Ranking Stocks and Returns: A Non-Parametric Analysis of Asset Pricing Anomalies *

Denys Maslov[†] and Oleg Rytchkov [‡]

PRELIMINARY. COMMENTS WELCOME.
February 14, 2010

Abstract

In this paper, we apply a non-parametric rank-based technique to analyze nine asset pricing anomalies. We demonstrate that the relation between almost every anomalous characteristic and abnormal returns is non-monotonic. In particular, many anomalies are detectable only for high values of characteristics. We argue that due to the presence of non-monotonicity the similarity between anomalous characteristics should be examined separately for different ranges of each variable. We demonstrate how the standard linear regression i) may significantly overstate or understate the strength of relation between characteristics and returns; ii) may fail to recognize the similarity as well as the difference between anomalies. We also introduce the distance between asset pricing anomalies and perform a cluster analysis in the anomaly space. We find that for stocks with extreme characteristics almost all considered anomalies appear to be statistically different.

Keywords: asset pricing, anomaly, Spearman rank correlation
JEL classification: G12

*We are grateful to Andres Almazan, Jonathan Cohn, Shimon Kogan, Igor Makarov, Jeremy Page, Clemens Sialm, Sheridan Titman, Chishen Wei, Yuzhao Zhang, and seminar participants at Temple University and University of Texas at Austin for very helpful comments. Nathaniel Light provided very good assistance in gathering information on anomalies. All remaining errors are our own.

[†]McCombs School of Business, University of Texas at Austin, 1 University Station, B6600 Austin, TX 78759. E-mail: denys.maslov@phd.mccombs.utexas.edu.

[‡]Fox School of Business, Temple University, 1801 Liacouras Walk, 423 Alter Hall (006-01), Philadelphia, PA 19122. Phone: (215) 204-4146. E-mail: rytchkov@temple.edu.

1 Introduction

An asset pricing anomaly is a pattern in expected stock returns which cannot be explained by a particular asset pricing model. Many anomalies are associated with firm characteristics and typically thought of as a monotonic relation between the characteristic and future abnormal stock returns. Among the existing asset pricing models, the Fama-French three-factor model (Fama and French, 1993) is the most popular, but even with respect to this model the number of discovered anomalies is overwhelming. Subrahmanyam (2009) reviews the literature and documents more than fifty variables explaining the cross-section of stock returns. Given a so long list of anomalies, several important questions arise. Are the documented relations between characteristics and abnormal returns indeed monotonic? How different are these anomalies? What is the best method to differentiate between several anomalies and classify them if they may be non-linear and even non-monotonic? The purpose of this paper is to give at least partial answers to these questions.

As a major tool for analysis of anomalies we propose a non-parametric approach, which hinges on two ideas. First, we suggest to characterize each potentially anomalous variable by the ranking of stocks it produces. If it is a real anomaly, this ranking must be similar to the ranking based on stock alphas. Thus, to test how anomalous the given characteristic is we use rank correlations between characteristics and abnormal returns. Second, we examine the relation between each anomalous characteristic and stock returns separately for various ranges of the characteristic. Being robust to non-linearities, the rank-based approach allows us to examine anomalies on the stock level *inside* various quintile portfolios. Such analysis reveals if the anomaly is monotonic or not.

We advocate for using rank-based correlations computed for individual securities since in the presence of non-linearities the standard linear regressions (including the Fama-MacBeth regression) may produce substantially misleading results. Using simple theoretical examples we demonstrate that the strength of a linear correlation between some characteristic and abnormal stock returns cannot be used for judging how well the characteristic ranks stocks with respect to their alphas. Also, the strength of a linear correlation between two characteristics is not appropriate for answering the question if these characteristics rank stocks similarly or not. Finally, we demonstrate that in the presence of non-linearities the regression is not the appropriate tool for distinguishing two anomalies.

Since the list of known anomalies is too long to be comprehensively studied in one paper, we limit our analysis to a small subset of them. Along with the book-to-market anomaly (Rosenberg, Reid, and Lanstein, 1984; Fama and French, 1992) and the size anomaly (Banz, 1981) which motivated the Fama-French three-factor model, we examine the analysts' forecasts anomaly (Diether, Malloy, and Scherbina, 2002), the idiosyncratic volatility anomaly (Ang, Hodrick, Xing, and Zhang, 2006), the total asset growth anomaly (Cooper, Gulen, and Schill, 2008), the abnormal capital investments anomaly (Titman, Wei, and Xie, 2004), the investments-to-assets ratio anomaly (Lyandres, Sun, Zhang, 2008), net stock issues anomaly (Fama and French, 2008), and the composite stock issuance anomaly (Daniel and Titman, 2006). Our choice is motivated by two factors. On one hand, these anomalies attracted much attention in the literature and many of them already have received theoretical explanations (e.g., Johnson, 2004; Li, Livdan, and Zhang, 2009; Avramov, Cederburg, and Hore, 2009). On the other hand, we try to simultaneously consider anomalies which are likely to be unrelated (e.g, analysts' forecasts anomaly and total asset growth anomaly) as well as anomalies which are likely to have the same nature (e.g., net stock issues anomaly and composite stock issuance anomaly).

The paper contains several empirical results. First, we document that many anomalies are not monotonic, i.e. the sign and the strength of the relation between the anomalous characteristic and abnormal returns significantly depends on the value of the characteristic itself. We demonstrate that many anomalies exist for extreme values of characteristics only and, surprisingly, the ranking correlations between characteristics and returns can change its sign as we move from one extreme to the other. For instance, idiosyncratic volatility has a negative relation to abnormal returns in quintile 5 (where idiosyncratic volatility is high) with the Spearman t-statistic -14.94 but it is *positively* related to abnormal returns in quintile 1 (where idiosyncratic volatility is low) with t-statistic 3.81. All anomalies except size are more pronounced for stocks with high values of characteristics, and in general are less evident in the intermediate range of anomalous characteristics. It should be emphasized that this result is not driven by higher dispersion of characteristics in the extreme portfolios, which may produce the heterogeneity of precision with which ranking correlations can be computed across different portfolios. Thus, the statements saying that some characteristic is positively/negatively related to expected returns is misleading for most anomalies.

It is well known that anomalies may be more pronounced on certain groups of stocks. For instance, Fama and French (2008) examine several anomalies across size groups and argue that the net stock issues anomaly shows up in all size groups whereas the asset growth anomaly exists on microcaps and small stocks only. Li and Zhang (2009) analyze the impact of financing constraints on anomalies and demonstrate that they amplify the investment-to-assets and asset growth anomalies. A similar effect is produced by idiosyncratic volatility, interpreted as a measure of arbitrage costs (Pontiff, 2006), on the book-to-market anomaly (Ali, Hwang, and Trombley, 2003) and the asset growth anomaly (Lipson, Mortal, and Schill, 2009). Cao (2009) examines the relation between idiosyncratic volatility and stock returns separately for undervalued and overvalued stocks and finds the opposite relations. According to our knowledge, our paper is the first to consider systematically the variation of the strength of anomaly along the anomalous characteristic itself.

While examining the rank correlations inside quintile portfolios, we effectively perform a test for monotonicity in the relation between characteristics and returns on individual securities. Having established that in the top and bottom portfolios there are statistically significant relations between characteristic and returns with opposite signs (positive in a bottom portfolio, negative in the top portfolio) we can reject the null hypothesis of a monotonic pattern in expected returns. Thus our paper contributes to the literature studying the monotonicity in asset pricing (e.g., Patton and Timmermann, 2009). However, our test seems to be more powerful than the portfolio-based tests. Although portfolio formation allows us to estimate expected returns more precisely, we also lose a lot of information about the dispersion of the characteristic inside portfolios (especially inside extreme portfolios) and this results in lower power of monotonicity tests.[1]

Next, we examine the similarity among the nine selected anomalies. We compute the rank correlations between characteristics for all stocks as well as for the stocks in each quintile portfolio. More importantly, to see how anomalies are related to each other, we introduce the distance in the anomaly space based on the rank correlations and perform a cluster analysis. As discussed by Subrahmanyam

[1]A similar point in the context of estimation of betas and factor loadings was recently made by Ang, Liu, and Schwarz (2008).

(2009), the lack of unified framework for analysis of anomalies significantly hampers our ability to compare and classify them. The introduction of the anomaly space with a natural metric on it may be a first step towards such unification enabling better understanding of interrelations between anomalies.

We have two findings here. First, the distance between anomalous characteristics significantly depends on the range of the characteristic for which it is computed. Specifically, we discover a tendency of anomalous characteristics to have higher correlation with each other among extreme stocks. Second, the cluster analysis represents a meaningful way to classify anomalies and in our case it is able to recognize asset growth related anomalies and put them into one cluster.

Last, we use partial rank correlations to test whether each characteristic contains additional information about ranking of abnormal returns after conditioning on other characteristics. This analysis reveals that almost all considered anomalies are statistically different among extreme stocks (where anomalies are especially strong), but the results are mixed if all stocks are considered simultaneously. Specifically, the anomalies based on asset growth, abnormal capital investments, and investments-to-assets ratio do not appear to be robust to conditioning on other variables. This result highlights the importance of comparing characteristics on various ranges.

Our results have several implications. Since the discovered anomalous strategies are relatively complicated and usually involve constant portfolio rebalancing, it is sometimes very difficult to say if the newly minted regularity in expected stock returns is really a new phenomenon or already a known anomaly discovered under a new name. The answer to this question has a paramount importance from the theoretical as well as the practical point of view. Indeed, if two anomalies constructed using different characteristics result in similar portfolios, there is no need to explain these anomalies separately. In the rational asset pricing paradigm any cross-sectional variation in stock returns is produced by the variation in loadings of stock returns on risk factors. Before delving into construction of new risk factors whose loadings would explain the anomaly we want to be sure that we deal with a new phenomenon. From the practical point of view, the ability to say whether two anomalies are different or not is important for efficient allocation of resources among money managers. Is it possible that different managers who use seemingly different strategies based on different anomalies ultimately end up with very similar portfolios just because they actually exploit the same anomaly? The answer to this question may significantly affect how resources should be allocated between different money managers.

The rest of the paper is organized as follows. Section 2 summarizes the existing approaches to identifying anomalous characteristics and testing if two anomalies are really different. In Section 3 we introduce our rank based methodology and define the distance in the space of all anomalies. Section 4 is devoted to comparison of our methodology with the linear regression. Sections 5 contains our main empirical results. Section 6 concludes by summarizing main contributions of the paper and discussing potential directions for future research.

2 Asset pricing anomalies

An asset pricing anomaly is a pattern in the cross-section of expected stock returns that is not explained by an asset pricing model. Historically, the CAPM was the first benchmark and all violations of CAPM were considered anomalies (e.g., Banz, 1981; Bhandari, 1988; Rosenberg, Reid, and Lanstein, 1985).

However, later the Fama-French three-factor model (Fama and French, 1993) replaced it and raised the bar, so it is presently standard to call a pattern in expected returns anomalous only if it cannot be explained by loadings on the Fama-French factors. For consistency with the literature, in our empirical analysis we also take the Fama-French three-factor model as a benchmark.

For any asset pricing model, we can always pose three different questions. The first is whether all alphas are zero, i.e. whether the model is consistent with the cross-section of expected returns. Having a negative answer to this question, we can inquire if alphas are associated with certain firm characteristics like valuation ratios, investment activity, etc. The characteristics containing information about alphas are called anomalous. The third question arises if we have two different characteristics related to abnormal returns. Do these characteristics capture the same anomaly and having one of them we do not need the other? The answer to this question is of paramount importance for understanding the cross-section of stock returns and developing asset pricing models.

To test if the given characteristic is anomalous, two major procedures are employed in the literature. The first approach is to assign stocks to portfolios based on the characteristic and examine alphas on these portfolios (e.g., Fama and French, 1993; Daniel and Titman, 1997). Specifically, it is common to form five or ten portfolios and test if the difference in abnormal returns on the top and bottom portfolios is statistically significant. Many papers also use the GRS statistic (Gibbons, Ross, and Shanken, 1989) to test the joint insignificance of portfolio alphas (e.g., Fama and French, 1996). Although portfolio formation reduces the error in the estimates of expected returns, the inference may still be rather imprecise since the portfolio-based tests ignore the variation of characteristics and returns inside portfolios. Moreover, the result may be sensitive to how stocks are weighted inside portfolios (e.g., Bali and Cakici, 2006). The GRS test being designed to uncover non-zero alphas is silent about the relation between characteristics and alphas.

The second approach is to run a Fama-MacBeth regression (Fama and MacBeth, 1973) of realized returns on betas and characteristics. The significance of the characteristic implies the existence of anomaly. Although this test is likely to be more powerful, it assumes a linear relation between characteristics and expected returns, which may be a serious misspecification. Moreover, a linear regression is sensitive to outliers and may produce misleading results especially when characteristics are highly skewed. In this paper, we develop an alternative methodology, which is based on the rank correlation between characteristics and returns.

One of the main questions to be asked after discovering a new anomaly is whether the new anomaly is substantially different from those already known. Obviously, the answer to this question depends on how we define "substantial difference". In analogy to testing individual anomalies, there are two major methods to assess the novelty of the anomaly. The first one is to do two-way or three-way independent sorts based on anomalous characteristics and compare abnormal returns on extreme portfolios (e.g., Lakonishok, Shleifer, and Vishny, 1994). Unfortunately, the applicability of this approach is rather limited. The main problem is that sorts on more than three characteristics are generally infeasible since portfolios become very thin (and even empty if the anomalies are absolutely identical). In this case, because of the small number of stocks in each portfolio, the estimates of expected returns are very imprecise, making the comparison of extreme portfolios very difficult. This problem is especially severe when two anomalies are indeed identical and the rankings generated by both characteristics coincide.

The second way to disentangle anomalies is to run a Fama-MacBeth regression of future individual stock returns on various anomalous characteristics and test if the significance of the given characteristic is subsumed by others (e.g., Boyer, Mitton, Vorkink, 2009; Fama and French, 2008; Lipson, Mortal, and Schill, 2009; Pontiff and Woodgate, 2008). Although the cross-sectional regression is more powerful and more universal, it also suffers from some drawbacks. Besides an already mentioned sensitivity to outliers, the assumption of a linear relation between characteristics and returns may result in a wrong identification of two characteristics as corresponding to different anomalies when they actually reflect the same anomaly and vice versa. Moreover, the characteristics may be highly correlated and the multicollinearity problem may make the inference very imprecise. In this paper, we introduce the distance between anomalies based on the rank correlation between characteristics. It allows us to quantify how different the anomalies are and, in addition, to use partial rank correlations to test the statistical significance of differences among anomalies.

3 Anomalies as stock rankings

3.1 Definitions

In this paper, we introduce a novel rank based characterization of anomalies. We say that a characteristic is anomalous relative to a certain asset pricing model if stocks ranked according to this characteristic also appear to be ranked with respect to abnormal expected returns. Our focus is mostly a statistical analysis of anomalies, so we stay agnostic regarding the origin of the anomaly and do not try to identify whether the anomaly is caused by missing risk factors or behavioral biases. To test the statistical significance of the ability of the characteristic to reveal stocks with high alphas and to measure how strong the anomaly is we suggest to use the Spearman rank correlation coefficient and rank tests.[2]

Our definition of an anomaly captures the general intuition that an anomalous characteristic should predict returns, but it also has several advantages over standard techniques described above. First, our approach is non-parametric and, consequently, quite flexible. It does not impose any restrictions on the functional form of returns and goes far beyond a simple linear relation between anomaly variables and stock returns, which is almost always implicitly assumed. Using rankings we might lose some accuracy but gain robustness, since the inference based on rankings is largely insensitive to the exact functional relation between characteristics and returns. Moreover, since expected returns are measured rather imprecisely, the loss of accuracy is likely to be minimal.

In rational asset pricing models, expected returns are exclusively determined by loadings on risk factors. However, often these loadings are unobservable and characteristics serve as proxies for them. Also, characteristics may be helpful for explaining expected returns if the dynamics of factor loadings are misspecified (Berk, Green, and Naik, 1999) or conditional factor loadings are measured imprecisely (Gomes, Kogan, and Zhang, 2003). In such cases, although the theoretical relation between expected

[2] The class of rank statistics designed to test independence between two random variables is quite large (Hájek, Šidák, and Sen, 1999). We use the Spearman correlation as one the simplest and intuitive. Another advantage of the Spearman rank correlation is that it assigns higher weights to those objects which are located distantly according to two rankings (as opposed to the Kendall rank correlation for example, which counts only the pairs of objects ordered differently in two rankings ignoring the quantitative difference in ranks).

stock returns and anomalous characteristics is almost always monotonic, it is typically nonlinear. For instance, Livdan, Sapriza, and Zhang (2009) demonstrate how financial constraints produce a convex relation between market leverage and expected returns. In such cases, the linear specification of anomaly is obviously inappropriate whereas the ranking approach should work quite well.

Second, our definition does not assume that there is a strict monotonic relation between characteristics and expected returns. Although a positive rank correlation means that a higher value of characteristic on average must correspond to higher expected return, stocks with the same expected returns are allowed to have different values of the characteristic even in population. It means that locally (for some intervals of expected returns) the characteristics may be linked to expected returns non-monotonically.

Third, the rank-based approach appears to be more robust to outliers, ensuring that the conclusions are not driven by unusual stocks with extremely high or low characteristics. Empirically, most anomalous characteristics have very skewed distributions (e.g., the unconditional skewness of asset growth is 383) resulting in a potentially high impact of extreme stocks. If the anomaly is uncovered using the rank correlation, it is more likely to be valid for a broader group of stocks.

Fourth, our definition of anomalies is consistent with the standard portfolio-based definition: if rankings produced by the characteristic and expected returns are similar, the portfolio of stocks with high value of characteristic will outperform the portfolio with low value of characteristic. However, as we demonstrate below, our definition leads to more powerful test of anomaly existence than the standard comparison of quintile portfolios.

From a practical point of view, it is extremely hard to measure precisely expected returns and the estimates of the functional relation between characteristics and future stock returns are very noisy. Even if we have the functional form of the anomaly estimated, the best way to exploit the anomaly is to find a set of stocks with high expected returns and to form a portfolio using a simple rule of thumb (e.g., take stocks equally weighted or value weighted). Thus, the inability of the rank based approach to predict quantitatively stock returns (it can only say where returns on the given stock will be relative to other stocks) is not a big hurdle for exploiting the anomaly.

The rank-based definition of anomalies naturally implies a certain anomaly classification, i.e. the way to define which anomalies are identical and which are distinct. First, we view two anomalous characteristics as corresponding to the same anomaly in the population if they rank all stocks identically. This classification convention is motivated by the following argument. If we treated two variables producing the same ranking of stocks as different anomalies, then any monotonic transformation of an anomalous characteristic would produce a new anomaly (for example, size and size^3 should be treated as separate anomalies). Since the number of possible monotonic transformations is infinite, any anomaly would immediately generate an infinite number of other anomalies and classification of them would not make any sense.

Second, although this factorization works well in the population, it is not applicable in practice. Indeed, because of measurement errors it would be very naive to expect that two given variables rank the stocks in exactly the same way. Moreover, the relation between the anomaly variable and alphas may be imperfect even in population, and two variables despite ranking stocks differently may contain the same information about abnormal returns. Thus, we extend our concept of equivalent anomalies and will say that two anomaly variables are indistinguishable if condition on ranking produced by one variable, the other variable has zero rank correlation with abnormal stock returns. In our empirical analysis, we will use sample partial rank correlations to test if two anomalies are identical.

3.2 Rank correlation in a panel

To make the definitions from the previous section operational, we introduce rank correlations in the context of panel data. Conceptually, rank correlations are designed to compare two given rankings of objects and test if these rankings are independent. As explained in the previous section, we use the Spearman rank correlation. Given two rankings p_i^1 and p_i^2 of N objects the Spearman rank correlation is defined as (Kendall, 1970)

$$\rho_{12} = 1 - \frac{6}{N^3 - N} \sum_{i=1}^{N} (p_i^1 - p_i^2)^2. \tag{1}$$

Since we observe time series of realized returns and stock characteristics, we effectively have multiple pairs of rankings with each pair corresponding to one period of time. To exploit all available information, we use a Fama-MacBeth-type procedure. Specifically, having T periods we rank objects using two different criteria (e.g., rank stocks using two different characteristics) in each period and construct the time series of rank correlations ρ_t, $t = 1, \ldots, T$. The best estimate of the rank correlation for the whole panel ρ is the average of ρ_t across all periods.

Having a panel of returns for which the serial correlation in negligible, it is safe to assume that the estimated rank correlations from different periods are independent.[3] Thus, we compute the standard deviation of ρ as a standard deviation of ρ_t and use the t-statistic to test the hypothesis $\rho = 0$. Note that if the number of objects N_t in period t is sufficiently large and the rankings p_i^1 and p_i^2 are independent then ρ_t is normally distributed: $\rho_t \sim N(0, 1/(N_t - 1))$ (Hájek, Šidák, and Sen, 1999). Hence, ρ is also normally distributed and the t-statistic has a conventional distribution.

3.3 Distance between anomalies

The rank correlation discussed above may be used for quantifying how different two given anomalous characteristics are. For this purpose, we introduce the distance between anomalies, which effectively defines a metric in the space of all anomalies. Specifically, in our empirical analysis the distance between anomalies i and j is computed by the following procedure.

1. For each anomalous characteristic i and each time period t sort all available stocks according to the characteristic assuming that stocks with supposedly higher expected returns have a higher rank.

2. For the anomalies i and j find those stocks which are ranked relative to both characteristics i and j and construct the corresponding rankings $x_{nt}^{(i)}$ and $x_{nt}^{(j)}$ on the intersection.

3. Using Eq. (1) and the Fama-MacBeth-type procedure described in Section 3.2 compute the panel Spearman rank correlation ρ_{ij}.

[3] In our empirical analysis, we checked that this assumption is innocuous. Unreported estimations show that the results are essentially unaffected if the Newey and West (1987) standard errors are used for construction of t-statistics.

4. The distance between anomalies is defined as

$$d_{ij} = \sqrt{1 - \rho_{ij}}. \tag{2}$$

This procedure deserves several comments. First, although the difference between anomalies can be measured directly by the correlation (for instance, like $1 - \rho_{ij}$) we define the distance by less straightforward Eq. (2), which guarantees that the distance d_{ij} is indeed a metric on the anomaly space and satisfies all standard conditions.[4]

Next, while computing rank correlations, special attention should be given to the possibility of having tied ranks. Since all characteristics we use in our analysis are continuous, the probability of having ties is negligible and can be safely ignored.

We compute the distance using stocks for which both characteristics are well defined. In our empirical analysis, the overlap is typically quite substantial and there is no reason to believe that the stocks for which only one characteristic is available are significantly different. However, it may be useful to keep in mind that the analysis of distances uses a slightly different universe of stocks relative to the analysis of stock returns. This is especially important if the distance is computed for anomalies determined for a small subsample of stocks only.

3.4 Partial rank correlations

Although the distance between anomalous characteristics convey information on how different these anomalies are, it does not allow to test statistically the significance of this difference. Indeed, it would be very naive to expect that two characteristics rank stocks absolutely identically, and if it is not the case the distance between characteristics is unambiguously positive. However, it could be that even if characteristics are completely different, they contain the same information about stock alphas, and having one characteristic we do not need the other to reconstruct the ranking of abnormal returns. Thus, for each pair of characteristics we need to answer the following question: having a ranking of stocks produced by one characteristic and conditioning upon this ranking, what is the correlation between the ranking according to the other characteristic and stock returns? The answer to this question is provided by the partial rank correlation coefficient.

Assume that we have three rankings p^1, p^2, and p^3. The partial rank correlation between rankings p^1, p^2 conditioned on the ranking p^3 is defined as

$$\rho_{12|3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{(1 - \rho_{13}^2)(1 - \rho_{23}^2)}}, \tag{3}$$

where ρ_{ij} are rank correlations between rankings p^i and p^j (Lehmann, 1977). Intuitively, $\rho_{12|3}$ measures what we can learn about the first ranking from the second ranking after taking into account all knowledge from the third ranking. Note that Eq. (3) is exactly the same as the formula for the partial correlation coefficient in the standard correlation analysis (e.g., Poirier, 1995).

[4] To define a metric, the function d_{ij} must be non-negative ($d_{ij} > 0$), symmetric ($d_{ij} = d_{ji}$), and satisfy the triangle inequality: $d_{ij} \leq d_{ik} + d_{kj}$ for any i, j, and k. In general, the last condition is not trivial and needs special verification. For the distance defined by Eq. (2) its validity is ensured by the triangle inequality for the Euclidean metric defined on the vectors of rankings (see Eq. (1)). A similar definition of distance based on the standard correlation coefficient was also used by Ormerod and Mounfield (2000) and Ahn, Conrad, and Dittmar (2007).

In our tests rankings p^2 and p^3 will correspond to two different characteristics, whereas p^1 will be generated by realized stock returns. Thus, the partial rank correlation quantitatively measures what we can learn about expected returns from one characteristic controlling for the other. Clearly, if the first characteristic is redundant and brings no new information the partial correlation is zero.

Similar to the case of unconditional rank correlations, to estimate partial rank correlations in a panel we use a Fama-MacBeth-type procedure. First, we apply Eq. (3) period by period and construct a time series of rank correlations $\rho_{12|3,t}$. Reported partial rank correlations are averages of correlations computed in each time period. To calculate the t-statistics, we use the time series of $\rho_{12|3,t}$ and its standard deviation.

4 Rank correlations vs. linear regressions: comparative analysis

As we mentioned above, one of the most popular ways to discover anomalies as well as to distinguish them is to run a linear regression of realized returns on betas and characteristics. In this section, we compare it with the rank based approach and show that under some circumstances the results of the linear regressions can be misleading. The propositions we make are proved by simple examples, which although may look too abstract and artificial, clearly illustrate our main points.

To simplify the arguments, we consider the following theoretical setup. There is an infinite number of stocks with several characteristics: $X_0, X_1, \ldots$ Assume that the characteristic X_0 is uniformly distributed on the interval $[0, 1]$ (stocks are equidistantly located between 0 and 1). Other characteristics can be thought of as functions of X_0. It is convenient to identify X_0 with anomalous expected returns and think of the characteristics $X_1, X_2, \ldots$ as associated with anomalies $A_1, A_2, \ldots$. For the sake of simplicity, we abstract from all sampling errors and perform the analysis in population.

Since we have an infinite number of stocks, we need to extend the definition of the Spearman rank correlation to this case. For any fixed N among all stocks from the interval $[0, 1]$ choose N stocks located in $1/N$, $2/N$, …, $(N-1)/N$, 1. According to the characteristic X_0 such stocks are naturally ranked as $1, 2, \ldots, N$. This ranking is denoted $x^{(0)}$. Next, for any other characteristic $X_1(X_0)$ compute $X_1(1/N)$, $X_1(2/N)$, …, $X_1((N-1)/N)$, $X_1(1)$ and create a new ranking of stocks $x^{(1)}$ based on this sequence. Since both rankings are finite, the Spearman rank correlation between them is given by Eq. (1): $\rho_N = \rho(x^{(0)}, x^{(1)})$. The Spearman rank correlation for an infinite number of stocks is defined as the limit: $\rho_\infty = \lim_{N\to\infty} \rho_N$.

In population, the magnitude of the anomaly is naturally measured by the rank correlation between the characteristic and expected returns. However, the linear regression effectively delivers the Pearson correlation. As stated in Propositions 1 and 2, under some circumstances the conclusions based on the Pearson correlation may be misleading.

Proposition 1

1. High Pearson correlation between an anomalous characteristic X and (abnormal) expected returns does not imply high magnitude of the anomaly A.

2. High Pearson correlation between anomalous characteristics X_1 and X_2 does not imply small distance between anomalies A_1 and A_2.

We prove the second statement only, since the first one follows from the second if one of the characteristics is identified with expected returns. The proof hinges on the following example. Assume that the characteristic X_2 can be represented as a function of X_1 which has the following form:

$$X_2(X_1) = \begin{cases} \frac{1}{\varepsilon}\left(\frac{1}{2}+a\right)X_1, & 0 \le X_1 \le \varepsilon \\ \frac{1}{2} - \frac{2a}{1-2\varepsilon}\left(X_1 - \frac{1}{2}\right), & \varepsilon \le X_1 \le 1-\varepsilon \\ b + \frac{1}{\varepsilon}\left(a+b-\frac{1}{2}\right)(X_1-1), & 1-\varepsilon \le X_1 \le 1, \end{cases} \tag{4}$$

where $\varepsilon \in [0, 1/2]$, $a \in [0, 1/2]$, $b > 1$. A typical graph of the function $X_2(X_1)$ is plotted in Panel A of Figure 1. Essentially, this functional form assumes that both characteristics are aligned in the extremes but produce opposite rankings in the intermediate region. The Pearson correlation between characteristics X_1 and X_2 is $cov(X_1, X_2)/\sqrt{var(X_1)var(X_2)}$, where variances and the covariance are given by the following integrals:

$$cov(X_1, X_2) = \int_0^1 \left(u - \frac{1}{2}\right)(X_2(u) - \bar{X}_2)du, \tag{5}$$

$$var(X_1) = \int_0^1 \left(u - \frac{1}{2}\right)^2 du = \frac{1}{12}, \qquad var(X_2) = \int_0^1 (X_2(u) - \bar{X}_2)^2 du, \tag{6}$$

and

$$\bar{X}_2 = \int_0^1 X_2(u)du.$$

Although due to the piece-wise linearity of the function $X_2(X_1)$ all integrals above can be computed analytically, the formulas are cumbersome and specific numerical examples are more insightful. Moreover, the Spearman rank correlation is also found numerically according to the limiting procedure described above. Both correlations are plotted in Panel B of Figure 1. Specifically, we set $a = 0.01$, $b = 1$, and graph the correlations as functions of ε. Not surprisingly, due to a general increasing trend of $X_2(X_1)$ the Pearson correlation for almost all values of ε is quite high. However, the characteristics generate opposite rankings in the intermediate region $\varepsilon \le X_1 \le 1 - \varepsilon$, and this region dominates as $\varepsilon \to 0$. This is captured by the rank correlation, which can become significantly negative. Thus, for a relatively small ε a high and positive Pearson correlation is consistent with a negative Spearman correlation, which implies a large distance between anomalies defined by Eq. (2). This completes the proof.

The next proposition describes an opposite situation when the Pearson correlation (or linear regression) fails to uncover the relation between characteristics or the characteristic and returns.

Proposition 2

1. Low Pearson correlation between anomalous characteristic X and (abnormal) expected returns does not imply the weakness of the anomaly A.

2. Low Pearson correlation between anomalous characteristics X_1 and X_2 does not imply large distance between anomalies A_1 and A_2.

Similar to Proposition 1, the first statement follows from the second one by identifying one of the characteristics with expected returns, and we prove these statements using a specific functional example. Again, assume that the characteristic X_2 can be represented as a non-linear function of X_1:

$$X_2(X_1) = \begin{cases} \frac{a}{1-\varepsilon} X_1, & 0 \leq X_1 \leq 1-\varepsilon \\ \frac{1}{\varepsilon}\left(b\varepsilon + (a-b)(1-X_1)\right), & 1-\varepsilon \leq X_1 \leq 1, \end{cases} \tag{7}$$

where $\varepsilon \in (0,1)$, $a > 0$. To make the example as sharp as possible we also set b as

$$b = -a\frac{1-2\varepsilon}{\varepsilon(3-2\varepsilon)}. \tag{8}$$

It is not difficult to check that this choice guarantees a zero correlation between characteristics: $cov(X_1, X_2) = 0$. As a specific example, we choose $a = 0.5$, $\varepsilon = 0.25$ and plot the function $X_2(X_1)$ in Panel A of Figure 2. This function is non-monotonic with a positive relation between characteristics for $X_1 < 1-\varepsilon$ and negative for $X_1 > 1-\varepsilon$. Next, we compute the Spearman rank correlation for different values of ε. The results are presented in Panel B of Figure 2. Remarkably, the rank correlation captures the relation between the characteristics even though they have a zero Pearson correlation. As ε decreases and the interval with an inverse relation between characteristics shrinks, the rank correlation becomes quite high and approaches 1 in the limit $\varepsilon \to 0$. If the characteristic X_2 is interpreted as expected returns, this result admits another formulation: the characteristic that fails to predict returns in the linear regression can still be very useful for ranking stocks.

The next proposition is of a bit different nature. It emphasizes a potential failure of the multivariate linear regression to assess the difference between anomalies.

Proposition 3 *The equivalence (zero distance) between anomalies A_1 and A_2 does not imply that in the linear regression either of the characteristics X_1 and X_2 subsumes the statistical significance of the other.*

To establish this proposition, again consider an infinite number of stocks. Assume that their (rescaled) anomalous expected returns E are uniformly distributed on the interval $[-1, 1]$. There are two characteristics of the stocks X_1 and X_2, which are related to anomalous expected returns as

$$X(E; b, c) = \begin{cases} -1 + \frac{1-c}{1-b}(E+1), & -1 \leq E \leq -b \\ \frac{c}{b}E, & -b \leq E \leq b \\ 1 + \frac{1-c}{1-b}(E-1), & b \leq E \leq 1, \end{cases}, \tag{9}$$

with different constant parameters $b > 0$ and $c > 0$: $X_1 = X(E; b_1, c_1)$, $X_2 = X(E; b_2, c_2)$. Without losing generality, $b_1 < b_2$. Typical graphs of X_1 and X_2 are depicted in Figure 3. Clearly, both characteristics reflect the anomaly since a high value of either X_1 or X_2 is associated with high expected returns. Moreover, $X_1(E)$ and $X_2(E)$ are monotonically increasing, so both characteristics rank all stocks in the same way. Thus, according to our definition there is only one anomaly in this example and the anomalous characteristics X_1 and X_2 are equivalent.

However, any linear regression (including the standard Fama-MacBeth regression) of returns on these characteristics can easily fail to recognize the equivalence. As a specific example, consider the following set of parameters: $b_1 = 0.2$, $b_2 = 0.8$, $c_1 = 0.9$, $c_2 = 0.1$. First, a straightforward computation shows that the correlation between the characteristics X_1 and X_2 is far from perfect. Indeed, $E(X_1) = E(X_2) = 0$ and

$$cov(X_1, X_2) = \frac{1}{2}\int_{-1}^{1} X(E; 0.2, 0.9)\, X(E; 0.8, 0.1)\, dE = 0.146,$$

$$var(X_1) = \frac{1}{2}\int_{-1}^{1} X(E; 0.2, 0.9)^2\, dE = 0.776, \qquad var(X_2) = \frac{1}{2}\int_{-1}^{1} X(E; 0.8, 0.1)^2\, dE = 0.076,$$

so the correlation coefficient is

$$corr(X_1, X_2) = \frac{cov(X_1, X_2)}{\sqrt{var(X_1)var(X_2)}} \approx 0.6.$$

Thus, although the characteristics describe exactly the same anomaly, they appear to be not highly correlated. This is another facet of Proposition 2.

Next, the theoretical regression of expected returns on X_1 and X_2 reveals that both characteristics appear to have non-zero slopes. Indeed,

$$\beta = \begin{pmatrix} var(X_1) & cov(X_1, X_2) \\ cov(X_1, X_2) & var(X_2) \end{pmatrix}^{-1} \begin{pmatrix} cov(X_1, E) \\ cov(X_2, E) \end{pmatrix} \approx \begin{pmatrix} 0.48 \\ 0.70 \end{pmatrix}. \tag{10}$$

Hence, a linear regression produces a misleading result making a researcher to think that the anomalies based on X_1 and X_2 are different and each characteristic contributes to explaining expected returns.

Note that this result substantially relies on the non-linearity of both characteristics. Thus, it does not work in the betas vs. characteristics tests since expected returns are supposed to be proportional to betas. However, it can explain the significance of both betas and characteristics if in violation of risk-based models expected returns appear to be non-linearly related to betas.

5 Empirical results

5.1 Data

Data on stock returns, stock prices and number of shares outstanding are from CRSP monthly files, while accounting data are from Compustat Fundamentals annual files. In the characteristic definitions below the accounting variables are explained by using both Compustat Fundamentals and Industrial files notations. Financial firms (SIC code between 6000 and 6999) are excluded from the sample. Only NYSE, Amex, and NASDAQ firms with common stocks (SHCD 10 or 11) are considered. Accounting data used for construction of characteristics in calendar year t are taken from statements with the fiscal year end in year $t-1$.

Returns and risk-adjusted returns. Returns are monthly stock returns which include dividends. In our analysis, we employ returns on individual securities as opposed to returns on portfolios. To adjust for the cross-sectional variation in expected returns explained by the Fama and French (1993) model, we compute risk-adjusted returns following Brennan, Chordia, and Subrahmanyam (1998) and

Avramov and Chordia (2006). Specifically, the risk-adjusted return $\tilde{r}_{it}$ on each security i for each month t is calculated as

$$\tilde{r}_{it} = r_{it} - r_t^f - \beta_i^{MKT} \times MKT_t - \beta_i^{HML} \times HML_t - \beta_i^{SMB} \times SMB_t.$$

Individual stock betas are estimated over non-overlapping five year periods by regressing excess returns on a constant and Fama-French factors MKT, HML, and SMB.

Book-to-Market (B/M). Book-to-Market is constructed following Fama and French (1992), who define it as a logarithm of the ratio of Book Value over Market Value. Market Value is equal to Compustat CSHO×PRCH_C where CSHO is a number of shares outstanding and PRCH_C is a stock price at the end of the calendar year t-1. Book Value is defined as Compustat SEQ - (Compustat PSTKL or PSTKRV or PSTK in this order of availability) + Compustat TXDITC (if not missing), where SEQ is stockholders' equity (Compustat item 216), PSTKL is preferred stock liquidating value (Compustat item 10), PSTKRV is preferred stock redemption value (Compustat item 56), PSTK is preferred stock par value (Compustat item 130), and TXDITC is balance sheet deferred taxes and investment tax credit (Compustat item 35). If Compustat SEQ is missing, Compustat CEQ + PSTK is used, where CEQ is common equity (Compustat item 60). If all previous variables are missing, Compustat AT - LT is used, where AT is total assets (Compustat item 6), and LT is total liabilities (Compustat item 181).

Size (S). Following Fama and French (1992), Size is defined as a logarithm of market capitalization of the firm. The latter is the product of the share price (CRSP variable PRC) and the number of shares outstanding recorded at the end of the previous month (CRSP variable SHROUT).

Analysts' forecasts dispersion (D). Dispersion in analyst forecasts is from Diether, Malloy, and Scherbina (2002) and is defined as the standard deviation of IBES next quarter earnings forecast divided by mean earnings forecast.

Idiosyncratic volatility ($IdVol$). This is an anomaly from Ang, Hodrick, Xing, and Zhang (2006) who define $IdVol$ as the standard deviation of residuals in the regression of daily CRSP returns on daily Fama - French factors. In month t the idiosyncratic volatility is computed using daily data for the previous month, so $IdVol$ is updated on a monthly basis.

Total asset growth ($ASSETG$). The anomaly based on this characteristic was described by Cooper, Gulen, and Schill (2008) and we construct $ASSETG$ following this paper. Specifically, $ASSETG$ is defined as

$$ASSETG_t = \frac{AT_{t-1} - AT_{t-2}}{AT_{t-2}},$$

where AT_t is total assets (Compustat annual item 6) in fiscal year ending in calendar year t.

Abnormal capital investments (CI). Following Titman, Wei, and Xie (2004) we define the measure of abnormal capital investments as

$$CI_{t-1} = \frac{CE_{t-1}}{(CE_{t-2} + CE_{t-3} + CE_{t-3})/3} - 1,$$

where CE_t is a firm's capital expenditures (Compustat annual item 128) scaled by its sales: $CE_t = CAPX_t/SALE_t$.

Investments-to-assets ratio ($INV/ASSET$). This characteristic is from Lyandres, Sun, and Zhang (2008) and is defined as

$$INV/ASSET_t = \frac{INVT_{t-1} - INVT_{t-2} + PPEGT_{t-1} - PPEGT_{t-2}}{AT_{t-2}},$$

where $INVT_t$ is intentories (Compustat data 3), $PPEGT_t$ is gross property, plant, and equipment, and AT_t is total assets (Compustat item 6) in fiscal year ending in calendar year t.

Net Stock Issue (NS). This is the anomaly highlighted in Fama and French (2008) and Pontiff and Woodgate (2008). Net Stock Issue is defined as

$$NS_t = \log\left(\frac{SASO_{t-1}}{SASO_{t-2}}\right)$$

where $SASO_t$ is the split-adjusted shares outstanding from the fiscal year-end in t: $SASO_t = CSHO_t \times AJEX_t$ where $CSHO_t$ is the shares outstanding (Compustat item 25) and $AJEX_t$ is the cumulative factor to adjust shares (Compustat item 27) from the fiscal year-end in t.

Composite Stock Issuance (ι). This anomaly is from Daniel and Titman (2006). Composite stock issuance is defined as

$$\iota_t = \log\left(\frac{ME_{t-1}}{ME_{t-6}}\right) - r(t-6, t-1),$$

where ME_t is the Market Equity at the end of calendar year t and $r(t-\tau, t)$ is the cumulative log return between the last trading day of calendar year $t-\tau$ and the last trading day of calendar year t. Effectively, composite stock issuance is net stock issue computed for a five-year period.

5.2 Individual anomalies

In this section, we describe the empirical results pertaining individually to each of the nine characteristics described above. Although we mostly work with individual securities, to understand how the relation between returns and characteristics depends on the characteristic itself we group firms into five portfolios by sorting them on a given characteristic. For book-to-market, asset growth, abnormal capital investments, investment-to-asset ratio, net stock issues, and composite stock issuance the portfolios are formed once a year at the end of June. They are held for one year, and at the end of next June are rebalanced. Portfolios based on size, idiosyncratic volatility, and dispersion in analysts' forecasts are created at the end of each month. For all characteristics, portfolio breakpoints are determined using all stocks for which the characteristic is available at the moment of portfolio formation. The sample period is from 1965 to 2007 for all characteristics except the dispersion in analysts' forecasts, for which the sample period is from 1983 to 2007.

From the previous research we already know that all characteristics under consideration except book-to-market are supposed to be negatively related to abnormal stock returns. For the uniformity of the analysis, we switch the sign of B/M and number the portfolios so that the first portfolio has the highest return and the fifth portfolio the lowest.

After assigning stocks to portfolios, we first check that the characteristic is indeed anomalous on a large scale, i.e. that returns on the portfolios line up according to the characteristic. Since in our subsequent analysis we examine individual stocks, we consider equal-weighted portfolios.

Panel A of Table 1 reports averages of monthly raw returns on five quintile portfolios. As expected, almost all characteristics appear to be negatively related to raw stock returns. In particular, in almost all cases portfolio 1 earns substantially lower return than portfolio 5, and the difference is highly statistically significant. Although for some characteristics (B/M, D, $ASSETG$, CI, and $INV/ASSET$) there is a strictly monotonic relation between the portfolio and average returns, for others the relation is more complicated and probably not statistically significant in the intermediate range of characteristics. The only characteristic that does not produce a meaningful dispersion of returns across equal-weighted portfolios is idiosyncratic volatility, and this result is consistent with Bali and Cakici (2006).

Panel B of Table 1 shows averages of returns adjusted for the Fama-French three-factor model, which measure the cross-sectional variation in expected returns not captured by the loadings on the market, HML, and SMB factors. Although the risk adjustment substantially reduces average returns, the differences in returns on extreme portfolios stay almost the same, and even notably increase for some characteristics (D, $IdVol$, ι). As a result, they are all statistically significant confirming the conclusions of the literature that the characteristics under consideration indeed capture the pattern in non-zero alphas. Moreover, the risk adjustment preserves the monotonicity where it was observed in the case of raw returns. Surprisingly, the correction for Fama-French factors does not eliminate the dispersion of returns across size and book-to-market portfolios. This result echoes the findings of Brennan, Chordia, and Subrahmanyam (1998) and seems to be another manifestation of the sensitivity of outcome to whether the risk adjustment is conducted on the portfolio level or on the level of individual securities.

One of the main innovations of our paper is that we look inside characteristic portfolios and measure the association between characteristics and returns using rank correlations. To show that the linear correlations (which are related to the standard Fama-MacBeth regressions of returns on characteristics) and Spearman correlations can tell completely different stories about the relation between characteristics and stock returns, we report both Pearson and Spearman correlations.

Panel A of Table 2 shows the Pearson correlation coefficients between nine characteristics and raw returns for all stocks as well as for quintile portfolios. For all stocks, the results are as expected: all characteristics are negatively and statistically significantly correlated with stock returns. This is exactly what the test of anomaly based on the standard linear Fama-MacBeth regression would show, and the results are consistent with the literature. Panel B of Table 2 reports Spearman rank correlations computed as explained in Section 3.2. Strikingly, for some characteristics the results are completely different. For instance, for size and abnormal capital investments the rank correlation with returns is statistically significant and *positive*. To understand why Pearson and Spearman correlations may be so different, we also compute these correlations inside quintile portfolios.

Even a brief inspection of t-statistics for quintile portfolios reveals the following pattern. In portfolio 5 (the portfolio with supposedly low stock returns) both Spearman and Pearson correlations are negative for all characteristics and almost all are significant (except size and idiosyncratic volatility for the Pearson correlation and only size for the Spearman correlation). However, in portfolio 1 most correlations are positive. For instance, the Spearman correlation is positive for all characteristics except size and is strongly significant for $IdVol$, $ASSETG$, CI, $INV/ASSET$, and ι. It means that as we take stocks with more extreme values of characteristics (no matter high or low!) expected returns tend to decline. In the intermediate region many anomalies are not statistically detectable. This is stark evidence of non-linearity and even non-monotonicity in the relations between characteristics and returns.

The non-monotonicity can explain why the rank correlation can be positive whereas the linear correlation is negative. As shown in the example illustrating Proposition 2, if the characteristic X_2 (now interpreted as expected returns) decline in both extremes of characteristic X_1 the Spearman correlation can be high and positive whereas the Pearson correlation is zero or negative. Although our example is quite stylized, it captures what we observe in real data.

A special pattern is observed for size. It is strongly negatively related to returns in portfolio 1 containing small stocks, but the relation is opposite (and has a statistically significant Spearman correlation) for medium stocks. The intermediate region dominates in the correlation including all stocks, and this explains the positive Spearman correlation coefficient in the whole sample. Thus, we confirm the conclusion of Knez and Ready (1997) who argue that the size effect is driven by extreme positive returns on a very limited number of small stocks. When the impact of such influential points is eliminated, the relation between size and returns appears to be positive, i.e. most small firms actually do worse than larger firms.

Another interesting point is related to idiosyncratic volatility. It fails to produce a statistically significant spread in returns across portfolios (see Panel A of Table 1), but its Spearman correlation with returns is negative and statistically significant in all portfolios except the bottom, where it is positive and significant. This illustrates the difference in the power between portfolio returns based approach and the rank-based approach.

Despite also showing the tendency to be positive in bottom portfolios and negative in top portfolios, Pearson correlations in general are less conclusive than Spearman correlations and many of them (beyond portfolio 5) are not statistically significant. Thus, Spearman correlations appear to be a more sensitive tool for examining individual securities inside portfolios.

In Table 3 we repeat the analysis for returns adjusted for the Fama-French three-factor model. Under the null hypothesis of exact pricing, the predictive ability of characteristics in the cross-section should be statistically insignificant. However, qualitatively all conclusions are very similar to those obtained for raw returns even for size and book-to-market. Thus, we confirm the results of Brennan, Chordia, and Subrahmanyam (1998) who demonstrate that the correction for the Fama-French factors cannot eliminate the size and book-to-market effects. The risk adjustment reduces the Spearman t-statistics for asset growth and investments-to-assets ratio, making then insignificant in the whole sample. However, they are still significant in the quintile portfolios with a very strong negative sign in the top portfolio and a strong positive sign in the bottom portfolio. As for the Pearson correlations, they all preserve their signs and significance when computed for all stocks, and many of them switch the sign in portfolio 1.

We argue that the majority of anomalies are pronounced in the extremes, and the overall relation between characteristics and returns is determined by portfolio 5. However, portfolio 5 typically has the largest dispersion of the characteristic, which makes the detection of the anomaly easier. To check if the anomalies are indeed stronger among stocks in the right tail of the characteristic distribution, we examine two portfolios that are designed to have similar cross-sectional dispersions of characteristics. Every period we sort all stocks according to the characteristic and discard top 5% of them. This eliminates highly unusual stocks and reduces the overall dispersion of the characteristic. Then, we break the rest of the stocks into two portfolios, which are referred to as Top Portfolio and Bottom Portfolio. The Bottom Portfolio must contain at least 50% of stocks and have the cross-sectional dispersion of the characteristic as close as possible to that of the Top Portfolio. Various statistics for the resulting portfolios are reported in Table 4.

As desired, for most characteristics the volatilities of dispersion in both portfolios have the same order. The exception is book-to-market, which is significantly more dispersed in the Bottom Portfolio. Except for size and book-to-market, the Top Portfolio contains fewer stocks indicating positive skewness of the characteristics. Consistent with Table 1, excluding idiosyncratic volatility the average return on the Bottom Portfolio appears to be higher than the return on the Top Portfolio.

The comparison of the Spearman and Pearson correlations across portfolios tend to confirm our claim that the overall relation between many anomaly variables and risk-adjusted stocks returns is produced mostly by stocks with high values of characteristics. In the Top Portfolio, the Spearman correlations are negative and significant for book-to-market, idiosyncratic volatility, asset growth, abnormal capital investments, investments-to-assets ratio, and composite stock issuance, but only idiosyncratic volatility and composite stock issuance preserve the sign and statistical significance in the Bottom Portfolio. Only net stock issues and analysts' forecasts dispersion gain statistically significant negative coefficients in the Bottom Portfolio being insignificant in the Top Portfolio. This is consistent with results in Table 3, which reports positive Spearman correlations in the first portfolio only. The comparison of Pearson correlations is also quite compelling: $IndVol$, $ASSETG$, CI, $INV/ASSET$, and ι lose their significance in the Bottom Portfolio.

Overall, we can conclude that all considered characteristics are indeed anomalous, but anomalies are non-monotonic, prevail in the extreme portfolios only and have opposite signs in the opposite extremes.

5.3 Distance between anomalies

Now we turn to the analysis of relations between different anomalies and focus on characteristics per se putting aside their ability to predict returns. Table 5 reports pairwise Spearman correlations between characteristics computed as described in Section 3.2. In Panel A, for each pair of characteristics the correlations are calculated using all stocks for which both characteristics are available at the given moment of time. To examine which stocks contribute more significantly to the correlations, we also consider quintile portfolios formed individually for each characteristic and report the results for portfolios 1, 3, and 5 in Panels B, C, and D, respectively. For example, in Panel B the column corresponding to B/M contains the Spearman correlations between B/M and other characteristics computed using the stocks from the bottom quintile sorted by book-to-market. Note that we do not require the stocks to be in the bottom quintile with respect to the other characteristic with which the correlation is computed. This convention explains why the correlation matrices in Panels B, C, and D are not symmetric.

The first observation from Table 5 is that the vast majority of t-statistics are strongly significant indicating that characteristics are correlated with each other. However, the correlation coefficients are relatively large for all stocks, quite substantial for portfolios 1 and 5, and rather small for portfolio 3.

This indicates the tendency of characteristics to be correlated on the unusual stocks and two characteristics which move together in the extreme portfolios may decouple in the intermediate portfolios. For instance, investments-to-assets ratio has a very strong correlation with asset growth in portfolio 5 (0.65) and in the whole sample (0.73), but a relatively weak correlation in portfolio 3 (0.17). Although all characteristics are supposed to be negatively related to stock returns, the pairwise correlations between them exhibit both signs. Moreover, many correlations switch the sign from one portfolio to another also indicating an instability of the relations between characteristics. The Spearman correlations reported in Table 5 admit a nice visualization. As described in Section 3.3, we can use them to introduce the distance between characteristics defined by Eq. (2). This distance converts the space of all anomalies into a well-defined metric space and allows us to examine how far each anomaly is located from the others and if different anomalies form clusters. Since we have only a very limited number of anomalies, we use a hierarchical clustering. As a result, we report the whole hierarchical tree which is effectively a hierarchy of clusters and shows how different clusters at a lower level are joined into a single cluster at the higher level. Reporting the whole tree instead of the structure of clusters enables us to preserve as much information as possible and to avoid a tricky question about the appropriate number of clusters.

Figure 4 shows the hierarchical trees of clusters computed for all stocks as well as for five quintile portfolios. Along the horizontal axis we plot the distance between anomalies and clusters. Eq. (2) implies that the distance between uncorrelated characteristics is 1. Although the distance described in Section 3.3 directly applies to the case of all stocks, it should be augmented with additional conventions when computed for portfolios. Indeed, the correlation matrices for portfolios in Table 5 are non-symmetric and depend on which characteristic was used for sorting. Correspondingly, for each pair of characteristics we can introduce two types of distances (when stocks are sorted by each characteristic) and define the metric on the anomaly space for portfolios as an average of these distances.

Figure 4 confirms our conclusion that the correlation between characteristics is mostly driven by extreme stocks (portfolios 1 and 5). In these portfolios many distances are less than 1 and the structure of clusters is more pronounced. In the intermediate portfolios 2, 3, and 4 the anomalies are located much more equidistantly and clusters are blurred indicating the lack of commonality between characteristics.

The hierarchical tree for portfolio 5, which as we have already shown is crucial for almost all anomalies, reveals interesting patterns. The asset growth related characteristics ι, NS, $INV/ASSET$, and $ASSETG$ represent one cluster indicating that these anomalies may have much in common. Notably, $INV/ASSET$ and $ASSETG$ represent the closest pair and always form a cluster. Another quite stable pair is formed by idiosyncratic volatility and dispersion of analysts' forecasts.

5.4 Partial rank correlations between characteristics and risk-adjusted returns

The cluster analysis gives a very good idea about how differently various characteristics rank stocks and how far from each other anomalies are. However, it is silent about the additional information on abnormal stock returns each characteristic contains relative to others. For instance, it can be the case that two characteristics are different in terms of stock ranking, but nevertheless contain the same information about returns. Hence, having one of them we do not need to use the other. To test if this is indeed the case we compute partial Spearman correlations as explained in Section 3.4.

The results are reported in Table 6. To save space, we show the results for the risk-adjusted returns only. As in the previous section, we find partial correlations for all stocks and quintile portfolios. In Panel A, the partial correlation between a characteristic and abnormal returns conditioned on another characteristic is computed using all stocks for which both characteristics are available at the given moment. Panels B, C, and D report partial correlations for quintile portfolios 1, 3, and 5, respectively. Portfolios are formed by sorting stocks with respect to the characteristic which correlation with returns is computed (not with respect to the characteristic upon which the correlation is conditioned). In all panels, rows correspond to the conditioning variables. For example, in Panel B the first row reports correlations between various anomaly variables and risk-adjusted returns conditioned on B/M.

Table 6 reveals several interesting patterns. First, t-statistics in Panel D indicate that high values of all anomalies (except size) can bring new information about alphas even after controlling for the peers. Thus, in this region (and this is exactly the region where anomalies are especially strong) they are statistically different despite higher correlation between anomalous variables themselves. Probably, the only notable exclusion is $INV/ASSET$, which is subsumed by $ASSETG$. Remarkably, as demonstrated by our cluster analysis, these two anomalies unambiguously form a cluster.

Second, confirming our previous results t-statistics in portfolio 3 for the majority of anomaly variables are low and the correlation coefficients are insignificant. The only exception is ι, which is significant even after controlling for many other characteristics.

Third, from Table 3 we know that B/M, $IdVol$, $ASSETG$, CI, $INV/ASSET$, and ι are positively related to returns in Portfolio 1. Panel B of 6 shows that $ASSETG$, CI, and $INV/ASSET$ mostly preserve this property in the conditional correlations. For B/M and $IdVol$ the results are mixed and the conditioning can change their relation to stock returns.

Fourth, the story becomes more complicated if we consider all stocks (Panel A). D, $IdVol$, NS, and ι are statistically significant with correct signs even after controlling for other characteristics and unambiguously represent different anomalies. In consistency with the results in Tables 2 and 3, size appears to be positively related to returns in the whole sample and this result is not invalidated by conditioning. The evidence for $ASSETG$, CI, and $INV/ASSET$ is mixed. Conditioning substantially affects their ability to track alphas, and in some cases after taking into account other variables their correlation coefficients switch the sign. This is another piece of evidence for necessity to look at various ranges of anomalous characteristics and examine them separately.

6 Conclusion

The main message of this paper is that many anomalous characteristics are related to stock returns non-monotonically. Thus, it is incorrect to think about anomalies in terms of a simple increasing or decreasing relation between some variables and expected stock returns. Moreover, linear correlations and rank correlations can tell us different stories, and in the presence of potential non-monotonicity the latter are more appropriate for the analysis of anomalies. The introduced anomaly space augmented with the metric seems to be a very convenient framework for anomaly unification, along with the cluster analysis which allows us to visualize the classification of existing anomalies. Overall, a rank-based non-parametric analysis can cast new light on the structure of asset pricing anomalies and may bring us closer to understanding of their origin.

In this paper, we considered only nine anomaly variables. It would be interesting to extend the analysis to other documented anomalies and examine if they also reverse their predictions in the bottom portfolio. Also, it would be fruitful to put other documented anomalies in the constructed anomaly space and see if they are really different and located distantly from other anomalies. We leave this analysis for future research.

References

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.

Ahn, Dong-Hyun, Jennifer Conrad, and Robert F. Dittmar, 2007, Basis assets, *Review of Financial Studies*, forthcoming.

Ang, Andrew, Robert Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259-99.

Ang, Andrew, Jun Liu, and Krista Schwarz, 2008, Using stocks or portfolios in tests of factor models, Working paper.

Avramov, Doron, Scott Cederburg, and Satadru Hore, 2009, Cross-sectional asset pricing puzzles: An equilibrium perspective, Working paper.

Avramov, Doron and Tarun Chordia, 2006, Asset pricing models and financial market anomalies, *Review of Financial Studies* 19, 1001-1040.

Bali, Turan G. and Nusret Cakici, 2006, Idiosyncratic volatility and the cross-section of expected returns, *Journal of Financial and Quantitative Analysis*, forthcoming.

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

Berk, Jonathan B., Richard C. Green, and Vasant Naik, 1999, Optimal Investment, Growth Options, and Security Returns, *Journal of Finance* 54, 1553-1607.

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

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.

Boyer, Brian, Todd Mitton, and Keith Vorkink, 2009, Expected idiosyncratic skewness, *Review of Financial Studies*, forthcoming.

Cao, Jie, 2009, Idiosyncratic risk, costly arbitrage, and the cross-section of stock returns, Working paper.

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.

Diether, Karl, Christopher J. Malloy, and Anna Scherbina, 2002, Difference of opinion and the cross-section of stock returns, *Journal of Finance* 57, 21132141.

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

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

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

Fama, Eugene F., and Kenneth R. French, 2008, Dissecting anomalies, *Journal of Finance* 63, 1653-1677.

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

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121-1152.

Gomes, Joao, Leonid Kogan, and Lu Zhang, 2003, Equilibrium Cross Section of Returns, *The Journal of Political Economy* 111, 693-732.

Hájek, Jaroslav, Zbyněk Šidák, and Pranab K. Sen, 1999, Theory of rank tests, Academic Press, San Diego.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65-92.

Johnson, Timothy C., 2004, Forecast dispersion and the cross section of expected returns, *Journal of Finance* 59, 1957-1978.

Knez, Peter J. and Mark J. Ready, 1997, On the robustness of size and book-to-market in cross-sectional regressions, *Journal of Finance* 52, 1355-1382.

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

Lehmann, R., 1977, General derivation of partial and multiple rank correlation coefficients, *Biometrical Journal* 19, 229-236.

Li, Dongmei, and Lu Zhang, 2009, Do investment frictions affect anomalies in the cross-section of returns? Working paper.

Li, Erica X. N., Dmitry Livdan, and Lu Zhang, 2009, Anomalies, *Review of Financial Studies* 22, 4301-4334.

Lipson, Marc L., Sandra Mortal, and Michael J. Schill, 2009, What explains the asset growth effect in stock returns? Working paper.

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

Lyanders, Evgeny, Le Sun, and Lu Zhang, 2008, The new issues puzzle: Testing the investment-based explanation, *Review of Financial Studies* 21, 2825-2855.

Kendall, Maurice G., 1970, Rank correlation methods, London, Griffin.

Newey, Whitney K. and Kenneth D. West, 1987, A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.

Ormerod, Paul and Craig Mounfield, 2000, Localised structures in the temporal evolution of asset prices, in New Approaches to Financial Economics, Santa Fe Conference.

Patton Andrew J. and Allan Timmermann, 2009, Monotonicity in asset returns: New tests with applications to the term structure, the CAPM and portfolio sorts, Working paper.

Poirier, Dale J., 1995, Intermediate statistics and econometrics, MIT Press, Cambridge, Massachusetts.

Pontiff, Jeffrey, 2006, Costly arbitrage and the myth of idiosyncratic risk, *Journal of Accounting and Economics* 42, 35-52.

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

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive Evidence of Market Inefficiency, *Journal of Portfolio Management* 11, 9-16.

Subrahmanyam, Avanidhar, 2009, The cross-section of expected stock returns: What have we learnt from the past twenty-five years of research? Working paper.

Table 1
Portfolio Raw Returns and Fama-French Risk-Adjusted Returns

This table shows averages of monthly equal-weighted stocks returns (Panel A) and returns adjusted using the Fama-French 3-factor model (Panel B) for five quintile portfolios constructed using nine anomalous characteristics. B/M is book-to-market, S is size, D is analysts' forecasts dispersion, $IdVol$ is idiosyncratic volatility, $ASSETG$ is total asset growth, CI is abnormal capital investments, $INV/ASSET$ is investments-to-assets ratio, NS is net stock issues, ι is composite stock issuance. More detailed description of characteristics is given in Section 5.1. The column (5-1) refers to the difference between returns on portfolio 5 and portfolio 1. The sample covers the period from January 1965 to December 2007 for all characteristics except the analysts' forecasts dispersion for which the sample is January 1983 - December 2007. All coefficients are multiplied by 100.

Panel A: Raw Returns

	Coefficient						t-stats					
	1	2	3	4	5	(5-1)	1	2	3	4	5	(5-1)
B/M	1.73	1.48	1.31	1.15	0.87	0.86	8.11	6.77	5.35	4.03	2.46	3.93
S	2.12	1.06	1.08	1.13	1.02	1.09	6.31	3.71	3.94	4.36	4.76	4.23
D	1.39	1.31	1.21	1.08	0.69	0.71	5.33	4.66	3.85	3.00	1.64	2.85
$IdVol$	1.17	1.40	1.42	1.28	1.14	0.03	7.39	6.64	5.45	4.00	2.81	0.09
$ASSETG$	1.80	1.51	1.36	1.26	0.75	1.05	5.49	6.58	6.10	5.05	2.37	7.81
CI	1.52	1.43	1.36	1.32	1.23	0.29	5.17	5.73	5.87	5.63	4.62	3.78
$INV/ASSET$	1.73	1.51	1.42	1.24	0.86	0.87	5.87	6.11	6.13	4.97	2.92	9.00
NS	1.34	1.35	1.40	1.13	0.72	0.63	5.70	5.41	5.10	3.92	2.29	5.03
ι	1.44	1.38	1.41	1.41	1.07	0.38	7.66	6.57	5.75	4.72	3.34	2.07

Panel B: Risk-Adjusted Returns

	Coefficient						t-stats					
	1	2	3	4	5	(5-1)	1	2	3	4	5	(5-1)
B/M	0.42	0.23	0.15	0.10	-0.05	0.47	6.20	4.10	2.45	1.37	-0.43	4.25
S	0.88	-0.03	-0.03	0.00	-0.04	0.92	4.81	-0.27	-0.45	0.03	-0.78	4.58
D	0.26	0.15	0.05	-0.04	-0.45	0.72	2.74	1.78	0.60	-0.39	-3.12	4.34
$IdVol$	0.13	0.24	0.27	0.14	0.00	0.13	2.43	5.09	4.83	1.76	0.00	0.65
$ASSETG$	0.47	0.28	0.21	0.11	-0.25	0.72	3.37	3.93	3.84	1.77	-2.74	6.84
CI	0.29	0.24	0.19	0.14	0.02	0.27	2.62	3.23	3.02	2.26	0.29	3.68
$INV/ASSET$	0.38	0.27	0.23	0.13	-0.19	0.58	3.36	3.58	4.10	1.89	-2.36	6.56
NS	0.14	0.20	0.21	0.05	-0.35	0.48	2.05	3.02	2.85	0.59	-3.11	5.88
ι	0.21	0.19	0.25	0.19	-0.18	0.39	4.46	3.49	3.27	2.03	-1.53	3.37

Table 2
Pearson and Spearman Correlations Between Characteristics and Raw Stock Returns

This table reports Pearson correlations (Panel A) and Spearman correlations (Panel B) between several anomalous characteristics and raw stock returns for five quintile portfolios. The column All reports correlations between characteristics and stock returns for the whole sample. B/M is book-to-market, S is size, D is analysts' forecasts dispersion, $IdVol$ is idiosyncratic volatility, $ASSETG$ is total asset growth, CI is abnormal capital investments, $INV/ASSET$ is investments-to-assets ratio, NS is net stock issues, ι is composite stock issuance. More detailed description of characteristics can be found in Section 5.1. The sample covers the period from January 1965 to December 2007 for all characteristics except the analysts' forecasts dispersion for which the sample is January 1983 - December 2007.

Panel A: Pearson Correlation

	Coefficient						t-stats					
	1	2	3	4	5	All	1	2	3	4	5	All
B/M	-0.0074	-0.0044	-0.0069	-0.0080	-0.0122	-0.0320	-2.93	-2.15	-3.21	-3.73	-4.41	-7.12
S	-0.0539	0.0018	0.0019	0.0021	-0.0085	-0.0146	-16.82	0.94	0.89	1.00	-1.95	-2.78
D	-0.0012	-0.0010	-0.0006	-0.0086	-0.0116	-0.0118	-0.31	-0.41	-0.20	-3.16	-3.95	-5.22
$IdVol$	0.0110	0.0017	-0.0043	-0.0059	-0.0051	-0.0164	2.84	0.76	-1.98	-2.87	-1.51	-2.84
$ASSETG$	0.0003	-0.0002	-0.0036	-0.0071	-0.0207	-0.0187	0.08	-0.09	-1.80	-3.43	-8.05	-9.23
CI	0.0010	0.0017	-0.0009	-0.0001	-0.0080	-0.0047	0.29	0.69	-0.36	-0.06	-2.68	-3.65
$INV/ASSET$	-0.0030	-0.0004	-0.0012	-0.0037	-0.0187	-0.0196	-0.97	-0.16	-0.59	-1.65	-6.62	-10.49
NS	0.0026	-0.0008	-0.0025	-0.0115	-0.0080	-0.0163	1.09	-0.36	-1.05	-4.54	-2.50	-7.66
ι	0.0039	-0.0011	-0.0013	0.0051	-0.0098	-0.0173	1.17	-0.44	-0.50	1.66	-3.43	-5.40

Panel B: Spearman Correlation

	Coefficient						t-stats					
	1	2	3	4	5	All	1	2	3	4	5	All
B/M	0.0044	-0.0049	-0.0108	-0.0157	-0.0358	-0.0518	1.75	-2.36	-4.98	-7.10	-12.12	-10.45
S	-0.0094	0.0127	0.0123	0.0101	-0.0013	0.0464	-3.23	6.22	5.62	4.65	-0.28	7.92
D	0.0036	-0.0042	-0.0042	-0.0154	-0.0239	-0.0466	1.12	-1.62	-1.45	-5.69	-7.09	-8.37
$IdVol$	0.0182	-0.0054	-0.0156	-0.0217	-0.0494	-0.0761	3.86	-2.32	-7.17	-10.81	-17.84	-11.80
$ASSETG$	0.0335	0.0096	-0.0028	-0.0106	-0.0374	-0.0060	8.46	4.04	-1.44	-4.94	-12.79	-2.04
CI	0.0235	0.0085	0.0007	-0.0029	-0.0286	0.0051	7.21	3.47	0.28	-1.26	-9.40	3.12
$INV/ASSET$	0.0167	0.0051	-0.0004	-0.0068	-0.0280	-0.0100	5.59	2.22	-0.20	-2.94	-10.09	-4.28
NS	0.0025	-0.0001	-0.0028	-0.0157	-0.0251	-0.0320	1.05	-0.05	-1.23	-6.07	-8.14	-11.40
ι	0.0080	-0.0113	-0.0089	0.0046	-0.0255	-0.0455	2.86	-4.50	-3.37	1.37	-9.51	-10.06

Table 3
Pearson and Spearman Correlations Between Characteristics and Fama-French Risk-Adjusted Returns

This table reports Pearson correlations (Panel A) and Spearman correlations (Panel B) between several anomalous characteristics and returns adjusted for risk using Fama-French 3-factor model for five quintile portfolios. The column All reports correlations between characteristics and risk-adjusted returns for the whole sample. B/M is book-to-market, S is size, D is analysts' forecasts dispersion, $IdVol$ is idiosyncratic volatility, $ASSETG$ is total asset growth, CI is abnormal capital investments, $INV/ASSET$ is investments-to-assets ratio, NS is net stock issues, ι is composite stock issuance. More detailed description of characteristics and risk-adjusted returns can be found in Section 5.1. The sample covers the period from January 1965 to December 2007 for all characteristics except the analysts' forecasts dispersion for which the sample is January 1983 - December 2007.

Panel A: Pearson Correlation

	Coefficient						t-stats					
	1	2	3	4	5	All	1	2	3	4	5	All
B/M	-0.0049	0.0007	-0.0033	-0.0041	-0.0069	-0.0148	-1.92	0.34	-1.71	-1.88	-2.30	-5.43
S	-0.0486	0.0020	0.0009	-0.0011	0.0034	-0.0149	-14.40	0.94	0.39	-0.51	1.34	-3.81
D	-0.0056	-0.0035	0.0038	-0.0079	-0.0125	-0.0129	-1.63	-1.18	1.26	-2.54	-4.25	-5.61
$IdVol$	0.0089	0.0038	-0.0024	-0.0047	-0.0049	-0.0128	2.92	1.90	-1.25	-2.22	-1.32	-3.01
$ASSETG$	0.0008	-0.0012	-0.0028	-0.0055	-0.0163	-0.0130	0.25	-0.51	-1.28	-2.70	-6.25	-7.32
CI	0.0010	0.0033	-0.0015	0.0000	-0.0044	-0.0054	0.31	1.33	-0.62	-0.02	-1.36	-3.83
$INV/ASSET$	-0.0014	0.0013	-0.0019	-0.0047	-0.0123	-0.0117	-0.49	0.59	-0.81	-2.13	-4.33	-6.27
NS	0.0009	0.0009	0.0009	-0.0086	-0.0073	-0.0129	0.36	0.35	0.37	-3.38	-2.30	-6.90
ι	0.0025	0.0012	-0.0010	0.0022	-0.0093	-0.0132	0.82	0.49	-0.43	0.84	-3.52	-5.90

Panel B: Spearman Correlation

	Coefficient						t-stats					
	1	2	3	4	5	All	1	2	3	4	5	All
B/M	0.0076	0.0010	-0.0052	-0.0085	-0.0249	-0.0246	3.12	0.48	-2.60	-3.70	-8.11	-9.17
S	-0.0067	0.0113	0.0074	0.0052	0.0084	0.0395	-2.25	4.93	3.15	2.40	3.17	9.15
D	-0.0010	-0.0059	-0.0005	-0.0129	-0.0270	-0.0438	-0.33	-2.00	-0.15	-4.19	-6.91	-9.65
$IdVol$	0.0116	-0.0027	-0.0105	-0.0179	-0.0452	-0.0650	3.81	-1.35	-5.51	-8.57	-14.94	-17.36
$ASSETG$	0.0277	0.0062	-0.0010	-0.0081	-0.0258	0.0016	8.95	2.55	-0.47	-3.99	-10.13	0.69
CI	0.0221	0.0105	0.0008	-0.0024	-0.0251	0.0039	7.25	4.33	0.32	-1.03	-8.50	2.39
$INV/ASSET$	0.0161	0.0090	-0.0007	-0.0078	-0.0190	0.0000	6.23	3.89	-0.32	-3.53	-7.46	0.00
NS	0.0017	0.0013	0.0010	-0.0127	-0.0187	-0.0231	0.66	0.52	0.44	-4.88	-6.92	-11.71
ι	0.0086	-0.0079	-0.0082	0.0002	-0.0232	-0.0362	3.35	-3.17	-3.41	0.08	-9.35	-12.76

Table 4
Portfolios With Comparable Cross-Sectional Dispersion of Characteristics

This table shows various statistics for portfolios designed to have comparable cross-sectional dispersion of characteristics. Every period we sort all stocks according to the characteristic and break the bottom 95% of them into two portfolios, which are referred to as Top Portfolio and Bottom Portfolio. The Bottom Portfolio must contain at least 50% of stocks and have the cross-sectional dispersion of each characteristic as close as possible to that of the Top Portfolio. To form portfolios, we use nine anomaly variables: B/M is book-to-market, S is size, D is analysts' forecasts dispersion, $IdVol$ is idiosyncratic volatility, $ASSETG$ is total asset growth, CI is abnormal capital investments, $INV/ASSET$ is investments-to-assets ratio, NS is net stock issues, ι is composite stock issuance. For each portfolio the table reports the cross-sectional volatility of the characteristic, the fraction of firms in the portfolio, average return on the portfolio, Spearman and Pearson correlations between characteristics and risk-adjusted stock returns inside portfolios as well as their t-statistics. The sample covers the period from January 1965 to December 2007 for all characteristics except the analysts' forecasts dispersion for which the sample is January 1983 - December 2007.

Panel A: Top Portfolio

	Volatility	Fraction of Firms	Average return	Spearman correlation	Spearman t-stat	Pearson correlation	Pearson t-stat
B/M	0.1605	0.5001	0.0106	-0.0208	-5.7944	-0.0173	-5.6029
S	0.9847	0.4972	0.0109	0.0201	4.7589	0.0009	0.2162
D	0.0943	0.0477	0.0097	-0.0085	-1.4659	-0.0038	-0.6454
$IdVol$	0.0087	0.2663	0.0133	-0.0294	-12.0077	-0.0053	-2.1813
$ASSETG$	0.1548	0.3520	0.0096	-0.0228	-7.7459	-0.0154	-5.5169
CI	0.3490	0.2902	0.0131	-0.0112	-4.6235	-0.0054	-2.3052
$INV/ASSET$	0.0804	0.4886	0.0118	-0.0210	-7.9547	-0.0159	-6.4220
NS	0.0492	0.0881	0.0071	-0.0051	-1.3230	-0.0016	-0.4155
ι	0.2212	0.4678	0.0130	-0.0201	-7.4847	-0.0116	-4.6445

Panel B: Bottom Portfolio

	Volatility	Fraction of Firms	Average return	Spearman correlation	Spearman t-stat	Pearson correlation	Pearson t-stat
B/M	2.9583	0.4999	0.0162	0.0137	4.2061	-0.0142	-5.5287
S	0.8949	0.5028	0.0154	0.0197	5.8849	-0.0246	-7.6314
D	0.0758	0.9523	0.0111	-0.0263	-4.9447	-0.0097	-2.0028
$IdVol$	0.0070	0.7337	0.0127	-0.0292	-5.6135	-0.0064	-1.3320
$ASSETG$	0.1350	0.6480	0.0156	0.0268	7.7337	-0.0069	-1.8819
CI	0.2793	0.7098	0.0142	0.0244	8.5435	-0.0001	-0.0386
$INV/ASSET$	0.1032	0.5114	0.0159	0.0183	6.3078	-0.0057	-1.9661
NS	0.0395	0.9119	0.0127	-0.0191	-7.5201	-0.0150	-7.2995
ι	0.3132	0.5322	0.0140	-0.0169	-6.0047	-0.0027	-1.3046

Table 5
Spearman Correlations Between Characteristics

This table collects time series averages of Spearman correlations between nine characteristics and their t-statistics. B/M is book-to-market, S is size, D is analysts' forecasts dispersion, $IdVol$ is idiosyncratic volatility, $ASSETG$ is total asset growth, CI is abnormal capital investments, $INV/ASSET$ is investments-to-assets ratio, NS is net stock issues, ι is composite stock issuance. Panel A shows correlations computed using all available stocks for each pair of characteristics. Each column of Panels B, C, and D reports Spearman correlations between the chosen characteristic and other characteristics computed for stocks with the chosen characteristic in the bottom quintile, middle quintile, and top quintile, respectively. The sample covers the period from January 1965 to December 2007 for all characteristics except the analysts' forecasts dispersion for which the sample is January 1983 - December 2007.

Panel A: All stocks

	Spearman correlation									t-stats								
	B/M	S	D	$IdVol$	$ASSETG$	CI	$INV/ASSET$	NS	ι	B/M	S	D	$IdVol$	$ASSETG$	CI	$INV/ASSET$	NS	ι
B/M	1.00	0.07	0.02	0.22	0.25	0.04	0.23	0.18	0.31	0.0	11.2	3.5	37.7	57.8	12.9	60.2	47.9	95.6
S	0.07	1.00	-0.23	-0.45	0.18	0.11	0.15	-0.08	-0.12	11.2	0.0	-92.5	-79.2	55.6	40.5	55.7	-28.5	-22.4
D	0.02	-0.23	1.00	0.29	-0.10	-0.06	-0.05	0.03	0.20	3.5	-92.5	0.0	124.8	-25.4	-20.4	-15.6	11.5	69.4
$IdVol$	0.22	-0.45	0.29	1.00	-0.05	-0.08	-0.04	0.14	0.32	37.7	-79.2	124.8	0.0	-11.7	-30.3	-13.3	40.0	92.2
$ASSETG$	0.25	0.18	-0.10	-0.05	1.00	0.26	0.73	0.28	0.14	57.8	55.6	-25.4	-11.7	0.0	101.2	264.4	64.1	38.0
CI	0.04	0.11	-0.06	-0.08	0.26	1.00	0.35	0.02	-0.03	12.9	40.5	-20.4	-30.3	101.2	0.0	173.3	6.6	-15.6
$INV/ASSET$	0.23	0.15	-0.05	-0.04	0.73	0.35	1.00	0.20	0.12	60.2	55.7	-15.6	-13.3	264.4	173.3	0.0	54.5	37.2
NS	0.18	-0.08	0.03	0.14	0.28	0.02	0.20	1.00	0.38	47.9	-28.5	11.5	40.0	64.1	6.6	54.5	0.0	80.7
ι	0.31	-0.12	0.20	0.32	0.14	-0.03	0.12	0.38	1.00	95.6	-22.4	69.4	92.2	38.0	-15.6	37.2	80.7	0.0

Panel B: Portfolio 1

	Spearman correlation									t-stats								
	B/M	S	D	$IdVol$	$ASSETG$	CI	$INV/ASSET$	NS	ι	B/M	S	D	$IdVol$	$ASSETG$	CI	$INV/ASSET$	NS	ι
B/M	1.00	0.05	-0.01	0.08	-0.24	-0.09	-0.07	0.05	0.01	0.0	17.4	-1.7	30.1	-46.3	-18.6	-25.8	17.6	1.7
S	0.20	1.00	0.25	0.16	0.23	0.22	0.16	-0.01	0.16	52.9	0.0	53.5	15.5	54.5	78.2	56.1	-3.5	58.9
D	-0.10	-0.15	1.00	0.06	-0.17	-0.14	-0.14	-0.05	-0.06	-25.5	-45.0	0.0	19.7	-40.8	-38.1	-33.4	-13.7	-13.9
$IdVol$	-0.08	-0.15	-0.13	1.00	-0.24	-0.16	-0.16	0.01	-0.08	-25.0	-20.2	-29.7	0.0	-64.1	-56.2	-55.0	5.8	-26.6
$ASSETG$	0.15	0.14	0.02	0.09	1.00	0.14	0.39	0.05	0.09	72.4	58.4	4.7	29.0	0.0	50.4	122.8	21.3	27.0
CI	0.07	0.05	0.02	0.03	0.10	1.00	0.03	0.00	0.02	25.6	17.4	4.7	8.8	36.2	0.0	10.4	-1.2	6.0
$INV/ASSET$	0.15	0.11	0.02	0.06	0.37	0.17	1.00	0.03	0.07	61.2	47.5	5.6	20.2	64.8	68.2	0.0	11.2	23.1
NS	0.08	-0.05	0.01	0.02	-0.13	-0.09	-0.08	1.00	-0.04	22.1	-18.7	2.1	5.3	-48.6	-20.1	-26.0	0.0	-11.1
ι	0.02	-0.04	-0.03	0.07	-0.17	-0.12	-0.12	0.04	1.00	6.7	-15.6	-9.1	17.8	-50.3	-32.2	-41.0	12.8	0.0

Table 5 (continued)

Panel C: Portfolio 3

	Spearman correlation									t-stats								
	B/M	S	D	$IdVol$	$ASSETG$	CI	$INV/ASSET$	NS	ι	B/M	S	D	$IdVol$	$ASSETG$	CI	$INV/ASSET$	NS	ι
B/M	1.00	0.01	0.00	0.05	0.09	0.01	0.04	0.03	0.10	0.0	4.9	-0.7	22.2	39.6	3.0	19.9	11.9	37.4
S	0.00	1.00	-0.05	-0.11	0.02	0.03	0.02	-0.01	-0.05	0.5	0.0	-22.4	-61.0	10.6	8.6	8.0	-4.5	-9.4
D	0.00	-0.04	1.00	0.05	-0.02	-0.02	0.01	0.00	0.03	0.8	-16.5	0.0	18.1	-5.7	-4.6	1.7	0.0	9.0
$IdVol$	0.06	-0.09	0.06	1.00	0.02	-0.02	0.00	0.03	0.09	34.2	-39.9	23.4	0.0	7.1	-7.3	-1.1	13.0	22.7
$ASSETG$	0.05	0.04	-0.02	0.00	1.00	0.05	0.17	0.05	0.03	25.9	19.5	-6.1	-2.2	0.0	23.0	60.0	18.4	8.7
CI	0.01	0.03	-0.01	-0.02	0.07	1.00	0.07	0.00	-0.01	3.2	12.9	-1.7	-6.7	28.1	0.0	27.7	0.8	-3.2
$INV/ASSET$	0.04	0.03	-0.01	0.00	0.20	0.06	1.00	0.03	0.00	21.4	16.8	-2.5	-1.1	100.0	24.9	0.0	13.0	1.4
NS	0.02	-0.01	0.00	0.03	0.06	0.00	0.04	1.00	0.11	9.2	-3.2	0.4	13.8	21.0	-1.8	17.9	0.0	24.4
ι	0.07	-0.02	0.04	0.06	0.05	-0.01	0.02	0.07	1.00	25.3	-8.1	10.8	27.2	20.9	-4.7	8.0	22.7	0.0

Panel D: Portfolio 5

	Spearman correlation									t-stats								
	B/M	S	D	$IdVol$	$ASSETG$	CI	$INV/ASSET$	NS	ι	B/M	S	D	$IdVol$	$ASSETG$	CI	$INV/ASSET$	NS	ι
B/M	1.00	0.06	0.03	0.05	0.18	0.06	0.15	0.04	0.09	0.0	17.1	8.7	18.5	51.5	14.9	36.8	11.5	29.0
S	-0.11	1.00	-0.12	-0.30	-0.07	-0.23	-0.05	-0.15	-0.13	-19.5	0.0	-37.2	-79.9	-26.3	-88.8	-17.8	-39.6	-37.6
D	0.09	-0.12	1.00	0.13	0.12	0.12	0.09	0.04	0.09	40.0	-38.4	0.0	34.0	44.0	36.7	27.1	10.3	21.4
$IdVol$	0.15	-0.25	0.14	1.00	0.15	0.15	0.12	0.11	0.13	46.8	-80.4	39.6	0.0	55.5	41.7	39.2	32.6	40.5
$ASSETG$	-0.03	-0.02	-0.04	-0.11	1.00	0.12	0.65	0.19	0.06	-6.5	-5.1	-13.7	-44.5	0.0	34.8	189.5	33.1	15.7
CI	-0.05	0.02	-0.02	-0.05	0.04	1.00	0.09	0.00	-0.01	-16.7	6.9	-6.7	-19.6	11.5	0.0	29.0	1.3	-2.8
$INV/ASSET$	-0.03	-0.01	-0.02	-0.09	0.48	0.16	1.00	0.13	0.03	-10.5	-3.6	-6.8	-38.8	83.9	60.3	0.0	23.0	6.6
NS	0.08	-0.07	0.02	0.06	0.34	0.10	0.27	1.00	0.18	19.8	-33.2	6.9	23.7	71.8	30.9	59.9	0.0	40.6
ι	0.17	-0.04	0.08	0.08	0.25	0.09	0.18	0.33	1.00	41.7	-10.5	21.5	29.0	51.7	17.1	48.7	59.1	0.0

Table 6
Partial Spearman Correlations Between Characteristics and Risk-Adjusted Returns

This table collects time series averages of partial Spearman correlations between nine characteristics (horizontal dimension) and risk-adjusted returns controlling for one of other characteristics (vertical dimension) as well as their t-statistics. B/M is book-to-market, S is size, D is analysts' forecasts dispersion, $IdVol$ is idiosyncratic volatility, $ASSETG$ is total asset growth, CI is abnormal capital investments, $INV/ASSET$ is investments-to-assets ratio, NS is net stock issues, ι is composite stock issuance. Panel A shows correlations computed using all available stocks for each pair of characteristics. Each column of Panels B, C, and D reports partial Spearman correlations between the chosen characteristic and risk-adjusted returns computed for stocks with the chosen characteristic in the bottom quintile, middle quintile, and top quintile, respectively. The sample covers the period from January 1965 to December 2007 for all characteristics except the analysts' forecasts dispersion for which the sample is January 1983 - December 2007.

Panel A: All stocks

	Partial Spearman correlation									t-stats								
	B/M	S	D	$IdVol$	$ASSETG$	CI	$INV/ASSET$	NS	ι	B/M	S	D	$IdVol$	$ASSETG$	CI	$INV/ASSET$	NS	ι
B/M	1.0000	0.0473	-0.0433	-0.0630	0.0007	0.0032	-0.0013	-0.0228	-0.0355	1.00	10.93	-9.44	-16.80	0.32	2.07	-0.74	-11.98	-12.38
S	-0.0269	1.0000	-0.0394	-0.0550	-0.0065	-0.0005	-0.0069	-0.0207	-0.0309	-9.62	1.00	-9.26	-18.40	-3.21	-0.35	-3.77	-10.61	-12.18
D	-0.0108	0.0122	1.0000	-0.0291	-0.0079	-0.0019	-0.0069	-0.0118	-0.0174	-3.30	3.46	1.00	-8.33	-2.65	-0.83	-2.36	-4.39	-5.12
$IdVol$	-0.0050	0.0086	-0.0321	1.0000	-0.0021	-0.0010	-0.0038	-0.0154	-0.0180	-1.79	2.29	-8.40	1.00	-0.97	-0.67	-1.90	-8.68	-8.41
$ASSETG$	0.0038	0.0408	-0.0421	-0.0607	1.0000	0.0036	-0.0006	-0.0250	-0.0350	1.44	9.73	-9.07	-16.39	1.00	2.52	-0.41	-11.74	-11.79
CI	0.0065	0.0384	-0.0414	-0.0585	0.0024	1.0000	-0.0005	-0.0204	-0.0329	2.22	9.21	-8.73	-15.84	1.06	1.00	-0.25	-9.52	-11.15
$INV/ASSET$	0.0055	0.0411	-0.0418	-0.0614	0.0023	0.0040	1.0000	-0.0227	-0.0333	2.00	9.63	-8.71	-16.50	1.31	2.81	1.00	-10.91	-11.32
NS	0.0019	0.0419	-0.0436	-0.0625	0.0073	0.0018	0.0049	1.0000	-0.0299	0.70	9.92	-9.28	-15.97	3.06	0.96	2.33	1.00	-9.99
ι	0.0123	0.0368	-0.0365	-0.0524	0.0095	0.0015	0.0081	-0.0078	1.0000	4.34	8.98	-8.07	-15.53	3.92	0.93	3.73	-3.78	1.00

Panel B: Portfolio 1

	Partial Spearman correlation									t-stats								
	B/M	S	D	$IdVol$	$ASSETG$	CI	$INV/ASSET$	NS	ι	B/M	S	D	$IdVol$	$ASSETG$	CI	$INV/ASSET$	NS	ι
B/M	1.0000	0.0065	-0.0021	0.0073	0.0274	0.0218	0.0158	0.0018	0.0062	1.00	1.77	-0.63	2.63	8.96	7.20	6.06	0.71	2.45
S	-0.0032	1.0000	-0.0023	0.0047	0.0221	0.0157	0.0118	0.0025	0.0040	-1.23	1.00	-0.73	1.84	8.30	5.50	5.07	0.99	1.62
D	0.0017	0.0051	1.0000	0.0012	0.0222	0.0103	0.0045	0.0012	0.0036	0.47	1.18	1.00	0.26	5.82	2.82	1.34	0.35	0.92
$IdVol$	-0.0018	-0.0193	-0.0033	1.0000	0.0127	0.0116	0.0059	0.0020	0.0048	-0.74	-7.02	-1.09	1.00	4.66	4.08	2.45	0.80	1.91
$ASSETG$	0.0054	-0.0103	0.0015	0.0109	1.0000	0.0200	0.0060	0.0029	0.0061	2.06	-3.33	0.43	3.46	1.00	6.79	2.60	1.11	2.41
CI	0.0061	-0.0074	0.0006	0.0099	0.0256	1.0000	0.0139	0.0019	0.0064	1.95	-2.31	0.16	2.88	8.17	1.00	5.07	0.70	2.28
$INV/ASSET$	0.0056	-0.0101	-0.0014	0.0115	0.0257	0.0200	1.0000	0.0001	0.0071	2.02	-3.22	-0.36	3.41	8.15	6.56	1.00	0.05	2.65
NS	0.0068	-0.0032	0.0008	0.0063	0.0293	0.0224	0.0208	1.0000	0.0098	2.09	-0.96	0.22	1.80	8.06	6.14	6.40	1.00	3.10
ι	0.0062	-0.0047	-0.0015	0.0117	0.0206	0.0185	0.0108	0.0036	1.0000	2.10	-1.41	-0.41	3.51	6.77	6.02	4.06	1.17	1.00

Table 6 (continued)

Panel C: Portfolio 3

	Partial Spearman correlation									t-stats								
	B/M	S	D	$IdVol$	$ASSETG$	CI	$INV/ASSET$	NS	ι	B/M	S	D	$IdVol$	$ASSETG$	CI	$INV/ASSET$	NS	ι
B/M	1.0000	0.0128	-0.0009	-0.0102	-0.0019	0.0014	-0.0012	0.0014	-0.0088	1.00	5.07	-0.28	-5.25	-0.90	0.55	-0.51	0.58	-3.46
S	-0.0059	1.0000	-0.0011	-0.0100	-0.0023	-0.0007	-0.0013	0.0024	-0.0053	-2.81	1.00	-0.35	-5.31	-1.06	-0.28	-0.58	1.00	-2.24
D	-0.0014	0.0047	1.0000	0.0013	0.0028	-0.0001	-0.0026	0.0039	-0.0011	-0.49	1.59	1.00	0.41	0.85	-0.02	-0.80	1.12	-0.31
$IdVol$	-0.0020	-0.0001	0.0004	1.0000	-0.0012	-0.0005	-0.0015	0.0036	-0.0034	-0.95	-0.03	0.12	1.00	-0.57	-0.21	-0.68	1.45	-1.45
$ASSETG$	0.0030	0.0113	0.0015	-0.0119	1.0000	0.0007	-0.0012	0.0016	-0.0097	1.29	4.70	0.46	-5.54	1.00	0.30	-0.52	0.62	-3.53
CI	0.0032	0.0039	0.0027	-0.0122	0.0023	1.0000	0.0028	0.0024	-0.0059	1.23	1.40	0.79	-4.94	0.99	1.00	1.10	0.82	-2.03
$INV/ASSET$	-0.0001	0.0118	0.0032	-0.0091	-0.0006	0.0028	1.0000	-0.0003	-0.0061	-0.05	4.59	0.90	-3.93	-0.26	1.10	1.00	-0.11	-2.10
NS	0.0011	0.0106	0.0004	-0.0101	0.0062	-0.0027	-0.0001	1.0000	-0.0148	0.39	3.74	0.12	-3.85	2.11	-0.92	-0.02	1.00	-4.31
ι	0.0028	0.0089	0.0016	-0.0116	0.0026	-0.0008	0.0022	0.0018	1.0000	1.18	3.44	0.45	-5.24	1.09	-0.29	0.84	0.63	1.00

Panel D: Portfolio 5

	Partial Spearman correlation									t-stats								
	B/M	S	D	$IdVol$	$ASSETG$	CI	$INV/ASSET$	NS	ι	B/M	S	D	$IdVol$	$ASSETG$	CI	$INV/ASSET$	NS	ι
B/M	1.0000	0.0051	-0.0278	-0.0488	-0.0265	-0.0252	-0.0191	-0.0207	-0.0234	1.00	1.76	-7.19	-15.91	-10.90	-8.26	-7.91	-7.69	-9.09
S	-0.0144	1.0000	-0.0216	-0.0539	-0.0227	-0.0189	-0.0167	-0.0101	-0.0168	-5.16	1.00	-5.77	-21.19	-9.06	-6.65	-6.58	-3.93	-7.09
D	-0.0059	0.0052	1.0000	-0.0376	-0.0189	-0.0166	-0.0101	-0.0191	-0.0221	-1.94	1.21	1.00	-9.93	-5.01	-4.18	-2.60	-5.30	-5.90
$IdVol$	-0.0054	0.0050	-0.0167	1.0000	-0.0159	-0.0178	-0.0107	-0.0083	-0.0134	-2.01	1.84	-4.59	1.00	-6.68	-6.22	-4.37	-3.23	-5.67
$ASSETG$	-0.0135	0.0078	-0.0258	-0.0459	1.0000	-0.0247	-0.0024	-0.0184	-0.0240	-4.54	2.62	-6.28	-15.22	1.00	-8.08	-0.91	-6.45	-9.09
CI	-0.0120	0.0110	-0.0254	-0.0438	-0.0226	1.0000	-0.0166	-0.0193	-0.0231	-3.65	3.21	-6.10	-13.44	-7.78	1.00	-5.92	-6.44	-8.17
$INV/ASSET$	-0.0137	0.0082	-0.0273	-0.0445	-0.0229	-0.0244	1.0000	-0.0197	-0.0279	-4.50	2.58	-6.09	-13.99	-8.23	-7.94	1.00	-6.81	-9.74
NS	-0.0134	0.0058	-0.0260	-0.0498	-0.0209	-0.0281	-0.0112	1.0000	-0.0217	-3.90	1.69	-5.98	-14.60	-7.18	-8.04	-3.61	1.00	-6.84
ι	-0.0099	0.0090	-0.0240	-0.0439	-0.0172	-0.0224	-0.0109	-0.0047	1.0000	-3.08	2.80	-5.43	-13.58	-6.47	-7.43	-4.07	-1.54	1.00

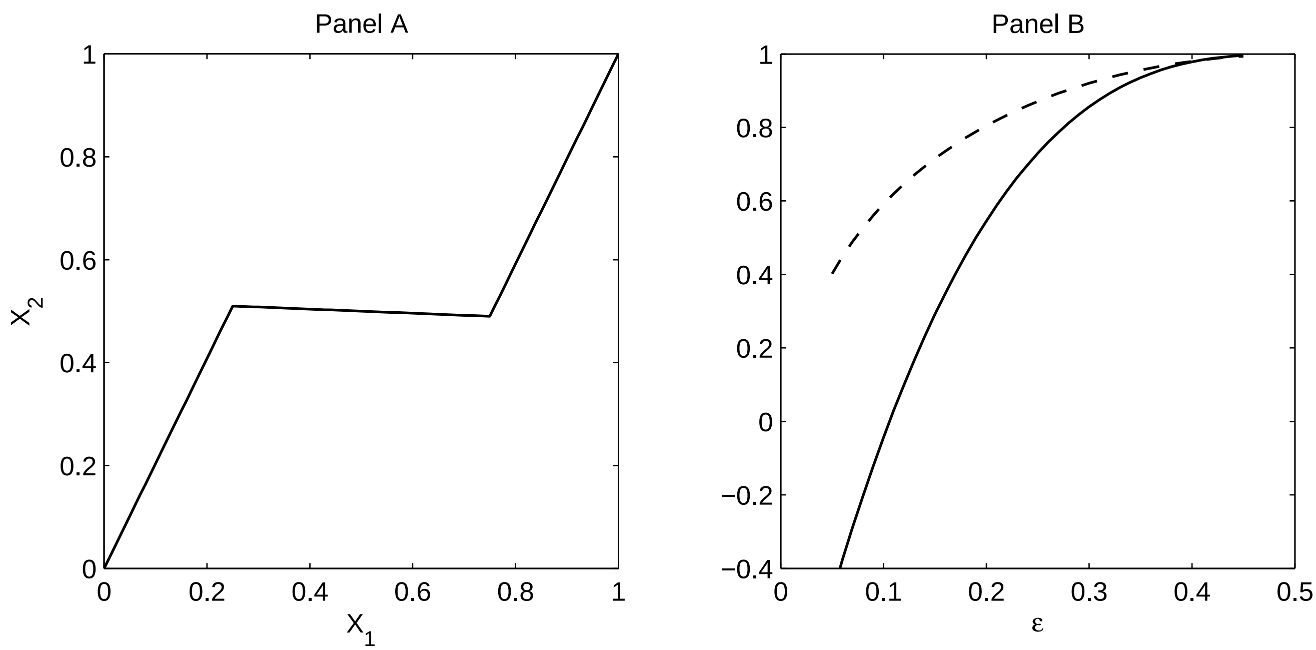

Figure 1. Illustration of Proposition 1. Panel A shows a typical graph of the function $X_2(X_1)$ defined by Eq. (4). The chosen parameters are $\varepsilon = 0.25$, $a = 0.01$, $b = 1$. Panel B plots the Pearson correlation between characteristics X_1 and X_2 (dashed line) as well as the Spearman rank correlation between stock rankings based on the same characteristics (solid line) as functions of ε.

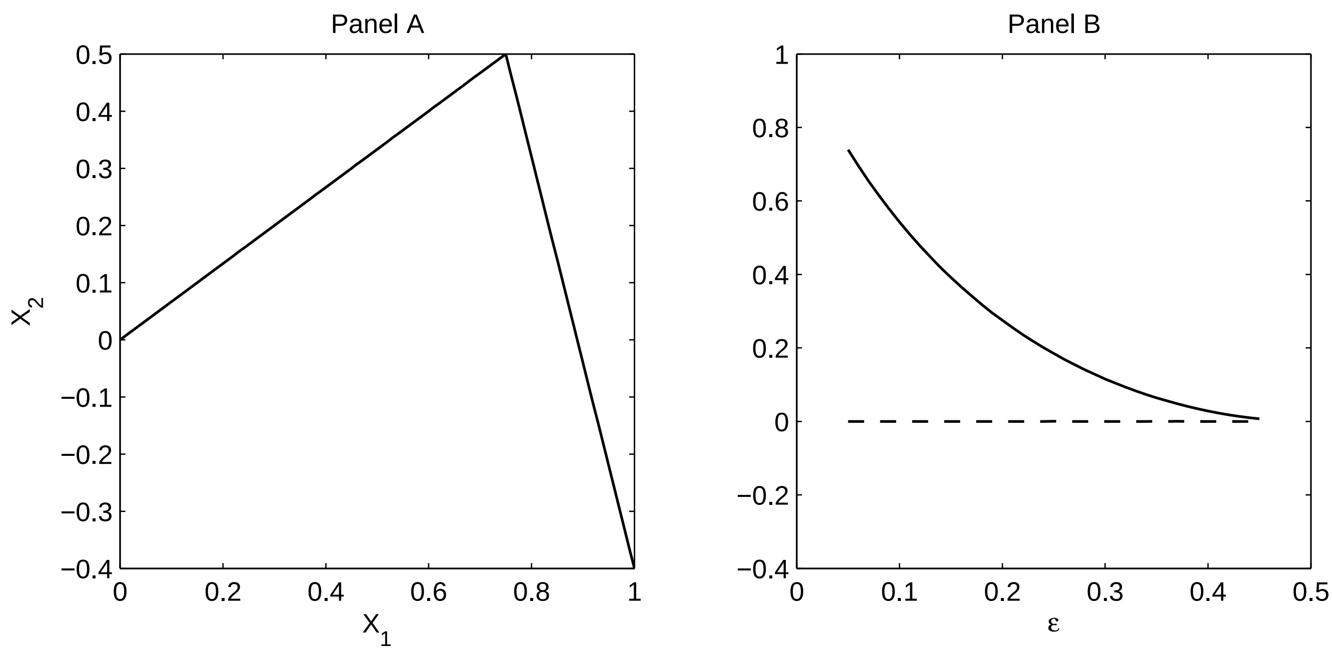

Figure 2. Illustration of Proposition 2. Panel A presents a typical graph of the function $X_2(X_1)$ defined by Eq. (7). The chosen parameters are $\varepsilon = 0.25$, $a = 0.5$, and the parameter b is set to -0.4 to get $cov(X_1, X_2) = 0$ (see Eq. (8)). Panel B plots the Pearson correlation between characteristics X_1 and X_2 (dashed line) as well as the Spearman rank correlation between stock rankings based on the same characteristics (solid line) as functions of ε.

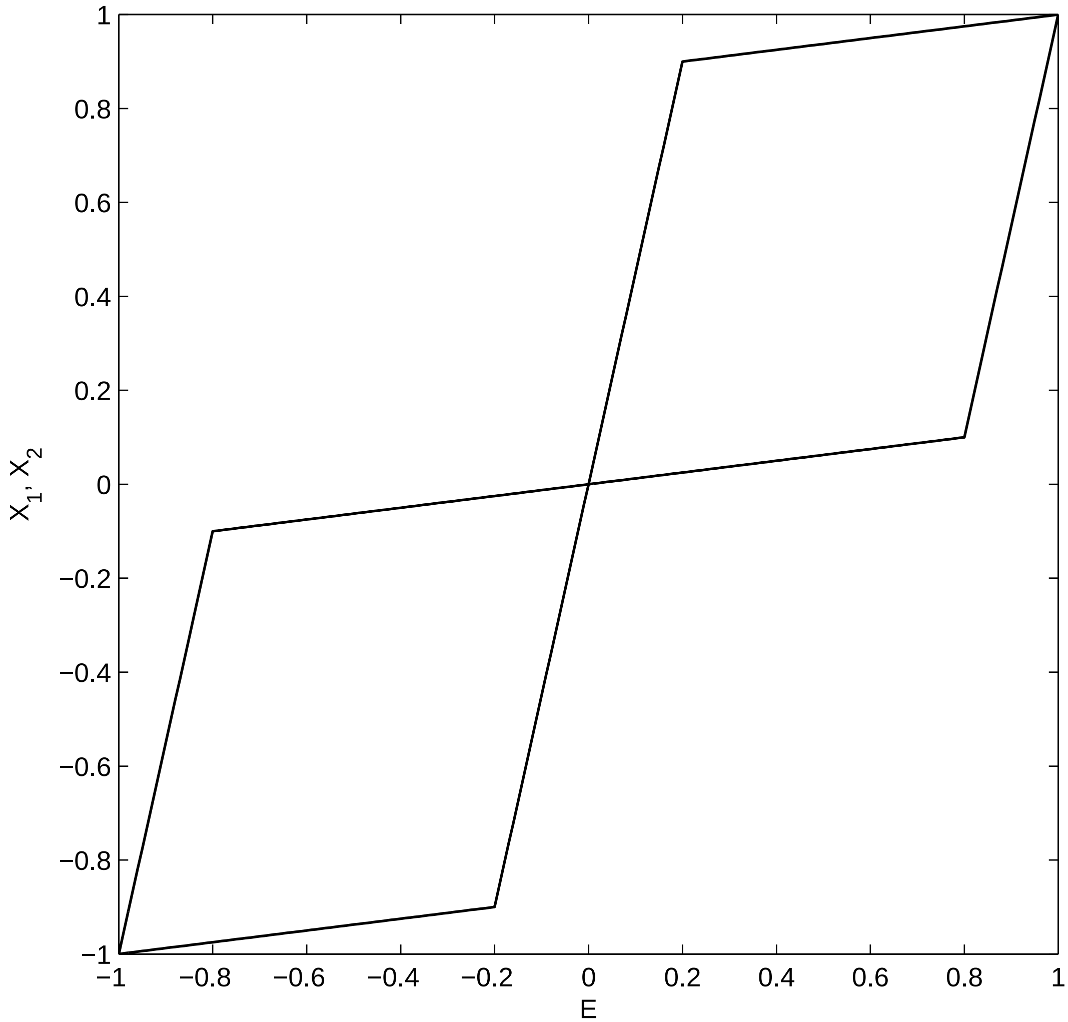

Figure 3. Illustration of Proposition 3. This Figure presents the graphs of the functions $X_1(E) = X(E; b_1, c_1)$, $X_2(E) = X(E; b_2, c_2)$, where $X(E; b, c)$ is defined by Eq. (9). The parameters are set as follows: $b_1 = 0.2$, $b_2 = 0.8$, $c_1 = 0.9$, $c_2 = 0.1$.

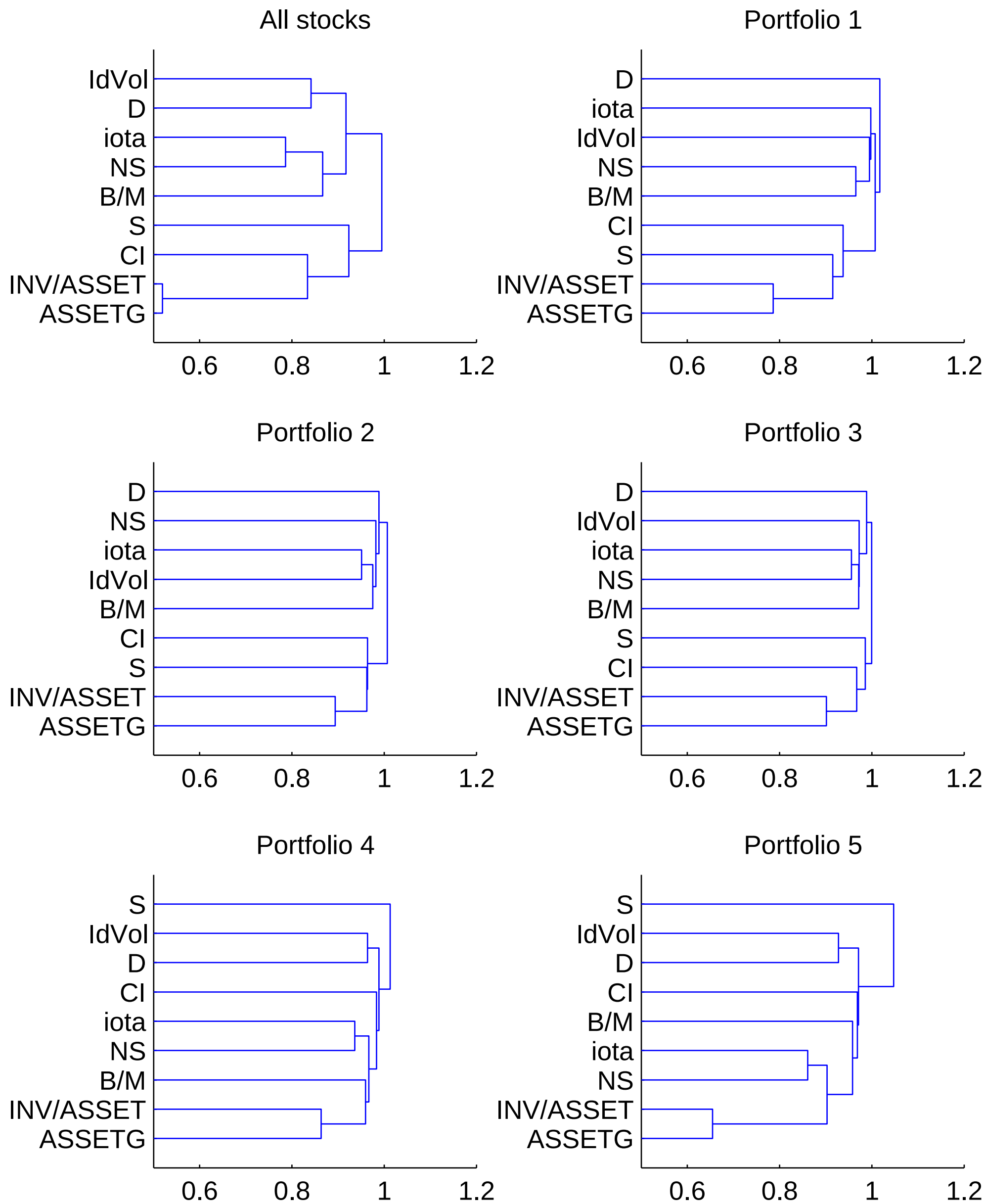

Figure 4. Cluster hierarchy of anomalies. This Figure plots the hierarchical tree of anomaly clusters in the metric space of anomalies. The distance between individual anomalies is given by Eq. (2) and the distance between clusters is defined as an average distance between all pairs of anomalies in any two clusters. The hierarchical trees are shown for all stocks as well as for quintile portfolios. B/M is book-to-market, S is size, D is analysts' forecasts dispersion, $IdVol$ is idiosyncratic volatility, $ASSETG$ is total asset growth, CI is abnormal capital investments, $INV/ASSET$ is investments-to-assets ratio, NS is net stock issues, ι is composite stock issuance.