Dissecting Market Efficiency*

Rasa Karapandza[†] Jose M. Marin[‡]

This version: March 15th, 2012

ABSTRACT

In this paper we introduce a new methodology to test market efficiency and to assess the performance of the most widely accepted asset pricing models. We use this methodology to test the semi-strong form of market efficiency in the context of publicly available accounting information. Instead of testing for a single accounting-based firm characteristic that can generate abnormal excess returns under a given asset pricing model, we simultaneously test a large set of public accounting information in the context of several asset pricing models. Our results cast considerable doubt on both the efficient markets hypothesis and the standard asset pricing models used to test that hypothesis. The evidence clearly suggests (a) that market inefficiency is not explained by a few known pricing anomalies but rather is a broad and general phenomena and (b) that standard asset pricing models fail to meet the minimum predictive requirements for such models.

JEL classification: G12.

Keywords: market efficiency, EMH, Fama–French, pricing anomalies, momentum.

*We thank to seminar participants at European Business School, University of St. Gallen and Universidad de Carlos III. Both authors acknowledge help from the Spanish Ministry of Science and Innovation (MICINN Grant ECO2008-05140) and Barcelona Super Computing Center Grants FI-2009-3-0014 and FI-2010-1-0019.

[†]European Business School (EBS), Wiesbaden, Germany; Email: rasa.karapandza@ebs.edu; web site: www.rasakarapandza.com

[‡]Universidad Carlos III de Madrid, Madrid, Spain; Email: josemaria.marin@uc3m.es; web site: www.josemarin.com

I. Introduction

The efficient markets hypothesis (EMH) is one of the pillars of modern finance. Individual positions about this hypothesis in academia range from believing in full efficiency to thinking there is no efficiency, but there is a general consensus that markets are largely efficient overall while coexisting with some inefficiencies (anomalies). Paraphrasing Malkiel (2005), the evidence is overwhelming that markets are far more efficient (albeit far less predictable) than some research would lead us to believe. This paper introduces a new methodology to test market efficiency and to assess the performance of the most widely accepted asset pricing models. We use this methodology to test the semi-strong form of market efficiency in the context of publicly available accounting information. The results of the study run strongly counter to both the EMH and the standard asset pricing models that are used to test it. In particular, our evidence indicates that market inefficiency is not restricted to isolated pricing anomalies and is instead a broad, general phenomena; our results also suggest that the standard asset pricing models are faulty.

There is a vast literature testing the existence of pricing anomalies associated with publicly available accounting information.[1] In this literature, price efficiency is most frequently tested in the context of the Fama and French (1993) methodology, which dominates empirical research in this area (Cochrane 2005). These tests are constructed as follows. Firms are sorted based on a particular characteristic; then tests evaluate the mispricing of portfolios formed with firms in the extremes of the sorts (say, in the the top and the bottom deciles) while assuming a linear factor model of expected returns. Each result in this literature reveals a piece of the puzzle but says nothing about the puzzle's overall size. In other words: assessing market efficiency by way of anomalies allows us to establish that markets are not fully (semi-strongly) efficient, but it does not tell us just *how* inefficient they are.[2] Moreover, most of the anomalies discovered by researchers disappear over time. As Malkiel (2005) notes, in the end we are never sure whether the anomalies were mere statistical bias or were true effects that, once publicized, faded out as traders adjusted their strategies to "arbitrage away" each anomaly.[3]

[1] For a recent review, see Richardson et al. (2010).

[2] We remark that such conjectures regarding market efficiency are made under the assumption that the test's asset pricing model is the true model; it is well known that the EMH cannot be tested independently of the asset pricing model.

[3] McLean and Pontiff (2012) claim that twice as many alphas disappear over time because of trading than are accounted for by revealed statistical bias.

Those findings motivate the present paper, in which our main goal is to assess the extent of market inefficiency—that is, the size of the puzzle.

Rather than starting with a single firm characteristic as a potential source of mispricing, in this paper we start with a large universe of randomly selected potential sources of mispricing and test them jointly. More specifically, we examine a set of firm characteristics created from arithmetic combinations of all accounting items reported in Compustat (from three to ten accounting items combined via addition, subtraction, multiplication, and/or division). We then sort firms according to these firm characteristics and compute the returns of dynamic portfolios created from the extreme deciles of those sorts. Finally, we test for the mispricing of these portfolios using several linear asset pricing models.

The setup proceeds as follows. We begin by randomly selecting a number N, which can range from 3 to 10; then we randomly select N items from the Compustat database and perform $N-1$ randomly selected arithmetic operations. The result of each sequence of operations is, in fact, a firm characteristic. For example, a random firm characteristic so created might be Total assets $-$ Cash and short-term investments $+$ Debt in current liabilities. In the next step we calculate the mispricing of dynamic portfolios that are created from the top and bottom deciles of the sort on Total assets $-$ Cash and short-term investments $+$ Debt in current liabilities. These constitute our *Informed* portfolios because they are constructed using accounting information. We repeat this procedure for many randomly created firm characteristics, which results in a large set of *Informed* portfolios and their associated alphas.

The output of this exercise is a distribution of alphas for each particular asset pricing model used. If we assume that the particular pricing model used is the true one, then these alphas may correspond to true anomalies (either new or previously identified) or may be nothing more than data artifacts. The distribution of alphas per se is informative but is not the final word for several reasons. The alphas obtained in our experiment could (1) be spurious owing to pure randomness in security prices, (2) reflect known or unknown pricing anomalies that are unrelated to accounting information, (3) reflect known or unknown accounting-based anomalies, or (4) reflect a few unknown anomalies that are picked up multiple times by one or more of our randomly created firm characteristics. In order to position our analysis with respect to the state of the art in asset pricing—and to assess accurately the extent of market efficiency—we must sort out these possible sources of mispricing. That is to say, if one tests many different firm characteristics then it is nearly certain some of these characteristics will turn out to be statistically significant even though they do not apply in the real world. It could also be the case that all alphas obtained in the exercise correspond to known pricing anomalies and are not necessarily related to accounting information. For instance, it is widely recognized that the Fama–French model fails to price portfolios accurately in some of the 25 cells of the well-known 5×5 sort of companies by size and book-to-market ratio (B/M).

Suppose we discover that all the alphas obtained in our exercise, which relies on public accounting information, are associated with small companies. Such a finding would be of interest in its own right, as it would suggest that accounting information is not an additional source of market inefficiency (beyond what is already known) and also that failures of the Fama–French model are a function of how investors process the accounting information of small companies. However, this would add little to what is already known. Another possibility is that all the alphas obtained in the exercise correspond to known accounting-based asset pricing anomalies. Such a result would generate interest by suggesting that the extant literature has already addressed the full extent of price inefficiency associated with public accounting information; yet this would not add much to the debate, either. Finally, it might also be that, in many of our randomly created variables, there is a Compustat variable that dominates and determines many portfolio sorts. A single anomaly present in many of the regressions could thereby artificially inflate the strength of our results.

To address these possible problems, we proceed as follows. In order to control for spurious alphas and alphas of known anomalies, we compare the distribution of the alphas of our *Informed* portfolios with the distribution of alphas generated from two sets of control portfolios. The *Spurious Control* portfolios are generated in the same way as the *Informed* portfolios except that no information is used when sorting and forming portfolios. These portfolios are formed by replacing each firm in the *Informed* portfolios with a randomly selected company from the universe of all firms. The *FF25 Control* portfolios are created by replacing each firm in the *Informed* portfolios with a randomly selected company positioned in the corresponding cell of the 5×5 Fama–French matrix. This procedure allows for comparisons because each portfolio in the two control groups shares all other characteristics (number of securities, number of each security's periods in the dynamic portfolio, etc.) with its counterpart *Informed* portfolio while differing along one particular dimension.

The *Informed* portfolios are created conditional on the information set of publicly available accounting information; in contrast, the *Spurious Control* portfolios are completely uninformative since all firms in these portfolios are selected randomly. In between these extremes are the *FF25 Control* portfolios, which incorporate information but only on firm size and B/M—although neither variable need be related to accounting information. The advantage of this setup is that it allows us to compare the distributions of alphas in the *Informed* versus *Spurious Control* portfolios and thereby asses the extent of mispricing that is associated with pure accounting information. We can also compare the *Informed* and *FF25 Control* portfolios and thus identify potential anomalies that are not associated with known sources of mispricing (e.g., firm size and B/M).

In order to control for known accounting-based anomalies, we compare the distribution of alphas for two alternative sets of accounting variables: the *original* Compustat data set and the *extended* Compustat data set. The latter augments the former with items related to known accounting-based pricing anomalies. Finally, to show that our results are not driven by either one or a few unknown anomalies, we create an additional robustness check. Namely, we exclude from our analysis any portfolio created from sorts on our artificial firm characteristics that is pairwise related to any portfolio created from sorts on other such characteristics.

From our analysis there follow several important conclusions regarding market efficiency. First, the set of anomalies derived from publicly available Compustat information is much larger than the set of accounting-based anomalies that have been disclosed by existing research. This newly identified mispricing is severe. Publicly available information contained in Compustat can generate abnormal excess returns exceeding 26.5% each month, which amounts to more than 300% annually. In more than three decades of research on the subject, academia has been able to uncover only the tip of a large iceberg. Second, the accounting-based anomalies reported here are explained neither by known sources of mispricing (e.g., size and value) nor by randomness in asset returns; hence they are not spurious. Furthermore, these anomalies survive all the standard robustness checks required in the anomalies literature. Thus, the distribution we describe consists of unique anomalies created by sorts on unrelated variables. Our study therefore demonstrates that focusing on only one potential source of information (here public accounting information) is enough to conclude that market inefficiency is a broader and more extensive phenomenon than previously believed. Future applications of our methodology to sources of possible mispricing *other* than public accounting information may well reinforce that conclusion.

In this paper we also provide some characterization of the accounting-based anomalies identified. By construction, our methodology allows us to compare the extent of mispricing associated with the different information sets used for portfolio creation, different time intervals, different factor models, and so forth. In particular, we analyze the relationship between the number of anomalies and the complexity of trading rules. With increases in the number of accounting variables involved (in the arithmetic operations used to generate firm characteristics and sort portfolios), we find both more and significantly larger alphas. In other words, we find that mispricing increases monotonically as the firm characteristics used to create *Informed* portfolios become increasingly complex (i.e., are created from a greater number of Compustat variables). Similarly, we establish that markets have become more efficient over time. This finding is consistent with investors using increasingly complex strategies to exploit market inefficiencies, which in turn makes them disappear.[4] We also identify the most common accounting variables involved in abnormal returns. This analysis not only corroborates the role played by the usual suspects identified in the literature but also points to some new ones.

As for asset pricing models, in this paper we focus on those that are most central to and most widely used in the literature: the CAPM, the Fama and French three-factor model, and three versions of the four-factor model that include a momentum factor (one using Fama–French momentum, one using the original Carhart factor, and one using our own construction of the Carhart factor). We have two reasons for analyzing these three alternative specifications of the momentum factor. First, there is some controversy regarding whether or not momentum should be viewed as a risk factor. Second, researchers routinely pick one of these specifications but without considering that the choice itself might bias the analysis. Because our analysis is already fairly complex, we do not analyze extensions of the four-factor model—for instance, those including the liquidity factors of Pastor and Stambaugh (2003) and Sadka (2010).

We restrict our attention to alpha as a measure of performance. It is noteworthy that the five models just described are nested (exception when we are comparing the three four-factor models with each

[4]This evidence seems to be consistent with changes in trading technologies over time. In particular, the increase in trading speed ushered in by computerized trading, when combined with the increasing prominence of high-frequency trading, may have eliminated anomalies regardless of how complex the trading rules became.

other). Given this property, if we assume that the linear factor model is the true asset pricing model and that the factors used are true risk factors, then we should expect model performance to increase with the number of factors. In particular we should expect a decrease in both the number and significance of alphas as we sequentially add factors to the model. Our study, however, reveals extreme violations of this property. Furthermore, the violations differ when considering the case of equally weighted versus value-weighted portfolios. The evidence thus clearly indicates that the linear asset pricing model is substantially flawed. Finally, our comparative study of the three alternative models that include a momentum factor shows that the one employing our reconstructed Carhart factor is always the best performer and that, with only one exception, the model employing the Fama–French momentum factor is the worst performer.

It is most likely that we have, in this paper, performed the largest disclosed data mining exercise known to date.[5] However, the goal of this exercise is precisely opposite to the goal of data miners. Whereas data miners perform a similar exercise but disclose only the selected output, here we disclose and analyze the complete output. Indeed, this is what allows us to assess the extent of data mining in the literature. Our analysis identifies many pricing anomalies not previously reported, and one implication is that they were mainly inferred by pure economic reasoning (and not by mining data).

The rest of the paper is organized as follows. In Section II we relate our work to the literature on the EMH and the empirical methods used to test it. In Section III we describe the data used in the paper, as well as the new methodology to test the EMH and to evaluate the performance of the most widely accepted asset pricing models used in empirical research. Section IV is dedicated to testing the semi-strong form of market efficiency in the context of accounting information. In Section V we further explore market efficiency and relate it to the degree of complexity of trading strategies, and Section VI compares the performance of the different asset pricing models used in the study. Section VII is devoted to analyzing the anomalies' most common factors. Finally, Section VIII offers some concluding remarks.

[5]More than 6 terabytes of data were generated and analyzed; this figure includes more than 15 billion regressions and more than 1.3 trillion estimated coefficients. Performing the exercise was crucially aided by the supercomputer Mare Nostrum.

II. Related Literature

In this paper we propose a new methodology to test the Efficient Market Hypothesis (EMH). The new methodology allows to reassess both the status of the hypothesis and the performance of the most widely accepted asset pricing model used in empirical research to test it. Hence, our paper is related to both the MEH and the asset pricing models used in empirical research. First, we place our study in the context of the market efficiency literature and next we relate our method to the debate on asset pricing models in empirical research.[6]

Fama (1970) defines an efficient market as "a market in which prices always "fully reflect" available information". Formally, a market is said to be efficient with respect to an information set, θ_t, if the price 'fully reflects' that information set (Fama, 1970). 'Fully reflects' means that the information is incorporated in the price or, as Malkiel characterizes it, that the price would be unaffected by revealing the information set to all market participants (Malkiel, 1992). Another, more operative, characterization of market efficiency originally provided by Jensen (1978) is: "a market is efficient with respect to information set θ_t if it is impossible to make economic profits by trading on the basis of information set θ_t". The EMH asserts that markets are efficient. The trust on the EMH (or the believe of a high degree of efficiency) is key for understanding key aspects of financial markets. First, it allows us to rely on the market mechanism to achieve economic efficiency. Second, it rationalizes regulation on items such as insider trading, and accounting rules based on "reasonable values", that is, market prices. Finally, it even permeates the legal system that accept those "reasonable values" as the basis for the quantification of damages in legal proceedings. The world would be very different without a high degree of trust on the EMH.

Attending to the type of information included in the Information sets, we have the classical distinction between weak (θ_t only contains past prices), semi-strong (θ_t contains all public information) and strong (θ_t contains both public and private information) forms of market efficiency (Robert (1967), Fama (1970)). Although the three forms have interest on their own and the debate is alive on all of

[6]Our goal here is not to review an extremely large literature, but to place our work in context and to highlight our contributions. For a full review of the literature on the MEH see survey articles such as Lo (11997), Dimson and Mussavian (1998), Lo (2008), Yen and Lee (2008) and Sewell (2011). For a review of the empirical methods used in testing it, for instance, see Campbell et al. (1996) and Cochrane (2005).

them, perhaps the semi-strong and the weak forms are the key ones when assessing the MEH. Indeed, the strong form is not even expected to hold as it would be inconsistent with market equilibrium when information gathering activities are costly and traders are price takers (Grossman and Stiglitz (1980)). This paper is set in the context of the semi-strong form and, consequently, at the center of the MEH controversy.

The literature has used several approaches to test the semi-strong form of market efficiency. Roughly, studies can be classified in two broad categories: the event studies approach, originally introduced by Ball and Brown (1968), and the portfolio approach of Fama and French (1993). Rather than directly testing the hypothesis, this literature has grown adding pieces of evidence, more or less controversial, of failures of the hypothesis: the so called market anomalies. The present paper challenges this approach and proposes a new one. We believe that, by construction, this approach is insufficient to address even the first and more important question: how (semi-strongly) inefficient are markets? Is the failure restricted to a set of isolated cases or is it a broader phenomena? The answer to this question in the current approach would depend on the skills of researchers to identify or not the remaining potential anomalies. Although not guaranteed, there may exist a limit to this search process, but the question will remain unanswered all along this lengthy process. In summary, while the standard approach provides evidence in favor of the rejection of the MEH, it is silent on how inefficient markets indeed are. Rather than starting from a potential source of failure, in our approach we start with a large enough set of potential sources of failures and let the data confirm the severity of the problem. More formally, each contribution in the traditional approach uses an information set, θ_t, that includes a single variable or firm characteristic, and then test for the existence of anomalous returns associated to that characteristic. In our approach, we consider a multidimensional information set that includes all possible firm characteristics (or, at least, a large enough set that allows for inferences) and then jointly tests for the existence of anomalous pricing associated to that broad area of public information. In the present study we apply the new approach to the case of public accounting information, which constitutes the richest source of public information about firms, and use the portfolio approach for the test of the MEH.

It is well known since Jensen (1978) that the MEH can not be tested in isolation, but that every test is a joint test of the hypothesis and the particular asset pricing model used in the test. For a long time, the CAPM (single factor model) placed that role in empirical tests, besides strong theoretical arguments in favor of more sources of common variation in asset returns being needed, beyond the market portfolio (proxied in empirical applications by some broad portfolio). Fama and French (1993) provided what at the time seemed to be the required set of extra factors. The so-called Fama and French 3-factor model, has become "one of the most popular multifactor model that now dominate empirical research" (Cochrane (2005)) and, hence, the dominating model that is commonly used in tests of the semi-strong form of the MEH. In recent years, the model has been enlarged, adding to the market, size and value factors, some new factors such as the momentum factor (Carhart (1997)) or the liquidity factor (Pastor and Stambaugh (2003)). Irrespectively of the controversy surrounding these factors as to whether they represent risk or something else, the fact is that for many years and currently anomalies are identified using the machinery developed by Fama and French.

Our study represents a unique opportunity to evaluate the performance of the many versions of these linear factor models, what constitutes the second motivation for the present paper. In the traditional approach, a given source of potential mispricing becomes an anomaly when it survives the factors. That is, to enter the universe of anomalies we require that the mispricing associated to the anomaly is not explained by the market, size, value, and liquidity premiums. Further, we also require that the profits associated to the anomaly are not explained by momentum in asset returns. But this unidirectional inspection, where factors are filters for a given potential anomaly, does not allow us to see how the factors themselves affect the possible generation of anomalies themselves. This is important as all multifactor models are proxies for an unknown set of "true factors". It may be the case that the addition or subtraction of some of the factors we routinely use in empirical research artificially enlarges the set of anomalies that can be obtained in a given data set. Given the extreme large scale of our study, we can perfectly observe how these interactions operate in the data and identify possible biases associated to some particular factor.

III. Data and Methodology

In this section we describe the data and the methodology used in the present paper to assess market efficiency and to compare different asset pricing models.

A. Data

The inputs in this paper are accounting and security prices data. All data comes from the merged CRSP-Compustat database provided by Wharton Research Data Services (WRDS). We use Compustat data for fiscal years from 1962 to 2009 and CRSP data from July 1963 till June 2010. The filtering and cleaning of the data mainly follows Fama and French (2008). In Appendix *A* we provide a detailed account of the data treatment and highlight some controversial issues that have been missed in the literature.

Compustat includes a total of 331 variables for each company. To ensure that our analysis does not miss known accounting based anomalies, we use an enlarged version of Compustat where some extra variables are included. The *enlarged* Compustat dataset adds to *original* Compustat two type of variables. First, following the large literature on accounting-based anomalies, we add to the Compustat data set entries to capture anomalies related to firm profitability, accruals, net stock issuance, asset growth, etc. Second, given that some variables are only relevant when expressed relative to some proxy for the size of the company, we also add columns normalized my market size. In Appendix *B* we provide the definition of the new variables and the details of its explicit construction.

B. Methodology

We analyze accounting based pricing anomalies using the standard portfolio methodology. Based on some accounting information of interest, in this methodology we create dynamic portfolios highly sensitive to that information, and then test for their mispricing in the context of well-known asset pricing models. The main innovation of the analysis here is that we do not analyze a given accounting information and the portfolio that may fail to reflect that information, but a large universe of accounting information and the set of portfolios based on that information that can be potentially mispriced.

To properly asses the extent of market efficiency in the context of accounting information, we create control portfolios. In the next subsection we give the details of the construction of the *Informed* or accounting-based portfolios, as well as the *Spurious Control* and *FF25 Control* portfolios. Finally, in the last part of this section we give the details of the different asset pricing models used in the present paper.

B.1. Portfolio construction

We start with the construction of the *Informed* portfolios. We create portfolios based on randomly selected arithmetic operations (additions, subtractions, multiplications and divisions) on 3 to up to 10 randomly selected Compustat variables or items. Every one of these combinations (firm characteristic) constitutes a portfolio sorting variable. For example, one sorting variable could be Assets−Cash+Debt, where subtraction and addition are randomly chosen as well as variables (Assets, Cash and Debt) themselves. For a fixed sorting variable, the construction is as follows. Using the accounting information of year $t-1$, we create two portfolios. One includes all companies in the top 10% of the sorting variable, and the other one the bottom 10% companies from the same sort. Companies for which the sorting variable is more than five standard deviations from the annual mean are not included in the portfolios. We create the monthly portfolio return series by value-weighting or equally-weighting the returns of the companies in the portfolio from July of year t to June of year $t+1$. We follow Fama and French (1993) in choosing July of year t as the portfolio formation date to ensure that the accounting information for the final year ending in year $t-1$ is available to the market. The treatment of missing returns and of returns from delisted firms is as described by Beaver et al. (2007). Portfolios are reformed annually. In the main analysis the monthly portfolio returns range from July 1963 to June 2010. In the sub-sample analysis we use 1990 as the break year. Hence in this analysis, the monthly portfolio returns run from July 1963 to June 1990 (Early subsample) and from July 1990 to June 2010 (Late subsample). This process is replicated for every sorting variable.

As previously stated, we also construct two type of control portfolios in this paper. The *Spurious Control* portfolios are constructed in a similar fashion, but just substituting each company in each of the *Informed* portfolios with a randomly selected company from the whole universe of firms.The *FF25 Control* portfolios are constructed by substituting each company in each of the *Informed* portfolios by randomly selected company with a randomly selected company positioned in the corresponding cell of the 5 x 5 Fama–French matrix, recalculated annually. The main analysis is done using the whole CRSP universe of firms. But we notice that in the robustness section, and in order to avoid suspicions on our results being driven by the presence of microcap stocks, we replicate the analysis separately in each of the 25 cells of the 5 x 5 Fama and French matrix.

C. Asset Pricing

We use five different models to test for pricing anomalies. Given its prevalence in the history of asset pricing we include the Market and the Fama and French three factors models. We also consider three different specifications of the four factor model that adds a momentum factor to the Fama and French three factor model. The reason for analyzing these three alternative specifications of the momentum factor is two-fold. First, there is a controversy as to whether or not Momentum should be considered a risk factor. Second, there are three alternative specifications and academics routinely pick one of them without considering if the choice itself introduces some specific bias in the analysis. The most widely used specification of the momentum factor is the one available in Kenneth French's webpage. The second one is the one provided by Mike Carhart (available by personal request). This factor, even though updated by Mike Crahart, is not revised backwardly, which means that some of the past returns are not obtainable given the updated and corrected set of companies and returns reported in CRSP. The third version of Momentum factor is the factor that we reconstruct by ourself following Carhart (1997) methodology with the updated version of CRSP. The reality is that the time series of the returns of the three factors are quite different. Regression of Kenneth French's momentum factor on Carhart's momentum factor yields R^2 of 55% while in a regression of original Carhart's factor on the factor that we construct we obtain R^2 of 88%. Using one or the other may have a non-trivial impact on the assessment of pricing anomalies. In this paper we wish to shed some light on this issue. More specifically, in this paper we consider the following asset pricing models:

- The Market Model (M):

  $$R_{it} - R_f = \alpha_i + \beta_{i1} EXM_t + \varepsilon_{it}, \tag{1}$$

  where R_{it} is the return of a portfolio i in month t, R_f is one-month Treasury bill rate and EXM_t is the excess return of the market portfolio in month t, proxied with Fama and French market factor.

- Fama and French 3 factor model (FF3):

$$R_{it} - R_f = \alpha_i + \beta_{i1} EXM_t + \beta_{i2} HML_t + \beta_{i3} SMB_t + \varepsilon_{it}, \tag{2}$$

  where, HML_t and SMB_t are Fama and French value and size factors, obtained from Kenneth French's webpage.

- Fama and French 4 factor model (FF4):

$$R_{it} - R_f = \alpha_i + \beta_{i1} EXM_t + \beta_{i2} HML_t + \beta_{i3} SMB_t + \beta_{i4} MOM_t + \varepsilon_{it}, \tag{3}$$

  where, MOM_t is the Fama and French momentum factor, obtained from Kenneth French's webpage.

- Four Factor Model with Carhart's original factor (C1):

$$R_{it} - R_f = \alpha_i + \beta_{i1} EXM_t + \beta_{i2} HML_t + \beta_{i3} SMB_t + \beta_{i4} MOMC1_t + \varepsilon_{it}, \tag{4}$$

  where, $MOMC1_t$ is Carhart's original factor obtained from XXX.

- Four Factor Model with Carhart's reconstructed factor (C1):

$$R_{it} - R_f = \alpha_i + \beta_{i1} EXM_t + \beta_{i2} HML_t + \beta_{i3} SMB_t + \beta_{i4} MOMC2_t + \varepsilon_{it}, \tag{5}$$

  where, $MOMC2_t$ is Carhart's reconstructed factor following Carhart (1997) methodology.

D. Summary statistics

In Table 1 we report some of the main characteristics of the present study. We start with a merged Compustat-CRSP data set that covers 12,228 firms and 331 accounting related variables for each firm. The original Compustat data set is enlarged with 356 variables. Out of all possible combinations of 3 to 10 extended Compustat columns we randomly construct 44.8 million sorting variables. In the portfolio analysis we use two methods of portfolio weighting: equally and value weighted portfolios.

The portfolios belong to one of three types: *Informed*, *Spurious Control* and *FF25 Control*. Three type of samples have been used: the Whole sample that covers the period July 1963 to June 2010, the Early sub-sample that covers the period July 1963 to June 1990 and the Late sub-sample that goes from July 1990 to June 2010. The experiment are conducted among all stock and inside each of 25 Fama and French portfolios sorted on size and book to market value. Finally, each portfolio has been assessed in the context of 5 alternative asset pricing models. This results in a total of 15.6 billion regressions and 1.3 trillion parameter estimated (regression coefficients and corresponding t-statistics, standard errors, p-values and R^2).

[Insert Table 1: Summary Statistics of the Experiment]

IV. Market Efficiency

For simplicity in the presentation we start assuming the Fama and French 4-factor model (F4) as the true asset pricing model that explains the cross section of asset returns. We advance that all results reported in this section hold when using any of the alternative models.[7]. Also for simplicity in the exposition we only report the case of value weighted portfolios. We start analyzing the extent of miss-pricing in our *Informed* portfolios (those that are created using the Extended-Compustat data set) using the whole sample, July 1963 to June 2010. In Figure 1 we plot the leading results of the experiment. Panel A plots the density function of the alphas, and Panel B the density function of the absolute value of the alphas which are statistically significant at the 99% confidence level. Table 1 reports some statistics of the previous two distributions. The results obtained are impressive. The distribution of alphas has an average of 0.049% per month and a variance of $0.16\%^2$. We obtain more that 2.7 million alphas statistically significant at the 99% confidence level and that is more than 6 times more than one would expect by pure chance. Some alphas are as large as 26.52% per month, which amounts to more that 300% annually. Even more impressive, we obtain 164 significant alphas at the 99% level which are larger than 10% monthly in absolute value.

[Insert Figure 1: Alphas of the *Informed* Portfolios Using the F4 Model in the Whole Sample]

[7]The complete set of results is available in the authors' webpages.

[Insert Table 2: Summary statistics of Alphas of the *Informed* using the F4 Model in the Whole Sample]

These results already point at a large extent of market inefficiency associated to public accounting information, in the context of the most widely accepted asset pricing model. But, as argued in the introduction, the results are not final as they can be driven by well-known anomalies or noise in the data not related to accounting information. To properly asses the extend of mispricing we need to control for these factors.

A. Informed versus control portfolios

The first issue is whether the alphas obtained in the *Informed* portfolios arise from pure noise in security prices that is picked as anomalous pricing in the context of the Fama French 4 factor model and are just statistical artifacts. To address this issue we compare the distribution of alphas of the accounting based portfolios against the distribution of alphas obtained with the *Spurious Control* portfolios. These *Spurious Control* portfolios have the same number of securities and run for the same sample period as the *Informed* portfolios. But, while the *Informed* are constructed based on public accounting information, the *Spurious Control* portfolios are non informative, as every security included in the portfolio is randomly selected. Figure 2 and Table 3 report the results. The results are conclusive: the mispricing arising from public accounting information are not explained by anomalous mispricing in non-informative *Spurious Control* portfolios. Panel A of Figure 2 plots the density of the alphas of the *Informed* versus *Spurious Control* portfolios. The distribution of the alphas of the *Informed* portfolios is much heavier in the tails and much lighter in the center than the distribution of the alphas in the *Spurious Control* experiment. The differences in the distributions are even more remarkable when we look at the distributions of the absolute value of the statistically significant alphas at the 99% confidence level in Panel B of Figure 2. The distribution of the alphas of the *Spurious Control* portfolios is completely dominated by the distribution of the alphas of *Informed* portfolios. These results are statistically verified using both parametric and non-parametric tests of differences in distributions. In particular, as before, the mean and the variance of the alphas of the *Informed* portfolios are larger than the ones of the *Spurious Control* portfolios (the t-test for differences in the mean has a t-statistics of -239 and a

p-value of 0, and the F-test for differences in the variance yields a F-Statistic of 1.37 and a p-value of 0). The non-parametric Kolmogorov-Smirnov test has a $K-S$-statistic of 0.054 and a p-value of 0. The differences in distributions are also economically significant: while the largest mispricing in the control experiment is 10.17% monthly, the use of accounting information can deliver almost 3 time more mispricing (up to 26.52% monthly).

[Insert Figure 2: Alphas of the *Informed* vs alphas of *Spurious Control* portfolios]

[Insert Table 3: Alphas of the *Informed* vs alphas of *Spurious Control* portfolios]

The Second issue is whether the alphas obtained in the *Informed* portfolios are associated to known pricing anomalies, not necessarily related to accounting information. To address this issue we balance the distribution of alphas of the accounting based portfolios against the distribution of alphas obtained with the *FF25 Control* portfolios. As stated previously, these *FF25 Control* portfolios are constructed by substituting each firm in an *Informed* portfolio by a firm similar in size and value (a firm that belongs to the same cell as the original one in the 5x5 Fama and French sort of firms by size and value). The results are conclusive: the mispricing arising from public accounting information are not explained by known anomalies such as size and value.

Pannel A of Figure 3 plots the density of the alphas of the *Informed* versus *FF25 Control* portfolios. The distribution of the alphas of the *Informed* portfolios is much heavier in the tails and lighter in the center than the distribution of the alphas in the *FF25 Control* experiment. The differences in the distributions are even more remarkable when we look at the distributions of the absolute value of the statistically significant alphas al the 99% confidence level. The distribution of the *FF25 Control* experiment is completely dominated by the distribution of the *Informed* experiment as presented in Panel B of Figure 3. These results are statistically verified using both parametric and non-parametric tests of differences in distributions. In particular, the average of the alphas is significantly larger in the *Informed* portfolios than in the *FF25 Control* portfolios (the t-test for differences in the mean has a t-statistics of 56.2 and a p-value of 0) and the variance is almost double and also highly statistically significant (the F-test for differences in the variance yields an F-Statistic of 1.74 and a p-value of 0). While the result on the mean is not relevant in the present context (as in our construction the long sort portfolio can arbitrarily yield positive or negative alphas), the result on the variance is extremely relevant. On the other hand, the non-parametric Kolmogorov-Smirnov test has a $K-S$-statistic of 0.059 and a p-value of 0. The differences in distributions are also economically significant: while the largest mispricing in the control experiment is 10.13% monthly, the use of accounting information can deliver almost three time more mispricing (up to 26.52% monthly).

[Insert Figure 3: Alphas in the *Informed* versus *FF25 Control* portfolios]

[Insert Table 4: Summary statistics of Alphas in the *Informed* versus *FF25 Control* portfoliso]

Finally, the third issue is whether the alphas obtained in the *Informed* portfolios are associated to known accounting based anomalies or to new phenomena. To address this point we repeat the same experiment but excluding the variables that explicitly capture known anomalies. In Figure 4 and Table 5 we report the results. The results presented so far stay qualitatively the same.

[Insert Figure 4: Alphas in the *Informed* Experiment with Compustat vs. Extended Compustat]

[Insert Table 5: Summary statistics of Alphas in the *Informed* Experiment with Compustat vs. Extended Compustat]

In light of the evidence reported in this section, we can state two very important conclusions. First, there is clear evidence that in more than 30 years of research on the subject, the academia has been only able to point at a very small subset of accounting-based pricing anomalies. Second, the failure of the Market Efficient Hypothesis is not anymore a matter of some isolated cases of rejection. The present study makes very clear that just focussing on a single potential source of failure (the case of public accounting information) is enough to conclude that the failure of the MEH is a large and broad phenomena. Needless to say, that these two conclusions are conditional on the assumption that the standard asset pricing models are the true models for asset valuation, an assumption that, as we argue in Section VI, seems to be in contradiction with the evidence.

B. Robustness checks

In this section we practice several robustness checks on the previous results. We focus on three checks: elimination of redundant firm characteristics, subsample robustness and robustness to the exclusion of microcap stocks. The first robustness check is specific to our methodology, and the other two are standard in the anomalies literature.

B.1. Redundant firm characteristics

Our methodology uses a random devise for the generation of firm characteristics. Such a devise is bound to generate firm characteristics that relate to the same source of accounting information. For instance, some compustat items are quantitatively very small and when added or subtracted to some arithmetic combinations do not change significantly the sorting of firms included in the portfolios. I n these cases, the original combination and the one obtained adding or substracting the above mentioned compustat item essentially represent the same firm characteristic. The previous construction suffer from the presence of these redundant firm characteristics. Their presence affects the distribution of alphas in two ways. On the one hand, when the redundancy is related to mispriced accounting information (signifficant alphas), the presence of redundant characteristics artificially increases the mass of mispriced *informed* portfolios both in absolute terms and relative to the control portfolios; but, on the other hand, when the redundancy is related to rightly priced accounting information, the presence of redundant characteristics operates in the opposite direction. *Ex ante* we do not know which case dominates and, hence, we do not know how their presence bias our results. But, given that any redundant alpha (significant or not) should not be in our study in the first place, we should verify that all our results survive the exclusion of these redundant firm characteristics.

To eliminate redundant firm characteristics, we examine the pairwise relationship between firm characteristics and exclude all those sorting variables that are even marginally related to each other. More especifically, we exclude all sorting variables that are related to some other sorting variable at 90% confidence interval independently of the R^2 of the regression. This is indeed a very restrictive criteria since some variables could be only marginally related. We show that our results do survive this test and that our results stay qualitatively the same. This is statistically verified using both parametric and non-parametric tests of differences in distributions. In particular, as before, the mean and the variance of the alphas of the *Informed* portfolios are larger than the ones of the *Spurious Control* portfolios (the t-test for differences in the mean has a t-statistics of -170 and a p-value of 0, and the F-test for differences in the variance yields a F-Statistic of 1.35 and a p-value of 0). The non-parametric Kolmogorov-Smirnov test has a $K-S$-statistic of 0.056 and a p-value of 0. The differences in distributions are also economically significant: while the largest mispricing in the control experiment is 9.56% monthly, the use of accounting information can deliver almost 3 time more mispricing (up to 26.52% monthly).

B.2. Subsample analysis

Our sample covers the period 1963-2010. We should not expect an anomaly to survive such a long period of more than 45 years. For this reason, and in order to properly assess robustness to subsample analysis, we have split the sample in halves and quarters. More specifically, we first split the data in the two subsamples (the Early and Late subsamples) and identify the set of mispriced portfolios that survive both subsamples. Removing those anomalies that are nonrobust to this sample split does not change our results qualitatively.

We verify this using both parametric and non-parametric tests of differences in distributions. In particular, as before, the mean and the variance of the alphas of the *Informed* portfolios are larger than the ones of the *Spurious Control* portfolios (the t-test for differences in the mean has a t-statistics of -120 and a p-value of 0, and the F-test for differences in the variance yields a F-Statistic of 1.33 and a p-value of 0). The non-parametric Kolmogorov-Smirnov test has a $K-S$-statistic of 0.048 and a p-value of 0. The differences in distributions are also economically significant: while the largest mispricing in the control experiment is 10.01% monthly, the use of accounting information can deliver almost 3 time more mispricing (up to 26.52% monthly).

B.3. Microcap stocks

Fama and French (2008) argues that most of the anomalies identified in the data are associated to microcap stocks. Further, he sustains that these microcaps are special in many dimensions and should not be taken into account in the anomalies debate. He proposes to exclude these stocks in all empirical studies related to the EMH. Although Fama and French's position is hard to swallow, as it implies excluding more than 60% of the stocks on average and indirectly acknowledging that the Fama and French machinery is of no much help in the debate surrounding the EMH, we find necessary verifying that our results are not driven by the presence of microcaps.

At this point, it is important to notice that our methodology already (at least partially) accounts for this. In particular, the firms in our *Spurious Control* portfolios are randomly selected and, given the preeminence of microcaps in the sample, the alphas of such portfolios already contains all the mispricing associated to the presence of these small firms. But in any case, in the present section we want to spell any remaining suspicion on the role played by microcaps on our results by means of an even stronger test. The test consists on replicating the previous analysis 'within' each of the 25 cells of the 5 x 5 Fama and French matrix. The results of the test are conclusive: all our results survive when the analysis is performed separately in each cell. This means that the mispricing of the informed portfolios not only survives in the cells that exclude small stocks, but also in the cells where that include the smallest firms in the economy. The first results proves that our results survive the exclusion of microcaps; the second, that the mispricing of accounting information is present in microcaps too, precisely the stocks Fama is reluctant to include in the analysis.

V. Complexity and Price Efficiency

In this section we tackle a different angle of price efficiency. The aim is to relate the degree of market efficiency to the level of complexity of trading strategies. *Ex ante* we should expect that outstanding performance must be associated to the more highly sophisticated trading strategies. This can be easily rationalized both in a world of costly information acquisition, as well as in a world of rational inattention. In these two scenarios we should expect that abnormal returns are hard to get when trading on simple strategies (say, trading based on sales or profits of companies) and relatively easier to get when trading on sophisticated rules (say trading based on accruals of small value stocks). The design of our experiment allows us to explicitly test for this hypothesis.

Using the output of our regular experiment, we analyze the extent of mispricing as a function of the number of the *extended* Compustat items used in the arithmetic operations. In Table 5 we report the results. As expected, as we increase the number of variables (*extended* Compustat columns) involved in the arithmetic operations, the variance of the distribution of alphas steadily increases. Furthermore, the number of alphas statistically significant at the 99% confidence level also increases. All the variances are significantly different with a p-value of 0. The results are true for both equally and value weighted portfolios. Hence, the evidence corroborates the expected outcome in a world of costly information acquisition or rational inattention.

[Insert Table 6: Anomalies and Degree of Complexity]

The previous insight can also be tested using a historical perspective. Have markets become more efficient over time? At a theoretical level, the question does not have a definitive answer. Arguably, the level of sophistication of traders and markets has been increasing steadily over the last decades. This evolution involves two countervailing forces. On the one hand, according to the previous argument, more sophisticated traders using more complex strategies would imply that market efficiency must have increased over time. But, on the other hand, markets themselves have become more complex. For instance, the amount and complexity of information that companies disclose is much larger today than 50 years ago. So it is the tax code and the type of securities issued by companies. This points at exactly the opposite direction. More information and of a more complex nature or more costly to process is available and, hence, potentially more likely to be missed by investors. This can potentially result in less informationally efficient markets. It is an empirical issue then to establish what has increased in complexity faster, markets or investors' trading strategies. To partially address this issue we have split our sample in two subsamples. The Early subsample covers the period July 1963 to June 1990 and the Late subsample, the period July 1990 to June 2010. In Figure 5 and Table 7 we report the results of the exercise. As we can see in Panel C of Figure 5, markets have become more efficient over time. Moreover Panels A and C show that our main conjecture about the market efficiency from Section 4 are robust irrespective of the sample period.[8]

[8]Notice that in our methodology we randomly the number 3 to 10 from a uniform distribution. Hence we have almost the same number of firm characteristics in each group and that given our large sample size these differences in the number of firm characteristics do not affect the mean and the variance reported in Table 7.

[Insert Figure 5: Subsample Analysis]

[Insert Table 7: Summary statistics of Subsample Analysis]

VI. Comparing Factor Models

Our experiment provides also an excellent setting to compare the performance of different asset pricing models. In this section we compare the distributions of alphas of the accounting based portfolios across the 5 different asset pricing models defined in equations (1) to (5), for both equally and value weighted portfolios. Our approach is minimalist, as we only focus on alphas and do not enter into other measures of model performance such as R^2 or the Gibbons-Ross-Shanken (Gibbons et al. (1989)) performance measure. But still our exercise is revealing of some failures of the most widely accepted asset pricing models in assessing pricing anomalies. In particular, the analysis highlights that all the models are misspecified and that the momentum factor does not seem to belong to the set of legitimate risk factors. Given that we consider models that range from one to 4 factors, we should expect some regularities, such as a decrease in the absolute value of the average alpha and the variance of alpha as we increase the number of factors included in the asset pricing model. On the other hand, *ex ante* we do not know which of the three different four factor models should perform best. That is, we should expect the following preference ordering for the models (1) to (5):

$$C2 \simeq^? C1 \simeq^? F4 \succeq F3 \succeq M,$$

but, unfortunately, the evidence clearly rejects this conjecture.

In Table 8 we report the mean and variance of alphas and their associated t, F and $K-S$ statistics for pairwise comparisons across the 5 different asset pricing models: the market Model (M), the Fama and French 3 factor Model (F3), the Fama and French 4 factor model (F4) that includes Fama and French momentum factor, and the Carhart based 4 factor models, one including the original Carhart momentum factor (C1) and the other, the reconstructed Carhart momentum factor (C2). Both the parametric and the non-parametric tests reveal significant differences in parameters and distributions in all pairwise comparisons.

[Insert Table 8: Comparison of Asset Pricing Models]

Using the variance as the metric for model performance we obtain the following ordering:

- In the case of equally weighted portfolios, the market model (M) delivers the largest variance, and the four factor with the reconstructed Carhart momentum factor (C2) the smallest. The complete ordering is as follows:

$$C2 \succeq C1 \succeq F3 \succeq F4 \succeq M.$$

  In this case the conjecture is rejected because the Fama and French 3 factor model performs better than the Fama and French 4 factor model. The result also suggests that the model using the reconstructed Carhart momentum factor (C2) is the best performer among the there four factor models analyzed.

- In the case of value weighted portfolios, the Fama and French 3 factor model (F3) delivers the largest mean alpha, and the Market model (M) the smallest. The complete ordering is as follows:

$$C2 \succeq F4 \succeq M \succeq C1 \succeq F3.$$

  In this case the conjecture is rejected because the Market model (M) performs better than two of the models that include more factors (C1 and F3). As before, the model with the Carhart reconstructed momentum factor is the best performer.

Using the mean of the absolute value of alphas as the metric for model performance we obtain:

- In the case of equally weighed portfolios, the Fama and French three factor model (F3) delivers the largest mean, and the market model (M), the smallest. The complete ordering is as follows:

$$M \succeq C2 \succeq C1 \succeq F4 \succeq F3.$$

  Again, in this case the conjecture fails to hold because the one factor model (M) performs better than all the others. In this case, the Carhart based momentum factor loses its superiority.

- In the case of value weighted portfolios, the Fama and French 4 factor model (F4) delivers the largest mean, and the Fama and French 3 factor model (F3), the smallest. The complete ordering is as follows:

$$F3 \succeq C2 \succeq C1 \succeq M \succeq F4.$$

  The irregularities in this case are too many to list here.

The picture that arises from the previous exercise is quite negative. In particular, we have identified many cases in which the addition of an extra factor in the asset pricing model enlarges, rather than reduces, the extent of mispricing. On a more dramatic tone and in light of this result, we wonder how to interpret the existing literature on anomalies that builds on the Fama and French paradigm and uses one of the previous model specifications. Are they anomalies of just model induced mispricing that arises from the inclusion or exclusion of some? On the other hand, the three alternative versions of the momentum factor widely used in the literature seem to capture very different things, what raises serious concerns on momentum itself as a risk factor and the way we control for it in asset pricing. The only regularity that we find in the present study regarding the three models that include momentum is that the model including the reconstructed Carhart factor is always the best performer, and that in all but one case the model with the Fama and French momentum factor is the worse performer

VII. New Anomalies?

As stated in the introduction, this paper has an explicit goal of not practicing data mining. Also, in light of the previous results, we wonder if there is a point on addressing new anomalies in the context of faulty asset pricing models irrespectively of whether these come from economic reasoning or data mining. But in any case, given the output produced in the present experiment we could not resist the temptation to disclose the main common accounting variables behind the significant alphas. The analysis that follows is just illustrative and must be taken with caution. We are associating significant alphas arising from faulty models and without imposing economic reasoning behind it.

To perform this exercise we start grouping the Compustat variables into broad areas of accounting. Then we compute the percentage of times that any of the items in an given accounting area enter in arithmetic combinations used in portfolios that exhibit statistically significant alphas (at the 99% confidence level). The results are reported in Figure 6.

[Insert Figure 6: Accounting Factors Behind Anomalies]

Figure 6 reveals that the most dominating factors in the mispricing are firms': Liabilities, Property, Depreciation, Taxes, Earnings and Income. Some of these variables have been pointed out in the literature as a source of mispricing, but some others are new. In any case, given the concerns raised before, we rather stop at this point than continue elaborating on doubtful output.

VIII. Concluding Remarks

In this paper we introduce a new methodology to test market efficiency and to compare the performance of alternative asset pricing models. The methodology is used to test the semi-strong form of market efficiency in the context of publicly available accounting information. The analysis reveals a large degree of market inefficiency not documented in previous literature and not explained by known pricing anomalies or randomness in asset prices. Hence, the first application of the new methodology is already quite devastating for the Market Efficiency Hypothesis. Furthermore we also provide evidence that shows that the traditional asset pricing model used in empirical research fail to meet minimum requirements of true linear asset pricing models.

As a premier on the application of the new methodology, the analysis in this paper is minimalist in many dimensions. First, the analysis of market efficiency is single out to the case of public accounting information. But there are many other sources of public information that can be tested. Second, the comparison of the performance of the different asset pricing models is done using alpha as the metric for performance. We do not study other variables, such as R^2. These two are extension we plan to analyze in future research.

References

Ball, Ray and Philip Brown, 1968, An empirical evaluation of accounting income numbers, *Journal of Accounting Research*, 6 (2), 159-178.

Beaver, William, Maureen McNichols, and Richard Price, 2007, Delisting Returns and their Effect on Accounting-based, Market Anomalies, *Journal of Accounting and Economics*, 43, 341-368.

Carhart, Mark M., 1997, On Persistence in Mutual Fund Performance, *The Journal of Finance*, 52, 57-82.

Cochrane, John H., 2005, *Asset Pricing*, Princeton University Press, New Jersey.

Fama, Eugene F., 1970, Efficient capital markets: A review of theory and empirical work, *The Journal of Finance*, 25 (2), 383-417.

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, 2008, Dissecting Anomalies, *The Journal of Finance*, Vol. LXIII, No. 4, 1653-1678.

Gibbons, Michael, Stephen Ross, and Jay Shanken, 1989, A test of the Efficiency of a Given Portfolio, *Econometrica* 57, 1121-1152.

Grossman, Sanford J. and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *The American Economic Review*, 70 (3), 393-408.

Jensen, Michael C., 1978, Some anomalous evidence regarding market efficiency, *Journal of Financial Economics*, 6 (2- 3), 95-101. McLean, R. David and Pontiff, Jeffrey E., 2012, Does Academic Research Destroy Stock Return Predictability?. AFFI/EUROFIDAI, Paris December 2012 Finance Meetings Paper. Available at SSRN: http://ssrn.com/abstract=2156623.

Malkiel, Burton G., 1992, Efficient market hypothesis, in P. Newman, M. Milgate and J. Eatwell (eds.), *New Palgrave Dictionary of Money and Finance*, Macmillan, London.

Malkiel, Burton G., 2005, Reflections on the Efficient Market Hypothesis: 30 Years Later, *The Financial Review*, 40 (1), 1-9.

Pastor, Lubos and Robert F. Stambaugh, 2003, Liquidity Risk and Expected Stock Returns, *Journal of Political Economy*, 111, 642-685.

Richardson, Scott, Irem Tuna and Peter Wysocki, 2010, Accounting Anomalies and Fundamental Analysis: A Review of Recent Research Advances, *Journal of Accounting and Economics*, Vol. 50, Issues 2-3, 410-454.

Roberts, H. (1967), Statistical versus clinical prediction of the stock market, Unpublished manuscript.

Sadka, Ronnie, 2010, Liquidity Risk and the Cross-section of Hedge-fund Returns, *Journal of Financial Economics*, 98, 54-71.

Appendix

A. Data Filtering and Cleansing

Following Fama and French (2008) we impose the following assumptions in the treatment of the data:

- We include only ordinary common shares that were trading on the New York Stock Exchange (NYSE), the American Stock Exchange (Amex), and NASDAQ. (SHRCD=10 or 11, EXCHCD=1,2 or 3)
- We exlude observations from Compustat that have a missing entry for FYEAR
- We exclude financial firms (Standard Industrial Classification codes between 6000 and 6999)
- We exclude firms with negative book equity at $t-1$.
- We exclude firms until they have been in Compustat for two years; this reduces the survival bias inherent in the way Compustat adds firms to its sample (Banz and Breen (1986))
- When faced with the duplicate entry in Compustat we keep the observations with the value off variable linkprim="C", which means that we keep observations assigned by CRSP to resolve ranges of overlapping or missing primary markers from Compustat in order to produce one primary security throughout the company history.
- The treatment of missing returns and of returns from delisted firms is as described by Beaver et al. (2007):
  - We calculate returns as as RET when RET is available and DLRET is not available
  - We calculate returns as as DLRET when RET is not available and DLRET is available
  - We calculate returns as as ((1+RET)(1+DLRET)-1) when RET is available and DLRET is available
  - We exclude firms when both RET and DLRET are not available

B. The Enlarged-Compustat data set

Table BI: Definition of Variables

In this table we define the list of firm characteristics that we use in our study besides those present in the Compustat database. The first column reports the name of the firm characteristic while the second column gives a description on how we construct this new the firm characteristic using variables from Compustat. All variables that we use come from Compustat database except PRC and SHROUT that are from CRSP.

Variable name	Variable expressed in terms of Compustat and CRSP variable names	Reference
BVFF	$(\mathrm{AT})-(\mathrm{LT})+(\mathrm{TXDITC})-$ $[(\mathrm{PSTKL})$ or (PSTKRV) or $(\mathrm{PSTK})]$	Fama and French (2008)
MVFF	$\mathrm{SHROUT}\times\lvert\mathrm{PRC}\rvert$ (end of December)	Fama and French (2008)
MVB	$\mathrm{CSHO}\times\mathrm{PRCC}\_F$	Bernhard et al. (1996)
extnd_comp_OWK	$(\mathrm{ACT})-(\mathrm{CHE})-(\mathrm{LCT})+(\mathrm{DLC})$	Fama and French (2008), Francis et al. (2003)
extnd_comp_TCA	$\text{extnd\_comp\_OWK}_t-\text{extnd\_comp\_OWK}_{t-1}$	Fama and French (2008), Francis et al. (2003)
extnd_comp_TAC	$\text{extnd\_comp\_TCA}-\mathrm{DP}$	Francis et al. (2003)
BMV	$\frac{\mathrm{BVFF}}{\mathrm{MVFF}}$	
extnd_comp_BTM	$\frac{\mathrm{CEQ}}{\mathrm{MVB}}$	
extnd_comp_BTMFF	$\ln(\mathrm{BMV})$	Fama and French (2008)
extnd_comp_BTMB	$\ln(\text{extnd\_comp\_BTM})$	Bernhard et al. (1996)
extnd_comp_NSIFF	$\ln\left(\frac{[\mathrm{CSHO}\times\mathrm{ADJEX\_C}]_t}{[\mathrm{CSHO}\times\mathrm{ADJEX\_C}]_{t-1}}\right)$	Fama and French (2008)
extnd_comp_TCAFF	$\frac{\left[\frac{\mathrm{OWK}}{\mathrm{CSHO}\times ADJEX\_C}\right]_{t-}-\left[\frac{\mathrm{OWK}}{\mathrm{CSHO}\times ADJEX\_C}\right]_{t-1}}{\left[\frac{\mathrm{BVFF}}{\mathrm{CSHO}\times\mathrm{ADJEX\_C}}\right]_t}$	Fama and French (2008)
extnd_comp_TCAFF1	$\frac{[\mathrm{OWK}]_t-[\mathrm{OWK}]_{t-1}}{[\mathrm{AT}]_t}$	Francis et al. (2003)
extnd_comp_TCAFF2	$\frac{[\mathrm{OWK}]_t-[\mathrm{OWK}]_{t-1}-[\mathrm{DP}]_t}{[\mathrm{AT}]_t}$	Francis et al. (2003)
extnd_comp_TAH	$\frac{[\mathrm{OWK}]_t-[\mathrm{OWK}]_{t-1}+[\mathrm{TXP}]_t-[\mathrm{TXP}]_{t-1}-[\mathrm{DP}]_t}{[\mathrm{AT}]_t}$	Hirshleifer et al. (2004)
extnd_comp_AGFF	$\ln\left(\frac{[\mathrm{AT}]_t}{[\mathrm{AT}]_t-1})\right)-[\mathrm{NSIFF}]_t$	Fama and French (2008)
extend_comp_PFF	$\frac{\mathrm{IB}-\mathrm{DVP}+\mathrm{TXDI}}{\mathrm{BVFF}}$	Fama and French (2008)
extend_comp_PF1	$\frac{\mathrm{IB}}{\mathrm{MVFF}}$	Francis et al. (2003)
extend_comp_PB	$\frac{\mathrm{EPSPX}}{\mathrm{BVFF}}$	Bernhard et al. (1996)
extend_comp_PF2	$\frac{\mathrm{IB}-\mathrm{TCA}+\mathrm{DP}}{\mathrm{MVFF}}$	Francis et al. (2003)
extend_comp_PH	$\frac{[\mathrm{OIADP}]_t}{[\mathrm{AT}]_{t-1}}-[\mathrm{TAH}]_t$	Hirshleifer et al. (2004)
extend_comp_NOAH	$\frac{-[\mathrm{CHE}]_t-[\mathrm{DLC}]_t-[\mathrm{DLTT}]_t-[\mathrm{MIB}]_t-[\mathrm{PSTK}]_t-[\mathrm{CEQ}]_t}{[\mathrm{AT}]_{t-1}}$	Hirshleifer et al. (2004)
extend_comp_Dequity	$\frac{\mathrm{SSTK}-\mathrm{PRSTKC}-\mathrm{CAPX}}{\frac{[\mathrm{AT}]_t+[\mathrm{AT}]_{t-1}}{2}}$	Bradshow, Richardson and Sloan (2006)
extend_comp_Ddebt	$\frac{\mathrm{DLTIS}-\mathrm{DLTR}-\mathrm{DLCCH}}{\frac{[\mathrm{AT}]_t+[\mathrm{AT}]_{t-1}}{2}}$	Bradshow, Richardson and Sloan (2006)
extend_comp_Dfin	extnd_comp_Dequity + extnd_comp_Ddebt	Bernhard et al. (1996)
extend_comp_CI	$\frac{[\mathrm{CAPX}]_t}{\frac{[\mathrm{CAPX}]_{t-1}+[\mathrm{CAPX}]_{t-2}+[\mathrm{CAPX}]_{t-3}}{3}}-1$	?? ?Titman, Wei and Xie (2004)

Figures

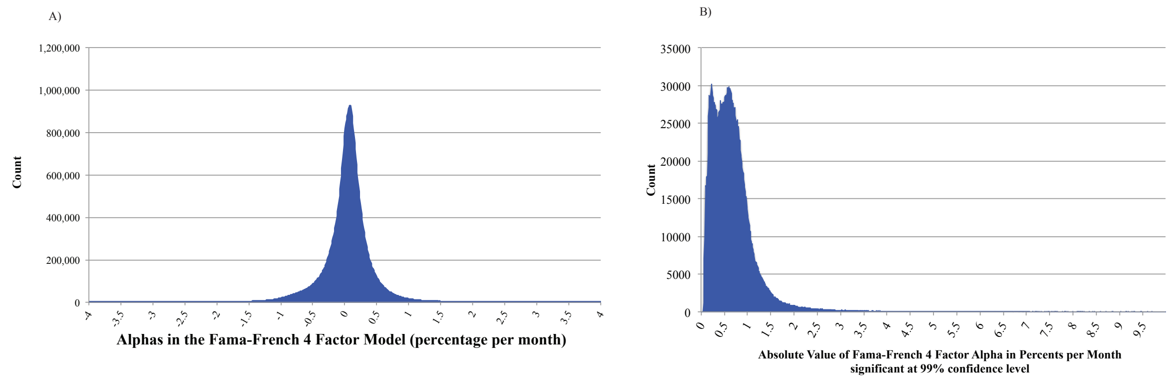

Figure 1. Distribution of Alphas of the *Informed* Portfolios in the Chart's Four-Factor Model
We begin by randomly selecting a number N, which can range from 3 to 10; then we randomly select N items from the Compustat database and perform $N-1$ randomly selected arithmetic operations. The result of each sequence of operations is, in fact, a firm characteristic. On 1 July of year t we sort firms according to this synthetically created firm characteristics and create a value-weighted portfolios from bottom and top decile of that sort. We keep these portfolios until the next July, when we rebalance them using the same rules. We regress excess returns of these portfolios on four factors (excess return on market, High $-$ Low B/M, Small $-$ Big and momentum). We repeat this procedure for more than 44 million different, synthetically created, firm characteristics. In the left panel we plot intercepts from these more than 44 million regressions. The right panel plots the distribution of only those intercepts that are significant at 99% confidence interval.

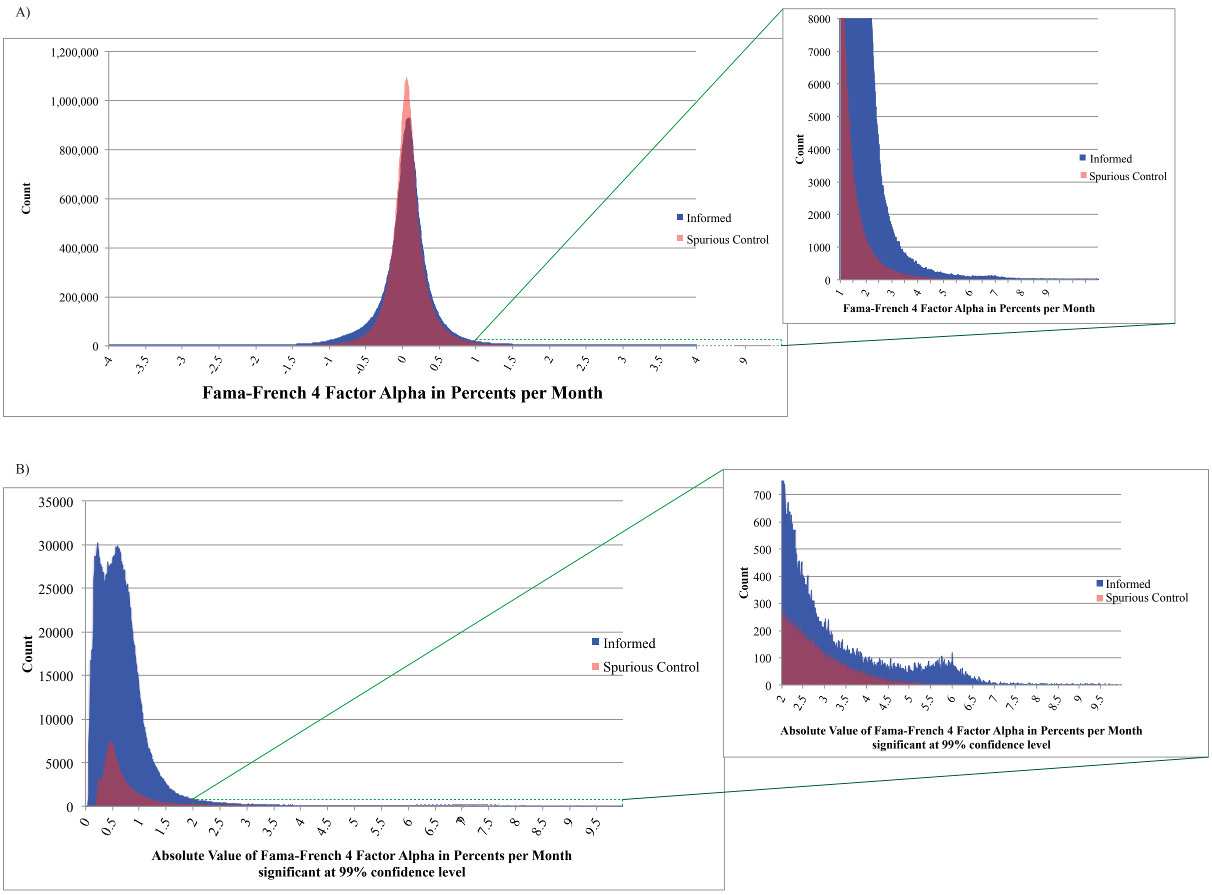

Figure 2. Alphas of *Informed* vs *Spurios Control* Portfolios

Panel A) compares the distributions of alphas in four factor Carhart model from *Informed* portfolios to those from *Spurious Control* portfolios. The *Informed* portfolios are formed in the same way as in Figure 1. The *Spurious Control* portfolios are generated in the same way as the *Informed* portfolios except that no information is used when sorting and forming portfolios. These portfolios are formed by replacing each firm in the *Informed* portfolios with a randomly selected company from the universe of all firms. Panel B) is similar except that it compares absolute values of only those alphas that are significant at 99% confidence level.

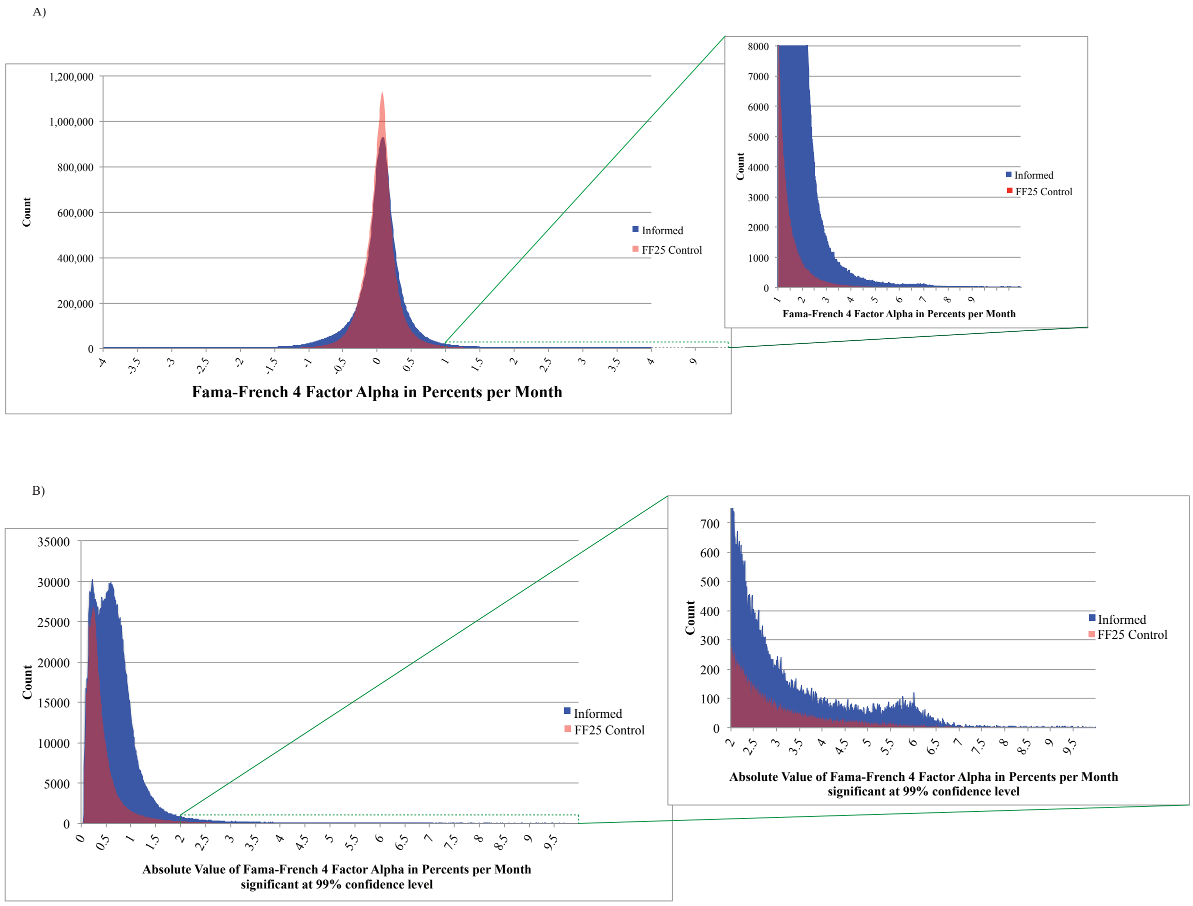

Figure 3. Alphas of *Informed* versus *FF25 Control* portfolios

Panel A) compares the distributions of alphas in four factor Carhart model from *Informed* portfolios to those from *FF25 Control* portfolios. The *Informed* portfolios are formed in the same way as in Figure 1. The *FF25 Control* portfolios are created contain firms of the similar size and book-to-market value as their corresponding *Informed* portfolios but no information is used when sorting and forming portfolios. The *FF25 Control* portfolios are created by replacing each firm in the *Informed* portfolios with a randomly selected company positioned in the corresponding cell of the 5×5 Fama–French matrix. Panel B) is similar except that it compares absolute values of only those alphas that are significant at 99% confidence level.

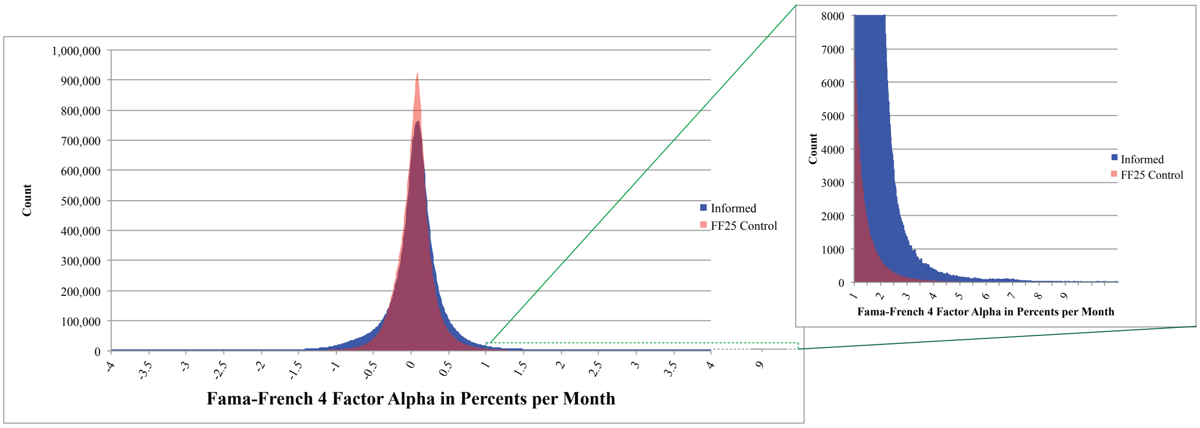

Figure 4. Alphas of *Informed* portfolios with Compustat vs. Extended Compustat

This figure is the same as Figure 3 except that we exclude all portfolios that were created using the variables used to construct anomalies previously known to the literature.

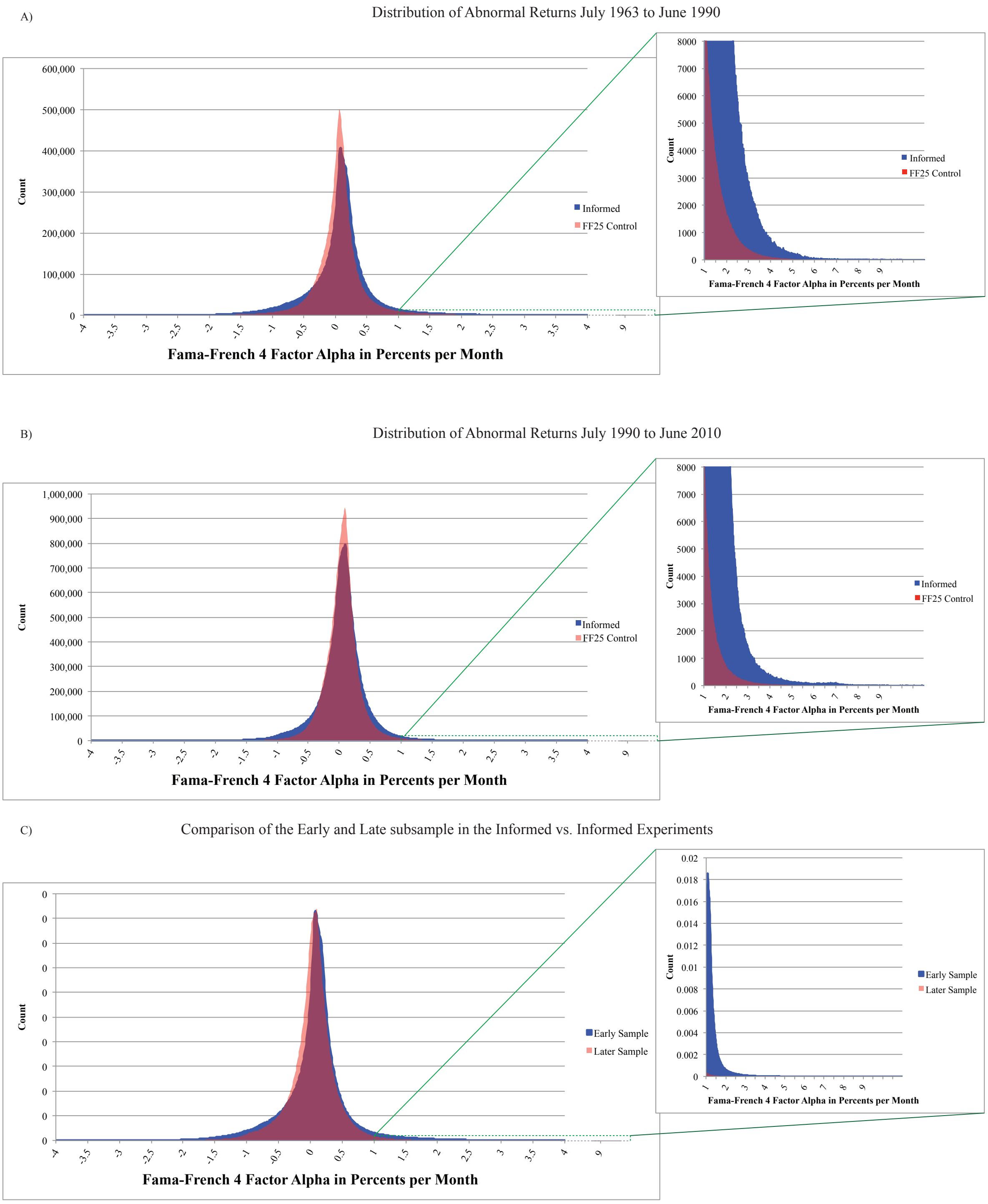

Figure 5. Subsample Analysis

Panels A) and B) are the same as Figure 3 except that they cover different time periods. The panel C) compares alphas in the four-factor Carhart's model of *Informed* portfolios for two distinct time periods. The early sample goes from 1963-1990 while the late sample covers the period from 1990 till 2010.

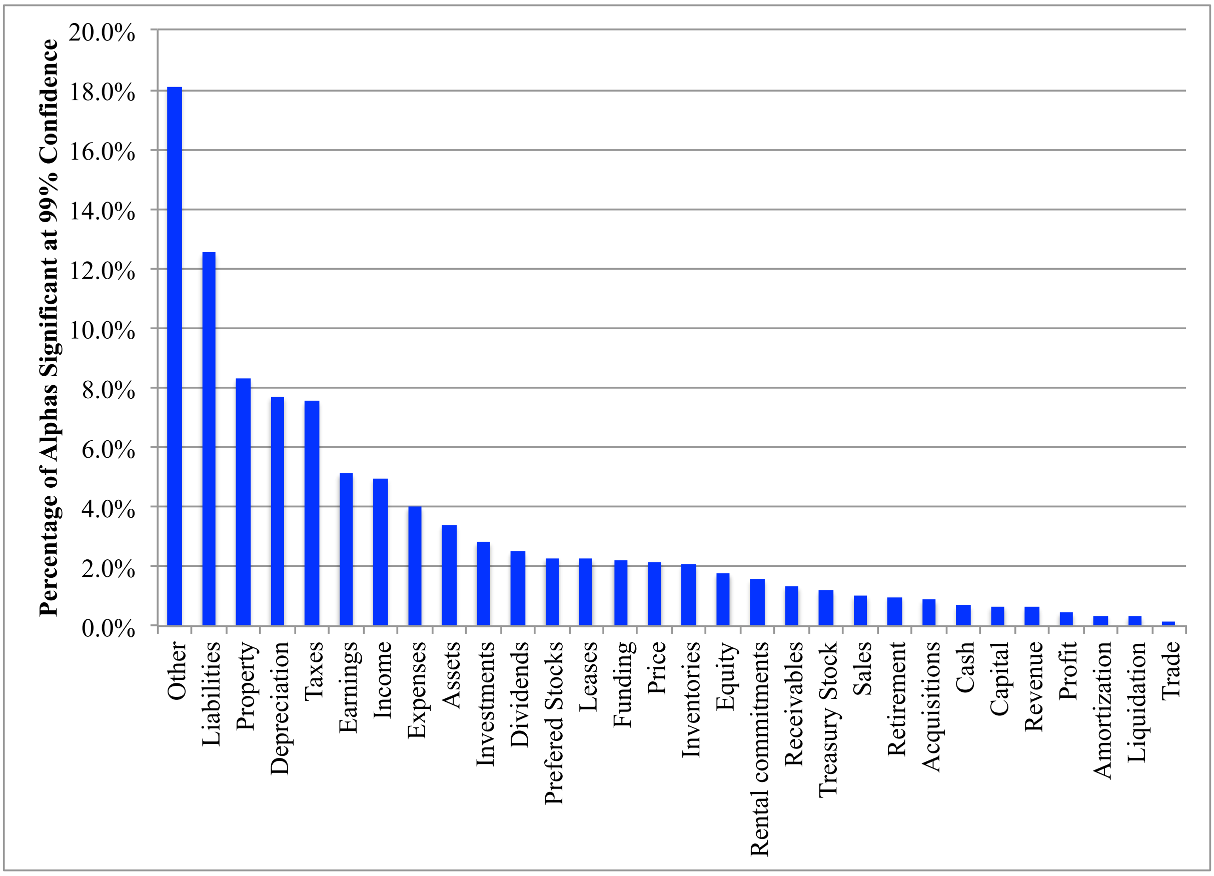

Figure 6. Accounting Factors Behind Anomalies
We count appearances of all Compustat variables used for creation of firm characteristics that produce significant miss pricing in the right panel of Figure 1. We sort all Compustat variables in some groups. This chart plots the frequencies of the categories whit which they generate misspriings.

Tables

Table I
Summary Statistics

This table summarizes the size of the quest undertaken in this paper. It is most likely that we have, performed the largest disclosed, finance related, data mining exercise known to date. More than 6 terabytes of data were generated and analyzed; this figure includes more than 15 billion regressions and more than 1.3 trillion estimated coefficients.

Number of Variables in Compustat	331
Number of Variables in Enlarged-Compustat	687
Number of Firms	12,228
Number of Firm/Years	125,575
Number of Firm/Months	1,452,011
Number of Portfolio Weighting Methodologies	2
Number of Experiment Types	3
Number of Considered Time Periods	3
Number of Factor Models of Expected Returns	5
Number of Stock Subsamples	26
Number of Firm Characteristics Analyzed	44,878,276
Number of Regressions Estimated	15,667,707,706
Number of estimated coefficients and corresponding t-statistics, standard errors, p-values and R-squares	1,331,755,155,010

Table II
Summary Statistics for *Informed* Portfolios

We begin by randomly selecting a number N, which can range from 3 to 10; then we randomly select N items from the Compustat database and perform $N-1$ randomly selected arithmetic operations. The result of each sequence of operations is, in fact, a firm characteristic. On 1 July of year t we sort firms according to this synthetically created firm characteristics and create a value-weighted portfolios from bottom and top decile of that sort. We keep these portfolios until the next July, when we rebalance them using the same rules. We regress excess returns of these portfolios on four factors (excess return on market, High $-$ Low B/M, Small $-$ Big and momentum). We repeat this procedure for more than 44 million different, synthetically created, firm characteristics. This table shows summary statistics for intercepts obtained in such regressions.

Mean	0.049
Variance	0.168
No. of observations	44,837,210
Number of Alphas Statistically Significant at 99% Confidence	2,748,609
Number of Alphas Statistically Significant at 99% Confidence with Absolute Value Larger than 1% per month	447,620
Number of Alphas Statistically Significant at 99% Confidence with Absolute Value Larger than 5% per Month	9,668
Number of Alphas Statistically Significant Alphas at 99% Confidence with Absolute Value Larger than 10% per month	164
Maximum Value of ABS(alpha) in % per Month	26.52

Table III
Comparison of alphas of *Informed* and *Spurious Control* portfolios

This table compares the distributions of alphas in four factor Carhart model from *Informed* portfolios to those from *Spurious Control* portfolios. The *Informed* portfolios are formed in the same way as in Table II. The *Spurious Control* portfolios are generated in the same way as the *Informed* portfolios except that no information is used when sorting and forming portfolios. These portfolios are formed by replacing each firm in the *Informed* portfolios with a randomly selected company from the universe of all firms.

	Informed	**Spurious Control**	**Test Statistic**	**p-value**
Mean	0.049	0.045	t-stat=-239	0.0000
Variance	0.168	0.097	f-stat=1.37	0.0000
Kolmogorov-Smirnov test			K-S-stat=0.054	0.0000
No. of observations	44,837,210	44,837,210	z-stat=820	0.0000
Number of Alphas Statistically Significant at 99% Confidence	2,748,609	504,333	z-stat=1300	0.0000
Number of Alphas Statistically Significant at 99% Confidence with Absolute Value Larger than 1% per month	447,620	75,450	z-stat=516	0.0000
Number of Alphas Statistically Significant at 99% Confidence with Absolute Value Larger than 5% per Month	9,668	2,161	z-stat=69	0.0000
Number of Alphas Statistically Significant Alphas at 99% Confidence with Absolute Value Larger than 10% per month	164	1	z-stat=13	0.0000
Maximum Value of ABS(alpha) in % per Month	26.52	10.17		

Table IV
Comparison of alphas of *Informed* and *FF25 Control* portfolios

This table compares the distributions of alphas in four factor Carhart model from *Informed* portfolios to those from *FF25 Control* portfolios. The *Informed* portfolios are formed in the same way as in Table II. The *FF25 Control* portfolios are created contain firms of the similar size and book-to-market value as their corresponding *Informed* portfolios but no information is used when sorting and forming portfolios. The *FF25 Control* portfolios are created by replacing each firm in the *Informed* portfolios with a randomly selected company positioned in the corresponding cell of the 5×5 Fama–French matrix.

	Informed	**FF25 Control**	**Test Statistic**	**p-value**
Mean	0.049	0.044	t-stat=56	0.0000
Variance	0.168	0.096	f-stat=1.75	0.0000
Kolmogorov-Smirnov test			K-S-stat=0.059	0.0000
No. of observations	44,837,210	44,837,210	z-stat=820	0.0000
Number of Alphas Statistically Significant at 99% Confidence	2,748,609	1,162,350	z-stat=820	0.0000
Number of Alphas Statistically Significant at 99% Confidence with Absolute Value Larger than 1% per month	447,620	92,511	z-stat=485	0.0000
Number of Alphas Statistically Significant at 99% Confidence with Absolute Value Larger than 5% per Month	9,668	1,794	z-stat=74	0.0000
Number of Alphas Statistically Significant Alphas at 99% Confidence with Absolute Value Larger than 10% per month	164	5	z-stat=12	0.0000
Maximum Value of ABS(alpha) in % per Month	26.52	10.13		

Table V
Compustat without the anomalies

This table is the same as Table IV except that we exclude all portfolios that were created using the variables used to construct anomalies previously known to the literature.

	Informed	FF25 Control	Test Statistic	p-value
Mean	0.047	0.043	t-stat=44	0.0000
Variance	0.169	0.096	f-stat=1.76	0.0000
Kolmogorov-Smirnov test			K-S-stat=0.058	0.0000
No. of observations	36,590,720	36,590,720		
Number of Alphas Statistically Significant at 99% Confidence	2,252,384	950,578	z-stat=744	0.0000
Number of Alphas Statistically Significant at 99% Confidence with Absolute Value Larger than 1% per month	374,547	74,860	z-stat=448	0.0000
Number of Alphas Statistically Significant at 99% Confidence with Absolute Value Larger than 5% per Month	7,998	1,476	z-stat=67	0.0000
Number of Alphas Statistically Significant Alphas at 99% Confidence with Absolute Value Larger than 10% per month	142	3	z-stat=12	0.0000
Maximum Value of ABS(alpha) in % per Month	26.52%	10.11		

Table VI
Anomalies and Degree of Complexity

This table shows means and variances of 14 different distributions of alphas in a four-factor Carharts model with *Informed* portfolios. All these distributions are created in the same way as in Table II with a difference that the first 7 are alphas of equally weighted portfolios and the remaining 7 are from the value weighted portfolios. Within each of these two groups these 7 distribution differ by how complex information was used to create *Informed* portfolios. The sorting firm characteristic in the least complex case is made of 3 Compustat variables, while the most complex one is made of 10 Compustat variables. Both, variances and means increase monotonically with complexity and these pairwise differences are always significant at 99% confidence level.

Number of Compustat Columns	Mean	Variance	Portfolio Weights
3	0.058	0.103	
4	0.073	0.128	
5	0.083	0.151	
6	0.092	0.170	EW
7	0.098	0.190	
8	0.101	0.207	
9	0.105	0.221	
10	0.107	0.234	
3	0.035	0.103	
4	0.041	0.131	
5	0.046	0.158	
6	0.052	0.182	VW
7	0.057	0.206	
8	0.061	0.226	
9	0.066	0.247	
10	0.069	0.260	

Table VII
Comparison of Sample periods

This table compares alphas in the four-factor Carhart's model of *Informed* portfolios for two distinct time periods. The early sample goes from 1963-1990 while the late sample covers the period from 1990 till 2010. The *Informed* portfolios are formed in the same way as in Table II.

	1963-1990	**1990-2010**	**Test Statistic**	**p-value**
Mean	0.071	0.044	t-stat=213	0.0000
Variance	0.333	0.177	f-stat=1.88	0.0000
Kolmogorov-Smirnov test			K-S-stat=0.069	0.0000
No. of observations	21,930,130	42,366,962		
Number of Alphas Statistically Significant at 99% Confidence	1,829,461	1,659,831	z-stat=742	0.0000
Number of Alphas Statistically Significant at 99% Confidence with Absolute Value Larger than 1% per month	553,762	426,301	z-stat=471	0.0000
Number of Alphas Statistically Significant at 99% Confidence with Absolute Value Larger than 5% per Month	2,185	9,173	z-stat=33	0.0000
Number of Alphas Statistically Significant Alphas at 99% Confidence with Absolute Value Larger than 10% per month	36	139	z-stat=3.77	0.0002
Maximum Value of ABS(alpha) in % per Month	16.20%	26.52%		

Table VIII
Comparison of Models of Expected Returns

Model1	Model2	Mean (Model1)	Mean (Model2)	Mean Comparison t-statistics	Mean Comparison p-value	Variance (Model1)	Variance (Model2)	Variance comparison p-value	Distribution Comparison KS p-value	Portfolio Weights	Sample
M	F3	0.015	-0.092	1,214	0.0000	0.202	0.150	0.0000	0.0000		
M	F4	0.015	0.083	-758	0.0000	0.202	0.158	0.0000	0.0000		
M	C	0.015	0.061	-522	0.0000	0.202	0.140	0.0000	0.0000		
M	C2	0.015	0.032	-190	0.0000	0.202	0.131	0.0000	0.0000		
F3	F4	-0.092	0.083	-2,116	0.0000	0.150	0.158	0.0000	0.0000	EW	
F3	C	-0.092	0.061	-1,905	0.0000	0.150	0.140	0.0000	0.0000		
F3	C2	-0.092	0.032	-1,565	0.0000	0.150	0.131	0.0000	0.0000		
F4	C	0.083	0.061	275	0.0000	0.158	0.140	0.0000	0.0000		
F4	C2	0.083	0.032	642	0.0000	0.158	0.131	0.0000	0.0000		
C	C2	0.061	0.032	376	0.0000	0.140	0.131	0.0000	0.0000		*Informed*
M	F3	-0.043	0.011	-604	0.0000	0.173	0.177	0.0000	0.0000		
M	F4	-0.043	0.049	-1,053	0.0000	0.173	0.168	0.0000	0.0000		
M	C	-0.043	0.039	-926	0.0000	0.173	0.175	0.0000	0.0000		
M	C2	-0.043	0.027	-802	0.0000	0.173	0.168	0.0000	0.0000		
F3	F4	0.011	0.049	-439	0.0000	0.177	0.168	0.0000	0.0000	VW	
F3	C	0.011	0.039	-318	0.0000	0.177	0.175	0.0000	0.0000		
F3	C2	0.011	0.027	-189	0.0000	0.177	0.168	0.0000	0.0000		
F4	C	0.049	0.039	119	0.0000	0.168	0.175	0.0000	0.0000		
F4	C2	0.049	0.027	254	0.0000	0.168	0.168	0.0000	0.0000		
C	C2	0.039	0.027	133	0.0000	0.175	0.168	0.0000	0.0000		
M	F3	-0.004	-0.115	1,705	0.0000	0.120	0.071	0.0000	0.0000		
M	F4	-0.004	0.050	-780	0.0000	0.120	0.094	0.0000	0.0000		
M	C	-0.004	0.044	-698	0.0000	0.120	0.086	0.0000	0.0000		
M	C2	-0.004	0.016	-303	0.0000	0.120	0.075	0.0000	0.0000		
F3	F4	-0.115	0.050	-2,726	0.0000	0.071	0.094	0.0000	0.0000	EW	
F3	C	-0.115	0.044	-2,683	0.0000	0.071	0.086	0.0000	0.0000		
F3	C2	-0.115	0.016	-2,301	0.0000	0.071	0.075	0.0000	0.0000		
F4	C	0.050	0.044	103	0.0000	0.094	0.086	0.0000	0.0000		
F4	C2	0.050	0.016	552	0.0000	0.094	0.075	0.0000	0.0000		
C	C2	0.044	0.016	457	0.0000	0.086	0.075	0.0000	0.0000		*Spurious Control*
M	F3	0.038	0.050	-171	0.0000	0.118	0.128	0.0000	0.0000		
M	F4	0.038	0.068	-417	0.0000	0.118	0.123	0.0000	0.0000		
M	C	0.038	0.056	-238	0.0000	0.118	0.135	0.0000	0.0000		
M	C2	0.038	0.051	-180	0.0000	0.118	0.128	0.0000	0.0000		
F3	F4	0.050	0.068	-238	0.0000	0.128	0.123	0.0000	0.0000	VW	
F3	C	0.050	0.056	-67	0.0000	0.128	0.135	0.0000	0.0000		
F3	C2	0.050	0.051	-9	0.0000	0.128	0.128	0.0000	0.0000		
F4	C	0.068	0.056	167	0.0000	0.123	0.135	0.0000	0.0000		
F4	C2	0.068	0.051	230	0.0000	0.123	0.128	0.0000	0.0000		
C	C2	0.056	0.051	59	0.0000	0.135	0.128	0.0000	0.0000		