

Portfolio Sorts and Tests of Cross-Sectional Patterns in Expected Returns*

Andrew J. Patton
University of Oxford

Allan Timmermann
University of California San Diego

12 June 2008

Abstract

Portfolio sorts are used extensively in finance to explore the relation between firm characteristics and expected returns. Existing empirical tests have largely been limited, however, to comparing expected returns of the top and bottom ranked portfolios. We propose a new and general approach to test for a monotonic relationship between expected returns and variables used to sort stocks and form portfolios, such as firm characteristics or past returns. Empirical evidence suggests that inference from standard t -tests can be overturned by our monotonicity tests. In particular, the short-term reversal and momentum effects and the declining relation between firm size and average returns are shown not to hold monotonically in post-1963 data. Conversely, the evidence in favor of other patterns, such as the value effect, is strengthened by our tests.

*We thank Susan Christoffersen, Erik Kole, Robert Kosowski, Jun Liu, Igor Makarov, 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 Society for Financial Econometrics Inaugural Conference at NYU for helpful comments and suggestions. Matlab code for the methods presented in this paper is available from <http://www.economics.ox.ac.uk/members/andrew.patton/code.html>. Patton: Department of Economics and Oxford-Man Institute of Quantitative Finance, University of Oxford, Manor Road, Oxford OX1 3UQ, United Kingdom. Email: andrew.patton@economics.ox.ac.uk. Timmermann: Rady School of Management and Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0553, USA. Email: atimmerm@ucsd.edu.

Portfolio sorts are used extensively throughout finance to establish and test for systematic cross-sectional patterns in expected stock returns related to firm or stock characteristics such as book-to-market or earnings-price ratios (Basu (1977, 1983) and Fama and French (1992, 2006)), firm size (Banz (1981) and Reinganum (1981)), co-skew with the market portfolio (Harvey and Siddique (2000)), financial constraints (Lamont, Polk and Saa-Requejo (2001)), liquidity (Pastor and Stambaugh (2003)), default risk (Vassalou and Xing (2004)), volatility (Ang, Hodrick, Xing and Zhang (2005)) and ‘downside’ risk (Ang, Chen and Xing (2006)). Such sorts are also frequently used to explore time-series dependencies in stock returns related to momentum or reversals (Jegadeesh and Titman (1993)) or to test for skill and performance persistence among mutual funds, (Carhart (1997)). Sorts on a single characteristic such as firm size are very common, but double sorts on multiple characteristics such as momentum and size (Rouwenhorst (1998)), momentum and trading volume (Lee and Swaminathan (2000)), financial constraints and R&D expenditures (Li (2007)) or payout policy and leverage (Nielsen (2007))—are also in widespread use, and triple sorts are not uncommon (Daniel, Grinblatt, Titman and Wermers (1997) and Vassalou and Xing (2004)).

While portfolio sorts are very common in empirical finance, the full set of implications of such sorts is generally not explored. Standard practice is to sort stocks (or funds) into multiple portfolios, often ten or more, at some formation date and study the patterns emerging in the average returns over the subsequent holding period going from the low end to the top end of the portfolios ranked on the variable(s) of interest. Formal tests, however, only consider the mean return differential between the top and bottom portfolios. 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 the sorting variable. For example, suppose that ‘small’, ‘medium’ and ‘large’ stocks have expected returns of 1%, 1.2% and 0.8% per month. Then a comparison of small and large stocks would lead to the conclusion that small firms earn higher mean returns than large ones. If taken to indicate a negative relationship between size and expected returns, this clearly ignores that medium-sized firms earn higher expected returns than small firms in this example. Only a test that simultaneously considers the expected returns on all three portfolios can identify the non-monotonic relationship between firm size and expected returns.

Conversely, in cases where a genuinely monotonic relationship emerges between expected returns and portfolio ranks across several portfolios, a test that only considers the difference between the top and bottom ranked portfolios does not utilize the full strength of the empirical evidence. If interest

is limited to establishing a trading strategy and it is possible to short the lowest ranked stocks, then a comparison between the expected returns on the top and bottom portfolios—as is done by the standard t -test—may suffice. However, when considering the implications of theories that rank stocks based on variables proxying for risk or liquidity, the complete pattern in expected returns across all portfolios is relevant. Furthermore, a formal test can determine whether deviations from a strict monotonic pattern constitute significant evidence against monotonicity or not.

This paper proposes new ways to test for monotonic relations between expected returns and portfolios ranked by some stock characteristic. Our tests are nonparametric, powerful, 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. This is important because for many economic models the underlying hypothesis is only that there is a monotonic relationship between, on the one hand, one or more firm characteristics that proxy for risk exposures and, on the other hand, expected returns. For example, recent studies have developed single-factor equilibrium models where expected returns are determined by firms' betas which depend monotonically on size and/or value as a result of differences in the risk characteristics of assets in place versus those of growth options (Berk, Green and Naik (1999) and Gomes, Kogan and Zhang (2003)), costly reversibility in investments (Zhang (2005)) or non-separability in the utility derived from nondurable and durable consumption (Yogo (2006)). Behavioral models and models where market prices deviate from fundamental values by a noise component can lead to similar monotonic relations between expected returns and firms' price-earnings ratio or size (Arnott et al. (2007).)

In common with the conventional top-minus-bottom t -test, our monotonic relation (MR) test holds that expected returns are identical under the null, while under the alternative we maintain a monotonically decreasing relation. Thus a rejection of the null of no relationship in favor of the hypothesized relationship (i.e. a finding of 'statistical significance) represents a strong empirical endorsement of the theory. The converse approach of maintaining a monotonically decreasing relation under the null versus no such relationship under the alternative has been developed by Wolak (1987, 1989a,b). Which approach is best depends on the research question. Richardson et al. (1992) used the Wolak test to see if there was evidence against an upwardsloping term structure of interest rates. One important empirical drawback, however, of entertaining the hypothesized monotonic relationship under the null is that a confirmation of a theory from a failure to reject the null may simply be due to limited power for the test. This turns out to be empirically important

since the Wolak test fails to reject the null for all of the portfolio sorts we consider in our empirical application, while the t -test and the MR test are able to differentiate between variables that find support in the data and those that do not.

Depending on the research question and the economic framework, one may prefer to entertain the presence of a monotonic relationship under the null or under the alternative. One important empirical drawback, however, of entertaining the hypothesized monotonic relationship under the null is that limited power (due to a short time series of data, or due to noisy data) will make 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.

To gain intuition for how our tests work, consider Figure 1 which plots average returns on value-weighted portfolios sorted by cash flow-to-price (upper window) or past returns (short-term reversal shown in the lower window), in both cases calculated over the period 1963-2006. For the cash flow-sorted stocks both the t -test comparing the returns on the top and bottom portfolios and our test for a monotonic relation (MR) across all ten portfolios reject the null of equal expected returns against an alternative of monotonically increasing expected returns. This holds despite the presence of two segments with declining mean returns (between portfolios 3 and 4 and portfolios 5 and 6) which our ‘MR’ test, in this case, ascribes to random sampling variation. In contrast, the two tests come up with opposite conclusions for the stocks sorted on short-term reversal. Here the t -test shows that the mean return difference between winner and loser stocks is highly significant. However, our test shows that the relationship between expected returns and past short-term performance is not monotonically declining, primarily due to the large increase in returns when moving from the loser portfolio to the second portfolio. Our test suggests that this increase is too large (14 basis points per month) to be explained by random sampling variation. In both cases the Wolak test fails to reject. A comparison of these two cases clearly illustrates the need to have a formal test that can tell when non-monotonic patterns in portfolio sorts, such as those observed in Figure 1, can be explained by sampling variation and when they can not.

We show how to generalize our tests when expected returns are hypothesized to be monotonically related (increasing or decreasing) to two variables such as firm size and book-to-market value. When applied to two-way sorts, we demonstrate that our new test for monotonic relations between expected returns and the row and column sorting variables gives rise to a surprisingly large number of inequality relations that are part of the test. For example, monotonicity in a 5×5 two-way

sort implies 40 adjacent inequality relations while a 10×10 two-way sort implies 180 adjacent inequalities. This gives intuition for why our test is far more informative than existing pair-wise t -tests conducted on the spread in the returns of top and bottom portfolios.

Empirically, it can make a big difference whether one uses our new test or conventional top-minus-bottom portfolio t -tests. 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 portfolios is highly significant, but the pattern in average returns across multiple portfolios is non-monotonic. This holds, for example, in post 1963-data for decile portfolios sorted on short-term reversal, momentum or firm size. These results suggest revisiting the empirical evidence for many portfolio sorts used in the finance literature.

Our approach is not restricted to testing for monotonic patterns in the expected returns on portfolios 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. It can also be adopted to test for piece-wise monotonic patterns, as in the case of the U-shaped relationship between fee waivers and mutual fund performance reported by Christoffersen (2001). Finally, using methods for converting conditional moments into unconditional ones along the lines of Boudoukh et al. (1999), we show that the approach can be used to conduct conditional tests of monotonicity.

The outline of the paper is as follows. Section 1 describes our new approach to testing for monotonic patterns in expected returns on portfolios sorted on one or more variables. Section 2 uses the proposed methodology to analyze a range of portfolio sorts from the empirical finance literature. Section 3 uses Monte Carlo simulations to shed light on the behavior of the new tests as well as that of standard tests under a set of controlled experiments and provides some extensions. Finally, Section 4 concludes.

1 Methodology

1.1 Portfolio Sorts and Monotonicity in Expected Returns

Going back as far as the early tests of the CAPM (Black, Jensen and Scholes (1972)), portfolio sorts have been used extensively in finance as a means to test the relationship between expected stock returns and variables used to rank stocks. Indeed, there are many advantages to considering cross-sectional return patterns at the portfolio level. By achieving diversification across idiosyn-

cratic return movements, portfolios are less affected by noise and outliers than individual stocks. Individual stocks often drop out of the sample as a result of mergers, bankruptcy or de-listings. This leads to unbalanced data structures at the individual stock level. In contrast, provided that a minimal number of firms are available each period, portfolios of stocks can always be formed and so periods with missing observations do not pose the same challenge to time-series tests when conducted at the portfolio level.

Another advantage of using portfolio sorts in uncovering patterns in expected returns is that linearity between the sorting variable and expected returns does not have to be imposed on empirical tests. If the variables used to sort stocks proxy for their loadings on some risk factor, we should expect a monotonic relationship between portfolio rank and expected returns although the relationship need not be linear and will in general depend on the distribution of factor loadings across firms as well as on the relationship between the (observed) sorting variable and the (unobserved) factor loading. Effectively the variable used to sort stocks takes the role of an instrument whose main requirement is that it maps monotonically into the factor loadings. The functional form of this mapping is not required to be known, nor does it have to be estimated. Conversely, if we reject monotonicity, one interpretation is that there cannot exist a one-to-one relationship between the loadings on a risk factor with a non-zero risk premium and the variable used to sort stocks.

This flexibility of monotonicity tests is important in practice. For example, in their analysis of returns on portfolios sorted on firm size and book-to-market ratio, Daniel and Titman (1997, page 6) note that “*Although the returns are almost monotonic in size, there are no significant differences in the returns of the small and medium size firms within any of the book-to-market groupings. However, the largest quintile of firms do have significantly lower returns than the rest, with this being especially true for the high book-to-market stocks. One implication of this is that a simple linear or log-linear regression of returns on capitalization and book-to-market ratios will not adequately characterize observed stock returns.*” Importantly, tests for monotonic relationships between expected returns and portfolio ranks are not subject to this criticism.

Many theoretical models imply a monotonic relationship between expected returns and economic state variables such as size, book-to-market ratio, price-dividend ratio or past returns. However, due to problems associated with measuring the true value of the state variables proposed by theory or due to an unknown functional form of the mapping from these variables to expected returns, conventional asset pricing tests can be difficult to implement. What is needed in such situations is

an approach that can test for monotonicity between portfolios sorted on variables proxying for risk or liquidity and expected returns. We next describe existing tests and then propose an alternative approach.

1.2 Existing Tests

Early work on multivariate one-sided tests includes Bartholomew (1961) and Kudo (1963) who considered the null hypothesis that a set of parameters satisfy $\mu_i = 0$, $i = 1, \dots, N$ against the alternative that $\mu_i \geq 0$, $i = 1, \dots, N$, where at least one inequality is strict. Kudo derived a test statistic whose distribution 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. Intuitively, the weights represent the probability that the sampling statistic has exactly i positive elements. Under the assumption that the covariance matrix is known, Kudo characterized the weights analytically in cases with up to four constraints.

Gourieroux et al. (1982) propose simulation methods to compute critical values for multivariate tests of inequality constraints. This has to be done on a case by case basis due to the dependence of the weights on the covariance matrix of the parameter estimates. To avoid this problem, Kodde and Palm (1986) derive lower and upper bounds on the critical values for the test. Unfortunately, if the test statistic falls between these values, the test is inconclusive.

In a series of papers Wolak (1987, 1989a,b) shows how to construct critical values for multivariate one-sided tests. Under normality of the distribution of the underlying parameter estimates, Wolak (1987) shows that an exact test can be derived. These conditions are not applicable here, however, since stock returns are highly non-normal and in such cases Wolak's result only holds asymptotically. The Wolak test requires estimating both the unconstrained model and the inequality-constrained model, with the latter typically requiring nonlinear optimization. Comparing the unconstrained to the constrained model, a likelihood ratio test is then formed. As in the earlier literature, this is distributed as a weighted sum of chi-squared variables. Critical values of this test statistic are generally not known in closed form, but a set of approximate values can be calculated through Monte Carlo simulation. This procedure is quite complicated and it is not clear that it is tractable in the presence of large numbers of inequalities. As a result, the test has only found limited use in finance. Richardson et al. (1992) applied the method in Wolak (1989a) to test for monotonicity of the term premium, and in our empirical work below we present the results of Wolak's test applied

to our data for comparison.

1.3 A New Approach

To address the shortcomings of existing tests, we next propose a new and simple non-parametric approach that tests directly for the presence of a monotonic relation between expected returns and the underlying sorting variables but does not otherwise require that this relationship be specified or known. This can be a great advantage in situations where 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 the observable firm characteristic used to sort stocks into portfolios and expected returns and thus whether the sorting variable can proxy for stocks' exposure to systematic risk or liquidity sources.

We first explain how the approach works for one-way sorts and then tackle two-way sorts.¹ We take the number of portfolios, N , as given and then show how a test can be conducted that accounts for the relationship between the complete set of portfolios (not just the top and bottom) and their expected returns. In one of our robustness checks, see Section 3, we examine the impact of N on the finite-sample size and power of our test and the t-test.

1.4 One-way Sorts

Let μ_i be the population value of the expected return on the i th portfolio ($i = 1, \dots, N$) obtained from a ranking of stocks using a single sorting variable such as the book-to-market ratio, market capitalization (size) or past returns. Theoretical considerations often dictate that the expected return of the ranked portfolios should be increasing ($\mu_i > \mu_{i-1}$) or decreasing ($\mu_i < \mu_{i-1}$) in the order of the portfolio for $i = 2, \dots, N$.

Suppose we wish to test the null that there is no difference in the expected returns of the sorted portfolios against the alternative that the expected return is decreasing as we move from portfolio 1 through N . This suggests testing the null of a flat pattern in expected returns when moving from portfolio one through portfolio N :

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_N, \tag{1}$$

¹Sorts on three or more variables are not explicitly treated here since they are uncommon in practice, though Daniel, Grinblatt, Titman and Wermers (1997) and Vassalou and Xing (2004) present three-way sorts. Our framework is easily generalized to cover higher-dimensional sorts.

against the alternative of an decreasing pattern

$$H_1 : \mu_1 > \mu_2 > \dots > \mu_N. \tag{2}$$

We write the null hypothesis (H_0) as the absence of a relationship between the portfolio rank and expected returns, and we specify the alternative hypothesis (H_1) as the relationship predicted by economic theory. Rejections of the null therefore support the theory. This is parallel to the standard empirical practice of testing for statistical significance of the coefficient of a variable hypothesized to have a significant effect in a regression. Setting up the problem in this way ensures that the theory is only endorsed if the data provides statistically significant evidence of the hypothesized relationship.

Stated slightly differently, we only look for rejections of H_0 “in the direction” of the hypothesized relationship. That is, we do *not* look for evidence against H_0 in the direction of a non-monotonic relationship, nor do we look for a monotonic relationship that is the opposite of that given by the theory. For example, findings indicating $\mu_1 < \mu_2 < \dots < \mu_N$ would not lead to a rejection of (1) in favor of (2). This means that a rejection of H_0 is evidence of a relationship consistent with economic theory.

Although we write the null hypothesis as an equality in (1), note that if for some portfolios we have $\mu_j < \mu_{j+1}$, this will make rejections of the null against the alternative in (2) even less likely: If in fact expected returns are increasing for some of the portfolios, our test becomes less likely to find the patterns hypothesized under the alternative. Effectively H_1 is set up so that rejections of the null against this alternative provide strong evidence of genuinely monotonically decreasing expected returns.

A test of (1) - (2) maps naturally into the case with portfolios sorted on market capitalization where, as the size increases and we move from small to large stocks, the prediction is that the expected return should decrease. Note that this has to hold for *all* pairs of portfolios and not just when comparing the top and bottom portfolios, although of course (2) also implies that the top portfolio should have higher expected returns than the bottom one.

To test this hypothesis, define the expected return differential of adjacent sorted portfolios

$\Delta_i = \mu_i - \mu_{i-1}$, $i = 2, \dots, N$ and their sample analogs based on portfolio returns $\{r_{it}\}_{t=1}^T$

$$\hat{\Delta}_i = \hat{\mu}_i - \hat{\mu}_{i-1} \quad i = 2, \dots, N$$

where $\hat{\mu}_i \equiv \frac{1}{T} \sum_{t=1}^T r_{it}$.

Then the null and the alternative can be rewritten as

$$H_0 : \Delta_i = 0, \quad i = 2, \dots, N \quad (3)$$

$$H_1 : \max_{i=2, \dots, N} \Delta_i < 0.$$

To see this, note that if the *largest* value of $\Delta_i = \mu_i - \mu_{i-1} > 0$, then we must obviously have that $\mu_i < \mu_{i-1}$ for *all* portfolios $i = 2, \dots, N$. This motivates our choice of test statistic:

$$J_T = \max_{i=2, \dots, N} \hat{\Delta}_i. \quad (4)$$

In Section 1.6 we discuss how we obtain appropriate critical values for this test statistic using a bootstrap procedure. We shall refer to the tests associated with hypotheses such as those in (3) as Monotonic Relationship (MR) tests. Note that testing the converse proposition—namely that expected returns are monotonically increasing—can be done simply by reordering the portfolios.

A naïve approach to testing the hypotheses in (3) would be to conduct a set of pair-wise t -tests that would test that Δ_i is positive for each $i = 2, \dots, N$. The obvious problem with this approach—and the reason why it is not in widespread use—is that it is not clear how to summarize information from these $N - 1$ test statistics into a single number since the test statistics are likely to be correlated and their joint distribution is unknown.

It is worth noting an important difference between our testing approach and that of Wolak (1989a). Corollary 1 of Wolak (1989a) provides a means of testing the following null and alternative hypotheses (in contrast with those in equation 3 above):²

$$H_0 : \max_{i=2, \dots, N} \Delta_i \leq 0 \quad (5)$$

$$H_1 : \max_{i=2, \dots, N} \Delta_i > 0$$

²Corollary 2 of Wolak (1989a) is not applicable to the problem of testing for a uniformly monotonic pattern as it may reject even if just a single pair of portfolios (as opposed to all pairs) produces evidence against the null and thus we do not consider it here.

Thus in Wolak’s framework, the null hypothesis is that there is a (weakly) monotonic relationship between expected returns and the sorting variable, while it is the alternative hypothesis which contains the case of no such monotonic relationship. Depending on the research question and the economic framework, one may prefer to entertain the presence of a monotonic relationship under the null or under the alternative. One important empirical drawback, however, of entertaining the hypothesized monotonic relationship under the null is that limited power (due to a short time series of data, or due to noisy data) will make 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. Our approach, on the other hand, contains the monotonic relationship under the alternative, and thus a rejection of the null of *no* relationship in favor of the hypothesized relationship represents a strong empirical endorsement of the theory. Furthermore, the null in Wolak’s test includes the case of *no* relationship (when $\max \Delta_i = 0$) and so a failure to reject the null could actually be the result of the absence of a relationship between expected returns and the sorting variable.

The null and alternative hypotheses of the Wolak approach can also be tested using simple Bonferroni bounds (see, e.g. Fama (1984)). This is implemented by analyzing whether the largest value of Δ_i generates a (positive) t-statistic that is beyond the critical value obtained by using a Bonferroni bound on the probability of a Type I error. Such an approach will, as usual, be a conservative test of this null hypothesis, and this is confirmed in a Monte Carlo study reported in Section 3 which compares the Bonferroni bound-based test with Wolak’s test and with our proposed MR test.

Note that the test in (3) is based on all ‘non-superfluous’ pairs, a choice that makes our test directly comparable to that in Wolak (1989a). Another possibility is to include all possible pairwise comparisons, $\Delta_{ij} = \mu_i - \mu_j$ for $i > j$. Instead of having $(N - 1) = 9$ pairwise comparisons, this gives $N(N - 1)/2 = 45$ comparisons when $N = 10$. An advantage of the more general setup is that it nests the inequality $\Delta_{10,1} = \mu_{10} - \mu_1$ which underlies the conventional t -test. We compare the adjacent pairs test to the “all pairs” test in the empirical analysis and Monte Carlo simulations.

1.5 Two-way Sorts

It is common practice to sort stocks into portfolios based on two variables such as book-to-market ratio and firm size and to present estimates of average portfolio returns in a two-dimensional table. Our approach is easily generalized to cover such two-way sorts. To see this, suppose that the

outcome of the two-way sort is reported in an $N \times N$ table with sorts according to one variable ordered across rows and sorts by the other variable listed down the columns. Suppose we are interested in testing the hypothesis that expected returns decrease both down the column and along the rows. The proposition of no systematic relationship – which we seek to reject – is entertained under the null. To formalize the test, let the expected value of the return on the row i , column j portfolio be denoted μ_{ij} :

$$H_0 : \mu_{ij} = \mu_{kl} \quad \text{for all } i, j, k, l \in [1, \dots, N]. \quad (6)$$

The alternative hypothesis is that the expected return is decreasing in both the row and column index:

$$H_1 : \mu_{ij} > \mu_{kl} \quad \text{for all } k \geq i \text{ and } l \geq j. \quad (7)$$

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

$$\begin{aligned} H_0 & : \Delta_{ij}^r = \Delta_{ij}^c = 0, \quad \text{for all } i, j \\ H_1 & : \Delta_{ij}^r < 0 \text{ and } \Delta_{ij}^c < 0, \quad \text{for all } i, j, \end{aligned} \quad (8)$$

or, equivalently,

$$H_1 : \max_{i,j=2,\dots,N} \Delta_{ij}^r < 0 \quad \text{and} \quad \max_{i,j=2,\dots,N} \Delta_{ij}^c < 0. \quad (9)$$

In parallel with the one-way sort, this gives rise to a test statistic

$$J_T = \max_{i,j=2,\dots,N} \{\hat{\Delta}_{ij}^r, \hat{\Delta}_{ij}^c\}. \quad (10)$$

To see how many inequalities the alternative hypothesis gives rise to, note that the hypothesis that expected returns decline along the first row of an $N \times N$ table of portfolio sorts yields $N - 1$ inequalities:

$$\mu_{1,j} < \mu_{1,j-1} \quad j = 2, \dots, N. \quad (11)$$

Similarly, the hypothesis that expected returns decline along the first column implies another $N - 1$ inequalities

$$\mu_{i,1} < \mu_{i-1,1} \quad i = 2, \dots, N. \quad (12)$$

Finally, for all elements of the table, μ_{ij} , that are not in the first row or in the first column, two restrictions must hold simultaneously provided that expected returns are decreasing in both the

row and column index:

$$\mu_{i,j-1} > \mu_{i,j}, \mu_{i-1,j} > \mu_{i,j} \quad i, j = 2, \dots, N. \quad (13)$$

This gives rise to another $2(N-1)^2$ inequalities.³ When combined with the first $2(N-1)$ inequalities in (11) and (12), the assumption of a monotonic relationship along the two dimensions gives rise to a total of $2N(N-1)$ inequalities.

To see what these results imply in practice, note that for a 5×5 table, 40 inequalities are implied by the theory of a monotonic relationship in expected returns along both row and column dimensions, whereas for a 10×10 table 180 inequalities are implied. This shows both how potentially complicated and how rich the full set of relations implied by theory can be when applied to two-way portfolio sorts.

Our results are easily generalized to cases where the number of rows and columns differs. For an $N \times K$ table, there will be $2NK - K - N$ inequalities to test. Moreover, for a D -dimensional sort, with N portfolios in each direction, the total number of inequalities amounts to $DN^{D-1}(N-1)$.

1.6 A Bootstrap Approach

Under standard conditions, provided in detail below, the vector of sample average returns for the N portfolios, $[\hat{\mu}_{1T}, \dots, \hat{\mu}_{NT}]'$, will asymptotically follow a normal distribution, i.e., in large samples, T ,

$$\sqrt{T}([\hat{\mu}_{1T}, \dots, \hat{\mu}_{NT}]' - [\mu_1, \dots, \mu_N]') \stackrel{a}{\sim} N(\mathbf{0}, \Omega). \quad (14)$$

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, Ω . These parameters influence the distribution of the test statistic even though we are not otherwise interested in them. Unfortunately, when the set of portfolios involved in the test grows large, the number of parameters increases significantly—e.g., 5,050 covariance parameters are required for a 10×10 portfolio sort—and it can be difficult to estimate these parameters with much precision.

As shown in (3) and (9), we are interested in studying the maximum value of a multivariate vector of sample moments which is asymptotically normally distributed. Unfortunately, there are no tabulated critical values for such maximum or minimum values—precisely because these would

³Note that we do not need to add the inequality $\mu_{i-1,j-1} < \mu_{i,j}$ since this is already implied by the inequalities in equation (13).

depend on the entire covariance matrix, Ω . Furthermore, the asymptotic distribution may not provide good guidance to finite sample distributions 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 et al. (2006) and use a bootstrap methodology. As pointed out by White, a major advantage of this approach is that it does not require estimating Ω directly. To see how the approach works in practice, let $\{r_{it}, t = 1, \dots, T; i = 1, \dots, N\}$ be the original set of returns data recorded for N portfolios 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 $\{\hat{r}_{i\tau(t)}^{(b)}, \tau(1), \dots, \tau(T); i = 1, \dots, N\}$, 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 in order to preserve any cross-sectional dependencies in returns. Finally, b is an indicator for the bootstrap number which runs from $b = 1$ to B . The number of bootstrap replications, B , is chosen to be sufficiently large that the results do not depend on Monte Carlo errors. In our application we set $B = 1000$. 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. In our empirical work below we choose the average block length to be ten months, which seems appropriate for returns data which display limited time-series dependencies at the monthly horizon.

The randomly drawn return series are recentered by subtracting the original sample means $(\hat{\mu}_1, \dots, \hat{\mu}_N)'$, thus ensuring that the bootstrapped data satisfy the null hypothesis by construction. As argued by Wolak (1987), this is the uniquely least favorable value of μ under the null.⁴ We then count the number of times where a pattern at least as unfavorable (i.e. yielding at least as small 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 statistic J_T and allows us to conduct inference⁵. We implement a “studentized” version of this bootstrap, as advocated by Hansen (2005) and Romano and Wolf (2005), which eliminates the impact of cross-sectional heteroskedasticity in

⁴By analogy, the least favorable value of the mean of a Gaussian variable is zero in a test of the null hypothesis that the mean is greater than or equal to zero.

⁵Matlab code to implement the tests proposed in this paper is available from <http://www.economics.ox.ac.uk/members/andrew.patton/code.html>.

the portfolio returns, a feature that is prominent in some common portfolio sorts (those based on market capitalization, for example), and may lead to gains in power.

Proposition 1 below provides a formal justification for the application of the bootstrap to our problem. Its proof is presented in the Appendix.

Proposition 1 *Let $\mathbf{r}_t \equiv [r_{1t}, r_{2t}, \dots, r_{Nt}]'$ be a strictly stationary process and define $\boldsymbol{\mu} \equiv E[\mathbf{r}_t]$, $\hat{\boldsymbol{\mu}}_{iT} \equiv T^{-1} \sum_{t=1}^T \bar{r}_{it}$. Assume $E[|r_{it}|^{6+\varepsilon}] < \infty$ for some $\varepsilon > 0$ for all $i = 1, 2, \dots, N$, and that $\{\mathbf{r}_t\}$ is α -mixing of size $-3(6+\varepsilon)/\varepsilon$. Then as $T \rightarrow \infty$*

$$\sqrt{T}(\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}) \Rightarrow N(0, \Omega)$$

$$\text{and } \max_{i=2, \dots, N} \sqrt{T} \{(\hat{\boldsymbol{\mu}}_{iT} - \hat{\boldsymbol{\mu}}_{i-1,T}) - (\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i-1})\} \Rightarrow W \equiv \max_{i=2, \dots, N} \{Z_i - Z_{i-1}\}$$

where $[Z_1, \dots, Z_N]' \sim N(0, \Omega)$, and \Rightarrow denotes convergence in distribution. Further, if b_T is the length of the average block in the stationary bootstrap, and $b_T \rightarrow \infty$ and $b_T/T \rightarrow 0$ as $T \rightarrow \infty$, then

$$\begin{aligned} & \sup_z |P^* [\|\hat{\boldsymbol{\mu}}_T^* - \hat{\boldsymbol{\mu}}_T\| \leq z] - P[\|\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}\| \leq z]| \rightarrow^p 0, \text{ as } T \rightarrow \infty \\ \text{and } & \sup_z \left| P^* \left[\min_{i=2, \dots, N} \sqrt{T} \{(\hat{\boldsymbol{\mu}}_{iT}^* - \hat{\boldsymbol{\mu}}_{i-1,T}^*) - (\hat{\boldsymbol{\mu}}_{iT} - \hat{\boldsymbol{\mu}}_{i-1,T})\} \leq z \right] \right. \\ & \left. - P \left[\min_{i=2, \dots, N} \sqrt{T} \{(\hat{\boldsymbol{\mu}}_{iT} - \hat{\boldsymbol{\mu}}_{i-1,T}) - (\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i-1})\} \leq z \right] \right| \rightarrow^p 0, \text{ as } T \rightarrow \infty \end{aligned}$$

where $\|\cdot\|$ represents a norm on \mathbb{R}^N and P^* is the probability measure induced by the bootstrap conditional on the original data.

In words, under a standard set of moment and mixing conditions on returns, the appropriately scaled vector of sample moments converges to a multivariate normal distribution. Moreover, inference about the maximum of this distribution can be conducted by means of the stationary bootstrap provided that the average block length does not grow too fast. The approach is non-parametric and does not require making distributional or functional form assumptions about the relationship between the sorting variable and expected returns.

2 Empirical Results

Having introduced our methodology in the previous section, we next revisit a range of empirical findings from the finance literature using portfolio sorts. We compare the outcome of tests based

on our new monotonic relationship approach to those based on standard t -tests that compare the average return differential between top and bottom ranked portfolios.

2.1 One-way Portfolio sorts

We consider five portfolios sorted on firm characteristics such as market equity (size), book-to-market ratio, cashflow-price ratio, earnings-price ratio and the dividend yield. The first variable measures market value, whereas the last four variables capture some measure of firms' fundamental value relative to their market price. In addition, we study three portfolios related to patterns in past stock returns, namely short-term reversal, momentum and long-term reversal.

Data on returns from these portfolio sorts are obtained from Ken French's web site at Dartmouth college.⁶ Portfolios comprise stocks listed on NYSE, AMEX and NASDAQ. Returns on value-weighted portfolios up to the end of 2006 are used, but results are very similar when using equal-weighted portfolios. Data begin in 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. The portfolios formed on the basis of past returns consider returns during the previous month (short-term reversal), the previous year (momentum) and the previous five years (long-term reversal).

For some sorts the hypothesis is that there should be a declining pattern in expected returns (e.g. related to firm size), while in other cases it is that of an increasing pattern (e.g. expected returns versus book-to-market ratio). To make the results easier to interpret and compare across portfolio sorts, whenever appropriate we therefore reverse the order of the portfolios so the expected return of the first ("top") decile portfolio should always exceed that of the tenth ("bottom") decile portfolio. This is purely a normalization and has no effect on the results.⁷ In all cases we are therefore interested in testing for a monotonically declining pattern in expected returns.

Table 1 reports estimates of expected returns for the decile portfolios sorted on the eight variables listed in the columns. The original findings on cross-sectional return patterns in portfolios sorted on various firm characteristics reported by Fama and French (1992) were based on data

⁶Details on how these portfolios are formed are available from

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

⁷This affects the portfolios sorted on the book-to-market, cashflow-price, earnings-price or dividend-price ratios and the momentum-sorted portfolios for which we reverse the order of the ranks.

starting in July 1963. Subsequently, there has been much discussion of whether results for this period are different to those for earlier periods. For this reason we present results for both the full sample and the post-July 1963 period.

First consider the full sample results (Panel A). For the portfolios sorted on size (ME) a monotonic pattern is observed with a strong decline in average returns from the top to the bottom ranked stocks. For the portfolios sorted on book-to-market value (BE-ME), again a strong and nearly monotonic pattern in average returns emerges only interrupted by a slight increase in the average returns from decile seven to eight by less than one basis point per month. Similar results obtain for the portfolios sorted on the cashflow-price (CF-P) or earnings-price (E-P) ratios – in the latter case the decline in average returns across deciles is again monotonic.

In contrast with the strong monotonic or near-monotonic declines in average returns observed for the first four portfolios, column 5 shows that the portfolios sorted on the dividend-price ratio experience several reversals in the expected return pattern of up to 12 basis points. Clearly the evidence of a monotonically decreasing pattern in expected returns is much weaker when stocks are sorted on the dividend-price ratio.

Turning to the portfolios sorted on past returns (columns 6-8), those sorted on momentum display a monotonic pattern in average returns while portfolios sorted on short-term reversal show just one deviation from a monotonically declining pattern. In contrast, the average returns of the portfolios sorted on long-term reversal increase by seven basis points per month from decile four to five and by three basis points from decile six to seven.

In the post-1963 sample there are large reversals (increases) 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 increase from portfolio six to seven) and short-term reversal (an increase of 14 basis points from portfolio one to two). For the portfolios sorted on the book-to-market, cash flow-price or earnings-price ratios or long term reversal there are either no reversals or few and smaller ones compared with those observed for the other sorts.

In conclusion, for some portfolio sorts a monotonic pattern is observed in average returns while for other portfolio sorts non-monotonic patterns of varying degree arise. A key question is clearly how strong the deviations from a monotonic pattern must be in order for us to fail to reject the null hypothesis and establish a monotonically declining pattern in expected returns.

To answer this question, Table 2 presents test results for the eight portfolios. Again Panel A uses the full sample, while Panel B uses the post-1963 data. The first column shows the spread—in basis points per month—in the average return recorded for the top and bottom portfolios. In the full sample this ranges from 17 basis points per month for the portfolios sorted on the dividend yield to 100 and 125 basis points per month for the portfolios sorted on short-term reversal and momentum, respectively. Spreads between 50 and 70 basis points per month are recorded for the five remaining portfolios.

The second column of Table 2 reports the t -statistic for testing the significance of the difference in expected returns between the top and bottom portfolios based on heteroskedasticity and autocorrelation consistent standard errors (Newey and West (1987)). In the full sample the t -statistics range from 0.94 for the portfolios sorted on the dividend yield to more than five for the portfolios sorted on short-term reversal or momentum. The third column gives the associated p -value. All p -values fall below 1% with exception of the portfolios sorted on the dividend yield. Such strong results are not really surprising— we are, after all, focusing on portfolios deemed sufficiently significant by previous researchers to stand out in a very large empirical literature.

Does this evidence carry over to our more comprehensive test (denoted the MR test) for a monotonic relationship in expected returns across all ten portfolios? To see if this is the case, the fifth column in Table 2 reports the bootstrapped p -values associated with the test for a monotonically decreasing pattern in expected returns as described in (3). For six of eight portfolios the p -value is indeed below one percent in the full sample, confirming and strengthening the original conclusion about a significant monotonic relationship between expected returns and sorting variables such as short-term reversal, momentum, book-to-market ratio, size, cashflow-price and earnings-price ratios.

For the portfolios using long-term reversal as the sorting variable, the outcome based on the new MR test differs markedly from that based on the traditional t -test. Here a non-monotonic pattern emerges as witnessed by the p -value of 0.22. As shown in Table 1, this is related to the large increase in average returns from decile four to five and the smaller increase from decile six to seven. Wolak’s (1989a) test fails to reject the null of a (weakly) monotonic relationship for all eight portfolios.⁸ Even the conventional t -test found no evidence of a monotonic pattern for the

⁸The results reported here are based on 1000 Monte Carlo simulations to obtain the critical values of the Wolak test.

portfolios sorted on the dividend yield and so this illustrates the difficulty in interpreting the Wolak test: Failure to find evidence against a weakly monotonic pattern in the sorted portfolio returns may simply reflect weak power. The Bonferroni-based version of Wolak's test yields the same conclusions as Wolak's test for these series: no evidence against the null of a weakly monotonic pattern in expected returns.

Our MR test, of course, accounts for the effects of random sampling variation. This has important implications economically, in addition to the usual statistical motivations. For example, for the portfolios sorted on book-to-market where one or two reversals appear in the ordering of average returns, the test still rejects very strongly because these reversals are small in magnitude (less than one basis point per decile portfolio) relative to the sampling variability of the average returns. In contrast, the relationship between expected returns and the dividend yield continues to be insignificant with a p -value of 0.54. This is to be expected given the large reversals in the mean return patterns observed for the portfolios sorted on this variable.

Turning to the test results for the post-1963 sample, Panel B of Table 2 shows that the spread in average returns between the top and bottom portfolios now ranges 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. The associated t -statistics for the top minus bottom ranked portfolios now range from 0.30 (for portfolios sorted on the dividend yield) to 5.7 in the case of the momentum-sorted portfolios. Once again the portfolios sorted on the dividend yield fail to produce a statistically significant top-bottom spread. Moreover, the spread between the average returns on the smallest and largest stocks is now borderline insignificant with a p -value of 0.06. Once again, the Wolak test fails to reject for all eight portfolios.

Very different results follow from our MR test which shows that, for the post-1963 returns data, there is no monotonic relationship 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 final column in Table 2 shows that in every single case (both in the full sample and in the post-1963 sample), the MR tests based on comparing only the adjacent portfolios versus comparing the full set of portfolios lead to the same inference.

We conclude from our comparison of the portfolios sorted on short-term reversal (p -values of 0.26 under the MR test versus 0.01 under the t -test using the monthly data), momentum (0.29 versus 0.00) and size (0.27 versus 0.06) that very different inference can emerge from a simple t -test which only considers the expected returns of the top and bottom portfolios versus that from the proposed test for a monotonic pattern in the full set of expected returns. Clearly the more rigorous test for a monotonic relationship between firm characteristics and size that we propose is more difficult to satisfy—and ultimately more informative—than standard t -tests based only on the top-bottom spread.

2.2 Two-way Portfolio Sorts

We next consider a range of two-way sorts that combine portfolios sorted on firm size with portfolios sorted on one of four variables, namely book-to-market ratio, momentum, short-term reversal and long-term reversal. In all cases we study 5×5 portfolio sorts. The sample begins in 1926 or 1927 for the 5×5 sorts on size and book-to-market ratio, momentum or short-term reversal. For the sort on size and long-term reversal, the sample begins in July 1963.

Average returns for these 5×5 sorted portfolios are reported in Table 3, while the bottom right corners of the panels in Table 4 report the associated monotonicity test results. If necessary we have reordered the portfolio sorts so the relevant hypothesis is that expected returns should be declining along both rows and columns. In all cases the rows represent portfolios sorted on firm size going from the smallest to the largest quintile of firms.

For the portfolios sorted on size and book-to-market ratio (panel A) there is some evidence of a non-monotonic pattern with increases of up to 13 basis points per month observed in the fourth and fifth columns (which capture growth firms with a low B/M ratio). Overall, however, in the vast number of cases the mean return pattern is in line with the prediction that expected returns should be declining in firm size and increasing in the book-to-market ratio. Consistent with this, the MR test results in Table 4 show that the null of no systematic relationship between size and book-to-market ratio on the one hand and expected returns on the other hand is rejected against the alternative of a monotonic relationship for the 5×5 sort with an MR test p -value of 0.01.⁹

⁹We also considered a 10×10 two-way sort on size and book-to-market ratio which begins in July 1963. For the sort on this sample period the test generated an MR p -value of 0.39 and could not be rejected. This seemingly contradictory finding is easily explained by the fact that the 10×10 sort is based on post-1963 data, due to the

For the portfolios sorted on size and momentum, panel B shows only slight reversals, less than five basis points, when moving from rows two to three or rows three to four in the fifth quintile of the momentum-sorted portfolios. As expected, the MR test in Table 6 produces strong evidence of a monotonic relation between size, momentum and expected portfolio returns, strongly rejecting the null of no such relationship (p -value of 0.001).

Finally, the portfolios sorted on size and short- or long-term reversal (panels C and D) show examples of large deviations from a monotonic decline in the average returns of the fifth column when moving across rows (increasing instead by 60 basis points when moving from the smallest to the biggest firms in the fifth short-term reversal quintile). Hence for these sorts there is either a larger number of cases with non-monotonic patterns in average returns (as in the case of the size versus long-term reversal sort) or examples of very large reversals in the patterns (as in the case of the size versus short-term reversal sort). This helps explain why the MR test fails to reject for both the size and long-term reversal two-way sort (p -value of 0.10) and for the size and short-term reversal sort (p -value of 0.82).

Results were distinctly different under the top-minus-bottom t -test—which rejected the null in all four cases with p -values below 0.1%—and the Wolak (1989a) test which failed to reject the null of a weakly monotonic relationship in all the two-way sorts considered here.

When conducted on the two-way sorted portfolios the monotonic relationship tests typically involve a large number of inequalities. It is therefore useful to decompose the overall (joint) test into a series of conditional tests that help identify the economic source of the results. Towards this end, Table 4 looks further into the two-way sort results, using the MR test to examine patterns in expected returns keeping one sorting variable (e.g. size) constant while varying another (e.g. book-to-market ratio) or vice versa. For example, the penultimate row of the first panel of Table 4 presents the p -values from the tests for a monotonic relationship between the portfolio sorting variable in the row (size) and expected returns, conditional on keeping the column portfolio (the book-to-market ratio) fixed. Hence the p -value on a test for the size effect, conditional on being in

limited number of firms available in the older data, while the 5×5 sort uses the longer data going back to 1926. As shown in Tables 2 and 4, the relationship between size and expected returns appears to be uniformly monotonic in the longer data sample but not in the shorter (post-1963) sample. Hence we would expect to find a non-monotonic relationship for the 10×10 sort based on the shorter sample and a monotonic relationship for the 5×5 sort based on the longer sample – which is exactly what we find.

the top book-to-market quintile (value stocks) is 0.002, while it is 0.431 if we condition on being in 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 relationship between the sorting variable in the row, across all column portfolios. The penultimate column presents corresponding results from tests for a monotonic relationship between the portfolio sorting variable in the columns and expected returns, conditional on the row portfolio, representing firm size.

The results in Table 4 confirm some existing findings, and shed new light on others. For example, panel A reveals that the value effect holds within all size portfolios (with p -values ranging from 0.00 to 0.02). Consistent with Loughran (1997), the book-to-market effect, as measured by the spread in the top-minus-bottom portfolios' average returns, is much stronger for the smallest stocks than for the largest stocks. However, at least for the sample considered here (1926-2006), the statistical evidence appears to support the conclusion in Fama and French (2006) that there is a value effect even among large stocks. Conversely, the size effect is only significant for value firms; for the bottom two book-to-market portfolios the size effect is non-monotonic, mainly due to low average returns among small growth firms.

The results for the size and momentum two-way sorts provide strong support for the momentum effect. The p -values in the penultimate column of panel B in Table 4 are all less than 0.01, indicating that the momentum effect is present in all five size portfolios. Thus although the spread on the momentum portfolios is greater for small caps than large caps (1.05% versus 0.60% per month), the effect is strongly significant for all size portfolios. A similar result holds for the short-term reversal sort in panel C, although the joint test fails to reject the null of no systematic relationship, which we can see is entirely due to the reversal of the size effect in the fifth short-term reversal portfolio. In fact, expected returns in that portfolio are monotonically *increasing* in size, rather than decreasing as in the other portfolios.

Finally, the results for the size versus long-term reversal sorts in Panel D reveal that although the long-term reversal spreads are greatest for small stocks, the evidence for a monotonic relationship between expected returns and the long-term reversal factor is strongest for the large stocks. In fact, for all but the largest quintile of stocks the MR test fails to identify a monotonic pattern in expected returns, thus helping to explain why the MR test fails to reject the null against a monotonically declining relation.

To summarize, the empirical results show that for a number of portfolio sorts in common use

the MR test reveals evidence of a non-monotonic relation between portfolio rank and expected returns which standard t -tests conducted on the top versus bottom portfolios fail to identify. Our monotonic relationship test delivers in a single number (p -value) a summary measure of whether a large number of inequalities hold jointly for two-way sorted portfolios. Moreover, the test can be decomposed which allows us to identify which portfolios give rise to evidence in contradiction with the underlying theory.

3 Robustness Analysis and Extensions

This section studies the performance of the MR test using Monte Carlo simulations and bootstrap analysis. We also analyze the impact of the choice of the number of portfolios to use on the size and power of these tests. Finally, we show how our approach can be extended to test for monotonicity in the patterns of individual regression coefficients (e.g. factor loadings) and piece-wise monotonic patterns in expected returns.

3.1 Monte Carlo Study

The hypothesis tests proposed here are non-standard. Moreover, unlike the standard t -test for equal expected returns, there are no optimality results or closed-form distributions against which test statistics such as (4) or (10) can be compared and from which critical values can be computed. Since we are effectively in uncharted territory, we next undertake a series of Monte Carlo simulation experiments that offer insights into the finite-sample behavior of the proposed tests.

The first set of scenarios covers situations where the hypothesized theory is true and there is a monotonic relationship between portfolio rank and the portfolios' true expected returns. Good tests reject the null of no systematic relationship in this situation 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 1 through portfolio 5, and then increase by the remaining 20% of the total across the remaining 5 portfolios. Experiment III assumes a single large increase in the expected return from decile 1 to decile two, equal to 50% of the total increase, and then spreads the remaining 50% of the increase across the 9 remaining portfolios. These three patterns are illustrated in

the first column of Figure 2 and all have in common that the theory of a monotonic relation holds.

The second set of scenarios covers situations where the theory fails to hold and there is in fact a non-monotonic relationship between portfolio ranks and expected returns. For these cases, a comprehensive test that considers all portfolios jointly should not reject the null hypothesis. However, tests that only consider the relative expected return of the top and bottom portfolios may reject. This outcome illustrates the limited conclusions that can be drawn from traditional pair-wise means tests.

Experiments IV-VI all break the monotonic pattern in expected returns in some way. Experiment IV assumes a symmetric Λ -shaped pattern where expected returns first rise and then decline so the expected return of the first and tenth deciles are identical. 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. Finally, experiment VI assumes a non-monotonic pattern with declines in expected returns for every second decile.

Each pattern is multiplied by a step size which varies from a single basis point per month to two, five and ten 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 the one-dimensional sorts with $N = 10$ portfolios. 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 new returns data using 2500 Monte Carlo simulations which reshuffle the original returns. For each simulated data set we then use $B = 1000$ replications of the stationary bootstrap of Politis and Romano (1994).

3.1.1 Analytical results under Normality

In order to obtain simple analytical results, we first make the assumption that the estimated differences in portfolio returns are independently and normally distributed:

$$\Delta\hat{\mu}_i \sim N\left(\Delta\mu_i, \frac{1}{T}\sigma_i^2\right), \text{ for } i = 2, \dots, N \quad (15)$$

$$\text{Corr}[\Delta\hat{\mu}_i, \Delta\hat{\mu}_j] = 0 \text{ for all } i \neq j. \quad (16)$$

This setup allows us to present formulas for the power of the tests and establish intuition for which results to expect. Under the assumptions given in equations (15) and (16), 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, which is given by

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

Assuming that the variances are known, the t -statistic will thus be

$$tstat \equiv \frac{\sqrt{T}(\hat{\mu}_N - \hat{\mu}_1)}{\sqrt{\sum_{i=2}^N \sigma_i^2}} \sim N \left(\sqrt{T} \frac{\mu_N - \mu_1}{\sqrt{\sum_{i=2}^N \sigma_i^2}}, 1 \right). \quad (17)$$

Under the null we have $\mu_N = \mu_1$ and so the t -statistic has the usual $N(0, 1)$ distribution. Under the alternative hypothesis that $\mu_N > \mu_1$ the t -statistic will, as usual, diverge 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

$$\Pr [tstat > 1.645] = \Phi \left(\sqrt{T} \frac{\mu_N - \mu_1}{\sqrt{\sum_{i=2}^N \sigma_i^2}} - 1.645 \right), \quad (18)$$

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

For our proposed test of monotonically increasing expected returns, the power is obtained as follows. Recall our test statistic:

$$J_T \equiv \min_{i=2, \dots, N} \Delta \hat{\mu}_i.$$

To obtain the distribution of this statistic under the null, we use 100,000 simulated draws to compute critical values, denoted $J_T^*(\boldsymbol{\sigma})$. The power of our test is then simply

$$\Pr [J_T(\Delta \boldsymbol{\mu}, \boldsymbol{\sigma}) > J_T^*(\boldsymbol{\sigma})].$$

We compute this power using 100,000 simulated draws and set $T = 966$, which is the number of monthly returns on the “size” portfolios since 1926.

Results for this benchmark case are presented in the first column of Table 5 labeled “normal simulation”. These results show that the power of the t -test converges to one as the step size grows from 1 to 10 basis points, for experiments I, II and III. The probability of rejecting the null

also approaches one for experiment VI, which assumes a non-monotonic but increasing pattern of expected returns.

The new MR test has slightly lower power than the t -test for the three experiments in which both tests should reject the null (experiments I, II and III). When the step size is 10 basis points the power of the bootstrap test is 0.85, 0.68 and 0.73 in these experiments.

Compensating for this modest reduction in power, we observe that the probability that the bootstrap rejects the null in experiments IV, V and VI—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 exceeds seven percent. Thus the bootstrap test is very unlikely to falsely reject the null hypothesis. In contrast, the standard t -test very frequently rejects the null particularly under experiment VI when the step size is comparable to that observed for the majority of portfolio sorts in the empirical analysis, i.e. 5-10 basis points. Of course, the t -test is not “wrong”; however it is limited in its scope since it only compares the top and bottom portfolios and thus fails to detect non-monotonic patterns in the full portfolio sorts.

3.1.2 Bootstrap simulation results

The second set of columns in Table 5 present the results from the simulation based on bootstrap draws of monthly returns on the size sorted decile portfolios, so again $T = 966$. Broadly stated, the results from the t -test and MR test from these simulations are comparable to those obtained under the Normality assumption described previously. For comparison, we also present the results from applying Wolak’s (1989a) test and a test based on Bonferroni bounds to these simulated returns. Recall that Wolak’s test and the Bonferroni-based test have a weakly monotonic relationship under the null hypothesis, and a non-monotonic relationship under the alternative. Thus, in contrast with the t -test and the MR test, these tests should *not* reject the null for returns generated under experiments I-III, while they should reject the null hypothesis for returns generated using experiments IV-VI.

The first panel, with step size set to zero, shows that the Bonferroni, Wolak, MR and t -tests have roughly the correct size when there is genuinely no relationship between expected returns and portfolio rank, although the latter two tests slightly over-reject the null hypothesis. This is not an unusual finding and mirrors those in simulation studies of the finite-sample size of asset pricing tests, see e.g. Campbell, Lo and MacKinlay (1997), section 5.4.

Under experiment I, the t -test and MR test have similar power for all step sizes, though the t -test has slightly greater power. 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. This rejection rate increases to around 20% under the two basis point differential and to a 50% rejection rate under the five basis point return differential. Compared with the empirical estimates of the spreads in Table 2, this experiment appears to be the empirically most relevant example for four of the portfolios. Finally, under the largest step size with a 10 basis point return differential per portfolio, the rejection rate is 84%-97%. For this case the average t -statistic is close to four and both the t -test and the MR test are quite powerful. Wolak's test and the Bonferroni-based test should not reject the null under experiment I and this is indeed what is found in the simulation results.

In experiments II and III, the expected return pattern is monotonic but non-linear, and both tests should again reject the null hypothesis. The t -test is of course unaffected by the presence of a kink in the expected returns, as it only reflects the difference in expected returns between portfolio 1 and portfolio 10. In contrast, the power of the MR test is influenced by the presence of small segments to the extent that they affect the size of the smallest return differential ($\min_i \Delta\mu_i$). For a step size equal to 5 basis points, for example, the power of the MR test is around 30% in experiments II and III, compared with 47% in experiment I. For a step size equal to 10 basis points, the power of the MR test is 57% for experiment II and 65% in experiment III, compared with 84% in experiment I.

Turning to the second set of experiments involving a non-monotonic relationship between expected returns and portfolio ranks, 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 VI with the largest step size. These are cases where we do not want a test to reject if the theory implies a monotonic relationship between portfolio rank and expected returns. For these experiments we expect Wolak's test and the Bonferroni-based test to reject the null hypothesis of a weakly monotonic relationship: for step sizes less than 5 basis points neither of these tests exhibit much power, but for step sizes of 5 and particularly 10 basis points these two tests do detect the non-monotonic relationship, with Wolak's test having between 2% and 15% better power than the Bonferonni-based test.

Note that the three tests consider different hypotheses and so their size and power are not

directly comparable: The t -test only compares the top and bottom portfolio; the MR does consider all portfolios and continues to have equality of means as the null and inequality as the alternative; finally, the Wolak test has 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 most 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-VI, 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, as the empirical results and simulations clearly illustrate, this could simply be due to the test having weak power. Finally, the MR test does not reject the null when the evidence contradicts the theory as in experiments IV-VI, and it appears to have good power for step sizes that match the data.

The final column in Table 5 labeled ‘MR^{all}’ shows the result of using the full set of possible pair-wise inequalities in the test. As noted earlier, for the one-way table with $N = 10$, this entails comparing 45 rather than 9 pairs of portfolio mean returns. There does appear to be a small gain in power from including the full set of inequalities, although this may in part reflect that this approach leads to a slightly oversized test.

3.2 Choosing the number of portfolios

So far we have taken the number of portfolios as given, however a key question when implementing portfolio sorts is how many portfolios to use. Intuition suggests not using too many - grouping stocks averages out idiosyncratic effects and allows the use of data from stocks with unequal return histories. At the same time, one should not use too few portfolios since including too many stocks in a portfolio will make it difficult to find the effect of interest. In practice, the finance literature has largely settled on using a standard number of portfolios in empirical tests such as 10 for one-way sorts, and 5×5 or 2×3 portfolios for two-way sorts.

Nevertheless, it is of interest to see how sensitive the tests are to this design parameter. To address this point, we maintain the simulation setup with 10 portfolios from Table 5 but further aggregate returns on these portfolios to consider values of N equal to two, three or five. The case with two portfolios is formed by aggregating the returns on portfolios 1-5 and 6-10; the case with three portfolios aggregates portfolios 1-3, 4-7 and 8-10; and the case with five portfolios aggregates portfolios 1-2, 3-4, 5-6, 7-8 and 9-10.

Table 6 shows the outcome of our simulations, based on the Normal approximation described in Section 3.1.1. Similar results (available upon request) were obtained using a bootstrap-based simulation. First consider the t -test that compares the mean returns of the top and bottom ranked stocks. We find a non-monotonic relationship between the power of this test and the number of portfolios, N . This is due to two effects that pull in opposite directions: Reducing N means aggregating over larger cross-sections of stocks and leads to smaller standard errors in t -tests of mean return differentials, thus improving the power of the tests for a fixed return spread. However, reducing N also means a smaller difference in the return spreads since it leads to the inclusion of stocks with larger mean returns in the bottom portfolio and the inclusion of stocks with smaller mean returns in the top portfolio, thus smoothing out return differentials.

Turning to the MR test, Table 6 suggests that the power of this test is generally declining in the number of portfolios, N , although the decline is often quite modest.¹⁰ Against this stands the finding that too much cross-sectional aggregation—that is, using too few portfolios—may cover up truly non-monotonic patterns. To see this, Figure 3 shows the patterns in expected returns for different numbers of portfolios for experiments IV-VI, all of which involve non-monotonic patterns in the underlying set of ten expected return series. When $N = 2$, the pattern in expected returns appears monotonic, while it is clearly non-monotonic for $N > 2$. This also happens for $N = 3$ and $N = 5$ in experiment VI. In these cases the MR test would report a monotonic pattern in expected returns when this is in fact false at the disaggregated portfolio level.

These results shed light on the basic trade-off that researchers should consider when choosing the number of portfolios to use in a sorting experiment: Modest gains in power can be made by reducing the number of portfolios, however using a small number of portfolios may ‘smooth out’ genuinely non-monotonic patterns in expected returns, and thus increase the possibility of falsely rejecting the null hypothesis.

¹⁰This finding is consistent with the power gains that Campbell, Lo and MacKinlay (1997) report for F -tests of the CAPM as N is reduced, see their Table 5.2 for example.

3.3 Extensions

3.3.1 Monotonic Patterns in Risk-Adjusted Returns or Factor Loadings

In common with other tests based on portfolio sorts, our approach does not test whether asset pricing models hold exactly in the sense that the mean of the residuals from a beta pricing model equals zero for all portfolios. This is best done in a second step provided that the first step corroborates that the sorting variables are systematically related to expected returns in the first place. In a sense our approach tests a necessary but not a sufficient condition that must hold for any asset pricing model in which expected returns are monotonically related to the sorting variable.

The methodology can, however, 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, obtained via a maintained asset pricing model, are monotonically increasing (or decreasing) in past performance. Alternatively, a corporate finance model may imply that the sensitivity of returns (or sales, or free cash flow) to a credit constraint factor is monotonically decreasing in firm size. These examples, and our original specification above, are nested in the more general framework

$$\begin{aligned} r_{it} &= \beta_i' \mathbf{F}_t + e_{it}, \quad i = 1, 2, \dots, N \\ \beta_i &\equiv [\beta_{0i}, \beta_{1i}, \dots, \beta_{Ki}]', \end{aligned} \tag{19}$$

with the associated hypotheses

$$\begin{aligned} H_0 &: \beta_{j1} = \beta_{j2} = \dots = \beta_{jN}, \\ H_1 &: \beta_{j1} < \beta_{j2} < \dots < \beta_{jN} \quad (0 \leq j \leq K). \end{aligned} \tag{20}$$

Our framework in the previous sections corresponds to regressing each portfolio return onto a constant and so emerges when $K = 0$ and $F_{0t} = 1$ for all t . A test for monotonic risk-adjusted returns could be conducted by regressing the portfolio 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 the portfolio return on a constant and the factor of interest, and possibly other “control” factors, and then testing that the coefficient on the factor of interest is monotonically increasing or decreasing.

In this general case, the bootstrap test is obtained by estimating the regression on the bootstrapped data

$$\hat{r}_{i\tau(t)}^{(b)} = \boldsymbol{\beta}_i^{(b)'} \mathbf{F}_{i\tau(t)}^{(b)} + e_{i\tau(t)}^{(b)}, \quad i = 1, 2, \dots, N. \quad (21)$$

Note that the explanatory variables in this regression are also shuffled using the same time index as the portfolio returns. For each bootstrap sample an estimate of the coefficient vector is obtained. This is re-centered by subtracting the corresponding estimate from the original data. From the re-centered bootstrapped estimates, $\hat{\beta}_i^{(b)}$, a vector of the estimated differences in the parameter of interest (j) is computed:

$$\begin{aligned} \Delta \hat{\boldsymbol{\beta}}_j^{(b)} &\equiv \left[\Delta \hat{\beta}_{j2}^{(b)}, \dots, \Delta \hat{\beta}_{jN}^{(b)} \right]' \\ \text{where } \Delta \hat{\beta}_{jk}^{(b)} &\equiv \hat{\beta}_{jk}^{(b)} - \hat{\beta}_{j,k-1}^{(b)}. \end{aligned} \quad (22)$$

Finally, imposing the null hypothesis that the parameter of interest is constant across portfolios, a test statistic for the bootstrap sample can be computed:

$$J_{j,T}^{(b)} \equiv \min_{k=2, \dots, N} \Delta \hat{\beta}_{jk}^{(b)} \quad (23)$$

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 1. Corollary 2.6 of White (2000) justifies the above approach as the parameter estimates in this case are differentiable functions of averages of random variables.

If estimates of β are available, alternatively the test proposed by Gibbons, Ross and Shanken (1989) can be used. This is a multivariate test of whether the $N \times 1$ vector of intercepts are jointly zero against a general non-zero alternative and thus does not use any ranking information. If the ranking information is correct, our test, by directing power towards the ranked portfolios is likely to be more powerful, while it will be less so if deviations from the asset pricing model take a different form. If applied in an asset pricing context, our approach could be used to test for a monotonic pattern in the alphas or betas of a set of ranked portfolios. In both cases, inevitably the results will be joint tests of the asset pricing model and the particular specification of the β -model. If the beta model is misspecified, such tests may wrongly reject.

3.3.2 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 et al. (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 term structure being upward-sloping during economic recessions) but could take other forms as well. The first step of a conditional version of our test would consist of pre-multiplying 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 $\tilde{r}_{it} = r_{it} \otimes z_t$ along the lines proposed above.

4 Conclusion

Economic theory often predicts the existence of a monotonic relation between variables such as book-to-market ratio, firm size or past returns on the one hand and expected returns on the other hand. This is a natural hypothesis to consider in cases where the variables used to rank stocks are believed to capture a risk or liquidity premium. If the theory is correct in these cases, stocks in the portfolio with the lowest value of the variable capturing risk or illiquidity should also earn the smallest expected return. As we move to portfolios comprising firms that are riskier or more illiquid, the expected return should increase monotonically, though not necessarily linearly.

In this paper we propose a test that reveals whether a null hypothesis of no systematic relationship between expected returns and some “sorting” variable can be rejected in favor of a monotonic relationship predicted by economic theory. Our test summarizes in a single number whether the relationship is monotonic or not. Moreover, it is non-parametric and does not require making any assumptions about the functional form of the relationship between the sorting variables and expected returns. This is a big advantage since monotonicity of the relationship between portfolio rank and expected returns is preserved under very general conditions, including non-linear mappings between sorting variables and risk factor loadings. Perhaps most importantly, our test is extremely easy to use and so it is our hope that it will find widespread use.

The traditional pair-wise t -test and the proposed monotonic relationship test inspect very different implications of the theoretical proposition of a relationship between portfolio ranks and expected returns. Moreover, as we have shown, which test is adopted matters substantially in practice, particularly for stocks sorted on short-term reversal, momentum or size. If interest is limited to establishing a trading strategy and it is possible to short the lowest ranked stocks, then a comparison between the expected returns on the top and bottom portfolios—as is done by the t -test—may suffice. However, when considering the implications of theories that rank stocks based on variables proxying for exposure to risk or liquidity, the complete cross-sectional pattern in expected returns should be considered and our test is well-suited for this purpose.

We see two principal uses for our new test. First, it can be used as a descriptive statistic for monotonicity in the expected returns of stocks grouped into portfolios according to one or more sorting variables. Besides providing a single summary statistic for monotonicity, our approach allows researchers to decompose the results to see, e.g., if the book-to-market ratio effect is monotonic for all size-sorted portfolios or, conversely, if size effects are present for both value and growth stocks. Second, if theoretical considerations suggest a monotonic relationship between the sorting variable and the loadings on possibly unobserved risk factors, then our approach can be used to formally test asset pricing implications in a flexible manner. Lack of monotonicity may imply that the conjectured theoretical model is not an adequate description of the data.

Appendix: Proof of Proposition 1.

The first part of proposition 1—convergence of the vector of sample means to a multivariate normal distribution under the stated moment and mixing conditions—follows from Theorem 3 in Politis and Romano (1994). As shown in Proposition 2.2 in White (2000), since the maximum of the vector of sample means is a continuous function of the elements of the vector, by the continuous mapping theorem we have that

$$\max_{i=2,\dots,N} \sqrt{T} \{(\hat{\mu}_{iT} - \hat{\mu}_{i-1,T}) - (\mu_i - \mu_{i-1})\} \Rightarrow W \equiv \max_{i=2,\dots,N} \{Z_i - Z_{i-1}\},$$

where $[Z_1, \dots, Z_N]' \sim N(\mathbf{0}, \mathbf{\Omega})$. The final part of Proposition 1, which justifies use of the bootstrap, follows from Corollary 2.4 in White (2000), noting that we do not need any further assumptions than stated in the proposition due to the fact that there are no estimated parameters here. ■

References

- [1] Ang, A. and J. Chen, 2007, The CAPM over the Long Run: 1926-2001. *Journal of Empirical Finance* 14, 1-40.
- [2] Ang, A., J. Chen and Y. Xing, 2006, Downside Risk. *Review of Financial Studies* 19, 1191-1239.
- [3] Ang, A., R.J. Hodrick, Y. Xing and X. Zhang, 2006, The Cross-section of Volatility and Expected Returns. *Journal of Finance* 51, 259-299.
- [4] Arnott, R., J. Hsu, J. Liu and H. Markowitz, 2007. Does Noise Create the Size and Value Effects? Manuscript, University of California, San Diego.
- [5] Banz, R., 1981, The Relationship Between Returns and Market Value of Common Stocks. *Journal of Financial Economics* 9, 3-18.
- [6] Bartholomew, D.J., 1961, A Test of Homogeneity of Means under Restricted Alternatives. *Journal of the Royal Statistical Society B* 23, 239-281.
- [7] Basu, S. 1977, Investment Performance of Common Stocks in Relation to their Price Earnings Ratios: A Test of the Efficient Market Hypothesis. *Journal of Finance* 32, 663-682.
- [8] Basu, S., 1983, The Relationship between Earnings Yield, Market Value and Return for NYSE Common Stocks: Further Evidence. *Journal of Financial Economics* 12, 129-156.
- [9] Berk, J.B., R. C. Green and V. Naik, 1999, Optimal Investment, Growth Options, and Security Returns. *Journal of Finance* 54, 1553-1607.
- [10] Black, F., M.C. Jensen and M. Scholes, 1972, The Capital Asset Pricing Model: Some Empirical Tests. In Michael C. Jensen (ed.): *Studies in the Theory of Capital Markets*. Praeger Publishers, New York.
- [11] Boudoukh, J., M. Richardson, T. Smith and R.F. Whitelaw, 1999, Ex Ante Bond Returns and the Liquidity Preference Hypothesis. *Journal of Finance* 54, 1153-1167.
- [12] Campbell, J.Y., A.W. Lo and A.C. MacKinlay, 1997, *The Econometrics of Financial Markets*. Princeton University Press: New Jersey.
- [13] Carhart, M., 1997, On Persistence in Mutual Fund Performance, *Journal of Finance* 52, 57-82.
- [14] Daniel, K. and S. Titman, 1997, Evidence on the Characteristics of Cross-sectional Variation in Stock Returns. *Journal of Finance* 52, 1-33.
- [15] Daniel, K., M. Grinblatt, S. Titman and R. Wermers, 1997, Measuring Mutual Fund Performance with Characteristic Based Benchmarks. *Journal of Finance* 52 1035-1058.
- [16] Fama, E.F. and K.R. French, 1992, The Cross-Section of Expected Stock Returns. *Journal of Finance* 47, 427-465.
- [17] Fama, E.F. and K.R. French, 2006, The Value Premium and the CAPM. *Journal of Finance* 61, 2163-2186.

- [18] Gibbons, M., S. Ross and J. Shanken, 1989, A Test of the Efficiency of a Given Portfolio. *Econometrica* 57, 1121-1152.
- [19] Gomes, J., L. Kogan and L. Zhang, 2003, Equilibrium Cross Section of Returns. *Journal of Political Economy* 111, 693-732.
- [20] Gourieroux, C., A. Holly and A. Monfort, 1982, Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters. *Econometrica* 50, 63-80.
- [21] Hansen, P.R., 2005, A Test for Superior Predictive Ability. *Journal of Business and Economic Statistics* 23, 365-380.
- [22] Harvey, C., and A., Siddique, 2000, Conditional Skewness in Asset Pricing Tests, *Journal of Finance* 55, 1263-1295.
- [23] Jegadeesh, N. and S. Titman, 1993, Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance* 48, 65-92.
- [24] Kodde, D.A. and F.C. Palm, 1986, Wald Criteria for Jointly Testing Equality and Inequality Restrictions. *Econometrica* 54, 1243-1248.
- [25] Kosowski, R., A. Timmermann, R. Wermers and H. White, 2006, Can Mutual Fund “Stars” Really Pick Stocks? New Evidence from a Bootstrap Analysis. *Journal of Finance* 61, 2551-2595.
- [26] Kudo, A., 1963, A Multivariate Analogue of the One-sided Test. *Biometrika* 50, 403-418.
- [27] Lamont, O., C. Polk and J. Saa-Requejo, 2001, Financial Constraints and Stock Returns. *Review of Financial Studies* 14, 529-554.
- [28] Lee, C.M.C. and B. Swaminathan, 2000, Price Momentum and Trading Volume. *Journal of Finance* 55, 2017-2069.
- [29] Li, D. , 2007, Financial Constraints, R&D Investment and Stock Returns: Theory and Evidence. Mimeo, Wharton School.
- [30] Loughran, T., 1997, Book-to-Market Across Firm Size, Exchange, and Seasonality. *Journal of Financial and Quantitative Analysis* 32, 249-268.
- [31] Newey, W. and K.D. West, 1987, A Simple, Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55, 703-708.
- [32] Nielsen, A.E., 2007, Corporate Governance, Capital Structure Choice and Equity Prices. Mimeo, Princeton University.
- [33] Pastor, L. and R.F. Stambaugh, 2003, Liquidity Risk and Expected Stock Returns. *Journal of Political Economics* 111, 642-685.
- [34] Politis, D.N. and J.P. Romano, 1994, The Stationary Bootstrap. *Journal of the American Statistical Association* 89, 1303-1313.

- [35] Reinganum, M., 1981, Misspecification of Capital Asset Pricing: Empirical Anomalies Based on Earnings Yields and Market Values. *Journal of Financial Economics* 9, 19-46.
- [36] Richardson, M., P. Richardson and T. Smith, 1992, The Monotonicity of the Term Premium: Another Look. *Journal of Financial Economics* 31, 97-106.
- [37] Romano, J.P. and M. Wolf, 2005, Stepwise Multiple Testing as Formalized Data Snooping. *Econometrica* 73, 1237-1282.
- [38] Rouwenhorst, K.G., 1998, International Momentum Strategies. *Journal of Finance* 53, 267-284.
- [39] Sullivan, R., A. Timmermann and H. White, 1999, Data-snooping, Technical Trading Rules and the Bootstrap. *Journal of Finance* 54, 1647-1692.
- [40] Vassalou, M. and Y. Xing, 2004, Default Risk in Equity Returns. *Journal of Finance* 59, 831-868.
- [41] White, H., 2000, A Reality Check for Data Snooping. *Econometrica* 68, 1097-1126.
- [42] 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.
- [43] Wolak, F.A., 1989a, Testing Inequality Constraints in Linear Econometric Models. *Journal of Econometrics* 31, 205-235.
- [44] Wolak, F.A., 1989b, Local and Global Testing of Linear and Nonlinear Inequality Constraints in Nonlinear Econometric Models. *Econometric Theory* 5, 1-35.
- [45] Yogo, M., 2006, A Consumption-Based Explanation of Expected Stock Returns. *Journal of Finance* 61, 539-580
- [46] Zhang, L., 2005, The Value Premium. *Journal of Finance* 60, 67-103.

Table 1: 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, 1926:07), book-to-market value (BE-ME, 1926:07), cash flow-price (CF-P, 1951:07), earnings-price (E-P, 1951:07), dividend-price (D-P, 1927:07), momentum (1927:01), short term reversal (ST reversal, 1926:02) and long term reversal (LT reversal, 1931:01). The start of the sample period is listed in bracket after the sorting variable. All data series end in 2006:12 and are taken from Ken French's web site at Dartmouth. In Panel B the results for the period 1963:07 - 2006:12 are reported. The order of the portfolios has been normalized so we would expect the mean return to decline as we move down from the top to the bottom ranked portfolio. In particular, the order of the portfolios sorted on the book-to-market, cash flow-price, earnings-price or dividend-price ratios and momentum has been reversed.

	ME	BE-ME	CF-P	E-P	D-P	Momentum	ST reversal	LT reversal
A: Full sample								
top	1.5201	1.4095	1.4369	1.5487	1.0945	1.5911	1.4935	1.5042
2	1.3261	1.3064	1.3973	1.4027	1.1089	1.2681	1.2167	1.3154
3	1.2981	1.2719	1.2120	1.3440	1.1446	1.1634	1.1679	1.2865
4	1.2487	1.1147	1.1888	1.2647	1.0865	1.0322	1.0387	1.0937
5	1.2111	1.1006	1.0862	1.1897	0.9908	0.9435	1.0642	1.1643
6	1.1828	1.0535	1.1162	1.0197	0.9019	0.8689	1.0167	1.0440
7	1.1476	0.9703	1.0030	1.0100	1.0239	0.8650	0.9721	1.0715
8	1.0926	0.9769	0.9863	1.0100	0.9338	0.7412	0.9322	1.0472
9	1.0401	0.9728	0.9392	0.8641	0.9843	0.7316	0.8285	0.9222
bottom	0.9066	0.8715	0.8630	0.8516	0.9239	0.3376	0.4869	0.9023
B: Data from 1963:07 – 2006:12								
top	1.2728	1.3962	1.3343	1.4289	1.0737	1.6482	1.1473	1.3949
2	1.2058	1.2695	1.3289	1.2838	1.1220	1.2376	1.2813	1.2439
3	1.2414	1.2163	1.1268	1.2290	1.1328	1.1449	1.2551	1.2141
4	1.1853	1.1877	1.0896	1.2337	1.0452	0.9414	1.0550	1.1042
5	1.2090	1.1098	1.0297	1.0751	1.0056	0.8992	1.0197	1.1273
6	1.0975	1.0136	1.0685	0.9426	0.9124	0.8007	0.9354	1.0739
7	1.1511	1.0134	0.9600	0.9593	1.0063	0.9028	0.8901	1.0701
8	1.0953	0.9894	0.9746	0.9745	1.0293	0.8624	0.9406	0.9775
9	1.0260	0.9476	0.8976	0.8448	0.9377	0.7357	0.7477	0.8970
bottom	0.8863	0.8245	0.8497	0.8292	1.0021	0.1766	0.6834	0.8824

Table 2: Test statistics for portfolios sorted on a single variable

For each set of decile portfolios sorted on the variables listed in the rows, 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-West heteroskedasticity and autocorrelation consistent standard errors), while column 3 shows the associated p -value. Columns 4 and 5 report the p -values from a test based on Bonferroni bounds and from Wolak's (1989a) test. Columns 6 and 7 present the p -values from the monotonic relationship (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. The sorting variables are market equity (ME, 1926:07), book-to-market value (BE-ME, 1926:07), cash flow-price (CF-P, 1951:07), earnings-price (E-P, 1951:07), dividend-price (D-P, 1927:07), momentum (1927:01), short term reversal (ST reversal, 1926:02) and long term reversal (LT reversal, 1931:01). The start of the sample period is listed in bracket after the sorting variable. All data series are taken from Ken French's web site at Dartmouth and are based on value-weighted portfolios with mean returns reported in percent. Panel A shows results using all available data on each series, and Panel B shows the results for the 1963:07 - 2006:12 sub-sample.

	top minus bottom	t-test		Wolak	Bonf.	MR test	MR ^{all} test
		t -stat	p -value				
A: Full sample							
ME	0.6135	2.3495	0.0094	0.9980	1.0000	0.0020	0.0020
BE-ME	0.5380	2.4394	0.0074	0.9984	1.0000	0.0000	0.0000
CF-P	0.5739	3.2442	0.0006	0.9941	1.0000	0.0180	0.0120
E-P	0.6971	3.6645	0.0001	1.0000	1.0000	0.0000	0.0000
D-P	0.1705	0.9385	0.1740	0.5281	0.3607	0.6920	0.4920
Momentum	1.2536	5.4340	0.0000	0.9990	1.0000	0.0020	0.0020
ST reversal	1.0066	5.0601	0.0000	0.9952	1.0000	0.0120	0.0160
LT reversal	0.6019	2.4858	0.0065	0.8813	1.0000	0.2210	0.1580
B: Data from 1963:07 – 2006:12							
ME	0.3865	1.5358	0.0623	0.7191	1.0000	0.2740	0.2370
BE-ME	0.5717	2.5437	0.0055	1.0000	1.0000	0.0000	0.0000
CF-P	0.4847	2.4037	0.0081	0.9884	1.0000	0.0240	0.0120
E-P	0.5997	2.6832	0.0036	0.9907	1.0000	0.0080	0.0210
D-P	0.0716	0.2947	0.3841	0.8185	1.0000	0.3360	0.2560
Momentum	1.4715	5.6709	0.0000	0.8679	1.0000	0.2910	0.2420
ST reversal	0.4639	2.3635	0.0091	0.8723	1.0000	0.2580	0.1700
LT reversal	0.5125	2.2052	0.0137	0.9971	1.0000	0.0020	0.0020

Table 3: Estimates of expected returns for 5×5 double-sorted portfolios

This table shows mean returns for stock portfolios using 5 × 5 two-way sorts. The sorting variables are market equity (ME), which is always listed in the rows and one of either book-to-market value (BE/ME, 1926:07), momentum (1927:01), short term reversal (ST reversal, 1926:02) or long term reversal (LT reversal, 1963:01). The start of the sample period is listed in bracket after the sorting variable. All data series end in 2006:12 and are taken from Ken French’s web site at Dartmouth. Portfolios are value-weighted and mean returns are reported in percent per month. The order of the portfolios has been normalized so the hypothesis is that the mean return declines as we move down the rows or along the columns from the top towards the bottom ranked portfolios.

	top	2	3	4	bottom
A: ME × BE/ME					
small	1.8037	1.5757	1.4162	1.1685	0.8729
2	1.5878	1.4859	1.4012	1.2963	0.9307
3	1.5051	1.3630	1.3260	1.2599	1.0160
4	1.4558	1.3365	1.2486	1.0806	1.0245
big	1.2870	1.1042	1.0575	0.9623	0.9381
B: ME × momentum					
small	2.0131	1.7895	1.6691	1.4778	0.9662
2	1.7836	1.5059	1.2809	1.2007	0.6051
3	1.6617	1.2584	1.1311	1.0031	0.6077
4	1.6114	1.2120	1.0271	0.8833	0.6554
big	1.2506	1.0257	0.8418	0.7614	0.6546
C: ME × ST reversal					
small	2.5712	1.6264	1.4189	1.0423	0.1296
2	1.9531	1.5322	1.3367	1.0480	0.4129
3	1.7393	1.4035	1.2858	0.9786	0.5285
4	1.4821	1.3214	1.1882	0.9696	0.6890
big	1.1104	1.0332	0.9648	0.9421	0.7279
D: ME × LT reversal					
small	1.5748	1.3528	1.4400	1.2901	0.8732
2	1.3397	1.2483	1.2819	1.2258	1.0330
3	1.2282	1.1987	1.0905	1.1918	0.9974
4	1.2021	1.0750	1.0230	1.0645	0.9580
big	1.0484	0.9889	0.8924	0.8717	0.7575

Table 4: Conditional and joint monotonicity tests for double-sorted portfolios

This table shows mean returns for stock portfolios using 5×5 two-way sorts. The sorting variables are market equity (ME), which is always listed in the row, and one of either book-to-market value (BE/ME, 1926:07), momentum (1927:01), short term reversal (ST reversal, 1926:02) or long term reversal (LT reversal, 1963:01). The start of the sample period is listed in bracket after the sorting variable. All data series end in 2006:12 and are taken from Ken French’s web site at Dartmouth. Portfolios are value-weighted and mean returns are reported in percent per month. The order of the portfolios has been normalized so the hypothesis is that the expected return declines as we move down the rows or along the columns from the top towards the bottom ranked portfolios.

The penultimate row in each panel presents the p -values from tests for a monotonic relationship (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 top book-to-market portfolio is 0.002, while it is 0.431 if we condition on being in the bottom book-to-market portfolio. The final row in each panel presents the p -value from a joint test for a monotonic relationship between the portfolio sorting variable in the row, across all column portfolios. The penultimate and final columns present p -values from tests for a monotonic relationship 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 relationship in both variables.

	top	2	3	4	bottom	MR p -value	Joint MR p -value
A: Market equity \times Book-to-market ratio							
Market equity	Book-to-market ratio						
small	1.8037	1.5757	1.4162	1.1685	0.8729	0.000	
2	1.5878	1.4859	1.4012	1.2963	0.9307	0.000	
3	1.5051	1.3630	1.3260	1.2599	1.0160	0.005	0.000
4	1.4558	1.3365	1.2486	1.0806	1.0245	0.003	
big	1.2870	1.1042	1.0575	0.9623	0.9381	0.015	
MR p -value	0.002	0.013	0.059	0.383	0.431		
Joint MR p -value			0.0870				0.005

Table 4: Conditional and joint monotonicity tests for double-sorted portfolios

Continued from previous page.

	top	2	3	4	bottom	MR <i>p</i> -value	Joint MR <i>p</i> -value
B: Market value \times momentum							
Market equity	Momentum						
small	2.0131	1.7895	1.6691	1.4778	0.9662	0.000	
2	1.7836	1.5059	1.2809	1.2007	0.6051	0.000	
3	1.6617	1.2584	1.1311	1.0031	0.6077	0.000	0.000
4	1.6114	1.2120	1.0271	0.8833	0.6554	0.000	
big	1.2506	1.0257	0.8418	0.7614	0.6546	0.002	
MR <i>p</i> -value	0.003	0.022	0.000	0.007	0.241		
Joint MR <i>p</i> -value			0.045				0.001
C: Market equity \times short-term reversal							
Market equity	Short-term reversal						
small	2.5712	1.6264	1.4189	1.0423	0.1296	0.000	
2	1.9531	1.5322	1.3367	1.0480	0.4129	0.000	
3	1.7393	1.4035	1.2858	0.9786	0.5285	0.000	0.000
4	1.4821	1.3214	1.1882	0.9696	0.6890	0.000	
big	1.1104	1.0332	0.9648	0.9421	0.7279	0.037	
MR <i>p</i> -value	0.000	0.024	0.023	0.098	0.939		
Joint MR <i>p</i> -value			0.890				0.823
D: Market equity \times long-term reversal							
Market equity	Long-term reversal						
small	1.5748	1.3528	1.4400	1.2901	0.8732	0.459	
2	1.3397	1.2483	1.2819	1.2258	1.0330	0.177	
3	1.2282	1.1987	1.0905	1.1918	0.9974	0.695	0.232
4	1.2021	1.0750	1.0230	1.0645	0.9580	0.329	
big	1.0484	0.9889	0.8924	0.8717	0.7575	0.010	
MR <i>p</i> -value	0.029	0.002	0.002	0.016	0.697		
Joint MR <i>p</i> -value			0.291				0.100

Table 5: Monte Carlo simulation results

Using 2500 Monte Carlo simulations and 1000 bootstraps for each simulation, this table reports rejection frequencies for simulated data repeatedly drawn from the value-weighted decile portfolios sorted on market equity. 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 relationship (MR) test, respectively. These results are obtained under the normality assumption discussed in Section 3. Columns 3 and 6 show the corresponding proportions of rejections using bootstrapped portfolio data based on the value-weighted decile portfolios sorted on market equity. Columns 4 and 5 reports the proportion of rejections of the null of a weakly monotonic relationship using a test based on Bonferroni bounds and from Wolak’s (1989a) test, with critical values for Wolak’s test based on 1000 simulations per replication. The last column presents the proportion of rejections using the MR test based on all possible inequalities implied by monotonicity, rather than the minimal set considered in the base-case MR test. “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, while experiments IV-VI impose a non-monotonic pattern. Figure 2 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.

Table 5 is on the following page.

Table 5: Monte Carlo simulation results.*See description on the previous page.*

Experiment	Normal simulation		Bootstrap simulation				
	t-test rej%	MR rej%	t-test rej%	Bonf. rej%	Wolak rej%	MR rej%	MR ^{all} rej%
Step=0							
I	0.0500	0.0500	0.0736	0.0340	0.0356	0.0644	0.0902
Step=0.01							
I	0.1263	0.0992	0.1344	0.0212	0.0164	0.1144	0.1361
II	0.1263	0.0956	0.1344	0.0228	0.0208	0.1000	0.1500
III	0.1263	0.0926	0.1344	0.0252	0.0220	0.0992	0.1205
IV	0.0500	0.0452	0.0736	0.0460	0.0508	0.0536	0.0598
V	0.0560	0.0483	0.0820	0.0440	0.0456	0.0564	0.0631
VI	0.0949	0.0671	0.1120	0.0368	0.0380	0.0708	0.0926
Step=0.02							
I	0.2600	0.1708	0.2208	0.0104	0.0064	0.1764	0.2262
II	0.2600	0.1562	0.2208	0.0180	0.0116	0.1504	0.1943
III	0.2600	0.1518	0.2208	0.0172	0.0112	0.1420	0.1689
IV	0.0500	0.0349	0.0736	0.0632	0.0832	0.0356	0.0484
V	0.0626	0.0416	0.0868	0.0616	0.0732	0.0432	0.0533
VI	0.1642	0.0639	0.1660	0.0584	0.0552	0.0612	0.0779
Step=0.05							
I	0.8047	0.4792	0.5936	0.0008	0.0008	0.4680	0.5148
II	0.8047	0.3815	0.5936	0.0064	0.0020	0.3024	0.3221
III	0.8047	0.3765	0.5936	0.0048	0.0016	0.3096	0.3918
IV	0.0500	0.0067	0.0736	0.2216	0.3744	0.0032	0.0057
V	0.0859	0.0092	0.1044	0.2196	0.3524	0.0044	0.0057
VI	0.5096	0.0144	0.3876	0.2004	0.2132	0.0108	0.0189
Step=0.10							
I	0.9996	0.8472	0.9660	0.0000	0.0000	0.8368	0.8033
II	0.9996	0.6810	0.9660	0.0020	0.0012	0.5712	0.5959
III	0.9996	0.7294	0.9660	0.0000	0.0000	0.6484	0.7025
IV	0.0500	0.0001	0.0736	0.8140	0.9600	0.0000	0.0000
V	0.1382	0.0000	0.1460	0.8140	0.9584	0.0000	0.0000
VI	0.9548	0.0000	0.7680	0.7292	0.8060	0.0000	0.0000

Table 6: Monte Carlo simulation for varying numbers of portfolios

This table reports the proportion of rejections from a t -test applied to the spread in expected returns between the top and bottom portfolios (columns 1-4) and from the monotonic relationship (MR) test (columns 5-8), for varying numbers of portfolios, N . The results for $N = 10$ are identical to the results in columns 3-4 of Table 5 and are presented here for ease of reference. The case with two portfolios is formed by aggregating the returns on portfolios 1-5 and 6-10; the case with three portfolios aggregates portfolios 1-3, 4-7 and 8-10; the case with five portfolios aggregates portfolios 1-2, 3-4, 5-6, 7-8 and 9-10. These results are obtained under the normality assumption discussed in Section 3. “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, while experiments IV-VI impose a non-monotonic pattern. Figure 3 displays the shapes of experiments IV-VI for the various values of N ; the shapes of experiments I-III across different values of N are easily inferred from Figure 2. The experiments are identical when the step size is zero and so only the results for Experiment I are presented in that case.

Table 6 is on the following page.

Table 6: Monte Carlo simulation for varying numbers of portfolios*See description on the previous page.*

N	t-test				MR test			
	2	3	5	10	2	3	5	10
step=0								
I	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500
step=0.01								
I	0.1333	0.1352	0.1339	0.1263	0.1342	0.1293	0.1193	0.0992
II	0.1266	0.1327	0.1328	0.1263	0.1272	0.1150	0.1091	0.0956
III	0.1077	0.1115	0.1158	0.1263	0.1099	0.1057	0.1005	0.0926
IV	0.0500	0.0500	0.0500	0.0500	0.0511	0.0378	0.0404	0.0452
V	0.0620	0.0585	0.0573	0.0560	0.0650	0.0464	0.0471	0.0483
VI	0.0892	0.0910	0.0843	0.0949	0.0905	0.0888	0.0799	0.0671
step=0.02								
I	0.2821	0.2880	0.2837	0.2600	0.2841	0.2704	0.2348	0.1708
II	0.2608	0.2801	0.2803	0.2600	0.2661	0.2117	0.1920	0.1562
III	0.2024	0.2142	0.2273	0.2600	0.2042	0.1900	0.1781	0.1518
IV	0.0500	0.0500	0.0500	0.0500	0.0510	0.0262	0.0272	0.0349
V	0.0762	0.0681	0.0654	0.0626	0.0771	0.0357	0.0341	0.0416
VI	0.1476	0.1528	0.1339	0.1642	0.1487	0.1475	0.1206	0.0639
step=0.05								
I	0.8474	0.8575	0.8503	0.8047	0.8490	0.8030	0.6986	0.4792
II	0.8065	0.8440	0.8444	0.8047	0.8098	0.5676	0.4828	0.3815
III	0.6497	0.6870	0.7249	0.8047	0.6528	0.6000	0.5221	0.3765
IV	0.0500	0.0500	0.0500	0.0500	0.0512	0.0026	0.0026	0.0067
V	0.1333	0.1043	0.0952	0.0859	0.1333	0.0072	0.0049	0.0092
VI	0.4406	0.4625	0.3807	0.5096	0.4451	0.4281	0.3079	0.0144
step=0.10								
I	0.9999	0.9999	0.9999	0.9996	0.9999	0.9985	0.9802	0.8472
II	0.9996	0.9999	0.9999	0.9996	0.9996	0.9190	0.7951	0.6810
III	0.9921	0.9956	0.9977	0.9996	0.9921	0.9764	0.9260	0.7294
IV	0.0500	0.0500	0.0500	0.0500	0.0508	0.0000	0.0000	0.0001
V	0.2821	0.1923	0.1650	0.1382	0.2854	0.0001	0.0000	0.0000
VI	0.9108	0.9274	0.8503	0.9548	0.9134	0.8823	0.6978	0.0000

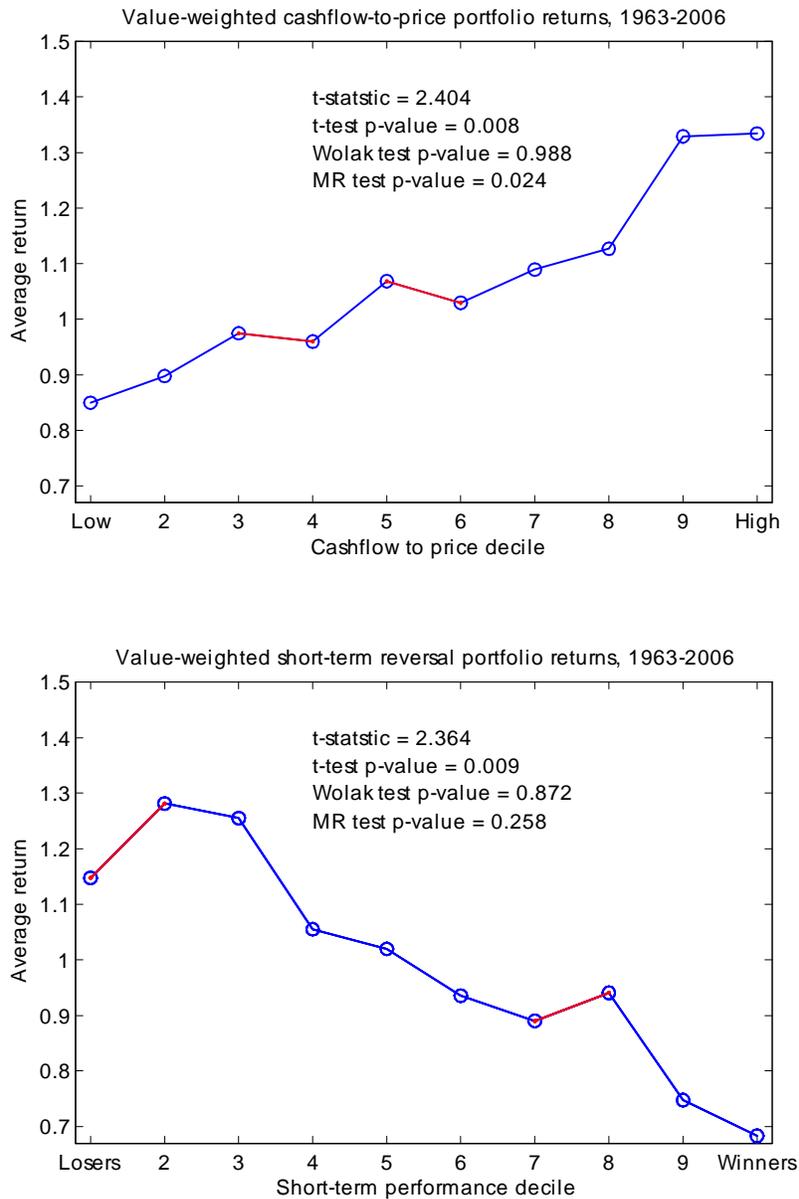


Figure 1: Average monthly returns on the cashflow-to-price (upper graph) and short-term reversal (lower graph) decile portfolios, from July 1963 to December 2006. Also presented are the test statistics and associated p-values from t-tests on these two data sets, the p-value from Wolak's (1989a) test, as well as the p-value from the proposed 'MR' test.

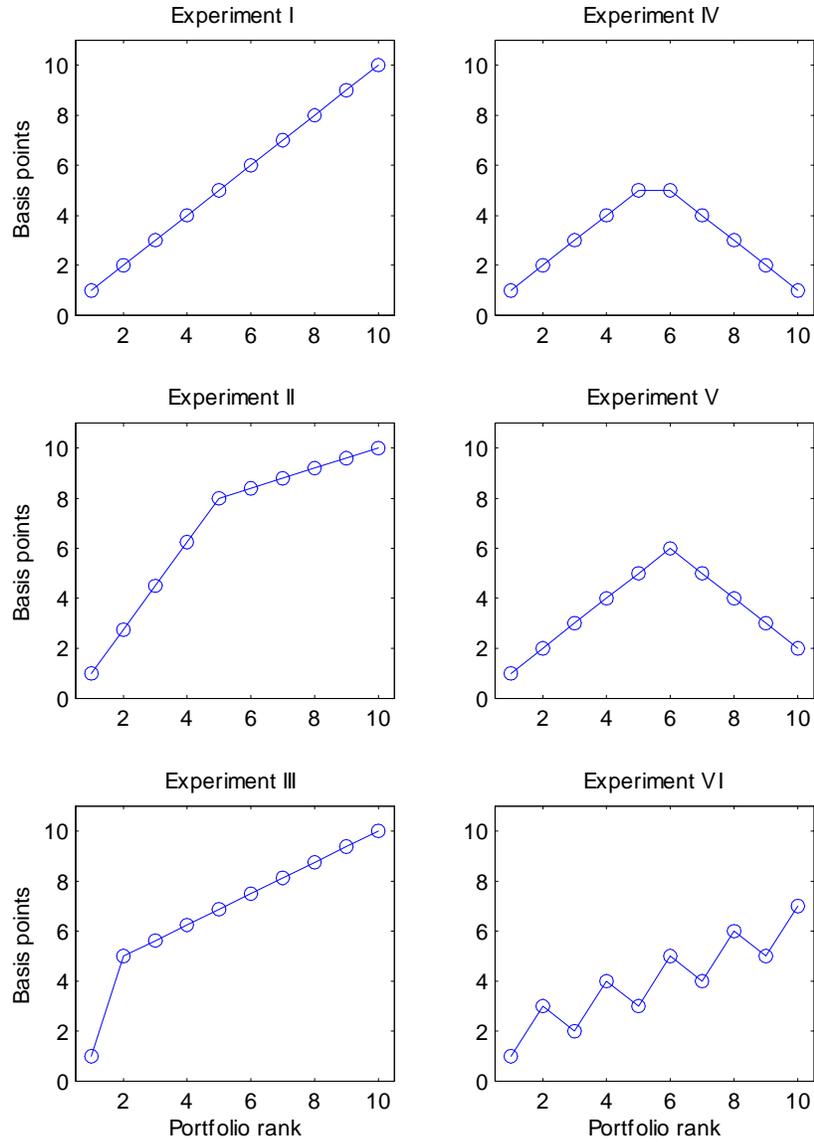


Figure 2: Patterns in expected returns under the six experiments considered in the Monte Carlo simulations for a step size of 1 basis point.

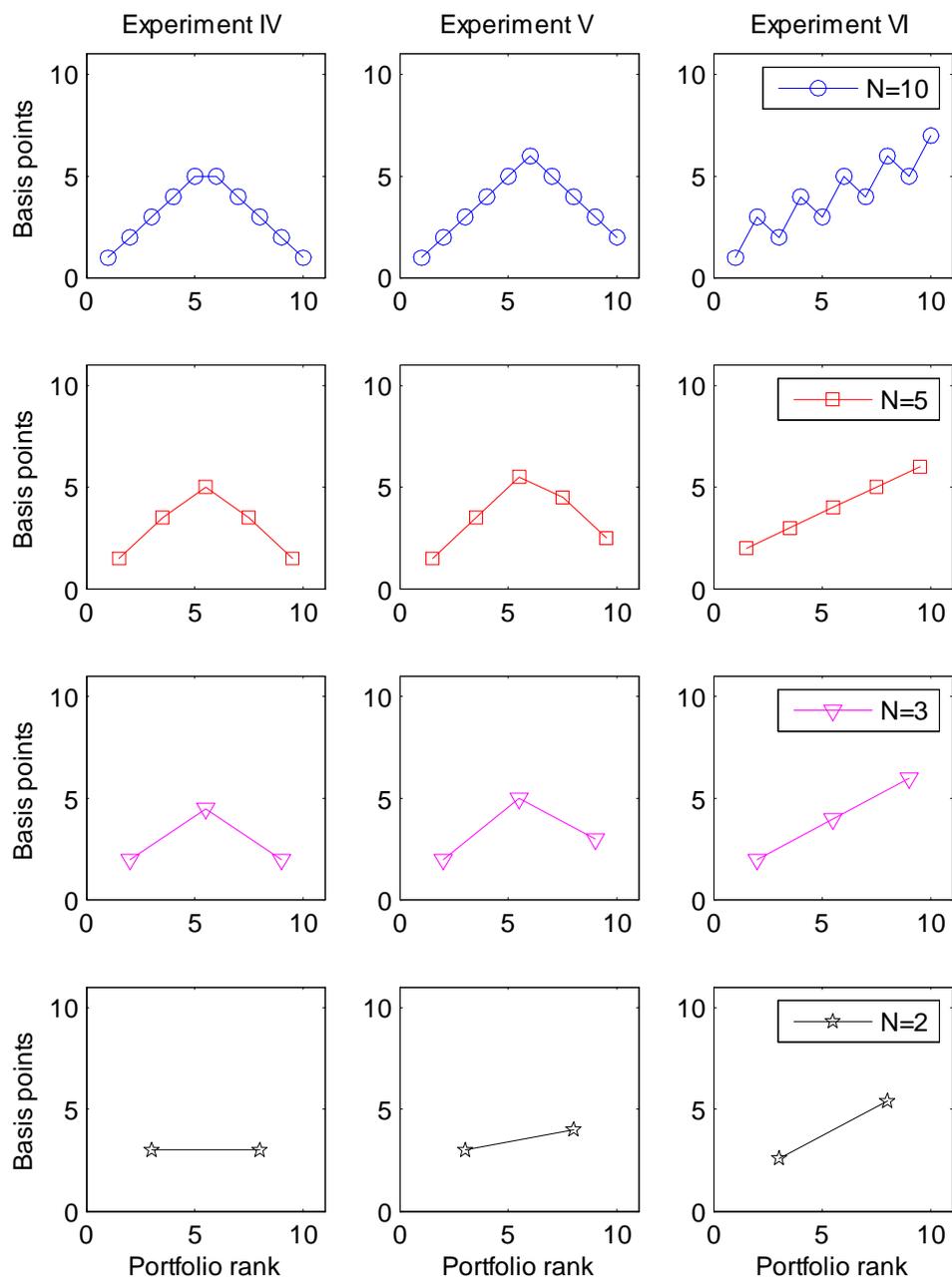


Figure 3: Patterns in expected returns considered in the Monte Carlo simulations, with a step size of 1 basis point, under experiments IV-VI, with the number of portfolios, N , equal to 10, 5, 3 and 2.