NBER WORKING PAPER SERIES

# . . . AND THE CROSS-SECTION OF EXPECTED RETURNS

Campbell R. Harvey
Yan Liu
Heqing Zhu

Working Paper 20592
http://www.nber.org/papers/w20592

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2014

We appreciate the comments of Viral Acharya, Jawad Addoum, Tobias Adrian, Andrew Ang, Ravi Bansal, Mehmet Beceren, Itzhak Ben-David, Bernard Black, Jules van Binsbergen, Oliver Boguth, Tim Bollerslev, Alon Brav, Ian Dew-Becker, Robert Dittmar, Jennifer Conrad, Michael Cooper, Andres Donangelo, Wayne Ferson, Simon Gervais, Bing Han, John Hand, Andrew Karolyi, Abby Yeon Kyeong Kim, Lars-Alexander Kuehn, Sophia Li, Harry Markowitz, Kyle Matoba, David McLean, Marcelo Ochoa, Peter Park, Lubos Pastor, Andrew Patton, Lasse Heje Pedersen, Tapio Pekkala, Jeff Pontiff, Ryan Pratt, Alexandru Rosoiu, Tim Simin, Avanidhar Subrahmanyam, Ivo Welch, Basil Williams, Yuhang Xing, Josef Zechner and Xiaofei Zhao as well as seminar participants at the 2014 New Frontiers in Finance Conference at Vanderbilt University, the 2014 Inquire Europe-UK meeting in Vienna, the 2014 WFA meetings, and seminars at Duke University, University of Utah, and Penn State university. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

. . . and the Cross-Section of Expected Returns

Campbell R. Harvey, Yan Liu, and Heqing Zhu

NBER Working Paper No. 20592
October 2014
JEL No. C01,C58,G0,G1,G12,G3

## ABSTRACT

Hundreds of papers and hundreds of factors attempt to explain the cross-section of expected returns. Given this extensive data mining, it does not make any economic or statistical sense to use the usual significance criteria for a newly discovered factor, e.g., a t-ratio greater than 2.0. However, what hurdle should be used for current research? Our paper introduces a multiple testing framework and provides a time series of historical significance cutoffs from the first empirical tests in 1967 to today. Our new method allows for correlation among the tests as well as missing data. We also project forward 20 years assuming the rate of factor production remains similar to the experience of the last few years. The estimation of our model suggests that a newly discovered factor needs to clear a much higher hurdle, with a t-ratio greater than 3.0. Echoing a recent disturbing conclusion in the medical literature, we argue that most claimed research findings in financial economics are likely false.

Campbell R. Harvey
Duke University
Fuqua School of Business
Durham, NC 27708-0120
and NBER
cam.harvey@duke.edu

Yan Liu
Texas A&M University,
College Station, TX 77843
y-liu@mays.tamu.edu

Heqing Zhu
University of Oklahoma
Norman, OK 73019 USA
carolinehzhu@ou.edu

# 1 Introduction

Forty years ago, one of the first tests of the Capital Asset Pricing Model (CAPM) found that the market beta was a significant explanator of the cross-section of expected returns. The reported t-ratio of 2.57 in Fama and MacBeth (1973) comfortably exceeded the usual cutoff of 2.0. However, since that time, hundreds of papers have tried to explain the cross-section of expected returns. Given the known number of factors that have been tried and the reasonable assumption that many more factors have been tried but did not make it to publication, the usual cutoff levels for statistical significance are not appropriate. We present a new framework that allows for multiple tests and derive recommended statistical significance levels for current research in asset pricing.

We begin with 313 papers that study cross-sectional return patterns published in a selection of journals. We provide recommended p-values from the first empirical tests in 1967 through to present day. We also project minimum t-ratios through 2032 assuming the rate of "factor production" remains similar to the recent experience. We present a taxonomy of historical factors as well as definitions.[1]

Our research is related to a recent paper by McLean and Pontiff (2014) who argue that certain stock market anomalies are less anomalous after being published.[2] Their paper tests the statistical biases emphasized in Leamer (1978), Ross (1989), Lo and MacKinlay (1990), Fama (1991) and Schwert (2003).

Our paper also adds to the recent literature on biases and inefficiencies in cross-sectional regression studies. Lewellen, Nagel and Shanken (2010) critique the usual practice of using cross-sectional $R^2$s and pricing errors to judge the success of a work and show that the explanatory powers of many previously documented factors are spurious.[3] Balduzzi and Robotti (2008) challenge the traditional approach of estimating factor risk premia via cross-sectional regressions and advocate a factor projection approach. Our work focuses on evaluating the statistical significance of a factor given the previous tests on other factors. Our goal is to use a multiple testing framework to both re-evaluate past research and to provide a new benchmark for current and future research.

We tackle multiple hypothesis testing from the frequentist perspective. Bayesian approaches on multiple testing and variable selection also exist. However, the high dimensionality of the problem combined with the fact that we do not observe all the factors that have been tried poses a big challenge for Bayesian methods. While we

[1] We also provide a link to a file with full references and hyperlinks to the original articles: *http://faculty.fuqua.duke.edu/~charvey/Factor-List.xlsx.*

[2] Other recent papers that systematically study the cross-sectional return patterns include Subrahmanyam (2010), Green, Hand and Zhang (2012, 2013).

[3] A related work by Daniel and Titman (2012) constructs more powerful statistical tests and rejects several recently proposed factor models.

propose a frequentist approach to overcome this missing data issue, it is unclear how to do this in the Bayesian framework. Nonetheless, we provide a detailed discussion of Bayesian methods in the paper.

There are limitations to our framework. First, should all factor discoveries be treated equally? We think no. A factor derived from a theory should have a lower hurdle than a factor discovered from a purely empirical exercise. Nevertheless, whether suggested by theory or empirical work, a t-ratio of 2.0 is too low. Second, our tests focus on unconditional tests. It is possible that a particular factor is very important in certain economic environments and not important in other environments. The unconditional test might conclude the factor is marginal. These two caveats need to be taken into account when using our recommended significance levels for current asset pricing research.

While our focus is on the cross-section of equity returns, our message applies to many different areas of finance. For instance, Frank and Goyal (2009) investigate around 30 variables that have been documented to explain capital structure decisions of public firms. Welch and Goyal (2004) examine the performance of a dozen variables that have been shown to predict market excess returns. These two applications are ideal settings to employ multiple testing methods.[4]

Our paper is organized as follows. In the second section, we provide a chronology of the "discovered" factors. The third section presents a categorization of the factors. Next, we introduce some multiple testing frameworks and suggest appropriate cutoffs for both past and future asset pricing tests. Some concluding remarks are offered in the final section.

## 2 The Search Process

Our goal is not to catalogue every asset pricing paper ever published. We narrow the focus to papers that propose and test new factors. For example, Sharpe (1964), Lintner (1965) and Mossin (1966) all theoretically proposed (at roughly the same time), a single factor model — the Capital Asset Pricing Model (CAPM). Beginning with Black, Jensen and Scholes (1972), there are hundreds of papers that test the CAPM. We include the theoretical papers as well as the first paper to empirically test the model, in this case, Black, Jensen and Scholes (1972). We do not include the hundreds of papers that test the CAPM in different contexts, e.g., international markets, different time periods. We do, however, include papers, such as Fama and MacBeth (1973) which tests the market factor as well as two additional factors.

Sometimes different papers propose different empirical proxies for the same type of economic risk. Although they may look similar from a theoretical standpoint, we still include them. An example is the empirical proxies for idiosyncratic financial constraints risk. While Lamont, Polk and Saa-Requejo (2001) use the Kaplan and Zingales (1997) index to proxy for firm-level financial constraint, Whited and Wu

[4]Harvey and Liu (2014a) show how to adjust Sharpe Ratios used in performance evaluation for multiple tests.

(2006) estimate their own constraint index based on the first order conditions of firms' optimization problem. We include both even though they are likely highly correlated.

Since our focus is on factors that can broadly explain asset market return patterns, we omit papers that focus on a small group of stocks or for a short period of time. This will, for example, exclude a substantial amount of empirical corporate finance research that studies event-driven return movements.

Certain theoretical models lack immediate empirical content. Although they could be empirically relevant once suitable proxies are constructed, we choose to exclude them.

With these rules in mind, we narrow our search to generally the top journals in finance, economics and accounting. To include the most recent research, we search for working papers on SSRN. Working papers pose a challenge because there are thousands of them and they are not refereed. We choose a subset of papers that we suspect are in review at top journals or have been presented at top conferences or are due to be presented at top conferences. We end up using 63 working papers. In total, we focus on 313 published works and selected working papers. We catalogue 316 different factors.[5]

Our collection of 316 factors likely under-represents the factor population. First, we generally only consider top journals. Second, we are very selective in choosing only a handful of working papers. Third, and perhaps most importantly, we should be measuring the number of factors tested (which is unobservable) — that is, we do not observe the factors that were tested but failed to pass the usual significance levels and were never published (see Fama (1991)).

# 3 Factor Taxonomy

To facilitate our analysis, we group the factors into different categories. We start with two broad categories: "common" and "individual". "Common" means the factor can be viewed as a proxy for a common source of risk. Risk exposure to this factor or its innovations is supposed to help explain cross-sectional return patterns. "Individual" means the factor is specific to the security or portfolio. A good example is Fama and MacBeth (1973). While the beta against the market return is systematic (exposure to a common risk factor), the standard deviation of the market model residual is security specific and hence an idiosyncratic or individual risk. Many of the individual factors we identify are referred to in the literature as "characteristics".

Based on the unique properties of the proposed factors, we further divide the "common" and "individual" groups into finer categories. In particular, we divide "common" into: "financial", "macro", "microstructure", "behavioral", "accounting" and "other". We divide "individual" into the same categories — except we omit the

[5]As already mentioned, some of these factors are highly correlated. For example, we include two versions of idiosyncratic volatility — where the residual is defined by different time-series regressions.

"macro" classification, which is common, by definition. The following table provides further details on the definitions of these sub-categories and gives examples for each.

Table 1: **Factor Classification**

| Risk type | | Description | Examples |
|---|---|---|---|
| **Common** (113) | **Financial** (46) | Proxy for aggregate financial market movement, including market portfolio returns, volatility, squared market returns, etc. | Sharpe (1964): market returns; Kraus and Litzenberger (1976): squared market returns |
| | **Macro** (40) | Proxy for movement in macroeconomic fundamentals, including consumption, investment, inflation, etc. | Breeden (1979): consumption growth; Cochrane (1991): investment returns |
| | **Microstructure** (11) | Proxy for aggregate movements in market microstructure or financial market frictions, including liquidity, transaction costs, etc. | Pastor and Stambaugh (2003): market liquidity; Lo and Wang (2006): market trading volume |
| | **Behavioral** (3) | Proxy for aggregate movements in investor behavior, sentiment or behavior-driven systematic mispricing | Baker and Wurgler (2006): investor sentiment; Hirshleifer and Jiang (2010): market mispricing |
| | **Accounting** (8) | Proxy for aggregate movement in firm-level accounting variables, including payout yield, cash flow, etc. | Fama and French (1992): size and book-to-market; Da and Warachka (2009): cash flow |
| | **Other** (5) | Proxy for aggregate movements that do not fall into the above categories, including momentum, investors' beliefs, etc. | Carhart (1997): return momentum; Ozoguz (2008): investors' beliefs |
| **Individual** (202) | **Financial** (61) | Proxy for firm-level idiosyncratic financial risks, including volatility, extreme returns, etc. | Ang, Hodrick, Xing and Zhang (2006): idiosyncratic volatility; Bali, Cakici and Whitelaw (2011): extreme stock returns |
| | **Microstructure** (28) | Proxy for firm-level financial market frictions, including short sale restrictions, transaction costs, etc. | Jarrow (1980): short sale restrictions; Mayshar (1981): transaction costs |
| | **Behavioral** (3) | Proxy for firm-level behavioral biases, including analyst dispersion, media coverage, etc. | Diether, Malloy and Scherbina (2002): analyst dispersion; Fang and Peress (2009): media coverage |
| | **Accounting** (87) | Proxy for firm-level accounting variables, including PE ratio, debt to equity ratio, etc. | Basu (1977): PE ratio; Bhandari (1988): debt to equity ratio |
| | **Other** (24) | Proxy for firm-level variables that do not fall into the above categories, including political campaign contributions, ranking-related firm intangibles, etc. | Cooper, Gulen and Ovtchinnikov (2010): political campaign contributions; Edmans (2011): intangibles |

Numbers in parentheses represent the number of factors identified. See Table 5 for details.

# 4 Adjusted T-ratios in Multiple Testing

## 4.1 Why Multiple Testing?

Given so many papers have attempted to explain the same cross-section of expected returns,[6] statistical inference should not be based on a "single" test perspective.[7] Our goal is to provide guidance as to the appropriate significance level using a multiple testing framework.

We want to emphasize that there are many forces that make our guidance lenient, that is, a credible case can be made for even lower p-values. We have already mentioned that we only sample a subset of research papers and the "publication bias/hidden tests" issue (i.e. it is difficult to publish a non-result).[8] However, there is another publication bias that is more subtle. In many scientific fields, replication studies routinely appear in top journals. That is, a factor is discovered, and others try to replicate it. In finance and economics, it is very difficult to publish replication studies. Hence, there is a bias towards publishing "new" factors rather than rigorously verifying the existence of discovered factors.

There are two ways to deal with the bias introduced by multiple testing: out-of-sample validation and using a statistical framework that allows for multiple testing.[9] When feasible, out-of-sample testing is the cleanest way to rule out spurious factors. In their study of anomalies, McLean and Pontiff (2014) take the out-of-sample approach. Their results show a degradation of performance of identified anomalies after publication which is consistent with the statistical bias. It is possible that this degradation is larger than they document. In particular, they drop 10 of their 82 anomalies because they could not replicate the in-sample performance of published studies. Given these non-replicable anomalies were not even able to survive routine data revisions, they are likely to be insignificant strategies, either in-sample or out-of-sample. The degradation from the original published "alpha" is 100% for these strategies — which would lead to a higher average rate of degradation for the 82 strategies.

[6] Strictly speaking, different papers study different sample periods and hence focus on "different" cross-sections of expected returns. However, the bulk of the papers we consider have substantial overlapping sample periods. Also, if one believes that cross-sectional return patterns are stationary, then these papers are studying roughly the same cross-section of expected returns.

[7] When just one hypothesis is tested, we use the term "individual test", "single test" and "independent test" interchangeably. The last term should not be confused with any sort of stochastic independence.

[8] See Rosenthal (1979) for one of the earliest and most influential works on publication bias.

[9] Another approach to test factor robustness is to look at multiple asset classes. This approach has been followed in several recent papers, e.g., Frazzini and Pedersen (2012) and Koijen, Moskowitz, Pedersen and Vrugt (2012).

While the out-of-sample approach has many strengths, it has one important drawback: it cannot be used in real time.[10] In contrast to many tests in the physical sciences, we often need years of data to do an out-of-sample test. We pursue the multiple testing framework because it yields immediate guidance on whether a discovered factor is real.

## 4.2 A Multiple Testing Framework

In statistics, multiple testing refers to simultaneous testing more than one hypothesis. The statistics literature was aware of this multiplicity problem at least 50 years ago.[11] Early generations of multiple testing procedures focus on the control of the *family-wise error rate* (see Section 4.3.1). More recently, increasing interest in multiple testing from the medical literature has spurred the development of methods that control the *false-discovery rate* (see Section 4.3.2). Nowadays, multiple testing is an active research area in both the statistics and the medical literature.[12]

Despite the rapid development of multiple testing methods, they have not attracted much attention in the finance literature.[13] Moreover, most of the research that does involve multiple testing focuses on the Bonferroni adjustment, which is known to be too stringent. Our paper aims to fill this gap by systematically introducing the multiple testing framework.

First, we introduce a hypothetical example to motivate a more general framework. In Table 5, we categorize the possible outcomes of a multiple testing exercise. Panel A displays an example of what the literature could have discovered and Panel B

[10]To make real time assessment in the out-of-sample approach, it is common to hold out some data. However, this is not genuine out-of-sample testing as all the data are observable to researchers. A real out-of-sample test needs data in the future.

[11]For early research on multiple testing, see Scheffé's method (Scheffé (1959)) for adjusting significance levels in a multiple regression context and Tukey's range test (Tukey (1977)) on pairwise mean differences.

[12]See Shaffer (1995) for a review of multiple testing procedures that control for the *family-wise error rate*. See Farcomeni (2008) for a review that focuses on procedures that control the *false-discovery rate*.

[13]For the literature on multiple testing corrections for data snooping biases, see Sullivan, Timmermann and White (1999, 2001) and White (2000). For research on data snooping and variable selection in predictive regressions, see Foster, Smith and Whaley (1997), Cooper and Gulen (2006) and Lynch and Vital-Ahuja (2012). For applications of multiple testing approach in the finance literature, see for example Shanken (1990), Ferson and Harvey (1999), Boudoukh et al. (2007) and Patton and Timmermann (2010). More recently, the *false discovery rate* and its extensions have been used to study technical trading and mutual fund performance, see for example Barras, Scaillet and Wermers (2010), Bajgrowicz and Scaillet (2012) and Kosowski, Timmermann, White and Wermers (2006). Conrad, Cooper and Kaul (2003) point out that data snooping accounts for a large proportion of the return differential between equity portfolios that are sorted by firm characteristics. Bajgrowicz, Scaillet and Treccani (2013) show that multiple testing methods help eliminate a large proportion of spurious jumps detected using conventional test statistics for high-frequency data. Holland, Basu and Sun (2010) emphasize the importance of multiple testing in accounting research.

notationalizes Panel A to ease our subsequent definition of the general Type I error rate — the chance of making at least one false discovery or the expected fraction of false discoveries.

Table 2: **Contingency Table in Testing M Hypotheses.**

Panel A shows a hypothetical example for factor testing. Panel B presents the corresponding notation in a standard multiple testing framework.

Panel A: An Example

| | Unpublished | Published | Total |
|---|---|---|---|
| Truly insignificant | 500 | 50 | 550 |
| Truly significant | 100 | 50 | 150 |
| Total | 600 | 100 | 700 |

Panel B: The Testing Framework

| | $H_0$ not rejected | $H_0$ rejected | Total |
|---|---|---|---|
| $H_0$ True | $N_{0\|a}$ | $N_{0\|r}$ | $M_0$ |
| $H_0$ False | $N_{1\|a}$ | $N_{1\|r}$ | $M_1$ |
| Total | $M-R$ | $R$ | $M$ |

Our example in Panel A assumes 100 published factors (denoted as $R$). Among these factors, suppose 50 are false discoveries and the rest are real ones. In addition, researchers have tried 600 other factors but none of them were found to be significant. Among them, 500 are truly insignificant but the other 100 are true factors. The total number of tests ($M$) is 700. Two types of mistakes are made in this process: 50 factors are falsely discovered to be true while 100 true factors are buried in unpublished work. Usual statistical control in a multiple testing context aims at reducing "50" or "50/100", the absolute or proportionate occurrence of false discoveries, respectively. Of course, we only observe published factors because factors that are tried and found to be insignificant rarely make it to publication.[14] This poses a challenge since the usual statistical techniques only handle the case where all test results are observable.

Panel B defines the corresponding terms in a formal statistical testing framework. In a factor testing exercise, the typical null hypothesis is that a factor is not significant. Therefore, a factor is insignificant means the null hypothesis is "true". Using "0" ("1") to indicate the null is true (false) and "a" ("r") to indicate acceptance (rejection), we can easily summarize Panel A. For instance, $N_{0|r}$ measures the number of rejections

[14]Examples of publication of unsuccessful factors include Fama and MacBeth (1973) and Ferson and Harvey (1993). Fama and MacBeth (1973) show that squared beta and standard deviation of the market model residual have an insignificant role in explaining the cross-section of expected returns. Overall, it is rare to publish "non-results" and all instances of published non-results are coupled with significant results for other factors.

when the null is true (i.e. the number of false discoveries) and $N_{1|a}$ measures the number of acceptances when the null is false (i.e. the number of missed discoveries). To avoid confusion, we try not to use standard statistical language in describing our notation but rather words unique to our factor testing context. The generic notation in Panel B is convenient for us to formally define different types of errors and describe adjustment procedures in subsequent sections.

## 4.3 Type I and Type II Errors

For a single hypothesis test, a value $\alpha$ is used to control Type I error: the probability of finding a factor to be significant when it is not. In a multiple testing framework, restricting each individual test's Type I error rate at $\alpha$ is not enough to control the overall probability of false discoveries. The intuition is that, under the null that all factors are insignificant, it is very likely for an event with $\alpha$ probability to occur when many factors are tested. In multiple hypothesis testing, we need measures of the Type I error that help us simultaneously evaluate the outcomes of many individual tests.

To gain some intuition on plausible measures of Type I error, we return to Panel B of Table 5. $N_{0|r}$ and $N_{1|a}$ count the total number of the two types of errors: $N_{0|r}$ counts false discoveries while $N_{1|a}$ counts missed discoveries. As generalized from single hypothesis testing, the Type I error in multiple hypothesis testing is also related to false discoveries — concluding a factor is "significant" when it is not. But, by definition, we must draw several conclusions in multiple hypothesis testing, and there is a possible false discovery for each. Therefore, plausible definitions of the Type I error should take into account the joint occurrence of false discoveries.

The literature has adopted at least two ways of summarizing the "joint occurrence". One approach is to count the total number of occurrences $N_{0|r}$. $N_{0|r}$ greater than zero suggests incorrect statistical inference for the overall multiple testing problem — the occurrence of which we should limit. Therefore, the probability of event $N_{0|r} > 0$ should be a meaningful quantity for us to control. Indeed, this is the intuition behind the *family-wise error rate* introduced later. On the other hand, when the total number of discoveries $R$ is large, one or even a few false discoveries may be tolerable. In this case, $N_{0|r}$ is no longer a suitable measure; a certain *false discovery proportion* may be more desirable. Unsurprisingly, the expected value of $N_{0|r}/R$ is the focus of *false discovery rate*, the second type of control.

The two aforementioned measures are the most widely used in the statistics literature. Moreover, many other techniques can be viewed as extensions of these measures.[15] We now describe each measure in detail.

[15]Holm (1979) is the first to formally define the *family-wise error rate*. Benjamini and Hochberg (1995) define and study the *false discovery rate*. Alternative definitions of error rate include *per comparison error rate* (Saville, 1990), *positive false discovery rate* (Storey, 2003) and *generalized false discovery rate* (Sarkar and Guo, 2009).

### 4.3.1 Family-wise Error Rate

The *family-wise error rate* (FWER) is the probability of at least one Type I error:

$$\text{FWER} = Pr(N_{0|r} \geq 1).$$

FWER measures the probability of even a single false discovery, irrespective of the total number of tests. For instance, researchers might test 100 factors; FWER measures the probability of incorrectly identifying one or more factors to be significant. Given significance or threshold level $\alpha$, we explore two existing methods (Bonferroni and Holm's adjustment) to ensure FWER does not exceed $\alpha$. Even as the number of trials increases, FWER still measures the probability of a single false discovery. This absolute control is in contrast to the proportionate control afforded by the *false discovery rate* (FDR), defined below.

### 4.3.2 False Discovery Rate

The *false discovery proportion* (FDP) is the proportion of Type I errors:

$$\text{FDP} = \begin{cases} \dfrac{N_{0|r}}{R} & \text{if } R > 0, \\ 0 & \text{if } R = 0. \end{cases}$$

The *false discovery rate* (FDR) is defined as:

$$\text{FDR} = E[FDP].$$

FDR measures the expected proportion of false discoveries among all discoveries. It is less stringent than FWER and usually much less so when many tests are performed.[16]

[16]There is a natural ordering between FDR and FWER. Theoretically, FDR is always bounded above by FWER, i.e., $FDR \leq FWER$. To see this, by definition,

$$\begin{aligned} \text{FDR} &= E[\frac{N_{0|r}}{R}|R > 0]Pr(R > 0) \\ &\leq E[I_{(N_{0|r} \geq 1)}|R > 0]Pr(R > 0) \\ &= Pr((N_{0|r} \geq 1) \cap (R > 0)) \\ &\leq Pr(N_{0|r} \geq 1) = \text{FWER}, \end{aligned}$$

where $I_{(N_{0|r} \geq 1)}$ is an indicator function of event $N_{0|r} \geq 1$. This implies that procedures that control FWER under a certain significance level automatically control FDR under the same significance level. In our context, a factor discovery criterion that controls FWER at $\alpha$ also controls FDR at $\alpha$.

Intuitively, this is because FDR allows $N_{0|r}$ to grow in proportion to $R$ whereas FWER measures the probability of making even a single Type I error.

Returning to Example A, Panel A shows that a false discovery event has occurred under FWER since $N_{0|r} = 50 \geq 1$ and the realized *FDP* is high, $50/100 = 50\%$. This suggests that the *probability* of false discoveries (FWER) and the *expected* proportion of false discoveries (FDR) may be high.[17] The remedy, as suggested by many FWER and FDR adjustment procedures, would be to lower p-value thresholds for these hypotheses. In terms of Panel A, this would turn some of the 50 false discoveries insignificant and push them into the "Unpublished" category. Hopefully the 50 true discoveries would survive this p-value "upgrade" and remain significant, which is only possible if their p-values are relatively large.

On the other hand, Type II errors — the mistake of missing true factors — are also important in multiple hypothesis testing. Similar to Type I errors, both the total number of missed discoveries $N_{1|a}$ and the fraction of missed discoveries among all abandoned tests $N_{1|a}/(M-R)$ are frequently used to measure the severity of Type II errors.[18] Ideally, one would like to simultaneously minimize the chance of committing a Type I error and that of committing a Type II error. In our context, we would like to include as few insignificant factors (i.e., as low a Type I error rate) as possible and simultaneously as many significant ones (i.e., as low a Type II error rate) as possible. Unfortunately, this is not feasible: as in single hypothesis testing, a decrease in the Type I error rate often leads to an increase in the Type II error rate and vice versa. We therefore seek a balance between the two types of errors. A standard approach is to specify a significance level $\alpha$ for the Type I error rate and derive testing procedures that aim to minimize the Type II error rate, i.e., maximize power, among the class of tests with Type I error rate at most $\alpha$.

When comparing two testing procedures that can both achieve a significance level $\alpha$, it seems reasonable to use their Type II error rates. However, the exact Type II error rate typically depends on a set of unknown parameters and is therefore difficult to assess.[19] To overcome this difficulty, researchers frequently use distance of the

[17] Panel A only shows one realization of the testing outcome for a certain testing procedure (e.g., independent tests). To evaluate FWER and FDR, both of which are expectations and hence depend on the underlying joint distribution of the testing statistics, we need to know the population of the testing outcomes. To give an example that is compatible with Example A, we assume that the t-statistics for the 700 hypotheses are independent. Moreover, we assume the t-statistic for a true factor follows a normal distribution with mean zero and variance one, i.e., $\mathcal{N}(0,1)$; for a false factor, we assume its t-statistic follows $\mathcal{N}(2,1)$. Under these assumptions about the joint distribution of the test statistics, we find via simulations that FWER is 100% and FDR is 26%, both exceeding 5%.

[18] See Simes (1986) for one example of Type II error in simulation studies and Farcomeni (2008) for another example in medical experiments.

[19] In single hypothesis testing, a typical Type II error rate is a function of the realization of the alternative hypothesis. Since it depends on unknown parameter values in the alternative hypothesis, it is difficult to measure directly. The situation is exacerbated in multiple hypothesis testing because the Type II error rate now depends on a multi-dimensional unknown parameter vector. See Zehetmayer and Posch (2010) for power estimation in large-scale multiple testing problems.

actual Type I error rate to some pre-specified significance level as the measure for a procedure's efficiency. Intuitively, if a procedure's actual Type I error rate is strictly below $\alpha$, we can probably push this error rate closer to $\alpha$ by making the testing procedure less stringent, i.e., higher p-value threshold so there will be more discoveries. In doing so, the Type II error rate is presumably lowered given the inverse relation between the two types of error rates. Therefore, once a procedure's actual Type I error rate falls below a pre-specified significance level, we want it to be as close as possible to that significance level in order to achieve the smallest Type II error rate. Ideally, we would like a procedure's actual Type I error rate to be exactly the same as the given significance level.[20]

Both FWER and FDR are important concepts that have wide applications in many scientific fields. However, based on specific applications, one may be preferred over the other. When the number of tests is very large, FWER controlling procedures tend to become very tough and eventually lead to a very limited number of discoveries, if any. Conversely, FWER control is more desirable when the number of tests is relatively small, in which case more discoveries can be achieved and at the same time trusted. In the context of our paper, it is difficult to judge whether the number of tests in the finance literature is large. First, we are unsure of the true number of factors that have been tried. Although there are around 300 published ones, hundreds or even thousands of factors might have been constructed and tested. Second, 300 may seem a large number to researchers in finance but is very small compared to the number of tests conducted in medical research.[21] Given this difficulty, we do not take a stand on the relative appropriateness of these two measures but instead provide adjusted p-values for both. Researchers can compare their p-values with these benchmarks to see whether FDR or even FWER is satisfied.

Related to the false discovery rate, recent research by Lehmann and Romano (2005) tries to control the probability of the realized FDP exceeding a certain threshold value, i.e., $P(FDP > \gamma) \leq \alpha$, where $\gamma$ is the threshold FDP value and $\alpha$ is the significance level.[22] Instead of the expected FDP (i.e., the FDR), Lehmann and Romano's method allows one to make a statement concerning the realized FDP, which might be more desirable in certain applications. For example, targeting the realized FDP is a loss control method and seems more appropriate for risk management or insurance. For our asset pricing applications, we choose to focus on the FDR. In addition, it is difficult to tell whether controlling the realized FDP at $\gamma = 0.1$ with a

[20] In our framework, individual p-values are sufficient statistics for us to make adjustment for multiple tests. Each individual p-value represents the probability of having a t-statistic that is at least as large as the observed one under the null hypothesis. What happens under the alternative hypotheses (e.g., the power of the tests) does not directly come into play because hypothesis testing in the frequentist framework has a primary focus on the Type I error rate. When we deviate from the frequentist framework and consider Bayesian methods, the power of the tests becomes important because Bayesian odds ratios put the Type I and Type II error rate on the same footing.

[21] For instance, tens of thousands of tests are performed in the analysis of DNA microarrays. See Farcomeni (2008) for more applications of multiple testing in medical research.

[22] Also see Romano and Shaikh (2006) and Romano, Shaikh and Wolf (2008).

significance level of $\alpha = 0.05$ is more stringent than controlling FDP at $\gamma = 0.2$ with a significance level of $\alpha = 0.01$. While we use the FDR in our application, we provide some details on the FDP methods in the Appendix.

## 4.4 P-value Adjustment: Three Approaches

The statistics literature has developed many methods to control both FWER and FDR.[23] We choose to present the three most well-known adjustments: Bonferroni, Holm, and Benjamini, Hochberg and Yekutieli (BHY). Both Bonferroni and Holm control FWER, and BHY controls FDR. Depending on how the adjustment is implemented, they can be categorized into two general types of corrections: a "Single step" correction equally adjusts each p-value and a "sequential" correction is an adaptive procedure that depends on the entire distribution of p-values. Bonferroni is a single-step procedure whereas Holm and BHY are sequential procedures. Table 3 summarizes the two properties of the three methods.

Table 3: **A Summary of p-value Adjustments**

| Adjustment type | Single/Sequential | Multiple test |
|---|---|---|
| Bonferroni | Single | FWER |
| Holm | Sequential | FWER |
| Benjamini, Hochberg and Yekutieli (BHY) | Sequential | FDR |

In the usual multiple testing framework, we observe the outcomes of all test statistics, those rejected as well as not rejected. In our context, however, successful factors are more likely to be published and their p-values observed. This missing observations problem is the main obstacle in applying existing adjustment procedures. In appendix A, we propose a new general methodology to overcome this problem. For now, we assume that all tests and their associated p-values are observed and detail the steps for the three types of adjustments.

Suppose there are in total $M$ tests and we choose to set FWER at $\alpha_w$ and FDR at $\alpha_d$. In particular, we consider an example with the total number of tests $M = 10$ to illustrate how different adjustment procedures work. For our main results, we set both $\alpha_w$ and $\alpha_d$ at 5%. Table 4, Panel A lists the t-ratios and the corresponding p-values for 10 hypothetical tests. The numbers in the table are broadly consistent with the magnitude of t-ratios that researchers report for factor significance. Note

[23] Methods that control FWER include Holm (1979), Hochberg (1988) and Hommel (1988). Methods that control FDR include Benjamini and Hochberg(1995), Benjamini and Liu (1999) and Benjamini and Yekutieli (2001).

that all 10 factors will be "discovered" if we test one hypothesis at a time. Multiple testing adjustments will usually generate different results.

Table 4: **An Example of Multiple Testing**

Panel A displays 10 t-ratios and their associated p-values for a hypothetical example. Panel B and C explain Holm's and BHY's adjustment procedure, respectively. Bold numbers in each panel are associated with significant factors under the specific adjustment procedure of that panel. $M$ represents the total number of tests (10) and $c(M) = \sum_{j=1}^{M} 1/j$. $k$ is the order of p-values from lowest to highest. $\alpha_w$ is the significance level for Bonferroni's and Holm's procedure and $\alpha_d$ is the significance level for BHY's procedure. Both numbers are set at 5%. The threshold t-ratio for Bonferroni is 0.05%, for Holm 0.60% and for BHY 0.85%.

| Panel A: 10 Hypothetical t-ratios and Bonferroni "significant" factors | | | | | | | | | | | # of discoveries |
|---|---|---|---|---|---|---|---|---|---|---|---|
| k | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| t-ratio | 1.99 | 2.63 | 2.21 | 3.43 | 2.17 | 2.64 | 4.56 | 5.34 | 2.75 | 2.49 | 3 |
| p-value (%) | 4.66 | 0.85 | 2.71 | **0.05** | 3.00 | 0.84 | **0.00** | **0.00** | 0.60 | 1.28 | |
| Panel B: Holm adjusted p-values and "significant" factors | | | | | | | | | | | |
| New order (k) | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | 4 |
| Old order k | 8 | 7 | 4 | 9 | 6 | 2 | 10 | 3 | 5 | 1 | |
| p-value (%) | 0.00 | 0.00 | 0.05 | 0.60 | 0.84 | 0.85 | 1.28 | 2.71 | 3.00 | 4.66 | |
| $\alpha_w/(M+1-k)$ | **0.50** | **0.56** | **0.63** | **0.71** | 0.83 | 1.00 | 1.25 | 1.67 | 2.50 | 5.00 | |
| Panel C: BHY adjusted p-values and "significant" factors | | | | | | | | | | | |
| New order (k) | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | 6 |
| Old order k | 8 | 7 | 4 | 9 | 6 | 2 | 10 | 3 | 5 | 1 | |
| p-value (%) | 0.00 | 0.00 | 0.05 | 0.60 | 0.84 | 0.85 | 1.28 | 2.71 | 3.00 | 4.66 | |
| $(k \cdot \alpha_d)/(M \times c(M))$ $\alpha_d = 5\%$ | **0.15** | **0.21** | **0.50** | **0.70** | **0.85** | **1.00** | 1.20 | 1.35 | 1.55 | 1.70 | |

### 4.4.1 Bonferroni's Adjustment

*Bonferroni*'s adjustment is as follows:

- Reject any hypothesis with p-value $\leq \frac{\alpha_w}{M}$:

$$p_i^{Bonferroni} = \min[Mp_i, 1].$$

Bonferroni applies the same adjustment to each test. It inflates the original p-value by the number of tests $M$; the adjusted p-value is compared with the threshold value $\alpha_w$.

**Example 4.4.1** To apply Bonferroni's adjustment to the example in Table 4, we simply multiply all the p-values by 10 and compare the new p-values with $\alpha_w = 5\%$.

Equivalently, we can look at the original p-values and consider the cutoff of 0.5%(= $\alpha_w/10$). This leaves the t-ratio of tests 4,7 and 8 as significant.

Using the notation in Panel B of Table 5 and assuming $M_0$ of the $M$ null hypotheses are true, Bonferroni operates as a single step procedure that can be shown to restrict FWER at levels less than or equal to $M_0\alpha_w/M$, without any assumption on the dependence structure of the p-values. Since $M_0 \leq M$, Bonferroni also controls FWER at level $\alpha_w$.[24]

### 4.4.2 Holm's Adjustment

Sequential methods have recently been proposed to adjust p-values in multiple hypothesis testing. They are motivated by a seminal paper by Schweder and Spjotvoll (1982), who suggest a graphical presentation of the multiple testing p-values. In particular, using $N_p$ to denote the number of tests that have a p-value exceeding $p$, Schweder and Spjotvoll (1982) suggest plotting $N_p$ against $(1-p)$. When $p$ is not very small, it is very likely that the associated test is from the null hypothesis. In this case, the p-value for a null test can be shown to be uniformly distributed between 0 and 1. It then follows that for a large $p$ and under independence among tests, the expected number of tests with a p-value exceeding $p$ equals $T_0(1-p)$, where $T_0$ is the number of null hypotheses, i.e., $E(N_p) = T_0(1-p)$. By plotting $N_p$ against $(1-p)$, the graph should be approximately linear with slope $T_0$ for large p-values. Points on the graph that substantially deviate from this linear pattern should correspond to non-null hypotheses, i.e., discoveries. The gist of this argument — large and small p-values should be treated differently — have been distilled into many variations of sequential adjustment methods, among which we will introduce Holm's method that controls FWER and BHY's method that controls FDR.

*Holm*'s adjustment is as follows:

- Order the original p-values such that $p_{(1)} \leq p_{(2)} \leq \cdots p_{(k)} \leq \cdots \leq p_{(M)}$ and let associated null hypotheses be $H_{(1)}, H_{(2)}, \cdots H_{(k)} \cdots, H_{(M)}$.
- Let $k$ be the minimum index such that $p_{(k)} > \frac{\alpha_w}{M+1-k}$.
- Reject null hypotheses $H_{(1)} \cdots H_{(k-1)}$ (i.e., declare these factors significant) but not $H_{(k)} \cdots H_{(M)}$.

[24]The number of true nulls $M_0$ is inherently unknown, so we usually cannot make Bonferroni more powerful by increasing $\alpha_w$ to $\hat{\alpha} = M\alpha_w/M_0$ (note that $M_0\hat{\alpha}/M = \alpha_w$). However some papers, including Schweder and Spjotvoll (1982) and Hochberg and Benjamini (1990), try to improve the power of Bonferroni by estimating $M_0$. We try to achieve the same goal by using either Holm's procedure which also controls FWER or procedures that control FDR, an alternative definition of Type I error rate.

The equivalent adjusted p-value is therefore

$$p_{(i)}^{Holm} = \min[\max_{j \leq i}\{(M - j + 1)p_{(j)}\}, 1].$$

Holm's adjustment is a step-down procedure: [25] for the ordered p-values, we start from the smallest p-value and go down to the largest one. If $k$ is the smallest index that satisfies $p_{(k)} > \frac{\alpha_w}{M+1-k}$, we will reject all tests whose ordered index is below $k$.

To explore how Holm's adjustment procedure works, suppose $k_0$ is the smallest index such that $p_{(k)} > \frac{\alpha_w}{M+1-k}$. This means that for $k < k_0$, $p_{(k)} \leq \frac{\alpha_w}{M+1-k}$. In particular, for $k = 1$, Bonferroni = Holm, i.e., $\frac{\alpha_w}{M} = \frac{\alpha_w}{M+1-(k=1)}$; for $k = 2$, $\frac{\alpha_w}{M} < \frac{\alpha_w}{M+1-(k=2)}$, so Holm is less stringent than Bonferroni. Since less stringent hurdles are applied to the second to the $(k_0 - 1)$th p-values, more discoveries are generated under Holm's than Bonferroni's adjustment.

**Example 4.4.2** To apply Holm's adjustment to the example in Table 4, we first order the p-values in ascending order and try to locate the smallest index $k$ that makes $p_{(k)} > \frac{\alpha_w}{M+1-k}$. Table 4, Panel B shows the ordered p-values and the associated $\frac{\alpha_w}{M+1-k}$'s. Starting from the smallest p-value and going up, we see that $p_{(k)}$ is below $\frac{\alpha_w}{M+1-k}$ until $k = 5$, at which $p_{(5)}$ is above $\frac{\alpha_w}{10+1-7}$. Therefore, the smallest $k$ that satisfies $p_{(k)} > \frac{\alpha_w}{M+1-k}$ is 5 and we reject the null hypothesis for the first four ordered tests (we discover four factors) and fail to reject the null for the remaining six tests. The original labels for the rejected tests are in the second row in Panel B. Compared to Bonferroni, one more factor (9) is discovered, that is, four factors rather than three are significant. In general, Holm's approach leads to more discoveries and all discoveries under Bonferroni are also discoveries under Holm's.

Like Bonferroni, Holm also restricts FWER at $\alpha_w$ without any requirement on the dependence structure of p-values. It can also be shown that Holm is uniformly more powerful than Bonferroni in that tests rejected (factors discovered) under Bonferroni are always rejected under Holm but not vice versa. In other words, Holm leads to at least as many discoveries as Bonferroni. Given the dominance of Holm over Bonferroni, one might opt to only use Holm. We include Bonferroni because it is the most widely used adjustment and a simple single-step procedure.

### 4.4.3 Benjamini, Hochberg and Yekutieli's Adjustment

*Benjamini, Hochberg and Yekutieli* (BHY)'s adjustment is as follows:

[25] Viewing small p-values as "up" and large p-values as "down", Holm's procedure is a "step-down" procedure in that it goes from small p-values to large ones. This terminology is consistent with the statistics literature. Of course, small p-values are associated with "large" values of the test statistics.

- As with Holm's procedure, order the original p-values such that $p_{(1)} \leq p_{(2)} \leq \cdots p_{(k)} \leq \cdots \leq p_{(M)}$ and let associated null hypotheses be $H_{(1)}, H_{(2)}, \cdots H_{(k)} \cdots, H_{(M)}$.
- Let $k$ be the maximum index such that $p_{(k)} \leq \frac{k}{M \times c(M)} \alpha_d$.
- Reject null hypotheses $H_{(1)} \cdots H_{(k)}$ but not $H_{(k+1)} \cdots H_{(M)}$.

The equivalent adjusted p-value is defined sequentially as:

$$p_{(i)}^{BHY} = \begin{cases} p_{(M)} & \text{if } i = M, \\ \min[p_{(i+1)}^{BHY}, \frac{M \times c(M)}{i} p_{(i)}] & \text{if } i \leq M-1. \end{cases}$$

where, $c(M)$ is a function of the total number of tests $M$ and controls for the generality of the test. We adopt the choice in Benjamini and Yekutieli (2001) and set $c(M)$ at

$$c(M) = \sum_{j=1}^{M} \frac{1}{j},$$

a value at which the procedure works under arbitrary dependence structure among the p-values. We discuss alternative specifications of $c(M)$ shortly.

In contrast to Holm's, BHY's method starts with the largest p-value and goes up to the smallest one. If $k$ is the largest index that satisfies $p_{(k)} \leq \frac{k}{M \times c(M)} \alpha_d$, we will reject all tests (discover factors) whose ordered index is below or equal to $k$. Also, note that $\alpha_d$ (significance level for FDR) is chosen to be a smaller number than $\alpha_w$ (significance level for FWER). The reason for such a choice is discussed in Section 4.6.

To explore how BHY works, let $k_0$ be the largest index such that $p_{(k)} \leq \frac{k}{M \times c(M)} \alpha_d$. This means that for $k > k_0$, $p_{(k)} > \frac{k}{M \times c(M)} \alpha_d$. In particular, we have $p_{(k_0+1)} > \frac{(k_0+1)}{M \times c(M)} \alpha_d$, $p_{(k_0+2)} > \frac{(k_0+2)}{M \times c(M)} \alpha_d$, $\ldots$, $p_{(M)} > \frac{M}{M \times c(M)} \alpha_d$. We see that the $(k_0+1)$th to the last null hypotheses, not rejected, are compared to numbers smaller than $\alpha_d$, the usual significance level in single hypothesis testing. By being stricter than single hypothesis tests, BHY guarantees that the *false discovery rate* is below the pre-specified significance level under arbitrary dependence structure among the p-values. See Benjamini and Yekutieli (2001) for details on the proof.

**Example 4.4.3** To apply BHY's adjustment to the example in Table 4, we first order the p-values in ascending order and try to locate the largest index $k$ that satisfies $p_{(k)} \leq \frac{k}{M \times c(M)} \alpha_d$. Table 4, Panel C shows the ordered p-values and the associated $\frac{k}{M \times c(M)} \alpha_d$'s. Starting from the largest p-value and going down, we see that $p_{(k)}$ is above $\frac{k}{M \times c(M)} \alpha_d$ until $k = 6$, at which $p_{(6)}$ is below $\frac{k}{10 \times 2.93} \alpha_d$. Therefore, the smallest

$k$ that satisfies $p_{(k)} \leq \frac{k}{M \times c(M)} \alpha_d$ is 6 and we reject the null hypothesis for the first six ordered tests and fail to reject for the remaining four tests. In the end, BHY leads to six significant factors (8,7,4,9,6 and 2), three more than Bonferroni and two more than Holm.[26]

Under independence among p-values, we can gain insight into the choice of $p_{(i)}^{BHY}$ by interpreting $p_{(i)}^{BHY}$ as the solution to a post-experiment maximization problem.[27] In particular, assume all individual hypotheses are performed and their p-values collected. It can be shown that $p_{(i)}^{BHY}$ is the solution to the following problem:

$$\begin{aligned} &\textit{Objective} : \text{Choose } \hat{p} \text{ that maximizes the number of discoveries } n(\hat{p}), \\ &\textit{Constraint} : \hat{p}M/n(\hat{p}) \leq \alpha_d. \end{aligned}$$

We first interpret the constraint. Under independence and when each hypothesis is tested individually at level $\hat{p}$, the expected number of false discoveries satisfies $E(N_{0|r}) \leq \hat{p}M$. Hence, after observing the outcome of the experiment and thus conditional on having $n(\hat{p})$ discoveries, the FDR is no greater than $\hat{p}M/n(\hat{p})$. The constraint therefore requires the post-experiment FDR to satisfy the pre-specified significance level. Under this constraint, the objective is to choose $\hat{p}$ to maximize the number of discoveries. Since the constraint is satisfied for each realized p-value sequence of the experiment, it is satisfied in expectation as well. In sum, $p_{(i)}^{BHY}$ is the optimal cutoff p-value (i.e., maximal number of discoveries) that satisfies the FDR constraint for each outcome of the experiment.

The choice of $c(M)$ determines the generality of BHY's procedure. Intuitively, the larger $c(M)$ is, the more difficult it is to satisfy the inequality $p_{(k)} \leq \frac{k}{M \times c(M)} \alpha_d$ and hence there will be fewer discoveries. This makes it easier to restrict *the false discovery rate* below a given significance level since fewer discoveries are made. In the original work that develops the concept of *false discovery rate* and related testing procedures, $c(M)$ is set equal to one. It turns out that under this choice, BHY is only valid when the test statistics are independent or positively dependent.[28] With our choice of $c(M)$, BHY is valid under any form of dependence among the p-values.[29] Note with $c(M) > 1$, this reduces the size of $\frac{k}{M \times c(M)} \alpha_d$ and it is tougher to satisfy

[26]For independent tests, 10/10 are discovered. For BHY, the effective cutoff is 0.85%, for Bonferroni 0.50% and for Holm 0.60%. The cutoffs are all far smaller than the usual 5%.

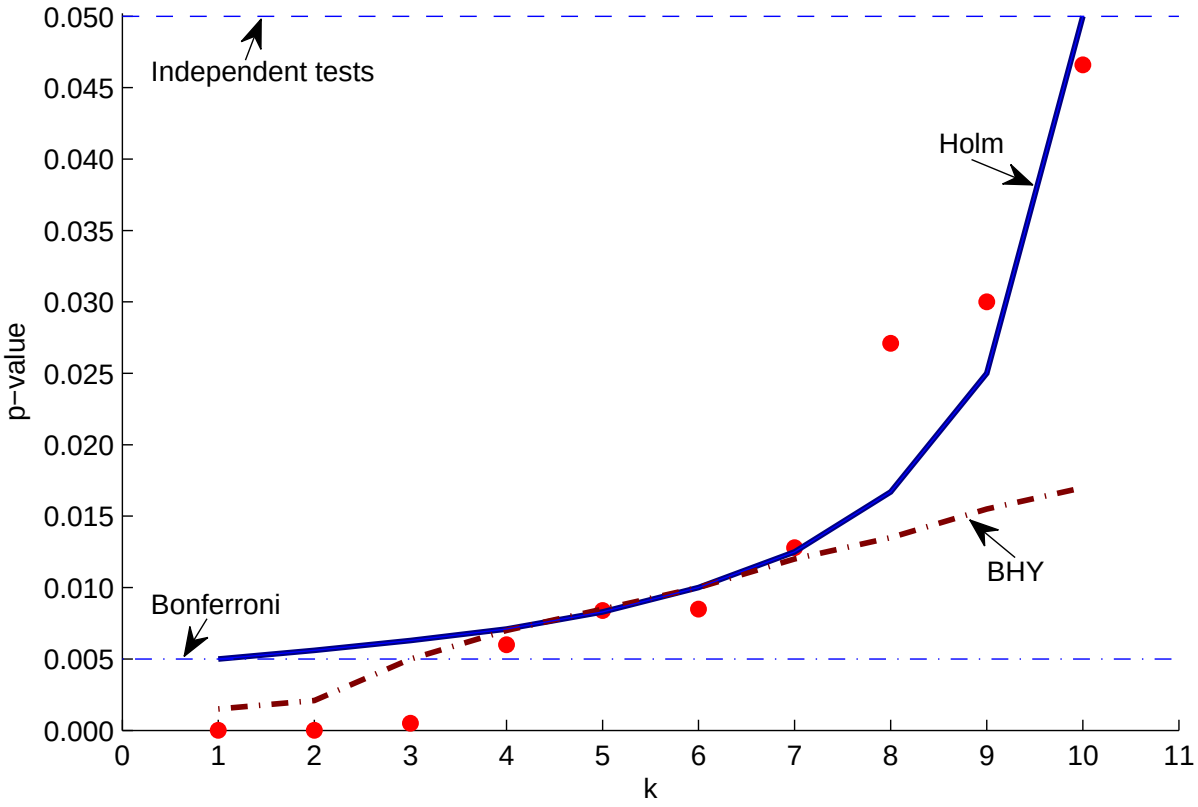
[27]This interpretation is shown in Benjamini and Hochberg (1995). Under independence, $c(M) \equiv 1$ is sufficient for BHY to work. See our subsequent discussions on the choice of $c(M)$.

[28]Benjamini and Hochberg (1995) is the original paper that proposes FDR and sets $c(M) \equiv 1$. They show their procedures restricts the FDR below the pre-specified significance level under independence. Benjamini and Yekutieli (2001) and Sarkar (2002) later show that the choice of $c(M) \equiv 1$ also works under positive dependence. For recent studies that assume specific dependence structure to improve on BHY, see Yekutieli and Benjamini (1999), Troendle (2000), Dudoit and Van der Laan (2008) and Romano, Shaikh and Wolf (2008). For a modified Type I error rate definition that is analogous to FDR and its connection to Bayesian hypothesis testing, see Storey (2003).

[29]See Benjamini and Yekutieli (2001) for the proof.

the inequality $p_{(k)} \leq \frac{k}{M \times c(M)} \alpha_d$. That is, there will be fewer factors found to be significant.

Figure 1: **Multiple Test Thresholds for Example A**




The 10 p-values for Example A and the threshold p-value lines for various adjustment procedures. All 10 factors are discovered under independent tests, three under Bonferroni, four under Holm and six under BHY. The significance level is set at 5% for each adjustment method.

Figure 1 summarizes Example A. It plots the original p-value sample as well as threshold p-value lines for various adjustment procedures. We see the stark difference in outcomes between multiple and single hypothesis testing. While all 10 factors would be discovered under single hypothesis testing, only three to six factors would be discovered under a multiple hypothesis test. Although single hypothesis testing guarantees the Type I error of each test meets a given significance level, meeting the more stringent FWER or FDR bound will lead us to discard a number of factors.

To summarize the properties of the three adjustment procedures, Bonferroni's adjustment is the simplest and inflates the original p-value by the total number of tests. Holm's adjustment is a refinement of Bonferroni but involves ordering of p-values and thus depends on the entire distribution of p-values. BHY's adjustment,

unlike that of Bonferroni or Holm, aims to control the false discovery rate and also depends on the distribution of p-values. Importantly, all three methods allow for general dependence among the test statistics.

## 4.5 Summary Statistics

Figure 2 shows the history of discovered factors and publications.[30] We observe a dramatic increase in factor discoveries during the last decade. In the early period from 1980 to 1991, only about one factor is discovered per year. This number has grown to around five in the 1991-2003 period, during which a number of papers, such as Fama and French (1992), Carhart (1997) and Pastor and Stambaugh (2003), spurred interest in studying cross-sectional return patterns. In the last nine years, the annual factor discovery rate has increased sharply to around 18. In total, 164 factors were discovered in the past nine years, roughly doubling the 84 factors discovered in all previous years. We do not include working papers in Figure 2. In our sample, there are 63 working papers covering 68 factors.

We obtain t-ratios for each of the 316 factors discovered, including the ones in working papers.[31] The overwhelming majority of t-ratios exceed the 1.96 benchmark for 5% significance.[32] The non-significant ones typically belong to papers that propose a number of factors. These likely represent only a small sub-sample of non-significant t-ratios for all tried factors. Importantly, we take published t-ratios as given. That is, we assume they are econometrically sound with respect to the usual suspects (data errors, coding errors, misalignment, heteroskedasticity, autocorrelation, outliers, etc.).

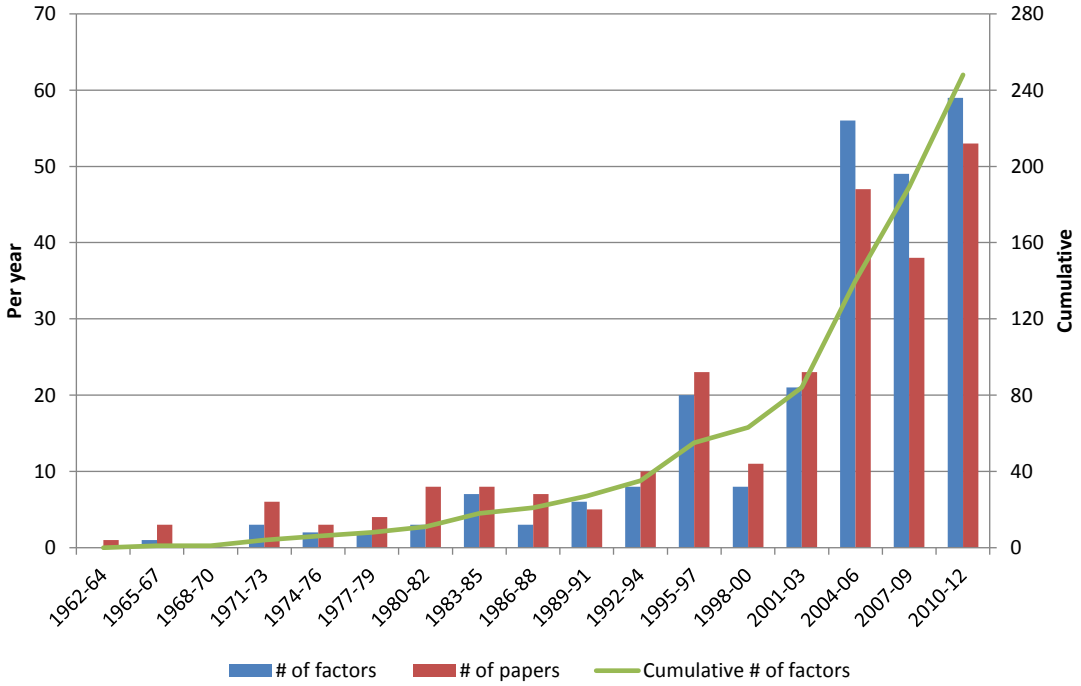
## 4.6 P-value Adjustment When All Tests Are Published ($M = R$)

We now apply the three adjustment methods previously introduced to the observed factor tests, under the assumption that test results of all tried factors are available. We know that this assumption is false since our sample under-represents all insignificant factors by conventional significance standards: we only observe those insignificant

[30] To be specific, we only count those that have t-ratios or equivalent statistics reported. Roughly 20 new factors fail to satisfy this requirement. For details on these, see factors in Table 6 marked with ‡.

[31] The sign of a t-ratio depends on the source of risk or the direction of the long/short strategy. We usually calculate p-values based on two-sided t-tests, so the sign does not matter. Therefore we use absolute values of these t-ratios.

[32] The multiple testing framework is robust to outliers. The procedures are based on either the total number of tests (Bonferroni) or the order statistics of t-ratios (Holm and BHY).




factors that are published alongside significant ones. We design methods to handle this missing data issue later.

Despite some limitations, our results in this section are useful for at least two purposes. First, the benchmark t-ratio based on our incomplete sample provides a lower bound of the true t-ratio benchmark. In other words, if $M$(total number of tests) $>R$ (total number of discoveries), then we would accept fewer factors than when $M = R$,[33], so future t-ratios need to at least surpass our benchmark to claim significance. Second, results in this section can be rationalized within a Bayesian or hierarchical testing framework.[34] Factors in our list constitute an "elite" group: they have survived academia's scrutiny for publication. Placing a high prior on this group in a Bayesian testing framework or viewing this group as a cluster in a hierarchical testing framework, one can interpret results in this section as the first step factor selection within an a priori group.

[33]This is always true for Bonferroni's adjustment but not always true for the other two types of adjustments. The Bonferroni adjusted t-ratio is monotonically increasing in the number of trials so the t-ratio benchmark will only rise if there are more factors. Holm and BHY depend on the exact t-ratio distribution so more factors do not necessarily imply a higher t-ratio benchmark.

[34]See Wagenmakers and Grünwald (2006) and Storey (2003) on Bayesian interpretations of traditional hypothesis testing. See Meinshausen (2008) for a hierarchical approach on variable selection.

Based on our sample of observed t-ratios of published factors,[35] we obtain three benchmark t-ratios. In particular, at each point in time, we transform the set of available t-ratios into p-values. We then apply the three adjustment methods to obtain benchmark p-values. Finally, these p-value benchmarks are transformed back into t-ratios, assuming that standard normal distribution well approximates the t-distribution. To guide future research, we extrapolate our benchmark t-ratios 20 years into the future.

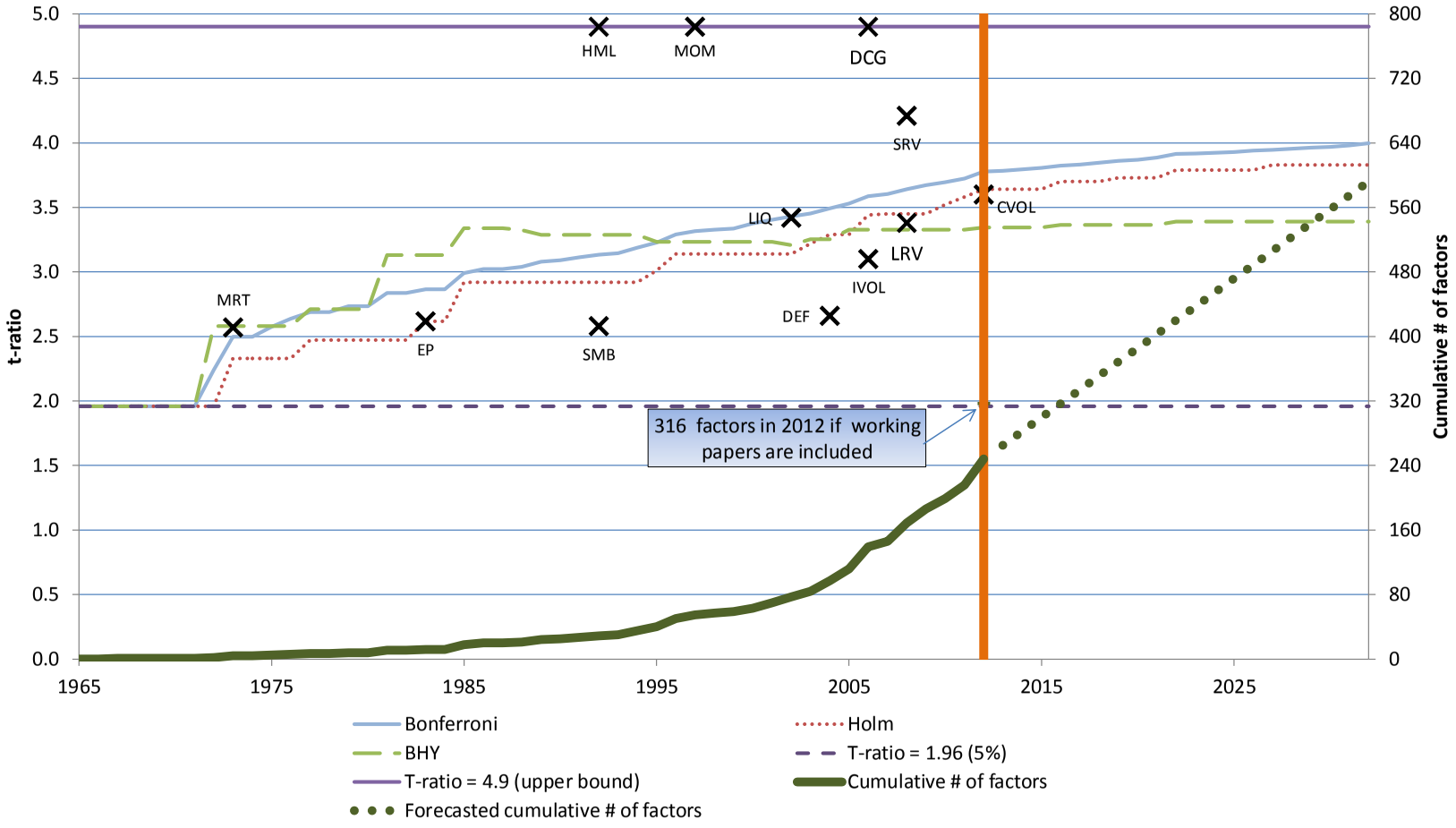
We choose to set $\alpha_w$ at 5% (Holm, FWER) and $\alpha_d$ at 1% (BHY, FDR) for our main results. Significance level is subjective, as in individual hypothesis testing where conventional significance levels are usually adopted. Since FWER is a special case of the Type I error in individual testing and 5% seems the default significance level in cross-sectional studies, we set $\alpha_w$ at 5%. On the other hand, FDR is a weaker control relative to FWER; moreover, it has no power in further screening individual tests if FDR is set greater than or equal to the significance level of individual tests.[36] We therefore set FDR at 1% but will explain what happens when $\alpha_d$ is increased to 5%.

Figure 3 presents the three benchmark t-ratios. Both Bonferroni and Holm adjusted benchmark t-ratios are monotonically increasing in the number of discoveries. For Bonferroni, the benchmark t-ratio starts at 1.96 and increases to 3.78 by 2012. It reaches 4.00 in 2032. The corresponding p-values for 3.78 and 4.00 are 0.02% and 0.01% respectively, much lower than the starting level of 5%. Holm implied t-ratios always fall below Bonferroni t-ratios, consistent with the fact that Bonferroni always results in fewer discoveries than Holm. However, Holm tracks Bonferroni closely and their differences are small. BHY implied benchmarks, on the other hand, are not monotonic. They fluctuate before year 2000 and stabilize at 3.39 (p-value = 0.07%) after 2010. This stationarity feature of BHY implied t-ratios, inherent in the definition of FDR, is in contrast to Bonferroni and Holm. Intuitively, at any fixed significance level $\alpha$, the Law of Large Numbers forces the false discovery rate (FDR)

[35]There are at least two ways to generate t-ratios for a risk factor. One way is to show that factor related sorting results in cross-sectional return patterns that are not explained by standard risk factors. The t-ratio for the intercept of the long/short strategy returns regressed on common risk factors is usually reported. The other way is to use factor loadings as explanatory variables and show that they are related to the cross-section of expected returns after controlling for standard risk factors. Individual stocks or stylized portfolios (e.g., Fama-French 25 portfolios) are used as dependent variables. The t-ratio for the factor risk premium is taken as the t-ratio for the factor. In sum, depending on where the new risk factor or factor returns enter the regressions, the first way can be thought of as the left hand side (LHS) approach and the second the right hand side (RHS) approach. For our data collection, we choose to use the RHS t-ratios. When they are not available, we use the LHS t-ratios or simply the t-ratios for the average returns of long/short strategies if the authors do not control for other risk factors.

[36]When tests are all significant based on single testing and for Benjamini and Hochberg (1995)'s original adjustment algorithm (i.e., $c(M) \equiv 1$), BHY yields the same results as single testing. To see this, notice that the threshold for the largest p-value becomes $\alpha_d$ in BHY's method. As a result, if all tests are individually significant at level $\alpha_d$, the largest p-value would satisfy $p_{(M)} \leq \alpha_d$. Based on BHY's procedure, this means we reject all null hypotheses. In our context, the p-values for published factors are all below 5% due to hidden tests. Therefore, under $c(M) \equiv 1$, if we set $\alpha_d$ equal to 5%, all of these factors will be still be declared as significant under multiple testing.

Figure 3: **Adjusted t-ratios, 1965-2032**




The green solid curve shows the historical cumulative number of factors discovered, excluding those from working papers. Forecasts (dotted green line) are based on a linear extrapolation. The dark crosses mark selected factors proposed by the literature. They are MRT (market beta; Fama and MacBeth (1973)), EP (earnings-price ratio; Basu (1983)), SMB and HML (size and book-to-market; Fama and French (1992)), MOM (momentum; Carhart (1997)), LIQ (liquidity; Pastor and Stambaugh (2003)), DEF (default likelihood; Vassalou and Xing (2004)), IVOL (idiosyncratic volatility; Ang, Hodrick, Xing and Zhang (2006)); DCG (durable consumption goods; Yogo (2006)); SRV and LRV (short-run and long-run volatility; Adrian and Rosenberg (2008)) and CVOL (consumption volatility; Boguth and Kuehn (2012)). T-ratios over 4.9 are truncated at 4.9. For detailed descriptions of these factors, see Table 6.

to converge to a constant.[37] If we change $\alpha_d$ to 5%, the corresponding BHY implied benchmark t-ratio is 2.78 (p-value = 0.54%) in 2012 and 2.81 (p-value = 0.50%) in 2032, still much higher than the 1.96 staring value. In sum, taking into account of testing multiplicity, we believe the minimum threshold t-ratio for 5% significance is about 2.8, which corresponds to a p-value of 0.5%.

To see how the new t-ratio benchmarks better differentiate the statistical significance of factors, in Figure 3 we mark the t-ratios of a few prominent factors. Among these factors, HML, MOM, DCG, SRV and MRT are significant across all types of t-ratio adjustments, EP, LIQ and CVOL are sometimes significant and the rest are never significant.

## 4.7 Robustness

Our three adjustment methods are able to control their Type I error rates (FWER for Bonferroni and Holm; FDR for BHY) under arbitrary distributional assumptions about the test statistics. However, if the test statistics are positively correlated, then all three methods might be conservative in that too few factors are discovered. Then again, counteracting this conservatism is our incomplete coverage of tried factors. By adding factors to our current sample, certain adjusted threshold t-ratios (e.g., Bonferroni) will increase, making our current estimates less conservative. We discuss the dependence issue in this section and address the incomplete coverage issue in the Appendix.

### 4.7.1 Test statistics dependence

In theory, under independence, Bonferroni and Holm approximately achieve the prespecified significance level $\alpha$ when the number of tests is large.[38] On the other hand, both procedures tend to generate fewer discoveries than desired when there is a certain degree of dependence among the tests. Intuitively, in the extreme case where all tests are the same (i.e., correlation = 1.0), we do not need to adjust at all: FWER is the

[37] This intuition is precise for the case when tests are independent. When there is dependence, we need the dependence to be weak to apply the Law of Large Numbers.

[38] To see this for Bonferroni, suppose tests are independent and all null hypotheses are true. We have

$$\begin{aligned} FWER &= Pr(N_{0|r} \geq 1) \\ &= 1 - Pr(N_{0|r} = 0) \\ &= 1 - (1 - \alpha/n)^n \\ &\xrightarrow{n \to \infty} 1 - \exp(-\alpha) \approx \alpha \end{aligned}$$

where $n$ denotes the number of tests. The last step approximation is true when $\alpha$ is small.

same as the Type I error rate for single tests. Hence, the usual single hypothesis test is sufficient. Under either independence or positive dependence, the actual Type I error rate of BHY is strictly less than the pre-specified significance level, i.e., BHY is too stringent in that too few factors are discovered.[39]

Having discussed assumptions for the testing methods to work efficiently, we now try to think of scenarios that can potentially violate these assumptions. First, factors that proxy for the same type of risk may be dependent. Moreover, returns of long-short portfolios designed to achieve exposure to a particular type of factor may be correlated. For example, hedge portfolios based on dividend yield, earnings yield and book-to-market are correlated. Other examples include risk factors that reflect financial constraints risk, market-wide liquidity and uncertainty risk. If this type of positive dependence exists among test statistics, all three methods would likely to generate fewer significant factors than desired. There is definitely some dependence in our sample. As mentioned previously, there are a number of factors with price in the denominator which are naturally correlated. Another example is that we count four different idiosyncratic volatility factors. On the other hand, most often factors need to "stand their ground" to be publishable. In the end, if you think we are overcounting at 316, consider taking a haircut to 113 factors (the number of "common" factors). Figure 3 shows that our main conclusions do not materially change. For example, the Holm at 113 factors is 3.29 (p-value = 0.10%) while Holm at 316 factors is 3.64 (p-value = 0.03%).

Second, research studying the same factor but based on different samples will generate highly dependent test statistics. Examples include the sequence of papers studying the size effect. We try to minimize this concern by including, with a few exceptions, only the original paper that proposes the factor. To the extent that our list includes few such duplicate factors, our method greatly reduces the dependence that would be introduced by including all papers studying the same factor but for different sample periods.

Finally, when dependence among test statistics can be captured by Pearson correlations among contemporaneous strategy returns, we present a new model in Section 5 to systematically incorporate the information in test correlations.

### 4.7.2 The Case When $M > R$

To deal with the hidden tests issue when $M > R$, we propose in Appendix A a simulation framework to estimate benchmark t-ratios. The idea is to first back out the underlying distribution for the t-statistics of all tried factors; then, to generate

[39] See footnote 4.4.3 and the references therein.

benchmark t-ratio estimates, apply the three adjustment procedures to simulated t-statistics samples.[40]

Based on our estimates, 71% of all tried factors are missing. The new benchmark t-ratios for Bonferroni and Holm are estimated to be 4.01 and 3.96, respectively; both slightly higher than when $M = R$. This is as expected because more factors are tried under this framework. The BHY implied t-ratio increases from 3.39 to 3.68 at 1% significance and from 2.78 to 3.18 at 5% significance. In sum, across various scenarios, we think the minimum threshold t-ratio is 3.18, corresponding to BHY's adjustment for $M > R$ at 5% significance. Alternative cases all result in even higher benchmark t-ratios. Please refer to Appendix A for the details.

### 4.7.3 A Bayesian Hypothesis Testing Framework

We can also study multiple hypothesis testing within a Bayesian framework. One major obstacle of applying Bayesian methods in our context is the unobservability of all tried factors. While we propose new frequentist methods to handle this missing data problem, it is not clear how to structure the Bayesian framework in this context. In addition, the high dimensionality of the problem raises concerns on both the accuracy and the computational burden of Bayesian methods.

Nevertheless, ignoring the missing data issue, we outline a standard Bayesian multiple hypothesis testing framework in Appendix B and explain how it relates to our multiple testing framework. We discuss in detail the pros and cons of the Bayesian approach. In contrast to the frequentist approach, which uses generalized Type I error rates to guide multiple testing, the Bayesian approach relies on the posterior likelihood function and thus contains a natural penalty term for multiplicity. However, this simplicity comes at the expense of having a restrictive hierarchical model structure and independence assumptions that may not be realistic for our factor testing problem. Although extensions incorporating certain forms of dependence are possible, it is unclear what precisely we should do for the 316 factors in our list. In addition, even for the Bayesian approach, final reject/accept decision still involves threshold choice. Finally, as the number of tests becomes large, the Bayesian approach gets computationally challenging.[41] Due to these concerns, we choose not to implement the Bayesian approach and instead discuss it briefly. We leave extensions of the basic Bayesian framework that could possibly alleviate the above concerns to future research.

[40] The underlying assumption for the model in Appendix A is the independence among t-statistics, which may not be plausible given our previous discussions on test dependence. In that case, our structural model proposed in Section 5 provides a more realistic data generating process for the cross-section of test statistics.

[41] The calculation of the posterior likelihood function involves multiple integrals. As the number of tests becomes large, simulation approaches such as importance sampling may become unstable in calculating these high-dimensional integrals.

### 4.7.4 Methods Controlling the FDP

Instead of FDR, recent research by Lehmann and Romano (2005) develops methods to directly control the realized FDP. In particular, they propose a stepdown method to control the probability of FDP exceeding a threshold value. Since their definition of Type I error (i.e., $P(FDP > \gamma)$ where $\gamma$ is the threshold FDP value) is different from either FWER or FDR, results based on their methods are not comparable to ours. However, the main conclusion is the same. For instance, when $\gamma = 0.10$ and $\alpha = 0.05$, the benchmark t-ratio is 2.70 (p-value = 0.69%), much lower than the conventional cutoff of 1.96. The details are presented in Appendix C.

# 5 Correlation Among Test Statistics

Although the BHY method is robust to arbitrary dependence among test statistics, it does not use any information about the dependence structure. Such information, when appropriately incorporated, can be helpful in making the method more accurate (i.e., less conservative). We focus on the type of dependence that can be captured by Pearson correlation. As one way to generate correlation among test statistics, we focus on the correlation among contemporaneous variables (i.e., factor returns) that constitute the test statistics. This is perhaps the most important source of correlation as contemporaneous returns are certainly affected by the movements of the same set of macroeconomic and market variables. Therefore, in our context, the dependence among test statistics is equivalent to the correlation among strategy returns.

Multiple testing corrections in the presence of correlation has only been considered in the recent statistics literature. Existing methods include bootstrap based permutation tests and direct statistical modeling. Permutation tests resample the entire dataset and construct an empirical distribution for the pool of test statistics.[42] Through resampling, the correlation structure in the data is taken into account and no model is needed. In contrast, direct statistical modeling makes additional distributional assumptions on the data generating process. These assumptions are usually case dependent as different kinds of correlations are more plausible under different circumstances.[43]

[42] Westfall(1993) and Ge et al. (2003) are the early papers that suggest the permutation resampling approach in multiple testing. Later development of the permutation approach tries to reduce computational burden by proposing efficient alternative approaches. Examples include Lin (2005), Conneely and Boehnke (2007) and Han, Kang and Eskin (2009).

[43] See Sun and Cai (2008) and Wei et al. (2009).

In addition, recent research in finance explores bootstrap procedures to assess the statistical significance of individual tests.[44] Most of these studies focus on mutual fund evaluation. They bootstrap the time-series of mutual fund returns and obtain an empirical distribution for the t-ratio for each fund. In contrast, our approach focuses on the joint distribution of the t-ratios, as both FWER and FDR depend on the cross-section of t-ratios. As such, we are able to apply a multiple testing framework to the cross-section of factor tests.

Our data pose a challenge to existing methods both in finance and statistics because we do not always observe the time-series of strategy returns (when a t-ratio is based on long-short strategy returns) or the time-series of slopes in cross-sectional regressions (when a t-ratio is based on the slope coefficients in cross-sectional regressions). Often all we have is the single t-statistic that summarizes the significance of a factor. We propose a novel approach to overcome this missing data problem. It is in essence a direct modeling approach but does not require the full information of the return series based on which the t-statistic is constructed. In addition, our approach is flexible enough to incorporate various kinds of distributional assumptions. We expect it to be a valuable addition to the multiple testing literature, especially when only test statistics are observable.

Our method first proposes a structural model to describe the data generating process for the cross-section of returns. It highlights the key statistical properties for returns in our context and is flexible enough to incorporate various kinds of dependence. Through the structural model, we link Type I error rates in multiple testing to the few structural parameters in the model. Finally, we estimate the model using the t-statistics for published factors and provide multiple testing adjusted t-ratios based on the estimated structural model.[45]

## 5.1 A Model with Correlations

For each factor, suppose researchers construct a corresponding long-short trading strategy and normalize the return standard deviation to be $\sigma = 15\%$ per year, which is close to the annual volatility of the market index.[46] In particular, let the normalized

[44] See Efron (1979) for the original work in the statistics literature. For recent finance applications, see Kosowski, Timmermann, White, and Wermers (2006), Kosowski, Naik and Teo (2007), Fama and French (2010) and Cao, Chen, Liang and Lo (2013).

[45] See Harvey and Liu (2014b) for further details of our approach.

[46] Notice that this assumption is not necessary for our approach. Fixing the standard deviations of different strategies eliminates the need to separately model them, which can be done through a joint modeling of the mean and variance of the cross-section of returns. See Harvey and Liu (2014b) for further discussions on this.

strategy return in period $t$ for the $i$-th discovered strategy be $X_{i,t}$. Then the t-stat for testing the significance of this strategy is:

$$T_i = (\sum_{t=1}^{N} X_{i,t}/N)/(\sigma/\sqrt{N}).$$

Assuming joint normality and zero serial correlation for strategy returns, this t-stat has a normal distribution

$$T_i \sim N(\mu_i/(\sigma/\sqrt{N}), 1),$$

where $\mu_i$ denotes the population mean of the strategy. The $\mu_i$'s are unobservable and hypothesis testing under this framework amounts to testing $\mu_i > 0$. We assume that each $\mu_i$ is an independent draw from the following mixture distribution:

$$\mu_i \sim p_0 I_{\{\mu=0\}} + (1 - p_0)\text{Exp}(\lambda),$$

where $I_{\{\mu=0\}}$ is the distribution that has a point mass at zero, $\text{Exp}(\lambda)$ is the exponential distribution that has a mean parameter $\lambda$ and $p_0$ is the probability of drawing from the point mass distribution. This mixture distribution assumption is the core component for Bayesian multiple testing[47] and succinctly captures the idea of hypothesis testing in the traditional frequentist's view: while there are a range of possible values for the means of truly profitable strategies, a proportion of strategies should have a mean that is indistinguishable from zero. The exponential assumption is not essential for our model as more sophisticated distributions (e.g., a Gamma distribution featuring two free parameters) can be used. We use the exponential distribution for its simplicity[48] and perhaps more importantly, for it being consistent with the intuition that more profitable strategies are less likely to exist. An exponential distribution captures this intuition by having a monotonically decreasing probability density function.

Next, we incorporate correlations into the above framework. Among the various sources of correlations, the cross-sectional correlations among contemporaneous returns are the most important for us to take into account. This is because, unlike time-series correlations for individual return series, cross-sectional return correlations are caused by macroeconomic or market movements and can have a significant impact on multiple testing correction. Other kinds of correlations can be easily embedded into our framework as well.[49]

[47] See Appendix B for a brief discussion on the Bayesian approach for multiple testing.

[48] As shown later, we need to estimate the parameters in the mixture model based on our t-statistics sample. An over-parameterized distribution for the continuous distribution in the mixture model, albeit flexible, may result in imprecise estimates. We therefore use the simple one-parameter exponential distribution family.

[49] To incorporate the serial correlation for individual strategies, we can model them as simple autoregressive processes. To incorporate the spatial structure in the way that factors are discovered

As a starting point, we assume that the contemporaneous correlation between two strategies' returns is $\rho$. The non-contemporaneous correlations are assumed to be zero. That is,

$$\begin{aligned} \text{Corr}(X_{i,t}, X_{j,t}) &= \rho, \quad i \neq j, \\ \text{Corr}(X_{i,t}, X_{j,s}) &= 0, \quad t \neq s. \end{aligned}$$

Finally, to incorporate the impact of hidden tests, we assume that $M$ factors are tried but only factors that exceed a certain t-ratio threshold are published. We set the threshold t-statistic at 1.96 and focus on the sub-sample of factors that have a t-statistic larger than 1.96. However, as shown in Appendix A, factors with marginal t-ratios (i.e., t-ratios just above 1.96) are less likely to be published than those with larger t-ratios. Therefore, our sub-sample of published t-ratios only covers a fraction of t-ratios above 1.96 for tried factors. To overcome this missing data problem, we assume that our sample covers a fraction $r$ of t-ratios in between 1.96 and 2.57 and that all t-ratios above 2.57 are covered. We bootstrap from the existing t-ratio sample to construct the full sample. For instance, when $r = 1/2$, we simply duplicate the sample of t-ratios in between 1.96 and 2.57 and maintain the sample of t-ratios above 2.57 to construct the full sample. For the baseline case, we set $r = 1/2$, consistent with the analysis in Appendix A. We try alternative values of $r$ to see how the results change.[50]

Given the correlation structure and the sampling distribution for the means of returns, we can fully characterize the distributional properties of the cross-section of returns. We can also determine the distribution for the cross-section of t-statistics as they are functions of returns. Based our sample of t-statistics for published research, we match key sample statistics with their population counterparts in the model.

The sample statistics we choose to match are the quantiles of the sample of t-statistics and the sample size (i.e., the total number of discoveries). Two concerns motivate us to use quantiles. First, sample quantiles are less susceptible to outliers

(i.e., a group of factors discovered during a certain period can be related to each other due to the increased research intensity on that group for that period), we can impose a Markov structure on the time-series of $\mu_i$'s. See Sun and Cai (2008) for an example of spatial dependence for the null hypotheses. Lastly, to accommodate the intuition that factors within a class should be more correlated than factors across classes, we can use a block diagonal structure for the correlation matrix for strategy returns. See Harvey and Liu (2014b) for further discussion of the kinds of correlation structures that our model is able to incorporate.

[50]Our choice of the threshold t-ratio is smaller than the 2.57 threshold in Appendix A. This is for model identification purposes. With a large t-ratio threshold (e.g., t= 2.57), factors that are generated under the null hypothesis (i.e., false discoveries) are observed with a low probability as their t-ratios rarely exceed the threshold. With little presence of these factors in the sample, certain parameters (e.g., $p_0$) are poorly identified. In short, we cannot estimate the probability of drawing from the null hypothesis accurately if we rarely observe a factor that is generated from the null hypothesis. We therefore lower the threshold to allow a better estimation of the model. For more details on the selection of the threshold t-ratio, see Harvey and Liu (2014b).

compared to means and other moment-related sample statistics. Our t-ratio sample does have a few very large observations and we expect quantiles to be more useful descriptive statistics than the mean and the standard deviation. Second, simulation studies show that quantiles in our model are more sensitive to changes in parameters than other statistics. To offer a more efficient estimation of the model, we choose to focus on quantiles.

In particular, the quantities we choose to match and their values for the baseline sample (i.e., $r = 1/2$) are given by:

$$\begin{cases} \widehat{T} = \text{Total number of discoveries} = 353, \\ \widehat{Q}_1 = \text{The 20th percentile of the sample of t-statistics} = 2.39, \\ \widehat{Q}_2 = \text{The 50th percentile of the sample of t-statistics} = 3.16, \\ \widehat{Q}_3 = \text{The 90th percentile of the sample of t-statistics} = 6.34. \end{cases}$$

These three quantiles are representative of the spectrum of quantiles and can be shown to be most sensitive to parameter changes in our model. Fixing the model parameters, we can also obtain the model implied sample statistics $T, Q_1, Q_2$, and $Q_3$ through simulations.[51] The estimation works by seeking to find the set of parameters that minimizes the following objective function:

$$D(\lambda, p_0, M, \rho) = w_0(T - \widehat{T})^2 + \sum_{i=1}^{3} w_i(Q_i - \widehat{Q}_i)^2$$

where $w_0$ and $\{w_i\}_{i=1}^3$ are the weights associated with the squared distances. Motivated by the optimal weighting for the Generalized Method of Moments (GMM) estimators, we set these weights at $w_0 = 1$ and $w_1 = w_2 = w_3 = 10,000$. They can be shown to have the same magnitude as the inverses of the variances of the corresponding model implied sample statistics across a wide range of parameter values and should help improve estimation efficiency.[52]

[51] Model implied quantiles are difficult (and most likely infeasible) to calculate analytically. We obtain them through simulations. In particular, for a fixed set of parameters, we simulate 5,000 independent samples of t-statistics. For each sample, we calculate the four summary statistics. The median of these summary statistics across the 5,000 simulations are taken as the model implied statistics.

[52] We do not pursue a likelihood-based estimation. Our framework models each factor to have a mean return that follows a mixture distribution. Depending on the factor being true or not, the indicator variable in the mixture distribution can take one of two values. However, we never know if a factor is indeed true or not so the indicator variable is latent. It needs to be integrated out in a likelihood-based approach. Given that we have more than a thousand factors, the calculation of the integral for the likelihood function becomes infeasible. Existing methods such as the EM (Expectation Maximization) algorithm may mitigate the computational burden but there is an additional issue. Our model involves a search for the number of factors so the number of variables that need to be integrated out is itself random. These difficulties lead us to a GMM-based approach.

We estimate the three parameters ($\lambda, p_0$, and $M$) in the model and choose to calibrate the correlation coefficient $\rho$. In particular, for a given level of correlation $\rho$, we numerically search for the model parameters $(\lambda, p_0, M)$ that minimize the objective function $D(\lambda, p_0, M, \rho)$.

We choose to calibrate the amount of correlation because the correlation coefficient is likely to be weakly identified in this framework. Ideally, to have a better identification of $\rho$, we would like to have t-statistics that are generated from samples that have varying degrees of overlap.[53] We do not allow this in either our estimation framework (i.e., all t-statistics are generated from samples that cover the same period) or our data (we do not record the specific period for which the t-statistic is generated). As a result, our results are best interpreted as the estimated t-ratio thresholds for a hypothetical level of correlation. Nonetheless, we provide a brief discussion on the plausible levels of correlation in later sections. For additional details about the estimation method and its performance, we refer the readers to Harvey and Liu (2014b).

To investigate how correlation affects multiple testing, we follow an intuitive simulation procedure. In particular, fixing $\lambda$, $p_0$ and $M$ at their estimates, we know the data generating process for the cross-section of returns. Through simulations, we are able to calculate the previously defined Type I error rates (i.e., FWER and FDR) for any given threshold t-ratio. We search for the optimal threshold t-ratio that exactly achieves a pre-specified error rate.

## 5.2 Results

Our estimation framework assumes a balanced panel with $M$ factors and $N$ periods of returns. We need to assign a value to $N$. Returns for published works usually cover a period ranging from twenty to fifty years. In our framework, the choice of $N$ does not affect the distribution of $T_i$ under the null hypothesis (i.e., $\mu_i = 0$) but will affect $T_i$ under the alternative hypothesis (i.e., $\mu_i > 0$). When $\mu_i$ is different from zero, $T_i$ has a mean of $\mu_i/(\sigma/\sqrt{N})$. A larger $N$ reduces the noise in returns and makes it more likely for $T_i$ to be significant. To be conservative (i.e., less likely to generate significant t-ratios under the alternative hypotheses), we set $N$ at 240 (i.e., twenty years). Other specifications of $N$ change the estimate of $\lambda$ but leave the other parameters almost intact. In particular, the threshold t-ratios are little changed for alternative values of $N$.

The results are presented in Table 5. Across different correlation levels, $\lambda$ (the mean parameter for the exponential distribution that represents the mean returns

[53] Intuitively, t-statistics that are based on similar sample periods are more correlated than t-statistics that are based on distinct sample periods. Therefore, the degree of overlap in sample period helps identify the correlation coefficient. See Ferson and Chen (2013) for a similar argument on measuring the correlations among fund returns.

for true factors) is consistently estimated at 0.55% per month. This corresponds to an annual factor return of 6.6%. Therefore, we estimate the average mean returns for truly significant factors to be 6.6% per annum. Given that we standardize factor returns by an annual volatility of 15%, the average annual Sharpe ratio for these factors is 0.44 (or monthly Sharpe ratio of 0.13).[54]

For the other parameter estimates, both $p_0$ and $M$ are increasing in $\rho$. Focusing on the baseline case in Panel A and at $\rho = 0$, we estimate that researchers have tried $M = 1297$ factors and 60.4% ($= 1 - 0.396$) are true discoveries. When $\rho$ is increased to 0.60, we estimate that a total of $M = 1775$ factors have been tried and around 39.9% ($= 1 - 0.601$) are true factors. Notice that we can estimate the average total number of discoveries by $M \times (1 - p_0)$ if we were able to observe which distribution the factor mean is drawn from. This estimate is around 750 when the level of correlation is not too high (i.e., $\rho < 0.8$). Of course, in reality we cannot observe the underlying distribution for the factor mean and have to rely on the t-statistics. As a result, a significant fraction of these 750 factors are discarded because their associated t-statistics cannot overcome the threshold t-ratio.

Turning to the estimates of threshold t-ratios and focusing on FWER, we see that they are not monotonic in the level of correlation. Intuitively, two forces are at work in driving these threshold t-ratios. On the one hand, both $p_0$ and $M$ are increasing in the level of correlation. Therefore, more factors — both in absolute value and in proportion — are drawn from the null hypothesis. To control the occurrences of false discoveries based on these factors, we need a higher threshold t-ratio. On the other hand, a higher correlation among test-statistics reduces the required threshold t-ratio. In the extreme case when all test statistics are perfectly correlated, we do not need multiple testing adjustment at all. These two forces work against each other and result in the non-monotonic pattern for the threshold t-ratios under FWER. For FDR, it appears that the impact of larger $p_0$ and $M$ dominates so that the threshold t-ratios are increasing in the level of correlation.

Across various correlation specifications, our estimates show that in general a t-ratio of 3.9 and 3.0 is needed to control FWER at 5% and FDR at 1%, respectively.[55] Notice that these numbers are not far away from our previous estimates of 3.78 (Holm adjustment that controls FWER at 5%) and 3.38 (BHY adjustment that controls FDR at 1%). However, these seemingly similar numbers are generated through different mechanisms. Our current estimate assumes a certain level of correlation among returns and relies on an estimate of more than 1,300 for the total number of trials. On

[54] Our estimates are robust to the sample percentiles that we choose to match. For instance, fixing the level of correlation at 0.2, when we use the 10th together with the 50th and 90th percentiles of the sample of t-statistics, our parameter estimate is $(p_0, \lambda, M) = (0.390, 0.548, 1287)$. Alternatively, when we use the 80th together with the 20th and 50th percentiles of the sample of t-statistics, our parameter estimate is $(p_0, \lambda, M) = (0.514, 0.579, 1493)$. Both estimates are in the neighborhood of our baseline model estimate.

[55] To save space, we choose not to discuss the performance of our estimation method. Harvey and Liu (2014b) provide a detailed simulation study of our model.

the other hand, our previous calculation assumes that the 316 published factors are all the factors that have been tried but does not specify a correlation structure.

Table 5: **Estimation Results: A Model with Correlations**

We estimate the model with correlations. $r$ is the assumed proportion of missing observations for factors with a t-ratio in between 1.96 and 2.57. Panel A shows the results for the baseline case when $r = 1/2$ and Panel B shows the results for the case when $r = 2/3$. $\rho$ is the correlation coefficient between two strategy returns in the same period. $p_0$ is the probability of having a strategy that has a mean of zero. $\lambda$ is the mean parameter of the exponential distribution for the means of the true factors. $M$ is the total number of trials.

| | | | | t-ratio | | | |
|---|---|---|---|---|---|---|---|
| $\rho$ | $p_0$ | $\lambda$(%) (monthly) | $M$ | FWER(5%) | FWER(1%) | FDR(5%) | FDR(1%) |
| Panel A: $r = 1/2$ (Baseline) | | | | | | | |
| 0 | 0.396 | 0.550 | 1,297 | 3.89 | 4.28 | 2.16 | 2.88 |
| 0.2 | 0.444 | 0.555 | 1,378 | 3.91 | 4.30 | 2.27 | 2.95 |
| 0.4 | 0.485 | 0.554 | 1,477 | 3.81 | 4.23 | 2.34 | 3.05 |
| 0.6 | 0.601 | 0.555 | 1,775 | 3.67 | 4.15 | 2.43 | 3.09 |
| 0.8 | 0.840 | 0.560 | 3,110 | 3.35 | 3.89 | 2.59 | 3.25 |
| Panel B: $r = 2/3$ | | | | | | | |
| 0 | 0.683 | 0.550 | 2,458 | 4.17 | 4.55 | 2.69 | 3.30 |
| 0.2 | 0.722 | 0.551 | 2,696 | 4.15 | 4.54 | 2.76 | 3.38 |
| 0.4 | 0.773 | 0.552 | 3,031 | 4.06 | 4.45 | 2.80 | 3.40 |
| 0.6 | 0.885 | 0.562 | 4,339 | 3.86 | 4.29 | 2.91 | 3.55 |
| 0.8 | 0.922 | 0.532 | 5,392 | 3.44 | 4.00 | 2.75 | 3.39 |

## 5.3 How Large Is $\rho$?

Our sample has limitations in making a direct inference on the level of correlation. To give some guidance, we provide indirect evidence on the plausible levels of $\rho$.

First, the value of the optimized objective function sheds light on the level of $\rho$. Intuitively, a value of $\rho$ that is more consistent with the data generating process should result in a lower optimized objective function. Across the various specifications of $\rho$

in Table 5, we find that the optimized objective function reaches its lowest point when $\rho = 0.2$. Therefore, our t-ratio sample suggests a low level of correlation. However, this evidence is only suggestive given the weak identification of $\rho$ in our model.

Second, we draw on external data source to provide inference. In particular, we gain access to the S&P CAPITAL IQ database, which includes detailed information on the time-series of returns of over 400 factors for the US equity market. Calculating the average correlation among these equity risk factors for the 1985-2014 period, we estimate $\rho$ to be around 0.15.

Finally, existing studies in the literature provide guidance on the level of correlation. McLean and Pontiff (2014) estimate the correlation among anomaly returns to be around 0.05. Green, Hand and Zhang (2012) focus on accounting-based factors and find the average correlation to be between 0.06 and 0.20. Focusing on mutual fund returns, Barras, Scaillet and Wermers (2010) argue for a correlation of zero among fund returns while Ferson and Chen (2013) calibrate this number to be between 0.04 and 0.09.

Overall, we believe that the average correlation among factor returns should be low, possibly in the neighborhood of 0.20.

## 5.4 How Many True Factors Are There?

The number of true discoveries using our method seems high given that most of us have a prior that there are only a handful of true systematic risk factors. However, many of these factors that our method deems statistically true have tiny Sharpe Ratios. For example, around 70% of them have a Sharpe Ratio that is less than 0.5. From a modeling perspective, we impose a monotonic exponential density for the mean returns of true factors. Hence, by assumption the number of discoveries will be decreasing in the mean return.

Overall, statistical evidence can only get us this far in terms of getting rid of the false discoveries. It is hard to further reduce the number of discoveries to the level that we believe is true. This is a limitation not only to our framework but probably any statistical framework that relies on individual p-values. To see this, suppose the smallest t-ratio among true risk factors is 3.0 and assume our sample covers 50 risk factors that all have a t-ratio above 3.0. Then based on statistical evidence only, it is impossible to rule out any of these 50 factors from the list of true risk factors.

We agree that a further scrutiny of the factor universe is a highly meaningful exercise. There are at least two routes we can take. One route is to introduce additional testable assumptions that a systematic risk factor has to satisfy to claim significance. Pukthuanthong and Roll (2014) use the principle components of the cross-section of realized returns to impose such assumptions. The other route is to incrementally increase the factor list by thoroughly evaluating the economic contribution of a risk

factor. Harvey and Liu (2014d) provide such a framework. We expect both lines of research to help in culling the number of factors.

# 6 Conclusion

At least 316 factors have been tested to explain the cross-section of expected returns. Most of these factors have been proposed over the last ten years. Indeed, Cochrane (2011) refers to this as "a zoo of new factors". Our paper argues that it is a serious mistake to use the usual statistical significance cutoffs (e.g., a t-ratio exceeding 2.0) in asset pricing tests. Given the plethora of factors and the inevitable data mining, many of the historically discovered factors would be deemed "significant" by chance.

Our paper presents three conventional multiple testing frameworks and proposes a new one that particularly suits research in financial economics. While these frameworks differ in their assumptions, they are consistent in their conclusions. We argue that a newly discovered factor today should have a t-ratio that exceeds 3.0. We provide a time-series of recommended "cutoffs" from the first empirical test in 1967 through to present day. Many published factors fail to exceed our recommended cutoffs.

While a ratio of 3.0 (which corresponds to a p-value of 0.27%) seems like a very high hurdle, we also argue that there are good reasons to expect that 3.0 is too low. First, we only count factors that are published in prominent journals and we sample only a small fraction of the working papers. Second, there are surely many factors that were tried by empiricists, failed, and never made it to publication or even a working paper. Indeed, the culture in financial economics is to focus on the discovery of new factors. In contrast to other fields such as medical science, it is rare to publish replication studies of existing factors. Given that our count of 316 tested factors is surely too low, this means the t-ratio cutoff is likely even higher.[56]

Should a t-ratio of 3.0 be used for every factor proposed in the future? Probably not. A case can be made that a factor developed from first principles should have a lower threshold t-ratio than a factor that is discovered as a purely empirical exercise. Nevertheless, a t-ratio of 2.0 is no longer appropriate — even for factors that are derived from theory.

In medical research, the recognition of the multiple testing problem has led to the disturbing conclusion that "most claimed research findings are false" (Ioannidis (2005)). Our analysis of factor discoveries leads to the same conclusion – many of the factors discovered in the field of finance are likely false discoveries: of the 296 published

[56]In astronomy and physics, even higher threshold t-ratios are often used to control for testing multiplicity. For instance, the high profile discovery of Higgs Boson has a t-ratio of more than 5 (p-value less than 0.0001%). See ATLAS Collaboration (2012), CMS Collaboration (2012), and Harvey and Liu (2014c).

significant factors, 158 would be considered false discoveries under Bonferonni, 142 under Holm, 132 under BHY (1%) and 80 under BHY (5%). In addition, the idea that there are so many factors is inconsistent with the principal component analysis, where, perhaps there are five "statistical" common factors driving time-series variation in equity returns (Ahn, Horenstein and Wang (2012)).

The assumption that researchers follow the rules of classical statistics (e.g., randomization, unbiased reporting, etc.) is at odds with the notion of individual incentives which, ironically, is one of the fundamental premises in economics. Importantly, the optimal amount of data mining is not zero since some data mining produces knowledge. The key, as argued by Glaeser (2008), is to design appropriate statistical methods to adjust for biases, not to eliminate research initiatives. The multiple testing framework detailed in our paper is true to this advice.

Our research quantifies the warnings of both Fama (1991) and Schwert (2003). We attempt to navigate the zoo and establish new benchmarks to guide empirical asset pricing tests.

Table 6: **Factor List: Factors Sorted by Year**

An augmented version of this table is available for download and resorting. The main table includes full citations as well as hyperlinks to each of the cited articles. See http://faculty.fuqua.duke.edu/~charvey/Factor-List.xlsx.

| Year | Common # | Indi. # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 1964 | | | Market return | THEORY | Common financial | *Journal of Finance* | Sharpe (1964) |
| 1965 | | | Market return | THEORY | Common financial | *Journal of Finance* | Lintner (1965) |
| 1966 | | | Market return | THEORY | Common financial | *Econometrica* | Mossin (1966) |
| 1967 | | 1 | Total volatility | Individual stock return volatility | Individual financial | *Yale Economic Essays* | Douglas (1967) |
| 1972 | | | Market return | THEORY | Common financial | *Journal of Finance* | Heckerman (1972) |
| | | | Relative prices of consumption goods | THEORY | Common macro | | |
| 1972 | 1 | | Market return | Equity index return | Common financial | *Studies in the Theory of Capital Markets* | Black, Jensen and Scholes (1972) |
| 1972 | | | Market return | THEORY | Common financial | *Journal of Business* | Black (1972) |
| 1973 | | | State variables representing future investment opportunity | THEORY | Common financial and macro | *Econometrica* | Merton (1973) |
| 1973 | | | Market return† | Equity index return | Common financial | *Journal of Political Economy* | Fama and MacBeth (1973) |
| | 2 | | Beta squared* | Square of market beta | Common financial | | |
| | | 2 | Idiosyncratic volatility* | Residual stock volatility from CAPM | Individual financial | | |
| 1973 | | | High order market return | THEORY | Common financial | *Journal of Financial and Quantitative Analysis* | Rubinstein (1973) |
| 1974 | | | World market return | THEORY | Common financial | *Journal of Economic Theory* | Solnik (1974) |
| 1974 | | | Individual investor resources | THEORY | Common financial | *Journal of Financial Economics* | Rubinstein (1974) |
| 1975 | | 3 | Earnings growth expectations | Projecting firm earnings growth based on market beta, firm size, dividend payout ratio, leverage and earnings variability | Individual accounting | *Journal of Finance* | Gupta and Ofer (1975) |
| 1976 | | | Market return† | Equity index return | Common financial | *Journal of Finance* | Kraus and Litzenberger (1976) |
| | 3 | | Squared market return* | Square of equity index return | Common financial | | |
| 1977 | | 4 | PE ratio | Firm price-to-earnings ratio | Individual accounting | *Journal of Finance* | Basu (1977) |
| 1978 | | | Marginal rate of substitution | THEORY | Common macro | *Econometrica* | Lucas (1978) |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 1979 | | 5 | Dividend yield | Dividend per share divided by share price | Individual accounting | *Journal of Financial Economics* | Litzenberger and Ramaswamy (1979) |
| | | | Market return† | Equity index return | Common financial | | |
| 1979 | | | Aggregate real consumption growth | THEORY | Common macro | *Journal of Financial Economics* | Breeden (1979) |
| 1980 | | | Short sale restrictions | THEORY | Individual microstructure | *Journal of Finance* | Jarrow (1980) |
| 1981 | | | Market return†‡ | Equity index return | Common financial | *Journal of Finance* | Fogler, John and Tipton (1981)[a] |
| | | | Treasury bond return‡ | 3-month US Treasury bill return | Common financial | | |
| | | | Corporate bond return‡ | Index of long-term Aa utility bonds with deferred calls returns | Common financial | | |
| 1981 | 4 | | Treasury bill return | Principal components extracted from returns of Treasury bills | Common financial | *Journal of Finance* | Oldfield and Rogalski (1981) |
| 1981 | | | World consumption | THEORY | Common macro | *Journal of Financial Economics* | Stulz (1981) |
| 1981 | | | Transaction costs | THEORY | Individual microstructure | *Journal of Finance* | Mayshar (1981) |
| 1981 | | 6 | Firm size | Market value of firm stocks | Individual financial | *Journal of Financial Economics* | Banz (1981) |
| 1981 | | 7 | Short interest | Equity short interest | Individual microstructure | *Journal of Financial and Quantitative Analysis* | Figlewski (1981) |
| 1982 | | | Individual consumer's wealth | THEORY | Common financial | *Journal of Business* | Constantinides (1982) |
| 1983 | | 8 | EP ratio | Firm earnings-to-price ratio | Individual accounting | *Journal of Financial Economics* | Basu (1983) |
| 1983 | | | Foreign exchange rate change | THEORY | Common financial | *Journal of Finance* | Adler and Dumas (1983) |
| 1983 | | | Institutional holding‡ | Institutional concentration rankings from Standard and Poor's | Individual other | *Financial Analyst Journal* | Arbel, Carvell and Strebel (1983) |
| 1984 | | | Earnings expectations‡ | Consensus earnings expectations | Individual accounting | *Financial Analyst Journal* | Hawkins, Chamberlin and Daniel (1984) |
| 1984 | | | New listings announcement‡ | Announcement that a company has filed a formal application to list on the NYSE | Individual accounting | *Financial Analyst Journal* | McConnell and Sanger (1984) |
| 1985 | | | Market return† | Equity index return | Common financial | *Journal of Financial Economics* | Chan, Chen and Hsieh (1985) |
| | 5 | | Industrial production growth | Seasonally adjusted monthly growth rate of industrial production | Common macro | | |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| | 6 | | Change in expected inflation* | Change in expected inflation as defined in Fama and Gibbons (1984) | Common macro | | |
| | 7 | | Unanticipated inflation | Realized minus expected inflation | Common macro | | |
| | 8 | | Credit premium | Risk premium measured as difference in return between "under Baa" bond portfolio and long-term government bond portfolio | Common financial | | |
| | 9 | | Term structure* | Yield curve slope measured as difference in return between long-term government bond and 1-month Treasury bill | Common financial | | |
| 1985 | | 9 | Long-term return reversal | Long-term past abnormal return | Individual other | *Journal of Finance* | Bondt and Thaler (1985) |
| 1985 | | | Investment opportunity change | THEORY | Common financial | *Econometrica* | Cox, Ingersoll and Ross (1985) |
| 1986 | | | Transaction costs | THEORY | Common microstructure | *Journal of Financial Economics* | Amihud and Mendelson (1986) |
| 1986 | | | Transaction costs | THEORY | Common microstructure | *Journal of Political Economy* | Constantinides (1986) |
| 1986 | | | Expected inflation | THEORY | Common macro | *Journal of Finance* | Stulz (1986) |
| 1986 | 10 | | Long-term interest rate | Change in the yield of long-term government bonds | Common financial | *Journal of Finance* | Sweeney and Warga (1986) |
| 1986 | | | Industrial production growth† | Seasonally adjusted monthly growth rate of industrial production | Common macro | *Journal of Business* | Chen, Roll and Ross (1986) |
| | | | Credit premium† | Risk premium measured as difference in return between "under Baa" bond portfolio and long-term government bond portfolio | Common financial | | |
| | | | Term structure† | Yield curve slope measured as difference in return between long-term government bond and 1-month Treasury bill | Common financial | | |
| | | | Unanticipated inflation† | Realized minus expected inflation | Common macro | | |
| | | | Change in expected inflation† | Changes in expected inflation as defined in Fama and Gibbons (1984) | Common macro | | |
| | 11 | | Change in oil prices* | Growth rate in oil prices | Common macro | | |
| 1988 | | 10 | Debt to equity ratio | Non-common equity liabilities to equity | Individual accounting | *Journal of Finance* | Bhandari (1988) |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 1988 | | | Long-term growth forecasts‡ | Long-term growth forecasts proxied by the five-year earnings per share growth rate forecasts | Individual accounting | *Financial Analyst Journal* | Bauman and Dowen (1988) |
| 1989 | 12 | | Consumption growth | Per capita real consumption growth | Common macro | *Journal of Finance* | Breeden, Gibbons and Litzenberger (1989) |
| 1989 | | 11 | Illiquidity | Illiquidity proxied by bid-ask spread | Individual microstructure | *Journal of Finance* | Amihud and Mendelson (1989) |
| 1989 | | 12 | Predicted earnings change | Predicted earnings change in one year based on a financial statement analysis that combines a large set of financial statement items | Individual accounting | *Journal of Accounting & Economics* | Ou and Penman (1989) |
| 1990 | | 13 | Return predictability | Short-term (one month) and long-term (twelve months) serial correlations in returns | Individual financial | *Journal of Finance* | Jegadeesh (1990) |
| 1991 | | | Market return† | Equity index return | Common financial | *Journal of Political Economy* | Ferson and Harvey (1991) |
| | | | Consumption growth† | Real per capita growth of personal consumption expenditures for nondurables & services | Common macro | | |
| | | | Credit spread† | Baa corporate bond return less monthly long-term government bond return | Common financial | | |
| | 13 | | Change in the slope of the yield curve | Change in the difference between a 10-year Treasury bond yield and a 3-month Treasury bill yield | Common financial | | |
| | | | Unexpected inflation† | Difference between actual and time-series forecasts of inflation rate | Common macro | | |
| | 14 | | Real short rate | One-month Treasury bill return less inflation rate | Common financial | | |
| 1992 | 15 | | Size | Return on a zero-investment portfolio long in small stocks and short in large stocks | Common accounting | *Journal of Finance* | Fama and French (1992)[b] |
| | 16 | | Value | Return on a zero-investment portfolio long in growth stocks and short in value stocks | Common accounting | | |
| 1992 | | | Return momentum‡ | Size and beta adjusted mean prior five-year returns | Individual financial | *Journal of Financial Economics* | Chopra, Lakonishok and Ritter (1992) |
| 1992 | | | Predicted return signs‡ | Return signs predicted by a logit model using financial ratios | Individual accounting | *Journal of Accounting & Economics* | Holthausen and Larcker (1992) |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 1993 | | 14 | Return momentum | Past stock returns | Individual other | *Journal of Finance* | Jegadeesh and Titman (1993) |
| 1993 | | | Returns on S&P stocks‡ | Returns on S&P stocks | Common financial | *Review of Financial Studies* | Elton, Gruber, Das and Hlavka (1993) |
| | | | Returns on non-S&P stocks‡ | Returns on non-S&P stocks | Common financial | | |
| 1993 | | | High order market and bond return‡ | High order equity index returns and bond returns | Common financial | *Journal of Finance* | Bansal and Viswanathan (1993)[c] |
| 1993 | | | Market return† | Equity index return | Common financial | *Journal of Financial Economics* | Fama and French (1993) |
| | | | Size† | Return on a zero-investment portfolio long in small stocks and short in large stocks | Common accounting | | |
| | | | Value† | Return on a zero-investment portfolio long in growth stocks and short in value stocks | Common accounting | | |
| | | | Term structure† | Difference in return between long-term government bond and one-month Treasury bill | Common financial | | |
| | | | Credit risk† | Difference in return between long-term corporate bond and long-term government bond | Common financial | | |
| 1993 | | | World equity return‡ | US dollar return of the MSCI world equity market in excess of a short-term interest rate | Common financial | *Review of Financial Studies* | Ferson and Harvey (1993)[d] |
| | | | Change in weighted exchange rate‡ | Log first difference of the trade-weighted US dollar price of ten industrialized countries' currencies | Common financial | | |
| | | | Change in long-term inflationary expectations‡ | Change in long-term inflationary expectations | Common macro | | |
| | | | Weighted real short-term interest rate‡ | GDP weighted average of short-term interest rates in G-7 countries | Common financial | | |
| | | | Change in oil price†‡ | Change in the monthly average US dollar price per barrel of crude oil | Common macro | | |
| | | | Change in the Eurodollar-Treasury yield spread‡ | First difference of the spread between the 90-day Eurodollar yield and the 90-day Treasury-bill yield | Common financial | | |
| | | | Change in G-7 industrial production‡ | Change in G-7 industrial production | Common macro | | |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| | | | Unexpected inflation for the G-7 countries‡ | Unexpected inflation based on a time-series model on an aggregate G-7 inflation rate | Common macro | | |
| 1994 | 17 | | World equity return | US dollar return of the MSCI world equity market in excess of a short-term interest rate | Common financial | *Journal of Banking and Finance* | Ferson and Harvey (1994) |
| | 18 | | Change in weighted exchange rate* | Log first difference of the trade-weighted US dollar price of ten industrialized countries' currencies | Common financial | | |
| | 19 | | Change in long-term inflationary expectations* | Change in long-term inflationary expectations | Common macro | | |
| | | | Change in oil price*† | Change in the monthly average US dollar price per barrel of crude oil | Common macro | | |
| 1994 | 20 | | Tax rate for capital gains | Short-term capital gains tax rate | Common accounting | *Journal of Finance* | Bossaerts and Dammon (1994) |
| | 21 | | Tax rate for dividend | Dividend tax rate | Common accounting | | |
| 1995 | 22 | | Change in expected inflation | Change in expectation from economic surveys | Common macro | *Journal of Finance* | Elton, Gruber and Blake (1995) |
| | 23 | | Change in expected GNP | Change in expectation from economic surveys | Common macro | | |
| 1995 | | 15 | New public stock issuance | New public stock issuance | Individual accounting | *Journal of Finance* | Loughran and Ritter (1995) |
| 1995 | | 16 | Dividend initiations | Initiations of cash dividend payments | Individual financial | *Journal of Finance* | Michaely, Thaler and Womack (1995) |
| | | 17 | Dividend omissions | Omissions of cash dividend payments | Individual financial | | |
| 1995 | | | Seasoned equity offerings‡ | Whether a firm makes seasoned equity offerings | Individual financial | *Journal of Financial Economics* | Spiess and Affleck-Graves (1995) |
| 1996 | 24 | | Money growth | M2 or M3 minus currency, divided by total population | Common macro | *Journal of Finance* | Chan, Foresi and Lang (1996) |
| 1996 | 25 | | Returns on physical investment | Inferred from investment data via a production function | Common macro | *Journal of Political Economy* | Cochrane (1996) |
| 1996 | | | Market return† | Equity index return | Common financial | *Journal of Political Economy* | Campbell (1996) |
| | 26 | | Labor income | Real labor income growth rate | Common macro | | |
| | | | Dividend yield† | Dividend yield on value-weighted index | Common financial | | |
| | | | Interest rate† | Treasury bill rate less 1-year moving average | Common financial | | |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| | | | Term structure† | Long-short government bond yield spread | Common financial | | |
| 1996 | | | Market return† | Equity index return | Common financial | *Journal of Finance* | Jagannathan and Wang (1996) |
| | | | Slope of yield curve† | Long-short government bond yield spread | Common financial | | |
| | | | Labor income† | Real labor income growth rate | Common macro | | |
| 1996 | | 18 | Earnings forecasts | Errors in analysts' forecasts on earnings growth | Individual accounting | *Journal of Finance* | Porta (1996) |
| 1996 | | 19 | R&D capital | R&D capital over total assets | Individual accounting | *Journal of Accounting & Economics* | Lev and Sougiannis (1996) |
| 1996 | | 20 | Accruals | Accruals defined by the change in non-cash current assets, less the change in current liabilities, less depreciation expense, all divided by average total assets | Individual accounting | *Accounting Review* | Sloan (1996) |
| 1996 | | 21 | Buy recommendations | Buy recommendations from security analysts | Individual financial | *Journal of Finance* | Womack (1996) |
| | | 22 | Sell recommendations | Sell recommendations from security analysts | Individual financial | | |
| 1996 | | 23 | Credit rating | Institutional investor country credit rating from semi-annual survey | Individual other | *Journal of Portfolio Management* | Erb, Harvey and Viskanta (1996) |
| 1996 | | 24 | Illiquidity | Derivative transaction price with respect to signed trade size | Individual microstructure | *Journal of Financial Economics* | Brennan and Subrahmanyam (1996) |
| 1997 | | | Nonlinear functions of consumption growth‡ | Low order orthonormal polynomials of current and future consumption growth | Common macro | *Journal of Finance* | Chapman (1997)[e] |
| 1997 | | | Opportunistic strategy return‡ | Return for hedge funds that follow an opportunistic strategy | Common financial | *Review of Financial Studies* | Fung and Hsieh (1997)[f] |
| | | | Global/macro strategy return‡ | Return for hedge funds that follow a global/macro strategy | Common financial | | |
| | | | Value strategy return‡ | Return for hedge funds that follow a value strategy | Common financial | | |
| | | | Trend following strategy return‡ | Return for hedge funds that follow a trend following strategy | Common financial | | |
| | | | Distressed investment strategy return‡ | Return for hedge funds that follow a distressed investment strategy | Common financial | | |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 1997 | | | Size† | Return on a zero-investment portfolio long in small stocks and short in large stocks | Common accounting | *Journal of Finance* | Carhart (1997) |
| | | | Value† | Return on a zero-investment portfolio long in growth stocks and short in value stocks | Common accounting | | |
| | | | Market return† | Equity index return | Common financial | | |
| | 27 | | Momentum | Return on a zero-investment portfolio long in past winners and short in past losers | Common other | | |
| 1997 | | | Size† | Market value of equity | Individual accounting | *Journal of Financial Economics* | Brennan, Chordia and Subrahmanyam (1997) |
| | | | Book-to-market ratio† | Book value of equity plus deferred taxes to market value of equity | Individual accounting | | |
| | | | Momentum† | Past cumulative stock return | Individual financial | | |
| | | 25 | Trading volume | Dollar volume traded per month | Individual microstructure | | |
| 1997 | | 26 | Disclosure level | Voluntary disclosure level of manufacturing firms' annual reports | Individual accounting | *Accounting Review* | Botosan (1997) |
| 1997 | | 27 | Earnings forecast uncertainty | Standard deviation of earnings forecasts | Individual accounting | *Journal of Financial Research* | Ackert and Athanassakos (1997) |
| 1997 | | | Size† | Market value of equity | Individual accounting | *Journal of Finance* | Daniel and Titman (1997) |
| | | | Value† | Book value of equity plus deferred taxes to market value of equity | Individual accounting | | |
| 1997 | | | Earnings management likelihood‡ | Earnings management likelihood obtained by regressiong realized violators of Generally Accepted Accounting Principles on firm characteristics | Individual accounting | *Journal of Accounting and Public Policy* | Beneish (1997) |
| 1997 | | 28 | Corporate acquisitions | Difference between stock mergers and cash tender offers for corporate acquisitions | Individual financial | *Journal of Finance* | Loughran and Vijh (1997) |
| 1998 | | | Fundamental analysis‡ | Investment signals constructed using a collection of variables that relate to contemporaneous changes in inventories, accounts receivables, gross margins, selling expenses, capital expenditures, effective tax rates, inventory methods, audit qualifications, and labor force sales productivity | Individual accounting | *Accounting Review* | Abarbanell and Bushee (1998) |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 1998 | | | Firm fundamental value‡ | Firms' fundamental values estimated from I/B/E/S consensus forecasts and a residual income model | Individual accounting | *Journal of Accounting and Economics* | Frankel and Lee (1998) |
| 1998 | | 29 | Bankruptcy risk | The probability of bankruptcy from Altman (1968) | Individual financial | *Journal of Finance* | Ilia (1998) |
| 1998 | | 30 | Illiquidity | Liquidity proxied by the turnover rate: number of shares traded as a fraction of the number of shares outstanding | Individual microstructure | *Journal of Financial Markets* | Datar, Naik and Radcliffe (1998) |
| 1999 | 28 | | Fitted return based on predictive regressions | Expected portfolio return obtained by projecting historical returns on lagged macro instruments, including term spreads, dividend yield, credit spread and short-term Treasury bill | Common financial | *Journal of Finance* | Ferson and Harvey (1999) |
| 1999 | | 31 | Industry momentum | Industry-wide momentum returns | Individual other | *Journal of Finance* | Moskowitz and Grinblatt (1999) |
| 1999 | | | Debt offerings‡ | Whether a firm makes straight and convertible debt offerings | Individual financial | *Journal of Financial Economics* | Spiess and Affleck-Graves (1999) |
| 2000 | 29 | | Entrepreneur income | Proprietary income of entrepreneurs | Common financial | *Journal of Finance* | Heaton and Lucas (2000) |
| 2000 | 30 | | Coskewness | Excess return on a portfolio which long stocks with low past coskewness | Common financial | *Journal of Finance* | Harvey and Siddique (2000) |
| 2000 | | 32 | Trading volume | Past trading volume | Individual microstructure | *Journal of Finance* | Lee and Swaminathan (2000) |
| 2000 | | 33 | Within-industry size | Difference between firm size and average firm size within the industry | Individual financial | *Working Paper* | Asness, Porter and Stevens (2000) |
| | | 34 | Within-industry value | Difference between firm book-to-market ratio and average book-to-market ratio within the industry | Individual accounting | | |
| | | 35 | Within-industry cashflow to price ratio | Difference between firm cashflow to price ratio and average cashflow to price ratio within the industry | Individual accounting | | |
| | | 36 | Within-industry percent change in employees | Difference between firm percent change in employees and average percent change in employees within the industry | Individual accounting | | |
| | | 37 | Within-industry momentum | Difference between firm past stock prices and average past stock prices within the industry | Individual financial | | |
| 2000 | | 38 | Financial statement information | A composite score based on historical financial statement that separates winners from losers | Individual accounting | *Journal of Accounting Research* | Piotroski (2000) |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2001 | | | Consumption growth† | Per capita real consumption growth rate | Common macro | *Journal of Political Economy* | Lettau and Ludvigson (2001) |
| | 31 | | Consumption-wealth ratio | Proxied by a weighted average of human and nonhuman wealth | Common macro | | |
| 2001 | | 39 | Level of liquidity | Level of dollar trading volume and share turnover | Individual microstructure | *Journal of Financial Economics* | Chordia, Subrahmanyam and Anshuman (2001) |
| | | 40 | Variability of liquidity | Volatility of dollar trading volume and share turnover | Individual microstructure | | |
| 2001 | | 41 | Financial constraints | Measure financial constraints with Kaplan and Zingales (1997) index | Individual financial | *Review of Financial Studies* | Lamont, Polk and Saa-Requejo (2001) |
| 2001 | | | Straddle return‡ | Lookback straddles' returns constructed based on option prices | Common financial | *Review of Financial Studies* | Fung and Hsieh (2001) |
| 2001 | | | Consensus recommendations* | Consensus recommendations measured by the average analyst recommendations | Individual accounting | *Journal of Finance* | Barber, Lehavy, McNichols and Trueman (2001) |
| 2001 | | 42 | Bond rating changes | Moody's bond ratings changes | Individual financial | *Journal of Finance* | Dichev and Piotroski (2001) |
| 2001 | | 43 | Analysts' forecasts | Financial analysts' forecasts of annual earnings | Individual accounting | *Accounting Review* | Pieter, Lo and Pfeiffer (2001) |
| 2001 | | 44 | Institutional ownership | Institutional holdings of firm assets | Individual accounting | *Quarterly Journal of Economics* | Gompers and Metrick (2001) |
| 2002 | | | Market return† | Equity index return | Common financial | *Journal of Finance* | Dittmar (2002) |
| | | | Squared market return† | Squared equity index return | Common financial | | |
| | | | Labor income growth† | Smoothed labor income growth rate | Common financial | | |
| | 32 | | Squared labor income growth | Squared smoothed labor income growth rate | Common financial | | |
| 2002 | | 45 | Distress risk | Distress risk as proxied by Ohlson's O-score | Individual financial | *Journal of Finance* | Griffin and Lemmon (2002) |
| 2002 | | 46 | Analyst dispersion | Dispersion in analysts' earnings forecasts | Individual behavioral | *Journal of Finance* | Diether, Malloy and Scherbina (2002) |
| 2002 | | 47 | Breadth of ownership | Ratio of the number of mutual funds holding long positions in the stock to total number of mutual funds | Individual microstructure | *Journal of Financial Economics* | Chen, Hong and Stein (2002) |
| 2002 | | 48 | Information risk | Probability of information-based trading for individual stock | Individual microstructure | *Journal of Finance* | Easley, Hvidkjaer and O'Hara (2002) |
| 2002 | | 49 | Short-sale constraints | Shorting costs for NYSE stocks | Individual microstructure | *Journal of Financial Economics* | Jones and Lamont (2002) |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2002 | | 50 | Earnings sustainability | A summary score based on firm fundamentals that informs about the sustainability of earning | Individual accounting | *Working Paper* | Penman and Zhang (2002) |
| 2002 | 33 | | Market illiquidity | Average over the year of the daily ratio of the stock's absolute return to its dollar trading volume | Common microstructure | *Journal of Financial Markets* | Amihud (2002) |
| 2003 | 34 | | GDP growth news | GDP growth news obtained from predictive regressions on lagged equity and fixed-income portfolios | Common macro | *Journal of Financial Economics* | Vassalou (2003) |
| 2003 | 35 | | Market liquidity | Aggregated liquidity based on firm future excess stock return regressed on current signed excess return times trading volume | Common microstructure | *Journal of Political Economy* | Pastor and Stambaugh (2003) |
| 2003 | | | Idiosyncratic return volatility† | Residual variance obtained by regressing daily stock returns on market index return | Individual financial | *Journal of Financial Economics* | Ali, Hwang and Trombley (2003) |
| | | | Transaction costs† | Bid-ask spread, volume, etc. | Individual microstructure | | |
| | | | Investor sophistication† | Number of analysts or institutional owners | Individual accounting | | |
| 2003 | | 51 | Shareholder rights | Shareholder rights as proxied by an index using 24 governance rules | Individual accounting | *Quarterly Journal of Economics* | Gompers, Ishii and Metrick (2003) |
| 2003 | | 52 | Excluded expenses | Excluded expenses in firm's earnings reports | Individual accounting | *Review of Accounting Studies* | Jeffrey, Lundholm and Soliman (2003) |
| 2003 | | 53 | Growth in long-term net operating assets | Growth in long-term net operating assets | Individual accounting | *Accounting Review* | Fairfield, Whisenant and Yohn (2003) |
| 2003 | | 54 | Order backlog | Order backlog divided by average total assets, transformed to a scaled-decile variable | Individual accounting | *Review of Accounting Studies* | Rajgopal, Shevlin and Venkatachalam (2003) |
| 2003 | | 55 | Return consistency | Consecutive returns with the same sign | Individual financial | *Journal of Behavioral Finance* | Watkins (2003) |
| 2004 | 36 | | Idiosyncratic consumption | Cross-sectional consumption growth variance | Common macro | *Journal of Finance* | Jacobs and Wang (2004) |
| 2004 | 37 | | Cash flow news | News about future market cash flow | Common financial | *American Economic Review* | Campbell and Vuolteenaho (2004) |
| | 38 | | Discount rate news | News about future market discount rate | Common financial | | |
| 2004 | | | Market return† | Equity index return | Common financial | *Review of Financial Studies* | Vanden (2004)[g] |
| | 39 | | Index option returns | Return on S&P 500 index option | Common financial | | |

... continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2004 | 40 | | Default risk | Firm default likelihood using Merton's option pricing model | Common financial | *Journal of Finance* | Vassalou and Xing (2004) |
| 2004 | 41 | | Real interest rate | Real interest rates extracted from a time-series model of bond yields and expected inflation | Common financial | *Journal of Finance* | Brennan, Wang and Xia (2004) |
| | 42 | | Maximum Sharpe ratio portfolio | Maximum Sharpe ratio portfolio extracted from a time-series model of bond yields and expected inflation | Common financial | | |
| 2004 | 43 | | Return reversals at the style level | Zero-investment portfolios sorted based on past return performance at the style level | Common other | *Journal of Financial Economics* | Teo and Woo (2004) |
| 2004 | | 56 | Unexpected change in R&D | Unexpected change in firm research and expenditures | Individual accounting | *Journal of Finance* | Allan, Maxwell and Siddique (2004) |
| 2004 | | 57 | 52-week high | Nearness to the 52-week high price | Individual financial | *Journal of Finance* | George and Hwang (2004) |
| 2004 | | 58 | Analysts' recommendations | Consensus analysts' recommendations from sell-side firms | Individual accounting | *Journal of Finance* | Jegadeesh, Kim, Krische and Lee (2004) |
| 2004 | | 59 | Put-call parity | Violations of put-call parity | Individual financial | *Journal of Financial Economics* | Ofek, Richardson and Whitelaw (2004) |
| 2004 | | 60 | Abnormal capital investment | Past year capital expenditures scaled by average capital expenditures for previous three years | Individual accounting | *Journal of Financial and Quantitative Analysis* | Titman, Wei and Xie (2004) |
| 2004 | | 61 | Balance sheet optimism | Net operating assets scaled by total assets | Individual accounting | *Journal of Accounting and Economics* | Hirshleifer, Hou, Teoh and Zhang (2004) |
| 2005 | 44 | | Long-horizon consumption growth | Three-year consumption growth rate | Common macro | *Journal of Political Economy* | Parker and Julliard (2005) |
| 2005 | 45 | | Long-run consumption | Cash flow risk measured by cointegration residual with aggregate consumption | Common macro | *Journal of Finance* | Bansal, Dittmar and Lundblad (2005) |
| 2005 | 46 | | Housing price ratio | Ratio of housing to human wealth | Common financial | *Journal of Finance* | Lustig and Nieuwerburgh (2005) |
| 2005 | | 62 | External corporate governance | Proxies for corporate control | Individual accounting | *Journal of Finance* | Cremers and Nair (2005) |
| | | 63 | Internal corporate governance | Proxies for share-holder activism | Individual accounting | | |
| 2005 | | | Market return† | Equity index return | Common financial | *Journal of Financial Economics* | Acharya and Pedersen (2005)[h] |
| | 47 | | Market liquidity* | Value-weighted individual stock illiquidity as defined in Amihud (2002) | Common microstructure | | |
| | | 64 | Individual stock liquidity | Individual stock illiquidity as defined in Amihud (2002) | Individual microstructure | | |

... continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2005 | | 65 | Price delay | Delay in a stock price's response to information | Individual microstructure | *Review of Financial Studies* | Hou and Moskowitz (2005) |
| 2005 | | 66 | Heterogeneous beliefs | Factors constructed from disagreement among analysts about expected short- and long-term earnings | Individual financial | *Review of Financial Studies* | Anderson, Ghysels and Juergens (2005) |
| 2005 | | 67 | Short-sale constraints | Short-sale constraint proxied by Institutional ownership | Individual microstructure | *Journal of Financial Economics* | Nagel (2005) |
| 2005 | | 68 | Short-sale constraints | Short-sale constraint proxied by short interest and institutional ownership | Individual microstructure | *Journal of Financial Economics* | Asquith, Pathak and Ritter (2005) |
| 2005 | | 69 | Patent citation | Change of patent citation impact deflated by average total assets | Individual other | *Journal of Accounting, Auditing & Finance* | Gu (2005) |
| 2005 | | 70 | Information uncertainty | Information uncertainty proxied by firm age, return volatility, trading volume or cash flow duration | Individual financial | *Review of Accounting Studies* | Jiang, Lee and Zhang (2005) |
| 2005 | | 71 | Adjusted R&D | Adjusted R&D that incorporates capitalization and amortization | Individual accounting | *Working Paper* | Lev, Nissim and Thomas (2005) |
| 2005 | | 72 | R&D reporting biases | R&D reporting biases proxied by the difference between R&D growth and earnings growth | Individual accounting | *Contemporary Accounting Research* | Lev, Sarath and Sougiannis (2005) |
| 2005 | | 73 | Growth index | A combined index constructed based on earnings, cash flows, earnings stability, growth stability and intensity of R&D, capital expenditure and advertising | Individual accounting | Review of Accounting Studies | Mohanram (2005) |
| 2006 | | | Market return† | Equity index return and its square | Common financial | *Review of Financial Studies* | Vanden (2006)[i] |
| | | | Index option return† | Index option return and its square | Common financial | | |
| | | | Interaction between index and option return‡ | Product of market and option returns | Common financial | | |
| 2006 | 48 | | Financing frictions | Default premium | Common financial | *Review of Financial Studies* | Gomes, Yaron and Zhang (2006) |
| 2006 | 49 | | Investment growth by households* | Household investment growth | Common macro | *Journal of Business* | Li, Vassalou and Xing (2006) |
| | 50 | | Investment growth by nonfarm nonfinancial corporate firms | Nonfarm nonfinancial corporate firms investment growth | Common macro | | |
| | 51 | | Investment growth by nonfarm noncorporate business | Nonfarm noncorporate business investment growth | Common macro | | |
| | 52 | | Investment growth by financial firms | Financial firms investment growth | Common macro | | |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2006 | | | Third to tenth power of market return‡ | Third to tenth power of market return | Common financial | *Journal of Business* | Chung, Johnson and Schill (2006)[j] |
| 2006 | | 74 | Financial constraints | Constraint index estimated from a firm's investment Euler equation | Individual financial | *Review of Financial Studies* | Whited and Wu (2006) |
| 2006 | 53 | | Downside risk | Correlation with index return conditional on index return being below a threshold value | Common financial | *Review of Financial Studies* | Ang, Chen and Xing (2006) |
| 2006 | 54 | | Systematic volatility | Aggregate volatility relative to Fama and French (1992) three-factor model | Common financial | *Journal of Finance* | Ang, Hodrick, Xing and Zhang (2006) |
| | | 75 | Idiosyncratic volatility | Idiosyncratic volatility relative to Fama and French (1992) three-factor model | Individual financial | | |
| 2006 | 55 | | Investor sentiment | Composite sentiment index based on various sentiment measures | Common behavioral | *Journal of Finance* | Baker and Wurgler (2006) |
| 2006 | 56 | | Retail investor sentiment | Systematic retail trading based on transaction data | Common behavioral | *Journal of Finance* | Kumar and Lee (2006) |
| 2006 | 57 | | Durable and nondurable consumption growth | Durable and nondurable consumption growth | Common macro | *Journal of Finance* | Yogo (2006) |
| 2006 | | | Market return† | Equity index return | Common financial | *Journal of Finance* | Lo and Wang (2006) |
| | 58 | | Trading volume | Return on a hedge portfolio constructed using trading volume and market returns | Common microstructure | | |
| 2006 | 59 | | Liquidity | Market-wide liquidity constructed first by decomposing firm-level liquidity into variable and fixed price effects then averaging the variable component | Common microstructure | *Journal of Financial Economics* | Sadka (2006) |
| 2006 | 60 | | Earnings | Return on a zero-investment portfolio long in stocks with high earnings surprises and short in stocks with low earnings surprises | Common accounting | *Journal of Financial Economics* | Chordia and Shivakumar (2006) |
| 2006 | 61 | | Liquidity | Turnover-adjusted number of days with zero trading over the prior 12 months | Common microstructure | *Journal of Financial Economics* | Liu (2006) |
| 2006 | | 76 | Capital investment | Capital expenditure growth | Individual accounting | *Journal of Finance* | Anderson and Garcia-Feijoo (2006) |
| 2006 | | 77 | Industry concentration | Industry concentration as proxied by the Herfindahl index | Individual accounting | *Journal of Finance* | Hou and Robinson (2006) |
| 2006 | | 78 | Environment indicator* | A composite index measuring a firm's environmental responsibility | Individual other | *Financial Management* | Brammer, Brooks and Pavelin (2006) |

... continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| | | 79 | Employment indicator* | A composite index measuring employee responsibility | Individual other | | |
| | | 80 | Community indicator* | A composite index measuring community responsiveness | Individual other | | |
| 2006 | | 81 | Intangible information | Residuals from cross-sectional regression of firm returns on fundamental growth measures | Individual accounting | *Journal of Finance* | Daniel and Titman (2006) |
| 2006 | | 82 | Profitability | Expected earnings growth | Individual accounting | *Journal of Financial Economics* | Fama and French (2006) |
| | | 83 | Investment* | Expected growth in book equity | Individual accounting | | |
| | | | Book-to-market† | Book value of equity plus deferred taxes to market value of equity | Individual accounting | | |
| 2006 | | 84 | Net financing | Net amount of cash flow received from external financing | Individual accounting | *Journal of Accounting and Economics* | Bradshaw, Richardson and Sloan (2006) |
| 2006 | | 85 | Forecasted earnings per share | Analysts' forecasted earnings per share | Individual accounting | *Working Paper* | Cen, Wei and Zhang (2006) |
| 2006 | | 86 | Pension plan funding | Pension plan funding status calculated as the difference between the fair value of plan assets and the projected benefit obligation, divided by market capitalization | Individual accounting | *Journal of Finance* | Franzoni and Marin (2006) |
| 2006 | | 87 | Acceleration | Firm's ranking on change in six-month momentum relative to the cross-section of other firms | Individual financial | *Working Paper* | Gettleman and Marks (2006) |
| 2006 | | 88 | Unexpected earnings' autocorrelations | Standardized unexpected earnings' autocorrelations via the sign of the most recent earnings realization | Individual accounting | *Journal of Accounting Research* | Narayanamoorthy (2006) |
| 2007 | 62 | | Payout yield | Return on a zero-investment portfolio long in high-yield stocks and short in low-yield stocks | Common accounting | *Journal of Finance* | Boudoukh, Michaely, Richardson and Roberts (2007) |
| 2007 | 63 | | Productivity | Productivity level as in King and Rebelo (2000) | Common macro | *Journal of Financial Economics* | Balvers and Huang (2007) |
| | 64 | | Capital stock | Quarterly capital stock interpolated from annual data | Common macro | | |
| 2007 | 65 | | Fourth-quarter to fourth-quarter consumption growth | Fourth-quarter to fourth-quarter consumption growth rate | Common macro | *Journal of Finance* | Jagannathan and Wang (2007) |
| 2007 | | 89 | Credit rating | S&P firm credit rating | Individual financial | *Journal of Finance* | Avramov, Chordia, Jostova and Philipov (2007) |
| 2007 | | 90 | Trader composition | Fraction of total trading volume of a stock from institutional trading | Individual microstructure | *Working Paper* | Shu (2007) |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2007 | | 91 | Change in order backlog | Change in order backlog | Individual accounting | *Seoul Journal of Business* | Baik and Ahn (2007) |
| 2007 | | 92 | Firm productivity | Firm productivity measured by returns on invested capital | Individual accounting | *Working Paper* | Brown and Rowe (2007) |
| 2007 | | 93 | Insider forecasts of firm volatility | Future firm volatility obtained from executive stock options | Individual financial | *Working Paper* | James, Fodor and Peterson (2007) |
| 2007 | | 94 | Ticker symbol | Creativity in stocks' ticker symbols | Individual other | *Quarterly Review of Economics & Finance* | Head, Smith and Wilson (2007) |
| 2007 | 66 | | Earnings cyclicality | Sensitivity of earnings to changes in aggregate total factor productivity | Common macro | *Working Paper* | Gourio (2007) |
| 2008 | 67 | | Market volatility innovation | Difference in monthly average of squared daily return differences | Common financial | *Review of Financial Studies* | Kumar, Sorescu, Boehme and Danielsen (2008) |
| | | 95 | Firm age | Firm's public listing age | Individual accounting | | |
| | | | Market return† | Equity index return | Common financial | | |
| | | 96 | Interaction between market volatility and firm age | Product of market volatility and firm age | Individual accounting | | |
| 2008 | 68 | | Short-run market volatility | High frequency volatility extracted from a time-series model of market returns | Common financial | *Journal of Finance* | Adrian and Rosenberg (2008) |
| | 69 | | Long-run market volatility | Low frequency volatility extracted from a time-series model of market returns | Common financial | | |
| 2008 | 70 | | Investment growth | Return on a zero-investment portfolio long in low investment growth firms and short in high investment growth firms | Common financial | *Review of Financial Studies* | Xing (2008) |
| 2008 | 71 | | Mean consumption growth | Across-state mean consumption growth rate | Common macro | *Review of Financial Studies* | Korniotis (2008) |
| | 72 | | Variance of consumption growth* | Across-state consumption growth variance | Common macro | | |
| | 73 | | Mean habit growth | Across-state mean habit growth rate | Common macro | | |
| | 74 | | Variance of habit growth | Across-state habit growth variance | Common macro | | |
| 2008 | 75 | | Liquidity | Systematic liquidity extracted from eight empirical liquidity measures | Common microstructure | *Journal of Financial Economics* | Korajczyk and Sadka (2008) |
| 2008 | | 97 | Country-level idiosyncratic volatility | Weighted average of variances and auto-covariances of firm-level idiosyncratic return shocks | Individual financial | *Review of Financial Studies* | Guo and Savickas (2008) |
| 2008 | | 98 | Distress | Distressed firm failure probability estimated based on a dynamic logit model | Individual financial | *Journal of Finance* | Campbell, Hilscher and Szilagyi (2008) |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2008 | | 99 | Shareholder advantage | Benefits from renegotiation upon default | Individual accounting | *Review of Financial Studies* | Garlappi, Shu and Yan (2008) |
| | | 100 | Interaction between shareholder advantage and implied market value of assets | Implied market value of assets provided by Moody's KMV | Individual accounting | | |
| 2008 | | 101 | Asset growth | Year-on-year percentage change in total assets | Individual accounting | *Journal of Finance* | Cooper, Gulen and Schill (2008) |
| 2008 | | 102 | Share issuance | Annual share issuance based on adjusted shares | Individual accounting | *Journal of Finance* | Pontiff and Woodgate (2008) |
| 2008 | | | Earnings announcement return‡ | Earnings announcement return capturing the market reaction to unexpected information contained in the firm's earnings release | Individual financial | *Working Paper* | Brandt, Kishore, Santa-Clara and Venkatachalam (2008) |
| 2008 | | 103 | Firm economic links | Economic links proxied by return of a portfolio of its major customers | Individual financial | *Journal of Finance* | Cohen and Frazzini (2008) |
| 2008 | | 104 | Sin stock | Stocks in the industry of adult services, alcohol, defense, gaming, medical and tobacco | Individual other | *Financial Analyst Journal* | Frank, Ma and Oliphant (2008) |
| 2008 | | 105 | Goodwill impairment | Buyers' overpriced shares at acquisition | Individual accounting | *Accounting Review* | Gu and Lev (2008) |
| 2008 | | 106 | Information in order backlog | Changes in order backlog on future profitability | Individual accounting | *Working Paper* | Gu, Wang and Ye (2008) |
| 2008 | | 107 | Investor recognition | Investor recognition proxied by the change in the breadth of institutional ownership | Individual other | *Review of Accounting Studies* | Lehavy and Sloan (2008) |
| 2008 | | 108 | DuPont analysis | Sales over net operating assets in DuPont analysis | Individual accounting | *Accounting Review* | Soliman (2008) |
| 2008 | | 109 | Small trades | Volume arising from small trades | Individual microstructure | *Review of Financial Studies* | Hvidkjaer (2008) |
| 2008 | 76 | | Idiosyncratic component of S&P 500 return | Residual of the linear projection of the S&P 500 return onto the CRSP value weighted index return | Common financial | *Working Paper* | Brennan and Li (2008) |
| 2009 | 77 | | Cash flow covariance with aggregate consumption | Cash flow covariance with aggregate consumption | Common macro | *Journal of Finance* | Da (2009) |
| | 78 | | Cash flow duration | Cash flow duration sensitivity to aggregate consumption | Common macro | | |
| 2009 | | | Financial constraints | THEORY | Common financial/macro | *Journal of Finance* | Livdan, Sapriza and Zhang (2009) |
| 2009 | 79 | | Long-run stockholder consumption growth | Aggregated microlevel stockholder consumption | Common macro | *Journal of Finance* | Malloy, Moskowitz and Vissing-Jorgensen (2009) |

... continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2009 | 80 | | Takeover likelihood | Estimated via a logit model of regressing ex-post acquisition indicator on various firm- and industry-level accounting variables | Common financial | *Review of Financial Studies* | Cremers, Nair and John (2009) |
| 2009 | 81 | | Illiquidity | Estimated using structural formula in line with Kyle's (1985) lambda | Common microstructure | *Review of Financial Studies* | Chordia, Huh and Subrahmanyam (2009) |
| 2009 | 82 | | Cash flow | Aggregate earnings based on revisions to analyst earnings forecasts | Common accounting | *Journal of Financial Economics* | Da and Warachka (2009) |
| 2009 | 83 | | Investors' beliefs* | Belief extracted from a two-state regime-switching model of aggregate market return and aggregate output | Common other | *Review of Financial Studies* | Ozoguz (2008) |
| | 84 | | Investors' uncertainty | Uncertainty extracted from a two-state regime-switching model of aggregate market return and aggregate output | Common other | | |
| 2009 | | 110 | Media coverage | Firm mass media coverage | Individual behavioral | *Journal of Finance* | Fang and Peress (2009) |
| 2009 | | 111 | Financial distress | Credit rating downgrades | Individual accounting | *Journal of Financial Economics* | Avramov, Chordia, Jostova and Philipov (2009) |
| 2009 | | 112 | Idiosyncratic volatility | Conditional expected idiosyncratic volatility estimated from a GARCH model | Individual accounting | *Journal of Financial Economics* | Fu (2009) |
| 2009 | | 113 | Debt capacity | Firm tangibility as in Almeida and Campello (2007) | Individual accounting | *Journal of Finance* | Hahn and Lee (2009) |
| 2009 | | 114 | Realized-implied volatility spread | Difference between past realized volatility and the average of call and put implied volatility | Individual financial | *Management Science* | Bali and Hovakimian (2009) |
| | | 115 | Call-put implied volatility spread | Difference between call and put implied volatility | Individual financial | | |
| 2009 | | 116 | Productivity of cash | Net present value of all the firm's present and future projects generated per dollar of cash holdings | Individual accounting | *Working Paper* | Chandrashekar and Rao (2009) |
| 2009 | | 117 | Advertising | Change in expenditures on advertising | Individual accounting | *Working Paper* | Chemmanur and Yan (2009) |
| 2009 | | 118 | Analyst forecasts optimism | Relative optimism and pessimism proxied by the difference between long-term and short-term analyst forecast of earnings growth | Individual financial | *Journal of Financial Markets* | Da and Warachka (2009) |
| 2009 | | 119 | Information revelation | Monthly estimate of the daily correlation between absolute returns and dollar volume | Individual microstructure | *Working Paper* | Gokcen (2009) |

... continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2009 | | 120 | Earnings volatility | Earnings volatility | Individual accounting | *Working Paper* | Gow and Taylor (2009) |
| 2009 | | 121 | Cash flow volatility | Rolling standard deviation of the standardized cashflow over the past sixteen quarters | Individual accounting | *Journal of Empirical Finance* | Huang (2009) |
| 2009 | | 122 | Local unemployment | Relative state unemployment | Individual other | *Working Paper* | Korniotis and Kumar (2009) |
| | | 123 | Local housing collateral | State-level housing collateral | Individual other | | |
| 2009 | | 124 | Efficiency score | Firm efficiency/inefficiency identified from the residual of the projection of firm market-to-book ratio onto various firm financial and accounting variables | Individual financial | *Journal of Financial and Quantitative Analysis* | Nguyen and Swanson (2009) |
| 2009 | | 125 | Order imbalance | Difference between buyer- and seller-initiated trades | Individual microstructure | *Review of Financial Studies* | Barber, Odean and Zhu (2009) |
| 2010 | 85 | | Market volatility and jumps | Estimated based on S&P index option returns | Common financial | *Working Paper* | Cremers, Halling and Weinbaum (2010) |
| 2010 | 86 | | Market mispricing | Zero-investment portfolio constructed from repurchasing and issuing firms | Common behavioral | *Review of Financial Studies* | Hirshleifer and Jiang (2010) |
| 2010 | | 126 | Idiosyncratic skewness | Skewness forecasted using firm level predictive variables | Individual financial | *Review of Financial Studies* | Boyer, Mitton and Vorkink (2010) |
| 2010 | | 127 | Political campaign contributions | Firm contributions to US political campaigns | Individual other | *Journal of Finance* | Cooper, Gulen and Ovtchinnikov (2010) |
| 2010 | | 128 | Real estate holdings | Real estate to total property, plant and equipment | Individual accounting | *Review of Financial Studies* | Tuzel (2010) |
| 2010 | | 129 | Realized skewness | Realized skewness obtained from high-frequency intraday prices | Individual financial | *Working Paper* | Amaya, Christoffersen, Jacobs and Vasquez (2011) |
| | | 130 | Realized kurtosis | Realized kurtosis obtained from high-frequency intraday prices | Individual financial | | |
| 2010 | | 131 | Excess multiple | Excess multiple calculated as the difference between the accounting multiple and the warranted multiple obtained by regressing the cross-section of firm multiples on accounting variables | Individual accounting | *Journal of Accounting, Auditing & Finance* | An, Bhojraj and Ng (2010) |
| 2010 | | 132 | Firm information quality | Firm information quality proxied by analyst forecasts, idiosyncratic volatility and standard errors of beta estimates | Individual financial/accounting | *Working Paper* | Armstrong, Banerjee and Corona (2010) |

... continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2010 | | 133 | Long-run idiosyncratic volatility | Long-run idiosyncratic volatility filtered from idiosyncratic volatility using HP filters | Individual financial | *Working Paper* | Cao and Xu (2010) |
| 2010 | 87 | | Private information | Return on a zero-investment portfolio long in high PIN stocks and short in low PIN stocks; PIN (private information) is the probability of information-based trade | Common microstructure | *Journal of Financial and Quantitative Analysis* | David, Hvidkjaer and O'Hara (2010) |
| 2010 | | 134 | Intra-industry return reversals | Intra-industry return reversals captured by the return difference between loser stocks and winners stocks based on relative monthly performance within the industry | Individual financial | *Working Paper* | Hameed, Huang and Mian (2010) |
| 2010 | | 135 | Related industry returns | Stock returns from economically related supplier and customer industries | Individual financial | *Journal of Finance* | Menzly and Ozbas (2010) |
| 2010 | | 136 | Earnings distributed to equity holders | Earnings distributed to equity holders | Individual accounting | *Review of Accounting & Finance* | Papanastasopoulos, Thomakos and Wang (2010) |
| | | 137 | Net cash distributed to equity holders | Dividends minus stock issues | Individual accounting | | |
| 2010 | | 138 | Excess cash | Most recently available ratio of cash to total assets | Individual accounting | *Financial Management* | Simutin (2010) |
| 2010 | | 139 | Extreme downside risk | Extreme downside risk proxied by the left tail index in the classical generalized extreme value distribution | Individual financial | *Journal of Banking and Finance* | Huang, Liu, Rhee and Wu (2010) |
| 2010 | | 140 | Volatility smirk | Steepness in individual option volatility smirk | Individual financial | *Journal of Financial and Quantitative Analysis* | Xing, Zhang and Zhao (2010) |
| 2010 | | | Exposure to financial distress costs | THEORY | Individual financial | *Journal of Financial Economics* | George and Hwang (2010) |
| 2011 | 88 | | Rare disasters | Disaster index based on international political crises | Common financial | *Journal of Financial Economics* | Berkman, Jacobsen and Lee (2011) |
| 2011 | | | Distress risk‡ | Aggregate distress risk obtained by projecting future business failure growth rates on a set of basis assets | Common financial | *Journal of Financial Economics* | Kapadia (2011)[k] |
| 2011 | | | Momentum† | Factor-mimicking portfolios based on momentum of international equity returns | Common other | *Review of Financial Studies* | Hou, Karolyi and Kho (2011) |
| | 89 | | Cash flow-to-price | Factor-mimicking portfolios based on cash flow-to-price of international equity returns | Common accounting | | |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2011 | | 141 | R&D investment | Firm's investment in research and development | Individual accounting | *Review of Financial Studies* | Li (2011) |
| | | | Financial constraints† | Kaplan and Zingales (1997) financial constraint index | Individual financial | | |
| 2011 | | 142 | Extreme stock returns | Portfolios sorted based on extreme past returns | Individual financial | *Journal of Financial Economics* | Bali, Cakici and Whitelaw (2011) |
| 2011 | | 143 | Jumps in individual stock returns | Average jump size proxied by slope of option implied volatility smile | Individual financial | *Journal of Financial Economics* | Yan (2011) |
| 2011 | | 144 | Intangibles | Employee satisfaction proxied by the list of "100 Best Companies to Work for in America" | Individual other | *Journal of Financial Economics* | Edmans (2011) |
| 2011 | | | Market return† | Equity index return | Common financial | *Working Paper* | Chen, Novy-Marx and Zhang (2011) |
| | 90 | | Investment portfolio return | Difference between returns of portfolios with low and high investment-to-asset ratio | Common financial | | |
| | 91 | | Return-on-equity portfolio return | Difference between returns of portfolios with high and low return on equity | Common financial | | |
| 2011 | | 145 | Volatility of liquidity | Measured by the price impact of trade as in Amihud (2002) | Individual microstructure | *Working Paper* | Akbas, Armstrong and Petkova (2011) |
| 2011 | | 146 | Dispersion in beliefs | Revealed through active holdings of fund managers | Individual behavioral | *Working Paper* | Jiang and Sun (2011) |
| 2011 | | 147 | Credit default swap spreads | Five-year spread less one-year spread | Individual financial | *Working Paper* | Han and Zhou (2011) |
| 2011 | | 148 | Organizational capital | Directly measured using Selling, General and Administrative expenditures | Individual accounting | *Working Paper* | Eisfeldt and Papanikolaou (2011) |
| 2011 | | 149 | Residual income | Firm residual income growth extracted from firm earnings growth | Individual accounting | *Review of Accounting Studies* | Balachandran and Mohanram (2011) |
| 2011 | | 150 | Accrual volatility | Firm accrual volatility measured by the standard deviation of the ratio of accruals to sales | Individual accounting | *Working Paper* | Bandyopadhyay, Huang and Wirjanto (2011) |
| 2011 | | 151 | Implied cost of capital | Implied cost of capital estimated using option contracts | Individual financial | *Working Paper* | Callen and Lyle (2011) |
| 2011 | | 152 | Non-accounting information quality | Average delay with which non-accounting information is impounded into stock price | Individual financial | *Contemporary Accounting Research* | Callen, Khan and Lu (2011) |
| | | 153 | Accounting information quality | Average delay with which accounting information is impounded into stock price | Individual financial | | |

... continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2011 | | 154 | Labor unions | Labor force unionization measured by the percentage of employed workers in a firm's primary Census industry Classification industry covered by unions in collective bargaining with employers | Individual other | *Journal of Financial and Quantitative Analysis* | Chen, Kacperczyk and Ortiz-Molina (2011) |
| 2011 | | 155 | Overreaction to nonfundamental price changes | Overreaction to within-industry discount rate shocks as captured by decomposing the short-term reversal into across-industry return momentum, within-industry variation in expected returns, under-reaction to within-industry cash flow news and overreaction to within-industry discount rate news | Individual other | *Working Paper* | Da, Liu and Schaumburg (2011) |
| 2011 | | 156 | Short interest | Short interest from short sellers | Individual financial | *Accounting Review* | Michael and Rees (2011) |
| 2011 | | 157 | Percent total accrual | Firm accruals scaled by earnings | Individual accounting | *Accounting Review* | Hafzalla, Lundholm and Van Winkle (2007) |
| 2011 | | | Projected earnings accuracy‡ | Skilled analysts identified by both past earnings forecasts accuracy and skills | Individual accounting | *Working Paper* | Hess, Kreutzmann and Pucker (2011) |
| 2011 | | 158 | Firm productivity | Firm level total factor productivity estimated from firm value added, employment and capital | Individual accounting | *Working Paper* | Imrohoroglu and Tuzel (2011) |
| 2011 | | 159 | Really dirty surplus | Really dirty surplus that happens when a firm issues or reacquires its own shares in a transaction that does not record the shares at fair market value | Individual accounting | *Accounting Review* | Landsman, Miller, Peasnell and Shu (2011) |
| 2011 | | 160 | Earnings forecast | Earnings forecast based on firm fundamentals | Individual accounting | *Review of Accounting Studies* | Li (2011) |
| 2011 | | 161 | Asset growth | Yearly percentage change in total balance sheet assets | Individual accounting | *Working Paper* | Nyberg and Poyry (2011) |
| 2011 | | 162 | Real asset liquidity | Number of potential buyers for a firm's assets from within the industry | Individual microstructure | *Working Paper* | Ortiz-Molina and Phillips (2011) |
| 2011 | | 163 | Customer-base concentration | Annual change in customer-base concentration | Individual other | *Working Paper* | Patatoukas (2011) |
| 2011 | | 164 | Tax expense surprises | Seasonally differenced quarterly tax expense | Individual accounting | *Journal of Accounting Research* | Thomas and Zhang (2011) |
| 2011 | | | Predicted earnings increase score‡ | Predicted earnings increase score based on financial statement information | Individual accounting | *Review of Accounting Studies* | Wahlen and Wieland (2011) |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2011 | | | Shareholder recovery | THEORY | Common financial | *Journal of Finance* | Garlappi and Yan (2011) |
| 2011 | 92 | | Garbage growth | Realized annual garbage growth | Common macro | Journal of Finance | Savov (2011) |
| 2012 | 93 | | Financial intermediary's wealth | Intermediary's marginal value of wealth proxied by shocks to leverage of securities broker-dealers | Common financial | *Journal of Finance* | Adrian, Etula and Muir (2012) |
| 2012 | 94 | | Stochastic volatility* | Estimated from a heteroscedastic VAR based on market and macro variables | Common financial | *Working Paper* | Campbell, Giglio, Polk and Turley (2012) |
| 2012 | 95 | | Average variance of equity returns | Decomposition of market variance into an average correlation component and an average variance component | Common financial | *Review of Financial Studies* | Chen and Petkova (2012) |
| 2012 | 96 | | Income growth for goods producing industries | Income growth for goods producing industries | Common macro | *Journal of Finance* | Eiling (2012) |
| | 97 | | Income growth for manufacturing industries | Income growth for manufacturing industries | Common macro | | |
| | 98 | | Income growth for distributive industries | Income growth for distributive industries | Common macro | | |
| | 99 | | Income growth for service industries* | Income growth for service industries | Common macro | | |
| | 100 | | Income growth for government* | Income growth for government | Common macro | | |
| 2012 | 101 | | Consumption volatility | Filtered consumption growth volatility from a Markov regime-switching model based on historical consumption data | Common macro | *Journal of Finance* | Boguth and Kuehn (2012) |
| 2012 | 102 | | Market skewness | Higher moments of market returns estimated from daily index options | Common financial | *Journal of Financial Economics* | Chang, Christoffersen and Jacobs (2012) |
| 2012 | 103 | | Learning* | Learning estimated from an investor's optimization problem under Knightian uncertainty | Common financial | *Working Paper* | Viale, Garcia-Feijoo and Giannetti (2011) |
| | 104 | | Knightian uncertainty | Knightian uncertainty estimated from an investor's optimization problem under Knightian uncertainty | Common financial | | |
| 2012 | 105 | | Market uncertainty | Proxied by variance risk premium | Common financial | *Working Paper* | Bali and Zhou (2012) |
| 2012 | | | Labor income‡ | Labor income at the census division level | Common macro | *Working Paper* | Gomez, Priestley and Zapatero (2012)[l] |
| 2012 | | 165 | Product price change | Cumulative product price changes since an industry enters the producer price index program | Individual financial | *Working Paper* | Van Binsbergen (2012) |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2012 | 106 | | Future growth in the opportunity cost of money | Opportunity cost of money as proxied by 3-month Treasury bill rate or effective Federal Funds rate | Common macro | *Working Paper* | Lioui and Maio (2012) |
| 2012 | | | Inter-cohort consumption differences | THEORY | Common macro | *Journal of Financial Economics* | Garleanu, Kogan and Panageas (2012) |
| 2012 | 107 | | Market-wide liquidity | Proxied by "noise" in Treasury prices | Common microstructure | *Working Paper* | Hu, Pan and Wang (2012) |
| 2012 | | 166 | Stock skewness | Ex ante stock risk-neutral skewness implied by option prices | Individual financial | *Journal of Finance* | Conrad, Dittmar and Ghysels (2012) |
| 2012 | | 167 | Expected return uncertainty | Proxied by the volatility of option-implied volatility | Individual financial | *Working Paper* | Baltussen, Van Bekkum and Van der Grient (2012) |
| 2012 | | 168 | Information intensity | Proxied by monthly frequency of current report filings | Individual microstructure | *Working Paper* | Zhao (2012) |
| 2012 | | 169 | Credit risk premia | Market implied credit risk premia based on the term structure of CDS spreads | Individual financial | *Working Paper* | Friewald, Wagner and Zechner (2012) |
| 2012 | | 170 | Geographic dispersion | Number of states in which a firm has business operations | Individual other | *Journal of Financial Economics* | Garcia and Norli (2012) |
| 2012 | | 171 | Political geography | Political proximity measured by political alignment index of each state's leading politicians with the ruling presidential party | Individual other | *Journal of Financial Economics* | Kim, Pantzalis and Park (2012) |
| 2012 | | 172 | Option to stock volume ratio | Option volume divided by stock volume | Individual microstructure | *Journal of Financial Economics* | Johnson and So (2012) |
| 2012 | | 173 | Cash holdings | Firm cash holdings | Individual accounting | *Journal of Financial Economics* | Palazzo (2012) |
| 2012 | | 174 | Labor mobility | Labor mobility based on average occupational dispersion of employees in an industry | Individual accounting | *Working Paper* | Donangelo (2012) |
| 2012 | | 175 | Debt covenant protection | Firm-level covenant index constructed based on 30 covenant categories | Individual accounting | *Working Paper* | Wang (2012) |
| 2012 | | 176 | Stock-cash flow sensitivity | Stock-cash flow sensitivity estimated from a structural one-factor contingent-claim model | Individual financial | *Working Paper* | Chen and Strebulaev (2012) |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2012 | 108 | | Jump beta | Discontinuous jump beta based on Todorov and Bollerslev (2010) | Common financial | *Working Paper* | Sophia Zhengzi Li (2012) |
| 2012 | | | Long-run consumption growth‡ | Long-run consumption growth rate identified from the risk-free rate and market price-dividend ratio based on Bansal and Yaron (2005)'s long-run risk model | Common macro | *Journal of Financial Economics* | Ferson, Nallareddy and Xie (2012)[m] |
| | | | Short-run consumption growth‡ | Short-run consumption growth rate identified from the risk-free rate and market price-dividend ratio based on Bansal and Yaron (2005)'s long-run risk model | Common macro | | |
| | | | Consumption growth volatility‡ | Consumption growth volatility shocks identified from the risk-free rate and market price-dividend ratio based on Bansal and Yaron (2005)'s long-run risk model | Common macro | | |
| 2012 | | 177 | Change in call implied volatility | Change in call implied volatility | Individual financial | *Working Paper* | Ang, Bali and Cakici (2012) |
| | | 178 | Change in put implied volatility | Change in put implied volatility | Individual financial | | |
| 2012 | | 179 | Firm hiring rate | Firm hiring rate measured by the change in the number of employees over the average number of employees | Individual other | *Working Paper* | Bazdresch, Belo and Lin (2012) |
| 2012 | | 180 | Information processing complexity | Past return for paired pseudo-conglomerates | Individual financial | *Journal of Financial Economics* | Cohen and Lou (2012) |
| 2012 | | 181 | Opportunistic buy | Prior month buy indicator for opportunistic traders who do not trade routinely | Individual microstructure | *Journal of Finance* | Cohen, Malloy and Pomorski (2012) |
| | | 182 | Opportunistic sell | Prior month sell indicator for opportunistic traders who do not trade routinely | Individual microstructure | | |
| 2012 | | 183 | Innovative efficiency | Patents/citations scaled by research and development expenditures | Individual other | *Journal of Financial Economics* | Hirshleifer, Hsu and Li (2012) |
| 2012 | | 184 | Abnormal operating cash flows | Abnormal operating cash flows | Individual accounting | *Working Paper* | Li (2012) |
| | | 185 | Abnormal production costs | Abnormal production costs | Individual accounting | | |
| 2012 | | 186 | Deferred revenues | Changes in the current deferred revenue liability | Individual accounting | *Contemporary Accounting Research* | Prakash and Sinha (2012) |
| 2012 | | 187 | Earnings conference calls | Sentiment of conference call wording | Individual other | *Journal of Banking and Finance* | Price, Doran, Peterson and Bliss (2012) |

... continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2012 | | 188 | Earnings forecast optimism | Difference between characteristic forecasts and analyst forecasts | Individual accounting | *Working Paper* | So (2012) |
| 2012 | 109 | | Commodity index | Open interest-weighted total index that aggregates 33 commodities | Common financial | *Working Paper* | Boons, Roon and Szymanowska (2012) |
| 2012 | | 189 | Time-series momentum | Time-series momentum strategy based on autocorrelations of scaled returns | Individual financial | *Journal of Financial Economics* | Moskowitz, Ooi and Pedersen (2012) |
| 2012 | | 190 | Carry | Expected return minus expected price appreciation | Individual financial | *Working Paper* | Koijen, Moskowitz, Pedersen and Vrugt (2012) |
| 2012 | | 191 | Expected return proxy | Logistic transformation of the fit ($R^2$) from a regression of returns on past prices | Individual financial | *Journal of Financial Economics* | Burlacu, Fontaine, Jimenez-Garces and Seasholes (2012) |
| 2012 | | 192 | Fraud probability | Probability of manipulation based on accounting variables | Individual accounting | *Financial Analysts Journal* | Beneish, Lee and Nichols (2013) |
| 2012 | | 193 | Buy orders | Sensitivity of price changes to sell orders | Individual microstructure | *Working Paper* | Brennan, Chordia, Subrahmanyam and Tong (2012) |
| | | 194 | Sell orders | Sensitivity of price changes to buy orders | Individual microstructure | | |
| 2013 | 110 | | Expected dividend level | Expected dividend level based on a macro time-series model | Common financial | *Working Paper* | Doskov, Pekkala and Ribeiro (2013) |
| | 111 | | Expected dividend growth | Expected dividend growth based on a macro time-series model | Common financial | | |
| 2013 | | 195 | Firm's ability to innovate | Rolling firm-by-firm regressions of firm-level sales growth on lagged R&D | Individual accounting | *Review of Financial Studies* | Cohen, Diether and Malloy (2013) |
| 2013 | | 196 | Board centrality | Board centrality measured by four basic dimensions of well-connectedness | Individual other | *Journal of Accounting and Economics* | Larcker, So and Wang (2013) |
| 2013 | | 197 | Gross profitability | Gross profits to assets | Individual accounting | *Journal of Financial Economics* | Novy-Marx (2013) |
| 2013 | | 198 | Betting-against-beta | Long leveraged low-beta assets and short high-beta assets | Individual financial | *Working Paper* | Frazzini and Pedersen (2013) |
| 2013 | | 199 | Secured debt | Proportion of secured to total debt | Individual accounting | *Working Paper* | Valta (2013) |
| | | 200 | Convertible debt | Proportion of convertible to total debt | Individual accounting | | |
| | | 201 | Convertible debt indicator | Dummy variable indicating whether a firm has convertible debt outstanding | Individual accounting | | |
| 2013 | 112 | | Cross-sectional pricing inefficiency | Pricing inefficiency proxied by returns to simulated trading strategies that capture momentum, profitability, value, earnings and reversal | Common microstructure | *Working Paper* | Akbas, Armstrong, Sorescu and Subrahmanyam (2013) |

. . . continued

| Year | # | # | Factor | Formation | Type | Journal | Short reference |
|---|---|---|---|---|---|---|---|
| 2013 | | 202 | Attenuated returns | Composite trading strategy returns where the weights are based on averaging percentile rank scores of various characteristics for each stock on portfolios | Individual financial | *Working Paper* | Chordia, Subrahmanyam and Tong (2013) |
| 2013 | | 203 | Bad private information | Decomposing the PIN measure of Easley, Hvidkjaer and O'Hara (2002) into two elements that reflect informed trading on good news and bad news | Individual microstructure | *Working Paper* | Brennan, Huh and Subrahmanyam (2013) |
| 2013 | 113 | | Trend signal | Return on a zero-investment portfolio long in past winners and short in past losers based on short-term, intermediate-term and long-term stock price trends | Common other | *Working Paper* | Han and Zhou (2013) |

Notes to Table

This table contains a summary of risk factors that explain the cross-section of expected returns. The column "Indi.(#)"("Common(#)") reports the cumulative number of empirical factors that are classified as individual (common) risk factors.
*: insignificant; †: duplicated; ‡: missing p-value.
a: No p-values reported for their factors constructed from principal component analysis.
b: Fama and French (1992) create zero-investment portfolios to test size and book-to-market effects. This is different from the testing approach in Banz (1981). We therefore count Fama and French (1992)'s test on size effect as a separate one.
c: No p-values reported for their high order equity index return factors.
d: No p-values reported for their eight risk factors that explain international equity returns.
e: No p-values reported for his high order return factors.
f: No p-values reported for their five hedge fund style return factors.
g: Vanden (2004) reports a t-statistic for each Fama-French 25 size and book-to-market sorted stock portfolios. We average these 25 t-statistics.
h: Acharya and Pedersen (2005) consider the illiquidity measure in Amihud (2002). This is different from the liquidity measure in Pastor and Stambaugh (2003). We therefore count their factor as a separate one.
i: No p-values reported for the interactions between market return and option returns.
j: No p-values reported for their co-moment betas.
k: No p-values reported for his distress tracking factor.
l: Gomez, Priestley and Zapatero (2012) study census division level labor income. However, most of the division level labor income have a non-significant t-statistic. We do not count their factors.
m: No p-values reported for their factors estimated from the long-run risk model.

# References


Abarbanell, J.S., and B.J. Bushee, 1998, Abnormal returns to a fundamental analysis strategy, *Accounting Review 73, 19-45.*

Acharya, Viral V. and Lasse Heje Pedersen, 2005, Asset pricing with liquidity risk, *Journal of Financial Economics 77, 375-410.*

Ackert, L., and G. Athanassakos, 1997, Prior uncertainty, analyst bias, and subsequent abnormal returns, *Journal of Financial Research 20, 263-273.*

Adler, Michael and Bernard Dumas, 1983, International portfolio choice and corporation finance: A synthesis, *Journal of Finance 38, 925-984.*

Adrian, Tobias, Erkko Etula and Tyler Muir, 2012, Financial intermediaries and the cross-section of asset returns, *Journal of Finance, Forthcoming.*

Adrian, Tobias and Joshua Rosenberg, 2008, Stock returns and volatility: Pricing the short-run and long-run components of market risk, *Journal of Finance 63, 2997-3030.*

Ahn, Seung C., Alex R. Horenstein and Na Wang, 2012, Determining rank of the beta matrix of a linear asset pricing model, *Working Paper, Arizona State University and Sogang University.*

Akbas, Ferhat, Will J. Armstrong and Ralitsa Petkova, 2011, The Volatility of liquidity and expected stock returns, *Working Paper, Purdue University.*

Akbas, Ferhat, Will Armstrong, Sorin Sorescu and Avanidhar Subrahmanyam, 2013, Time varying market efficiency in the cross-section of expected stock returns, *Working Paper, University of Kansas.*

Ali, Ashiq, Lee-Seok Hwang and Mark A. Trombley, 2003, Arbitrage risk and the book-to-market anomaly, *Journal of Financial Economics 69, 355-373.*

Almeida, Heitor and Murillo Campello, 2007, Financial constraints, asset tangibility, and corporate investment, *Review of Financial Studies 20, 1429-1460.*

Amaya, Diego, Peter Christoffersen, Kris Jacobs and Aurelio Vasquez, 2011, Do realized skewness and kurtosis predict the cross-section of equity returns, *Working Paper, University of Aarhus.*

Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets 5, 31-56.*

Amihud, Yakov and Haim Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics 17, 223-249.*

Amihud, Yakov and Haim Mendelson, 1989, The effects of beta, bid-ask spread, residual risk, and size on stock returns, *Journal of Finance 44, 479-486.*

An, Jiyoun, Sanjeev Bhojraj and David T. Ng, 2010, Warranted multiples and future returns, *Journal of Accounting, Auditing & Finance 25, 143-169.*

Anderson, Christopher W. and Luis Garcia-Feijóo, 2006, Empirical evidence on capital investment, growth options, and security returns, *Journal of Finance 61, 171-194.*

Anderson, Evan W., Eric Ghysels and Jennifer L. Juergens, 2005, Do heterogeneous beliefs matter for asset pricing, *Review of Financial Studies 18, 875-924.*

Ang, Andrew, Joseph Chen and Yuhang Xing, 2006, Downside risk, *Review of Financial Studies 19, 1191-1239.*

Ang, Andrew, Robert J. Hodrick, Yuhang Xing and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance 61, 259-299.*

Andrew Ang, Turan G. Bali and Nusret Cakici, 2012, The joint cross section of stocks and options, *Working Paper, Columbia University.*

Arbel, Avner, Steven Carvell and Paul Strebel, 1983, Giraffes, institutions and neglected firms, *Financial Analysts Journal 39, 57-63.*

Armstrong, Chris, Snehal Banerjee and Carlos Corona, 2010, Information quality and the cross-section of expected returns, *Working Paper, University of Pennsylvania.*

Asness, Clifford, R. Burt Porter and Ross Stevens, 2000, Predicting stock returns using industry-relative firm characteristics, *Working Paper, AQR Capital Management.*

Asquith, Paul, Parag A. Pathak and Jay R. Ritter, 2005, Short interest, institutional ownership and stock returns, *Journal of Financial Economics 78, 243-276.*

ATLAS Collaboration, 2012, Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, *Physics Letters B 716, 1-29.*

Avramov, Doron, Tarun Chordia, Gergana Jostova and Alexander Philipov, 2007, Momentum and credit rating, *Journal of Finance 62, 2503-2520.*

Avramov, Doron, Tarun Chordia, Gergana Jostova and Alexander Philipov, 2009, Dispersion in analysts' earnings forecasts and credit rating, *Journal of Financial Economics 91, 83-101.*

Baik, Bok and Tae Sik Ahn, 2007, Changes in order backlog and future returns, *Seoul Journal of Business 13, 105-126.*

Bajgrowicz, Pierre and Oliver Scaillet, 2012, Technical trading revisited: False discoveries, persistence tests, and transaction costs, *Journal of Financial Economics 106, 473-491.*

Bajgrowicz, Pierre, Oliver Scaillet and Adrien Treccani, 2013, Jumps in High-Frequency Data: Spurious Detections, Dynamics, and News, *Working Paper, University of Geneva.*

Baker, Malcolm and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *Journal of Finance 61, 1645-1680.*

Balachandran, Sudhakar and Partha Mohanram, 2011, Using residual income to refine the relationship between earnings growth and stock returns, *Review of Accounting Studies 17, 134-165.*

Balduzzi, P. and C. Robotti, 2008, Mimicking portfolios, economic risk premia, and tests of multi-beta models, *Journal of Business and Economic Statistics 26, 354-368.*

Bali, Turan G. and Armen Hovakimian, 2009, Volatility spreads and expected stock returns, *Management Science 55, 1797-1812.*

Bali, Turan G. and Hao Zhou, 2012, Risk, uncertainty, and expected returns, *Working Paper, Georgetown University.*

Bali, Turan G., Nusret Cakici and Robert F. Whitelaw, 2011, Maxing out: Stocks as lotteries and the cross-section of expected returns, *Journal of Financial Economics 99, 427-446.*

Baltussen, Guido, Sjoerd Van Bekkum and Bart Van der Grient, 2012, Unknown unknowns: Vol-of-vol and the cross section of stock returns, *Working Paper, Erasmus University.*

Balvers, Ronald J. and Dayong Huang, 2007, Productivity-based asset pricing: Theory and evidence, *Journal of Financial Economics 86, 405-445.*

Bandyopadhyay, Sati, Alan Huang and Tony Wirjanto, 2010, The accrual volatility anomaly, *Working Paper, University of Waterloo.*

Bansal, Ravi and Amir Yaron, 2005, Risks for the long run: a potential resolution of asset pricing puzzles, *Journal of Finance 59, 1481-1509.*

Bansal, Ravi, Robert F. Dittmar and Christian T. Lundblad, 2005, Consumption, dividends, and the cross section of equity returns, *Journal of Finance 60, 1639-1672.*

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

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

Barber, Brad, Reuven Lehavy, Maureen McNichols and Brett Trueman, 2001, Can investors profit from the prophets? Security analyst recommendations and stock returns, *Journal of Finance 56, 531-563.*

Barber, B., T. Odean and N. Zhu, 2009, Do retail trades move markets? *Review of Financial Studies 22, 152-186.*

Barras, Laurent, Oliver Scaillet and Russ Wermers, 2010, False discoveries in mutual fund performance: Measuring luck in estimated alphas, *Journal of Finance 65, 179-216.*

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.*

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.*

Bauman, Scott and Richard Dowen, 1988, Growth projections and common stock returns, *Financial Analyst Journal 44, 79-80.*

Bazdresch, Santiago, Frederico Belo and Xiaoji Lin, 2012, Labor hiring, investment, and stock return predictability in the cross section, *Working Paper, University of Minnesota.*

Begg, C.B. and J.A., Berlin, 1988, Publication bias: A problem in interpreting medical data, *Journal of the Royal Statistical Society, Series A, 419-463.*

Beneish, M.D., 1997, Detecting GAAP violation: Implications for assessing earnings management among firms with extreme financial performance, *Journal of Accounting and Public Policy 16, 271-309.*

Beneish, M.D., M.C. Lee and D. Craig Nichols, 2012, Fraud detection and expected returns, *Available at SSRN, 2012.*

Benjamini, Yoav and Daniel Yekutieli, 2001, The control of the false discovery rate in multiple testing under dependency, *Annals of Statistics 29, 1165-1188.*

Benjamini, Yoav and Wei Liu, 1999, A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence, *Journal of Statistical Planning and Inference 82, 163-170.*

Benjamini, Yoav and Yosef Hochberg, 1995, Controlling the false discovery rate: A practical and powerful approach to multiple testing, *Journal of the Royal Statistical Socitey, Series B, 289-300.*

Berardino, Palazzo, 2012, Cash holdings, risk, and expected returns, *Journal of Financial Economics 104, 162-185.*

Berkman, Henk, Ben Jacobsen and John B. Lee, 2011, Time-varying rare disaster risk and stock returns, *Journal of Financial Economics 101, 313-332.*

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

Black, Fischer, 1972, Capital market equilibrium with restricted borrowing, *Journal of Business 45, 444-454.*

Black, Fischer, Michael C. Jensen and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests. In *Studies in the theory of capital markets,* ed. Michael Jensen, pp. 79-121. New York: Praeger.

Bondt, Werner F.M. and Richard Thaler, 1985, Does the stock market overreact?,*Journal of Finance 40, 793-805.*

Brammer, Stephen, Chris Brooks and Stephen Pavelin, 2006, Corporate social performance and stock returns: UK evidence from disaggregate measures, *Financial Management 35, 97-116.*

Brandt, Michael, Runeet Kishore, Pedro Santa-Clara and Mohan Venkatachalam, 2008, Earnings announcements are full of surprises, *Working Paper, Duke University.*

Bradshaw, Mark, Scott Richardson and Richard Sloan, 2006, The relation between corporate financing activities, analysts' forecasts and stock returns, *Journal of Accounting and Economics 42, 53-85.*

Breeden, Douglas T., 1979, An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics 7, 265-296.*

Breeden, Douglas T., Michael R. Gibbons and Robert H. Litzenberger, 1989, Empirical Test of the Consumption-Oriented CAPM, *Journal of Finance 44, 231-262.*

Brennan, Michael J., Ashley W. Wang and Yihong Xia, 2004, Estimation and test of a simple model of intertemporal capital asset pricing, *Journal of Finance 59, 1743-1776.*

Brennan, Michael J. and Avanidhar Subrahmanyam, 1996, Market microstructure and asset pricing: On the compensation for illiquidity in stock returns, *Journal of Financial Economics 41, 441-464.*

Brennan, Michael and Feifei Li, 2008, Agency and asset pricing, *Working Paper, UCLA.*

Brennan, Michael, Sahn-Wook Huh and Avanidhar Subrahmanyam, 2013, The pricing of good and bad private information in the cross-section of expected stock returns, *Working Paper, University of California at Los Angeles.*

Brennan, Michael J., Tarun Chordia and Avanidhar Subrahmanyam, 1998, Alternative factor specifications, security characteristics, and the cross-section of expected stock returns, *Journal of Financial Economics 49, 345-373.*

Brennan, Michael J., Tarun Chordia, Avanidhar Subrahmanyam and Qing Tong, 2012, Sell-order liquidity and the cross-section of expected stock returns, *Journal of Financial Economics 105, 523-541.*

Brown, David and Bradford Rowe, 2007, The productivity premium in equity returns, *Working Paper, University of Wisconsin, Madison.*

Brown, D. Andrew, Nicole A. Lazar, Gauri S. Datta, Woncheol Jang, Jennifer E. McDowell, 2012, Incorporating spatial dependence into Bayesian multiple testing of statistical parametric maps in functional neuroimaging, *JSM.*

Boguth, Oliver and Lars-Alexander Kuehn, 2012, Consumption volatility risk, *Journal of Finance, Forthcoming.*

Boons, Martijn, Frans De Roon and Marta Szymanowska, 2012, The stock market price of commodity risk, *Working Paper, Tilburg University.*

Boudoukh, Jacob, Roni Michaely, Matthew Richardson and Michael R. Roberts, 2007, On the importance of measuring payout yield: implications for empirical asset pricing, *Journal of Finance 62, 877-915.*

Bossaerts, Peters and Robert M. Dammon, 1994, Tax-induced intertemporal restrictions on security returns, *Journal of Finance 49, 1347-1371.*

Botosan, Christine A., 1997, Disclosure level and the cost of equity capital, *Accounting Review 72, 323-349.*

Boudoukh, Jacob, Roni Michaely, Matthew Richardson and Michael R. Roberts, 2007, On the importance of measuring payout yield: implications for empirical asset pricing, *Journal of Finance 62, 877-915.*

Boyer, Brian, Todd Mitton and Keith Vorkink, 2010, Expected idiosyncratic skewness, *Review of Financial Studies 23, 170-202.*

Burlacu, Radu, Patrice Fontaine, Sonia Jimenez-Garces and Mark S. Seasholes, 2012, Risk and the cross section of stock returns, *Journal of Financial Economics 105, 511-522.*

Callen, Jeffrey and Matthew Lyle, 2011, The term structure of implied costs of equity capital, *Working Paper, University of Toronto.*

Callen, Jeffrey, Mozaffar Khan and Hai Lu, 2011, Accounting quality, stock price delay, and future stock returns, *Contemporary Accounting Research 30, 269-295.*

Campbell, John Y., 1996, Understanding risk and return, *Journal of Political Economy 104, 298-345.*

Campbell, John Y., Jens Hilscher and Jan Szilagyi, 2008, In search of distress risk, *Journal of Finance 63, 2899-2939.*

Campbell, John Y., Stefano Giglio, Christopher Polk and Robert Turley, 2012, An Intertemporal CAPM with Stochastic Volatility, *Working Paper, Harvard University.*

Campbell, John Y. and Tuomo Vuolteenaho, 2004, Bad beta, good beta, *American Economic Review 94, 1249-1275.*

Cao, Xuying and Yexiao Xu, 2010, Long-run idiosyncratic volatilities and cross-sectional stock returns, *Working Paper, University of Illinois at Urbana-Champaign.*

Cao, Charles, Yong Chen, Bing Liang and Andrew W. Lo, 2013, Can hedge funds time market liquidity?, *Journal of Financial Economics 109, 493-516.*

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance 52, 57-82.*

Cen, Ling, John Wei and Jie Zhang, 2006, Forecasted earnings per share and the cross section of expected stock returns, *Working Paper, Hong Kong University of Science & Technology.*

Chan, K. C., Nai-fu Chen and David A. Hsieh, 1985, An exploratory investigation of the firm size effect, *Journal of Financial Economics 14, 451-471.*

Chan, K.C., Silverio Foresi and Larry H.P. Lang, 1996, Does money explain returns? Theory and empirical analysis, *Journal of Finance 51, 345-361.*

Chandrashekar, Satyajit and Ramesh K.S. Rao, 2009, The productivity of corporate cash holdings and the cross-section of expected stock returns, *Working Paper, University of Texas at Austin.*

Chang, Bo Young, Peter Christoffersen and Kris Jacobs, 2012, Market skewness risk and the cross section of stock returns, *Journal of Financial Economics, Forthcoming.*

Chapman, David A., 1997, Approximating the asset pricing kernel, *Journal of Finance 52, 1383-1410.*

Chemmanur, Thomas and An Yan, 2009, Advertising, attention, and stock returns, *Working Paper, Boston College.*

Chen, Joseph, Harrison Hong and Jeremy C. Stein, 2002, Breadth of ownership and stock returns, *Journal of Financial Economics 66, 171-205.*

Chen, Huafeng, Marcin Kacperczyk and Hernan Ortiz-Molina, 2011, Labor unions, operating flexibility, and the cost of equity, *Journal of Financial and Quantitative Analysis 46, 25-58.*

Chen, Long, Robert Novy-Marx and Lu Zhang, 2011, An alternative three-factor model, *Working Paper.*

Chen, Nai-Fu, Richard Roll and Stephen A. Ross, 1986, Economic forces and stock market, *Journal of Business 59, 383-403.*

Chen, Zhanhui and Ralitsa Petkova, 2012, Does idiosyncratic volatility proxy for risk exposure?, *Review of Financial Studies 25, 2745-2787.*

Chen, Zhiyao and Ilya Strebulaev, 2012, Contingent-claim-based expected stock returns, *Working Paper, University of Reading.*

Chopra, Navin, Josef Lakonishok and Jay R. Ritter, 1992, Measuring abnormal performance: do stocks overreact?, *Journal of Financial Economics 31, 235-268.*

Chordia, Tarun, Avanidhar Subrahmanyam and V. Ravi Anshuman, 2001, Trading activity and expected stock returns, *Journal of Financial Economics 59, 3-32.*

Chordia, Tarun, Avanidhar Subrahmanyam and Qing Tong, 2013, Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *Working Paper, http://dx.doi.org/10.1016/j.jacceco.2014.06.001.*

Chordia, Tarun and Lakshmanan Shivakumar, 2006, Earnings and price momentum, *Journal of Financial Economics 80, 627-656.*

Chordia, Taurn, Sahn-Wook Huh and Avanidhar Subrahmanyam, 2009, Theory-based illiquidity and asset pricing, *Review of Financial Studies 22, 3629-3668.*

Chung, Y. Peter, Herb Johnson and Michael J. Schill, 2006, Asset pricing when returns are non-normal: Fama-French factors versus higher-order systematic comoments, *Journal of Business 79, 923-940.*

CMS Collaboration, 2012, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, *Physics Letters B 716, 30-61.*

Cochrane, John, 1991, Production-based asset pricing and the link between stock returns and economic fluctuations, *Journal of Finance 46, 209-237.*

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

Cochrane, John H., 2011, Presidential Address: Discount Rates, *Journal of Finance 66, 1047-1108.*

Cohen, Lauren and Andrea Frazzini, 2008, Economic links and predictable returns, *Journal of Finance 63, 1977-2011.*

Cohen, Lauren, Christopher Malloy and Lukasz Pomorski, 2012, Decoding inside information, *Journal of Finance 67, 1009-1043.*

Cohen, Lauren and Dong Lou, 2012, Complicated firms, *Journal of Financial Economics 104, 383-400.*

Cohen, Lauren, Karl Diether and Christopher Malloy, 2013, Misvaluing innovation, *Review of Financial Studies 26, 635-666.*

Cohen, Randy, Christopher Polk and Bernhard Silli, 2009, Best ideas, *Working Paper, Harvard Business School.*

Conrad, Jennifer, Michael Cooper and Gautam Kaul, 2003, Value versus glamour, *Journal of Finance 58, 1969-1996.*

Conrad, Jennifer, Robert F. Dittmar and Eric Ghysels, 2013, Ex ante skewness and expected stock returns, *Journal of Finance 68, 85-124.*

Constantinides, G., 1982, Intertemporal asset pricing with heterogeneous consumers and without demand aggregation, *Journal of Business 55, 253-267.*

Constantinides, G., 1986, Capital market equilibrium with transaction costs, *Journal of Political Economy 94, 842-862.*

Cooper, Michael J. and Huseyin Gulen, 2006, Is time-series-based predictability evident in real time? *Journal of Business 79, 1263-1292.*

Cooper, Michael J., Huseyin Gulen and Alexei V. Ovtchinnikov, 2010, Corporate political contributions and stock returns, *Journal of Finance 65, 687-724.*

Cooper, Michael J., Huseying Gulen and Michael J. Schill, 2008, Asset growth and the cross-section of stock returns, *Journal of Finance 63, 1609-1651.*

Cox, D.R., 1982, Statistical significance tests, *British Journal of Clinical Pharmacology 14, 325-331.*

Cox, John C., Jonathan E. Ingersoll, Jr. and Stephen A. Ross, An intertemporal general equilibrium model of asset pricing, *Econometrica 53, 363-384.*

Cremers, Martijn, Michael Halling and David Weinbaum, 2010, In search of aggregate jump and volatility risk in the cross-section of stock returns, *Working Paper, Yale University.*

Cremers, K.J. Martijn and Vinay B. Nair, 2005, Governance Mechanisms and equity prices, *Journal of Finance 60, 2859-2894.*

Cremers, K. J. Martijn, Vinay B. Nair and Kose John, 2009, Takeovers and the cross-section of returns, *Review of Financial Studies 22, 1409-1445.*

Da, Zhi, 2009, Cash flow, consumption risk, and the cross-section of stock returns, *Journal of Finance 64, 923-956.*

Da, Zhi and Ernst Schaumburg, 2011, Relative valuation and analyst target price forecasts, *Journal of Financial Markets 14, 161-192.*

Da, Zhi and Mitchell Craig Warachka, 2009, Cash flow risk, systematic earnings revisions, and the cross-section of stock returns, *Journal of Financial Economics 94, 448-468.*

Da, Zhi and Mitchell Craig Warachka, 2009, Long-term earnings growth forecasts, limited attention, and return predictability, *Working Paper, University of Notre Dame.*

Da, Zhi, Qianqiu Liu and Ernst Schaumburg, 2011, Decomposing short-term return reversal, *Working Paper, University of Notre Dame.*

Daniel, Kent and Sheridan Titman, 1997, Evidence on the characteristics of cross sectional variation in stock returns, *Journal of Finance 52, 1-33.*

Daniel, Kent and Sheridan Titman, 2006, Market reactions to tangible and intangible information, *Journal of Finance 61, 1605-1643.*

Daniel, Kent and Sheridan Titman, 2012, Testing factor-model explanations of market anomalies, *Critical Finance Review 1, 103-139.*

Datar, Vinay, Narayan Y Naik and Robert Radcliffe, 1998, Liquidity and stock returns: An alternative test, *Journal of Financial Markets 1, 203-219.*

Dichev, Ilia, 1998, Is the risk of bankruptcy a systematic risk? *Journal of Finance 53, 1131-1147.*

Dichev, Ilia and Joseph Piotroski, 2001, The long-run stock returns following bond ratings changes, *Journal of Finance 56, 173-203.*

Diether, Karl B., Christopher J. Malloy and Anna Scherbina, 2002, Differences of opinion and the cross section of stock returns, *Journal of Finance 57, 2113-2141.*

Dittmar, Robert F., 2002, Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns, *Journal of Finance 57, 369-403.*

Donangelo, Andres, 2012, Labor Mobility: Implications for Asset Pricing, *Working Paper, University of Texas at Austin.*

Doskov, Nikolay, Tapio Pekkala and Ruy M. Ribeiro, 2013, Tradable Macro Risk Factors and the Cross-Section of Stock Returns, *Working Paper, Norges Bank Investment Management.*

Douglas, G.W., 1967, Risk in the equity markets: An empirical appraisal of market efficiency, *Yale Economic Essays 9, 3-48.*

Doran, James, Andy Fodor and David Peterson, 2007, Insiders versus outsiders with employee stock options: Who knows best about future firm risk and implications for stock returns, *Working Paper, Florida State University.*

Doyle, Jeffrey, Russell Lundholm and Mark Soliman, 2003, The predictive value of expenses excluded from pro forma earnings, *Review of Accounting Studies 8, 145-174.*

Drake, Michael and Lynn Rees, 2011, Should investors follow the prophets or the bears? Evidence on the use of public information by analysts and short sellers, *Accounting Review 86, 101-130.*

Dudoit, S. and Van der Laan, M., 2008, Multiple testing procedures with applications to Genomics, *Springer Series in Statistics*, New York, USA.

Easley, David, Soeren Hvidkjaer and Maureen O'Hara, 2002, Is information risk a determinant of asset returns, *Journal of Finance 57, 2185-2221.*

Easley, David, Soeren Hvidkjaer and Maureen O'Hara, 2010, Factoring information into returns, *Journal of Financial and Quantitative Analysis 45, 293-309.*

Eberhart, Allan, William Maxwell and Akhtar Siddique, 2004, An examination of long-term abnormal stock returns and operating performance following R&D increases, *Journal of Finance 59, 623-650.*

Edmans, Alex, 2011, Does the stock market fully value intangibles? Employee satisfaction and equity prices, *Journal of Financial Economics 101, 621-640.*

Efron, Bradley, 1979, Bootstrap methods: another look at the jackknife, *Annuals of Statistics 7, 1-26.*

Efron, Bradley and Robert Tibshirani, 2002, Empirical Bayes methods and false discovery rates for microarrays, *Genetic Epidemiology 23, 70-86.*

Efron, Bradley, Robert Tibshirani, John Storey and Virginia Tusher, 2001, Empirical Bayes analysis of a microarray experiment, *Journal of the American Statistical Association 96, 1151-1160.*

Efron, Bradley, 2004, Large-scale simultaneous hypothesis testing: the choice of a null hypothesis, *Journal of the American Statistical Association 99, 96-104.*

Efron, Bradley, 2006, Microarrays, empirical Bayes, and the two-groups model, *Statistical Science 23, 2008.*

Eiling, Esther, 2012, Industry-specific human capital, idiosyncratic risk, and the cross-section of expected stock returns, *Journal of Finance 68, 43-84.*

Eisfeldt, Andrea L. and Dimitris Papanikolaou, 2011, Organization capital and the cross-section of expected returns, *Working Paper, UCLA.*

Elgers, Pieter T., May H. Lo and Ray J. Pfeiffer Jr., 2001, Delayed security price adjustments to financial analysts' forecasts of annual earnings, *Accounting Review 76, 613-632.*

Elton, Edwin J., Martin J. Gruber and Christopher R. Blake, 1995, Fundamental economic variables, expected returns, and bond fund performance, *Journal of Finance 50, 1229-1256.*

Elton, Edwin J., Martin J. Gruber, Sanjiv Das and Matthew Hlavka, 1993, Efficiency with costly information: A reinterpretation of evidence from managed portfolios, *Review of Financial Studies 6:1, 1-22.*

Erb, Claude, Campbell Harvey and Tadas Viskanta, 1996, Expected returns and volatility in 135 countries, *Journal of Portfolio Management 22, 46-58.*

Fabozzi, Frank J., K.C. Ma and Becky J. Oliphant, 2008, Sin stock returns, *Financial Analysts Journal Fall, 82-94.*

Fairfield, Patricia M., J. Scott Whisenant and Teri Lombardi Yohn, 2003, Accrued earnings and growth: implications for future profitability and market mispricing, *Accounting Review 78, 353-371.*

Fama, Eugene F., 1991, Efficient capital markets: II, *Journal of Finance 46, 1575-1617.*

Fama, Eugene F. and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy 81, 607-636.*

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 on stocks and bonds, *Journal of Financial Economics 33, 3-56.*

Fama, Eugene F. and Kenneth R. French, 2006, Profitability, investment and average returns, *Journal of Financial Economics 82, 491-518.*

Fama, E., K. French, 2010, Luck versus skill in the cross section of mutual fund returns, *Journal of Finance 65, 1915-1947.*

Fang, Lily and Joel Peress, 2009, Media coverage and the cross-section of stock returns, *Journal of Finance 64, 2023-2052.*

Farcomeni, Alessio, 2007, A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion, *Statistical Methods in Medical Research 17, 347-388.*

Ferson, Wayne E. and Campbell R. Harvey, 1991, The variation of economic risk premiums, *Journal of Political Economy 99, 385-415.*

Ferson, Wayne E. and Campbell R. Harvey, 1993, The risk and predictability of international equity returns, *Review of Financial Studies 6, 527-566.*

Ferson, Wayne E. and Campbell R. Harvey, 1994, Sources of risk and expected returns in global equity markets, *Journal of Banking and Finance 18, 775-803.*

Ferson, Wayne E. and Campbell R. Harvey, 1999, Conditioning variables and the cross section of stock returns, *Journal of Finance 54, 1325-1360.*

Ferson, Wayne and Yong Chen, 2013, How many good and bad fund managers are there, really? *Working Paper, University of Southern California.*

Ferson, Wayne E., Suresh Nallareddy and Biqin Xie, 2012, The "out-of-sample" performance of long run risk models, *Journal of Financial Economics, Forthcoming.*

Figlewski, Stephen, 1981, The informational effects of restrictions on short sales: some empirical evidence, *Journal of Financial and Quantitative Analysis 16, 463-476.*

Fogler, H. Russell, Rose John and James Tipton, 1981, Three factors, interest rate differentials and stock groups, *Journal of Finance 36, 323-335.*

Frank, Murray Z. and Vidhan K. Goyal, 2009, Capital structure decisions: Which factors are reliably important? *Financial Management 38, 1-37.*

Frankel, Richard and Charles Lee, 1998, Accounting valuation, market expectation, and cross-sectional stock returns, *Journal of Accounting and Economics 25, 283-319.*

Franzoni, Francesco and Jose M. Marin, 2006, Pension plan funding and stock market efficiency, *Journal of Finance 61, 921-956.*

Frazzini, Andrea and Lasse Heje Pedersen, 2013, Betting against beta, *Working Paper, AQR Capital Management.*

Friewald, Nils, Christian Wagner and Josef Zechner, 2012, The cross-section of credit risk premia and equity returns, *Working Paper, Vienna University.*

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

Fu, Fangjian, 2009, Idiosyncratic risk and the cross-section of expected stock returns, *Journal of Financial Economics 91, 24-37.*

Fung, William and David A. Hsieh, 1997, Empirical characteristics of dynamic trading strategies: The case of hedge funds, *Review of Financial Studies 10, 275-302.*

Fung, William and David A. Hsieh, 2001, The risk in hedge fund strategies: theory and evidence from trend followers, *Review of Financial Studies 14, 313-341.*

Garcia, Diego and Oyvind Norli, 2012, Geographic dispersion and stock returns, *Journal of Financial Economics, Forthcoming.*

Garlappi, Lorenzo and Hong Yan, 2011, Financial distress and the cross-section of equity returns, *Journal of Finance 66, 789-822.*

Garlappi, Lorenzo, Tao Shu and Hong Yan, 2008, Default risk, shareholder advantage, and stock returns, *Review of Financial Studies 21, 2743-2778.*

Gârleanu, Nicolae, Leonid Kogan and Stavros Panageas, 2012, Displacement risk and asset returns, *Journal of Financial Economics 105, 491-510.*

George, Thomas J. and Chuan-yang Hwang, 2004, The 52-week high and momentum investing, *Journal of Finance 59, 2145-2176.*

George, Thomas J. and Chuan-yang Hwang, 2010, A resolution of the distress risk and leverage puzzles in the cross section of stock returns, *Journal of Financial Economics 96, 56-79.*

Gettleman, Eric and Joseph M. Marks, 2006, Acceleration strategies, *Working Paper, Seton Hall Univeristy.*

Glaeser, Edward, 2008, Research incentives and empirical methods, *Chapter 13, The Foundations of Positive and Normative Economics: A Handbook, Oxford University Press.*

Gokcen, Umut, 2009, Information revelation and expected stock returns, *Working Paper, Boston College.*

Gomes, Joao F., Amir Yaron and Lu Zhang, 2006, Asset pricing implications of firms' financing constraints, *Review of Financial Studies 19, 1321-1356.*

Gómez, Juan-Pedro, Richard Priestley and Fernando Zapatero, 2012, Labor income, relative wealth concerns, and the cross-section of stock returns, *Working Paper, Instituto de Empresa Business School.*

Gompers, Paul A. and Andrew Metrick, 2001, Institutional investors and equity prices, *Quarterly Journal of Economics, 116, 229-259.*

Gompers, Paul A., Joy L. Ishii and Andrew Metrick, 2003, Corporate governance and equity prices, *Quarterly Journal of Economics 118, 107-155.*

Gourio, Francois, 2007, Labor leverage, firms' heterogeneous sensitivities to the business cycle, and the cross-section of expected returns, *Working Paper, Boston University.*

Gow, Ian and Daniel Taylor, 2009, Earnings volatility and the cross-section of returns, *Working Paper, Northwestern University.*

Green, Jeremiah, John RM Hand and X. Frank Zhang, 2012, The supraview of return predictive signals, *Review of Accounting Studies, Forthcoming.*

Green, Jeremiah, John RM Hand and X. Frank Zhang, 2013, The remarkable multidimensionality in the cross section of expected US stock returns, *Working Paper, Pennsylvania State University.*

Greene, William H., 2008, Econometric analysis, Prentice Hall.

Griffin, John M. and Michael L. Lemmon, 2002, Book-to-market equity, distress risk, and stock returns, *Journal of Finance 57, 2317-2336.*

Gu, Feng, 2005, Innovation, future earnings, and market efficiency, *Journal of Accounting, Auditing and Finance 20, 385-418.*

Gu, Feng and Baruch Lev, 2008, Overpriced shares, ill-advised acquisitions, and goodwill impairment, *Accounting Review 86, 1995-2022.*

Gu, Li, Zhiqiang Wang and Jianming Ye, 2008, Information in order backlog: change versus level, *Working Paper, Fordham University.*

Guo, Hui and Robert Savickas, 2008, Average idiosyncratic volatility in G7 countries, *Review of Financial Studies 21, 1259-1296.*

Hafzalla, Nader, Russell Lundholm and E. Matthew Van Winkle, 2011, Percent Accruals, *Accounting Review 86, 209-236.*

Hahn, Jaehoon and Hangyong Lee, 2009, Financial constraints, debt capacity, and the cross-section of stock returns, *Journal of Finance 64, 891-921.*

Hameed, Allaudeen, Joshua Huang and Mujtaba Mian, 2010, Industries and stock return reversals, *Working Paper, National University of Singapore.*

Han, Bing and Yi Zhou, 2011, Term structure of credit default swap spreads and cross-section of stock returns, *Working Paper, University of Texas at Austin.*

Han, Yufeng and Guofu Zhou, 2013, Trend factor: A new determinant of cross-section stock returns, *Working Paper, University of Colorado Denver.*

Harvey, Campbell R. and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *Journal of Finance 55, 1263-1295.*

Harvey, Campbell R. and Yan Liu, 2014a, Backtesting, *Working Paper, Duke University.*

Harvey, Campbell R. and Yan Liu, 2014b, Multiple testing in economics, *Working Paper, Duke University.*

Harvey, Campbell R. and Yan Liu, 2014c, Evaluating trading strategies, *Working Paper, Duke University.*

Harvey, Campbell R. and Yan Liu, 2014d, Incremental factors, *Working Paper, Duke University.*

Hawkins, Eugene H., Stanley C. Chamberlin and Wayne E. Daniel, 1984, Earnings expectations and security prices, *Financial Analysts Journal, 24-74.*

Head, Alex, Gary Smith and Julia Wilson, 2007, Would a stock by any other ticker smell as sweet? *Quarterly Review of Economics & Finance 49, 551-561.*

Heaton, John and Deborah Lucas, 2000, Portfolio choice and asset prices: The importance of entrepreneurial risk, *Journal of Finance 55, 1163-1198.*

Heckerman, Donald G., 1972, Portfolio selection and the structure of capital asset prices when relative prices of consumption goods may change, *Journal of Finance 27, 47-60.*

Heckman, James J., 1979, Sample selection bias as a specification error, *Econometrica 47, 153-161.*

Hess, Dieter, Daniel Kreutzmann and Oliver Pucker, Projected earnings accuracy and profitability of stock recommendations, *Working Paper, University of Cologne.*

Hirshleifer, David, Kewei Kou, Siew Hong Teoh and Yinglei Zhang, 2004, Do investors overvalue firms with bloated balance sheets?, *Journal of Accounting and Economics 38, 297-331.*

Hirshleifer, David and Danling Jiang, 2010, A financing-based misvaluation factor and the cross-section of expected returns, *Review of Financial Studies 23, 3401-3436.*

Hirshleifer, David, Po-Hsuan Hsu and Dongmei Li, 2012, Innovative efficiency and stock returns, *Journal of Financial Economics 107, 632-654.*

Hochberg, Yosef, 1988, A sharper Bonferroni procedure for multiple tests of significance, *Biometrika 75, 800-802.*

Hochberg, Yosef and Benjamini Y., 1990, More powerful procedures for multiple significance testing, *Statistics in Medicine 9, 811-818.*

Hochberg, Yosef and Tamhane, Ajit, 1987, Multiple comparison procedures, *John Wiley & Sons.*

Holland, Burt, Sudipta Basu and Fang Sun, 2010, Neglect of multiplicity when testing families of related hypotheses, *Working Paper, Temple University.*

Holm, Sture, 1979, A simple sequentially rejective multiple test procedure, *Scandinavian Journal of Statistics 6, 65-70.*

Holthausen, Robert W. and David F. Larcker, 1992, The prediction of stock returns using financial statement information, *Journal of Accounting & Economics 15, 373-411.*

Hommel, G., 1988, A stagewise rejective multiple test procedure based on a modified Bonferroni test, *Biometrika 75, 383-386.*

Hou, Kewei and David T. Robinson, 2006, Industry concentration and average stock returns, *Journal of Finance 61, 1927-1956.*

Hou, Kewei, G. Andrew Karolyi and Bong-Chan Kho, 2011, What factors drive global stock returns?, *Review of Financial Studies 24, 2527-2574.*

Hou, Kewei and Tobias J. Moskowitz, 2005, Market frictions, price delay, and the cross-section of expected returns, *Review of Financial Studies 18, 981-1020.*

Hu, Grace Xing, Jun Pan and Jiang Wang, 2012, Noise as information for illiquidity, *Working Paper, University of Hong Kong.*

Huang, Alan Guoming, 2009, The cross section of cashflow volatility and expected stock returns, *Journal of Empirical Finance 16, 409-429.*

Huang, Wei, Qianqiu Liu, Ghon Rhee and Feng Wu, 2010, Extreme downside risk and expected stock returns, *Journal of Banking & Finance 36, 1492-1502.*

Hvidkjaer, Soeren, 2008, Small trades and the cross-section of stock returns, *Review of Financial Studies 31, 1123-1151.*

Imrohoroglu, Ayse and Selale Tuzel, 2011, Firm level productivity, risk, and return, *Working Paper, University of Southern California.*

Ioannidis, J.P., 2005, Why most published research findings are false, *PLoS medicine 2, e124, 694-701*

Jacobs, Kris and Kevin Q. Wang, 2004, Idiosyncratic consumption risk and the cross section of asset returns, *Journal of Finance 59, 2211-2252.*

Jagannathan, Ravi and Yong Wang, 2007, Lazy investors, discretionary consumption, and the cross-section of stock returns, *Journal of Finance 62, 1623-1661.*

Jagannathan, Ravi and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance 51, 3-53.*

Jarrow, Robert, 1980, Heterogeneous expectations, restrictions on short sales, and equilibrium asset prices, *Journal of Finance 35, 1105-1113.*

Jefferys, William H. and James O. Berger, 1992, Ockham's razor and Bayesian analysis, *American Scientist 80, 64-72.*

Jegadeesh, Narasimhan, 1990, Evidence of predictable behavior of security returns, *Journal of Finance 45, 881-898.*

Jegadeesh, Narasimhan, Joonghyuk Kim, Suan D. Krische and Charles Lee, 2004, Analyzing the analysts: When do recommendations add value? *Journal of Finance 59, 1083-1124.*

Jegadeesh, Narasimhan and Sheridan Titman, 1993, Returns to buying winners and selling losers: implications for stock market efficiency, *Journal of Finance 48, 65-91.*

Jiang, Guohua, Charles MC Lee and Yi Zhang, 2005, Information uncertainty and expected returns, *Review of Accounting Studies 10, 185-221.*

Jiang, Hao and Zheng Sun, 2011, Dispersion in beliefs among active mutual funds and the cross-section of stock returns, *Working Paper, Erasmus University.*

Johnson, Travis L. and Eric C. So, 2012, The option to stock volume ratio and future returns, *Journal of Financial Economics 106, 262-286.*

Jones, Charles M. and Owen A. Lamont, 2002, Short-sale constraints and stock returns, *Journal of Financial Economics 66, 207-239.*

Kapadia, Nishad, 2011, Tracking down distress risk, *Journal of Financial Economics 102, 167-182.*

Kaplan, Steven N. and Luigi Zingales, 1997, Do investment-cash flow sensitivities provide useful measures of financing constraints?, *Quarterly Journal of Economics 112, 169-215.*

Kelly, Bryan and Seth Pruitt, 2011, The three-pass regression filter: A new approach to forecasting using many predictors, *Working Paper, University of Chicago.*

Kim, Chansog Francis, Christos Pantzalis and Jung Chul Park, 2012, Political geography and stock returns: The value and risk implications of proximity to political power, *Journal of Financial Economics 106, 196-228.*

Kraus, Alan and Robert H. Litzenberger, 1976, Skewness preference and the valuation of risk assets, *Journal of Finance 31, 1085-1100.*

Koijen, Ralph SJ, Tobias J. Moskowitz, Lasse Heje Pedersen and Evert B. Vrugt, 2012, Carry, *Working Paper, University of Chicago.*

Korajczyk, Robert A. and Ronnie Sadka, 2008, Pricing the commonality across alternative measures of liquidity, *Journal of Financial Economics 87, 45-72.*

Korniotis, George M., 2008, Habit formation, incomplete markets, and the significance of regional risk for expected returns, *Review of Financial Studies 21, 2139-2172.*

Korniotis, George M. and Alok Kumar, 2009, Long Georgia, short Colorado? The geography of return predictability, *Working Paper, Board of Governors of the Federal Reserve System.*

Kosowski, Robert, Allan Timmermann, Russ Wermers and Hal White, 2006, Can mutual fund "stars" really pick stocks? New evidence from a Bootstrap analysis, *Journal of Finance 61, 2551-2595.*

Kosowski, Robert, Narayan Y. Naik and Melvyn Teo, 2007, Do hedge funds deliver alpha? A Bayesian and bootstrap analysis, *Journal of Financial Economics 84, 229-264.*

Kumar, Alok and Charles MC Lee, 2006, Retail investor sentiment and return comovement, *Journal of Finance 61, 2451-2486.*

Kumar, Praveen, Sorin M. Sorescu, Rodney D. Boehme and Bartley R. Danielsen, 2008, Estimation risk, information, and the conditional CAPM: Theory and evidence, *Review of Financial Studies 21, 1037-1075.*

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica 53, 1315-1335.*

Lamont, Owen, Christopher Polk and Jesus Saa-Requejo, 2001, Financial constraints and stock returns, *Review of Financial Studies 14, 529-554.*

Landsman, Wayne R., Bruce L. Miller, Ken Peasnell and Shu Yeh, 2011, Do investors understand really dirty surplus? *Accounting Review 86, 237-258.*

Larcker, David F., Eric C. So and Charles CY Wang, 2013, Boardroom centrality and firm performance, *Journal of Accounting and Economics 55, 225-250.*

Leamer, Edward E., 1978, Specification searches: Ad hoc inference with nonexperimental data, *New York: John Wiley & Sons.*

Lee, Charles and Bhaskaran Swaminathan, 2000, Price momentum and trading volume, *Journal of Finance 55, 2017-2069.*

Lehavy, Reuven and Richard G. Sloan, 2008, Investor recognition and stock returns, *Review of Accounting Studies 13, 327-361.*

Lehmann, Erich Leo and Joseph P. Romano, 2005, Generalizations of the familywise error rate, *Springer US, 2012.*

Lettau, Martin and Sydney Ludvigson, 2001, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying, *Journal of Political Economy 109, 1238-1287.*

Lev, Baruch, Bharat Sarath and Theodore Sougiannis, 2005, R&D reporting biases and their consequences, *Contemporary Accounting Research 22, 977-1026.*

Lev, Baruch, Doron Nissim and Jacob Thomas, 2005, On the informational usefulness of R&D capitalization and amortization, *Working Paper, Columbia University.*

Lev, Baruch and Theodore Sougiannis, 1996, The capitalization, amortization, and value-relevance of R&D, *Journal of Accounting and Economics 21, 107-138.*

Lewellen, Jonathan, Stefan Nagel and Jay Shanken, 2010, A skeptical appraisal of asset pricing tests, *Journal of Financial Economics 96, 175-194.*

Liang, Yulan and Arpad Kelemen, 2008, Statistical advances and challenges for analyzing correlated high dimensional SNP data in genomic study for complex diseases, *Statist. Surv. 2, 43-60.*

Li, Dongmei, 2011, Financial constraints, R&D investment, and stock returns, *Review of Financial Studies 24, 2975-3007.*

Li, Kevin Ke, 2011, How well do investors understand loss persistence? *Review of Accounting Studies 16, 630-667.*

Li, Qing, Maria Vassalou and Yuhang Xing, 2006, Sector investment growth rates and the cross section of equity returns, *Journal of Business 79, 1637-1665.*

Li, Sophia Zhengzi, 2012, Continuous beta, discontinuous beta, and the cross-section of expected stock returns, *Working Paper, Duke University.*

Li, Xi, 2012, Real earnings management and subsequent stock returns, *Working Paper, Boston College.*

Lintner, John, 1965, Security prices, risk, and maximal gains from diversification, *Journal of Finance 20, 587-615.*

Lioui, Abraham and Paulo Maio, 2012, Interest rate risk and the cross-section of stock returns, *Working Paper, EDHEC Business School.*

Litzenberger, Robert H. and Krishna Ramaswamy, 1979, The effect of personal taxes and dividends on capital asset prices, *Journal of Financial Economics 7, 163-195.*

Liu, Weimin, 2006, A liquidity-augmented capital asset pricing model, *Journal of Financial Economics 82, 631-671.*

Livdan, Dmitry, Horacio Sapriza and Lu Zhang, 2009, Financially constrained stock returns, *Journal of Finance 64, 1827-1862.*

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

Lo, Andrew W. and Jiang Wang, 2006, Trading volume: Implications of an intertemporal capital asset pricing model, *Journal of Finance 61, 2805-2840.*

Loughran, Tim and Anand Vijh, 1997, Do long-term shareholders benefit from corporate acquisitions? *Journal of Finance 52, 1765-1790.*

Loughran, Tim and Jay R. Ritter, 1995, The new issues puzzle, *Journal of Finance 50, 23-51.*

Lucas Jr, Robert E., 1978, Asset prices in an exchange economy, *Econometrica 46, 1429-1445.*

Lustig, Hanno N. and Stijn G. Van Nieuwerburgh, 2005, Housing collateral, consumption insurance, and risk premia: An empirical perspective, *Journal of Finance 60, 1167-1219.*

Lynch, Anthony and Tania Vital-Ahuja, 2012, Can subsample evidence alleviate the data-snooping problem?: A comparison to the maximal $R^2$ cutoff test, *Working Paper, New York University.*

Malloy, Christopher J., Tobias J. Moskowitz and Annette Vissing-Jorgensen, 2009, Long-run stockholder consumption risk and asset returns, *Journal of Finance 64, 2427-2479.*

Markowitz, Harry H. and Gan Lin Xu, 1994, Data mining corrections, *Working Paper, Daiwa Securities Trust Company.*

Mayshar, Joram, 1981, Transaction costs and the pricing of assets, *Journal of Finance 36, 583-597.*

McConnell, John J. and Gary C. Sanger, 1984, A trading strategy for new listings on the NYSE, *Financial Analysts Journal, 34-38.*

McLean, R. David and Jeffrey Pontiff, 2014, Does academic research destroy stock return predictability? *Working Paper, University of Alberta.*

Meinshausen, Nicolai, 2008, Hierarchical testing of variable importance, *Biometrika 95, 265-278.*

Meng, Cliff YK and Arthur P. Dempster, 1987, A Bayesian approach to the multiplicity problem for significance testing with binomial data, *Biometrics 43, 301-311.*

Menzly, Lior and Oguzhan Ozbas, 2010, Market segmentation and cross-predictability of returns, *Journal of Finance 65, 1555-1580.*

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

Michaely, Roni, Richard H. Thaler and Kent L. Womack, 1995, Price reactions to dividend initiations and omissions: overreaction or drift? *Journal of Finance 50, 573-608.*

Mohanram, Partha S., 2005, Separating winners from losers among low book-to-market stocks using financial statement analysis, *Review of Accounting Studies 10, 133-170.*

Moskowitz, Tobias J. and Mark Grinblatt, 1999, Do industries explain momentum?, *Journal of Finance 54, 1249-1290.*

Moskowitz, Tobias J., Yao Hua Ooi and Lasse Heje Pedersen, 2012, Time series momentum, *Journal of Financial Economics 104, 228-250.*

Mossin, Jan, Equilibrium in a Capital Asset Market, *Econometrica 34, 768-783.*

Nagel, Stefan, 2005, Short sales, institutional investors and the cross-section of stock returns, *Journal of Financial Economics 78, 277-309.*

Narayanamoorthy, Ganapathi, 2006, Conservatism and cross-sectional variation in the post-earnings announcement drift, *Journal of Accounting Research 44, 763-789.*

Nguyen, Giao X. and Peggy E. Swanson, 2009, Firm characteristics, relative efficiency and equity returns, *Journal of Financial and Quantitative Analysis 44, 213-236.*

Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics 108, 1-28.*

Nyberg, Peter and Salla Pöyry, 2011, Firm expansion and stock price momentum, *Working Paper, Aalto University.*

Ofek, Eli, Matthew Richardson and Robert F. Whitelaw, 2004, Limited arbitrage and short sales restrictions: evidence from the options markets, *Journal of Financial Economics 74, 305-342.*

Gupta, Manak C. and Aharon R. Ofer, Investor's expectations of earnings growth, their accuracy and effects on the structure of realized rates of return, *Journal of Finance 30, 509-523.*

Oldfield, George S. and Richard J. Rogalski, 1981, Treasury bill factors and common stock returns, *Journal of Finance 36, 337-350.*

Ortiz-Molina, Hernán and Gordon M. Phillips, 2011, Real asset liquidity and the cost of capital, *Working Paper, University of British Columbia.*

Ou, Jane A. and Stephen H. Penman, 1989, Financial statement analysis and the prediction of stock returns, *Journal of Accounting & Economics 11, 295-329.*

Ozoguz, Arzu, 2009, Good times or bad times? Investor's uncertainty and stock returns, *Review of Financial Studies 22, 4377-4422.*

Papanastasopoulos, Georgios, Dimitrios Thomakos and Tao Wang, 2010, The implications of retained and distributed earnings for future profitability and stock returns, *Review of Accounting & Finance 9, 395-423.*

Parker, Jonathan A. and Christian Julliard, 2005, Consumption risk and the cross section of expected returns, *Journal of Political Economy 113, 185-222.*

Pastor, Lubos and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy 111, 643-685.*

Patatoukas, Panos N., 2011, Customer-base concentration: implications for firm performance and capital markets, *Working Paper, University of California Berkeley.*

Patton, Andrew J. and Allan Timmermann, 2010, Monotonicity in asset returns: New tests with applications to the term structure, the CAPM, and portfolio sorts, *Journal of Financial Economics 98, 605-625.*

Penman, Stephen and Xiao-jun Zhang, 2002, Modeling sustainable earnings and P/E ratios with financial statement analysis, *Working Paper, Columbia University.*

Pesaran, M. Hashem and Allan Timmermann, 2007, Selection of estimation window in the presence of breaks, *Journal of Econometrics 137, 134-161.*

Piotroski, Joseph D., 2000, Value investing: The use of historical financial statement information to separate winners from losers, *Journal of Accounting Research 38, 1-41.*

Pontiff, Jeffrey and Artemiza Woodgate, 2008, Share issuance and cross-sectional returns, *Journal of Finance 63, 921-945.*

Porta, Rafael, 1996, Expectations and the cross-section of stock returns, *Journal of Finance 51, 1715-1742.*

Prakash, Rachna and Nishi Sinha, 2012, Deferred revenues and the matching of revenues and expenses, *Contemporary Accounting Research, Forthcoming.*

Price, S. Mckay, James S. Doran, David R. Peterson and Barbara A. Bliss, Earnings conference calls and stock returns: The incremental informativeness of textual tone, *Journal of Banking and Finance 36, 992-1011.*

Pukthuanthong, Kuntara and Richard Roll, 2014, *Working Paper, University of Missouri, Columbia.*

Rajgopal, Shivaram, Terry Shevlin and Mohan Venkatachalam, 2012, Does the stock market fully appreciate the implications of leading indicators for future earnings? Evidence from order backlog, *Review of Accounting Studies 8, 461-492.*

Roll, Richard, 1988, $R^2$, *Journal of Finance 43, 541-566.*

Romano, Joseph P., Azeem M. Shaikh and Michael Wolf, 2008, Formalized data snooping based on generalized error rates, *Econometric Theory 24, 404-447.*

Rosenthal, Robert, 1979, The "file drawer problem" and tolerance for null results, *Psychological Bulletin 86, 638-641.*

Ross, Stephen A., 1989, Regression to the max, *Working Paper, Yale University.*

Romano, Joseph P., Azeem M. Shaikh and Michael Wolf, 2008, Control of the false discovery rate under dependence using the bootstrap and subsampling, *Test 17, 417-442.*

Rubinstein, Mark E., 1973, The fundamental theorem of parameter-preference security valuation, *Journal of Financial and Quantitative Analysis 8, 61-69.*

Rubinstein, Mark E., 1974, An aggregation theorem for securities markets, *Journal of Financial Economics 1, 225-244.*

Sarkar, Sanat K., 2002, Some results on false discovery rate in stepwise multiple testing procedure, *Annals of Statistics 30, 239-257.*

Sarkar, Sanat K. and Wenge Guo, 2009, On a generalized false discovery rate, *The Annals of Statistics 37, 1545-1565.*

Sadka, Ronnie, 2006, Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk, *Journal of Financial Economics 80, 309-349.*

Saville, Dave J., 1990, Multiple comparison procedures: The practical solution, *The American Statistician 44, 174-180.*

Savov, Alexi, 2011, Asset pricing with garbage, *Journal of Finance 66, 177-201.*

Scheffe, H., 1959, The Analysis of Variance, *Wiley, New York.*

Schweder, T. and Eil Spjøtvoll, 1982, Plots of p-values to evaluate many tests simultaneously, *Biometrika 69, 493-502.*

Schwert, G. William, 2003, Anomalies and market efficiency, *Handbook of the Economics of Finance*, edited by G.M. Constantinides, M. Haris and R. Stulz, Elsevier Science B.V.

Scott, James G., 2009, Bayesian adjustment for multiplicity, *Dissertation, Duke University.*

Scott, James G. and James O. Berger, 2006, An exploration of aspects of Bayesian multiple testing, *Journal of Statistical Planning and Inference 136, 2144-2162.*

Scott, James G. and James O. Berger, 2010, Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem, *Annuals of Statistics 38, 2587-2619.*

Shaffer, Juliet Popper, 1995, Multiple hypothesis testing, *Annual Review of Psychology 46, 561-584.*

Shanken, Jay, 1990, Intertemporal asset pricing: An empirical investigation, *Journal of Econometrics 45, 99-120.*

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance 19, 425-442.*

Shu, Tao, 2007, Trader composition, price efficiency, and the cross-section of stock returns, *Working Paper, University of Texas at Austin.*

Simes, R. John, 1986, An improved Bonferroni procedure for multiple tests of significance, *Biometrika 73, 751-754.*

Simutin, Mikhail, 2010, Excess cash and stock returns, *Financial Management 39, 1197-1222.*

Sloan, Richard G., 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting Review 71, 289-315.*

So, Eric C., 2012, A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts? *Working Paper, Stanford University.*

Soliman, Mark T., 2008, The use of DuPont analysis by market participants, *Accounting Review 83, 823-853.*

Solnik, Bruno H., 1974, An equilibrium model of the international capital market, *Journal of Economic Theory 8, 500-524*

Spiess, D. Katherine and John Affleck-Graves, 1995, Underperformance in long-run stock returns following seasoned equity offerings, *Journal of Financial Economics 38, 243-267.*

Spiess, Katherine and John Affleck-Graves, 1995, The long-run performance of stock returns following debt offerings, *Journal of Financial Economics 54, 45-73.*

Storey, John D., 2003, The positive false discovery ratee: A Bayesian interpretation and the q-value, *The Annals of Statistics 31, 2013-2035.*

Stulz, René M., 1981, A model of international asset pricing, *Journal of Financial Economics 9, 383-406.*

Stulz, René M., 1986, Asset Pricing and Expected Inflation, *Journal of Finance 41, 209-223.*

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 mining: The case of calendar effects in stock returns, *Journal of Econometrics 105, 249-286.*

Subrahmanyam, Avanidhar, 2010, The cross-section of expected stock returns: What have we learnt from the past twenty-five years of research? *Working Paper, UCLA.*

Sweeney, Richard J. and Arthur D. Warga, 1986, The pricing of interest-rate risk: evidence from the stock market, *Journal of Finance 41, 393-410.*

Vanden, Joel M., 2004, Options trading and the CAPM, *Review of Financial Studies 17, 207-238.*

Teo, Melvyn and Sung-Jun Woo, 2004, Style effects in the cross-section of stock returns, *Journal of Financial Economics 74, 367-398.*

Thomas, Jacob and Frank X. Zhang, 2011, Tax expense momentum, *Journal of Accounting Research 49, 791-821.*

Thornton, Alison and Peter Lee, 2000, Publication bias in meta-analysis: its causes and consequences, *Journal of Clinical Epidemiology 53, 207-216.*

Titman, Sheridan, Kuo-Chiang Wei and Feixue Xie, 2004, Capital investments and stock returns, *Journal of Financial and Quantitative Analysis 39, 677-700.*

Troendle, James F., 2000, Stepwise normal theory multiple test procedures controlling the false discovery rate, *Journal of Statistical Planning and Inference 84, 139-158.*

Todorov, Viktor and Tim Bollerslev, 2010, Jumps and betas: A new framework for disentangling and estimating systematic risks, *Journal of Econometrics 157, 220-235.*

Tukey, John W., 1977, Exploratory data analysis, *Addison-Wesley.*

Tuzel, Selale, 2010, Corporate real estate holdings and the cross-section of stock returns, *Review of Financial Studies 23, 2268-2302.*

Valta, Philip, 2013, Strategic default, debt structure, and stock returns, *Working Paper, University of Lausanne.*

Van Binsbergen, Jules H., 2009, Good-specific habit formation and the cross-section of expected returns, *Working Paper, Stanford University.*

Vanden, Joel M., 2006, Option coskewness and capital asset pricing, *Review of Financial Studies 19, 1279-1320.*

Vassalou, Maria, 2003, News related to future GDP growth as a risk factor in equity returns, *Journal of Financial Economics 68, 47-73.*

Vassalou, Maria and Yuhang Xing, 2004, Default risk in equity returns, *Journal of Finance 2004, 831-868.*

Viale, Ariel M., Luis Garcia-Feijoo and Antoine Giannetti, 2012, Safety first, robust dynamic asset pricing, and the cross-section of expected stock returns, *Working Paper, Florida Atlantic University.*

Wagenmakers, Eric-Jan and Peter Grünwald, 2005, A Bayesian perspective on hypothesis testing: A comment on Killeen (2005), *Psychological Science 17, 641-642.*

Wahlen, James M. and Matthew M. Wieland, 2011, Can financial statement analysis beat consensus analysts' recommendations? *Review of Accounting Studies 16, 89-115.*

Wang, Yuan, 2012, Debt covenants and cross-sectional equity returns, *Working Paper, Concordia University.*

Watkins, Boyce, 2003, Riding the wave of sentiment: An analysis of return consistency as a predictor of future returns, *Journal of Behavioral Finance 4, 191-200.*

Westfall, Peter H., 1993, Resampling-based multiple testing, *John Wiley & Sons.*

Welch, Ivo and Amit Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies 21, 1455-1508.*

White, Halbert, 2000, A reality check for data snooping, *Econometrica 68, 1097-1126.*

Whited, Toni M. and Guojun Wu, 2006, Financial constraints risk, *Review of Financial Studies 19, 531-559.*

Whittemore, Alice S., 2007, A Bayesian false discovery rate for multiple testing, *Journal of Applied Statistics 34, 1-9.*

Womack, Kent L., 1996, Do brokerage analysts' recommendations have investment value? *Journal of Finance 51, 137-167.*

Xing, Yuhang, 2008, Interpreting the value effect through the Q-Theory: An empirical investigation, *Review of Financial Studies 21, 1767-1795.*

Xing, Yuhang, Xiaoyan Zhang and Rui Zhao, 2010, What does the individual option volatility smirk tell us about future equity returns? *Journal of Financial & Quantitative Analysis 45, 641-662.*

Yan, Shu, 2011, Jump risk, stock returns, and slope of implied volatility smile, *Journal of Financial Economics 99, 216-233.*

Yekutieli, Daniel and Yoav Benjamini, 1999, Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics, *Journal of Statistical Planning and Inference 82, 171-196.*

Yogo, Motohiro, 2006, A consumption-based explanation of expected stock returns, *Journal of Finance 61, 539-580.*

Zehetmayer, Sonja and Martin Posch, 2010, Post hoc power estimation in large-scale multiple testing problems, *Bioinformatics 26, 1050-1056.*

Zhao, Xiaofei, 2012, Information intensity and the cross-section of stock returns, *Working Paper, University of Toronto.*

# A Multiple Testing When the Number of Tests ($M$) is Unknown

The empirical difficulty in applying standard p-value adjustments is that we do not observe factors that have been tried, found to be insignificant and then discarded. We attempt to overcome this difficulty using a simulation framework. The idea is first simulate the empirical distribution of p-values for all experiments (published and unpublished) and then adjust p-values based on these simulated samples.

First, we assume the test statistic (t-statistic, for instance) for any experiment follows a certain distribution $D$ (e.g., exponential distribution) and the set of published works is a truncated $D$ distribution. Based on the estimation framework for truncated distributions,[57] we estimate parameters of distribution $D$ and total number of trials $M$. Next we simulate many sequences of p-values, each corresponding to a plausible set of p-value realizations of all trials. To account for the uncertainty in parameter estimates of $D$ and $M$, we simulate p-value sequences based on the distribution of estimated $D$ and $M$. Finally, for each p-value, we calculate the adjusted p-value based on a sequence of simulated p-values. The median is taken as the final adjusted p-value.

## A.1 Using Truncated Exponential Distribution to Model the t-ratio Sample

Truncated distributions have been used to study hidden tests (i.e., publication bias) in medical research.[58] The idea is that studies reporting significant results are more likely to get published. Assuming a threshold significance level or t-statistic, researchers can to some extent infer the results of unpublished works and gain understanding of the overall effect of a drug or treatment. However, in medical research, insignificant results are still viewed as an indispensable part of the overall statistical evidence and are given much more prominence than in the financial economics research. As a result, medical publications tend to report more insignificant results. This makes applying the truncated distribution framework to medical studies difficult as there is no clear-cut threshold value.[59] In this sense, the truncated distributional framework suits our study better — 1.96 is the obvious hurdle that research needs to overcome to get published.

[57] See Heckman (1979) and Greene (2008), Chapter 24.

[58] See Begg and Berlin (1988) and Thornton and Lee (2000).

[59] When the threshold value is unknown, it must be estimated from the likelihood function. However, such estimation usually incurs large estimation errors.

On the other hand, not all tried factors with p-value above 1.96 are reported. In the quantitative asset management industry significant results are not published — they are considered "trade secrets". For the academic literature, factors with "borderline" t-ratios are difficult to get published. Thus, our sample is likely missing a number of factors that have t-ratios just over the bar of 1.96. To make our inference robust, for our baseline result, we assume all tried factors with t-ratios above 2.57 are observed and ignore those with t-ratios in the range of (1.96, 2.57). We experiment with alternative ways to handle t-ratios in this range.

Many distributions can be used to model the t-ratio sample. One restriction that we think any of these distributions should satisfy is the monotonicity of the density curve. Intuitively, it should be easier to find factors with small t-ratios than large ones.[60] We choose to use the simplest distribution that incorporates this monotonicity condition: the exponential distribution.

Panel A of Figure A.1 presents the histogram of the baseline t-ratio sample and the fitted truncated exponential curve.[61] The fitted density closely tracks the histogram and has a population mean of 2.07.[62] Panel B is a histogram of the original t-ratio sample which, as we discussed before, is likely to under-represent the sample with a t-ratio in the range of (1.96, 2.57). Panel C is the augmented t-ratio sample with the ad hoc assumption that our sample covers only half of all factors with t-ratios between 1.96 and 2.57. The population mean estimate is 2.22 in Panel B and 1.93 in Panel C. As expected, the under-representation of relatively small t-ratios results in a higher mean estimate for the t-ratio population. We think the baseline model is the best among all three models as it not only overcomes the missing data problem for the original sample, but also avoids guessing the fraction of missing observations in the 1.96-2.57 range. We use this model estimates for the follow-up analysis.

Using the baseline model, we calculate other interesting population characteristics that are key to multiple hypothesis testing. Assuming independence, we model observed t-ratios as draws from an exponential distribution with mean parameter $\hat{\lambda}$ and a known cutoff point of 2.57. The proportion of unobserved factors is then estimated as:

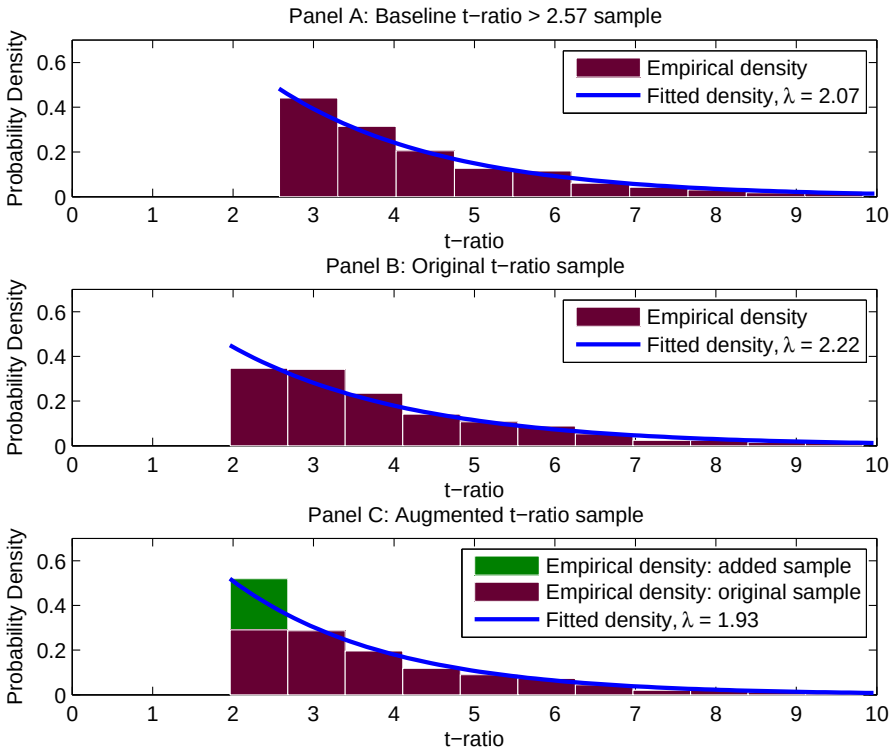
$$P(\text{unobserved}) = \Phi(2.57; \hat{\lambda}) = 1 - \exp(-2.57/\hat{\lambda}) = 71.1\% \tag{A.1}$$

[60] This basic scarcity assumption is also the key ingredient in our model in Section 5.

[61] There are a few very large t-ratios in our sample. We fit the truncated exponential model without dropping any large t-ratios. In contrast to the usual normal density, exponential distribution is better at modeling extreme observations. In addition, extreme values are pivotal statistics for heavy-tailed distributions and are key for model estimation. While extreme observations are included for model estimation, we exclude them in Figure A.1 to better focus on the main part of the t-ratio range.

[62] Our truncated exponential distribution framework allows a simple analytical estimate for the population mean of the exponential distribution. In particular, let c be the truncation point and the t-ratio sample be $\{t_i\}_{i=1}^N$. The mean estimate is given by $\hat{\lambda} = 1/(\bar{t} - c)$, where $\bar{t} = (\sum_{i=1}^N t_i)/N$ is the sample mean.

Figure A.1: **Density Plots for t-ratio**




Empirical density and fitted exponential density curves based on three different samples. Panel A is based on the baseline sample that includes all t-ratios above 2.57. Panel B is based on the original sample with all t-ratios above 1.96. Panel C is based on the augmented sample that adds the sub-sample of observations that fall in between 1.96 and 2.57 to the original t-ratio sample. It doubles the number of observations within the range of 1.96 and 2.57 in the original sample. $\lambda$ is the single parameter for the exponential curve. It gives the population mean for the unrestricted (i.e., non-truncated) distribution.

where $\Phi(c;\lambda)$ is the cumulative distribution function evaluated at $c$ for a exponential distribution with mean $\lambda$. Our estimates indicate that the mean absolute value of the t-ratio for the underlying factor population is 2.07 and about 71.1% of tried factors are discarded. Given that 238 out of the original 316 factors have a t-ratio exceeding 2.57, the total number of factor tests is estimated to be 824 ($= 238/(1-71.1\%)$) and the number of factors with a t-ratio between 1.96 and 2.57 is estimated to be 82.[63] Since our t-ratio sample covers only 57 such factors, roughly 30% (=(82-57)/82) of t-ratios between 1.96 and 2.57 are hidden.

[63]Directly applying our estimate framework to the original sample that includes all t-ratios above 1.96, the estimated total number of factor tests would be 713. Alternatively, assuming our sample only covers half of the factors with t-ratios between 1.96 and 2.57, the estimated number of factors is 971.

## A.2 Simulated Benchmark t-ratios Under Independence

The truncated exponential distribution framework helps us approximate the distribution of t-ratios for all factors, published and unpublished. We can then apply the aforementioned adjustment techniques to this distribution to generate new t-ratio benchmarks. However, there are two sources of sampling and estimation uncertainty that affect our results. First, our t-ratio sample may under-represent all factors with t-statistics exceeding 2.57.[64] Hence, our estimates of total trials are biased (too low), which affects our calculation of the benchmarks. Second, estimation error for the truncated exponential distribution can affect our benchmark t-ratios. Although we can approximate the estimation error through the usual asymptotic distribution theory for MLE, it is unclear how this error affects our benchmark t-ratios. This is because t-ratio adjustment procedures usually depend on the entire t-ratio distribution and so standard transformational techniques (e.g., the delta method) do not apply. Moreover, we are not sure whether our sample is large enough to trust the accuracy of asymptotic approximations.

Given these concerns, we propose a simulation framework that incorporates these uncertainties. We divide it into four steps:

**Step I Estimate $\lambda$ and $M$ based on a new t-ratio sample with size $r \times R$.**

Suppose our current t-ratio sample size is $R$ and it only covers a fraction of $1/r$ of all factors. We sample $r \times R$ t-ratios (with replacement) from the original t-ratio sample. Based on this new t-ratio sample, we apply the above truncated exponential distribution framework to the t-ratios and obtain the parameter estimates $\lambda$ for the exponential distribution. The truncation probability is calculated as $\hat{P} = \Phi(2.57; \hat{\lambda})$. We can then estimate the total number of trials by

$$\hat{M} = \frac{rR}{1 - \hat{P}}$$

**Step II Calculate the benchmark t-ratio based on a random sample generated from $\hat{\lambda}$ and $\hat{M}$ .**

Based on the previous step estimate of $\hat{\lambda}$ and $\hat{M}$, we generate a random sample of t-ratios for all tried factors. We then calculate the appropriate benchmark t-ratio based on this generated sample.

**Step III Repeat Step II 10,000 times to get the median benchmark t-ratio.**

Repeat Step II (based on the same $\hat{\lambda}$ and $\hat{M}$) 10,000 times to generate a collection of benchmark t-ratios. We take the median as the final benchmark t-ratio corresponding to the parameter estimate $(\hat{\lambda}, \hat{M})$.

[64] This will happen if we miss factors published by the academic literature or we do not have access to the "trade secrets" by industry practitioners.

**Step IV Repeat Step I-III 10,000 times to generate a distribution of benchmark t-ratios.**

Repeat Step I-III 10,000 times, each time with a newly generated t-ratio sample as in Step I. For each repetition, we obtain a benchmark t-ratio $t_i$ corresponding to the parameter estimates $(\hat{\lambda}_i, \hat{M}_i)$. In the end, we have a collection of benchmark t-ratios $\{t_i\}_{i=1}^{10000}$.

To see how our procedure works, notice that Steps II-III calculate the theoretical benchmark t-ratio for a t-ratio distribution characterized by $(\hat{\lambda}, \hat{M})$. As a result, the outcome is simply one number and there is no uncertainty around it. Uncertainties are incorporated in Steps I and IV. In particular, by sampling repeatedly from the original t-ratio sample and re-estimating $\lambda$ and $M$ each time, we take into account estimation error of the truncated exponential distribution. Also, under the assumption that neglected significant t-ratios follow the empirical distribution of our t-ratio sample, by varying $r$, we can assess how this under-representation of our t-ratio sample affects results.

Table A.2 shows estimates of $M$ and benchmark t-ratios. When $r = 1$, the median estimate for the total number of trials is 817,[65] almost the same as our previous estimate of 820 based on the original sample. Unsurprisingly, Bonferroni implied benchmark t-ratio (4.01) is larger than 3.78, which is what we get ignoring unpublished works. Holm implied t-ratio (3.96), while not necessarily increasing in the number of trials, is also higher than before (3.64). BHY implied t-ratio increases from 3.39 to 3.68 at 1% significance and from 2.78 to 3.18 at 5% significance. As $r$ increases, sample size $M$ and benchmark t-ratios for all four types of adjustments increase. When $r$ doubles, the estimate of $M$ also approximately doubles and Bonferroni and Holm implied t-ratios increase by about 0.2, whereas BHY implied t-ratios increase by around 0.03 (under both significance levels).

[65] Our previous estimate of 820 is a one-shot estimate based on the truncated sample. The results in Table A.2 are based on repeated estimates based on re-sampled data: we re-sample many times and 817 is the *median* of all these estimates. It is close to the one-shot estimate.

Table A.2: **Benchmark t-ratios When $M$ is Estimated**

Estimated total number of factors tried ($M$) and benchmark t-ratio percentiles based on a truncated exponential distribution framework. Our estimation is based on the original t-ratio sample truncated at 2.57. The sampling ratio is the assumed ratio of the true population size of t-ratios exceeding 2.57 over our current sample size. Both Bonferroni and Holm have a significance level of 5%.

| Sampling ratio | M | Bonferroni | Holm | BHY(1%) | BHY(5%) |
|---|---|---|---|---|---|
| ($r$) | [10% 90%] | [10% 90%] | [10% 90% ] | [10% 90%] | [10% 90%] |
| 1 | 817 | 4.01 | 3.96 | 3.68 | 3.17 |
| | [731 947 ] | [3.98 4.04 ] | [3.92 4.00] | [3.63 3.74 ] | [3.12 3.24] |
| 1.5 | 1234 | 4.11 | 4.06 | 3.70 | 3.20 |
| | [1128 1358 ] | [4.08 4.13 ] | [4.03 4.09] | [3.66 3.74 ] | [3.16 3.24] |
| 2 | 1646 | 4.17 | 4.13 | 3.71 | 3.21 |
| | [1531 1786 ] | [4.15 4.19 ] | [4.11 4.15] | [3.67 3.75 ] | [3.18 3.25] |

# B A Simple Bayesian Framework

The following framework is adopted from Scott and Berger (2006). It highlights the key issues in Bayesian multiple hypothesis testing.[66] More sophisticated generalizations modify the basic model but are unlikely to change the fundamental hierarchical testing structure.[67] We use this framework to explain the pros and cons of performing multiple testing in a Bayesian framework.

The hierarchical model is as follows:

**H1.** $(X_i|\mu_i, \sigma^2, \gamma_i) \overset{iid}{\sim} N(\gamma_i\mu_i, \sigma^2)$,

**H2.** $\mu_i|\tau^2 \overset{iid}{\sim} N(0, \tau^2), \gamma_i|p_0 \overset{iid}{\sim} Ber(1-p_0)$,

**H3.** $(\tau^2, \sigma^2) \sim \pi_1(\tau^2, \sigma^2), p_0 \sim \pi_2(p_0)$.

We explain each step in detail as well as the notation:

**H1.** $X_i$ denotes the average return generated from a long-short trading strategy based on a certain factor; $\mu_i$ is the unknown mean return; $\sigma^2$ is the common variance for returns and $\gamma_i$ is an indicator function, with $\gamma_i = 0$ indicating a zero factor mean. $\gamma_i$ is the counterpart of the reject/accept decision in the usual (frequentists') hypothesis testing framework.

H1 therefore says that factor returns are independent conditional on mean $\gamma_i\mu_i$ and common variance $\sigma^2$, with $\gamma_i = 0$ indicating that the factor is spurious. The common variance assumption may look restrictive but we can always scale factor returns by changing the dollar investment in the long-short strategy. The crucial assumption is conditional independence of average strategy returns. Certain form of conditional independence is unavoidable for Bayesian hierarchical modeling[68] — probably unrealistic for our application. We can easily think of scenarios where average returns of different strategies are correlated, even when population means are known. For example, it is well known that two of the most popular factors, the Fama and French (1992) HML and SMB are correlated.

[66] We choose to present the full Bayes approach. An alternative approach — the empirical-Bayes approach — is closely related to the BHY method that controls the *false-discovery rate* (FDR). See Storey (2003) and Efron and Tibshirani (2002) for the empirical-Bayes interpretation of FDR. For details on the empirical-Bayes method, see Efron, Tibshirani, Storey and Tusher (2001), Efron (2004) and Efron (2006). For an in-depth investigation of the differences between the full Bayes and the empirical-Bayes approach, see Scott and Berger (2010). For an application of the empirical-Bayes method in finance, see Markowitz and Xu (1994).

[67] See Meng and Dempster (1987) and Whittemore (2007) for more works on the Bayesian approach in hypothesis testing.

[68] Conditional independence is crucial for the Bayesian framework and the construction of posterior likelihoods. Although it can be extended to incorporate special dependence structures, there is no consensus on how to systematically handle dependence. See Brown et al. (2012) for a discussion of independence in Bayesian multiple testing. They also propose a spatial dependence structure into a Bayesian testing framework.

**H2.** The first step population parameters $\mu_i$'s and $\gamma_i$'s are assumed to be generated from two other parametric distributions: $\mu_i$'s are independently generated from a normal distribution and $\gamma_i$'s are simply generated from a Bernoulli distribution, i.e., $\gamma_i = 0$ with probability $p_0$.

The normality assumption for the $\mu_i$'s requires the reported $X_i$'s to randomly represent either long/short or short/long strategy returns. If researchers have a tendency to report positive abnormal returns, we need to randomly assign to these returns plus/minus signs. The normality assumptions in both H1 and H2 are important as they are necessary to guarantee the properness of the posterior distributions.

**H3.** Finally, the two variance variables $\tau^2$ and $\sigma^2$ follow a joint prior distribution $\pi_1$ and the probability $p_0$ follows a prior distribution $\pi_2$.

Objective or "neutral" priors for $\pi_1$ and $\pi_2$ can be specified as:

$$\begin{aligned} \pi_1(\tau^2, \sigma^2) &\propto (\tau^2 + \sigma^2)^{-2} \\ \pi_2(p_0) &= \text{Uniform}(0, 1) \end{aligned}$$

Under this framework, the joint conditional likelihood function for $X_i$'s is simply a product of individual normal likelihood functions and the posterior probability that $\gamma_i = 1$ (discovery) can be calculated by applying Bayes' law. When the number of trials is large, to calculate the posterior probability we need efficient methods such as importance sampling, which involves high dimensional integrals.

One benefit of a Bayesian framework for multiple testing is that the multiplicity penalty term is already embedded. In the frequentists' framework, this is done by introducing FWER or FDR. In a Bayesian framework, the so-called "Ockham's razor effect"[69] automatically adjusts the posterior probabilities when more factors are simultaneously tested.[70] Simulation studies in Scott and Berger (2006) show how the discovery probabilities for a few initial signals increase when more noise are added to the original sample.

However, there are several shortcomings for the Bayesian approach. Some of them are specific to the context of our application and the others are generic to the Bayesian multiple testing framework.

At least two issues arise when applying the Bayesian approach to our factor selection problem. First, we do not observe all tried factors. While we back out the distribution of hidden factors parametrically under the frequentist framework, it is not clear how the missing data and the multiple testing problems can be simultaneously

[69] See Jefferys and Berger (1992).

[70] Intuitively, more complex models are penalized because extra parameters involve additional sources of uncertainty. Simplicity is rewarded in a Bayesian framework as simple models produce sharp predictions. See the discussions in Scott (2009).

solved under the Bayesian framework. Second, the hierarchical testing framework may be overly restrictive. Both independence as well as normality assumptions can have a large impact on the posterior distributions. Although normality can be somewhat relaxed by using alternative distributions, the scope of alternative distributions is limited as there are only a few distributions that can guarantee the properness of the posterior distributions. Independence, as we previously discussed, is likely to be violated in our context. In contrast, the three adjustment procedures under the frequentists' framework are able to handle complex data structures since they rely on only fundamental probability inequalities to restrict their objective function — the Type I error rate.

There are a few general concerns about the Bayesian multiple testing framework. First, it is not clear what to do after obtaining the posterior probabilities for individual hypotheses. Presumably, we should find a cutoff probability $P$ and reject all hypotheses that have a posterior discovery probability larger than $P$. But then we come back to the initial problem of finding an appropriate cutoff p-value, which is not at all a clear task. Scott and Berger (2006) suggest a decision-theoretic approach that chooses the cutoff $P$ by minimizing a loss-function. The parameters of the loss-function, however, are again subjective. Second, the Bayesian posterior distributions are computationally challenging. We document three hundred factors but there are potentially many more if missing factors are taken into account. When $M$ gets large, importance sampling is a necessity. However, results of importance sampling rely on simulations and subjective choices of the centers of the probability distributions for random variables. Consequently, two researchers trying to calculate the same quantity might get very different results. Moreover, in multiple testing, the curse of dimensionality generates additional risks for Bayesian statistical inference.[71] These technical issues create additional hurdles for the application of the Bayesian approach.

[71] See Liang and Kelemen (2008) for a discussion on the computational issues in Bayesian multiple testing.

# C Method Controlling the FDP

We apply the methods developed in Lehmann and Romano (2005) to control the realized FDP. The objective is $P(FDP > \gamma) \leq \alpha$, where $\gamma$ is the threshold FDP value and $\alpha$ is the significance level. Fixing $\gamma$ and $\alpha$, we order the individual p-values from the smallest to the largest (i.e., $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(M)}$) and let the corresponding hypotheses be $H_{(1)}, H_{(2)}, \cdots, H_{(M)}$. We then reject the i-th hypothesis if $p_{(i)} \leq \alpha_i / C_{\lfloor \gamma M \rfloor + 1}$, where

$$\begin{aligned} \alpha_i &= \frac{(\lfloor \gamma i \rfloor + 1)\alpha}{M + \lfloor \gamma i \rfloor + 1 - i}, \\ C_k &= \sum_{j=1}^{k} \frac{1}{j}. \end{aligned}$$

Here, for a real number $x$, $\lfloor x \rfloor$ denotes the greatest integer that is no greater than $x$. Similar to $c(M)$ in BHY's adjustment, $C_{\lfloor \gamma M \rfloor + 1}$ allow one to control the FDP under arbitrary dependence structure of the p-values.

Table C.1 shows the benchmark t-ratios based on our sample of 316 factors for different levels of FDP thresholds and significance . The benchmark t-ratios are higher when the FDP thresholds are tougher (i.e., $\gamma$ is lower) or when the significance levels are lower (i.e., $\alpha$ is lower). For typical values of $\gamma$ and $\alpha$, the benchmark t-ratios are significantly lower than conventional values, consistent with previous results based on the FWER or FDR methods. For instance, when $\gamma = 0.10$ and $\alpha = 0.05$, the benchmark t-ratio is 2.70 (p-value = 0.69%), much higher than the conventional cutoff of 1.96.

Table C.1: **Benchmark t-ratios for Lehmann and Romano (2005)**

Estimated benchmark t-ratios based on Lehmann and Romano (2005). The objective is $P(FDP > \gamma) \leq \alpha$.

| | $\gamma = 0.05$ | $\gamma = 0.10$ | $\gamma = 0.20$ |
|---|---|---|---|
| $\alpha = 0.01$ | 3.70 | 3.46 | 3.25 |
| $\alpha = 0.05$ | 3.04 | 2.70 | 2.38 |
| $\alpha = 0.10$ | 2.38 | 2.16 | 2.16 |

# D FAQ

## D.1 Specific Questions

- *Why is FWER called "rate" when it is a single number? (Section 4.3)*

  FWER has been used by the statistics literature a long time ago, even before 1979. However, Holm (1979) seems to be the first one that formally defines the *family-wise error rate*. Terms used in Holm (1979) are different from our current presentation. Our "*family-wise*" terminology is likely first mentioned in Cox (1982) and later formally defined in Hochberg and Tamhane (1987). "Rate" is the standard terminology nowadays, though we are not sure of the historical reason for calling it "rate" instead of probability. But we notice people using "Type I error rate" instead of "Type I error probability" in single hypothesis testing. We think that "probability" can be used interchangeably with "rate" since "probability" is "rate" in frequentists' view.

- *How can we tell in real time the errors? (Section 4.3.2)*

  We never know the "true" errors. Even with out-of-sample testing, all we can tell is how likely it is for a factor to be "real" for one particular *realization* of historical returns.

- *Is it possible to set the actual Type I error rate to be exactly at the pre-specified level? (Section 4.3.2)*

  Any adjustment procedure has a theoretically implied upper bound on the Type I error rate. This bound is the "actual Type I error rate" (as opposed to the realized Type I error rate for a particular outcome of a multiple test) and usually achievable under specific distributional assumptions (e.g.,negative dependence among p-values as in BHY). We usually use the distance between this bound and the pre-specified significance level to measure the goodness of a procedure. In reality, for a particular sequence of p-value realizations, e.g., 316 p-values for our 316 factors, we cannot do much. By following a specific adjustment procedure, we can say what the maximal *expected* Type I error rate is if we repeat such multiple testing many times, each time with a different p-value sequence. Comparing two procedures A and B, we want to know whether the expected Type I error rate (after integrating out the randomness in the return data) under A is closer to the significance level than it is under B. It makes little sense to compare A and B based on a particular outcome (e.g., 316 p-values) of a multiple test.

- *Why doesn't the t-value go to something much larger than 3.5 after so many tests (Section 4.6)*

  We report t-ratios not p-values. Suppose you start with a cutoff p-value of 0.05. For a single test, the t-ratio needs to be 2.0 or above. Now consider a multiple testing framework. For simplicity, consider the Bonferroni test. If there are two tests, appropriate cutoff is a p-value of 0.025. For 10 tests, the p-value drops to 0.005. The table below shows the number of multiple tests necessary for certain levels of t-ratios. For example, if we had 87,214 tests, then the Bonferroni would require the factor to have a t-ratio of 5.0 to be deemed significant (p-value of 0.00000057).

Table D.1: **Bonferroni t-ratios and Required Number of Tests**

Bonferroni t-ratios, cut-off p-values and the required number of tests.

| t-ratio | p-value | # of Bonferroni tests |
|---|---|---|
| 2 | 0.05 | 1 |
| 3 | 0.0027 | 19 |
| 4 | 0.000063 | 789 |
| 5 | 0.00000057 | 87,214 |
| 6 | 0.0000000020 | 25,340,000 |
| 7 | $2.56 \times 10^{-12}$ | $1.95 \times 10^{10}$ |
| 8 | $1.33 \times 10^{-15}$ | $3.75 \times 10^{13}$ |

- *Why is there a drop for the time-series of BHY implied t-ratios? (Section 4.6)*

  In Figure 3, there seems to be a drop for BHY implied t-ratios around 1995. Unlike Bonferroni or Holm, BHY implied benchmark t-ratios are not necessarily monotonically increasing in the number of factors. This is because *false discovery rate* (FDR) is about the proportion of false discoveries among all discoveries. Given a set of t-ratios for the years before 1995, suppose we find the BHY implied adjusted t-ratio. In year 1995, suppose we observe a set of large t-ratios. These large t-ratios will likely increase the denominator of FDR (i.e., the number of discoveries $R$). At the same time, they are unlikely to increase the numerator (i.e., the number of false discoveries $N_{0|r}$). As a result, including this new set of large t-ratios into the previous t-ratio set, the new BHY implied benchmark t-ratio will likely decrease. The highly significant t-ratios for 1995 dilute the proportion of false discoveries made based on the t-ratios from previous years.

- *Is "...control their Type I error rates under arbitrary distributional assumptions" really true? Suppose we had 186 factors but they were 99% correlated — effectively just one factor. This seems to me to be a situation where independent test criterion is appropriate. (Section 4.7)*

The statement is correct and the concern is about the Type II rather than Type I error of the testing procedure. In the example, it is true that independent criterion makes more sense. But multiple testing procedures are also able to control the Type I error rates, albeit too much in this case. For instance, Bonferroni implies a threshold t-ratio of 3.8 when there are 316 factors. If most of the factors are perfectly correlated, then the FWER under Bonferroni's criterion is effectively zero. Since zero is less than any pre-specified significance level, the tests still control what they are supposed to control — Type I error rate (FWER or FDR). Of course, the power of the test, which, as previously discussed, can be measured by the distance between the actual Type I error rate and the pre-specified level, would be too low.

- *How does incomplete coverage of "significant" factors affect our results? (Section 4.7)*

  It is likely that our sample somewhat under-represents the population of significant factors. As previously discussed, there are a number of causes for this under-coverage. First, there are some truly significant factors that were tested as insignificant and never made it to publication. Second, we are highly selective in choosing among working papers. Third, we only consider the top three finance journals in choosing among published works. This under-coverage will impact our t-value cutoffs. To quantitatively evaluate this impact, we tabulate a new set of cutoffs: those generated under different degrees of under-representation of the population of significant factors. Table D.2 reports the cutoff t-statistics for 2012. Assuming a true population size over our sample size ratio of $r$, we report adjusted t-ratios for our three approaches.[72] The top row corresponds to a sample size ratio of one, i.e., our original sample. We see that when the true population is twice as large as our sample, Bonferroni implied benchmark t-ratio increases from 3.78 to 3.95 and Holm from 3.64 to 3.83. Relative to the percentage change in t-ratios, the corresponding change in p-values is large. For Bonferroni, p-value changes from 0.016% to 0.008%; for Holm from 0.027% to 0.013%. Both p-values drop by at least half. For BHY, however, the change is less dramatic. This is consistent with our previous discussion of the stationarity of BHY. In sum, we think a robust t-ratio range for Bonferroni and Holm is 3.6-4.0; for BHY, 3.3-3.4 when $\alpha_d = 1\%$ and 2.80-2.85 when $\alpha_d = 5\%$.

[72] Assuming this $r$ ratio and sample size $N$, we obtain Bonferroni adjusted t-ratios straightforwardly based on total number of factors $Nr$. For Holm and BHY, we sample (with replacement) $N(r-1)$ values from the recent 10 years' t-ratios sample. Together with the original sample, we have an augmented sample of $Nr$ t-ratios. We follow Holm or BHY to get the adjusted t-ratio benchmarks for each augmented sample. Finally, we generate $W = 1000$ such augmented samples and take the median as the final benchmark t-ratio. When $Nr$ or $N(r-1)$ are not integers, we use the smallest integer that is greater than or equal to $Nr$ and $N(r-1)$, respectively.

Table D.2: **Cutoff t-ratios for Alternative Sample Sizes**

Benchmark t-ratios and their associated p-values for the three multiple testing adjustments for 2012. Sample size ratio is true population size over our sample size. Both Bonferroni and Holm have a significance level of 5%.

| Sample size ratio ($r$) for significant factors | Bonferroni [p-value] | Holm [p-value] | BHY(1%) [p-value] | BHY(5%) [p-value] |
|---|---|---|---|---|
| 1 | 3.78 [0.016%] | 3.64 [0.027%] | 3.34 [0.08%] | 2.78 [0.544%] |
| 2 | 3.95 [0.008%] | 3.83 [0.013%] | 3.39 [0.070%] | 2.81 [0.495%] |
| 3 | 4.04 [0.005%] | 3.93 [0.008%] | 3.41 [0.065%] | 2.84 [0.451%] |

- *Haven't there been recent advances in Bayesian literature with respect to multiple testing? (Appendix B)*

  In papers that apply the Bayesian testing method, there are many new ways that try to handle inadequacies. For instance, to relax the independence assumption, Brown et al. (2012) introduce an autoregressive dependence structure because their data are obtained sequentially. But they have to assume that noise from the autoregressive processes is independent from the rest of the system. Conditional independence is key to Bayesian modeling. There are ways to circumvent it, but most methods are data-driven and not applicable in our context. For instance, it is unclear how to model dependence among the test statistics of factors in our list. The indeterminacy of the cutoff is mentioned in Scott and Berger (2006). There are many applied works that propose ad hoc methods to try to establish a threshold. Finally, computational difficulty is a longstanding issue in Bayesian literature. People often discard Bayesian methods because they incorporate a "subjective" prior (i.e.,generate random samples around a region where researchers "believe" the parameters should be concentrated) into their posterior calculation. Multiple testing introduces dimensionality concerns, and it is well-known that posterior distributions are hard to calculate accurately when the dimensionality is high. In sum, we think the above three issues are generic to the Bayesian multiple testing framework and for which there are no simple/systematic solutions.

## D.2 General Questions

- *What if the underlying data are non-stationary in that as anomalies are discovered, they are arbitraged away; some newer frictions/biases arise, they are discovered, and then arbitraged away, and so on? This seems like a possible alternative that would lead to the creation of more and more factors over time, without necessarily implying that the t-ratio ought to be raised for newer factors.*

  Our preferred view is that the factor universe is a combination of some stationary factors that cannot be arbitraged away (systematic risk) and some other transitory factors that can be arbitraged away once discovered. As time goes by and we accumulate more data, stationary factors tend to become more significant (t-ratio proportional to the square root of the number of time periods). Through our multiple testing framework, the adjusted benchmark t-ratio becomes higher. This higher bar helps screen newly discovered transitory factors. In other words, it should be harder to propose new transitory factors as longer sample is available. Without multiple testing, recent transitory factors are just as likely to be discovered as past transitory factors. This means that the discovery rate for transitory factors will remain high (if not higher) as time goes by. This is exactly the trend that we try to curb. Ideally, the finance profession should focus on systematic/stationary factors rather than transitory abnormalities.

- *What if many of the factors are highly correlated or at least within a "span" other than the common 3-4 factors like the Fama-French three factors and Momentum which are controlled for while finding new factors? That is, is it possible that the literature has just been rediscovering "new" factors but they remain spanned by other documented factors that did not become an "industry" like Fama-French three factors and Momentum factors?*

  This is possible, although as we mentioned in the paper, newly proposed factors often need to "stand their ground" against similar factors (not just Fama-French three factors and Momentum) that are previously proposed. All of our three adjustments are robust to correlations among the factors. This means that the Type I error (rate of false discoveries) is still under control. However, high correlations make our adjustment less powerful, that is, the benchmark t-ratio is too high for a new factor to overcome. However, given the hundreds of factors proposed, we think it is time to worry more about the Type I error than the power of the tests. A recent paper by Green, Hand and Zhang (2013) show that the correlations among strategy signals are low on average. This seems to suggest that new factors proposed in the literature are somewhat independent from past ones.

- *Should the benchmark t-ratios be higher simply because the number of data points has increased through time?*

For a single, independent test, the t-ratio threshold should remain constant through time. For a return series that has a mean and variance, it is true that its t-ratio will increase as we have more data points. However, this does not imply a higher t-value threshold for hypothesis testing. At 5% significance level, we should always use 2.0 for single test as it gives the correct Type I error rate under the null hypothesis that mean return is zero. As time goes by, truly significant factors are more likely to be identified as significant and false factors are more likely to be identified as insignificant. In other words, the power of the test is improved as we have more observations but this should not change the cutoff value.

In fact, when the sample size is extremely large, it becomes very easy to generate large t-statistics. In this case, people often use alternative statistics (e.g., odds ratios) to summarize strategy performance.

- *How does Kelly and Pruitt (2011, "The three-pass regression filter: A new approach to forecasting using many predictors") relate to our paper?*

  Kelly and Pruitt (2011) is related to our paper in that it also tries dimension reduction when there is a large cross-section. However, their paper is fundamentally different from ours. Kelly and Pruitt (2011) try to extract a few factors from the cross-section and use them to forecast other series. Therefore, the first-step extraction needs to be done in a way that increases the forecasting power in the second step. Our paper stops at the first stage: we look to condense the factor universe that can explain the cross-section of returns.