VALUE VERSUS GLAMOUR

Jennifer Conrad, Michael Cooper, and Gautam Kaul*

VALUE VERSUS GLAMOUR

ABSTRACT

The fragility of the CAPM has led to a resurgence of research that frequently uses trading strategies based on sorting procedures to uncover relations between firm characteristics (such as "value" or "glamour") and equity returns. We examine the propensity of these strategies to generate statistically and economically significant profits due to our familiarity with the data. Under plausible assumptions, data-snooping can account for up to 50 percent of the in-sample relations between firm characteristics and returns uncovered using single (one-way) sorts. The biases can be much larger if we simultaneously condition returns on two (or more) characteristics.

The debate around the empirical support for the one-factor Capital Asset Pricing Model (CAPM) in explaining the cross-section of expected returns of financial securities has led to resurgence of empirical research aimed at discovering variables that might better explain the behavior of returns. This research has, at least in part, been given theoretical validity by both the inter-temporal version of the CAPM [see, for example, Merton (1973)] and the Arbitrage Pricing Theory [see Connor (1984) and Ross (1976)]. However, with little guidance from theory on the identity of the "factors" that determine returns, for the most part the goal of this literature is the discovery of multiple characteristics, such as "value" versus "glamour," that are statistically related to asset returns.

The resurgence of this literature started about two decades ago with tests by Banz (1981) and Reinganum (1981) that use firm size, in addition to a firm's beta, to explain the cross-section of required returns on equity. More recently, a series of papers by Fama and French (1992, 1993, 1995, 1996, 1998) on value versus glamour stocks has made this issue a central focus of the profession. These papers, and numerous others, present evidence that multiple relative characteristics of value versus glamour stocks can explain large fractions of the variability in asset returns.[1]

Given this backdrop, a good question to ask is: To what extent are the findings on value versus glamour characteristics subject to data-snooping biases? The purpose of this paper is to gauge the impact of data-snooping on empirical findings that are based on a specific and commonly used sorting methodology to uncover relations between equity returns and multiple firm characteristics. Many studies sort variables into portfolios based on specific firm characteristics and report the subsequent period's returns/profits deriving from trading strategies that are long and short in some subset of these portfolios. Most significantly, based on the belief

that asset returns are determined by multiple factors, there is a uniformly increasing tendency for researchers to simultaneously sort returns on more than a single firm characteristic.[2] In fact, value and glamour have both come to signify multiple attributes.

Inadvertent snooping is inherent to this literature because: (a) any new research endeavor is conditioned on the collective knowledge built up to that point;[3] (b) there is no explicit guidance from theory regarding both the number and the identities of the characteristics related to average returns; and, consequently, (c) we are particularly susceptible to the bias of retaining the findings that "work" and discarding the ones that do not.

We choose 15 firm characteristics that have been successfully used in previous studies to uncover cross-sectional differences in returns. We measure the in-sample relations between these variables and returns measured over subsequent time periods, and interpret these relations to be a reflection of mispricing (rather than rewards for risk-taking). We conduct both one-way and two-way sorts on the 15 characteristics, recognizing that the analysis of the two-way sorts is more pertinent to the research findings presented in most studies. Since the chosen firm characteristics have already been shown to "work," we then attempt to gauge the effects of our collective snooping on the in-sample profitability of the trading strategies. We devise four measures of data-snooping that measure the proportion of "real" profits observed in-sample that can be generated by performing plausible number of searches over "random characteristics." The differences among the four measures hinge on the degree of familiarity that the researcher/trader is assumed to have with the data.

Our in-sample tests and simulation exercises collectively show that inadvertent snooping biases, perhaps based on our collective familiarity of the data, have the potential to explain a significant fraction, but not all, of the in-sample profits of strategies based on one-way sorts. Securities need to be sorted into an unrealistically large number of portfolios (50) for snooping

to explain more than 50 percent of the profits. Since most studies sort securities into 10 portfolios, we conclude that about 50 percent of the in-sample profits reported in single-sort studies are "real."

The evidence for the two-way sorts is more disturbing, however. Our simulation exercises suggest that potentially 80 percent to 100 percent of the profits to strategies based on this methodology can be explained by our prior familiarity with the data. The increase in the snooping bias is quite dramatic as the two-way sorts are increased from the 3x3 to 10x10 portfolios, all of which have been used in the literature. The trend toward simultaneous sorting of firms on multiple characteristics is understandable given the lack of support for a single-factor model. Our evidence suggests however that caution needs to be exercised in interpreting the predictive relations uncovered by such studies. The higher propensity for data snooping when a researcher moves from one-way to two-way (or even a three-way) sorts exists because it is much easier to generate a larger number of portfolios while simultaneously conditioning on multiple characteristics. We show that this tendency is exacerbated because multi-sort strategies that generate the highest profits are also ones that sort based on correlated firm characteristics.

We also evaluate the performance of typical out-of-sample tests used in empirical studies. Many out-of-sample tests amount to sub-period analyses because the entire data set is first used to show the in-sample effectiveness of particular firm characteristics, and then the same data is split up into smaller sets to conduct the "out-of-sample" tests. We find that the out-of-sample profitability of the trading strategies declines substantially; the remaining profits are statistically insignificant.

Another pattern in studies attempting to confirm the existing evidence "out-of-sample" is to append a significantly shorter time-series of data to the original data, and to use the resulting longer time-series to confirm the initial evidence. Virtually all studies that document the size

effect, for example, fall into this category. This "out-of-sample" experiment is also likely to be affected by any snooping bias that is present in the original results. To gauge the extent of such bias, we conduct a battery of tests that measure the rates of decay in the out-of-sample predictability of both the "real" and "simulated" firm characteristics. On average, the two decay rates are similar, again suggesting that snooping biases may have a non-trivial impact on our collective findings on the relations between equity returns and firm characteristics.

Section I explains the methodology used in the paper. Section II describes the data and presents the detailed evidence. Section III presents out-of-sample tests, and Section IV contains a brief summary and conclusion.

I. Methodology

Panel A of the Appendix lists the 15 commonly (and successfully) used value versus glamour firm characteristics, while Panel B contains a description of the COMPUSTAT and CRSP variables used to construct the characteristics. For both the in- and out-of-sample strategies, we use these firm characteristics to examine two types of trading rules popular in the literature: one-way and two-way sorts. For one-way sorts, we form 5 - 50 portfolios. For two-way sorts, we independently sort firms into 3-to-10 groups for each variable, resulting in a total of 9 -to-100 portfolios.

We compute the average cross-sectional differences in returns based on the predictive variables and methods that are commonly used in the literature. First, we assume that the investor sorts firms into portfolios each year based on values of the lagged predictive variables. Stocks in each portfolio are equally-weighted, and the monthly returns of the portfolios are calculated from July to June of the following year. For the one-way sorts, the investor next examines the mean returns (calculated over the entire sample) of all the ranked portfolios at the end of the sample period. She then calculates the "strategy return" to the combined zero-cost

portfolio that is long (short) in the extreme-ranked portfolios (for example, "10" or "1" for the 10-portfolio one-way sort) for a particular predictive variable. This procedure is followed for each of the firm characteristics, giving an average return for the long, short and combined portfolio for each of the sort combinations.

To choose among two-way-sort strategies, we first form nine (through 100) portfolios by sorting firms independently into three (through 10) groups based on the lagged values of *both* the firm characteristics under consideration. Average monthly returns from July to June of the following year are calculated for each of the portfolios. As before, the investor selects the extreme portfolios based on the two-way sorts (for example, "1" or "9" for the 3x3 two-way sort) and forms a zero-cost portfolio by buying one extreme portfolio using the proceeds from shorting the other extreme portfolio. This procedure is followed for each of the two-way sorts, giving an average return for the long, short and combined portfolios.[4]

We assume that the investor uses the entire sample period to examine the relation between predictor variables, or combinations thereof, at time t, and returns at time t+1. This technique is similar to both academic studies that analyze the predictive ability of firm characteristics and "back testing" used by financial institutions to analyze the effectiveness of proprietary trading strategies.

This procedure however has two potential sources of data-snooping bias however. The first and more obvious source of bias is that all 15 firm characteristics in our sample *successfully* capture cross-sectional differences in the in-sample returns of value versus glamour stocks. It is therefore possible that we discovered these variables after systematic searches over a wide range of potential variables over the past several decades. We attempt to provide the reader with a sense of the number of searches needed over randomly generated firm characteristics to create an illusion of predictability similar to the one documented in the literature.

The second source of snooping is subtler and directly related to the sorting procedures common to the asset pricing literature. It is assumed in most studies that each researcher ex ante determines both (a) the number of portfolios to sort securities into, and (b) the strategy to trade only the *extreme* portfolios based on the ex ante rankings of the firm characteristic involved. These choices may also suffer from snooping since prior familiarity with the firm characteristic is bound to unwittingly influence them.[5]

We create four measures to capture the biases from these potential sources of snooping. Each measure involves calculation of profits that are generated in datasets that, by design, have no predictability. To generate the random samples, we first generate random "firm characteristics." For each of the one-way simulations we use a non-repeating seed to generate a random factor $\sim N(0, 1)$, while for the two-way simulations, again using non-repeating seeds, we generate two random factors $\sim N(0, 1)$. The random firm characteristics have no time-series or cross-sectional relations with the firms' returns and, therefore, by design cannot be ex ante predictors of returns.[6]

All four data-snooping measures then use the random firm characteristics to answer the following question: What proportion of the real profits observed in the data (and reported in the literature) can be generated by implementing trading strategies based on one- and two-way sorts of the 15 "best" random variables selected from a pool of N potential characteristics? The number of random characteristics, N, over which the search is conducted is varied between 200 and 5,000. The main differences between the four measures hinge on the definition of the term "best," which reflects ex ante constraints the researcher or investor imposes on the strategies. The fewer constraints imposed, the greater the potential for data-snooping. To capture the flexibility available to the researcher regarding the ad hoc choice of the number of portfolios into which the sample of securities is sorted, we choose between 5 - 50 portfolios for the one-way

sort and between 9 - 100 portfolios for the two-way sort (resulting from 3x3 through 10x10 two-way analyses).

The first measure, "Nomono," is the most aggressive measure of snooping, where the profits to the trading strategy portfolio are calculated as the difference in returns to the ex post highest versus lowest performing portfolios. Contrary to the apparent practice used in the literature, "Nomono" implies that there is no ex ante monotonicity in the ranking requirements for the traded portfolios. Although this is an aggressive measure of snooping, given the numerous studies (published and unpublished, academic and practitioner) that have attempted to uncover predictive relations in returns data, it is plausible that we collectively have searched for the best (say 15) firm characteristics that also happen to have a monotonic relation with returns.[7]

The second measure of data-snooping, "Weakmono," is based on the two trading portfolios at each of the extremes of the monotonic ranking of the firm characteristic(s), but allows the researcher the flexibility to choose among them. For example, in a 10-portfolio one-way sort, the researcher can choose to trade portfolios 1 or 2 and 9 or 10, whichever yields the highest (lowest) returns and therefore maximizes the returns to the zero-cost portfolio. Such choices may reflect concerns about the reliability of data in the extreme portfolios, or judgments regarding nonlinearities in the data (for example, those related to negative earnings.) For examples of such patterns in published work, see Brennan, Jegadeesh and Swaminathan (1993), Daniel and Titman (1997), Fama and French (1992), Lakonishok, Shleifer, and Vishny (1994), and Tinic and West (1986).

The third measure, "Mono," maintains the ex ante monotonicity apparently used in the literature. Hence, in a 10-portfolio one-way sort, the trader can only be short (or long) in portfolios 1 or 10, where the ranking is determined ex ante. One could of course argue that in practice traders (or researchers) also *ex ante* stipulate which extreme portfolio (10 or 1) is bought

or sold, and that our measure therefore effectively allows for snooping on the extreme portfolios. Such a requirement would be most likely if the variables examined have theoretical justification. In the absence of rigorous guidelines for the choice of variables, however, we do not stipulate whether extreme portfolios should represent long or short positions.

Our fourth measure attempts to capture our familiarity with the data that is inherent in the pre-testing mentioned above. The measure, "Significant_beta," calculates the profits to the trading strategy as the difference in returns to portfolios based on the sorting of a firm characteristic that was found to have significant cross-sectional beta (|t-statistic| >1.96) in a prior regression of returns on the same characteristic. As in the Nomono measure, the "strategy return" is calculated as the difference in returns between the high and low portfolios based on ex post performance.

Our methodology for evaluating the impact of data-snooping on the profits of strategies based on firm characteristics of common stocks has some key advantages. First, in comparing in-sample strategy returns in the real and random data, we generate measures of snooping whose economic significance is easily evaluated since they are also measured in profits or returns. This is of particular interest in assessing the profitability of extant cross-sectional trading strategies where R-squared measures are less frequently used. Second, a practical advantage of our methodology is that we employ a portfolio formation technique that is directly comparable to the methods used in most of the recent predictability literature (namely, cross-sectional sorts). We do not develop our own in-sample predictors based on complex patterns detected with a "black box" technology, such as a sophisticated data-mining tool like a genetic algorithm or other forms of artificial intelligence. We instead employ the historical distributions of variables that have been purported to have meaningful economic links with expected returns. Finally, our bootstrapping methodology can be adapted to different contexts to gauge the effects of snooping.

II.   The Evidence

A.   Data

We construct a sample of non-financial firms that have returns listed in the 1995 CRSP

monthly files and data in the COMPUSTAT annual industrial files from 1955 through 1995.

Our sample includes firms that are listed on the NYSE, AMEX and NASDAQ exchanges. To

minimize the backfill bias [see, for example, Chan, Jegadeesh, and Lakonishok (1995)], we

require that firms have a minimum of two years of data available on COMPUSTAT before they

are included in the sample. We compute the values of each of the 15 firm characteristics used in

this study (see Appendix) for each company, including (the logs of) earnings-price (E/P),

dividend-price (D/P), cash flow-price (C/P), and book-to-market (B/M) ratios, as well as price,

market capitalization (size), and prior return measures. Following Breen and Korajczyk (1995),

we include four time-series variables in the sample; specifically, time-series versions of B/M,

D/P, past 12-month returns, and price per share.

In the construction of all the firm characteristics, we do not use any information that

would be unavailable to the investor at the time the portfolio decisions are made. We therefore

match COMPUSTAT fiscal year-end data from year t-1 with CRSP returns measured from July

of year t to June of year t+1 [see, for example, Fama and French (1992)]. Price and market

capitalization are calculated in June for the end of year t. Lagged 1- and 3-year returns are

calculated from the beginning of July of year t-j, where j=1, 3, to the end of June of year t.

Returns are calculated only when securities are traded, which should mitigate any effects of non-

trading in the sample. If a firm is de-listed during the year, we substitute the T-bill return for its

equity return for the remainder of the year.

B.   *In-sample profits: One-way sorts*

Table I presents the distribution of in-sample profits to the long, short, and combined (zero-cost) portfolios using one-way sorts. Panels A - D contain estimates for the 5 - through 50-portfolio sorts. The average (across the characteristics or "strategies") profits for the combined (zero-cost) portfolio for the five-portfolio sort are 0.42% per month and statistically significant with a t-statistic of 2.39. Not surprisingly, the average profits increase as the number of portfolios increase, and amount to 0.76% per month for the 50-portfolio strategy in Panel D. The profits of zero-cost (or combined) portfolios can obviously be scaled to any size. However, a comparison of the magnitude of these profits with the returns to the long and short portfolio strategies [for example, 1.32% per month and 0.89% per month for the 5-portfolio and 1.53% and 0.74% for the 50-portfolio strategies, respectively] shows that they are economically significant as well. The annualized returns to the trading strategies range between 5 percent and 9 percent for the 5- to- 50-portfolio strategies.

{INSERT TABLE I HERE}

For brevity, we do not report the performance of strategies based on each of the firm characteristics, but the strategies that perform the best are based on firm characteristics that are most commonly reported in the literature. For example, for the 10-portfolio sort, the strategies producing the maximum profits are book-to-market and cash-flow related variables. This pattern is consistent with two diametrically opposed possibilities: (a) commonly used variables in the literature are indeed "truly" related to the cross-section of expected returns; or, (b) these variables have emerged as "winners" following our collective mining of the data. Our paper is an attempt to determine the relative importance of these two hypotheses.

Figure 1, Panels A – D, presents simulation evidence on all four snooping measures that correspond directly to the real profitability of the trading strategies reported in Table I. Recall that the measures use random firm characteristics to answer the following question: What proportion of the real profits observed in the data can be generated by implementing trading strategies based on one-way sorts of the 15 "best" random variables searched from a pool of N potential characteristics, where N is varied between 200 and 5,000? The y-axis of Figure 1 shows average profits to the combined portfolios of the "best" 15 random factors as a percentage of the real average profits to the combined portfolios reported in Table I. The x-axis measures the number of searches that are conducted to obtain these 15 characteristics, and is scaled in increments of 200 up to 5,000.

{INSERT FIGURE 1 HERE}

If, for example, the reader believes that the entire profession conducted 200 independent searches to distill the 15 "real" firm characteristics, Panel A suggests that data-snooping could explain between 22% and 37% of the profits obtained in the 5-portfolio one-way sort. These proportions increase to 39% to 47% for 5,000 searches. The variation in the estimates (between 22% and 37% or between 39% and 47%) depends on the snooping measure used: "Weakmono," "Mono," "Significant_beta," or "Nomono." [8] This particular finding can of course be interpreted in several ways, but one particular interpretation is quite powerful. One can conclude that 200 researchers had to simultaneously conduct only one search each to uncover 15 firm characteristics that could generate "profits" on the order of 22% to 37% of the real profits reported in one-way 5-portfolio studies, even though the characteristics have no real predictive ability. Not surprisingly, the proportions in Figure 1 increase for all four measures with an increase in the number of searches.

There are two other noteworthy aspects of the evidence in Figure 1. First, the simulated profits grow at a higher rate than the real profits as the number of portfolios that the securities are sorted into increases (see Table I). This suggests that snooping effects increase with the fineness of the sort. Second, for each sorting procedure, the fraction of the profits explained increases fairly sharply for smaller values of N, and then stabilizes at proportions less than 100% of the profits observed in the real data. This suggests that data-snooping can explain some, but not all, of the predictability reported in the literature.

Most of the published analyses on one-way strategies use 10-portfolio sorts, making Figure 1, Panel B, the most relevant one for us to consider. For this specific type of strategy, the proportion of real profits that can be generated by random factors ranges between 30% and 50% for the 200 iteration case, and between 47% and 68 % for the 5,000 iteration case. This suggests that although a substantial fraction of the reported predictability may be spurious, it is reassuring that a large fraction of the profits appear to be "real." This conclusion is of course subject to the caveat that the profits have not been adjusted for transactions costs or risks involved in the execution of the strategies.

C.   *In-sample profits: Two-way sorts*

The evidence for the trading strategies based on two-way sorts is shown in Table II, with Panels A - D containing average profits for the long, short, and combined portfolios for the 3x3 through 10x10 sorts. The average profits to the combined portfolios are again positive and statistically significant at 0.57 % per month for the 3x3 sort. This amounts to an annualized return of seven percent. Again, there is a systematic and substantial increase in the profitability of the trading strategies from the 3x3-sort to the 10x10-sort, with returns to the latter being 1.02% per month (or, equivalently, over 12% on an annualized basis).[9]

{INSERT TABLE II HERE}

The findings in Table II are consistent with the evidence reported in the literature in two important respects. First, using relatively naïve "trading rules" that exploit the information contained in publicly available firm characteristics, we can apparently earn statistically and economically significant profits on zero-cost portfolios. Second, firm characteristics that exhibit the highest degree of predictive ability are also ones that have captured the attention of the profession, such as market capitalization [Keim (1983)], book-to-market [Fama and French (1992)] and cash flow-to-price ratios [Lakonishok, Shleifer, and Vishny (1994)].

{INSERT FIGURE 2 HERE}

The results from the simulation exercises for the 3x3 through 10x10 two-way sorts are presented in Figure 2, Panels A - D, which correspond to the real strategies presented for the two-way sorts in Table II, Panels A - D. The simulations for the two-way sorts are different in that 101 two-way combinations are chosen as the "best" among the N pairs considered in the simulation. [10] For the N = 200 simulation exercise presented in Panel A, the proportion of profits explained by the snooping measures ranges between 29% and 34% of the 0.57% per month profits in the real data. This percentage increases steadily as the number of firm characteristics

considered increases, with the estimates ranging between 39% and 45% for N = 5,000. The striking aspect of Figure 2 however is the dramatic increase in the snooping bias as the number of portfolios in the two-way sorts increases. Even for the case of N = 200, the proportions of profits generated by the simulated strategies vary between 45% and 52% for the 5x5 sort; between 58% and 70% for the 7x7 sort; and between 83% and 93% for the 10x10 sort. The estimates approach higher proportions as the number of searches increases to 1,000; for example, the proportion varies between 80% and 100% for the 10x10 case.

Given that up to 10x10 two-way sorts have been used in the literature, the estimates in Figure 2 are disturbing. Large proportions of the reported profits (or predictability) in the literature could be a result of our prior familiarity with the data. Ironically, in the absence of theory-based identification of the specific multiple asset-pricing factors, two-way sorts also seem to be the more legitimate and sensible course for empiricists to pursue. But, unlike the one-way sorts where a reasonable case can be made for genuine predictability in the data based on value versus glamour characteristics, our simulation evidence suggests that conclusions based on two-way sorts need to be interpreted with far more caution.

Why do two-way sorts have a greater potential for data-snooping bias? This propensity comes from two sources. First, it is simply easier for a researcher to generate a larger number of portfolios using two-way sorts. Specifically, a researcher who views a five-way sort as the minimum number of groups required to generate adequate cross-sectional dispersion in the characteristic of interest would quite naturally examine a minimum of 25 portfolios if two factors are considered. This seemingly innocuous practice will lead to a larger propensity to inadvertently snoop the data. We examine the magnitude of this bias by estimating the relation between average profits and the number of groups or portfolios used to categorize the securities. In both the real and simulated data, we regress the average profits across all characteristics in

one-way and two-way sorts on the number of portfolio groupings employed (P). These regressions are shown in Table III. In all cases, the number of groupings/portfolios is strongly and positively related to the level of profits. For example, the coefficient on P in the two-way real data of 0.0045 implies an increase in predicted profit of approximately seven basis points per month that can be achieved by merely moving from a 10-portfolio one-way sort to a 5x5-portfolio two-way sort.

{INSERT TABLE III HERE}

The second source of profits to the two-way strategies is related to the nature of the variables that researchers typically use. As mentioned earlier, the average correlation between the 15 firm characteristics in the sample is 0.13, but the correlation among the "successful" variables (those that are in the top decile of the two-way trading strategies' profits) more than doubles to 0.32. Consequently, we re-estimate the relation in two-way sorts between average profits and the number of portfolios for tercile correlation subgroups. The results are presented in Panel B of Table III.

Note that the coefficient on P, the number of portfolio groupings employed, doubles in the real data for highly correlated characteristics. Therefore, a move from a 10-portfolio one-way sort to a 5x5-portfolio two-way sort would lead to an increase in average monthly profits of ten basis points, or a 15% increase from the benchmark. The intuition for this is straightforward—the increase in the number of portfolios when the two characteristics are more strongly related is roughly equivalent to a finer sort on a single characteristic. More interestingly, the coefficient of P for the highly correlated characteristics in the real data is strikingly similar to the coefficient of P in the two-way sorts for highly correlated characteristics in the *simulated* data (0.0068 vs. 0.0076). This evidence suggests that, apart from being simply a finer cut on a (single) real factor, two-way sorts may have a greater tendency to generate spurious profits. We investigate

this issue in the section below, when we examine the decay rates of profits in the real and simulated data for one- and two-way sorts.

III.   Out-of-Sample Evidence

A common practice in the finance literature is to use "out-of-sample" tests to gauge the robustness of the predictive relations established within a given sample. In this section, we present traditional and some new out-of-sample tests to gauge the robustness of our conclusions based on the in-sample simulation exercises.

In the traditional out-of-sample tests, we construct our trading strategies by again ranking securities into portfolios based on firm characteristics, or combinations thereof. As before, we require that the strategies consist only of extreme portfolios, for example, portfolios 1 or 10 in a one-way 10-portfolio sort, or any corner combination in a two-way sort. Next, we identify portfolios formed at time t that earn the largest and smallest returns in year t+1 over a given five-year in-sample period. Last, we use these same portfolios (or "rules") in an adjacent future 5-year "out-of-sample" period, with the investor purchasing (shorting) the portfolio that had the highest (lowest) returns in the prior in-sample sub-period. We roll forward through the data, estimating optimal in-sample rules that are then used as the basis for strategies for subsequent out-of-sample periods. This process results in a non-overlapping time-series of long, short, and combined out-of-sample portfolio returns to each of the one- and two-way variable combinations.

The "out-of-sample" results for both the one- and two-way sorts are reported in Table IV. For brevity, we restrict our one-way sorts to the 10- and 50-portfolio cases and the two-way sorts to the 3x3- and 7x7-portfolio cases. Not surprisingly, the out-of-sample predictive ability of the 15 real firm characteristics weakens. Table IV, Panels A and B, show that out-of-sample profits

to the zero-cost combined portfolio using one-way sorts decline by 50% (from 0.52% to 0.26% per month) for the 10-portfolio case and by 55% (from 0.76% to 0.34% per month) for the 50-portfolio case. The differences between the in- and out-of-sample profits are statistically significant with p-values of less than 0.001. The results for the two-way sorts in Panels C and D show even larger percentage drops in the profitability of the trading strategies, with the out-of-sample profits dropping by 58% for both the 3x3 and 7x7 sorts. None of the trading strategies' profits reported in Table IV are statistically significant, with the highest t-statistic being 1.39. In addition, Spearman correlation analysis of the rank ordering of in- versus out-of-sample profits of trading strategies shows that, on average, a strategy's in-sample profitability is not a good predictor of its relative out-of-sample performance.

{INSERT TABLE IV HERE}

In spite of the statistically weak profitability of the strategies reported in Table IV, the profits remain economically significant; out-of-sample annualized returns are three to four percent for one-way sorts and about seven percent for two-way sorts. However, snooping is likely to affect even these "out-of-sample." Specifically, we have assumed that the 15 firm characteristics used in the trading strategies are chosen randomly. Previous studies, and our results in Tables I and II, however, have already shown that these firm characteristics are correlated with returns over the *entire* sample period. The "out-of-sample" profits reported in Table IV are therefore not strictly out-of-sample; they are based on sub-periods that are part of the overall sample used in the literature to generate the in-sample profits.[11]

To address the above concern that typical out-of-sample tests may not be truly out-of-sample, we examine another variant of such tests: the use of a single holdout period [see, for example, Davis, Fama, and French (2000)]. We divide our sample into two equal sub-periods. Using random firm characteristics in one-way simulations, we find 207 strategies out of 5,000

that have significant betas at the five percent level or better in the first sub-period. When we test

these 207 strategies in the second 'holdout' period, we find that only 11 have significant betas of

the same sign at the ten percent significance level, and eight significant betas at the five percent

significance level. This suggests that if an investigator chooses a successful strategy at the end of

the first sub-period, and genuinely has no knowledge of the holdout sample, then it is unlikely

that she would reject the null hypothesis (of no predictability) in the holdout period by chance (a

11/207 chance at the 10%, and a 8/207 chance at the five percent significance level).

Many "out-of-sample" tests however take a different form than the one discussed above.

Studies attempting to verify the existing evidence "out-of-sample" use additional time-series of

data to confirm the evidence initially established over a given time period. However, the

additional time-series is typically much shorter in length compared to the original data, thus

causing such experiments to be also affected by the snooping bias. Virtually all ongoing studies

on value versus glamour firm characteristics that use historical CRSP and COMPUSTAT data,

for example, fall into this category. It may therefore be informative to examine the *difference*

between the initial in-sample, and the shorter out-of-sample, tests.

To formally examine the holdout-period robustness in the real data, we examine the

differences in "decay rates" of the predictability of the real versus the simulated firm

characteristics.[12] The decay rate is simply a measure of the change in a specific trading strategy's

profits following the initial "discovery period." Specifically, for each of the 15 real firm

characteristics and the 15 best random characteristics from the 5,000 simulations, and two-way

combinations thereof, we form trading strategy portfolios starting in 1974 using all past available

data.[13] We restrict the combined portfolios to be formed using only the extreme portfolios, thus

imposing ex ante monotonicity. For each variable, we roll through the data in yearly increments,

and identify the first occurrence of statistically significant profits (that is, a return for the

combined portfolio with a t-statistic greater than 1.96). For each strategy, we then track the

subsequent average profits, and compute the decay rate as the level of average profits over a

particular period relative to the average profits during the discovery period. Thus, a decay rate of

100% means that a variable experiences no drop in average profits after first discovery, a rate of

zero implies that the profits drop to zero, and a rate of –100% means the relationship between the

firm characteristic and returns is completely reversed. For example, if book-to-market has a first

statistically significant profit of 0.40% in 1978 (using data from the 1965 to 1978), then we

would track the average return going forward in time, and compare it to 0.40 percent. If, for

example, over the 1979 to 1989 period the average monthly return happened to be 0.20 percent

for this strategy, then that would imply a decay rate of 50 percent over ten years.

For a given sort dimension, we stack the decay rates across variables using the date of

first discovery as the first observation. We then compute averages across variables for an "event

window" that runs 240 months forward from the date of first discovery. To draw inferences, we

then compare the decay rates in the real versus simulated data.[14] If profits in the real data decay

at a rate greater than or equal to the simulated data, then the evidence of predictability in the real

data is consistent with snooping. If, on the other hand, the decay rate in the real sample is lower,

then it suggests the existence of some genuine predictability (given the caveat that our real

variable set was determined in the year 2001). Again for brevity, we conduct this analysis for

only the 10- and 50-portfolio one-way sorts and the 3x3- and 7x7-portfolio two-way sorts.

Table V contains the decay rates in the profits of both the real and random firm

characteristics. For the real characteristics, the average decay rates range from a high of almost

60 percent for the 10-portfolio one-way sort to a low of 20 percent for the 7x7-portfolio two-way

sort. These estimates describe the out-of-sample profitability of the specific trading strategies

considered in this paper. For example, for the 10-portfolio one-way sort, the best variable in

terms of decay is book-to-market, with a decay rate of 146.3 percent (implying that profits from the use of this variable actually *increase* in the out-of-sample period) and the worst performing variable is Price, with a decay of –76.9 percent (implying a reversal in the relation observed in-sample between this characteristic and equity returns).

{INSERT TABLE V HERE}

A more interesting exercise however is to compare the decay rates in the profits of real trading strategies with the decay rates in the profits of strategies based on the 15 best random firm characteristics. The average decay rates across the four strategies reported in Table V appear to be quite similar. For the 10-portfolio one-way and the 7x7-portfolio two-way sorts the decay rates for the real versus simulated strategies are virtually identical. The test-statistics for the difference between the average decay rates of the profits of the real versus random characteristics are the statistically insignificant values of 0.01 and -1.30, respectively. The fact that the decay rates are non-zero in the random samples, suggests that the returns are positively autocorrelated. If stock returns were uncorrelated, there should be no out-of-sample predictability in the random samples. The non-zero decay rates for the random samples, and their similarity to the ones in the real data, also suggests that the out-of-sample predictability of the real factors may be a result of positive autocorrelation in returns, rather than the factors themselves containing predictive power for future returns.

A comparison of Panels B of Table V shows that profits of strategies based on the random data for the 50-portfolio sort actually exhibits a better decay rate than those based on real data. The random data portfolio earns 43.7 percent of the initial period profits, while the real data portfolio earns only 31.7 percent of its corresponding discovery period profits. The t-statistic for the difference between the average decay rates is -7.72. In fact, the 3x3-sorts in Panels C are the only set of trading strategies for which the real strategies have a better decay

rate than the strategies based on random firm characteristics. The average decay rates of the real versus random strategies are 35.1 percent and 27.2 percent, respectively, and the t-statistic for the difference is 6.11.[15]

{INSERT FIGURE 3 HERE}

Figure 3 plots the average decay rates of the profits to the real and random strategies over the 240 months after the initial discovery period. The time-series behavior of the average decay rates in the graphs, combined with the average estimates in Table IV, illustrate a number of important points. First, the behaviors of the decay rates of the real and random data are similar, both over time and at any particular point in time. Second, the decay rates are worse for the higher dimension sorts, which confirm our earlier findings that trading strategies based on multiple sorts are more likely to yield statistically significant in-sample profits simply due to chance. Third (and something that is not obvious from the graphs), if we ex post choose the best portfolios (highlighted in the "Max" columns in Table IV) from the decay studies, we find that in three out of four cases (the exception being the 7x7 sort), strategies based on the best random firm characteristics perform better "out-of-sample" than those based on the best real characteristics.[16]

Finally, there is also another set of "out-of-sample" tests conducted by researchers that involves testing the robustness of the findings in a particular country (typically the U.S.) by using data from other countries. Such "out-of-sample" tests have also been conducted in the "value versus glamour" literature, but with mixed results. For example, Fama and French (1998) find strong evidence for the "value premium" in several countries during the 1975 to 1995 period. On the other hand, in a recent paper, Fohlin and Bossaerts (2001) find that evidence from Germany from a very different time period, 1881 to 1913, shows that the size effect is entirely due to selection bias, the momentum effect does not exist, and although there is a strong

book-to-market effect, it is of the opposite sign of the effect uncovered in the post-war U.S. data. Although our methodology is well suited for testing the robustness of the international evidence to snooping biases, the execution of such a study is complicated by other factors. In particular, inadvertent snooping in "out-of-sample" international studies can occur not only because of correlations in the stock returns across markets, but also because of correlations among the various firm characteristics. Given the mixed evidence across different markets, especially when different time periods are used [see Fohlin and Bossaerts (2001)], and the increased complexity of our simulation exercises with its requirements of data availability, we defer such a study to the future.[17]

IV. Conclusion

The fragility of the one-factor Capital Asset Pricing Model (CAPM) in explaining the cross-section of expected returns of financial securities has led to a tremendous resurgence of research aimed at discovering variables that might explain the behavior of returns. Much of this research uses sorting procedures to uncover relations between combinations of "value versus glamour" firm characteristics and equity returns. In this paper, we examine the propensity of these strategies to generate statistically and economically significant profits due to data-snooping biases induced by our *collective* familiarity with the data. We are particularly interested in the potential effects of snooping because of our increasing tendency to sort securities simultaneously on multiple characteristics.

We construct four different measures of data-snooping, the differences among them hinging on the extent of snooping. All the measures are based on a simulation technique that can be easily adapted to other scenarios. Under plausible assumptions, our collective familiarity with the data could account for large fractions of relations between well-known value and glamour firm characteristics reported in the literature. This is particularly the case when multiple firm

characteristics are used simultaneously to sort firms, an arguably natural and understandable trend in the literature given the inability of the one-factor model to explain required returns on financial assets. The higher propensity for data-snooping in two-way (or even a three-way) sorts exists because it is much easier to generate a larger number of portfolios while simultaneously conditioning on multiple characteristics. This tendency is exacerbated because multi-sort strategies that generate the highest profits are also ones that sort firms based on correlated characteristics. We also present traditional and some new out-of-sample tests to confirm our in-sample findings.

Our research is most closely related to the recent work of Berk (1998), Lo and MacKinlay (1990), Foster, Smith and Whaley (1997), and Sullivan, Timmermann, and White (1999, 2001) who statistically address the issue of data-snooping, albeit with respect to different applications. Lo and MacKinlay (1990) demonstrate that conducting tests of asset pricing models on portfolios that have been formed on characteristics of the data can lead to substantially higher and potentially spurious rejection rates of the null hypothesis. Similarly, Berk (1998) shows that the common practice of sorting firms into portfolios and then running asset-pricing tests *within* each
group introduces a bias in favor of rejecting the model under consideration. Foster, Smith and Whaley (1997) also focus on tests of asset pricing models, but examine the biases in R^2 measures when researchers choose k predictors from a larger set of **m** possible variables. They propose variations in the traditional tests that researchers use to assess whether a particular variable is predictable. Sullivan, Timmermann, and White (1999) use a bootstrapping methodology to show the dominant effects of data-snooping on technical trading rules, and the same authors show how a "statistical reality check" on the numerous well-publicized calendar effects in stock returns can also reveal their fragility [see Sullivan, Timmermann, and White (2001)].

Although these papers warn researchers about excessive familiarity with the data when drawing inferences about tests of asset pricing models, most of the recent empirical literature on asset pricing draws conclusions about the importance of firm characteristics by emphasizing the economic profits or returns to trading strategies based on sorting procedures using these characteristics. In this important respect, our research is more pertinent to the large and rapidly growing literature on empirical asset pricing. Another important aspect of our work is that we assume that all the profits to trading strategies conditioned on well-known and readily observable value versus glamour firm characteristics are "excess" profits, over and above what should be earned as a reward for the risk(s) of the strategies. Even under this extreme assumption, we are able to provide support for the hypothesis that a large fraction of the reported relations between equity returns and firm characteristics could result from our collective familiarity with the data.[18] Depending on a researcher's interpretation, this finding could make it easier to detect the sources of true profitability, or allow a better focus on the risk-based explanations of such predictability [see, for example, Berk (1995), Berk, Green, and Naik (1999), Fama and French (1996), and Ferson, Sarkissian, and Simin (1999).]

APPENDIX

This appendix contains a list of all the firm characteristics with their definitions in (Panel A), and

the COMPUSTAT variables used in constructing them (Panel B).

Panel A

Cross-Sectional Variable	Construction
1. Price Per Share of Common Stock (P)	Center for Research in Security Prices (CRSP) price per share
2. Dividend-Price Ratio (D/P)	(item201)/(item199)
3. Earnings-Price Ratio (E/P)	(item58)/(item199)
4. Cashflow-Price Ratio (C/P)	((item18+item14)/ (item25*item199)
5. Market Value of Equity (ME)	CRSP price*CRSP number of shares
6. Total Book-Market Equity Ratio (B/ME)	(item60+item74+item208-item130)/(item25*item199)
7. Book Debt-Equity Ratio (D/E)	(item9)/(item60)
8. Average five-year growth rate in sales (GS)	(mean of the annual percent change in item12 from year t to year t-4)/(standard deviation of the annual percent change in item 12 from year t to year t-4)
9. Return on Investment (ROE)	(item258)/(item216)
10. One-Year Returns (1-YEAR)	12-month holding period return from CRSP
11. Three-Year Returns (3-YEAR)	36-month holding period return from CRSP
Time-series Variable	**Construction**
12. Price Per Share of Common Stock (P)	Same as 1. above
13. Dividend-Price Ratio (D/P)	Same as 2. above
14. Total Book-Market Equity Ratio (B/ME)	Same as 6. above
15. One-Year Returns (1-YEAR)	Same as 10. above

Panel B

Compustat Annual Data Items	Definition
9	Long-Term Total Book Debt
12	Net Sales
14	Depreciation and Amortization
18	Income Before Extraordinary Items
25	Common Shares Outstanding
58	Earnings per share
60	Total Common Book Equity
74	Deferred Taxes, Balance Sheet
130	Preferred Stock Par Value
199	Price of Common Stock, Fiscal Year End
201	Dividends Per Share
208	Investment Tax Credit, Balance Sheet
216	Stockholders' Equity
258	Net Income Adjusted for Common Stock Equivalents

REFERENCES

Ball, Ray, 1978, Anomalies in relationships between securities' yields and yield-surrogates, *Journal of Financial Economics* 6, 103-126.

Banz, Rolf W., 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3-18.

Basu, Sanjoy, 1983, The relationship between earnings yield, market value, and return for NYSE common stocks: Further evidence, *Journal of Financial Economics* 12, 129-156.

Berk, Jonathan B., 1995, A critique of size related anomalies, *Review of Financial Studies* 8, 275-286.

Berk, Jonathan B., 1998, Sorting out sorts, *Journal of Finance* 55, 407-427.

Berk, Jonathan B., Richard C. Green, and Vasant Naik, 1999, Optimal investment, growth options, and security returns, *Journal of Finance* 54, 1553-1607.

Bhandari, Laxmi Chand, 1988, Debt/equity ratio and expected common stock returns: Empirical evidence, *Journal of Finance* 43, 507-528.

Black, Fischer, 1993a, Beta and return, *Journal of Portfolio Management* 20, 8-18.

Black, Fischer, 1993b, Estimating expected return, *Financial Analyst Journal* 49, 36-38.

Black, Fischer, Michael C. Jensen, and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests, in M. Jensen, ed.: *Studies in the Theory of Capital Markets* (Praeger, New York).

Breen, William J., and Robert Korajczyk, 1995, On selection biases in book-to-market tests of asset pricing models, Working Paper, Kellogg Graduate School Of Management.

Brennan , Michael J., Jegadeesh, Narasimhan, and Bhaskaran Swaminathan, 1993, Investment analysis and the adjustment of stock prices to common information, *Review of Financial Studies* 6, 799-824.

Capaul, Carlo, Ian Rowley, and William F. Sharpe, 1993, International value and growth stock

returns, *Financial Analyst Journal*, January-February, 27-36.

Chan, Louis K.C., Narasimhan Jegadeesh, and Josef Lakonishok, 1995, Evaluating the

performance of value versus glamour stocks: The impact of selection bias, *Journal of*

Financial Economics 38, 269-296.

Chan, Louis K.C., Jason Karceski, and Josef Lakonishok, 1998, The risk and return from factors,

Journal of Financial and Quantitative Analysis 33, 159-188.

Chan, Louis K.C., Yasushi Hamao, and Josef Lakonishok, 1991, Fundamentals and stock returns

in japan, *Journal of Finance* 46, 1739-1789.

Chen, Nai-Fu, Richard Roll, and Stephen A. Ross, 1983, Economic forces and the stock market,

Journal of Business 59, 383-403.

Claessens, Stijn, Susmita Dasgupta, and Jack Glen, 1996, The cross-section of expected returns:

evidence from the emerging markets, Working Paper, International Finance Corporation.

Connor, Gregory, 1984, A unified beta pricing theory, *Journal of Economic Theory* 34, 13-31.

Connor, Gregory And Robert A. Korajczyk, 1986, Risk and return in an equilibrium APT:

Application of a new test methodology, *Journal of Financial Economics* 21, 255-289.

Cooper, Michael, Roberto Gutierrez, and William Marcum, 2001, On the predictability of stock

returns in real time, forthcoming, *The Journal of Business.*

Daniel, Kent, and Sheridan Titman, 1997, Evidence on the characteristics of cross sectional

variation in stock returns, *Journal of Finance* 53, 1-33.

Davis, James, 1994, The cross-section of realized stock returns: the pre-compustat evidence,

Journal of Finance 49, 1579-1593.

Davis, J., Fama, Eugene F., and Kenneth R. French, February 2000, Characteristics, covariances,

and average returns: 1029-1997, *Journal of Finance* 56, 389-406.

Denton, F., 1985, Data mining as an industry, *Review of Economics and Statistics* 67, 124-127.

Fama, Eugene F., and Kenneth R. French, 1992, The cross section of expected stock returns, *Journal of Finance* 47, 427-465.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns of stocks and bonds, *Journal of Financial Economics* 33, 3-56.

Fama, Eugene F., and Kenneth R. French, 1995, Size and book-to-market factors in earnings and returns, *Journal of Finance* 50, 131-155.

Fama, Eugene F., and Kenneth R. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55-84.

Fama, Eugene F., and Kenneth R. French, 1998, Value versus growth: The international evidence, *Journal of Finance* 53, 1975-1999.

Fama, Eugene F., and James Macbeth, 1973, Risk, return, and equilibrium: empirical tests, *Journal of Political Economy* 81, 607-63 6.

Ferson, Wayne E., Sergei Sarkissian, and Timothy Simin, 1999, The alpha factor asset pricing model: A parable, *Journal of Financial Markets* 2, 49-68.

Fohlin, Caroline, and Peter Bossaerts, 2001, Has the cross-section of average returns always been the same? Evidence from germany, 1881-1913, Caltech Social Sciences Working Paper No. 1084, California Institute of Technology, Pasadena, CA.

Foster, F. Douglas, Tom Smith, and Robert E. Whaley, 1997, Assessing the goodness-of-fit of asset pricing models: The distribution of the maximal R^2, *Journal of Finance* 52, 591-607.

Gibbons, Michael, 1982, Multivariate tests of financial models: A new approach, *Journal of Financial Economics* 10, 3-27.

 (Mcgraw-Hill Book Company, New York).

Jaffee, Jeffrey, Donald B. Keim, and Randolph Westerfield, 1989, Earnings yield, market values, and stock returns, *Journal of Finance* 44, 135-148.

Kothari, S.P., and Jay Shanken, 1997, Book-to-market, dividend yield, and expected market returns: A time-series analysis, *Journal of Financial Economics* 44, 169-203.

Kothari, S.P., Jay Shanken, and Richard G. Sloan, 1995, Another look at the cross-section of expected stock returns, *Journal of Finance* 50, 185-224.

Keim, Donald B., 1983, Size-related anomalies and stock return seasonalities: Further empirical evidence, *Journal of Financial Economics* 12, 13-32.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* 49, 1541-1578.

La Porta, Rafael, 1993, Survivorship bias and the predictability of stock returns in the compustat sample, Working Paper, Harvard University.

Leamer, Edward E., 1978, *Specification Searches* (Wiley, New York).

Lintner, John, 1965, The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13-37.

Lo, Andrew W., and A. Craig Mackinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies*, 3, 431-467.

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867-887.

Reinganum, Marc R., 1981, A new empirical perspective on the CAPM, *Journal of Financial and Quantitative Analysis* 16, 439-462.

Roll, Richard, and Stephen A. Ross, 1980, An empirical investigation of the arbitrage pricing theory, *Journal of Finance* 35, 1073-1103.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive evidence of market inefficiency, *Journal of Portfolio Management* 11, 9-17.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341-360.

Ross, Stephen A., 1987, Regression to the max, Working Paper, Yale School of Organization and Management.

Sharpe, William F., 1982, Factors in NYSE security returns, 1931-1979, *Journal Of Portfolio Management* 8, 5-19.

Stambaugh, Robert F., 1982, On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis, *Journal of Financial Economics* 10, 237-268.

Sullivan, Ryan, Allan Timmermann, and Halbert White, 1999, Data-snooping, technical trading rule performance, and the bootstrap, *Journal of Finance* 54, 1647-1691.

Sullivan, Ryan, Allan Timmermann, and Halbert White, 2001, Dangers of data-driven inferences: the case of calendar effects, *Journal of Econometrics,* 105 (1), 249-286.

TABLE I

In-Sample Profits to Trading Strategies Based on One-way Sorts of Real Data, 1965-1995

This table contains in-sample profits of trading strategies that are based on one-way sorts of 15 firm characteristics, 11 cross-sectional and four time-series. Securities are ranked into 5 (Panel A) through 50 (Panel D) categories using each of the characteristics and then combined into portfolios. Portfolio L (S) denotes an extreme portfolio (for example, 5 or 1, or 50 or 1, in Panels A and D, respectively) based on the ex ante sorting of each of the firm characteristic. The mean return of L (S) is the average return across all firm characteristics. The combined zero-cost portfolio, C, is long in L and short in S. Mean profits and returns are in percent, with corresponding t-statistics and the standard deviation of the returns and profits in the adjacent columns, and "Min" and "Max" are the highest and lowest average returns (profits) across individual firm characteristics over the 30-year (1965:07 - 1995:06) sample period.

Panel A: 5-Portfolio Sort

Portfolio	Mean	T-Stat.	Std	Min	Max
L	1.32	4.55	0.17	1.06	1.60
S	0.89	3.02	0.13	0.70	1.08
C	0.42	2.39	0.25	0.10	0.89

Panel B: 10-Portfolio Sort

Portfolio	Mean	T-Stat.	Std	Min	Max
L	1.37	4.30	0.20	0.99	1.63
S	0.84	2.75	0.13	0.64	1.04
C	0.52	2.41	0.28	0.19	0.98

Panel C: 30-Portfolio Sort

Portfolio	Mean	T-Stat.	Std	Min	Max
L	1.48	3.91	0.33	0.93	2.10
S	0.80	2.54	0.18	0.46	1.01
C	0.65	2.26	0.43	0.14	1.60

Panel D: 50-Portfolio Sort

Portfolio	Mean	T-Stat.	Std	Min	Max
L	1.53	3.73	0.43	0.94	2.48
S	0.74	2.30	0.20	0.33	1.00
C	0.76	2.16	0.52	0.00	2.07

TABLE II

In-Sample Profits to Trading Strategies Based on
Two-way Sorts of Real Data, 1965-1995

This table contains the profits of trading strategies that are based on two-way independent sorts of 15 firm characteristics, 11 cross-sectional and four time-series. We examine all two-way combinations of the 15 variables except for two-way combinations that use the cross-section and time-series of the same variable. Thus, the results below are reported for 101 (105 - 4) two-way combinations. Portfolios are formed by 3x3 (Panel A) through 10x10 (Panel D) sorts on each pair of the 15 firm characteristics. Portfolio L (S) denotes an extreme portfolio of the nine (or 100) portfolios in the 3x3 (or 10x10) ex ante two-way sorts of pairs of the firm characteristics. The mean return of L (S) is the average return across all firm characteristics. The combined zero-cost portfolio, C, is long in L and short in S. Mean profits and returns are in percent, with corresponding t-statistics and the standard deviation of the returns and profits in the adjacent columns, and "Min" and "Max" are the highest and lowest average returns (profits) across individual firm characteristics over the 30-year (1965:07 - 1995:06) sample period.

Panel A: 3x3 Sort

Portfolio	Mean	T-Stat.	Std	Min	Max
L	1.39	4.47	0.13	1.15	1.79
S	0.79	2.71	0.12	0.47	1.00
C	0.57	3.12	0.17	0.26	1.06

Panel B: 5x5 Sort

Portfolio	Mean	T-Stat.	Std	Min	Max
L	1.48	4.15	0.21	1.10	2.23
S	0.68	2.22	0.18	0.05	0.98
C	0.76	2.69	0.30	0.24	1.57

Panel C: 7x7 Sort

Portfolio	Mean	T-Stat.	Std	Min	Max
L	1.47	3.77	0.23	1.06	2.12
S	0.57	1.77	0.27	-0.71	0.98
C	0.84	2.40	0.35	0.04	2.07

Panel D: 10x10 Sort

Portfolio	Mean	T-Stat.	Std	Min	Max
L	1.53	3.50	0.24	0.92	2.23
S	0.46	1.27	0.33	-0.62	0.95
C	1.02	2.24	0.42	0.11	2.07

Table III

Regressions of Combined Portfolio Profits on the Number of Portfolios in a Sort

This table contains OLS regressions of combined portfolio profits on an intercept and the number of portfolios in a sort. For one-way sorts, we use the combined portfolio profits from 10, 25, 50, and 100 sorts. For two-way sorts, we use the combined profits from 3x3, 5x5, 7x7, and 10x10 sorts. In Panel A, we report regression results for real-data one-way sorts, real-data two-ways sorts, random-data one-way sorts, and random-data two-way sorts. In Panel B, we report regression results for two-way sorts for both real and random-data using subgroups based on tercile groupings of correlations between the firm characteristics employed in the sorts. For the random sorts, we carry out the analysis using the best 101 two-way random variable combinations chosen from the 5,000 simulations. We restrict the combined portfolios to be formed using only the extreme portfolios, thus imposing monotonicity. We multiply all point estimates by 100. T-statistics are in parentheses.

Panel A: One-way and Two-way Regressions

		Real Data	Simulated Data
One-Way Sorts	Coefficient on the Number of Portfolios in Sort (P)	0.0034 (2.01)	0.0075 (10.73)
Two-Way Sorts	Coefficient on the Number of Portfolios in Sort (P)	0.0045 (9.49)	0.0067 (30.43)

Panel B: Two-way Regressions by Levels of Correlation Between the Firm Characteristics of the Sorts

		Real Data	Simulated Data
Low Correlation Tercile Two-Way Sorts	Coefficient on the Number of Portfolios in Sort (P)	0.0033 (4.8)	0.0053 (17.05)
Middle Correlation Tercile Two-Way Sorts	Coefficient on the Number of Portfolios in Sort (P)	0.0035 (4.70)	0.0077 (21.98)
High Correlation Tercile Two-Way Sorts	Coefficient on the Number of Portfolios in Sort (P)	0.0068 (7.65)	0.0077 (20.88)

TABLE IV

**Out-of-Sample Profits to Trading Strategies Based on
One-and Two-Way Sorts of Real Data, 1965-1995**

This table contains the *out-of-sample* profits of trading strategies that are based on one-way and two-way sorts of 15 firm characteristics. Panels A and B contain *out-of-sample* returns and profits of portfolio combinations based on one-way sorts of securities using each of the firm characteristics. Securities are ranked into 10 (Panel A) and 50 (Panel B) categories using each of the characteristics and then combined into portfolios. Panels C and D contain out-of-sample returns and profits of portfolio combinations based on two-way sorts of securities using the firm characteristics. We examine all two-way combinations of the 15 variables except for two-way combinations that use the cross-section and time series of the same variable. Thus, the results below are reported for 105 minus 4 = 101 two-way combinations. Portfolios are formed by 3x3 (Panel C) and 7x7 (Panel D) sorts on each pair of the 15 firm characteristics. Portfolio L (S) denotes an extreme portfolio of the nine (49) portfolios in the 3x3 (7x7) ex ante two-way sorts of pairs of the firm characteristics. The mean return of L (S) reported in the table is the average *out-of-sample* annual return over the subsequent five years across all two-way combinations. The combined zero-cost portfolio, C, is long in L and short in S. The L and S portfolios are required to be a corner portfolio. The investor is assumed to rebalance the portfolios every year after updating the values of the firm characteristics that form the bases of the strategies. Mean profits and returns are in percent, with corresponding t-statistics and the standard deviation of the returns and profits in the adjacent columns; and "Min" and the "Max" are the highest and lowest average returns (profits) across individual firm characteristics over the 30-year (1965:07 - 1995:06) sample period.

Panel A: One-way (10-Portfolio) Sort

Portfolio	Mean	T-Stat.	Std	Min	Max
L	1.25	4.13	0.18	0.88	1.54
S	0.97	3.49	0.15	0.76	1.24
C	0.26	1.27	0.29	-0.12	0.73

Panel B: One-way (50-Portfolio) Sort

Portfolio	Mean	T-Stat.	Std	Min	Max
L	1.25	3.49	0.35	0.48	1.87
S	0.87	2.75	0.20	0.40	1.12
C	0.34	1.10	0.55	-0.74	1.40

Panel C: Two-way (3x3) Sort

Portfolio	Mean	T-Stat.	Std	Min	Max
L	1.22	4.37	0.13	0.94	1.50
S	0.93	3.32	0.18	0.55	1.26
C	0.24	1.39	0.18	-0.20	0.76

Panel D: Two-way (7x7) Sort

Portfolio	Mean	T-Stat.	Std	Min	Max
L	1.27	3.59	0.25	0.61	2.04
S	0.79	2.45	0.38	-0.21	1.80
C	0.35	1.05	0.38	-0.60	1.26

TABLE V

Decay Rates for Real and Random Data Sorts

This table contains decay rates for combined portfolio sorts based on real and random firm characteristics. Panels A and B contain decay rates for one-way 10- and 50-portfolio sorts, while panels C and D contain decay rates for 3x3- and 7x7-portfolio two-way sorts. The decay rate is derived from the combined portfolio profits between the first significant period for a sort variable and how that variable performs after the initial discovery period. Specifically, for each of the 15 real firm characteristics (or the 101 two-way combinations thereof), we form combined portfolios starting in 1974 (to allow for an adequate time-series of data for each real characteristic), using all past available data. For the random sorts, we carry out the same analysis on each of the best 15 random variables (or best 101 two-way random variable combinations) chosen from the 5000 simulations. We restrict the combined portfolios to be formed using only the extreme portfolios, thus imposing monotonicity. For each variable, we roll through the data in yearly increments using an expanding window, and keep track of the first occurrence of a significant combined portfolio (t-statistic greater than 1.96). We then track for each variable what the average profits are in the years after first discovery. For each sort variable, we compute a decay rate equal to the average profits over a particular period after first discovery relative to the average profits during the discovery period. Thus, a rate of 100% means that a variable experiences no drop in average profits after first discovery, a rate of zero implies that the profits drop to zero, and a rate of –100% means the relationship between the firm characteristic and returns is completely reversed. We line up the decay rates across variables using the date of first discovery as the first observation. We then compute the average decay rate across characteristics for an "event window" that runs 240 months out from date of first discovery. The "Mean" and "Std" (standard deviation) reported below are from the average monthly decay rate across all characteristics' decays. The "Min" and the "Max" are the highest and lowest decay rates per individual firm characteristic.

Real Sorts
Panel A: One-way (10-Portfolio) Sort

Portfolio	Mean	Std	Min	Max
Decay Rate	59.9%	22.8%	-76.9%	146.3%

Panel B: One-way (50-Portfolio) Sort

Portfolio	Mean	Std	Min	Max
Decay Rate	31.7%	20.1%	-69.6%	137.8%

Panel C: Two-way (3x3) Sort

Portfolio	Mean	Std	Min	Max
Decay Rate	35.1%	18.9%	-113.1%	149.8%

Panel D: Two-way (7x7) Sort

Portfolio	Mean	Std	Min	Max
Decay Rate	20.1%	10.5%	-187.9%	174.2%

Random Sorts
Panel A: One-way (10-Portfolio) Sort

Portfolio	Mean	Std	Min	Max
Decay Rate	59.5%	20.6%	21.1%	188.4%

Panel B: One-way (50-Portfolio) Sort

Portfolio	Mean	Std	Min	Max
Decay Rate	43.7%	18.6%	16.3%	156.2%

Panel C: Two-way (3x3) Sort

Portfolio	Mean	Std	Min	Max
Decay Rate	27.2%	5.1%	-172.0%	246.1%

Panel D: Two-way (7x7) Sort

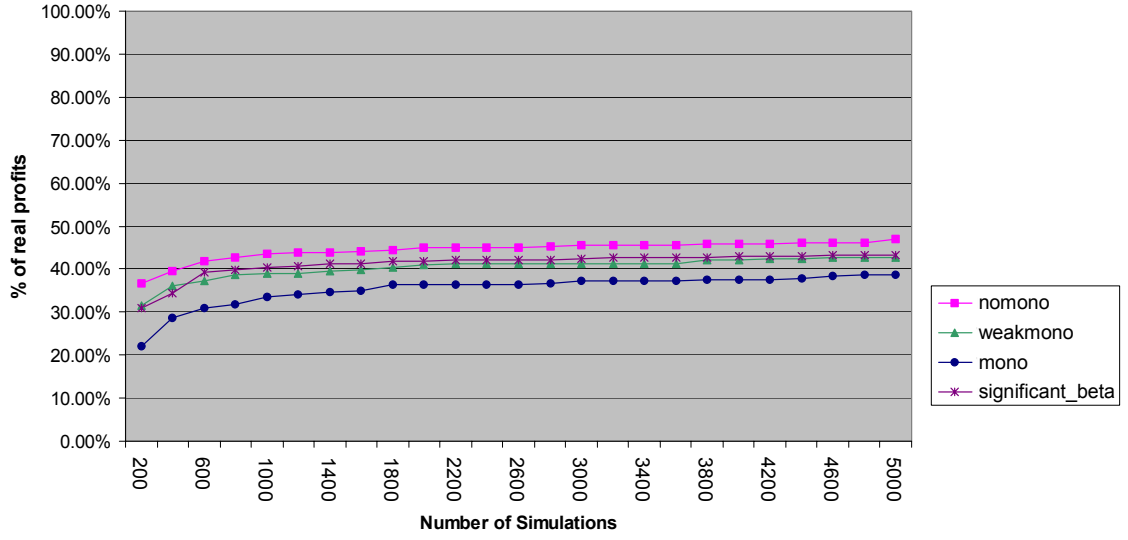| Portfolio	Mean	Std	Min	Max
Decay Rate	21.2%	9.5%	-52.9%	121.1%

FIGURE 1

Profits to One-Way Trading Strategies Implemented on Randomized Returns

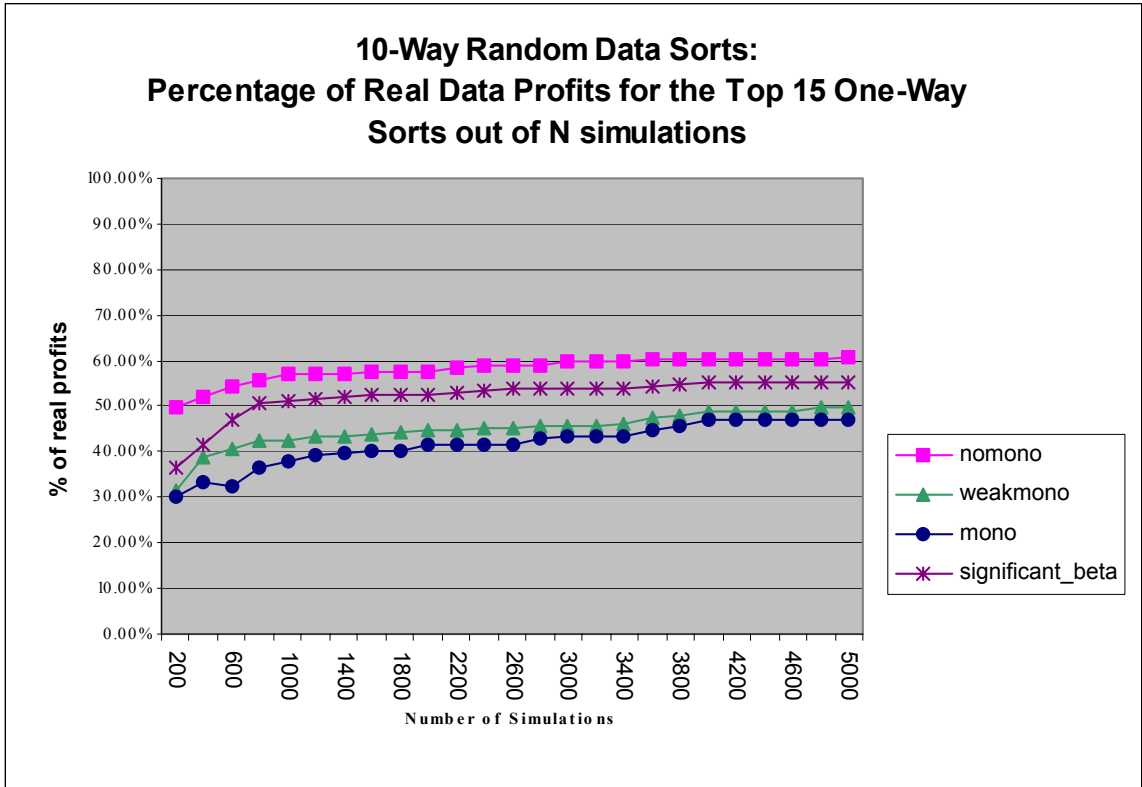The x-axis is the number of runs generating a random N (0,1) predictive variable. The increments are in 200 runs, ranging from 200 up to 5,000 runs. Panels A - D of the graph show the percentage of profits (i.e., 0.60 = 60%) obtained for the average of the combined portfolios for the "best" 15 random factors as a percentage of the average combined real profits found in the 5-, 10-, 30- and 50-portfolio one-way trading strategies reported in Table I. For the random data portfolios, "best" is defined in terms of four unique measures: "Nomono" profits are calculated as the difference in returns to the high and low portfolios where these portfolios are chosen based on ex post performance, so that they could be any portfolio of the firm characteristic sort; "Weakmono" profits are calculated as the difference in returns to the high and low portfolios where these portfolios are chosen, for example, for the 10-way sorts if they belong to deciles (1, 2) or (9, 10) of the firm characteristic; "Mono" profits are calculated as the difference in returns to the high and low portfolios where these portfolios are chosen if they belong to deciles 1 or 10 of the firm characteristic; and "Significant_beta" profits are calculated as the difference in returns to the high and low portfolios where these portfolios are chosen if the underlying firm characteristic was found to have significant cross-sectional beta (t-stat >1.96) in a prior regression of returns on the characteristic. For the Significant_beta category, profits are calculated as the difference in returns to the high and low portfolios where these portfolios are chosen based on ex post performance, so that they could be any deciles of the firm characteristic.
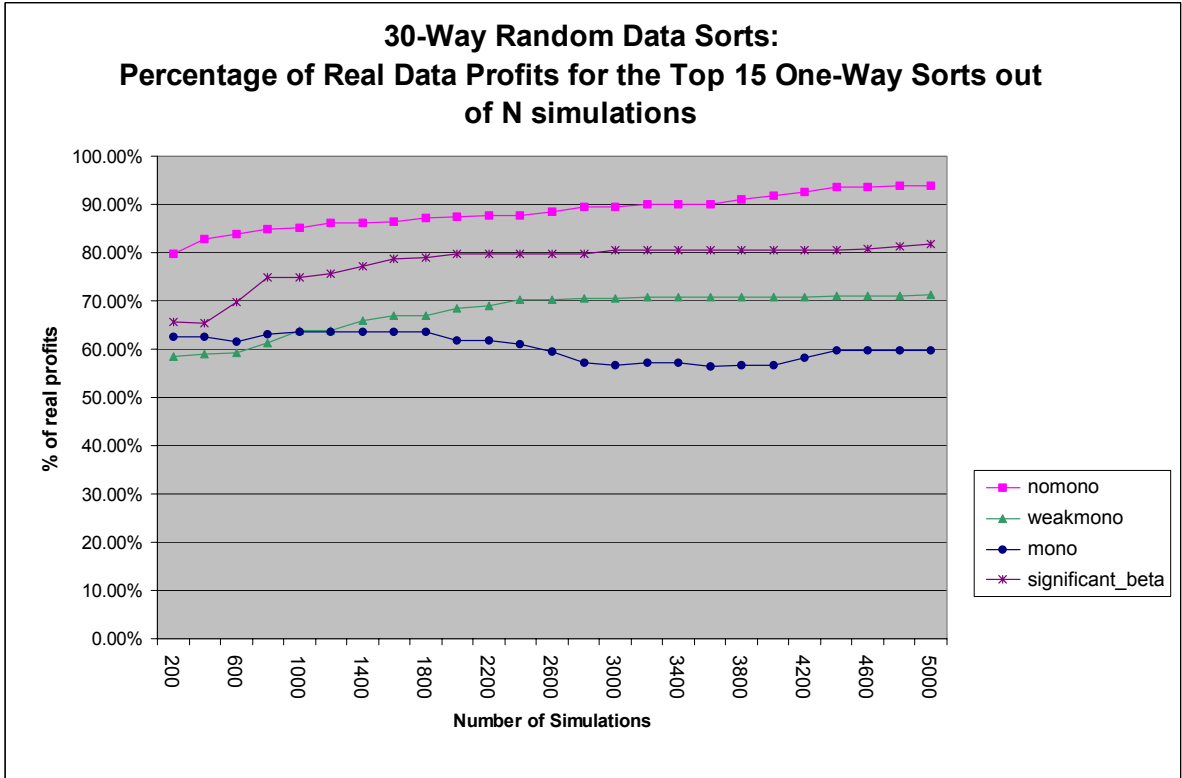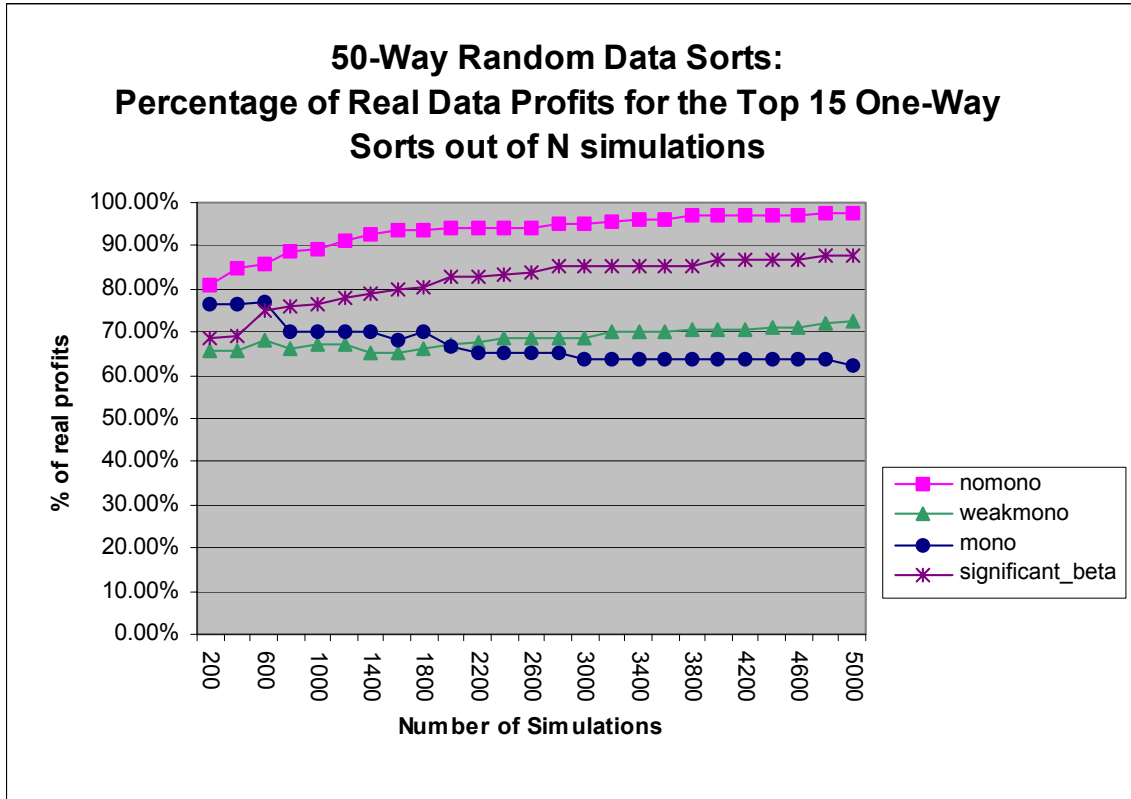
PANEL A

5-Way Random Data Sorts:
Percentage of Real Data Profits for the Top 15 One-Way Sorts
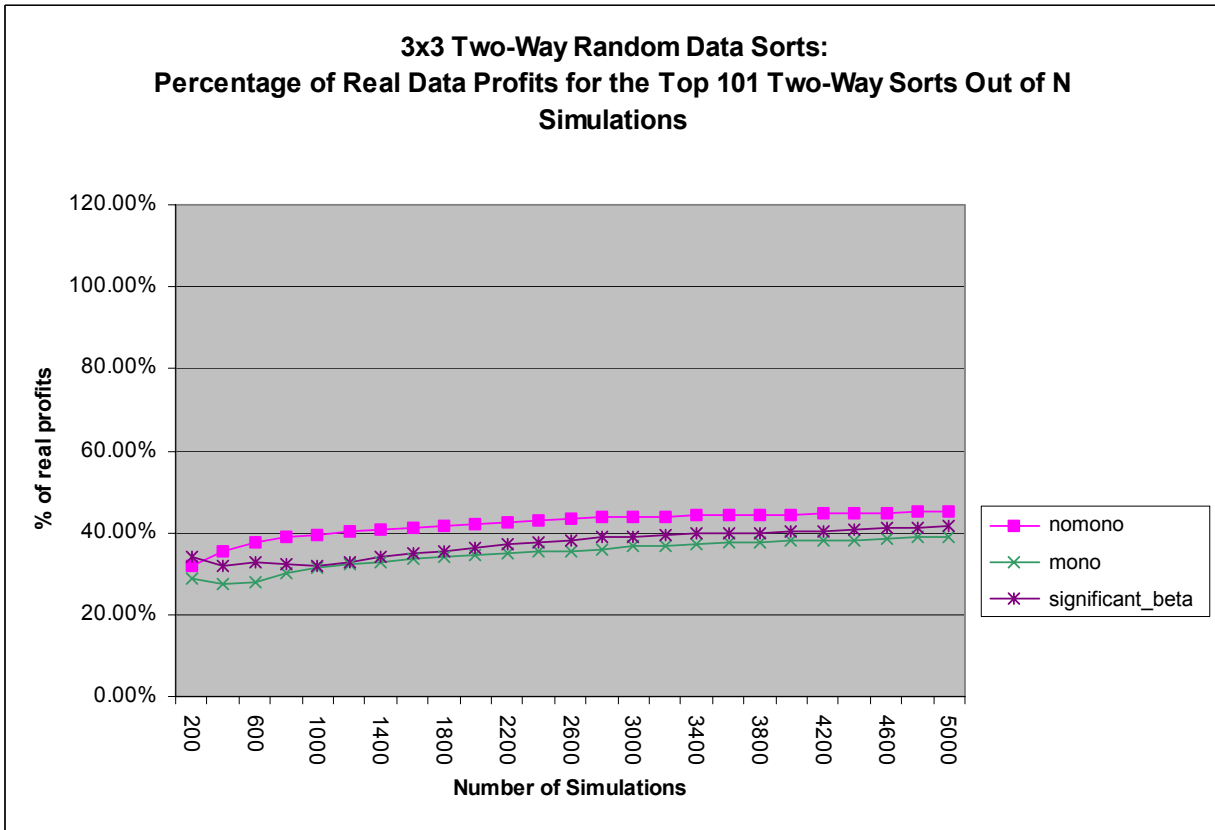**out of N simulations**

PANEL B

10-Way Random Data Sorts:
Percentage of Real Data Profits for the Top 15 One-Way
Sorts out of N simulations

PANEL C

30-Way Random Data Sorts:
Percentage of Real Data Profits for the Top 15 One-Way Sorts out
of N simulations

**50-Way Random Data Sorts:
Percentage of Real Data Profits for the Top 15 One-Way
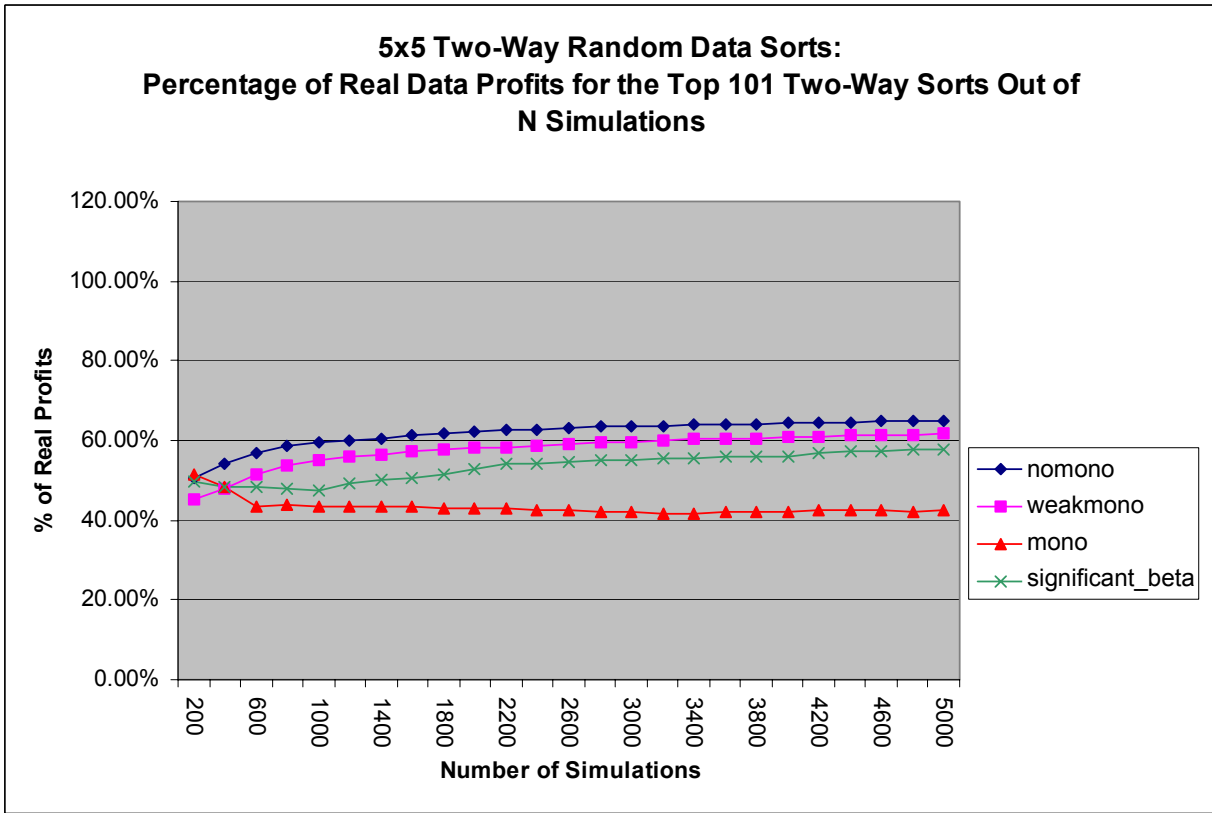Sorts out of N simulations**

FIGURE 2

Profits to Two-Way Trading Strategies Implemented on Randomized Returns

The x-axis is the number of runs generating a random N (0,1) predictive variable. The increments are in 200 runs, ranging from 200 up to 5,000 runs. Panels A - D of the graph shows the percentage of profits obtained for the average of the combined portfolios for the "best" 15 random factors (or the 101 two-way combinations thereof) as a percentage of the average combined real profits found in the 3x3 to 10x10 two-way trading strategies reported in Table II. For the random data portfolios, "best" is defined in terms of four unique measures: "Nomono" profits are calculated as the difference in returns to the high and low portfolios where these portfolios are chosen based on ex post performance, so that they could be any portfolio of the firm characteristic sort; "Weakmono" profits, not reported for the 3x3 sorts, are calculated as the difference in returns to the high and low portfolios where these portfolios are chosen if they belong to an extreme corner portfolio, or the portfolio adjacent to an extreme corner portfolio; "Mono" profits are calculated as the difference in returns to the high and low portfolios where these portfolios are chosen if they belong to an extreme corner portfolio; and "Significant_beta" profits are calculated as the difference in returns to the high and low portfolios where these portfolios are chosen if the underlying firm characteristic was found to have significant cross-sectional beta (t-stat >1.96) in a prior regression of returns on the characteristic. For the Significant_beta category, profits are calculated as the difference in returns to the high and low portfolios where these portfolios are chosen based on ex post performance, so that they could be any portfolio of the firm characteristic. At the lower number of runs and higher number of portfolios, 101 combinations do not always meet the definitions of "best." For example, for the 10x10 sorts and the "Weakmono" classification, we find five significant cases at 200 runs, 24 cases at 1000 runs, 52 cases at 200 runs, and 101 cases at 4000 and above. For "Mono," we find one case at 200 runs, four cases at 1000 runs, and 12 cases 4000 runs and above. For significant beta, we find 19 cases at 200 runs, 94 cases at 1000 runs, and 101 cases at 1200 runs and above.
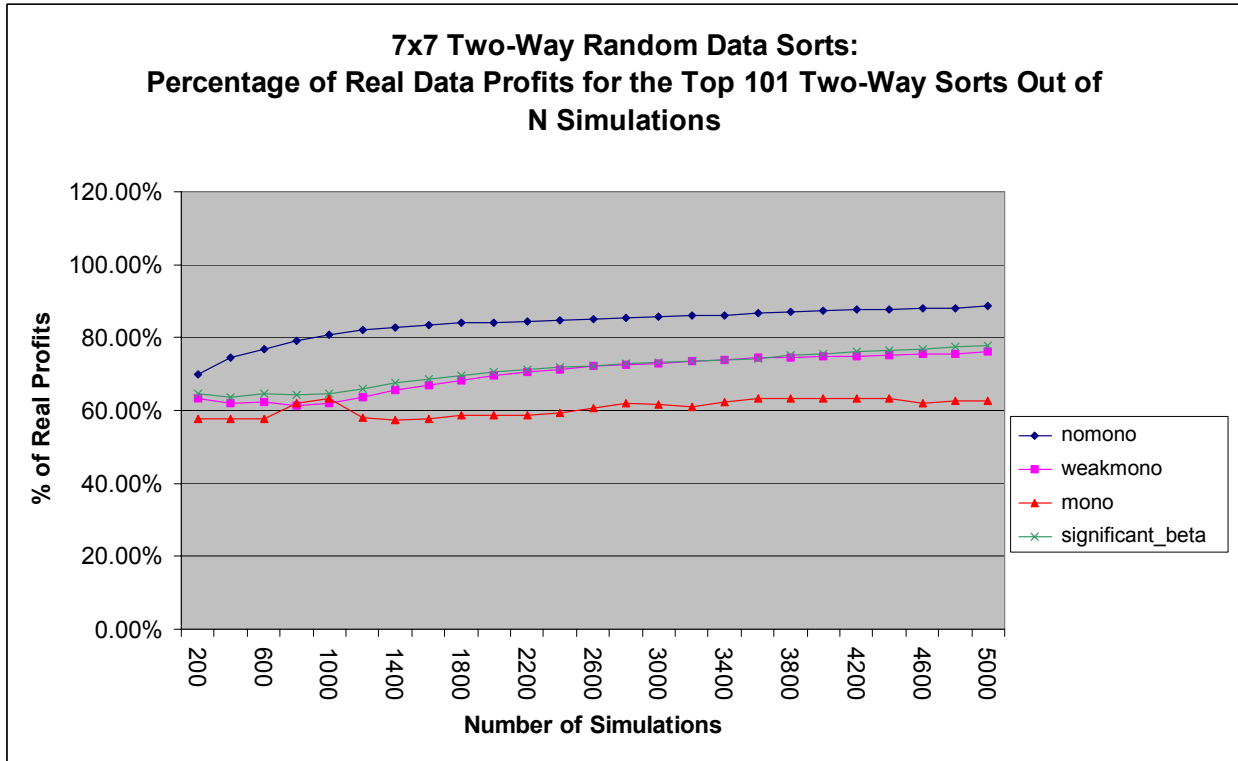
PANEL A

3x3 Two-Way Random Data Sorts:
Percentage of Real Data Profits for the Top 101 Two-Way Sorts Out of N
Simulations

PANEL B

5x5 Two-Way Random Data Sorts:
Percentage of Real Data Profits for the Top 101 Two-Way Sorts Out of
N Simulations

PANEL C

7x7 Two-Way Random Data Sorts:
Percentage of Real Data Profits for the Top 101 Two-Way Sorts Out of
N Simulations

PANEL D

10x10 Two-Way Random Data Sorts:
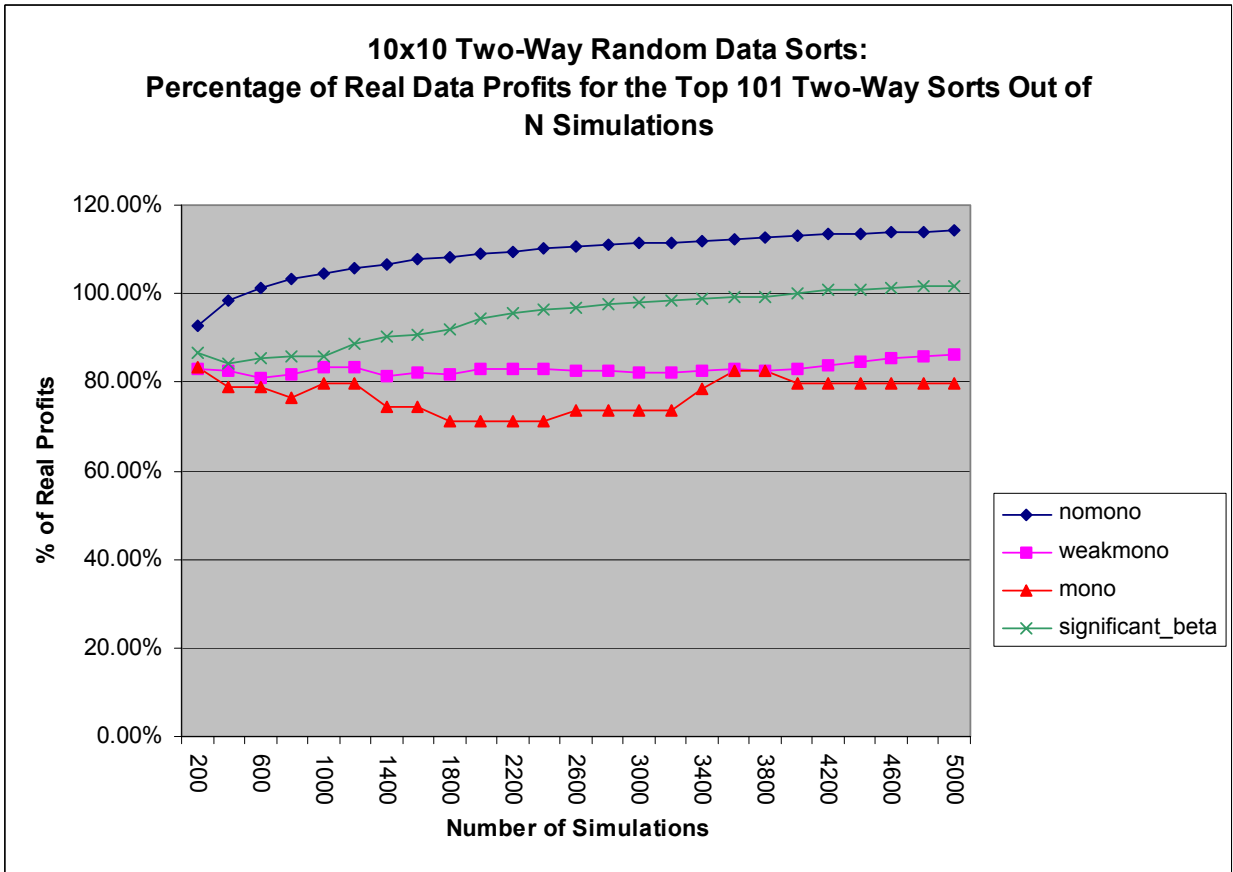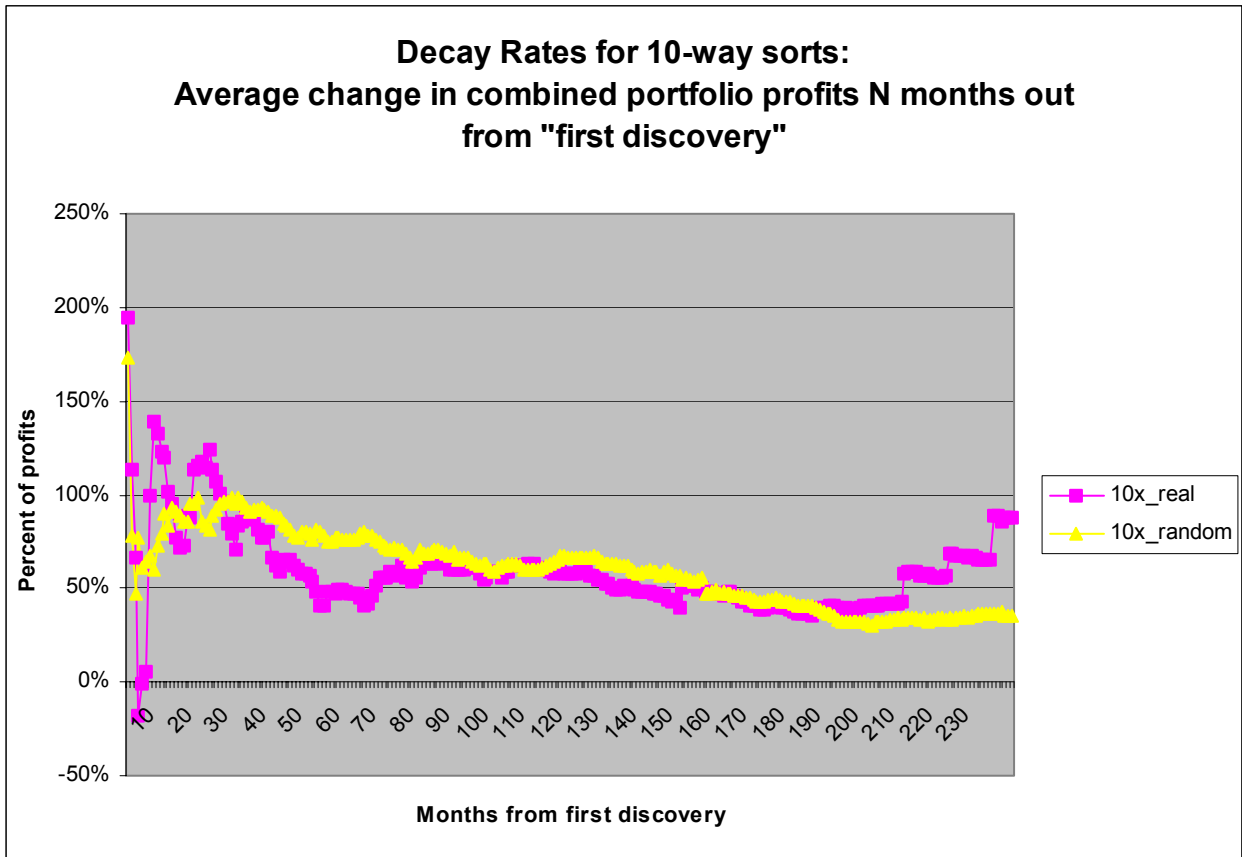Percentage of Real Data Profits for the Top 101 Two-Way Sorts Out of N Simulations
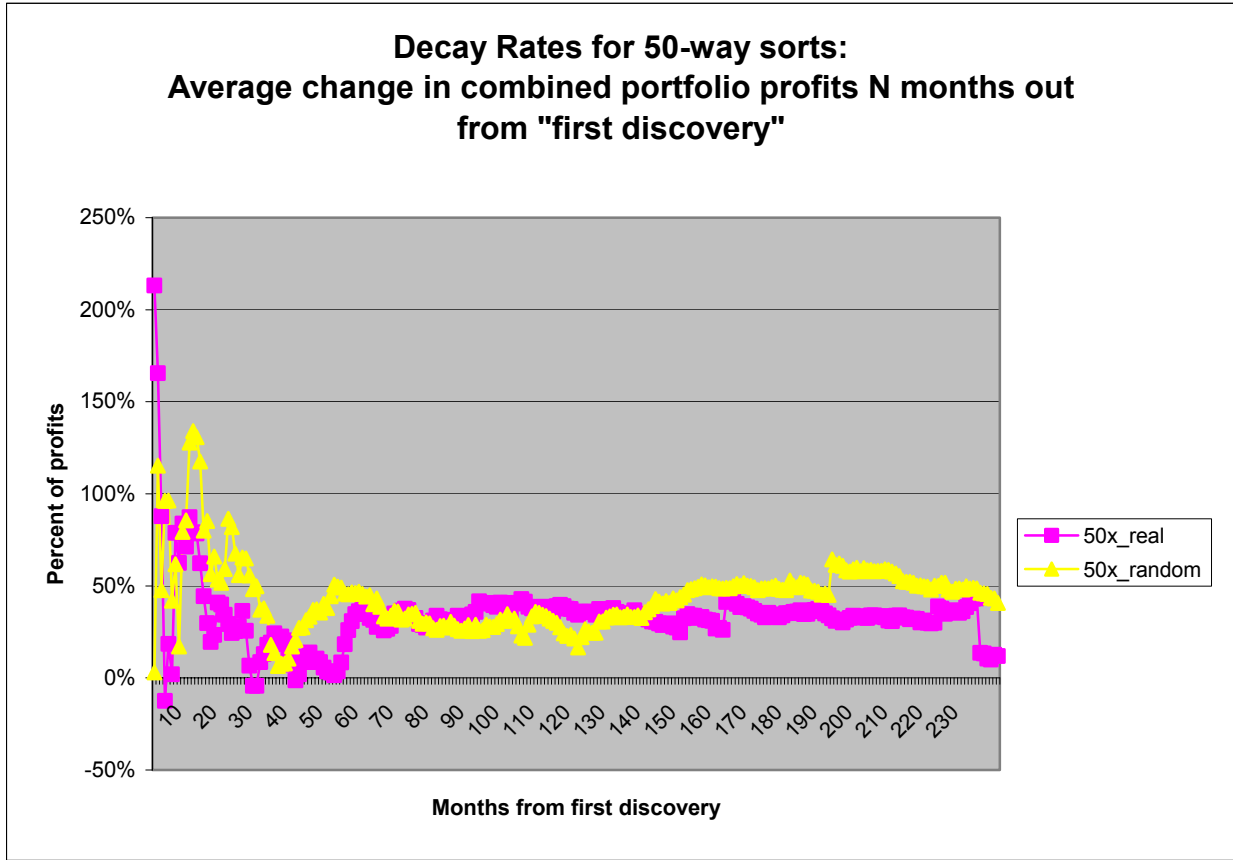
FIGURE 3

Decay Rate Event Study for Real and Random Data Sorts

This figure contains decay rates for combined portfolio sorts based on real and random firm characteristics for one-way 10- and 50-portfolio sorts, while panels C and D contain decay rates for 3x3- and 7x7-portfolio two-way sorts. The decay rate is derived from the combined portfolio profits between the first significant period for a sort variable and the performance of that variable thereafter. Specifically, for each of the 15 real firm characteristics (or the 101 two-way combinations thereof), we form combined portfolios starting in 1974 (to allow for an adequate time-series of data for each real characteristic), using all past available data. For the random sorts, we carry out the same analysis on each of the best 15 random variables (or best 101 two-way random variable combinations) chosen from the 5000 simulations. We restrict the combined portfolios to be formed using only the extreme portfolios, thus imposing monotonicity. For each variable, we roll through the data in yearly increments using an expanding window, and keep track of the first occurrence of a significant combined portfolio (t-statistic greater than 1.96). We then track for each variable what the average profits are in the years after first discovery. For each sort variable, we compute a decay rate equal to the average profits over a particular period after first discovery relative to the average profits during the discovery period. Thus, a rate of 100% means that a variable experiences no drop in average profits after first discovery, a rate of zero implies that the profits drop to zero, and a rate of −100% means the relationship between the firm characteristic and returns is completely reversed. We line up the decay rates across variables using the date of first discovery as the first observation. We then compute the average decay rate across characteristics for an "event window" that runs 240 months out from date of first discovery.
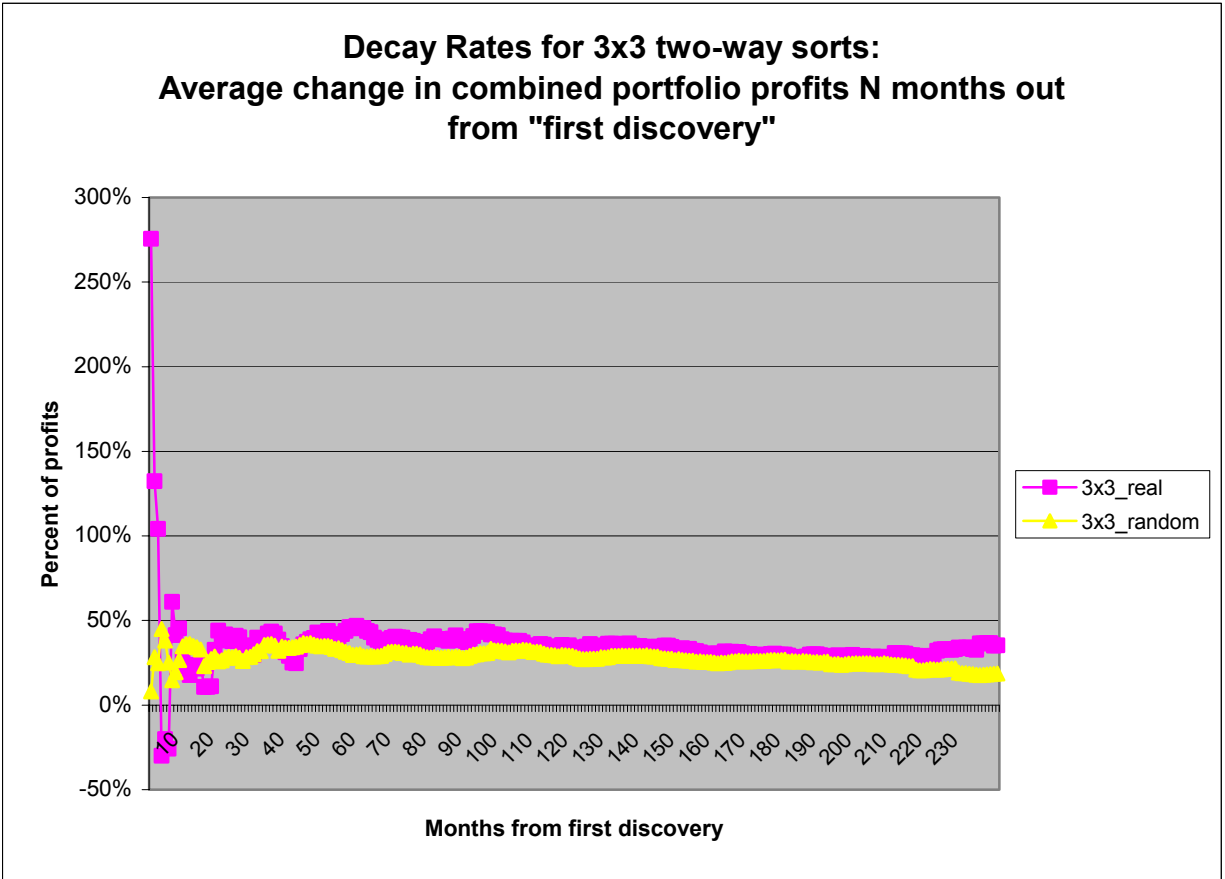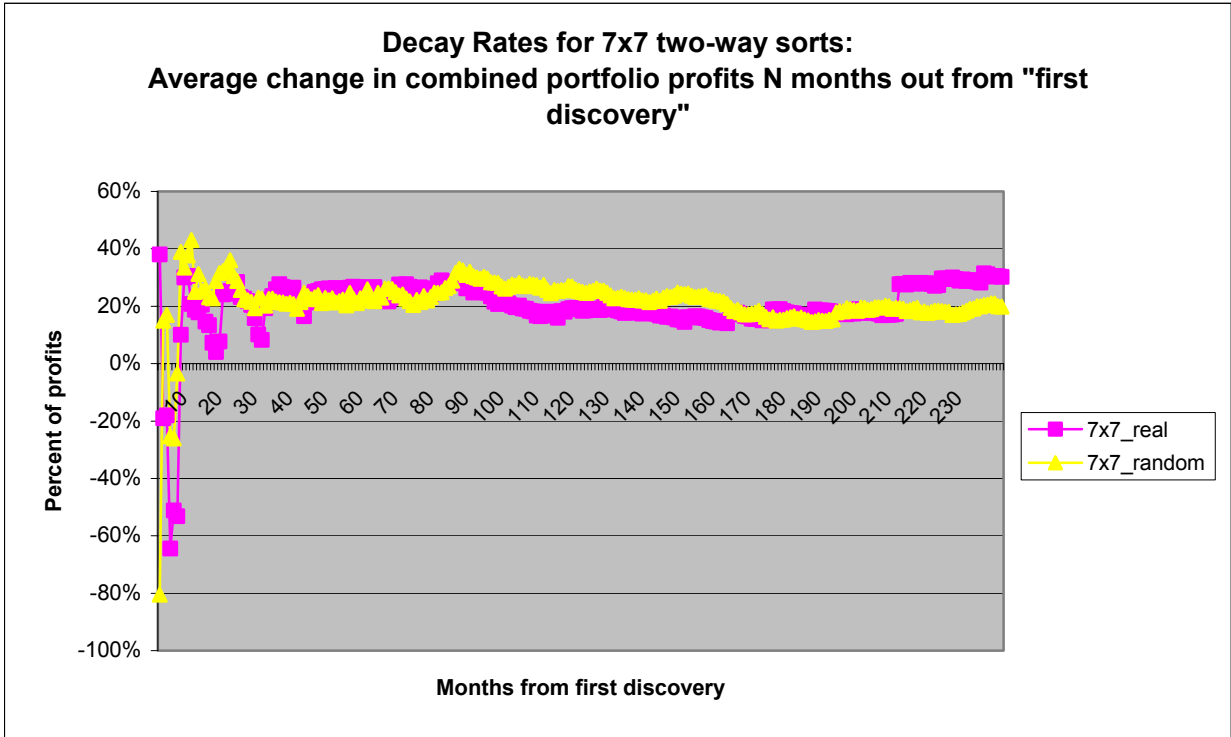
PANEL A

Decay Rates for 10-way sorts:
Average change in combined portfolio profits N months out
from "first discovery"

PANEL B

Decay Rates for 50-way sorts:
Average change in combined portfolio profits N months out from "first discovery"

Legend:
- 50x_real
- 50x_random

Y-axis: Percent of profits (−50% to 250%)
X-axis: Months from first discovery (10 to 230)

Decay Rates for 3x3 two-way sorts:
Average change in combined portfolio profits N months out
from "first discovery"

Decay Rates for 7x7 two-way sorts:
Average change in combined portfolio profits N months out from "first discovery"

FOOTNOTES

1. A *partial* list of papers in this literature includes Ball (1978), Banz (1981), Basu (1983), Bhandari (1988), Breen and Korajczyk (1995), Capaul, Rowley, Sharpe (1993), Chan, Jegadeesh, Lakonishok (1995), Chan, Karceski, and Lakonishok (1998), Chan, Hamao, and Lakonishok (1991), Chen, Roll, and Ross (1983), Claessens, Dasgupta, and Glen (1996), Daniel and Titman (1997), Davis (1994), Fama and French (1992, 1993, 1995, 1996, 1998), Jaffee, Keim, and Westerfield (1989), Kothari and Shanken (1997), Kothari, Shanken, and Sloan (1995), Keim (1983), Lakonishok, Shleifer, and Vishny (1994), La Porta (1993), Reinganum (1981), Rosenberg, Reid, and Lanstein (1985), and Sharpe (1982).We distinguish between papers that document the relation(s) between expected returns and cross-sectional variables, and direct tests of the CAPM/APT [see, for example, Black, Jensen, and Scholes (1972), Connor and Korajczyk (1986), Fama and MacBeth (1973), Gibbons (1982), Roll and Ross (1980), and Stambaugh (1982).]

2. While two-way sorts have become the norm, even three-way sorts have been used in the literature [see, for example, Chan, Hamao, and Lakonishok (1991) and Daniel and Titman (1997)].

3. Black (1993a, b), Denton (1985), Lo and MacKinlay (1990), Foster, Smith and Whaley (1997), Ross (1987), and Sullivan, Timmermann, and White (1999) also emphasize how we (usually out of sheer necessity) *collectively* condition our studies on existing empirical regularities with the unintended consequence of snooping the data.

4. We omit four combinations of variables, one for each of the cases where a time-series variable is matched with its own cross-section.

5. It is quite common to observe different sorting choices by different researchers. The number of portfolios that firms are sorted into typically range between 5-to-10 (see any study based on size or book-to-market sorts) using one-way sorts, to as many as 100 portfolios based on two-way sorts [see, for example, the seminal paper by Fama and French (1992), and also Banz (1981), Basu (1983), Daniel and Titman (1997), Fama and French (1995), Lakonishok, Shleifer, and Vishny (1994) and numerous others].

6. For our sorts, we require all returns in our sample to have non-missing values of the sorting variables in the Appendix. We do this to ensure that the unconditional return distributions of the real and random sorts are identical.

7. Virtually every study indulges in some unavoidable data snooping—certainly the authors of this study have (profitably) conducted analyses on portfolios sorted by size and price, for example.

8. Note that most of this increase occurs by about 1,200 searches, with the estimates ranging between 33 and 43 percent. After this point, there is minimal "advantage" from more familiarity with the data.

9. The differences in average profits across one- and two-way sorts (compare Tables I and II) are not large, and the variation in profits across strategies is smaller for two-way relative to one-way sorts. This suggests that two-way combinations of the characteristics do not provide substantially more information about cross-sectional differences in returns than the one-way sorts. This may be due to the fact that the firm characteristics are correlated. While the average correlation is only 0.126, the correlations in the tails of the distribution are large, 0.779 and -0.349.

10. Pair-wise combinations of the 15 real variables generate the 105 two-way combinations: (15*(15-1))/2 pair-wise combinations, less the four combinations where we sort on both the time-series and the cross-sectional counterparts of the same characteristic (because these sorts have virtually no observations).

11. Cooper, Gutierrez, and Marcum (2001) further explore such "real-time" issues inherent in out-of-sample tests by requiring the investor to endogenously determine in-sample the optimal predictor variables, rules relating those variables to future returns, and the dimensionality of the sort. Once they endogenize these portfolio investment parameters, it is difficult for an investor to outperform a passive buy-and-hold benchmark portfolio.

12. We thank the referee for this suggestion.

13. We start in 1974 to allow for adequate (up to ten years' worth) data for all the real firm characteristics.

14. If a given variable is not significant until later in the sample, then it will not have observations for all 240 months.

15. Of course, one possibility for why we observe worse decay rates in the real data is that investors are acting opportunistically to take advantage of any profit opportunities. Our method of comparing decay rates in real and simulated profits does not have the flexibility to distinguish between fleeting, but genuine, profit opportunities and (extreme) draws from a distribution centered around zero. An alternative approach, based on order statistics, could potentially be used to make such a distinction.

16.	Note that the decay rate analysis is conducted under the assumption that 5,000 searches are conducted to discover the 15 firm characteristics used. It is important to note however that in our in-sample analysis about 1,000 searches were sufficient to generate most of the in-sample profitability observed in the real data.

17.	It is fairly easy to show however that for reasonably high correlations across stock markets, the probability of finding similar firm characteristics that can predict returns increases.

18.	In related work, Ferson, Sarkissian, and Simin (1999) show that attribute-sorted portfolios of common stocks will appear to be "risk factors" if the attributes are chosen following an empirically observed relation to the cross-section of stock returns.