N$. In finite samples, it still pays to eliminate redundant conditioning variables. 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., $E(z_t e_{i,t+1} \mid x_t) = 0$, $i = 1, \dots, n$).[11] 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. Two alternative testing approaches

Like most nonparametric methods, the approach that we have proposed is not associated with any optimality property. In this section, we consider two alternative ways to test the conditional mean-variance efficiency hypothesis. They are also constructed with a nonparametric pricing kernel and they share the property that the underlying estimators converge at the parametric rate. This section briefly outlines the two alternative tests.

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

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

where

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

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

[10] This is nice as it suggests that bandwidth problems due to persistency of state variables (see, Pritsker, 1998; Chapman and Pearson, 2000) may not be a serious issue to our approach.

[11] In statistical term, the test is not consistent, for a given vector z_t, against certain processes that generate $e_{i,t+1}$. Chen and Fan (1999) recently propose a consistent test. The trade-off is that the limiting distribution of their test is nonstandard and, thus, it is complicated to implement in simulations and applications.

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

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

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

We estimate the parameter vector α_i by

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

where $\hat{w}_t=\hat{f}^2(x_t)\hat{\sigma}_p^2(x_t)$, $\hat{\sigma}_p^2(x_t)=\hat{g}_{pp}(x_t)-\hat{g}_p^2(x_t)$, 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 weighting function $\hat{w}_t$, 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 third-order generalized *U*-statistics. Under certain regularity conditions, if $Nh^{4k}\to\infty$ and $Nh^{4k+2}\to 0$, then the weighted least squares estimator $\hat{\alpha}_N=(\hat{\alpha}_1'\ldots\hat{\alpha}_n')'$ has a limiting normal distribution.

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

$$\hat{T}_\alpha=N\hat{\alpha}_N'\hat{\Sigma}_N^{-1}\hat{\alpha}_N \tag{16}$$

has a limiting $\chi^2(qn)$ distribution if the benchmark portfolio p is conditionally mean-variance efficient. The covariance matrix estimator $\hat{\Sigma}_N$ is given in Appendix C.

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

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

We estimate the vector λ of unconditional moments by

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

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

$$\hat{T}_\lambda=N\hat{\lambda}_N'\hat{\Lambda}_N^{-1}\hat{\lambda}_N, \tag{17}$$

where $\hat{\Lambda}_N$ is a covariance matrix estimator defined in Appendix C. Given Eqs. (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 mean-variance efficient. In the econometrics literature, Robinson (1989) provides a class of unconditional moments tests. The test given in Eq. (17) fits as a special case of his general methodology.

The regression test in Eq. (15) and the two alternative tests in Eqs. (16) and (17) are not equivalent. Analytical comparison of their performances is rather difficult. In the next section, we provide a simulation experiment to study whether these asymptotically justified methods have practical value for finite sample applications.

3. A Monte Carlo experiment

3.1. Returns and conditioning variables

Postwar monthly data on six excess returns and five forecasting variables are obtained from CRSP to design a Monte Carlo experiment. We use the value-weighted portfolio of New York Stock Exchange (NYSE) stocks as the benchmark portfolio (with the excess return denoted by VWR), and the value-weighted NYSE size decile 1, 3, 5, 7, and 9 portfolios as test assets. The conditioning variables that we use are the dividend-price ratio (DPR), the default premium (DEF), the 1-month Treasury bill rate (RTB), excess return on the NYSE equally weighted portfolio (EWR), and the term premium (TERM). All these are popular forecasting variables for stock returns. In terms of predicting the excess return VWR and VWR squared, we find by ordinary least squares regressions that the variable TERM appears to be redundant when the other four are all present. We keep TERM as a redundant state variable in simulations, just to check robustness of the results. Summary statistics of the data are provided in Table 1.

3.2. Data generation

In the simulations, the forecasting variables DPR, DEF, RTB, EWR, and TERM are generated through a VAR(1) model. Let $y_1 = \ln(\mathrm{DPR})$, $y_2 = \ln(\mathrm{DEF})$, $y_3 = \ln(\mathrm{RTB})$, $y_4 = \mathrm{EWR}$, and $y_5 = \mathrm{TERM}$. 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.[12]

The conditional distribution of the six excess returns is multivariate 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 deviation of the time-$(t+1)$ return is of the form

$$\sigma_{i,t} = |a_{i,0} + a_{i,1}\mathrm{DPR}_t + a_{i,2}\mathrm{DEF}_t + a_{i,3}\mathrm{RTB}_t + a_{i,4}\mathrm{EWR}_t|.$$

[12] Since DPR, DEF, and RTB are always positive throughout the sample, it seems more appropriate to let them be lognormal (instead of normal).

Table 1
Summary statistics of 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 1-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 Nth size decile NYSE stocks. 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.

Variable	Mean	Standard deviation	First autocorrelations	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
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	

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

$$E_t(\mathrm{VWR}_{t+1}) = b_0 + b_1 \mathrm{DPR}_t + b_2 \mathrm{DEF}_t + b_3 \mathrm{RTB}_t + b_4 \mathrm{EWR}_t .$$

We use four models, denoted H_0, H_1, H_2, and H_3, that differ in the generation of conditional expected returns (but they are identical otherwise). In the first model H_0, the NYSE value-weighted portfolio is conditionally mean-variance efficient. The conditional means of excess returns on the five size portfolios are determined by the beta pricing Eq. (3). In the other three models H_1, H_2, and H_3, the benchmark is not conditionally mean-variance efficient. We generate the conditional expected excess returns 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. In other words, conditional expected excess returns on the five size portfolios take the following forms under H_1, H_2, and H_3, respectively:

$\mathbf{H}_1$:

$$E_t r_{i,t+1} = b_{i,0} + b_{i,1} \mathrm{DPR}_t + b_{i,2} \mathrm{DEF}_t + b_{i,3} \mathrm{RTB}_t + b_{i,4} \mathrm{EWR}_t$$

H_2:

$$E_t r_{i,t+1} = b_{i,0} + b_{i,1}\ln(\mathrm{DPR}_t) + b_{i,2}\ln(\mathrm{DEF}_t) + b_{i,3}\ln(\mathrm{RTB}_t) + b_{i,4}\mathrm{EWR}_t$$

H_3:

$$E_t r_{i,t+1} = b_{i,0} + b_{i,1}\mathrm{DPR}_t + b_{i,2}\mathrm{DEF}_t + b_{i,3}\mathrm{RTB}_t + b_{i,4}\mathrm{EWR}_t + b_{i,5}\sigma^2_{i,t}$$

where $\sigma^2_{i,t}$ is the conditional variance of $r_{i,t+1}$.

The VAR process of conditioning variables and the conditional moments of asset returns are estimated with the historical data. The parameter estimates are taken to be the 'true' parameter values in the simulations.[13]

3.3. Simulation results

Table 2 presents the simulation results on finite sample performances of the nonparametric tests. In constructing the tests, the state variable vector is $x_t = (\mathrm{DPR}_t\ \mathrm{DEF}_t\ \mathrm{RTB}_t\ \mathrm{EWR}_t)'$ except for panel (C) where we add the redundant variable TERM. The instrument vector is $z_t = (1\ \ x'_t)'$ throughout the simulations. For notational convenience, we refer to the regression tests given in Eqs. (15) and (16) of Section 2 as R1-test and R2-test, respectively. We refer to the unconditional moments test given in Eq. (17) of Section 2 as UM-test. We also include a standard discount factor GMM approach which serves as a benchmark in the evaluation of the nonparametric tests. The discount factor is of the linear form as in Carhart et al. (1995) and Cochrane (1996), among others.[14]

Panel (A) records the rejection rates of the three nonparametric tests against conditional mean-variance efficiency of the NYSE benchmark portfolio at 5% and 10% levels. First, we look at the results based on the normal kernel. The R1-test is clearly the winner. Most impressive is its performance at sample size $N = 600$. Under H_0, it produces rejection rates very close to the corresponding significance levels. Meanwhile, the power of the test is nearly perfect under H_1, H_2, and H_3. At $N = 300$, the R1-test has significantly better power than the R2-test and the UM-test.

[13] Wang (1998) reported more details of the simulations and more results. Details such as parameter estimates of the four models and the VAR(1) are omitted for brevity, but they are available upon request.
Note that to use linear functional forms is primarily for convenience. There is no reason that such linear specifications would give any advantage to the nonparametric tests. The classic kernel regression estimates are designed to handle all functional forms. There is no reason that they perform better for linear functions. In fact, the tests tend to have better power in the log-linear model H_2 and quadratic model H_3.

[14] See Cochrane (1996) for a detailed discussion on the discount factor GMM approach, and Hansen (1982) for a rigorous presentation of the GMM.

Table 2
Simulation results

This table presents simulation results on performances of the nonparametric tests. In constructing the tests, the state variable vector is $x_t = (\mathrm{DPR}_t\ \mathrm{DEF}_t\ \mathrm{RTB}_t\ \mathrm{EWR}_t)'$ 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 panel (A). For the rest panels, the results are based on the normal kernel. Details of the kernel functions and the bandwidth are in Appendix D. All the simulation results are based on 1000 replications.

Panel (A) records the rejection rates in testing conditional mean-variance efficiency of the value-weighted NYSE portfolio under the null (H_0) and the three alternative hypotheses (H_1, H_2, and H_3) at the 5% and 10% levels. See Section 3.2 for details about the models H_0, H_1, H_2, and H_3. The R1-test, R2-test, and UM-test refer to the tests given in Eqs. (15), (16) and (17) of Section 2, respectively. Also included are results for a discount factor GMM approach. In the GMM case, the moment condition $Em_t r_t = 0$ is tested, where r_t is a vector of 25 scaled excess returns: the five excess asset returns SZ1, SZ3, SZ5, SZ7, SZ9, multiplied by the constant 1 and the four 1-month lagged instruments DPR, DEF, RTB, and EWR, respectively. The discount factor m_t is assumed to be as follows.

$$m_t = 1 + a_1\mathrm{DPR}_{t-1} + a_2\mathrm{DEF}_{t-1} + a_3\mathrm{RTB}_{t-1} + a_4\mathrm{EWR}_{t-1} \\ -(b_0 + b_1\mathrm{DPR}_{t-1} + b_2\mathrm{DEF}_{t-1} + b_3\mathrm{RTB}_{t-1} + b_4\mathrm{EWR}_{t-1})\mathrm{VWR}_t$$

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.

Panel (B) records the rejection rates 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 conditioning variables. The redundant variable TERM is added to the set, i.e., $x_t = (\mathrm{DPR}_t\ \mathrm{DEF}_t\ \mathrm{RTB}_t\ \mathrm{EWR}_t\ \mathrm{TERM}_t)'$. The test procedures are otherwise identical to those in (A).

(A) Rejection rates in the nonparametric testing

Sample	Test method	Kernel function	H_0		H_1		H_2		H_3	
			10%	5%	10%	5%	10%	5%	10%	5%
$N = 300$:	R1-test	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
	R2-test	n.k.	4.7	2.2	46.6	36.1	49.6	38.0	62.7	52.1
		h.k.	4.6	2.2	45.7	34.7	49.4	37.2	61.4	51.3
	UM-test	n.k.	0.4	0.1	31.7	16.4	25.9	13.2	49.8	32.4
		h.k.	0.3	0.0	16.2	7.6	19.1	9.3	32.3	18.9
	GMM		3.2	1.1	14.6	7.8	14.3	7.6	27.4	17.3
$N = 600$:	R1-test	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
	R2-test	n.k.	6.5	2.9	83.8	74.5	81.9	71.0	93.9	89.0
		h.k.	5.6	2.5	81.7	71.1	80.6	70.1	92.9	88.1
	UM-test	n.k.	2.9	0.6	92.2	84.4	88.2	77.2	98.0	95.1
		h.k.	0.8	0.1	73.1	58.8	79.8	65.9	89.9	81.1
	GMM		3.4	1.3	28.1	16.7	26.2	15.1	44.4	32.6

(B) Rejection rates under H_0

Test method	5%	10%	15%	20%	25%	50%	75%
R1-test	4.7	10.5	16.2	21.9	27.1	54.6	80.2
R2-test	2.9	6.5	10.7	15.5	20.7	48.5	77.7
UM-test	0.6	2.9	7.0	10.8	15.8	39.6	71.5

Table 2 (*continued*)

(C) Tests using a larger set of conditioning variables

$x_t = (\mathrm{DPR}_t\ \mathrm{DEF}_t\ \mathrm{RTB}_t\ \mathrm{EWR}_t\ \mathrm{TERM}_t)'$

Test method	Sample size	H_0		H_1		H_2		H_3	
		10%	5%	10%	5%	10%	5%	10%	5%
R1-test	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
R2-test	300	4.1	1.9	39.0	26.9	39.6	29.5	54.5	43.8
	600	3.3	1.2	70.5	58.5	69.4	56.8	85.7	78.0
UM-test	300	0.1	0.0	17.1	8.6	14.0	6.8	30.6	17.4
	600	0.7	0.0	80.1	67.8	75.4	60.8	93.7	86.3

Both the R2-test and the UM-test have reasonably good power at $N = 600$, with rejection rates all above 70%. The UM-test has slightly better power. At $N = 300$, however, the R2-test significantly outperforms the UM-test. The power of the UM-test is rather low for this sample size (with rejection probabilities below 35% in five out of the six cases), but even so it is no worse than the discount factor GMM approach. Overall, all the three nonparametric tests have better power than the GMM test.

The results based on the higher order kernel give the same conclusions on relative performance of the tests. In particular, the R1-test is still the best among the three. As one would expect, the rejection rates 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.[15]

Panel (B) presents a snapshot of distributions of the three test statistics under H_0, with sample size $N = 600$. It shows that the distribution of the R1-test statistic is surprisingly close to the predicted limiting distribution. The R2-test takes the second place. It is significantly better than the UM-test in this criterion.

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 the powers of the tests. Overall, the rejection rates drop, with an amount between 5% and 20% in most cases. The relative performances of the three tests remain in the same order as in panel (A). The R1-test remains the winner. Against the three alternative models, the R1-test still has impressive power at $N = 600$, with rejection rates all above 85%.

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

3.4. Extensions and variations

The Monte Carlo experiment is plain and simple. We now discuss some of the extensions and variations that we have investigated.[16]

First, we have looked at how sensitive the tests may be to the selection of conditioning variables. We implement the R1 tests with an *incorrect* set of conditioning variables. We study cases of missing variable: in each case we drop one of the four state variables DPR, DEF, RTB, and EWR. We also consider cases of mis-selecting variable: in each case we have one of the four state variables replaced by TERM. We find the nonparametric tests still have solid performance in all the cases.

We have computed the GMM tests using variations of the discount factor specification. We find that the tests tend to produce erroneous rejections when we use a parsimonious specification (such as the one of Cochrane, 1996), but they tend to have low power when we use a general form (such as the one given in Table 2). These results are consistent with the finding of Ferson and Foerster (1994) that GMM tests tend to have more serious power problems for models with a large number of parameters.

We have considered if the persistency in the variables DPR, DEF, and RTB affects the simulation results. These variables are so highly correlated over time that their first autocorrelations are very close to one (see Table 1). This is a question that may concern many, since recently Pritsker (1998) and Chapman and Pearson (2000) have shown, in the context of interest rate models, that extreme persistence in a time series can have important impacts on bandwidth choice in both pointwise kernel density and kernel regression estimation. To see if the persistency matters in our tests, we simply alter the parameter values in the VAR(1) such that the first autocorrelations of the variables DPR, DEF, and RTB drop to around 0.90 (so that they are no longer close to being nonstationary). Then, we regenerate data and reevaluate the R1 tests. We get quite similar test results, which shows that the simulation results are not driven by the fact that DPR, DEF, and RTB are extremely persistent over time.[17]

We have also checked if it makes a difference to allow for time-varying volatilities of the state variables. We use a simple ARCH VAR(1) model to incorporate time-varying volatilities. We estimate the model with the historical data to get parameter values and then use them to regenerate data. Again, the test results from the new simulated data are very similar to those being reported.

[16] Tables about details of the extensions and variations are omitted to keep things concise. They are available upon request.

[17] In contrast to pointwise density and regression function estimation, as noted in Section 2.2, both the kernel function and the bandwidth do not affect the limiting distribution of our test statistics. This implies that the problems for the pointwise nonparametric estimates may be largely irrelevant to our tests.

4. An illustrative application

We illustrate the nonparametric methods by applying them to testing the conditional CAPM. We first implement and compare the three testing methods. We then take a look at the pricing errors using the regression testing approach of Section 2.1.

4.1. Set-up

We use two sets of stock portfolios as test assets. We first focus on the set of the five NYSE size portfolios, the same as that for the simulations. The data are monthly excess returns from February 1947 to December 1995. To check the robustness of the results, we also use a second data set which contains the Fama and French (1993) 25 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, we use the NYSE value-weighted and equally weighted portfolios and the market portfolio of Fama and French. The conditioning variables used in the application are the dividend-price ratio (DPR), the default premium (DEF), the 1-month Treasury bill rate (RTB), and excess return on the NYSE equally weighted index (EWR). The vector of instruments in the tests is $z_t = (1\, x_t')'$, where x_t is the vector $(\mathrm{DPR}_t\ \mathrm{DEF}_t\ \mathrm{RTB}_t\ \mathrm{EWR}_t)'$. 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 Appendix D.

4.2. Plots of conditional betas

The testing methods that we propose in this paper completely avoid restrictions on functional forms of conditional betas. This can be very advantageous if the betas are not well approximated by linear models and if it is difficult to estimate the nonlinear beta functions. Note that on one hand there is no theoretical reason why the betas should be linear functions of conditioning variables. On the other hand, Ghysels (1998) finds strong empirical evidence against models of linear conditional betas. In this section, we present plots of conditional betas, which serve as an interesting way to illustrate nonlinearity in the time-varying betas.

We estimate the betas nonparametrically. The standard kernel regression method is applied to estimate the first two moments of returns and then obtain estimates of the betas. Confidence intervals are also obtained for the beta function estimates, using the stationary bootstrap method of Politis and Romano (1994). To cope with the curse of dimensionality, we focus on "one-dimensional snap shots". While betas are estimated as multivariate functions, we look at univariate functions, i.e., relations between one conditioning variable and betas, keeping all the other conditioning variables at their means. By doing so, we focus on areas where we have more data (as opposed to areas where there are few data points) and also avoid the problem of how to plot four-dimensional functions.

We obtain the conditional betas of the NYSE size portfolios. For brevity, we only report the beta of SZ5 (i.e., the size decile 5 portfolio, or a portfolio of medium size). The four panels in the figure correspond to four univariate functions. The solid curves are the estimated beta functions. The dashed ones represent the 95% confidence intervals. In each panel, we consider only the interval of the conditioning variable that ranges from two standard deviations below its mean to two standard deviations above the mean. To present more consistently, we rescale the interval to range from 0 to 200 (so that 100 represents the mean of the variable in each case).

Fig. 1 supports the rejections that Ghysels finds against linear conditional beta models.[18] Visually, the beta is highly nonlinear in the dividend yield (DPR): the slope changes drastically somewhere between 100 and 150. Also, the slope changes significantly between 50 and 100 in the panel associated with the stock index (EWR). In terms of the confidence interval bands, the nonlinearity between the beta and EWR seems highly significant. In the panels associated with the default premium (DEF) and the Treasury bill rate (RTB), the estimates are too noisy to give any clear conclusion, though they also suggest some nonlinearity. We have also inspected plots of betas for the other NYSE size portfolios and the Fama and French size and book-to-market portfolios. In sum, these plots suggest that nonlinearity in conditional betas is present but yet the beta function differs significantly across test assets. In addition, the beta estimates are noisy in most cases, showing that it is difficult to get precise nonparametric estimates of the nonlinear beta functions.

4.3. Tests of the conditional CAPM

First, we test the conditional Sharpe-Lintner CAPM using either the NYSE value-weighted portfolio or the NYSE equally weighted portfolio as proxy for the market. Test assets are the five NYSE size portfolios. The results are reported in panel (A) of Table 3. Next, we use the Fama–French market portfolio as the benchmark and the 25 size-BE/ME portfolios as test assets. Panel (B) of Table 3 present the test results, including six subset cases. In each of the cases, five out of the 25 size-BE/ME portfolios in a single size or BE/ME quantile are taken as test assets. The six quintiles that we choose are three size quintiles (small, medium, large) and three BE/ME quintiles (low, medium, high). In these two

[18] We perform the tests of Ghysels using our data and also get strong rejections against linear conditional beta specifications. Further details of the procedures to create the beta plots and to replicate the Ghysels' tests are omitted, but available upon request.

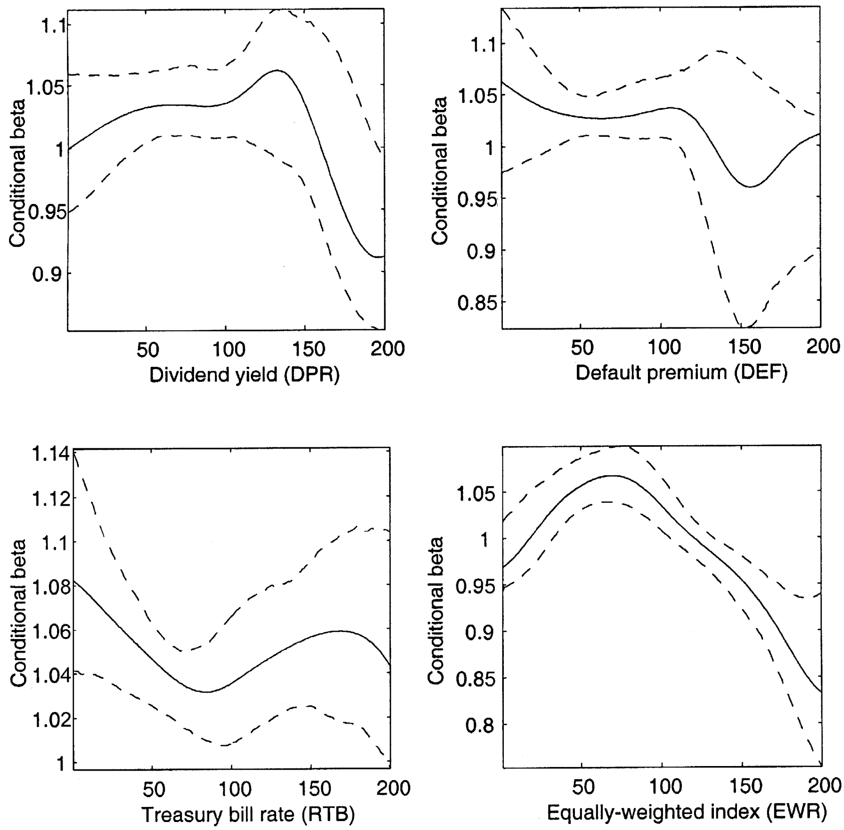

Fig. 1. Plots of the conditional beta of the NYSE size portfolio SZ5.

panels, results from all the three nonparametric tests are included for comparison. Like in the simulations, we refer to the three tests as the R1-test, the R2-test, and the UM-test, when comparing test results.[19]

The R1-tests strongly reject that either of the two NYSE market proxies is conditionally mean-variance efficient. The *p*-values are all below 3% for the full sample period. For the subsample period from February 1947 to December 1970, the tests for the value-weighted benchmark fail to reject at 10% level, but the *p*-values are fairly low (11.1% using the normal kernel). For every other subsample case, the tests all reject at 10% level. In panel (B), the R1-tests strongly reject that the Fama–French market proxy is conditionally mean-variance efficient. The

[19] Recall that the R1-test, R2-test, and UM-test stand for the tests defined in Eqs. (15), (16) and (17) of Section 2, respectively.

Table 3
Tests of the conditional CAPM

The vector of conditioning variables is $x_t = (\mathrm{DPR}_t\ \mathrm{DEF}_t\ \mathrm{RTB}_t\ \mathrm{EWR}_t)'$ 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 Appendix D. Results from all the three nonparametric tests are included in panels (A) and (B).

Panel (A) presents test results using the five NYSE size portfolios. Either the NYSE value-weighted portfolio (NYSE-VW) or equally weighted portfolio (NYSE-EW) is the benchmark, or proxy for the market. The R1-test, R2-test, and UM-test refer to the tests given in Eqs. (15), (16) and (17) of Section 2, respectively. 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 and book-to-market (BE/ME) portfolios. The benchmark portfolio is the Fama–French value-weighted market 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 a case of the WLS regression results for the sample period from February 1947 to December 1995. The non-constant 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. The pricing error at time t for asset i is estimated by $z_t'\hat{\delta}_i$ where $\hat{\delta}_i$ is defined by Eq. (7) of Section 2.

Panel (D) presents the means and standard deviations of estimated pricing errors of the conditional CAPM and the unconditional CAPM, where the benchmark portfolio is the NYSE value-weighted portfolio and the test assets are the five NYSE size portfolios. The method for estimating the errors is the same as in (C). For the pricing errors of the unconditional CAPM, the same regression method is applied (with the same regressors but b is constant).

(A) Tests using the NYSE size portfolios

Proxy for the market	Test method	Kernel function	2/47–12/95		2/47–12/70		1/71–12/95	
			χ^2-stat (25)	p-value (%)	χ^2-stat (25)	p-value (%)	χ^2-stat (25)	p-value (%)
NYSE-VW	R1-test	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
	R2-test	n.k.	34.0	10.8	32.0	15.7	41.0	2.3
		h.k.	33.3	12.3	31.7	16.6	40.9	2.4
	UM-test	n.k.	40.0	3.0	29.2	25.4	25.7	42.3
		h.k.	37.4	5.3	27.4	33.7	26.4	38.9
NYSE-EW	R1-test	n.k.	43.6	1.2	39.1	3.6	36.9	5.9
		h.k.	40.2	2.7	35.7	7.6	36.8	6.1
	R2-test	n.k.	32.9	13.4	37.1	5.6	40.7	2.5
		h.k.	32.1	15.5	36.6	6.4	41.0	2.3
	UM-test	n.k.	40.7	2.5	31.8	16.4	23.6	54.3
		h.k.	37.4	5.2	30.0	22.6	23.6	54.3

(B) Tests using the Size-BE/ME portfolios

	FF25		BM-Low		BM-Medium		BM-High	
	χ^2-stat (125)	*p*-value (%)	χ^2-stat (25)	*p*-value (%)	χ^2-stat (25)	*p*-value (%)	χ^2-stat (25)	*p*-value (%)
R1-test								
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
R2-test								
n.k.	244.1	0.0	33.4	12.2	34.4	9.9	39.3	3.4
h.k.	247.3	0.0	32.8	13.5	33.9	11.1	38.8	3.8
UM-test								
n.k.	125.8	46.2	19.6	76.5	32.4	14.6	36.6	6.3
h.k.	122.2	55.4	19.2	78.7	30.7	19.9	34.9	9.0

	SZ-Small		SZ-Medium		SZ-Large	
	χ^2-stat (25)	*p*-value (%)	χ^2-stat (25)	*p*-value (%)	χ^2-stat (25)	*p*-value (%)
R1-test						
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
R2-test						
n.k.	57.6	0.0	58.8	0.0	18.1	83.9
h.k.	56.5	0.0	58.3	0.0	17.9	84.7
UM-test						
n.k.	46.2	0.6	38.2	4.4	16.6	89.6
h.k.	44.6	0.9	36.6	6.3	16.7	89.3

(continued on next page)

Table 3 (*continued*)

(C) The WLS regression results using the size portfolios

	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 significance testing for individual regressors

	Int.	DPR	DEF	RTB	EWR
$\chi^2(5)$-stat	2.4	2.6	11.1	8.6	29.6
% p-value	79.3	75.8	4.9	12.7	0.0

Summary statistics of pricing errors

	Mean	Standard deviation	First autocorrelations	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

(D) Pricing errors estimated from the size portfolios

	Conditional CAPM $m_{t+1} = 1 - b(x_t)r_{p,t+1}$					Unconditional CAPM $m_{t+1} = 1 - b_0 r_{p,t+1}$				
	SZ1	SZ3	SZ5	SZ7	SZ9	SZ1	SZ3	SZ5	SZ7	SZ9
Benchmark return: VWR										
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

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.[20]

The pattern of p-values from the three test methods is consistent with the simulation results. In panel (A), the p-values from the UM-tests are smaller than those from the R2-tests for the full sample period. In every subsample case, however, the UM-tests produce much higher p-values than the R2-tests; the differences of the p-values range from 10.8% to 52.0%. Overall, the R1-tests have the most stable performance, producing the strongest rejections. In panel (B), the UM-tests produce large p-values (above 45%!) when all the 25 portfolios are used as test assets. On the contrary, the two regression testing approaches (R1 and R2) give strong rejections with 0.0% p-values. In every subset case, the UM-tests come up with the largest p-values while the R1-tests give the smallest.

4.4. Pricing errors

The regression testing approach of Section 2.1 provides a simple way to look into the conditional expected return errors or pricing errors. Panel (C) of Table 3 presents a case of the regression results. In this case, the benchmark is the NYSE value-weighted portfolio, and test assets are the five NYSE size portfolios. In these regressions, the four regressors DPR, DEF, RTB, and EWR are in deviation form (deviation from mean), so that the intercepts are just average pricing errors or biases of the conditional CAPM.[21]

The results of panel (C) 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 WLS regressions for small to medium size portfolios (SZ1, SZ3, and SZ5), 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. In the SZ7

[20] 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 conditioning variables: $x_t = (\mathrm{DPR}_t\ \mathrm{DEF}_t\ \mathrm{RTB}_t\ \mathrm{EWR}_t\ \mathrm{TERM}_t)'$. The test results are qualitatively robust. For brevity, these robustness checking results are omitted.

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

and SZ9 regressions, the two highly correlated regressors DEF and RTB have marginally significant slopes, while other coefficients are all insignificant.

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 with size from 1.08% to 0.31%. Now it is clear where the rejections come from. It is the volatilities of the pricing errors that have been detected by the regression tests. More specifically, the conditioning variable EWR has played a key role in the regression tests, by capturing a volatile component in pricing errors associated with the small to medium size portfolios. This component is not only statistically significant but also large in economic terms.

Finally, we compare the pricing errors of the conditional CAPM with those of the unconditional CAPM. Ghysels (1998) points out that if we correctly specify the dynamics of the beta risk, time-varying beta models are sure to outperform constant beta models. Yet he argues that if beta risk is misspecified, we may commit serious pricing errors which could be even larger than a constant beta model, and he shows that this has indeed happened in many cases. Does the nonparametric version of the conditional CAPM outperform the unconditional CAPM? We find that the conditional CAPM performs much better than the constant beta version, for each and every one of our test portfolios![22] Panel (D) reports the means and root mean squared errors (RMSE) of the pricing errors. Clearly, the conditional CAPM is a substantial improvement over the static CAPM. 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.[23]

5. Conclusion

In this paper, we propose three nonparametric methods for testing conditional mean-variance efficiency of a benchmark portfolio. These approaches are easy to

[22] Using $\hat{\varepsilon}_{i,t} = \hat{g}_i(x_t) - \hat{g}_p(x_t)\hat{g}_{ip}(x_t)/\hat{g}_{pp}(x_t)$ to estimate the pricing error at time t for asset i, we get the same conclusion, where $\hat{g}_p(x)$ and $\hat{g}_{pp}(x)$ are defined as in Section 2, and $\hat{g}_i(x)$ and $\hat{g}_{ip}(x)$ are kernel regression estimates defined similarly. For this method, we use the normal kernel defined in Appendix D with bandwidth $h = N^{-1/(k+4)}$.

[23] To check the robustness, we have studied the pricing errors of the conditional CAPM with the 25 size and book-to-market portfolios. The results are very similar and, hence, omitted for brevity. In short, we find in regressions as those in (C) that the EWR is the only statistically significant regressor. The EWR slopes have a negative relation to size, but no obvious relation to BE/ME. The standard deviations of the pricing errors with the 25 portfolios exhibit a strong negative relation to size, but no clear BE/ME pattern.

implement and robust to functional form misspecification. Moreover, they have a pleasant property that the estimators underlying the tests converge at the fast parametric rate. We derive limiting distributions of the test statistics and compare the three methods in a simulation study. In an application of the methods, we provide empirical results on testing the conditional CAPM.

This paper raises a number of challenging issues for future research. For example, Bansal and Viswanathan (1993), and Bansal et al. (1993) are the first to advocate the idea of using a flexible pricing kernel in empirical asset pricing. They propose a series expansion GMM approach to testing nonlinear APT models. It will be interesting to compare our testing methods with the series expansion GMM approach. Another challenging issue is if we can construct consistent nonparametric tests that are also easy to implement and powerful in finite sample applications. Finally, perhaps the most important issue is selection of conditioning variables. To cope with potential data snooping biases, an interesting project is to explore a more general test that combines the White's (2000) reality check bootstrap method with the nonparametric methods developed in this paper.

Acknowledgements

This paper is based upon part of my Ph.D. dissertation at the University of Chicago. I thank Yacine Aït-Sahalia, Ravi Bansal, Michael Brandt, Phelim Boyle, Xiaohong Chen, John Cochrane, George Constantinides, Philip Dybvig, David Hsieh, Lars Hansen, Kris Jacobs, Ravi Jagannathan, Raymond Kan, Ekaterini Kyriazidou, Kai Li, Robert McCulloch, Tom McCurdy, Ruey Tsay, Guofu Zhou, an anonymous referee for helpful comments and suggestions, and the Connaught Fund at the University of Toronto for research support.

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 (1976), 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 finance applications. In particular, restrictions on higher order moment existence are kept at a minimal level, due to concerns about possibly fat-tailed distributions of asset returns. In addition, simplicity of conditions on the bandwidth and the kernel is pursued to facilitate practical choices. Moreover, the β-mixing condition below requires no change for higher order extensions. Under this set of assumptions, we provide a large sample justification for the regression testing approach.

A.1. 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 $\{y_t, y_{t-1}, \dots\}$. In addition, define

$$W_1(t,s) \equiv \left(r_{p,s+1}^2 - r_{p,s+1}r_{p,t+1}\right)r_{t+1} + \left(r_{p,t+1}^2 - r_{p,t+1}r_{p,s+1}\right)r_{s+1},$$

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

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

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

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

A.2. 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, $n = 1, 2, \dots,$ 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) $\|W_1(t,\ s)\|_\rho < \infty$ and $\|W_2(t,\ s)\|_\rho < \infty$ for all $t < s$; (iii) $\|\eta(y_{t+1})\|_\rho < \infty$ 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

$$\int K(u)\mathrm{d}u = 1,$$

$$\int |u|^{j}|K(u)|\mathrm{d}u < \infty \qquad \text{if } 0 \leq j \leq k+1,$$

$$\int u_1^{l_1} \ldots u_k^{l_k} K(u)\mathrm{d}u = 0 \qquad \text{if } 0 < l_1 + \ldots + l_k < k+1,$$

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

Assumption A5. (i) The jth partial derivatives of fg_p, fg_{pp}, fg_r, fg_{pr}, and fg_{zz} exist for all $j \leq k+1$; (ii) The expectations $E[g_{pr}\nabla_{l_1,\ldots,l_j}(fg_p)], E[g_r\nabla_{l_1,\ldots,l_j}(fg_{pp})]$, $E[g_{pp}\nabla_{l_1,\ldots,l_j}(fg_r)], E[g_p\nabla_{l_1,\ldots,l_j}(fg_{pr})], E[g_{pp}\nabla_{l_1,\ldots,l_j}(fg_{zz})]$ exist for all $j \leq k+1$, 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 Assumption 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 the use of a higher order kernel (see, Powell et al., 1989).

Appendix B. Proofs of the propositions

We use three basic lemmas for the proofs of the propositions. Proofs of these lemmas are omitted but available upon request.

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, $1 \leq j \leq m-1$, and $t_1 < \ldots < t_m$, let $\xi_{j+1}, \ldots, \xi_m$ 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*

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

and

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

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

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

Define two generalized U-statistics of second order:

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

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

where

$$q_N(y_{t+1}, y_{s+1}) \equiv h^{-k} K\left(\frac{x_t - x_s}{h} \right) \left\{ W_1(t,s) - \left[\iota_n \otimes W_2(t,s) \right] \delta \right\},$$

$$\kappa_N(y_{t+1}, y_{s+1}) \equiv h^{-k} K\left(\frac{x_t - x_s}{h} \right) W_2(t,s).$$

Note that q_N and κ_N are symmetric and they vary with N through h.
For application of the Hoeffding projection technique, define

$$\gamma_N(y) \equiv \int q_N(y, y_{s+1}) \mathrm{d}F(y_{s+1}),$$

$$\hat{H}_{1N} \equiv \frac{1}{2} E\gamma_N(y_{t+1}) + \frac{1}{N} \sum_{t=1}^{N} \left[\gamma_N(y_{t+1}) - E\gamma_N(y_{t+1}) \right],$$

$$a_N(y) \equiv \int \kappa_N(y, y_{s+1}) \mathrm{d}F(y_{s+1}),$$

$$\hat{H}_{2N} \equiv \frac{1}{2} Ea_N(y_{t+1}) + \frac{1}{N} \sum_{t=1}^{N} \left[a_N(y_{t+1}) - Ea_N(y_{t+1}) \right],$$

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

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

$$NE\left[(U_{1N}-H_{1N})(U_{1N}-H_{1N})'\right]=o(1).$$

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

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

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 Proposition 1. From the definition of the WLS estimator $\hat{\delta}_N$, it follows that

$$\hat{A}_N(\hat{\delta}_N-\delta)=\left(1-\frac{1}{N}\right)U_{1N},$$

and

$$\hat{A}_N=\left(1-\frac{1}{N}\right)\iota_n\otimes U_{2N}.$$

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

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

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

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

This gives

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

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

$$\sqrt{N}\,(U_{1N} - EH_{1N}) = \sqrt{N}\,(H_{1N} - EH_{1N}) + o_p(1). \tag{18}$$

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),

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

Since $Ea(y_{t+1}) = 2E(w_t z_t z_t')$, it then follows that $\hat{A}_N \xrightarrow{p} A$.
Lemma 3(i) also gives

$$\begin{aligned} \sqrt{N}\,(H_{1N} - EH_{1N}) &\equiv \frac{1}{\sqrt{N}}\sum_{t=1}^{N}\left[\gamma_N(y_{t+1}) - E\gamma_N(y_{t+1})\right] \\ &= \frac{1}{\sqrt{N}}\sum_{t=1}^{N}\left[\gamma(y_{t+1}) - E\gamma(y_{t+1})\right] + o_p(1). \end{aligned} \tag{19}$$

By Eqs. (18), (19) and Lemma 3(ii),

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

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

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

where $e_{t+1} = (e_{1,t+1} \dots e_{n,t+1})' \otimes z_t$. Thus, $E\eta(y_{t+1}) - [\iota_n \otimes Ea(y_{t+1})]\delta = 0$.
By the central limit theorem of Doukhan et al. (1994)[24],

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

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

[24] 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 Proposition 2 (Proof of part (ii)). Define

$$\tilde{\gamma}_N(y_{t+1}) = \hat{\eta}_N(y_{t+1}) - [\iota_n \otimes \hat{a}_N(y_{t+1})]\,\delta.$$

Since $\hat{\delta}_N \xrightarrow{p} \delta$, it suffices for part (ii) to show that

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

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 $s_1 < s_2$

$$\int J_N(t,s_1)\,J_N(t,s_2)'\,\mathrm{d}F(z_{j+1}) = 0,$$

where $j = s_1$ if $s_1 < t$, and $j = s_2$ otherwise.
Applying Lemma 1, we have

$$\begin{aligned} N^2 E\{[\tilde{\gamma}_N(y_{t+1}) - \gamma_N(y_{t+1})][\tilde{\gamma}_N(y_{t+1}) - \gamma_N(y_{t+1})]'\} \\ \le \sum_{s_1=1}^{N} |EJ_N(t,s_1)\,J_N(t,s_1)'| + 2\sum_{1\le s_1<s_2\le N} |EJ_N(t,s_1)\,J_N(t,s_2)'| \\ \le NM_N + 8NM_N \sum_{n=1}^{N-1} \beta_n^{(\rho-2)/\rho} = o(N^2). \end{aligned}$$

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

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

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),

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

This establishes part (ii) of the theorem that $\hat{\Gamma}_j$ is consistent for Γ_j.
Proof of part (i): Given Eqs. (1) and (2) of Section 2,

$$E\left[\eta(y_{t+1})\,|\,I_t\right] = E(w_t e_{t+1}\,|\,I_t) + E(w_t e_{t+1}\,|\,x_t)$$
$$E\left[a(y_{t+1})\,|\,I_t\right] = w_t z_t z_t' + w_t g_{zz}(x_t)$$

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

$$E(w_t e_{t+1}\,|\,I_t) - \left[\iota_n \otimes (w_t z_t z_t')\right]\delta = 0,$$
$$E(w_t e_{t+1}\,|\,x_t) - \left[\iota_n \otimes (w_t g_{zz}(x_t))\right]\delta = 0.$$

Thus, $E[\gamma(y_{t+1})\,|\,I_t] = 0$ if Eq. (6) holds. It follows that $\Gamma_j = 0$ for $j \neq 0$.
Given Eqs. (1) and (2), the null hypothesis of conditional efficiency implies Eq. (6). So $\hat{\Omega}_N \xrightarrow{p} \Omega$. Then, with $\delta = 0$, Proposition 1 delivers the result: $N\hat{\delta}_N'\hat{\Omega}_N^{-1}\hat{\delta}_N \xrightarrow{d} \chi^2(qn)$. □

Appendix C. The covariance matrix estimators $\hat{\Sigma}_N$ and $\hat{\Lambda}_N$

The estimator $\hat{\Sigma}_N$ in Eq. (16) of Section 2 is

$$\hat{\Sigma}_N = \frac{1}{N}\sum_{t=1}^{N}\hat{\gamma}_N(y_{t+1})\hat{\gamma}_N(y_{t+1})',$$

where

$$\hat{\gamma}_N(y_{t+1}) = \hat{\eta}_N(y_{t+1}) - [\iota_n \otimes \hat{a}_N(y_{t+1})]\hat{a}_N,$$

$$\hat{\eta}_N(y_{t+1}) = \hat{f}^2(x_t)\Big[\hat{g}_{pp}(x_t)r_{t+1} - \hat{g}_p(x_t)r_{p,t+1}r_{t+1} + \hat{g}_r(x_t)r_{p,t+1}^2 - \hat{g}_{pr}(x_t)r_{p,t+1} + \hat{g}_{pp}(x_t)\hat{g}_r(x_t) - \hat{g}_p(x_t)\hat{g}_{pr}(x_t)\Big],$$

$$\hat{a}_N(y_{t+1}) = \hat{f}^2(x_t)\Big[\hat{\sigma}_p^2(x_t)z_t z_t' + \Big(r_{p,t+1}^2 - 2\hat{g}_p(x_t)r_{p,t+1} + \hat{g}_{pp}(x_t)\Big)\hat{g}_{zz}(x_t)\Big].$$

Here, $\hat{f}$, $\hat{g}_p$, $\hat{g}_{pp}$, $\hat{g}_r$, $\hat{g}_{pr}$, $\hat{g}_{zz}$, and r_{t+1} are defined as in Section 2. Note that $g_{zz}(x_t) = z_t z_t'$ when z_t is a fixed transformation of x_t (e.g., $z_t = (1\ x_t')'$). In such circumstances, one can simply use

$$\hat{a}_N(y_{t+1}) = \hat{f}^2(x_t)\Big[\hat{\sigma}_p^2(x_t) + r_{p,t+1}^2 - 2\hat{g}_p(x_t)r_{p,t+1} + \hat{g}_{pp}(x_t)\Big]z_t z_t'.$$

The estimator $\hat{\Lambda}_N$ in Eq. (17) of Section 2 is

$$\hat{\Lambda}_N = \frac{1}{N}\sum_{t=1}^{N}\hat{\gamma}_N(y_{t+1})\hat{\gamma}_N(y_{t+1})' - 4\hat{\lambda}_N\hat{\lambda}_N',$$

where

$$\hat{\gamma}_N(y_{t+1}) = \hat{f}(x_t)\Big[\hat{g}_{pp}(x_t)r_{t+1} - \hat{g}_p(x_t)r_{p,t+1}r_{t+1} + \hat{g}_r(x_t)r_{p,t+1}^2 - \hat{g}_{pr}(x_t)r_{p,t+1}\Big],$$

and $\hat{f}$, $\hat{g}_p$, $\hat{g}_{pp}$, $\hat{g}_r$, $\hat{g}_{pr}$, and r_{t+1} are defined as in Section 2.

Appendix D. Kernel and bandwidth

Two kernel functions are used in the nonparametric tests. 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 ith conditioning variable. This kernel is referred to as the normal kernel.

The second one is a "bias-corrected" kernel, which is constructed from the normal kernel as follows (see, Schucany and Summers, 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 $a = (a_1, \dots, a_k)'$, 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 $j = 1, \dots, k$. It is easy to verify that this is a kernel of order $k+1$ satisfying Assumption A4 in Appendix A.

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 in Proposition 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).[25]

The same kernels and bandwidth are used for the unconditional moments test given in Eq. (17) of Section 2. For the regression test in Eq. (16) of Section 2, a kernel of order $2k+1$ is required. We construct it similarly as above. The bandwidth is set to be $h = N^{-1/(4k+1)}$, to meet the convergence rate conditions.

References

Aït-Sahalia, Y., 1996. Nonparametric pricing of interest rate derivative securities. Econometrica 64, 527–560.

Bansal, R., Viswanathan, S., 1993. No arbitrage and arbitrage pricing: a new approach. Journal of Finance 48, 1231–1262.

Bansal, R., Hsieh, D.A., Viswanathan, S., 1993. A new approach to international arbitrage pricing. Journal of Finance 48, 1719–1747.

Bollerslev, T., Engle, R.F., Woodridge, J.M., 1988. A capital asset pricing model with time-varying covariances. Journal of Political Economy 96, 116–131.

[25] We have tried a cross-validation method to determine constant c in $h = cN^{-1/(2k+1)}$. For the postwar monthly data set, we found that $c = 1.04$, which is very close to one.

Brandt, M.W., 1999. Estimating portfolio and consumption choice: a conditional Euler equation approach. Journal of Finance 54, 1609–1645.

Carhart, M.M., Krail, R.J., Steven, R.L., Welch, K.D., 1995. Testing the conditional CAPM. Working Paper, Graduate School of Business, University of Chicago.

Chapman, D.A., Pearson, N.D., 2000. Is the short rate drift actually nonlinear? Journal of Finance 55, 355–388.

Chen, X., Fan, Y., 1999. Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. Journal of Econometrics 91, 373–401.

Christopherson, J.A., Ferson, W.E., Classman, D.A., 1998. Conditioning manager alphas on economic information: another look at the persistence of performance. Review of Financial Studies 11, 111–142.

Clare, A.D., Smith, P.N., Thomas, S.H., 1997. UK stock returns and robust test of mean variance efficiency. Journal of Banking and Finance 21, 641–660.

Cochrane, J.H., 1996. A cross-sectional test of an investment-based asset pricing model. Journal of Political Economy 104, 572–621.

Doukhan, P., Massart, P., Rio, E., 1994. The functional central limit theorem for strongly mixing processes. Annales de l'Institut Henri Poincare Probabilites et Statistiques 30, 63–82.

Engel, C., Frankel, J.A., Froot, K.A., Rorigues, A.P., 1995. Tests of conditional mean-variance efficiency of the U.S. stock market. Journal of Empirical Finance 2, 3–18.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on bonds and stocks. Journal of Financial Economics 33, 3–56.

Ferson, W.E., Foerster, S.R., 1994. Finite sample properties of the generalized method of moments in tests of conditional asset pricing models. Journal of Financial Economics 36, 29–55.

Ferson, W.E., Harvey, C.R., 1999. Conditioning variables and the cross section of stock returns. Journal of Finance 54, 1325–1360.

Ferson, W.E., Schadt, R.W., 1996. Measuring fund strategy and performance in changing economic conditions. Journal of Finance 51, 425–461.

Ferson, W.E., Siegel, A.F., 1998. Stochastic discount factor bounds with conditioning information. Working Paper, University of Washington.

Ferson, W.E., Kandel, S., Stambaugh, R.F., 1987. Tests of asset pricing with time-varying expected risk premiums and market betas. Journal of Finance 42, 201–220.

Ghysels, E., 1998. On stable factor structures in the pricing of risk: Do time varying betas help or hurt? Journal of Finance 53, 549–574.

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

Härdle, W., 1990. Applied Nonparametric Regression. Cambridge Univ. Press, New York.

Härdle, W., Stoker, T.M., 1989. Investigating smooth multiple regression by the method of average derivatives. Journal of the American Statistical Association 84, 986–995.

Harvey, C.R., 1989. Time-varying conditional covariances in tests of asset pricing models. Journal of Financial Economics 24, 289–317.

Harvey, C.R., 1991. The specification of conditional expectations. Working Paper, Fuqua School of Business, Duke University.

He, J., Kan, R., Ng, L., Zhang, C., 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., Linnik, Y.V., 1971. Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff, Groningen, Netherlands.

Jagannathan, R., Wang, Z., 1996. The conditional CAPM and the cross-section of expected returns. Journal of Finance 51, 3–53.

Khashimov, S.A., 1993. The central limit theorem for generalized *U*-statistics for weakly dependent vectors. Theory of Probability and Applications 38, 456–469.
Lee, B., 1992. A heteroskedasticity test robust to conditional mean misspecification. Econometrica 60, 159–171.
Lintner, J., 1965. The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.
McLeish, D.L., 1975. A maximal inequality and dependent strong laws. Annals of Probability 3, 829–839.
Pagan, A., Schwert, G.W., 1990. Alternative models for conditional stock volatility. Journal of Econometrics 45, 267–290.
Politis, D., Romano, J., 1994. The stationary bootstrap. Journal of the American Statistical Association 89, 1303–1313.
Powell, J.L., Stock, J.H., Stoker, T.M., 1989. Semiparametric estimation of index coefficients. Econometrica 57, 1403–1430.
Pritsker, M., 1998. Nonparametric density estimation and tests of continuous time interest rate models. Review of Financial Studies 11, 449–487.
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., Summers, J.P., 1977. Improvement of kernel-type density estimators. Journal of the American Statistical Association 72, 420–423.
Shanken, J., 1990. Intertemporal asset pricing: an empirical investigation. Journal of Econometrics 45, 99–120.
Sharpe, W.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. Chapman & Hall, London.
Wang, K.Q., 1998. Tests of conditional asset pricing models: a new approach. Ph.D. dissertation. Graduate School of Business, University of Chicago.
White, H., 2000. A reality check for data snooping. Econometrica 68, 1097–1126.
White, H., Domowitz, I., 1984. Nonlinear regressions with dependent observations. Econometrica 52, 143–161.
Yoshihara, K., 1976. Limiting behavior of U-statistics for stationary absolutely regular processes. Z. Wahrscheinlichkeitstheorie verw. Gebiete 35, 237–252.
Yoshihara, K., 1990. Limiting behavior of generalized quardratic forms generated by absolute regular sequences. Yokohama Mathematical Journal 37, 109–123.