



# Monotonicity in asset returns: New tests with applications to the term structure, the CAPM, and portfolio sorts<sup>☆</sup>

Andrew J. Patton<sup>a,b</sup>, Allan Timmermann<sup>c,d,\*</sup>

<sup>a</sup> Department of Economics, Duke University, Durham, NC 27708-0097, USA

<sup>b</sup> Oxford-Man Institute of Quantitative Finance, Oxford OX2 6ED, UK

<sup>c</sup> Rady School of Management, University of California San Diego, La Jolla, CA 92093-0553, USA

<sup>d</sup> CREATES, Denmark

## ARTICLE INFO

### Article history:

Received 17 October 2008

Received in revised form

30 April 2009

Accepted 25 November 2009

Available online 23 June 2010

### JEL classification:

G12

G14

### Keywords:

Asset pricing models

Portfolio sorts

CAPM

Monotonicity tests

## ABSTRACT

Many theories in finance imply monotonic patterns in expected returns and other financial variables. The liquidity preference hypothesis predicts higher expected returns for bonds with longer times to maturity; the Capital Asset Pricing Model (CAPM) implies higher expected returns for stocks with higher betas; and standard asset pricing models imply that the pricing kernel is declining in market returns. The full set of implications of monotonicity is generally not exploited in empirical work, however. This paper proposes new and simple ways to test for monotonicity in financial variables and compares the proposed tests with extant alternatives such as *t*-tests, Bonferroni bounds, and multivariate inequality tests through empirical applications and simulations.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

Finance contains many examples of theories implying that expected returns should be monotonically decreasing or monotonically increasing in securities' risk or liquidity characteristics. For example, under the liquidity preference hypothesis, expected returns on treasury securities should

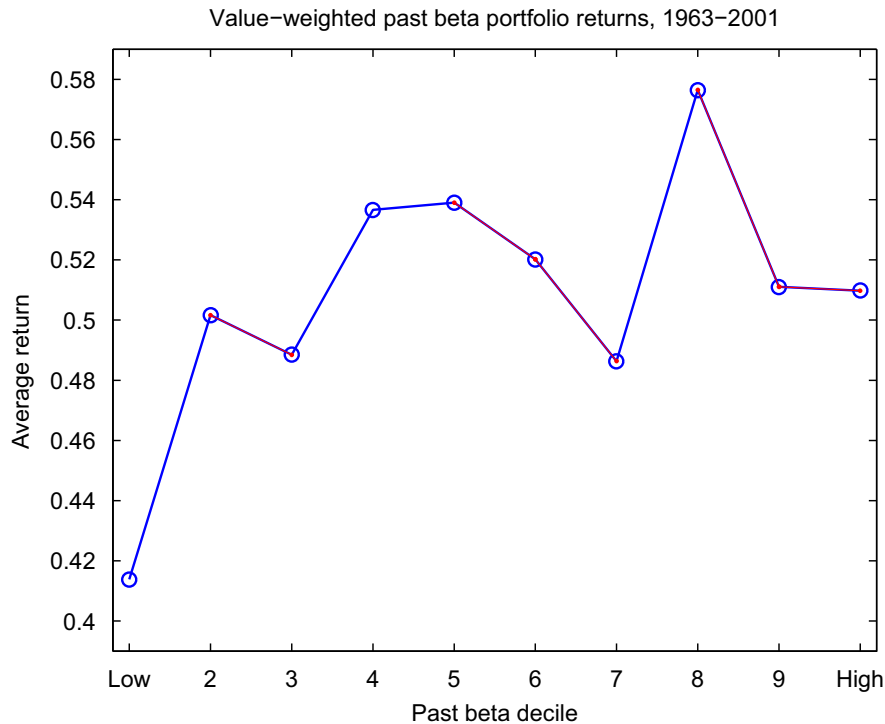
increase monotonically with their time to maturity. The Capital Asset Pricing Model (CAPM) implies a monotonically increasing pattern in the expected return of stocks ranked by their market betas. Another fundamental implication of finance theory is that the pricing kernel should be monotonically decreasing in investors' ranking of future states as measured, e.g., by market returns.

The full set of implications of such monotonic patterns is generally not explored, however, in empirical analysis. For example, when testing the CAPM, it is conventional practice to form portfolios of stocks ranked by their beta estimates. A *t*-test could then be used to consider the mean return spread between the portfolios with the highest and lowest betas. Yet comparing only the average returns on the top and bottom portfolios does not provide a sufficient way to test for a monotonic relation between expected returns and betas. As an illustration, Fig. 1 presents average monthly returns on stocks sorted into deciles according to their estimated betas. The mean return on the high-beta stocks exceeds that of the low-beta stocks, but a *t*-test on the top-

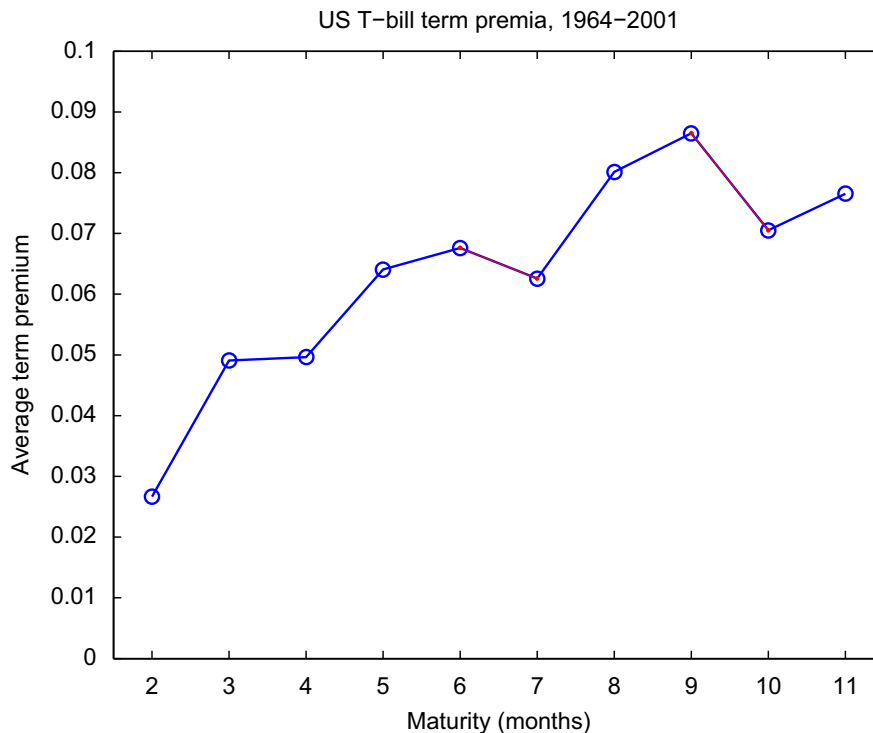
<sup>☆</sup> We thank an anonymous referee for many valuable suggestions. We also thank Susan Christoffersen, Erik Kole, Robert Kosowski, Jun Liu, Igor Makarov, Claudia Moise, Ross Valkanov, Simon van Norden, Michela Verardo, and seminar participants at HEC Montreal, the Adam Smith Asset Pricing workshop at Imperial College London, and the 2009 Western Finance Association conference for helpful comments and suggestions. Timmermann acknowledges support from CREATES, funded by the Danish National Research Foundation. This is a substantially revised version of the paper titled "Portfolio sorts and tests of cross-sectional patterns in expected returns."

\* Corresponding author at: Rady School of Management, University of California San Diego, La Jolla, CA 92093-0553, USA.

E-mail address: [atimmermann@ucsd.edu](mailto:atimmermann@ucsd.edu) (A. Timmermann).



**Fig. 1.** Average monthly returns on decile portfolios formed on past 12-month Capital Asset Pricing Model (CAPM) beta, from July 1963 to December 2001.



**Fig. 2.** Average monthly term premia for US T-bills, relative to a T-bill with one month to maturity, over the period January 1964 to December 2001.

minus-bottom return differential comes out insignificant. A test that considers only the return difference between the top and bottom ranked securities does not utilize the observation from Fig. 1 that none of the declining segments (which seemingly contradict the CAPM) appears to be particularly large, so the question arises whether the CAPM is, in fact, refuted by this evidence.

As a second illustration, Fig. 2 shows the term premia on T-bills with a maturity between two and 11 months. Clearly the overall pattern in the term premium is increasing, and this is confirmed by a *t*-test on the mean differential between the 11- and two-month bills, which comes out significant. However, there are also segments in which the term premium appears to be negative,

particularly between the nine and ten-month bills, so the question here is whether there are sufficiently many, and sufficiently large, negative segments to imply a rejection of the liquidity preference hypothesis. Only a test that simultaneously considers the mean returns across all maturities can answer this.

This paper proposes new ways to test for monotonicity in the expected returns of securities sorted by characteristics that theory predicts should earn a systematic premium. Our tests are nonparametric and easy to implement via bootstrap methods. Thus they do not require specifying the functional form (e.g., linearity) relating the sorting variables to expected returns or imposing distributional assumptions on returns. This is important because for many economic models the underlying hypothesis is only that expected returns should rise or decline monotonically in one or more security characteristics that proxy for risk exposures or liquidity.

In common with a conventional one-sided  $t$ -test, our monotonic relation (MR) test holds that expected returns are identical or weakly declining under the null, while under the alternative we maintain a monotonically increasing relation. (Testing for a monotonic decreasing relation can be accomplished by simply reordering the assets.) Thus a rejection of the null of no relation in favor of the hypothesized relation (i.e., a finding of statistical significance) represents a strong empirical endorsement of the theory. We also develop separate tests based on the sum of up and down moves. These tests combine information on both the number and magnitude of deviations from a flat pattern and so can help determine the direction of deviations in support of or against the theory.

The converse approach of maintaining a monotonically increasing relation under the null versus no such relation under the alternative has been developed by Wolak (1987, 1989) and was also adopted by Fama (1984) in the context of a Bonferroni bound test to summarize the outcome of several  $t$ -tests. Depending on the research question and the economic framework, one could prefer to entertain the presence of a monotonic relation under the null or under the alternative. For example, Richardson, Richardson and Smith (1992) use the Wolak test to see if there was evidence against an upward-sloping term structure of interest rates, as predicted by the liquidity preference hypothesis.

Because the MR and Wolak tests use different ways to test the theory, outcomes from such tests are not directly comparable. One drawback of entertaining the hypothesized monotonic relation under the null, as in the Wolak and Bonferroni tests, is that a confirmation of a theory from a failure to reject the null could simply be due to limited power for the test (because of a short time series of data or noisy data, for example). This turns out to be empirically important as the Wolak test sometimes fails to reject the null in cases in which the  $t$ -test and the MR test are able to differentiate between theories that find support in the data and those that do not. Conversely, in cases in which the MR test has weak power, it could fail to reject the null, and thus fail to support the theory, even for expected return patterns that appear to be monotonic.

Empirically, our tests reveal many interesting findings. For the CAPM example shown in Fig. 1, the MR test

strongly rejects the null in favor of a monotonically increasing relation between portfolio betas and expected returns. Consistent with this, the Bonferroni bound and Wolak tests fail to reject the null that expected returns increase in betas. Turning to the term structure example in Fig. 2, the MR, Bonferroni, and Wolak tests all fail to find evidence in support of the liquidity preference hypothesis as the term premia do not appear to be monotonically increasing in the maturity. Moreover, when applied to a range of portfolio sorts considered in the empirical finance literature, we find many examples where the difference in average returns between the top and bottom ranked portfolios is highly significant, but the pattern in average returns across multiple portfolios is non-monotonic. This holds, for example, for decile portfolios sorted on short-term reversal, momentum, or firm size.

Our tests are not restricted to monotonic patterns in the expected returns on securities sorted on one or more variables and can be generalized to test for monotonic patterns in risk-adjusted returns or in the factor loadings emerging from asset pricing models. They can also be adopted to test for piece-wise monotonic patterns, as in the case of the U-shaped relation between fee waivers and mutual fund performance reported by Christoffersen (2001) or the U-shaped pricing kernels considered by Bakshi, Madan, and Panayotov (2010). Finally, using methods for converting conditional moments into unconditional ones along the lines of Boudoukh, Richardson, Smith and Whitelaw (1999), we show that the approach can be used to conduct conditional tests of monotonicity.

The outline of the paper is as follows. Section 2 describes our new approach to testing for monotonic patterns in expected returns on securities ranked by one or more variables and compares it with extant methods. Section 3 uses Monte Carlo simulations to shed light on the behavior of the tests under a set of controlled experiments. Section 4 uses the various methodologies to analyze a range of return series from the empirical finance literature. Finally, Section 5 concludes.

## 2. Testing monotonicity

This section first provides some examples from finance to motivate monotonicity tests. We next introduce the monotonic relation test and then compare it with extant alternatives such as a student  $t$ -test based on top-minus-bottom return differentials, the multivariate inequality test proposed by Wolak (1989), and the Bonferroni bound.

### 2.1. Monotonicity tests in finance

One of the most basic implications of financial theory is that the pricing kernel should be monotonically decreasing in investors' ordering of future states (Shive and Shumway, 2009). In empirical work, this implication is typically tested by studying pricing kernels as a function of market returns. Using options data, Jackwerth (2000) finds that this prediction is supported by the data prior to the October 1987 crash, when risk aversion functions were monotonically declining. However, it

appears to no longer hold in post-crash data. [Rosenberg and Engle \(2002\)](#) also find evidence of a region with increasing marginal utility for small positive returns. These papers do not formally test monotonicity of the pricing kernel, however.

The practice of looking for monotonic patterns in expected returns on portfolios of stocks sorted by observables such as firm size or book-to-market ratio can be motivated by the fact that, although such variables are clearly not risk factors themselves, they could serve as proxies for unobserved risk exposures. For example, [Berk, Green, and Naik \(1999\)](#) develop a model of firms' optimal investment choices in which expected returns depend on a single risk factor. Estimating the true betas with regard to this risk factor requires knowing the covariance of each investment project in addition to the entire stock of ongoing projects, a task that is likely to prove infeasible. However, expected returns can be rewritten in terms of observable variables such as the book-to-market ratio and firm size, which become sufficient statistics for the risk of existing assets. Hence, expected returns on portfolios of stocks sorted on these variables should be monotonically increasing in book-to-market value and monotonically decreasing in firm size. Similar conclusions are drawn from the asset pricing model developed by [Carlson, Fisher and Giammarino \(2004\)](#).

As a second illustration, in a model of momentum effects in which growth rate risk rises with growth rates and has a positive price, [Johnson \(2002\)](#) shows that expected returns should be monotonically increasing in securities' past returns and uses portfolios to study this implication.

If factor loadings on risk factors are either observed or possible to estimate without much error, then tests based on linear asset pricing models could be preferable on efficiency grounds. However, in situations in which the nature (functional form) of the relation between expected returns and some observable variable used to rank or sort assets is unknown, the linear regression approach could be subject to misspecification biases. Hence, there is inherently a bias-efficiency trade-off between regression models that assume linearity, but make use of the full data, versus tests based on portfolio sorts that do not rely on this assumption. Tests of monotonicity between expected portfolio returns and observable stock characteristics such as book-to-market value or size offer a fairly robust way to evaluate asset pricing models, although they should be viewed as joint tests of the hypothesis that the sorting variable proxies for exposure to the unobserved risk factor and the validity of the underlying asset pricing model.

## 2.2. Monotonicity and inequality tests

The problem of testing for the presence or absence of a monotonic pattern in expected returns can be transformed into a test of inequality restrictions on estimated parameters. Consider a simple example in which decile portfolios have been formed by sorting stocks in ascending order based on their past estimated market betas. Let  $r_{1,t}, \dots, r_{10,t}$  be the associated returns on the decile portfolios listed. The CAPM implies that the expected

returns on these portfolios are increasing:

$$E[r_{10,t}] > E[r_{9,t}] > \dots > E[r_{1,t}]. \quad (1)$$

If we define  $\Delta_i \equiv E[r_{i,t}] - E[r_{i-1,t}]$ , for  $i=2, \dots, 10$ , this implication can be rewritten as

$$\Delta_i > 0 \quad \text{for } i = 2, \dots, 10. \quad (2)$$

Alternatively, consider a test of the liquidity premium hypothesis (LPH), as in [Richardson, Richardson and Smith \(1992\)](#) and [Boudoukh, Richardson, Smith and Whitelaw \(1999\)](#). If we define the term premium as  $E[r_t^{(\tau_i)} - r_t^{(1)}]$ , where  $r_t^{(\tau_i)}$  is the one-period return on a bond with maturity  $\tau_i$ , the simplest form of the LPH implies

$$E[r_t^{(\tau_i)} - r_t^{(1)}] > E[r_t^{(\tau_j)} - r_t^{(1)}], \quad \text{for all } \tau_i \geq \tau_j. \quad (3)$$

That is, term premia are increasing with maturity. If we define  $\Delta_i \equiv E[r_t^{(\tau_i)} - r_t^{(1)}] - E[r_t^{(\tau_{i-1})} - r_t^{(1)}]$ , then this prediction can be rewritten as

$$\Delta_i > 0 \quad \text{for } i = 2, \dots, N. \quad (4)$$

We next propose a new and simple nonparametric approach that tests directly for the presence of a monotonic relation between expected returns and the underlying sorting variable(s) but does not otherwise require that this relation be specified or known. This can be a great advantage in situations in which standard distributions are unreliable guides for the test statistics, difficult to compute, or simply unknown ([Ang and Chen, 2007](#)). Effectively our test allows us to examine whether there can exist a monotonic mapping from an observable characteristic used to sort stocks or bonds and their expected returns.

## 2.3. Testing for a monotonic relation: A new approach

Consider the ranking of expected returns on  $N+1$  securities. We take the number of securities,  $N+1$ , as given and then show how a test can be conducted that accounts for the relation between the complete set of securities (not just the top and bottom) and their expected returns. Denoting the expected returns by  $\mu = (\mu_0, \mu_1, \dots, \mu_N)'$ , and defining the associated return differentials as  $\Delta_i = \mu_i - \mu_{i-1}$ , we can use the link between monotonicity and inequality tests to consider tests on the parameter  $\Delta \equiv [\Delta_1, \dots, \Delta_N]'$ .

The approach proposed in this paper specifies a flat or weakly decreasing pattern under the null hypothesis and a strictly increasing pattern under the alternative, without requiring any maintained assumptions on  $\Delta^1$ :

$$H_0 : \Delta \leq \mathbf{0}$$

versus

$$H_1 : \Delta > \mathbf{0}. \quad (5)$$

The test is designed so that the alternative hypothesis is the one that the researcher hopes to prove, and in such cases it is sometimes called the research hypothesis ([Casella and Berger, 1990](#)). A theoretical prediction of a monotonic relation is therefore confirmed only if there is sufficient evidence in the data to support it. This is parallel

<sup>1</sup> Equalities and inequalities are interpreted as applying element by element for vectors.

to the standard empirical practice of testing the significance of the coefficient of a variable hypothesized to have a nonzero effect in a regression.

The null and alternative hypotheses in Eq. (5) can be rewritten as

$$H_0 : \Delta \leq \mathbf{0}$$

$$H_1 : \min_{i=1,\dots,N} \Delta_i > 0. \quad (6)$$

To see this, if the smallest value of  $\Delta_i > 0$ , then it must be that  $\Delta_i > 0$  for all  $i = 1, \dots, N$ . This motivates the following choice of test statistic:

$$J_T = \min_{i=1,\dots,N} \hat{\Delta}_i, \quad (7)$$

where  $\hat{\Delta}_i$  is based on the sample analogs  $\hat{\Delta}_i = \hat{\mu}_i - \hat{\mu}_{i-1}$ ,  $\hat{\mu}_i \equiv (1/T) \sum_{t=1}^T r_{it}$ , and  $\{r_{it}\}_{t=1}^T$  is the time series of returns on the  $i$ th security. We refer to the tests associated with hypotheses such as those in Eq. (6) as monotonic relation (MR) tests. Tests of monotonically decreasing expected returns simply reverse the order of the assets.

In Eq. (7) we consider all adjacent pairs of security returns. We could also consider all possible pair-wise comparisons,  $E[r_{i,t}] - E[r_{j,t}]$  for all  $i > j$ . The latter approach increases the number of parameter constraints, and the size of the vector  $\Delta$ , from  $N$  to  $N(N+1)/2$ . The adjacent pairs are sufficient for monotonicity to hold, but considering all possible comparisons could lead to empirical gains. We compare the adjacent pairs test to the “all pairs” test in our empirical analysis and Monte Carlo simulations. With the  $\Delta$  vector suitably modified, the theory presented below holds in both cases.

The proposed MR test for monotonicity is useful for detecting the presence or absence of a monotonic relation between expected returns and some economic variable. Because the test focuses on the smallest deviation from the null hypothesis, one can imagine patterns in expected returns for which the power of the test grows relatively slowly as the sample size expands. To address this concern, we propose two further measures, namely an Up and a Down statistic that account for both the frequency, magnitude, and direction of deviations from a flat pattern.<sup>2</sup>

<sup>2</sup> In particular, consider the following null and alternative:

$$H_0 : \Delta = \mathbf{0} \quad \text{versus} \quad H_1^- : \sum_{i=1}^N |\Delta_i| \mathbf{1}\{\Delta_i < 0\} > 0,$$

where the indicator  $\mathbf{1}\{\Delta_i < 0\}$  is one if  $\Delta_i < 0$  and is otherwise zero. Here the null is a flat pattern (no relation) and the alternative is that at least some parts of the pattern are strictly negative. By summing over all negative deviations, this statistic accounts for both the frequency and magnitude of deviations from a flat pattern. The natural test statistic is  $J_T^- = \sum_{i=1}^N |\hat{\Delta}_i| \mathbf{1}\{\hat{\Delta}_i < 0\}$ . As for the MR test, this Down test statistic does not have a standard limiting distribution under the null hypothesis, but critical values can be obtained using a bootstrap approach. The corresponding version of the test for cumulative evidence of an increasing pattern is

$$H_0 : \Delta = \mathbf{0} \quad \text{versus} \quad H_1^+ : \sum_{i=1}^N |\Delta_i| \mathbf{1}\{\Delta_i > 0\} > 0,$$

suggesting the Up test statistic  $J_T^+ = \sum_{i=1}^N |\hat{\Delta}_i| \mathbf{1}\{\hat{\Delta}_i > 0\}$ .

## 2.4. Extant tests and the choice of null and alternative hypotheses

The MR test is closely related to earlier work on multivariate inequality tests by Bartholomew (1961), Kudo (1963), Perlman (1969), Gouriéroux, Holly and Monfort (1982), and Wolak (1987, 1989). Wolak (1989) proposes a test that entertains (weak) monotonicity under the null hypothesis and specifies the alternative as non-monotonic:

$$H_0 : \Delta \geq \mathbf{0}$$

versus

$$H_1 : \Delta \text{ unrestricted.} \quad (8)$$

Here the theoretical prediction of monotonicity is contained in the null hypothesis and is rejected only if the data contain sufficient evidence against it. The test statistic in this approach is based on a comparison of an unconstrained estimate of  $\Delta$  with an estimate obtained by imposing weak monotonicity. Assuming that the data are normally distributed, Wolak (1987, 1989) shows that these test statistics have a distribution under the null that is a weighted sum of chi-squared variables,  $\sum_{i=1}^N \omega(N,i) \chi^2(i)$ , where  $\omega(N,i)$  are the weights and  $\chi^2(i)$  is a chi-squared variable with  $i$  degrees of freedom. Critical values are generally not known in closed form, but a set of approximate values can be calculated through Monte Carlo simulation. This procedure is computationally intensive and difficult to implement in the presence of large numbers of inequalities. As a result, the test has found only limited use in finance. Richardson, Richardson and Smith (1992) apply the method in Wolak (1989) to test for monotonicity of the term premium, and in our empirical work below we present the results of Wolak's test for comparison.

An important difference exists between the MR approach in Eq. (6) and that of Wolak (1989) in Eq. (8). In Wolak's framework, the null hypothesis is that a weakly monotonic relation exists between expected returns and the sorting variable, while the alternative hypothesis contains the case of no such monotonic relation. One potential drawback of entertaining the hypothesized monotonic relation under the null is that limited power (due to a short time series of data or noisy data) makes it difficult to reject the null hypothesis and thus difficult to have much confidence in a confirmation of a theory from a failure to reject the null.<sup>3</sup> The MR approach, meanwhile, contains the monotonic relation under the alternative, and thus a rejection of the null of no relation in favor of the hypothesized relation represents a strong empirical endorsement of the theory. Conversely, in cases in which the MR test has weak power, it could fail to reject the null and so incorrectly fail to support the theory entertained under the alternative hypothesis. In such cases the aforementioned Up and Down tests come in conveniently as they can help to diagnose if the problem is lack of power.

<sup>3</sup> Furthermore, the null in Eq. (8) includes the case of no relation (when  $\Delta = \mathbf{0}$ ) and so a failure to reject the null could be the result of the absence of a relation between expected returns and the sorting variable.



Because the setup of the null and alternative hypothesis under the MR test in Eq. (5) is the mirror image of that under the Wolak test in Eq. (8), one cannot draw universally valid conclusions about which approach is best. Instead, which test to use depends on the research question at hand. The MR test is more appropriate to use when the relevant question is: Does the data support the theory? Conversely, the Wolak setup is more appropriate for a researcher interested in finding out if there is significant evidence in the data against some theory. In cases where this distinction is not clear, one could even consider inspecting both types of tests. There is strong support for the theory if the MR test rejects while the Wolak test fails to reject. Conversely, if the Wolak test rejects while the MR test fails to reject, this constitutes strong evidence against theory. Cases in which both tests fail to reject constitute weak confirmation of the theory and could be due to the MR test having weak power. Finally, if both tests reject, they disagree about the evidence. We do not find a single case with this latter outcome in any of our empirical tests.

The MR test has greater apparent similarity to the setup of the multivariate one-sided tests considered by Bartholomew (1961), Kudo (1963), Perlman (1969), and Gouriéroux, Holly, and Monfort (1982), and is labeled “EI” in Wolak (1989):

$$H_0 : \Delta = \mathbf{0}$$

versus

$$H_1 : \Delta \geq \mathbf{0}, \quad (9)$$

with at least one inequality strict, under the maintained hypothesis  $H_m : \Delta \geq \mathbf{0}$ . The test statistic in this approach is based purely on an estimate of  $\Delta$  obtained by imposing the maintained assumption.<sup>4</sup> The main drawback of this framework, if one wishes to test for a monotonic relation, is that if the true relation is non-monotonic, then the behavior of the test is unknown, as the maintained hypothesis is then violated.<sup>5</sup> In a Monte Carlo study of this test (available upon request) we find that it performs well when the maintained hypothesis is satisfied. However, when this hypothesis is violated, the finite-sample size of the test tends to be very high, likely due to the fact that this test is not designed to work when the maintained hypothesis of weak monotonicity is violated. This leads the test to overreject and so we do not consider this test further here.

Lastly, a naïve approach to testing the hypotheses in Eq. (8) would be to conduct a set of pair-wise  $t$ -tests to see

<sup>4</sup> Kudo characterizes the weights analytically in cases with up to four constraints under the assumption that the covariance matrix of the parameter estimator is known. Gouriéroux, Holly and Monfort (1982) propose simulation methods to compute critical values when the covariance matrix is unknown. Kodde and Palm (1986) derive lower and upper bounds on the critical values for the test, which avoids the need for simulations.

<sup>5</sup> In contrast, the MR test is not derived from a statistic that imposes a maintained hypothesis in the estimation stage. Our test is therefore robust in the sense that it gives rise to the correct asymptotic distribution and inference in situations in which the data are from an unknown generating process.

if  $\Delta_i$  is positive for each  $i = 1, \dots, N$ . Unfortunately, it is not clear how to summarize information from these  $N$  tests into a single number because the test statistics are likely to be correlated and their joint distribution is unknown. To deal with this problem, Fama (1984) proposed using a Bonferroni bound. This method analyzes whether the smallest  $t$ -statistic on  $\hat{\Delta}_i$ ,  $i = 1, \dots, N$ , falls below the lower-tail critical value obtained by using a bound on the probability of a Type I error. The technique is simple to implement but tends to be a conservative test of the null hypothesis. This is confirmed in a Monte Carlo study reported in Section 3.

## 2.5. A bootstrap approach to the MR test

Under standard conditions, discussed in the Appendix, the estimated parameter vector  $\hat{\Delta} = [\hat{\Delta}_1, \dots, \hat{\Delta}_N]'$  asymptotically follows a normal distribution, i.e., in large samples ( $T \rightarrow \infty$ ),

$$\sqrt{T}([\hat{\Delta}_1, \dots, \hat{\Delta}_N]' - [\Delta_1, \dots, \Delta_N]') \overset{d}{\sim} \mathcal{N}(\mathbf{0}, \Omega). \quad (10)$$

Using this result would require knowledge, or estimation, of the full set of  $N(N+1)/2$  parameters of the covariance matrix for the sample moments,  $\Omega$ . These parameters influence the distribution of the test statistic even though we are not otherwise interested in them. Unfortunately, when the set of assets involved in the test grows large, the number of covariance parameters increases significantly and it can be difficult to estimate these parameters with much precision.

As shown in Eq. (7), we are interested in studying the minimum value of a multivariate vector of estimated parameters that is asymptotically normally distributed. There are no tabulated critical values for such minimum values, precisely because these would depend on the entire covariance matrix,  $\Omega$ . Furthermore, the asymptotic distribution might not provide reliable guidance to the finite sample behavior of the resulting tests.

To deal with the problem of not knowing the parameters of the covariance matrix or the critical values of the test statistic, we follow recent studies on financial time series such as Sullivan, Timmermann, and White (1999) and Kosowski, Timmermann, Wermers, and White (2006) and use a bootstrap methodology. As pointed out by White (2000), a major advantage of this approach is that it does not require estimating  $\Omega$  directly. While this approach dispenses with the need for making distributional assumptions on the data, conversely the approach might not be optimal in situations in which more information on the underlying return distribution is available.

To see how the approach works in practice, let  $\{r_{it}, t = 1, \dots, T; i = 0, 1, \dots, N\}$  be the original set of returns data recorded for  $N+1$  assets over  $T$  time periods. We first use the stationary bootstrap of Politis and Romano (1994) to randomly draw (with replacement) a new sample of returns  $\{\tilde{r}_{i\tau(t)}^{(b)}, \tau(1), \dots, \tau(T); i = 0, 1, \dots, N\}$ , where  $\tau(t)$  is the new time index, which is a random draw from the original set  $\{1, \dots, T\}$ . This randomized time index,  $\tau(t)$ , is common across

portfolios to preserve any cross-sectional dependencies in returns. Finally,  $b$  is an indicator for the bootstrap number which runs from  $b=1$  to  $b=B$ . The number of bootstrap replications,  $B$ , is chosen to be sufficiently large that the results do not depend on Monte Carlo errors. Time series dependencies in returns are accounted for by drawing returns data in blocks whose starting point and length are both random. The block length is drawn from a geometric distribution with a parameter that controls the average length of each block.

To implement the MR test, we need to obtain the bootstrap distribution of the parameter estimate  $\hat{\Delta}$  under the null hypothesis. The null in Eq. (6) is composite, and so, following White (2000), we choose the point in the null space least favorable to the alternative, namely,  $\Delta = \mathbf{0}$ .<sup>6</sup> The null is imposed by subtracting the estimated parameter  $\hat{\Delta}$  from the parameter estimate obtained on the bootstrapped return series,  $\hat{\Delta}^{(b)}$ . We then count the number of times when a pattern at least as unfavorable (i.e., yielding at least as large a value of  $J_T$ ) against the null as that observed in the real data emerges. When divided by the total number of bootstraps,  $B$ , this gives the  $p$ -value for the test and allows us to conduct inference:

$$J_T^{(b)} = \min_{i=1, \dots, N} (\hat{\Delta}_i^{(b)} - \hat{\Delta}_i), \quad b = 1, 2, \dots, B \quad (11)$$

and

$$\hat{p} = \frac{1}{B} \sum_{b=1}^B \mathbf{1}\{J_T^{(b)} > J_T\}. \quad (12)$$

When the bootstrap  $p$ -value is less than 0.05, we conclude that we have significant evidence against the null in favor of a monotonically increasing relation. We implement a studentized version of this bootstrap, as advocated by Hansen (2005) and Romano and Wolf (2005). This eliminates the impact of cross-sectional heteroskedasticity in the portfolio returns, a feature that is prominent for some securities and could lead to gains in power.

Theorem 1, given in the Appendix, provides a formal justification for the application of the bootstrap to our problem. Under a standard set of moment and mixing conditions on returns, the appropriately scaled vector of mean returns converges to a multivariate normal distribution. Hence, an important difference between the MR and Wolak tests is that, whereas the former does not make parametric assumptions on the distribution from which the data are drawn, the Wolak test assumes that the data are normally distributed. Moreover, inference about the minimum of a draw from the distribution of  $J_T$  can be conducted by means of the stationary bootstrap provided that the average block length grows with the sample size but at a slower rate.

<sup>6</sup> Analogously, in a simple one-sided test of a single parameter,  $H_0 : \beta \leq 0$  versus  $H_1 : \beta > 0$ , the point least favorable to the alternative under the null is zero.

## 2.6. Two-way sorts

Expected returns on financial securities are commonly modeled as depending on multiple risk or liquidity factors. In this subsection we show that the MR test is easily generalized to cover tests of monotonicity of expected returns based on two-way sorts.

Suppose that the outcome of the two-way sort is reported in an  $(N+1) \times (N+1)$  table with sorts according to one variable ordered across rows and sorts by the other variable listed along the columns. We are interested in testing the hypothesis that expected returns increase along both the columns and rows. The proposition of no systematic relation, which we seek to reject, is entertained under the null. To formalize the MR test in this case, let the expected value of the return on the row  $i$ , column  $j$  security be denoted  $\mu_{ij}$ :

$$H_0 : \mu_{ij} \leq \mu_{i-1,j}, \mu_{ij} \leq \mu_{i,j-1} \quad \text{for all } i,j. \quad (13)$$

The alternative hypothesis is that expected returns increase in both the row and column index:

$$H_1 : \mu_{ij} > \mu_{i-1,j}, \mu_{ij} > \mu_{i,j-1} \quad \text{for all } i,j. \quad (14)$$

Defining row  $\Delta_{ij}^r = \mu_{ij} - \mu_{i-1,j}$  and column  $\Delta_{ij}^c = \mu_{ij} - \mu_{i,j-1}$  differentials in expected returns, we can restate these hypotheses as

$$H_0 : \Delta_{ij}^r \leq 0, \Delta_{ij}^c \leq 0, \quad \text{for all } i,j$$

versus

$$H_1 : \Delta_{ij}^r > 0 \quad \text{and} \quad \Delta_{ij}^c > 0, \quad \text{for all } i,j, \quad (15)$$

or, equivalently,

$$H_1 : \min_{i,j=1, \dots, N} \{\Delta_{ij}^r, \Delta_{ij}^c\} > 0. \quad (16)$$

In parallel with the test for the one-way sort in Eq. (7), this gives rise to a test statistic

$$J_T = \min_{i,j=1, \dots, N} \{\hat{\Delta}_{ij}^r, \hat{\Delta}_{ij}^c\}. \quad (17)$$

The alternative hypothesis gives rise to  $2N(N-1)$  nonredundant inequalities. For a  $5 \times 5$  sort, this means 40 inequalities are implied by the theory of a monotonic relation in expected returns along both row and column dimensions; for a  $10 \times 10$  sort, 180 inequalities are implied. This shows both how potentially complicated and how rich the full set of relations implied by monotonicity can be when applied to returns sorted by two variables. If all pairs of returns are compared (not just the adjacent ones), we get  $(\frac{1}{2}N(N+1))^2 - N^2$  inequalities, which for a  $5 \times 5$  table yields 200 inequalities and for a  $10 \times 10$  table yields 2,925 inequalities.<sup>7</sup>

<sup>7</sup> These results are easily generalized to cases in which the number of rows and columns differs. For an  $N \times K$  table, there are  $2NK - K - N$  inequalities to test. Our results also generalize to sorts on three or more variables. For a  $D$ -dimensional sort, with  $N$  securities in each direction, the total number of inequalities amounts to  $DN^{D-1}(N-1)$ .

### 2.7. Monotonic patterns in risk-adjusted returns or factor loadings

The MR methodology can be extended to test for monotonic patterns in parameters other than the unconditional mean. For example, in a performance persistence study one might be interested in testing that risk-adjusted returns are monotonically increasing (or decreasing) in past performance. Alternatively, a corporate finance model could imply that the sensitivity of returns (or sales or free cash flows) to a credit constraint factor is monotonically decreasing in firm size. These examples are nested in the more general framework with  $K$  risk factors,  $\mathbf{F}_t = (F_{1t}, \dots, F_{Kt})'$ :

$$r_{it} = \beta_i' \mathbf{F}_t + e_{it}, \quad i = 0, 1, \dots, N$$

$$\beta_i \equiv (\beta_{i1}, \dots, \beta_{iK})', \quad (18)$$

with the associated hypotheses on the  $j$ th parameter in the above regression:

$$H_0 : \beta_{jN} \leq \beta_{jN-1} \leq \dots \leq \beta_{j0}$$

versus

$$H_1 : \beta_{jN} > \beta_{jN-1} > \dots > \beta_{j0} \quad (1 \leq j \leq K). \quad (19)$$

Our framework in the previous subsections corresponds to regressing each portfolio return onto a constant and so emerges when  $K=1$  and  $F_{1t}=1$  for all  $t$ . A test for monotonic risk-adjusted returns could be conducted by regressing returns onto a constant and a set of risk factors (for example, the Fama-French three-factor model) and then testing that the intercept (the alpha) from that regression is monotonically increasing. A test for monotonically increasing or decreasing factor sensitivity can be obtained by regressing returns on a constant, the factor of interest, and other control variables, and then testing that the coefficient on the relevant factor is monotonically increasing or decreasing.

The bootstrap regression for the general case takes the form

$$\tilde{r}_{it(t)}^{(b)} = \beta_i^{(b)'} \mathbf{F}_{it(t)}^{(b)} + e_{it(t)}^{(b)}, \quad i = 0, 1, \dots, N. \quad (20)$$

For each bootstrap sample an estimate of the coefficient vector is obtained. The null hypothesis is imposed by subtracting the corresponding estimate from the original data. From the recentered bootstrapped estimates,  $\hat{\beta}_i^{(b)} - \hat{\beta}_i$ , the test statistic for the bootstrap sample can be computed:

$$J_{j,T}^{(b)} \equiv \min_{i=1, \dots, N} [(\hat{\beta}_{j,i}^{(b)} - \hat{\beta}_{j,i}) - (\hat{\beta}_{j,i-1}^{(b)} - \hat{\beta}_{j,i-1})]. \quad (21)$$

By generating a large number of bootstrap samples the empirical distribution of  $J_{j,T}^{(b)}$  can be used to compute an estimate of the  $p$ -value for the null hypothesis, as in the simpler case presented in Section 2.5. The theorem in the Appendix covers this more general regression case and is based on the work of White (2000) and Politis and Romano (1994).

### 2.8. Conditional tests

Asset pricing models often take the form of conditional moment restrictions and so it is of interest to see how our tests can be generalized to this setting. Following Boudoukh, Richardson, Smith, and Whitelaw (1999), such a generalization is easily achieved by using the methods for converting conditional moment restrictions into unconditional moment restrictions commonly used in empirical finance.

To see how this works, let  $z_t$  be some instrument used to convert an unconditional moment condition into a conditional one. This instrument could take the form of an indicator variable that captures specific periods of interest corresponding to some condition being satisfied (e.g., the economy being in a recession) but could take other forms. The first step of the conditional MR test then pre-multiplies the set of returns,  $r_{it}$ , by  $z_t$ . In a second step, the test is conducted on the unconditional moments of the modified data  $\tilde{r}_{it} = r_{it} \times z_t$  along the lines proposed above.

## 3. Performance of the tests: A simulation study

The hypothesis tests proposed here are nonstandard. Unlike the standard  $t$ -test for equal expected returns, there are no optimality results or closed-form distributions against which test statistics such as those in Eq. (7) or Eq. (17) can be compared and from which critical values can be computed. To address this issue, we next undertake a series of Monte Carlo simulation experiments that offer insights into the finite-sample behavior of the proposed tests.

### 3.1. Monte Carlo setup

The first set of scenarios covers situations in which the hypothesized theory is valid and a monotonic relation exists between portfolio rank and the portfolios' true expected returns. We would like the MR tests to reject the null of no systematic relation in this situation (while the Wolak and Bonferroni tests should not reject) and the more often they reject, the more powerful they are.

Experiment I assumes monotonically increasing expected returns with identically sized increments between adjacent decile portfolios. Experiment II lets the expected return increase by 80% of the total from portfolio one through portfolio five, and then increase by the remaining 20% of the total across the remaining five portfolios. Experiment III assumes a single large increase in the expected return from decile one to decile two, equal to 50% of the total increase, and then spreads the remaining 50% of the increase across the nine remaining portfolios. These three patterns are illustrated in the first column of Fig. 3 and all have in common that the theory of a monotonic relation holds.

The second set of scenarios covers situations in which the theory fails to hold and a non-monotonic relation exists between portfolio ranks and expected returns, so



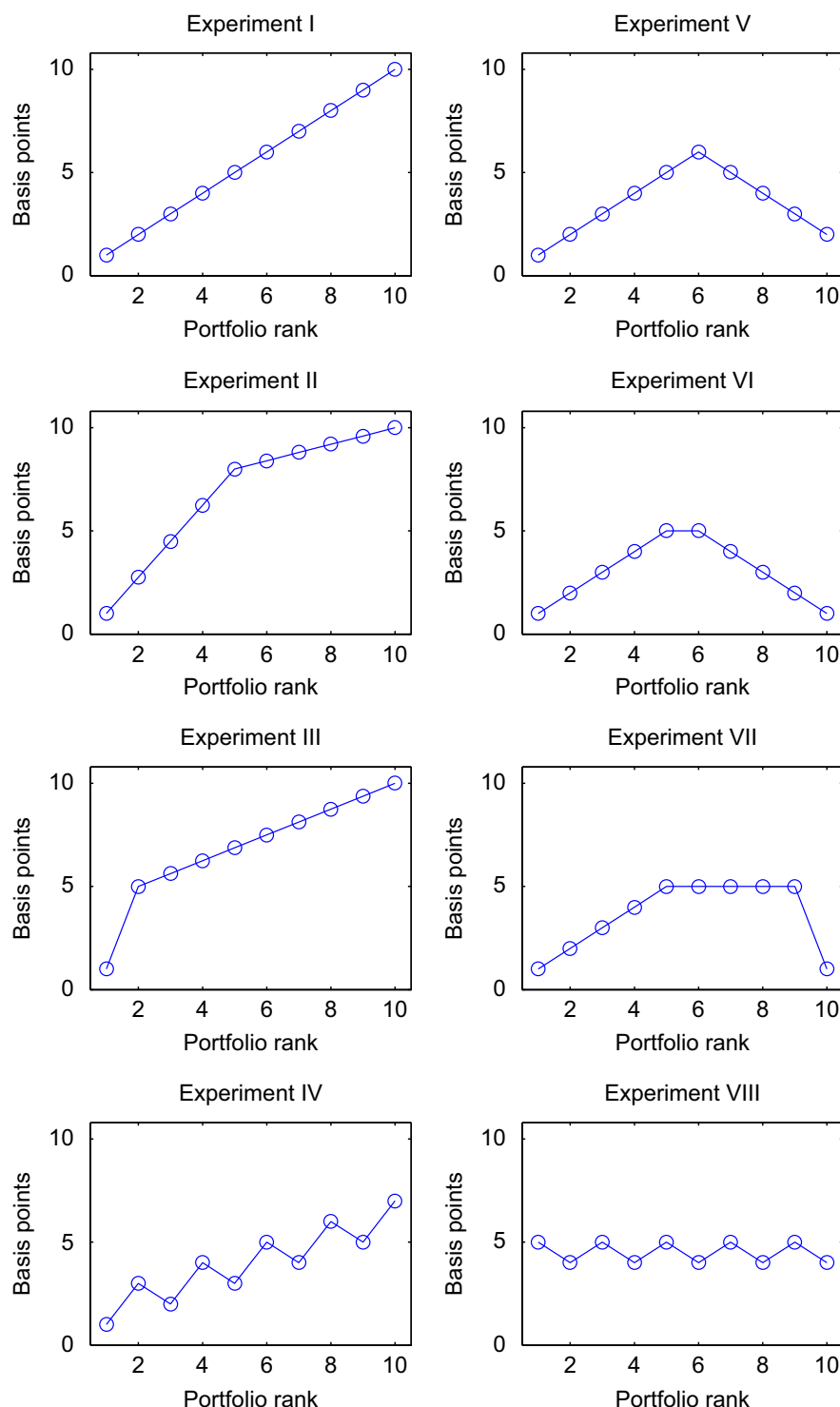


Fig. 3. Patterns in expected returns under the eight experiments considered in the Monte Carlo simulations for a step size of one basis point.

the MR test should not reject, while the Wolak and Bonferroni tests should reject.

Experiments IV–VIII all break the monotonic pattern in expected returns in some way. Experiment IV assumes an increasing but non-monotonic pattern with declines in expected returns for every second decile. Experiment V assumes a rising, then declining pattern in the expected return for a net gain in the expected return from the first

to the tenth portfolio. The next two experiments assume a pattern in which expected returns first rise and then decline so the expected return of the first and tenth deciles are identical, with the pattern being symmetric for Experiment VI and being smoothly increasing then flat and finally sharply decreasing for Experiment VII. Finally, Experiment VIII assumes a mostly flat, jagged pattern in expected returns.

Each pattern is multiplied by a step size that varies from a single basis point per month to two, five, and 10 basis point differentials in the expected returns.

To ensure that our experiments are computationally feasible and involve both a sufficiently large number of Monte Carlo draws of the original returns and a sufficient number of bootstrap iterations for each of these draws, we focus on a one-dimensional monotonic pattern with  $N=10$  assets. We present results based on two sets of assumptions. The first is based on a normality assumption, while the second set of results are based on more realistic data, where we use the bootstrap to reshuffle the true returns on the size-sorted decile portfolios and use these as our Monte Carlo simulation data. We draw 2,500 bootstrap samples of the original returns. For each simulated data set we employ  $B=1,000$  replications of the stationary bootstrap of Politis and Romano (1994) and use one thousand Monte Carlo simulations to get the weights required for the Wolak (1989) test. We consider two sample sizes for the simulation:  $T=966$  matches the full sample of data on size-sorted portfolios (1926–2006), and  $T=522$  matches the post-1963 sample of data on these portfolios.

### 3.2. Analytical results under normality

To obtain simple analytical results, we first make the assumption that the estimated differences in portfolio returns are independently and normally distributed, i.e.,  $\Delta\hat{\mu}_i \sim \mathcal{N}(\Delta\mu_i, (1/T)\sigma_i^2)$ , for  $i=2, \dots, N$ , and  $\text{Corr}[\Delta\hat{\mu}_i, \Delta\hat{\mu}_j] = 0$  for all  $i \neq j$ . This setup allows us to present formulas for the power of the tests and establish intuition for which results to expect. In particular, we can derive the power of the  $t$ -test analytically. First, note that the  $t$ -test is based on the difference between the mean returns of the  $N$ th and the first portfolios,

$$\hat{\mu}_N - \hat{\mu}_1 = \sum_{i=2}^N \Delta\hat{\mu}_i \sim \mathcal{N}\left(\sum_{i=2}^N \Delta\mu_i, \frac{1}{T} \sum_{i=2}^N \sigma_i^2\right). \quad (22)$$

Assuming that the variances are known, the  $t$ -statistic is

$$tstat \equiv \frac{\sqrt{T}(\hat{\mu}_N - \hat{\mu}_1)}{\sqrt{\sum_{i=2}^N \sigma_i^2}} \sim \mathcal{N}\left(\sqrt{T} \frac{\mu_N - \mu_1}{\sqrt{\sum_{i=2}^N \sigma_i^2}}, 1\right). \quad (23)$$

Under the null we have  $\mu_N = \mu_1$  and so the  $t$ -statistic has the usual  $\mathcal{N}(0,1)$  distribution. Under the alternative hypothesis that  $\mu_N > \mu_1$ , the  $t$ -statistic diverges as  $T \rightarrow \infty$ . For finite  $T$ , the probability of rejecting the null hypothesis using a one-sided test with a 5% critical value is then

$$\Pr[tstat > 1.645] = \Phi\left(\sqrt{T} \frac{\mu_N - \mu_1}{\sqrt{\sum_{i=2}^N \sigma_i^2}} - 1.645\right), \quad (24)$$

where  $\Phi(\cdot)$  is the *cdf* of a standard Normal distribution. Given a sample size,  $T$ , the vector of differences in expected returns  $\Delta\mu \equiv [\Delta\mu_2, \dots, \Delta\mu_N]'$ , and the vector of associated standard deviations  $\sigma \equiv [\sigma_2, \dots, \sigma_N]'$ , we can directly compute the power of the  $t$ -test.

For the MR test, the power is obtained as follows. To obtain the distribution of our test statistic,  $J_T \equiv \min_{i=2, \dots, N} \Delta\hat{\mu}_i$ , under the null, we use 100,000 simulated

draws to compute critical values, denoted  $J_T^*(\sigma)$ . The power of our test is then simply  $\Pr[J_T(\Delta\mu, \sigma) > J_T^*(\sigma)]$ . We compute this power using one thousand simulated draws.

Results for this benchmark case are presented in the first column of Table 1 labeled “Normal simulation.” For Experiments I, II, and III, the power of the  $t$ -test converges to one as the step size grows from one to 10 basis points. The probability of rejecting the null also approaches one for Experiment IV, which assumes a non-monotonic but increasing pattern of expected returns.

The MR test has somewhat lower power than the  $t$ -test for the three experiments in which both tests should reject the null (Experiments I, II, and III). Compensating for the reduction in power, we observe that the probability that the bootstrap rejects the null in Experiments IV–VIII, which would constitute a Type I error as these experiments do not have a monotonic pattern, goes to zero as the step size grows and never much exceeds the nominal size of the test. Thus the MR test is very unlikely to falsely reject the null hypothesis. In contrast, the  $t$ -test frequently rejects the null under Experiments IV and V when the step size is comparable to that observed for the majority of portfolio sorts in the empirical analysis in Section 3.3, i.e., five to 10 basis points. The  $t$ -test is not wrong. However, it has a limited scope because it compares only the top and bottom portfolios and thus fails to detect non-monotonic patterns in the full portfolio sorts.

### 3.3. Bootstrap simulation results

The second set of columns in Table 1 present the results from the simulation based on bootstrap draws of monthly returns on the size sorted decile portfolios presented in the empirical section using either the full sample (1926–2006, in Panel A) or a shorter subsample (1963–2006, in Panel B.) Results are first shown for the Wolak (1989) test and the Bonferroni bound test. Wolak’s test and the Bonferroni-based test have a weakly monotonic relation under the null hypothesis and a non-monotonic relation under the alternative. Hence, in contrast with the other tests, these tests should not reject the null for returns generated under Experiments I–III, while they should reject the null hypothesis under Experiments IV–VIII. We present the Up and Down tests along with the MR tests. We focus our discussion on the results for the long sample (Panel A), but the results are similar for the shorter sample.<sup>8</sup>

The first panel, with step size set to zero, shows that the Bonferroni, Wolak,  $t$ , Up, Down, MR, and MR<sup>all</sup> tests have roughly the correct size when there is genuinely no relation between expected returns and portfolio rank, although most of the tests slightly over-reject the null

<sup>8</sup> The generally small differences in power between the  $T=966$  and  $T=522$  simulations could indicate that the power of these tests is converging to unity relatively slowly. Differences in size and power across these two studies could arise from the fact that the former uses bootstrap shuffles of returns on the size-sorted portfolios from 1926 to 2006, while the latter uses returns from 1963 to 2006. Thus differences in the properties of this data across these sample periods could result in variations in the simulation results.

**Table 1**

Monte Carlo simulation results.

This table reports rejection frequencies for simulated data, using 2,500 Monte Carlo simulations. The columns report the proportion of rejections from tests of monotonicity, with nominal size of 0.05. Columns 1 and 2 report the proportion of rejections from a *t*-test applied to the spread in expected returns between the top and bottom portfolios and from the monotonic relation (MR) test, respectively. These results are obtained under the normality assumption discussed in Section 3. Columns 3–9 show the corresponding proportions of rejections using bootstrapped portfolio returns data obtained from Ken French's website and based on value-weighted decile portfolios sorted on market equity. The third and fourth columns (Wolak and Bonferroni) report the proportion of rejections of the null of a weakly monotonic relation using the Wolak (1989) test, with critical values based on one thousand simulations per replication, and a Bonferroni bound test. Column 5 is a standard *t*-test applied to the spread in expected returns between the top and bottom portfolios. The sixth and seventh columns report the proportion of rejections of the null of no relation against the alternative that the sum of positive (Up) or absolute values of negative (Down) differences in expected returns is nonzero. The final two columns report the proportion of rejections using the MR test based on the minimal set of possible inequalities (MR) or all possible inequalities implied by monotonicity (MR<sup>all</sup>). The latter four tests are based on the stationary bootstrap and use one thousand bootstrap replications.

“Step” refers to the step size in percent per month used in forming the spreads in expected returns on the decile portfolios. Experiments I–III impose a monotonic relation between portfolio rank and expected returns, and Experiments IV–VIII impose a non-monotonic pattern. Fig. 3 displays the shapes of the assumed patterns. The experiments are identical when the step size is zero and so only the results for Experiment I are presented in that case.

Simulated returns are generated as bootstrap samples from the monthly returns on size-sorted decile portfolios, using the stationary bootstrap with an average block length of 10 months. Panel A presents results when the full sample (1926–2006) of returns is used, generating  $T=966$  observations. Panel B presents results when a shorter sample (1963–2006) of returns is used, with  $T = 522$  observations.

Experiment	Normal simulation		Panel A: Monte Carlo simulation results for $T=966$						
	t-test	MR	Bootstrap simulation						
			Wolak	Bonferroni	t-test	Up	Down	MR	MR <sup>all</sup>
Step = 0									
I	0.050	0.050	0.036	0.034	0.074	0.078	0.032	0.064	0.090
Step = 0.01									
I	0.126	0.099	0.016	0.021	0.134	0.142	0.014	0.114	0.136
II	0.126	0.096	0.021	0.023	0.134	0.135	0.017	0.100	0.150
III	0.126	0.093	0.022	0.025	0.134	0.132	0.018	0.099	0.121
IV	0.095	0.067	0.038	0.037	0.112	0.124	0.021	0.071	0.093
V	0.056	0.048	0.046	0.044	0.082	0.083	0.035	0.056	0.063
VI	0.050	0.045	0.051	0.046	0.074	0.073	0.037	0.054	0.060
VII	0.050	0.048	0.027	0.036	0.074	0.075	0.036	0.046	0.046
VIII	0.045	0.047	0.036	0.048	0.080	0.068	0.036	0.064	0.066
Step = 0.02									
I	0.260	0.171	0.006	0.010	0.221	0.251	0.005	0.176	0.226
II	0.260	0.156	0.012	0.018	0.221	0.233	0.007	0.150	0.194
III	0.260	0.152	0.011	0.017	0.221	0.218	0.008	0.142	0.169
IV	0.164	0.064	0.055	0.058	0.166	0.212	0.017	0.061	0.078
V	0.063	0.042	0.073	0.062	0.087	0.091	0.039	0.043	0.053
VI	0.050	0.035	0.083	0.063	0.074	0.076	0.047	0.036	0.048
VII	0.050	0.025	0.075	0.121	0.074	0.096	0.039	0.032	0.032
VIII	0.040	0.038	0.050	0.055	0.075	0.073	0.039	0.066	0.066
Step = 0.05									
I	0.805	0.479	0.001	0.001	0.594	0.690	0.000	0.468	0.515
II	0.805	0.382	0.002	0.006	0.594	0.647	0.000	0.302	0.322
III	0.805	0.377	0.002	0.005	0.594	0.582	0.000	0.310	0.392

Table 1 (continued)

Experiment	Panel A: Monte Carlo simulation results for $T=966$								
	Normal simulation		Bootstrap simulation						
	t-test	MR	Wolak	Bonferroni	t-test	Up	Down	MR	MR <sup>all</sup>
IV	0.510	0.014	0.213	0.200	0.384	0.705	0.045	0.011	0.019
V	0.086	0.009	0.352	0.220	0.104	0.180	0.091	0.004	0.006
VI	0.050	0.007	0.374	0.222	0.074	0.114	0.013	0.003	0.006
VII	0.050	0.000	0.616	0.758	0.074	0.232	0.082	0.000	0.000
VIII	0.027	0.070	0.193	0.191	0.046	0.152	0.132	0.011	0.016
Step = 0.10									
I	1.000	0.847	0.000	0.000	0.966	0.984	0.000	0.837	0.803
II	1.000	0.681	0.001	0.002	0.966	0.979	0.000	0.571	0.596
III	1.000	0.729	0.000	0.000	0.966	0.950	0.000	0.648	0.703
IV	0.955	0.000	0.806	0.729	0.768	0.998	0.491	0.000	0.000
V	0.138	0.000	0.958	0.814	0.146	0.648	0.596	0.000	0.000
VI	0.050	0.000	0.960	0.814	0.074	0.401	0.657	0.000	0.000
VII	0.050	0.000	1.000	1.000	0.074	0.548	0.521	0.000	0.000
VIII	0.014	0.000	0.752	0.700	0.034	0.577	0.593	0.000	0.000
Panel B: Monte Carlo simulation results for $T=522$									
Step = 0									
I	0.050	0.050	0.072	0.058	0.056	0.064	0.056	0.044	0.041
Step = 0.01									
I	0.128	0.099	0.030	0.039	0.119	0.139	0.020	0.085	0.083
II	0.128	0.096	0.034	0.039	0.119	0.135	0.020	0.076	0.081
III	0.128	0.095	0.036	0.041	0.119	0.128	0.024	0.082	0.082
IV	0.096	0.068	0.053	0.048	0.096	0.113	0.040	0.056	0.058
V	0.056	0.051	0.074	0.058	0.063	0.075	0.055	0.045	0.045
VI	0.050	0.047	0.079	0.061	0.056	0.067	0.060	0.040	0.040
VII	0.050	0.047	0.068	0.065	0.056	0.073	0.057	0.046	0.047
VIII	0.044	0.044	0.080	0.064	0.049	0.065	0.060	0.042	0.040
Step = 0.02									
I	0.266	0.168	0.014	0.019	0.206	0.239	0.005	0.141	0.149
II	0.266	0.154	0.020	0.027	0.206	0.241	0.006	0.124	0.128
III	0.266	0.147	0.021	0.031	0.206	0.222	0.009	0.124	0.123
IV	0.167	0.067	0.050	0.066	0.142	0.202	0.031	0.057	0.057
V	0.063	0.044	0.084	0.067	0.072	0.092	0.054	0.033	0.033
VI	0.050	0.038	0.095	0.072	0.056	0.075	0.064	0.029	0.028
VII	0.050	0.035	0.089	0.098	0.056	0.096	0.059	0.039	0.041
VIII	0.039	0.032	0.097	0.085	0.046	0.070	0.073	0.032	0.033
Step = 0.05									
I	0.817	0.483	0.000	0.003	0.646	0.690	0.000	0.437	0.442
II	0.817	0.351	0.007	0.012	0.646	0.724	0.001	0.290	0.297
III	0.817	0.350	0.002	0.012	0.646	0.644	0.001	0.297	0.303
IV	0.521	0.018	0.186	0.202	0.386	0.752	0.067	0.017	0.020
V	0.087	0.013	0.290	0.200	0.089	0.208	0.085	0.007	0.006

VI	0.050	0.010	0.313	0.206	0.056	0.140	0.122	0.004	0.003
VII	0.050	0.003	0.367	0.505	0.056	0.229	0.087	0.004	0.004
VIII	0.027	0.006	0.281	0.251	0.031	0.130	0.186	0.011	0.012
Step = 0.10									
I	1.000	0.877	0.000	0.000	0.989	0.996	0.000	0.858	0.860
II	1.000	0.589	0.000	0.006	0.989	0.999	0.000	0.535	0.541
III	1.000	0.693	0.000	0.001	0.989	0.992	0.000	0.640	0.651
IV	0.960	0.000	0.730	0.674	0.846	1.000	0.526	0.000	0.000
V	0.140	0.000	0.870	0.699	0.125	0.750	0.480	0.000	0.000
VI	0.050	0.000	0.882	0.700	0.056	0.528	0.556	0.000	0.000
VII	0.050	0.000	0.979	0.997	0.056	0.667	0.354	0.000	0.000
VIII	0.014	0.000	0.834	0.714	0.018	0.544	0.709	0.000	0.000

hypothesis. This is not an unusual finding and mirrors results reported in simulation studies of the finite-sample size of asset pricing tests, see, e.g., Campbell, Lo, and MacKinlay (1997).

Under Experiment I, the Wolak and Bonferroni tests should not reject the null, and this is what we find. The bootstrapped results from the  $t$ -test and MR test are comparable to those obtained under the normality assumption. The  $t$ -test rejects slightly more frequently than the MR test. When the expected return differential increases by a single basis point per month for each decile portfolio, approximately 11–13% of the simulations correctly reject and this increases to around 20% under the two basis point differential. Under the five basis point return differential, the rejection rate is close to 50%. Finally, under the largest step size with a 10 basis point return differential per portfolio, the rejection rate is above 80%.

In Experiments II and III, the expected return pattern is monotonic but nonlinear, and both the  $t$  and MR tests should again reject the null hypothesis. The  $t$ -test uses only the difference in expected returns between portfolio one and portfolio 10 and so is unaffected by the presence of a kink in expected returns. The smallest step size affects the power of the MR test, which focuses on the minimum difference,  $\min_i \Delta \mu_i$ , and thus we expect this test to have lower power to detect patterns like Experiments II and II. This is what we find. For a step size equal to five basis points, for example, the power of the MR test is 30% in Experiments II and III, compared with 47% in Experiment I.

The column in Table 4 labeled “MR<sup>all</sup>” shows the result of using all possible pair-wise inequalities in the test. For a one-way sort with  $N=10$ , this entails comparing 45 instead of nine pairs of portfolio returns. There appears to be a small gain in power from including the full set of inequalities, although this could in part reflect that this approach leads to a slightly oversized test.

Turning to the second set of experiments involving a non-monotonic relation between expected returns and portfolio ranks, we expect Wolak’s test and the Bonferroni bound test to reject the null hypothesis of a weakly monotonic relation: For step sizes less than five basis points neither of these tests exhibit much power, but for step sizes of five and particularly 10 basis points they do detect the non-monotonic relation, with Wolak’s test in most cases having considerably better power than the Bonferroni bound test. Comparing Experiments VI and VII, we see that the power of both the Wolak and Bonferroni bound test is much greater in the presence of a single large deviation from the null compared with many small deviations that add up to the same total deviation. Importantly, in these experiments the MR tests very rarely reject, whereas the standard  $t$ -test does so frequently. For example, the  $t$ -test rejects 77% of the time in Experiment IV with the largest step size. These are cases in which we do not want a test to reject if the theory implies a monotonic relation between portfolio rank and expected returns.

Because the tests consider different hypotheses, their size and power are not directly comparable. The  $t$ -test compares only the top and bottom portfolio; the MR test considers all portfolios and continues to have equality of means as the null and inequality as the alternative; finally,



the Wolak and Bonferroni bound tests have weak inequality of expected returns under the null. Due to these differences, the tests embed different trade-offs in terms of size and power. While the  $t$ -test is powerful when expected returns are genuinely monotonically rising, this test cannot establish a uniformly monotonic pattern in expected returns across all portfolios and, as shown in Experiments IV–VIII, if used for this purpose can yield misleading conclusions. The Wolak test is not subject to this criticism. However, when this test fails to reject, this could simply be due to the test having weak power. Finally, the MR test has weak power for small steps in the direction hypothesized by the theory but appears to have good power for step sizes that match much of the empirical data considered below. Moreover, this test does not reject the null when the evidence contradicts the theory as in Experiments IV–VIII.<sup>9</sup>

#### 4. Empirical results

We finally revisit a range of examples from the finance literature. We compare the outcome of tests based on our new monotonic relation test or the Up and Down tests to a standard  $t$ -test, the Wolak (1989) test, and the Bonferroni bound.

Initially we consider empirical tests of the CAPM. An investor believing in this model would hold strong priors that expected stock returns and subsequent estimates of betas should be uniformly increasing in past estimates of betas, and so the CAPM is well suited to illustrate our methodology. We next consider the liquidity preference hypothesis, which conjectures that expected returns on Treasury securities rise monotonically with the time to maturity. Finally, we extend our analysis to a range of portfolio sorts previously considered in the empirical finance literature. In all cases we use one thousand bootstrap replications for the bootstrap tests and we choose the average block length to be 10 months, which seems appropriate for returns data that display limited time series dependencies at the monthly horizon. Finally, we employ one thousand Monte Carlo simulations to obtain the weight vector,  $\omega(N, i)$ , used to compute critical values in the Wolak (1989) test.

##### 4.1. Portfolio sorts on CAPM beta: Expected returns

We first test for an increasing relation between ex ante estimates of CAPM beta and subsequent returns, using the same data as in Ang, Chen, and Xing (2006), which runs from July 1963 to December 2001.<sup>10</sup> At the beginning of each month stocks are sorted into deciles on the basis of their beta estimated using one year of daily data, value-weighted portfolios are formed, and returns on these portfolios in the subsequent month are recorded. If the CAPM holds, we expect a monotonically increasing

pattern in average returns going from the low-beta to the high-beta portfolios.

A plot of the average returns on these portfolios is presented in Fig. 1, and the results of tests for an increasing relation between historical beta and subsequent returns are presented in Table 2. Although the high-beta portfolio has a larger mean return than the low-beta portfolio, the spread is not significant and generates a  $t$ -statistic of only 0.34. The MR test, however, does reject the null of no relation between past beta and expected returns in favor of a strictly increasing relation, with a  $p$ -value of 0.04.

One possible reason for the ability of our test to detect a uniform relation between beta and expected returns is that it considers all portfolio returns jointly. Another reason stems from the rising pattern in the standard deviation of beta decile portfolio returns. Low-beta portfolios have a much smaller standard deviation than high-beta portfolios. By using the studentized version of our MR test we are able to more efficiently estimate the pattern in these expected portfolio returns in a way that accounts for cross-sectional heteroskedasticity.

The Wolak and Bonferroni tests do not reject the null of a weakly increasing relation between betas and expected returns and so are consistent with the conclusion from our MR test.

##### 4.2. Portfolio sorts on CAPM beta: Post-ranked betas

As an illustration of the monotonicity tests on factor loadings, we next examine whether the post-ranked betas of portfolios ranked by their ex ante beta estimates are monotonically increasing across portfolios. Failure of this property would suggest that past beta estimates have little predictive content over future betas, perhaps due to instability, thus making them inadequate for the purpose of testing the CAPM. For this reason it is common to check monotonicity of the post-ranked betas, see, e.g., Fama and French (1992).

As above, at the beginning of each month stocks are sorted into deciles on the basis of their beta calculated using one year of daily data, value-weighted portfolios are formed, and returns on these portfolios in the subsequent month are recorded. If betas are stable over time and estimated without too much error, we would expect to see a monotonically increasing pattern in the post-ranked betas. We compute the post-ranked betas using the realized monthly returns on the decile portfolios ranked by ex ante betas. Specifically, denoting the return on the  $i$ th (ex ante sorted) decile portfolio as  $r_{it}$ , we estimate the following regressions:

$$r_{it} = \alpha_i + \beta_i r_{mt} + e_{it}, \quad i = 1, 2, \dots, 10 \quad (25)$$

and, using the theory discussed in Subsection 2.7, test the hypothesis

$$H_0 : \beta_{10} \leq \beta_9 \leq \dots \leq \beta_1$$

versus

$$H_1 : \beta_{10} > \beta_9 > \dots > \beta_1. \quad (26)$$

Panel C of Table 2 presents the results. As might be expected, the post-ranked beta estimates vary substantially

<sup>9</sup> In unreported simulation results, we impose identical pair-wise correlations across the test statistics and investigate how the results change when the correlation increased from 0 to 0.5 and 0.9. We find that the power of the MR, Wolak, and Bonferroni tests declines the higher the correlation.

<sup>10</sup> We thank the authors for providing us with these data.

**Table 2**

Test statistics for portfolios sorted by historical beta.

This table presents results of tests for a monotonic relation between estimates of Capital Asset Pricing Model (CAPM) beta and subsequent returns, using the same data as in [Ang, Chen, and Xing \(2006\)](#), which runs from July 1963 to December 2001. At the beginning of each month, stocks are sorted into deciles on the basis of their ex ante beta estimates calculated using one year of daily data, value-weighted portfolios are formed, and returns on these portfolios in the subsequent month are recorded. Panel A presents the average returns on each of these portfolios in percent per month along with the monthly standard deviation of portfolio returns. Panel B presents various tests of the monotonicity of average returns across portfolios. Column 1 reports the spread in the estimated expected return between the top and bottom ranked portfolio. Column 2 reports the *t*-statistic for this spread (using [Newey and West, 1987](#) heteroskedasticity and autocorrelation consistent standard errors), and Column 3 shows the associated *p*-value. Columns 4 and 5 present the *p*-values from the monotonic relation (MR) test applied to the decile portfolios, based either on the minimal set of portfolio comparisons or on all possible comparisons (MR<sup>all</sup>). Columns 6 and 7 show *p*-values associated with the Up and Down tests that consider signed deviations from a flat pattern. Columns 8 and 9 report the *p*-values from tests based on the [Wolak \(1989\)](#) test and a Bonferroni bound. Panel C shows post-ranked beta estimates for the decile portfolios ranked by ex ante betas. Finally, Panel D presents a *t*-test for the top-minus-bottom difference in post-ranked betas and *p*-values from the MR and MR<sup>all</sup> tests applied to the post-ranked beta estimates.

Panel A: Average returns on CAPM beta decile portfolios										
	Low	2	3	4	Past beta		7	8	9	High
Mean	0.414	0.502	0.488	0.537	0.539	0.520	0.486	0.576	0.511	0.510
Standard deviation	3.534	3.746	3.828	4.024	4.307	4.230	4.417	4.942	5.807	7.506
Panel B: Tests of monotonicity for returns on CAPM beta decile portfolios										
	Top minus	<i>t</i> -test		MR	MR <sup>all</sup>	Up	Down	Wolak	Bonferroni	
	bottom	<i>t</i> -statistic	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	
Statistic or <i>p</i> -value	0.096	0.339	0.367	0.039	0.040	0.648	0.920	0.958	1.000	
Panel C: Post-ranked betas on CAPM beta decile portfolios										
	Low	2	3	4	Past beta		7	8	9	High
Estimated beta	0.600	0.659	0.702	0.774	0.856	0.850	0.904	1.013	1.194	1.539
Panel D: Tests of monotonicity for post-ranked beta estimates										
	Top minus	<i>t</i> -test			MR	MR <sup>all</sup>				
	bottom	<i>t</i> -statistic	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value				
Statistic or <i>p</i> -value		0.938	9.486	0.000	0.003	0.003				

from a value of 0.60 for the stocks with the lowest historical beta estimates to a value of 1.54 for the stocks with the highest historical beta estimates. This difference in betas is large and significant with a *p*-value of 0.00. Moreover, the MR test in Panel D confirms that the pattern is monotonically increasing across the portfolios. The null of no relation is strongly rejected in favor of a monotonically increasing relation, with a *p*-value of 0.003.

#### 4.3. Testing monotonicity of the term premium

In a series of papers, [Fama \(1984\)](#), [McCulloch \(1987\)](#), and [Richardson, Richardson, and Smith \(1992\)](#) explore the implication of the liquidity preference hypothesis that expected returns on Treasury securities should be higher the longer their time to maturity. Eq. (3) shows that this fits in our framework.

To test this theory, [Fama \(1984\)](#) uses a Bonferroni bound based on individual *t*-tests applied to term premia on T-bills with a maturity up to 12 months. He finds evidence against monotonicity of the term premium as nine-month bills earned a higher premium than bills with longer maturity, particularly as compared with ten-month bills. [McCulloch \(1987\)](#) argues that this finding is explained by the unusual behavior of the bid-ask spread

of nine-month bills during 1964–1972. Subsequently, [Richardson, Richardson, and Smith \(1992\)](#) analyze monotonicity in the term structure using the Wolak test applied to bills with a maturity ranging from two to 11 months. For the period 1964–1990, they find that the Wolak test strongly rejects the null of a monotonically increasing pattern. However, this rejection appears to be confined to the 1964–1972 period because monotonicity is not rejected in subsamples covering the period 1973–1990.

We revisit the liquidity preference hypothesis by inspecting term premia on T-bills over the period 1964–2001, the longest available sample from the Center for Research in Security Prices monthly treasuries files. Like [Richardson, Richardson, and Smith \(1992\)](#), we restrict our analysis to maturities between two and 11 months. [Table 3](#) presents the results. Panel A shows the average term premia as a function of maturity. Over the full sample the spread in term premia between 11- and two-month bills is 0.05% per month or 0.6% per annum. Panel B shows that the associated *t*-statistic equals 2.42 and hence is statistically significant. Turning to the tests for monotonicity, both the Wolak and Bonferroni tests reject the null of an increasing term structure, while the MR test fails to find evidence in favor of a monotonically increasing term structure. These results are all consistent

**Table 3**

Test statistics for term premia.

This table presents results of tests for a monotonic relation (MR) between term premia on US Treasury bills (relative to the one-month Treasury bill) and the time to maturity, using data from the Center for Research in Security Prices monthly treasuries files, over the period January 1964–December 2001. Panel A reports the average term premia in percent per month for the full sample (1964–2001) and for two subsamples (1964–1972 and 1973–2001). Panel B reports various tests of the monotonicity of average term premia as a function of time to maturity. The first column presents the average difference between the longest and the shortest term premia, and the second and third columns present the *t*-statistic (using Newey and West, 1987 heteroskedasticity and autocorrelation consistent standard errors) and *p*-value corresponding to this difference. In the fourth and fifth columns we present the *p*-values from the MR test, based either on the minimal set of portfolio comparisons or on all possible comparisons. In the sixth and seventh columns we present our *p*-values from the tests for increasing (Up) and decreasing (Down) segments in term premia, and in the final two columns we present *p*-values from the Wolak (1989) test and a Bonferroni-based test of the null of weak monotonicity against an unconstrained alternative.

Panel A: Average term premia										
Sample	Maturity (in months)									
	2	3	4	5	6	7	8	9	10	11
1964–2001	0.027	0.049	0.050	0.064	0.068	0.063	0.080	0.086	0.071	0.077
1964–1972	0.023	0.040	0.038	0.052	0.054	0.052	0.069	0.069	0.018	0.050
1973–2001	0.028	0.052	0.053	0.068	0.072	0.066	0.084	0.092	0.087	0.085
Panel B: Tests of monotonicity of term premia										
	Top minus	<i>t</i> -test		MR	MR <sup>all</sup>	Up	Down	Wolak	Bonferroni	
Sample	bottom	<i>t</i> -statistic	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	<i>p</i> -value	
1964–2001	0.050	2.416	0.008	0.953	0.906	0.000	0.369	0.036	0.020	
1964–1972	0.026	0.908	0.182	0.983	0.991	0.003	0.375	0.007	0.004	
1973–2001	0.057	2.246	0.012	0.633	0.617	0.002	0.474	0.340	0.704	

with the presence of some declining segments in the term structure. Fig. 2 shows that, consistent with the earlier studies, the culprit appears to be the high term premium on nine-month bills.

When conducted separately on the subsample 1964–1972, very similar conclusions emerge. In this subsample, a monotonically rising term structure is clearly rejected by both the Wolak and Bonferroni tests and the MR test also finds no evidence to support uniformly increasing term premia. In sharp contrast, for the period 1973–2001, the Bonferroni and Wolak tests both fail to reject the null of an increasing term premium. The inability of the MR test to reject the null against an increasing term premium could simply reflect low power of this test in this sample. In support of this interpretation, notice that the Up test finds significant evidence of segments with a strictly increasing term premium, while the Down test fails to find significant evidence of decreasing segments.

#### 4.4. One-way portfolio sorts: Further evidence

It is common practice to inspect mean return patterns for portfolios sorted on firm or security characteristics. To illustrate how our approach can be used in this context, we therefore consider returns on a range of portfolios sorted on firm characteristics such as market equity (size), book-to-market ratio, cashflow-price ratio, earnings-price ratio, and the dividend yield or past returns over the previous month (short-term reversal), 12 months (momentum), or 60 months (long-term reversal). For some of these sorts, the implications of theory are

not as clear-cut as in the case of the CAPM and so should be viewed as joint tests that the sorting variable proxies for exposure to some unobserved risk factor and the validity of the underlying asset pricing model.

Data on value-weighted portfolio returns are obtained from Ken French's website at Dartmouth College<sup>11</sup> and are made up of stocks listed on NYSE, AMEX, and NASDAQ. The findings on cross-sectional return patterns in portfolios sorted on various firm characteristics reported by Fama and French (1992) were based on data starting in July 1963, and we keep this date as our starting point. However, we also consider the earliest starting point for each series, which is 1926 or 1927, except for the portfolios sorted on long-term reversal that begin in 1931 and the portfolios sorted on the earnings-price or cashflow-price ratios that begin in 1951. In all cases the data end in December 2006. We focus our discussion on the more recent sample from 1963 to 2006.

Panel A of Table 4 reports estimates of expected returns for the decile portfolios sorted on the eight variables listed in the columns. We preserve the order of the portfolios reported by Ken French. This means that we are interested in testing for an increasing relation between portfolio rank and expected returns for the portfolios sorted on book-to-market, cashflow-price, earnings-price, dividend yield, and momentum. We also are interested in testing for a decreasing relation for the portfolios sorted on size, short-term reversal, and long-term reversal.

<sup>11</sup> [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)

**Table 4**

Estimates of expected returns for decile portfolios.

This table reports mean returns (in percent per month) for stocks sorted into value-weighted decile portfolios. The sorting variables are market equity (ME), book-to-market value (BE-ME), cashflow-price (CF-P), earnings-price (E-P), dividend-price (D-P), momentum (M'tum), short-term reversal (ST Rev), and long-term reversal (LT Rev). All data series are taken from Ken French's website at Dartmouth College. Panels A and B report results for the period July 1963–December 2006, and Panels C and D report results for the earliest available starting point for each series, which is 1926 or 1927 except for the portfolios sorted on long-term reversal, which begin in 1931, and the portfolios sorted on the earnings-price or cashflow-price ratios, which begin in 1951. Panels B and D report various tests for monotonicity in the average returns on these portfolios described in Section 2. Row 1 reports the *t*-statistic for this spread (using Newey and West, 1987 heteroskedasticity and autocorrelation consistent standard errors), and Row 2 shows the associated *p*-value. Rows 3 and 4 present the *p*-values from the monotonic relation (MR) test between the portfolio sorting variables and expected returns across all decile portfolios, based either on the minimal set of portfolio comparisons or on all possible comparisons (MR<sup>all</sup>). Rows 5 and 6 report the *p*-values from the bootstrap tests for Up and Down changes in expected returns across adjacent portfolios. The second-last and last rows report the *p*-values from the Wolak (1989) test and a test based on Bonferroni bounds.

Portfolio	ME	BE-ME	CF-P	E-P	D-P	M'tum	ST Rev	LT Rev
Panel A: Average returns, 1963–2006								
Low	1.273	0.824	0.850	0.829	1.002	0.177	1.147	1.395
2	1.206	0.948	0.898	0.845	0.938	0.736	1.281	1.244
3	1.241	0.989	0.975	0.975	1.029	0.862	1.255	1.214
4	1.185	1.013	0.960	0.959	1.006	0.903	1.055	1.104
5	1.209	1.014	1.069	0.943	0.912	0.801	1.020	1.127
6	1.097	1.110	1.030	1.075	1.006	0.899	0.935	1.074
7	1.151	1.188	1.090	1.234	1.045	0.941	0.890	1.070
8	1.095	1.216	1.127	1.229	1.133	1.145	0.941	0.977
9	1.026	1.269	1.329	1.284	1.122	1.238	0.748	0.897
High	0.886	1.396	1.334	1.429	1.074	1.648	0.683	0.882
High–Low	–0.387	0.572	0.485	0.600	0.072	1.472	–0.464	–0.512
Panel B: Tests of monotonicity, 1963–2006								
<i>t</i> -statistic	–1.536	2.544	2.404	2.683	0.295	5.671	–2.364	–2.205
<i>t</i> -test <i>p</i> -value	0.062	0.005	0.008	0.004	0.384	0.000	0.009	0.014
MR <i>p</i> -value	0.274	0.000	0.024	0.008	0.336	0.291	0.258	0.002
MR <sup>all</sup> <i>p</i> -value	0.237	0.000	0.012	0.021	0.256	0.242	0.170	0.002
Up <i>p</i> -value	0.737	0.045	0.035	0.016	0.353	0.000	0.889	0.999
Down <i>p</i> -value	0.051	1.000	0.994	0.985	0.651	0.954	0.015	0.114
Wolak <i>p</i> -value	0.736	1.000	0.990	0.991	0.810	0.873	0.860	0.998
Bonferroni <i>p</i> -value	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Panel C: Average returns, full sample								
Low	1.520	0.872	0.863	0.852	0.924	0.338	1.493	1.504
2	1.326	0.973	0.939	0.864	0.984	0.732	1.217	1.315
3	1.298	0.977	0.986	1.010	0.934	0.741	1.168	1.286
4	1.249	0.970	1.003	1.010	1.024	0.865	1.039	1.094
5	1.211	1.054	1.116	1.020	0.902	0.869	1.064	1.164
6	1.183	1.101	1.086	1.190	0.991	0.943	1.017	1.044
7	1.148	1.115	1.189	1.265	1.086	1.032	0.972	1.072
8	1.093	1.272	1.212	1.344	1.145	1.163	0.932	1.047
9	1.040	1.306	1.397	1.403	1.109	1.268	0.828	0.922
High	0.907	1.410	1.437	1.549	1.094	1.591	0.487	0.902
High–Low	–0.613	0.538	0.574	0.697	0.171	1.254	–1.007	–0.602
Panel D: Tests of monotonicity, full sample								
<i>t</i> -statistic	–2.350	2.439	3.244	3.665	0.938	5.434	–5.060	–2.486
<i>t</i> -test <i>p</i> -value	0.009	0.007	0.001	0.000	0.174	0.000	0.000	0.006
MR <i>p</i> -value	0.002	0.000	0.018	0.000	0.692	0.002	0.012	0.221
MR <sup>all</sup> <i>p</i> -value	0.002	0.000	0.012	0.000	0.492	0.002	0.016	0.158
Up <i>p</i> -value	0.987	0.027	0.007	0.001	0.169	0.000	0.993	0.919
Down <i>p</i> -value	0.024	1.000	0.994	1.000	0.606	0.998	0.003	0.020
Wolak <i>p</i> -value	0.985	0.999	0.995	1.000	0.530	0.999	0.995	0.880
Bonferroni <i>p</i> -value	1.000	1.000	1.000	1.000	0.361	1.000	1.000	1.000

For the portfolios sorted on the book-to-market, cashflow-price or earnings-price ratios, or long-term reversal, there are either no reversals of the monotonic pattern or few and smaller ones compared with those observed for the other sorts. There are larger reversals in average returns for the size-sorted portfolios (three reversals of up to five basis points per month) as

well as for the portfolios sorted on dividend yield (five reversals of up to nine basis points), momentum (a 10 basis point decrease), and short-term reversal (an increase of 14 basis points). Hence, for some portfolio sorts a monotonic pattern is observed in average returns, while for other portfolio sorts non-monotonic patterns of varying degrees arise. Moreover, large differences emerge

in the magnitude of the top–bottom spreads, which range from a minimum of seven basis points per month for the portfolios sorted on the dividend yield to nearly 150 basis points for the portfolios sorted on momentum. A key question is clearly how strong the deviations from a monotonic pattern must be for us to reject the null hypothesis and establish monotonicity in expected returns.

To answer this question, Panel B presents test results for the eight portfolio sorts. The first row reports the  $t$ -statistic for testing the significance of the difference in expected returns between the top and bottom portfolios. These range from 0.30 (for portfolios sorted on the dividend yield) to 5.7 in the case of the momentum-sorted portfolios. The associated  $p$ -values show that the portfolios sorted on the dividend yield fail to produce a statistically significant top–bottom spread. Moreover, the spread in average returns on the size-sorted portfolios is borderline insignificant with a  $p$ -value of 0.06. The remaining portfolios generate significant spreads.

To see if the mean return patterns are consistent with monotonicity in expected returns, the third row in Panel B reports the bootstrapped  $p$ -values associated with the MR test. This test fails to find a monotonic relation between expected returns and portfolios ranked by short-term reversal ( $p$ -value of 0.26), momentum ( $p$ -value of 0.29), size ( $p$ -value of 0.27), or the dividend yield ( $p$ -value of 0.34). Only for the portfolios sorted on long-term reversal and the book-to-market, cash flow-price, and earnings-price ratios do we continue to find strong evidence of a monotonic pattern in expected returns. The fourth row in Panel B shows that the MR tests based on comparing only the adjacent portfolios versus comparing all possible pairs always lead to the same conclusions.

The MR test accounts for the effects of random sampling variation. This has important implications. For example, for the portfolios sorted on the earnings-price ratio where three reversals appear in the ordering of average returns, the test still rejects very strongly because these reversals are small in magnitude (less than two basis point per decile portfolio) relative to the sampling variability of the average returns. In contrast, the relation between expected returns and the dividend yield or momentum are insignificant. This is to be expected given the large reversals in the mean return patterns observed for the portfolios sorted on this variable.

For all eight portfolio sorts, the Wolak (1989) test fails to reject the null of a (weakly) monotonic relation. Moreover, this conclusion is supported by the multivariate inequality test based on the Bonferroni bound, which always equals one. Even the conventional  $t$ -test found no evidence of a monotonic pattern for the portfolios sorted on the dividend yield and so this evidence illustrates the difficulty that could arise in interpreting the Wolak and Bonferroni test. Failure to find evidence against a weakly monotonic pattern in the sorted portfolio returns could simply reflect weak power.

Similar results are obtained in the longer samples listed in Panels C and D, although the MR test now also finds evidence in support of monotonicity for the portfolios sorted on size, momentum, and short-term reversal but fails to reject the null for the portfolios sorted

on long-term reversal. As before, in each case the Bonferroni and Wolak tests fail to reject the null.

#### 4.5. Two-way portfolio sorts

As an illustration of how our methodology can be extended, we next consider two-way sorts that combine portfolios sorted on firm size with portfolios sorted on either the book-to-market ratio or momentum. In both cases we study  $5 \times 5$  portfolio sorts. Results, reported in Table 5, are for the period from July 1963 to December 2006. Inspecting the mean returns, there is some evidence of a non-monotonic pattern across size-sorted portfolios for the quintiles with a low book-to-market ratio tracking growth stocks (Panel A) or stocks with poor past performance (Panel B).

Because the two-way sorts involve a large number of inequalities, it is useful to decompose the overall (joint) test into a series of conditional tests that help identify the economic source of the results. Table 5 therefore uses the MR test to examine patterns in expected returns keeping one sorting variable (e.g., size) constant while varying another (e.g., the book-to-market ratio) or vice versa. For example, the penultimate row in each panel presents the  $p$ -values from the tests for a monotonic relation between the portfolio sorting variable in the row (size) and expected returns, conditional on keeping the column portfolio fixed. Hence, the  $p$ -value of a test for the size effect, conditional on being in the top book-to-market quintile (value stocks), is 0.031, while it is 0.687 for the bottom book-to-market portfolio (growth stocks). The final row in each panel presents the  $p$ -value from a joint test for a monotonic relation between the sorting variable in the row (size), computed across all column portfolios. Similarly, the penultimate column presents results from tests for a monotonic relation between the portfolio sorting variable in the columns and expected returns, conditional on the row portfolio, representing firm size.

The results in Table 5 shed new light on the earlier findings. For example, Panel A reveals that the value effect is strong among all size portfolios (with  $p$ -values ranging from 0.004 to 0.057). Hence, the statistical evidence appears to support the conclusion in Fama and French (2006) that there is a value effect even among large stocks, although the spread in the top-minus-bottom portfolios' average returns is much wider for the smallest stocks than for the largest stocks. Conversely, the size effect is significant only among value firms. For the other book-to-market sorted portfolios, the size effect is non-monotonic. Overall, the joint test fails to find evidence in support of a size and book-to-market effect.

Both the Wolak and Bonferroni tests failed to reject the null of monotonic patterns for both the size and book-to-market portfolios, with  $p$ -values close to one. Again this highlights the different conclusions that can emerge depending on whether monotonicity is entertained under the null or alternative hypothesis.

The results for the size and momentum two-way sorts provide strong support for a momentum effect among the four quintiles with the smallest stocks but fail to find a momentum effect for the largest stocks, for which the MR



**Table 5**

Conditional and joint monotonicity tests for double-sorted portfolios.

This table shows mean returns for stock portfolios using  $5 \times 5$  two-way sorts. The sorting variables are market equity, which is always listed in the row, and one of either book-to-market value or momentum. The sample period is July 1963–December 2006, and all data series are taken from Ken French's website at Dartmouth College. Portfolios are value-weighted and mean returns are reported in percent per month. The penultimate row in each panel presents the  $p$ -values from tests for a monotonic relation (MR) between the portfolio sorting variable in the row and expected returns, conditional on being in a given column portfolio. For example, the  $p$ -value on a test for the size effect, conditional on being in the Growth book-to-market portfolio is 0.687, while it is 0.031 if we condition on being in the Value book-to-market portfolio. The final row in each panel presents the  $p$ -value from a joint test for a monotonic relation between the portfolio sorting variable in the row, computed across all column portfolios. The penultimate and final columns present  $p$ -values from tests for a monotonic relation between the portfolio sorting variable in the column and expected returns, conditional on being in a given row portfolio (size). The bottom-right number in each panel is the  $p$ -value for the joint test for a monotonic relation in both variables.

Panel A: Market equity $\times$ book-to-market ratio						MR	Joint MR
Market equity	Book-to-market ratio					$p$ -value	$p$ -value
	Growth	2	3	4	Value		
Small	0.711	1.297	1.337	1.546	1.660	0.023	0.000
2	0.878	1.141	1.411	1.458	1.524	0.004	
3	0.889	1.205	1.210	1.334	1.506	0.057	
4	0.998	0.994	1.222	1.334	1.374	0.044	
Big	0.879	0.968	0.982	1.066	1.074	0.023	
MR $p$ -value	0.687	0.401	0.405	0.069	0.031		0.083
Joint MR $p$ -value			0.342				
Panel B: Market value $\times$ momentum						MR	Joint MR
Market equity	Momentum					$p$ -value	$p$ -value
	Losers	2	3	4	Winners		
Small	0.362	1.154	1.417	1.564	1.973	0.000	0.154
2	0.423	1.034	1.257	1.499	1.777	0.000	
3	0.601	0.979	1.123	1.228	1.728	0.000	
4	0.597	0.992	1.026	1.238	1.583	0.008	
Big	0.645	0.883	0.774	0.975	1.272	0.552	
MR $p$ -value	0.893	0.143	0.001	0.117	0.016		0.545
Joint MR $p$ -value			0.708				

test records a  $p$ -value of 0.552. Similarly, there is evidence of a size effect only for the three portfolios with the strongest past performance, but not among the two loser portfolios. Hence, it is not surprising that the tests for separate size and momentum effects both fail to reject, as does the overall joint test ( $p$ -value of 0.545).

Consistent with the results from the MR test, the Wolak test rejected the null that expected returns follow a monotonic pattern in both size and momentum. In contrast, the more conservative Bonferroni test failed to reject the null hypothesis of weak monotonicity.

To summarize, two-way sorts can be used to diagnose why empirical evidence could fail to support a hypothesized pattern in expected returns. Here our findings suggest that the size effect in expected returns is absent from growth firms and among loser stocks. Moreover, momentum effects seem strong for small and medium-size firms but not among the largest quintile of stocks.

## 5. Conclusion

Empirical research in finance often seeks to address whether a systematic relation exists between an asset's

expected return and some measure of the asset's risk or liquidity characteristics. In this paper we propose a test that reveals whether a null hypothesis of no systematic relation can be rejected in favor of a monotonic relation predicted by economic theory. The test summarizes in a single number whether the relation is monotonic or not. Moreover, it is nonparametric and does not require making any assumptions about the functional form of the relation between the variables used to sort securities and the corresponding expected returns. This is a big advantage because monotonicity in expected returns on securities sorted by some variable is preserved under very general conditions, including nonlinear mappings between sorting variables and risk factor loadings. Perhaps most important, our test is extremely easy to use.

We see two principal uses for the new test. First, it can be adopted as a descriptive statistic for monotonicity in the expected returns of individual securities or portfolios of securities ranked according to one or more sorting variables. Besides providing a single summary statistic for monotonicity, our test allows researchers to decompose empirical results to better diagnose the source of a rejection of (or failure to reject) the theory being tested. In general, it is good practice to consider the complete

cross-sectional pattern in expected returns on securities sorted by liquidity or risk characteristics and our test makes it easy to do this.

Second, if theoretical considerations suggest a monotonic relation between the sorting variable and expected returns, then the approach can be used to formally test asset pricing implications such as in the liquidity preference and CAPM examples covered here or in tests of whether the pricing kernel decreases monotonically in market returns. Moreover, when a model implies a particular ranking in the loadings of individual stocks on observed risk or liquidity factors, the monotonicity test can be based on the estimated asset betas. Lack of monotonicity in such cases could imply that the conjectured theoretical model is not an adequate description of the data.

## Appendix

The theorem below applies not just to sample means or differences in sample means, but also to slope coefficients, and so we use the general notation  $\beta$  as the coefficient of interest. The case of sample means is a special case of the result below, setting  $\mathbf{F}_{it} = 1$  for all  $i, t$ , i.e., regressing each of the asset returns simply on a constant. In the theorem we also consider the slightly more general case, relative to the discussion in Section 2.7, that the regressors in each equation can differ, so we index the regressors by both  $i$  and  $t$  instead of just  $t$ . This generalization could be of use in cases in which different control variables are needed for different assets, for example.

**Theorem 1.** Consider a set of regressions, with potentially different regressors in each equation:

$$r_{it} = \beta_i' \mathbf{F}_{it} + e_{it}, \quad i = 0, 1, \dots, N; \quad t = 1, 2, \dots, T,$$

and define

$$\mathbf{h}_t \equiv [\text{vech}(\mathbf{F}_{0t} \mathbf{F}_{0t}'), \dots, \text{vech}(\mathbf{F}_{Nt} \mathbf{F}_{Nt}'), r_{0t} \mathbf{F}_{0t}', \dots, r_{Nt} \mathbf{F}_{Nt}']',$$

where  $\text{vech}$  is the half-vec operator (see, e.g., Hamilton, 1994). Assume that (i)  $\mathbf{h}_t$  is a strictly stationary process, (ii)  $E[|h_{kt}|^{6+\varepsilon}] < \infty$  for some  $\varepsilon > 0$  for all  $k$ , where  $h_{kt}$  is the  $k$ th element of  $\mathbf{h}_t$ , (iii)  $\{\mathbf{h}_t\}$  is  $\alpha$ -mixing of size  $-3(6+\varepsilon)/\varepsilon$ , and (iv)  $E[\mathbf{F}_{it} \mathbf{F}_{it}']$  is invertible for all  $i=0, 1, \dots, N$ . Let  $\hat{\beta}_i$  denote the usual ordinary least squares (OLS) estimator of  $\beta_i$ . Let  $\theta_j \equiv [\beta_{j0}, \dots, \beta_{jN}]$  and  $\hat{\theta}_j \equiv [\hat{\beta}_{j0}, \dots, \hat{\beta}_{jN}]$  be the vector of the  $j$ th regression coefficient in each of the  $N+1$  regressions. Then as  $T \rightarrow \infty$

$$\sqrt{T}(\hat{\theta}_j - \theta_j) \Rightarrow \mathcal{N}(0, \Omega_j),$$

and

$$\begin{aligned} \min_{i=1, \dots, N} \sqrt{T} \{(\hat{\beta}_{j,i} - \hat{\beta}_{j,i-1}) - (\beta_{j,i} - \beta_{j,i-1})\} \\ \Rightarrow W_j \equiv \min_{i=1, \dots, N} \{Z_i - Z_{i-1}\}, \end{aligned}$$

where  $[Z_0, \dots, Z_N]' \sim \mathcal{N}(0, \Omega_j)$  and  $\Rightarrow$  denotes convergence in distribution. Further, if  $a_T$  is the length of the average block in the stationary bootstrap,  $a_T \rightarrow \infty$  and  $a_T/T \rightarrow 0$

as  $T \rightarrow \infty$ , then

$$\sup_z |P^*[\|\hat{\theta}_j - \theta_j\| \leq z] - P[\|\hat{\theta}_j - \theta_j\| \leq z]| \xrightarrow{P} 0, \quad \text{as } T \rightarrow \infty,$$

and

$$\begin{aligned} \sup_z \left| P^* \left[ \min_{i=1, \dots, N} \sqrt{T} \{(\hat{\beta}_{j,i}^* - \hat{\beta}_{j,i-1}^*) - (\hat{\beta}_{j,i} - \hat{\beta}_{j,i-1})\} \leq z \right] \right. \\ \left. - P \left[ \min_{i=1, \dots, N} \sqrt{T} \{(\hat{\beta}_{j,i} - \hat{\beta}_{j,i-1}) - (\beta_{j,i} - \beta_{j,i-1})\} \leq z \right] \right| \xrightarrow{P} 0, \quad \text{as } T \rightarrow \infty, \end{aligned}$$

where  $\|\cdot\|$  represents a norm on  $\mathbb{R}^{N+1}$ ,  $\xrightarrow{P}$  denotes convergence in probability, and  $P^*$  is the probability measure induced by the bootstrap conditional on the original data, so  $\hat{\beta}_{j,i}^*$  refers to the bootstrapped OLS estimate of the  $j$ th variable and the  $i$ th asset.

**Proof of Theorem 1.** The first part of Theorem 1, the asymptotic normality of the  $(N+1) \times 1$  vector of coefficients in each regression, follows from Theorem 4 in Politis and Romano (1994) under the stated conditions. As shown in Proposition 2.2 in White (2000), because the minimum of a vector of (differences in) parameters is a continuous function of the elements of the vector, by the continuous mapping theorem we have that

$$\begin{aligned} \min_{i=1, \dots, N} \sqrt{T} \{(\hat{\beta}_{j,i} - \hat{\beta}_{j,i-1}) - (\beta_{j,i} - \beta_{j,i-1})\} \\ \Rightarrow W_j \equiv \min_{i=1, \dots, N} \{Z_i - Z_{i-1}\}, \end{aligned} \quad (27)$$

where  $[Z_0, \dots, Z_N]' \sim \mathcal{N}(0, \Omega_j)$ . The final part of Theorem 1, which justifies use of the bootstrap, follows from Corollary 2.6 in White (2000), noting that we do not need any further assumptions than stated in the theorem due to the fact that there are no estimated parameters here. The Up and Down tests can be justified using the same reasoning as for the directional accuracy test in Section 4 of White (2000).  $\square$

## References

- Ang, A., Chen, J., 2007. The CAPM over the long run: 1926–2001. *Journal of Empirical Finance* 14, 1–40.
- Ang, A., Chen, J., Xing, Y., 2006. Downside risk. *Review of Financial Studies* 19, 1191–1239.
- Bakshi, G., Madan, D., Panayotov, G., 2010. Returns of claims on the upside and the viability of U-shaped pricing kernels. *Journal of Financial Economics* 97, 130–154.
- Bartholomew, D.J., 1961. A test of homogeneity of means under restricted alternatives. *Journal of the Royal Statistical Society B* 23, 239–281.
- Berk, J.B., Green, R.C., Naik, V., 1999. Optimal investment, growth options, and security returns. *Journal of Finance* 54, 1553–1607.
- Boudoukh, J., Richardson, M., Smith, T., Whitelaw, R.F., 1999. Ex ante bond returns and the liquidity preference hypothesis. *Journal of Finance* 54, 1153–1167.
- Campbell, J.Y., Lo, A.W., MacKinlay, A.C., 1997. *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ.
- Carlson, M., Fisher, A., Giammarino, R., 2004. Corporate investment and asset price dynamics: implications for the cross-section of returns. *Journal of Finance* 59, 2577–2603.
- Casella, G., Berger, R.L., 1990. *Statistical Inference*. Duxbury Press, Pacific Grove, CA.
- Christoffersen, S.E.K., 2001. Why do money fund managers voluntarily waive their fees? *Journal of Finance* 56, 1117–1140.
- Fama, E.F., 1984. Term premiums in bond returns. *Journal of Financial Economics* 13, 529–546.

- Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. *Journal of Finance* 47, 427–465.
- Fama, E.F., French, K.R., 2006. The value premium and the CAPM. *Journal of Finance* 61, 2163–2186.
- Gourieroux, C., Holly, A., Monfort, A., 1982. Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica* 50, 63–80.
- Hamilton, J.D., 1994. *Time Series Analysis*. Princeton University Press, Princeton, NJ.
- Hansen, P.R., 2005. A test for superior predictive ability. *Journal of Business and Economic Statistics* 23, 365–380.
- Jackwerth, J.C., 2000. Recovering risk aversion from option prices and realized returns. *Review of Financial Studies* 13, 433–451.
- Johnson, T.C., 2002. Rational momentum effects. *Journal of Finance* 57, 585–608.
- Kodde, D.A., Palm, F.C., 1986. Wald criteria for jointly testing equality and inequality restrictions. *Econometrica* 54, 1243–1248.
- Kosowski, R., Timmermann, A., Wermers, R., White, H., 2006. Can mutual fund “stars” really pick stocks? New evidence from a bootstrap analysis. *Journal of Finance* 61, 2551–2595.
- Kudo, A., 1963. A multivariate analogue of the one-sided test. *Biometrika* 50, 403–418.
- McCulloch, J.H., 1987. The monotonicity of the term premium: a closer look. *Journal of Financial Economics* 18, 185–192.
- Newey, W., West, K.D., 1987. A simple, positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Perlman, M.D., 1969. One-sided testing problems in multivariate analyses. *Annals of Mathematical Statistics* 40, 549–567.
- Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. *Journal of the American Statistical Association* 89, 1303–1313.
- Richardson, M., Richardson, P., Smith, T., 1992. The monotonicity of the term premium: another look. *Journal of Financial Economics* 31, 97–106.
- Romano, J.P., Wolf, M., 2005. Stepwise multiple testing as formalized data snooping. *Econometrica* 73, 1237–1282.
- Rosenberg, J.V., Engle, R.F., 2002. Empirical pricing kernels. *Journal of Financial Economics* 64, 341–372.
- Shive, S., Shumway, T., 2009. Is the Pricing Kernel Monotonic? Manuscript. University of Notre Dame, Notre Dame, IN.
- Sullivan, R., Timmermann, A., White, H., 1999. Data-snooping, technical trading rules, and the bootstrap. *Journal of Finance* 54, 1647–1692.
- White, H., 2000. A reality check for data snooping. *Econometrica* 68, 1097–1126.
- Wolak, F.A., 1987. An exact test for multiple inequality and equality constraints in the linear regression model. *Journal of the American Statistical Association* 82, 782–793.
- Wolak, F.A., 1989. Testing inequality constraints in linear econometric models. *Journal of Econometrics* 31, 205–235.