Does the Source of Fundamental Data Matter?

Ondrej Tobek* & Martin Hronec†

July 2018

ABSTRACT

We study the role of the choice of a fundamental database on the portfolio returns of a set of 74 fundamental anomalies. We benchmark Compustat by comparing it to Datastream in the US and find systematic differences in the raw financial statements across the databases. These differences only have a small effect on the returns of anomalies when they are constructed on stock-months existing in both databases. Different stock coverage across the databases, however, leads to large statistically and economically significant disparities in the returns. Profitability anomalies yield negative returns on the Datastream universe.

JEL classification: G11, G12, and G15.

*University of Cambridge, Department of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, UK; Email: ot234@cam.ac.uk; Phone: +44 (0) 7476 680075

†Charles University, Institute of Economic Studies, Opletalova 26, 110 00, Prague 1; Email: martin.hronec@fsv.cuni.cz; Phone: +420 606 681 623

Electronic copy available at: https://ssrn.com/abstract=3150654

A police officer sees a drunken man intently searching the ground near a lamppost and asks him the goal of his quest. The inebriate replies that he is looking for his car keys, and the officer helps for a few minutes without success. Then, he asks whether the man is certain that he dropped the keys near the lamppost. "No," is the reply, "I lost the keys somewhere across the street." "Why look here?" asks the surprised and irritated officer. "The light is much better here," the intoxicated man responds with aplomb.[1]

Most of the research in accounting and finance relies only on two databases, the Center for Research in Security Prices (CRSP) and Compustat, since they are the most easily available to academics. However, these databases are not error-proof. Can these errors create significant biases across studies or are the errors idiosyncratic and no cause for worry? We test this question by looking at the performance of 74 fundamental anomalies published in finance and accounting journals when they are constructed in the Compustat universe or alternatively in the Reuters Datastream universe.[2] The studied anomalies are, for example, accruals of Sloan (1996), earnings over price of Basu (1977), composite equity issuance of Daniel and Titman (2006), and R&D over Market Equity of Chan, Lakonishok, and Sougiannis (2001). We also test the role of trade data by comparing portfolio returns on the anomalies constructed with individual stock returns from Datastream or CRSP and fundamental signals constructed in Compustat.

Another crucial aspect of the individual databases is the composition of the universe of stocks there. Academic studies mostly focus only on common stocks listed on countries' main exchanges, but this focus requires a classification by data vendors that is often wrong in earlier years. Some databases might also suffer from incomplete coverage for the stocks with low capitalization and the less frequently traded stocks. We study the implications of these differences among the databases for quantitative strategies. This study is also loosely connected to the recent literature on data mining in the finance

[1]It is impossible to find the original source of this allegory. See https://quoteinvestigator.com/2013/04/11/better-light/ for an attempt to find the source.

[2]We sometimes call the Compustat universe as CRSP and Compustat universe since Compustat does not include trade data whereas Datastream contains both market and fundamental data. The fundamental sub-database in Datastream is called Worldscope and we denote it interchangeably as Datastream throughout this text.

Electronic copy available at: https://ssrn.com/abstract=3150654

literature, as it is expected that extensive data mining in CRSP and Compustat would lead to lower relative returns in Datastream.[3] We find some support for this notion and note that some recently discovered anomalies cannot be replicated in Datastream.

We first study the fundamental anomalies on a sample of stocks in CRSP that can be matched to fundamental data in both Datastream and Compustat. We start by comparing the individual raw items on the financial statements that are required for constructing the anomalies.[4] We find that the items can substantially differ across the two databases. There are some apparent patterns in the differences. They tend to cluster in areas where the data vendors require specific methodologies to be applied. Some examples include the treatment of short-term versus long-term debt, long-term leases, or financing items on cash flow statements. These substantial differences in raw items, however, mostly do not translate to differences in the portfolio returns on fundamental anomalies in the matched Datastream and Compustat sample of firms. The average correlation in portfolio returns on anomalies created from the two databases is 95.9%, and there are no apparent economically significant differences there.

The discrepancies are, however, substantially larger once we move outside the matched sample and construct anomalies on the full samples of companies in each fundamental database. We partially explain this outcome by the lower coverage of stocks with lower capitalization in Datastream in the earlier period, but some economically and statistically significant differences nonetheless remain.[5] The discrepancies are huge when the individual quantitative strategies are considered.[6] 41 of the 74 anomalies are significant at the 5% level in CRSP plus Compustat and 39 in Datastream over the 1990 to 2016 period. There are, however, only 29 anomalies that are significant in both. Inference for individual strategies thus suffers from large biases. The discrepancies are, however, much smaller for grouped anomalies. The average return on all 74 fundamental anomalies is al-

[3] See, for example, Hou, Xue, and Zhang (2017) or Harvey, Liu, and Zhu (2016) for data mining issues.

[4] This comparison was similarly performed in Ulbricht and Weiner (2005), who studied sample differences in fundamental variables in Datastream and Compustat in the US.

[5] Datastream covers 87.5% of the overall capitalization of stocks in Compustat in 1990, but this coverage has increased to essentially 100% since 2005. The two databases, however, continue to cover different sets of stocks labeled as common equity. The differences in returns on anomalies therefore remain substantial even after 2005.

[6] We provide detailed results for each anomaly in the online appendix.

Electronic copy available at: https://ssrn.com/abstract=3150654

most identical among the two databases. Datastream and other alternative data sources are thus safe to use in the aggregate analysis of returns on anomalies, especially when micro-caps are excluded from the sample.

The fundamental coverage in Datastream significantly predicts expected returns on stocks in CRSP. Stocks without the fundamental coverage significantly underperform those with the coverage. The fundamental coverage effect on expected returns in closely related to the number of analysts covering effect in Elgers, Lo, and Pfeiffer Jr (2001). The underperformance of stocks without the fundamental coverage is especially channeled to stocks with small operating profitability. Operating profitability anomaly yields substantially lower returns in Datastream because the low profitability stocks are less likely to be covered there. A value-weighted strategy shorting stocks without the fundamental coverage in Datastream that are in the lowest profitability decile in Compustat yields 28% annually over the 2000 to 2016 period.

There are three main sources of the differences in the returns on the anomalies. Firstly, the imperfect coverage causes disparity in portfolio breakpoints across the databases. Using breakpoints from NYSE, or large cap universe of stocks with full coverage in each region, elevates this problem. Secondly, the coverage of stocks within the population quantiles may differ. Value-weighting limits this problem since it shifts the focus on stocks that tend to have better coverage in all databases. Lastly, the databases may have idiosyncratic differences due to errors and design choices. Examples include different categorization of the individual securities and companies. These database-specific issues are the hardest to minimize and require a tailored solution every time.

The large discrepancies in the returns before 2005 have implications for international studies. We show that the problems with coverage are also prevalent in Europe, Japan, and Asia Pacific before 2000. Examples of studies that rely on Datastream and are thus affected include Ang, Hodrick, Xing, and Zhang (2009), McLean, Pontiff, and Watanabe (2009), Hou, Karolyi, and Kho (2011), Lee (2011), Titman, Wei, and Xie (2013), Watanabe, Xu, Yao, and Yu (2013), and Jacobs (2016).[7] The performance of individual

[7] Fama and French (2012) and Fama and French (2017) use fundamental data from Datastream to fill in gaps from Bloomberg, but similar patterns in coverage are also expected there.

Electronic copy available at: https://ssrn.com/abstract=3150654

strategies without any filters on the universe of stocks is very likely connected to large biases there. These biases are especially important for anomalies that are stronger in micro-caps as the coverage is significantly worse there. We test two new ways to construct portfolios that should lower the discrepancy. Both of them shift the focus on large cap stocks where the bias is smaller. The first method discards all the stocks that have capitalization smaller than the bottom decile of the NYSE. The second uses the breakpoint from the 1000 largest stocks in the region to construct the portfolios. We then use value-weighted returns in both of them. The correlation of portfolios between the two databases increases, but substantial differences remain. We conclude that the choice of the fundamental database used can have a large impact on tests of individual quantitative strategies, and researchers should be aware of this impact.

We next study the implications of the fundamental database choice for a selection of independently significant signals. There is a large amount of recent literature that attempts to shrink the number of anomalies by finding those that are independently significant after controlling for all the others.[8] Here, we follow the methodology from Green et al. (2017) and use Fama and MacBeth (1973) regressions of individual stock returns on rescaled fundamental characteristics and control for the false discovery rate. The results are overwhelming in the US, as there is only one significant anomaly out of 8 in Compustat that is common between the two databases. Both databases thus lead to very different discoveries. The differences in the US should translate to differences among selected anomalies in different global regions. Jacobs and Müller (2017) indeed show that significant anomalies are very different across the global regions, and our analysis thus explains this striking inconsistency. Any study attempting to distil which anomalies are significant should thus be aware that any selection procedure is very unstable and is dependent on the imperfections of the underlying data.

The conclusions of our study are not unique to Datastream but apply to all sources of historical fundamental data for international equities, given that none of them offers

[8]See, for example, Lewellen et al. (2015), Green, Hand, and Zhang (2017), Feng, Giglio, and Xiu (2017), and Freyberger, Neuhierl, and Weber (2017) for evidence from the US and Jacobs and Müller (2017) for international evidence.

Electronic copy available at: https://ssrn.com/abstract=3150654

perfect coverage of all listed stocks. Dai (2012) documents the gaps in coverage in FactSet Fundamentals, Compustat Global, and Bureau Van Dijk's international databases. Fama and French (2012) note gaps in the Bloomberg database. We focus only on anomalies created with fundamental data, but our conclusions are valid for trade data as well. Stocks covered in Datastream in 1990 correspond to 91.5% of the overall capitalization of all the stocks in CRSP, which is better than for fundamental data but is nowhere near perfect.

Ince and Porter (2006) have shown that the Datastream returns data has limitations, and some adjustments need to be applied to limit its errors. We propose several new ways to further limit the errors. We show that there are only a few discrepancies in returns with respect to CRSP after 2000. We recommend that the returns before 1990 should be winsorized at the 0.1% percentile and returns from 1990 to 1999 at the 0.01% percentile. We also propose a new way to correct the returns when there are stale quotes at the time of stock splits and other corporate events. Not implementing them can lead to erroneous returns of several thousand percent. This study is the first that evaluates the impact of not including delisting returns in Datastream. Shumway (1997) showed that missing delisting returns can have a large impact on the returns on some anomalies, such as size, but we find no such bias. Specifically, we note that omitting all delisting returns in CRSP leads to economically similar returns on our set of fundamental anomalies.

Our study is the closest to Ulbricht and Weiner (2005), who compared Compustat and Datastream in the US. They focused mainly on summary statistics for individual items on financial statements, while our study focuses on impacts for a large number of fundamental strategies. The studies are thus similar only in the initial step. The imperfect coverage of micro-caps in the US was also previously documented in Ulbricht and Weiner (2005), but we extend this coverage to international evidence and provide a wide assessment of the impacts of this imperfection. Our study is also related to Ince and Porter (2006) in that we propose new quality screens to shrink errors in Datastream.

We contribute to the literature in three ways. First, we propose new adjustments for the data from Datastream that decrease the number of errors there. These can be applied

Electronic copy available at: https://ssrn.com/abstract=3150654

to similar databases facing the same problems. Next, we provide robust evidence that the choice of Compustat as the main database in most of the finance and accounting literature is not a source of serious concern due to possible idiosyncratic errors there. Finally, we document the importance of coverage of listed stocks in fundamental databases. The coverage is especially important in the international setting where there is no single database with full coverage that spans a long time period. The partial coverage leads to biased and inconsistent results. This outcome should serve as a caveat for international studies where the fundamental data is important.

I. Data and initial adjustments

One of our sources for data on US stocks is the Merged CRSP/Compustat database from Wharton Research Data Services. The sample spans the 1963 to 2016 period and contains all NYSE, Amex, and NASDAQ common stocks (CRSP share code 10 or 11). We adjust the returns for delisting following guidance in Shumway (1997) and Hou et al. (2017).[9]

Our second source of US data, and primary source for international data, is Reuters Datastream (Worldscope). The database manual from 2007 states that: "The total universe of companies contained on the database has grown from approximately 4,000 in 1987, to over 51,100 at March 2007. This includes 33,300 currently active companies in developed and emerging markets, representing approximately 95% of global market capitalization." It should thus provide a good comparison for CRSP/Compustat in the US given its wide coverage. We filter the data following Ince and Porter (2006), Lee (2011), and Griffin, Kelly, and Nardari (2010). The procedure includes manually checking the names of the shares in the database for over 100 expressions that describe their share class.

[9]Specifically, we use the return over the month if the delisting is on the last day of the month. The relevant delisting return is then added as a return over the next month. Then, we use the delisting return (DLRET) from the monthly file if it is not missing. If it is missing, then we use $(1 + ret_{cum}) * (1 + DLRET_d) - 1$, where ret_{cum} is the cumulative return in the month of delisting and $DLRET_d$ is the delisting return from the daily file. Finally, we fill the gaps with $(1 + ret_{cum}) * (1 + DLRET_{avg}) - 1$, where $DLRET_{avg}$ is the average delisting return for stocks with the same first digit of the delisting code (DLSTCD). Hou et al. (2017) applies the average over the past 5 years, but we found this method to be very noisy and a single large outlier had a huge impact on the average value.

Electronic copy available at: https://ssrn.com/abstract=3150654

We leave only the primary quotes of ordinary shares of companies with few exceptions where the fundamental data in Datastream is linked with other share classes.[10] We also exclude all Real Estate Investment Trusts (REIT) We require the return index (RI) to be larger than 0.001 on the first day of the month for higher precision. All the returns in this study are converted to US dollars. We set the RI to missing if the price on the first day of the month is larger than $1 million. We delete daily returns for days when the stock market was closed in a given country.

We use the classification of Fama and French (2017), sorting developed countries into 4 groups: (1) North America (United States and Canada); (2) Europe (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom); (3) Japan; and (4) Asia Pacific (Australia, New Zealand, Hong Kong, and Singapore). The Datastream sample starts in 1990, where there was large enough coverage for the USA, Europe, and Japan. The stocks in individual countries are from the largest exchange in the given country with the exception of the US (NYSE, NASDAQ, and Amex) and Japan (Tokyo and Osaka).

A. *Merging Datastream and Compustat in the US*

We need to create a merged database from Datastream and Compustat for further analysis. Accordingly, we merge Datastream and CRSP on their main security level identifiers: DSCD and PERMNO. We do this rather than directly merging Datastream fundamental (Worldscope) and Compustat because it leads to a larger number of successfully matched stocks in the two databases. This better match is due to the design of Datastream where static data (for example industry classification or tickers) are separated from time-series data (for example prices). Static data then includes only the latest available entries so that if there are any changes over time, these changes are not

[10] We closely follow the description in Griffin et al. (2010) regarding what shares are not common. We also partially rely on the correct classification of stocks in CRSP, as we keep any stock that can be matched to CRSP by CUSIP and filtered by relevant filters there. This selection procedure is not very important in the current work, as stocks with fundamental coverage in Datastream are not plagued by as many errors or missing categorization compared to those without.

Electronic copy available at: https://ssrn.com/abstract=3150654

recorded. The CRSP and Compustat matching table in WRDS reflects the full history of changes. The fundamental data is related to the company and not only to particular share issues so that changes in the currently most relevant traded share class would cause a problem. DSCD is then related to particular share issue and it is assigned when it enters the Reuters platform, as is PERMNO in CRSP. Merging on DSCD and PERMNO thus leads to more precise results. We then connect Datastream with Datastream fundamental (done automatically by Reuters when downloading the data) and CRSP with Compustat (we use the Merged CRSP/Compustat database from WRDS) in the second stage.

We first connect the databases by the 8 digit Committee on Uniform Security Identification Procedures ticker (CUSIP) and then check if it was successful by comparing the exchange tickers and names in the two databases. We discard a few cases where it is evident that the merge was not successful. We then merge on 6 digit CUSIP and again manually check for the success of the merger. In the end, we get 130,000 merged PERMNO-year observations out of approximately 250,000 in Compustat over the 1980 to 2016 period. See Figure 1 for the number of firms in Datastream fundamental and Compustat and their merge success rate over time. It is evident from the figure that less than half of all firms in Compustat were in the merged sample in 1980. This level increased to approximately 95% in 2015.

[Place Figure 1 about here.]

B. Adjustments of returns in Datastream

Ince and Porter (2006) provided the first systematic treatment of data quality in the Datastream database. They suggested several adjustments to shrink the size of errors in the database. These adjustments include discarding extreme returns that revert the next month. They also note that dropping stocks with a price lower than $1 decreases the errors, as the mistakes tend to cluster in stocks with a low price. We have at least one decade worth of new data, so we revisit these issues.

Electronic copy available at: https://ssrn.com/abstract=3150654

Datastream provides stale prices when there is no trade during the day or when the stock is no longer traded so that the price of the last trade is repeated until there is a new trade. We thus delete all observations with stale prices at the end of our sample. We implement a new way to fix returns and prices when there is an event that affects the number of shares outstanding (e.g., stock split), but there are stale quotes of prices at that time.[11] We characterize this event by a concurrent daily return larger than 15% (lower than -15%), an increase in the daily adjustment factor (Datastream variable AR) by 15% (decrease by 15%), and zero volume (if Datastream variable UVO is missing). We delete the latest observations of price with no trading and backfill the correct prices from their first new quote if it arrives in less than 30 days after the event.

Following Ince and Porter (2006), we set as missing those monthly returns over 300% that revert back over the next month. We only discard returns which we failed to correct in our previously described procedure.[12] This adjustment leads to closer returns with respect to CRSP, and we have not found any way to improve it. We also set the RI to missing if the daily return is larger than 500%. We set any monthly return larger than 2000% as missing. There is only one such case in CRSP, but there are many in DS for the US.

Table I presents correlations between monthly returns in Datastream and CRSP depending on the percent of observations winsorized and the minimum price of a stock at the end of the previous month. We focus on three periods: 1980 to 1989, 1990 to 1999, and 2000 to 2016. We expect that the quality of data will increase over time so that lower adjustment amounts are needed. It is indeed the case and the most recent period does not require any filters or adjustments with 99.6% correlation of the returns. The most successful adjustment in the earliest period is winsorizing the highest and lowest 0.1% of all returns, or approximately 40 stocks, in a given month. We adjust only 0.01%, or approximately 4 observations every month, in the 1990 to 1999 period. There is no

[11] A natural reaction of price to the 1 to 10 split would be its decrease to 10% of the original price, but if there has been no trade since the split, the old price is still displayed in Datastream. This outcome results in an incorrectly displayed return of 900%.

[12] Specifically, we set as missing returns for two consecutive months if the return in the first was larger than 300% and the overall return over the two months was lower than 50%.

Electronic copy available at: https://ssrn.com/abstract=3150654

need for price filters in the latest period but limiting extreme returns on the stocks with the lowest price helps in the earlier periods. To summarize, we start with adjustments for large daily and monthly returns that revert back by first trying to fix them and then discarding the rest. We then winsorize the resulting returns at different levels depending on the period. Winsorization of the returns does not have a significant impact on our findings but it helps to make the comparison across Datastream and Compustat more robust since the results will not be as easily driven by few outliers.

[Place Table I about here.]

C. Construction of anomalies and portfolios

To study the role of the source of the accounting information, we primarily focus on the performance of fundamental anomalies. The main reason for this is that it is easy to quantify their differences across databases and this is possible in a systematic way across a large set of published studies. It should also be of the first order importance to any quantitative investor. We have tried to study the largest set of published anomalies possible. We have included all fundamental anomalies that we have found in the literature and that could be implemented in both Compustat and Datastream.[13] Specifically, we have tried to implement all anomalies documented in Harvey et al. (2016), McLean and Pontiff (2016), and Hou et al. (2017). This together constitutes 74 anomalies. We considered 93 anomalies initially, but excluded 19 that we failed to replicate within the original sample of the studies. We list only the remaining 74 anomalies in our analysis. We have grouped the anomalies into 5 categories and our main analysis then focuses only on these categories. The detailed results are provided in the online appendix. The groups are: accruals, profitability, value, investment, and intangibles. A detailed list of anomalies is provided in Table XIII in the Appendix. A detailed description of how we construct the anomalies is also provided in the section A in the online appendix.

We follow the original papers' guidance on the sample construction of individual

[13]Some anomalies cannot be replicated with Datastream because it does not contain some needed items. Examples are anomalies based on advertising expense.

Electronic copy available at: https://ssrn.com/abstract=3150654

anomalies. Most of the portfolios on the anomalies are equal-weighted except the cash-based operating profitability of Ball, Gerakos, Linnainmaa, and Nikolaev (2016), which is value-weighted. We construct returns on zero-cost portfolios as returns on stocks in the top quintile of each signal minus returns on the bottom quintile of each signal. We choose long legs as in the original studies.

Some anomalies require the classification of industries, such as Hou and Robinson (2006). The choice in the original papers is mostly with respect to Standard Industrial Classification (SIC) industry classification. We apply third level Datastream classification, which sorts industries into 19 groups instead for two main reasons. First, the coverage in Datastream is not the same as that in Compustat and this would create a huge difference for fundamental signals dependent on the industries if there are more than 100 industries. Second, the industry classification in Datastream is available only from the static file, which means that only the latest value is available. Variation over time for individual firms between closely related SIC codes would thus again cause problems. We provide the transition between SIC classification and Datastream classification in the online appendix.

D. Role of delisting returns

One shortfall of Datastream, and most of the other sources of returns for equities, is that it does not include the delisting return after the stock is removed from the exchange. Shumway (1997) showed that there could be a large bias in returns on portfolios constructed from CRSP data since it was missing many delisting returns from performance related delistings at the time of publication of his study. This has created an upward bias for returns on small cap stocks to the point that one half of size anomaly could be explained by it. The quality of CRSP has increased since his study so that most of the delisting returns are no longer missing. There are 20 680 delistings in CRSP, with just 2 742 of them missing as of 2017. We revisit this issue in this section by comparing the returns on portfolios sorted on our set of anomalies. We do not opt for the alternative data source on delisting, as Shumway did, but we will rather compare the returns on the

Electronic copy available at: https://ssrn.com/abstract=3150654

portfolios with all the adjustments in CRSP done correctly and with completely omitted delisting returns. The goal is to see if excluding them, as is tacitly done in Datastream, leads to systematic biases.

Table II provides the results for the 5 categories of fundamental anomalies.[14] It is apparent that there are some differences, but they are far smaller than Shumway (1997) suggested. They are not systematic in the sense that they would cluster in certain types of anomalies, with the exception of some profitability anomalies that tend to short stocks that go bankrupt with negative delisting returns. Omitting delisting returns then leads to approximately 10% lower estimated returns on them. The differences are small even for size and liquidity anomalies, where they are expected to be the largest. We can thus conclude that omitting delisting returns is not a cause for serious concern when using Datastream, and other factors play a far larger role. This is a different conclusion with respect to Shumway (1997), but it is hardly surprising. The average return over all delistings that were performance related is very close to zero in our sample, which is strikingly different from the -40% found in his study. His recommendation was to substitute the missing delisting returns for performance reasons by -30% return, which we do in our second comparison in the table.[15] The difference in returns is again tiny and the choice of how to adjust for delisting returns is thus not important.[16]

[Place Table II about here.]

II. Similarity of financial statements

We start our comparison of Compustat and Datastream by looking at raw financial statements. The corresponding items between fundamental databases should be very sim-

[14]The detailed results for each anomaly are provided in the online appendix.

[15]Delistings for a performance reason have the delisting codes: 500, 520, 551 to 574, 580, and 584 in CRSP.

[16]We have also tried several ways to interpolate the data on delistings from CRSP, but it did not lead to any meaningful improvements relative to omitting the delisting returns. It is possible to sort delistings in Datastream into several categories based on what is included in the names of the shares. Approximately half of all delisted stocks have some indication added to their name, such as 'DELIST' or 'MERGER'. Matching relevant firms in CRSP and computing the average delisting return for the categories, however, yields an average return that is close to zero.

Electronic copy available at: https://ssrn.com/abstract=3150654

ilar as most of the items can be obtained without any adjustment directly from statements provided by the companies in their regulatory filings. This, however, is not necessarily the case. We show that specific methodologies chosen by the data vendors can lead to large differences. We focus on reduced versions of the financial statements that include only items that were used in the construction of signals for fundamental anomalies in our reviewed literature. This is only a fraction of the variables, as there are 151 items in financial statements in Datastream with wide coverage from 1995 and over 200 items in Compustat. We focus only on the most important subset for the sake of brevity and because it is often difficult to find close matches for the other variables.

[Place Table III about here.]

Table III shows the time series averages of cross-sectional Pearson's and Spearman's correlations between items in the two databases. We also specify how we construct the corresponding items in Datastream in the last column. Some transitions can be done directly by simply matching items, but others have to be done by more complicated transformations. There are some visible patterns in the discrepancies between the databases. First, variables in the current working capital that are part of accruals tend to differ a great deal. Next, there are differences in the classification of leases in Property Plant and Equipment and the classification of long-term versus short-term debt. This is due to the different methodologies of data vendors and their interpretations of the raw statements provided by companies. Other notable differences are among the items in financing cash flows. This is again due to different methodologies by the vendors. To conclude, there are some notable differences across the databases that could create a systematic bias for the fundamental signals constructed from them.

III. Performance of anomalies in the same sample

The previous section has suggested some large differences in financial statements across the two databases. We will now investigate whether these differences translate into returns on the anomalies that are constructed from them. We start with a compari-

Electronic copy available at: https://ssrn.com/abstract=3150654

son within the sample of stocks that can be matched between the two databases in this section and follow with full samples in the individual databases in the next one.

We test the differences in two settings. First, we compare the similarities in the fundamental signals themselves, and then we turn to the returns on portfolios created based on them. Panel A of Table IV first looks at time series average of cross-sectional correlations between signals created from either Compustat or Datastream. Pearson's correlations can be very low for some signals, but the similarity in rankings based on the signals are much higher, with an average Spearman's correlation of 93.8%. This is mainly caused by outliers where few observations can completely dominate the correlations. The signals tend to have large tails and non-normal distribution so ranks are better at capturing the dependence structure.

[Place Table IV about here.]

We then turn to Fama-Mecbeth regressions to test the role of measurement errors in the fundamental data. The Fama-Mecbeth regressions provide an additional robustness since they measure predictability of the stock returns by the fundamental signals for all the stocks and not only those in the extreme quantiles, as in the portfolios. We consider two settings. First, we estimate standard Fama-Mecbeth regressions to explain the returns by cross-sectional quantiles of the fundamental signals from Compustat. That is, we estimate cross-sectional regressions of individual stock returns on the quantiles of signals for each month. We run the regressions individually for each signal. We then compute the time series mean and the corresponding t-statistic adjusted for heteroskedasticity and autocorrelation up to 12 lags with Newey and West (1987). Alternatively, we estimate the cross-sectional regressions with the quantiles of fundamental signals from Compustat instrumented by the quantiles of fundamental signals from Datastream.[17] This should allow us to see whether possible errors in Compustat have any systematic impact. The underlying assumption is that the errors in Datastream are independent of errors in Compustat.[18]

[17]We transform the fundamental signals to standardize them so that the presented results are more easily interpreted and their outliers play a smaller role.

[18]This assumption could be violated if, for example, the two data vendors copy the same mistakes

Electronic copy available at: https://ssrn.com/abstract=3150654

We first report the average coefficients from the cross-sectional regressions and then their corresponding t-statistics. There are no visible differences in either the returns or t-statistics before and after controlling for the measurement errors. The coefficients, and therefore also the returns on the anomalies, have increased by a slightly, but this is not economically meaningful.[19] Ten of the signals have differences in portfolio returns significantly different from zero. The differences are again not meaningful in economic terms. There are 41 significant anomalies with Compustat and CRSP, but there are two fewer with Datastream, out of which 38 are significant for both databases.

The rest of Panel A presents the discrepancies in the returns of the portfolios created either with CRSP and Compustat or with Datastream only. Panel B then decomposes the differences in the returns of the portfolios into two components created either by differences in returns (Compustat signals) or differences in signals (CRSP returns) across the two data sources. We do this by matching the fundamental signals from both Compustat and Datastream with the returns from CRSP. Alternatively, we take the fundamental signals from Compustat and merge them with the returns from either CRSP or Datastream. We then create portfolios and compare their returns. The table shows that there are some discrepancies for some signals, but they do not lead to any systematic biases. The lowest differences are in the value category, with a 99.4% average correlation occurring between the portfolios in this category. The largest differences are for intangibles. It is evident that the returns from both CRSP and Datastream provide almost identical portfolios for the same fundamental signals. This is documented by their average correlation of 99.4%. There are no strong systematic differences across the anomalies. The differences in quantitative portfolios between the two databases are mainly due to distinct fundamental signals in each of them.

One thing to notice is that the average return on profitability anomalies is zero for the joint sample of stocks from Datastream and Compustat. We will cover this discrepancy in greater detail in a Section VI.A later.

provided by the companies.

[19]The online appendix again provides the details for individual signals.

Electronic copy available at: https://ssrn.com/abstract=3150654

IV. Performance of anomalies in separate samples

We now turn to problems with distinct samples that emerge when Datastream and Compustat are not matched. That is, we look at differences across the two databases if portfolios are created solely from the data in each of them. We first start with the US and then widen the scope to international markets in the next section.

Table V compares the performance of the fundamental anomalies in the two databases without restriction on their joint coverage. We first focus on the case when there are no further filters on the universe of stocks and then try to test if the differences are smaller with some filters. The average return for all the anomalies is practically the same in Datastream and Compustat.[20] The average return is, however, not similar across all the categories. The returns on profitability anomalies drop the most with their average return going from 0.35% to -0.05% monthly and the average t-statistic going from 1.35 to 0.13. This drop is significant at the 5% level for both returns and t-statistics.[21] The changes in other categories are not statistically significant, but changes in individual anomalies can be substantial. A large difference is, for example, in operating profitability over assets, which would yield 0.93% monthly according to Compustat but only -0.11% monthly according to Datastream. This difference is significant at the 0.1% level.

43 of the anomalies have a difference in the return that is significant at the 5% level. There are 41 significant anomalies with Compustat and CRSP and 39 with Datastream. This is the same as in the common sample, but there are only 29 anomalies that are significant in both databases. Thus, one-quarter of all the anomalies cannot be consistently replicated across the two databases. This leads us to conclude that both databases can convey substantially different results due to their different coverage and classification of stocks when one considers individual anomalies. The differences are, however, much smaller if one focuses on groups of anomalies.

[20] We report standard errors that are computed from the variation of mean returns over all the anomalies in the given category. All the uncertainty is thus coming from variation across the anomalies, which we think is the more fundamental source of uncertainty. This is in line with the focus of our study, i.e., whether the choice of the fundamental database can have an impact on research conclusions. We have also estimated standard errors that are adjusted by clustering the on time and anomaly fixed effects in the panel setting, and they are always smaller. Our standard errors are thus more conservative.

[21] We test the significance of differences by t-test.

Electronic copy available at: https://ssrn.com/abstract=3150654

We next try to look at a reduced set of stocks that would suffer from smaller disparities. Figure 1 has documented that the coverage on Datastream was not ideal in earlier periods, especially for small stocks. Reuters provides different depths of fundamental coverage for companies in Datastream. Smaller companies that do not meet certain criteria are available only with a reduced set of items on their financial statements and all anomalies thus cannot be constructed for them. The Worldscope manual reports that $100 million market capitalization is the required threshold for the full coverage in some regions. This could be binding, especially historically. There are also differences in the way that Datastream and Compustat treat financial firms. The financial firms in Datastream have a special template for their financial statements, which is comprised of items that are different relative to industrial firms. This could lead to problems, as some signals cannot be constructed for them. Another important factor, which we consider, is time, as the coverage in Datastream has improved steadily.

We thus provide results for a restricted sample that contains only non-financial stocks with capitalization over $100 million and that spans the 2000 to 2016 period. The $100 million capitalization requirement is very similar to discarding the stocks with a size lower than the bottom decile in the NYSE, which has been widely used throughout the literature.[22] We then construct the fundamental signals on this reduced sample but create portfolios only from July 2010. Specifically, we censor all fundamental information from the time when the capitalization was lower than $100 million and before 2000 so that the signals are constructed only using a similar information set. This leads to samples in Compustat and Datastream that are very similar in size, and there are no obvious biases across capitalization quintiles in Datastream.

It is evident that the similarity of portfolios has increased, with the average correlations between returns increasing from 78.1% to 90.4%, but the differences remain substantial for some anomalies. 90.4% is still substantially smaller than 95.9% for stocks matched in the common sample, which implies that the classification of stocks in individual databases can have a substantial impact. The large difference in operating

[22] See, for example, Hou et al. (2017) and Green et al. (2017).

Electronic copy available at: https://ssrn.com/abstract=3150654

profitability over assets has virtually disappeared and would yield a 0.51% monthly average return according to Compustat and 0.44% according to Datastream. There are still 14 anomalies with differences in returns across the two databases that are significant at the 5% level. Significant anomalies again differ across the two databases. There are 6 significant anomalies with Compustat and 8 with Datastream, but only 4 of those are common across the two databases.

[Place Table V about here.]

A. *What drives the differences?*

We now study in more detail whether the missing fundamental coverage for stocks with smaller market capitalization can explain the discrepancy in the profitability of anomalies across the two databases. Figure 2 maps the proportion of stocks within a given size quintile in CRSP that has fundamental coverage in Datastream. We also include the lowest size quintile in Compustat for comparison. It is evident that the coverage has been very uneven over time and for different size quintiles. The smallest half of stocks suffered from insufficient fundamental coverage until 2000, and the full coverage only occurred around 2010.

Figure 4 further maps a smoothed histogram of the market cap of stocks with fundamental coverage in Compustat and Datastream in 1990 and 2015. It is apparent that the insufficient coverage in Datastream was throughout the whole distribution in 1990 but has virtually disappeared by 2015. There is thus no simple rule regarding how to discriminate based on size to eliminate all the differences in returns on the anomalies.

[Place Figure 2 and Figure 4 about here.]

Table VI tries to explain the differences in returns on anomalies across the two databases. We focus on the full samples without restrictions. We regress the difference in returns on the average cross-sectional quantile of the size of stocks in the respective portfolios. The quantiles are taken with respect to all the stocks in CRSP or Datastream.

Electronic copy available at: https://ssrn.com/abstract=3150654

We also regress the differences in returns on differences in average size. The regression is a simple pooled OLS with standard errors clustered on time periods and anomalies. Both size and difference in size are significant at the 5% level, both individually and jointly. The table thus documents that size is indeed important in explaining the differences and returns on anomalies that are more prevalent in larger stocks, which tend to differ less across the two databases.

[Place Table VI about here.]

B. *Sources of bias on the portfolio returns*

There would be no problems with the imperfect fundamental coverage if the stocks would be omitted randomly. The problem is that the coverage is not random, as documented earlier. There are three main sources of the biased returns on portfolios, and we will now cover them in detail.

Firstly, breakpoints on the portfolios are biased since the covered sample of stocks is not randomly sampled from the full population of stocks. The breakpoints are therefore valid only for a given database and not for the full population of the stocks or for other databases. The weighted average of stock returns for a subpopulation bounded by incorrectly specified breakpoints is biased if the bias in breakpoints is related to stock returns. That is, if the biased breakpoints cause omission or addition of stocks with different average return with respect to what subpopulation average for the given portfolio is. We will show that the likelihood of the fundamental coverage in Datastream depends on company size and number of analysts following among other factors. Both size and a number of analysts following has been linked in the literature to stock returns, see Banz (1981) and Elgers et al. (2001). Interactions between the anomalies and the variables driving the coverage is a source of another bias. Fama and French (1992) and Fama and French (2015), for example, document interactions of size and book to value, investments, and profitability. Bias coming from inappropriate breakpoints can be minimized by using breakpoint from large cap subpopulation of the stocks where there are only mild coverage

Electronic copy available at: https://ssrn.com/abstract=3150654

issues.

Secondly, imperfect coverage for stocks within a given subpopulation bounded by correct breakpoints can be a source of more bias. Suppose that it is possible to precisely specify population breakpoints and the bias discussed in the previous paragraph is completely dissolved. Non-random sampling could still cause problems if the likelihood of stocks omission is related to their expected returns. The argument for the bias is especially strong for interaction effects with size. Smaller stocks tend to be more illiquid and harder to trade in a significant quantity which limits the arbitrage opportunity. Any anomalies due to market frictions should therefore be stronger for the small cap stocks which creates interaction effects with size and problems with the non-random sampling.

Lastly, idiosyncratic differences across the databases can be a source of some bias. Classification of industries and treatment of static and time-series information are good examples. This aspect of the bias can be minimized only through specific treatment in the individual cases.

C. Portfolio constructions limiting the discrepancies

Is there any way to decrease the differences by choosing an appropriate methodology? This is not very important in the US, but it is of first order importance for international studies since Datastream is the most widely used database there. Figure 4 showed that there is a lower discrepancy in coverage for larger stocks. Specifically, the coverage for the 1000 largest stocks is very similar across the databases. We will now look at procedures that filter the universe of stocks based on their size to lower the bias.

Table VII presents the returns and t-statistics for value-weighted portfolios constructed on a large cap universe or with portfolio breakpoints from the largest 1000 stocks. The large cap universe is defined by stocks with a capitalization larger than that of the smallest decile at the NYSE. The logic behind the first adjustment is to truncate the whole distribution of stocks and discard the part where the difference is the largest. This should not cause any serious problems for measurement of profitability for implementable and scalable strategies as the small stocks constitute only a very small

Electronic copy available at: https://ssrn.com/abstract=3150654

proportion of the overall capitalization of the whole market and it is advocated, for example, in Hou et al. (2017). The second adjustment then again shifts the focus to large caps but does not discard the other stocks. The breakpoints based on the largest 1000 stocks and value-weighting guarantees that the largest stocks will dominate the returns of the portfolios. The use of breakpoints on large caps is very similar to the use of NYSE breakpoints, which has been applied in many studies and is advocated, for example, in Fama and French (2017).

Both methods lead to significant improvement in the correlation of portfolios across the two databases and provide very similar results. The average correlation has increased from 80.2% to approximately 86%. The discrepancy for the returns on profitability anomalies is now much lower as well, and the average absolute difference in the t-statistics decreased to almost one third. The difference in the inference on significance remains substantial nevertheless. There are 11 significant signals in Compustat and 12 in Datastream, but only 6 of those are common across the two databases for the large cap universe of stocks. There are 9 significant signals in Compustat and 9 in Datastream, but only 6 of those are common across the two databases for breakpoints based on the 1000 largest stocks. This is an even larger difference in relative terms with respect to considering all the stocks.

[Place Table VII and Figure 3 about here.]

V. Implications for studies of international markets

We have shown that fundamental coverage in Datastream in the US is not complete and this can have a large consequence on the measurement of performance of the anomalies. We will now focus on its coverage in different countries, as it is often the first database that researchers go to for international data. The US evidence serves as a great testing ground because it includes a large number of stocks, and its implications should be valid elsewhere as well. It is thus important to study imperfections in the coverage, as they could lead to biased estimates in these studies.

Electronic copy available at: https://ssrn.com/abstract=3150654

A. *Fundamental coverage around the globe*

Figure 4 presents a fraction of stocks with fundamental coverage depending on the size quintile in Japan, Europe, and Asia Pacific. It is evident that the imperfect coverage is as much present internationally as it is in the US. We next look for support of this imperfect fundamental coverage in Datastream and guidance regarding what patterns to expect from its manual. The Worldscope's manual states that: "In 1987, Worldscope established a second research center in Shannon, Ireland, to maintain and develop the database. In 1995, Worldscope established a third major research and data collection center in Bangalore, India. A fourth major research and data collection center in Manila, Philippines was added with Primark's 1999 acquisition of the Extel company database.... Today, the database operations group, which supports the Worldscope database, employs over 500 people mainly located in 3 collection centers located in Bangalore (India), Shannon (Ireland), and Manila (The Philippines)." It is thus very likely that the quality of data has been changing over time as new research centers have been established. We show precisely this in Figure 4. The coverage in Australia, New Zealand, Hong Kong, and Singapore was very uneven until 2001 and is close to 100% after that. Similarly, in Japan, Datastream fully covered only companies with large capitalization until 1998. The coverage is not complete even currently in a few European countries, but companies outside the lowest size quintile are generally fully covered from 1997. This is partly due to the inclusion of stocks outside the primary trading venue in each stock exchange. These stocks tend to be very illiquid and have only tiny market capitalization. They are thus not a source of serious concern, as any quantitative investor would exclude them from their investment universe anyway.

[Place Figure 4 about here.]

B. *Determinants of the coverage*

We have shown that the dependence of fundamental coverage on the market cap of individual stocks can have an impact on the measurement of performance of individual

Electronic copy available at: https://ssrn.com/abstract=3150654

anomalies in the case of the US. Are there any other confounding variables that a researcher should be aware of? The Worldscope manual from 2007 describes its content coverage in the following way: "A fully detailed analysis is required for all companies within the following countries: the United Kingdom, and the U.S. For all other countries, fully detailed analysis is required if any of the following criteria is fulfilled:

- Company is a constituent of the, FTSE ALL World, Dow Jones Global, MSCI World, MSCI EMF, S&P Global, S&P/Citigroup or a selected local index.
- Company has 5 or more broker estimates.
- Company has a market capitalization of greater than 100 million dollars (exception Japan, China & Taiwan).
- Legacy companies from Extel database[23]."

This description suggests that the number of analysts following can have a role very similar to size if it is related to expected returns on individual stocks. Elgers et al. (2001) show that this is indeed the case. Constituency in the indexes is more difficult to measure, but it is usually closely connected to size, which will capture most of its effect.

Table VIII presents logit regressions predicting fundamental coverage with the size quantile and analyst followings in individual countries.

$$Fundamental\ Coverage_{it} = \beta_0 + \beta_1 \mathbb{1}\{Size_{it} > \$100M\} + \beta_2 \mathbb{1}\{Analysts\ Following_{it} \geq 5\} + \beta_3 Size\ Quantile_{it} + \beta_4 (Size\ Quantile_{it} - Size\ Quantile_t^{\$100M}) \mathbb{1}\{Size_{it} > \$100M\} + \epsilon_{it} \tag{1}$$

All the standard errors in the reported t-statistics are HAC robust. The regressions should be useful in accessing where to expect problems with biases due to the confounding variables.

β_0 is proportional to unconditional coverage. That is, the higher it is, the better the fundamental coverage for stocks of all sizes. It has increased from the 1990-2002 period to the 2003-2016 period for almost all the regions, and the increase has been substantial

[23]The Extel database was acquired by Worldscope in 1999 and covered stocks in Asia.

Electronic copy available at: https://ssrn.com/abstract=3150654

for the US and countries in the Asia Pacific, as would be expected from the previous graphs. β_1 then captures the discrete change in coverage at approximately $100 million. It is insignificant or close to zero almost everywhere, with the exception of the US in the earlier period. This documents that the coverage of the stocks has indeed been only selective and not full in the US. On the other hand, the quantiles of size and more than 4 analysts following are significant almost everywhere. This means that both of them can lead to spurious results if the effect under study is somehow related to them. The size quantile tends to have a lower effect after the $100 million threshold, as is evident from a mostly insignificant $\beta_3 + \beta_4$ measuring slope on on size for stocks with capitalization larger than $100 million.

[Place Table VIII about here.]

C. Impact on selection of individually significant signals

The imperfect historical coverage in international markets has implications for returns on portfolios there in the very same way as in the US as argued in Section IV.B. Our study is therefore overwhelmingly showing that there could be a huge bias when looking at the performance of individual quantitative strategies in international markets. The bias can completely distort the statistical inference and lead to findings of patterns that are only its artifacts. The simple remedies of focusing on a large cap universe of stocks proposed earlier can correct for a part of the bias, but they cannot control for all of it. The bias is also much less important when all the anomalies are grouped together and the focus is only on their joint profitability.

D. Impact on selection of independently significant signals

The analysis so far has focused on returns on portfolios. We will here show that the same caveats apply in other settings as well. We follow the methodology from Green et al. (2017) to identify independently significant signals. Table IX presents anomalies that are significant in Fama and MacBeth (1973) regressions of individual stock returns

Electronic copy available at: https://ssrn.com/abstract=3150654

on rescaled anomalies. We put quantiles of the anomalies within each month-region instead of unscaled signals to limit the effect of outliers. All the signals are pooled in the regressions, as follows:

$$r_{i,t} = \beta_0 + \sum_{j=1}^{M} \beta_j Q_{i,j,t} + \epsilon_{i,t}. \tag{2}$$

for a given month t and number of signals M. $Q_{i,j,t}$ is the quantile of the fundamental signal that was available before the start of month t. We also remove binary variables and signals where the variance inflation factor is higher than 7.[24] We consider simple OLS regressions (E) and the value-weighted WLS regression (V). All the standard errors are HAC adjusted, as in Newey and West (1987), with 12 lags. We present the results for all the available stocks (All) and the restricted large cap stocks with sizes larger than the bottom decile in the NYSE (Large). U stands for all signals found to be significant while **A** stands for those that remain significant after a correction for a false discovery rate (FDR) at 5%.

The FDR correction is very important since one would tend to find one significant signal in 20 individual tests even if all of them are insignificant in reality. The FDR adjustment follows Benjamini and Yekutieli (2001) and proceeds by first sorting p-values from the smallest to the largest so that $p_1 \leq p_2 \ldots \leq p_i \ldots \leq p_M$. FDR adjusted p-values are determined with backward induction where $p_M^{FDR} = p_M \sum_{1 \leq j \leq M} \frac{1}{j}$ and

$$p_i^{FDR} = \min \left\{ p_{i+1}^{FDR}, p_i \frac{M}{i} \sum_{1 \leq j \leq M} \frac{1}{j} \right\} \tag{3}$$

The adjusted p-values p_i^{FDR} are then significant with an FDR of 5% if they are smaller than 5%.

The results for the US look staggering. There is only one common signal out of the 8 that is significant with FDR adjustment for Compustat for the full universe of stocks and OLS regressions. This does not change for large cap stocks, with one in 5 signals being common. Value-weighting helps as it selects only one significant signal

[24]The exclusion of signals is done iteratively, and we primarily discard signals that would not be significant for any specification in the US.

Electronic copy available at: https://ssrn.com/abstract=3150654

that is common across all the specifications for both the databases. The one commonly significant anomaly is the earnings predictability of Francis, LaFond, Olsson, and Schipper (2004), which is surprisingly not related to any commonly used factor. Omitting FDR correction does not change the inference and there are still huge differences. This suggests that it is virtually impossible to select independently significant signals in the same country using different datasets.

This then translates to large discrepancies for the international sample. It is apparent that some of the signals are common for the regions here, but the variability is again great, as in the case of the US. Jacobs and Müller (2017) conducted a similar exercise in international markets and found only a few signals that would be significant across all the regions. Our analysis here suggests that this result is a consequence of the imperfect coverage of Datastream in the individual regions. It serves as an important caveat that the population of stocks in individual regions and its coverage by data vendors has a substantial impact on research findings and anyone working with international data should be aware of this.

[Place Table IX about here.]

VI. Fundamental coverage and expected returns

We have shown that fundamental coverage in Datastream is related to variables that are themselves related to expected returns on stocks. We will now study if the fundamental coverage is itself related to the expected returns. Negative relationship of the fundamental coverage to size of the stocks suggests that stocks with fundamental coverage have lower returns than those without (see Banz (1981)). The positive relationship to the number of analysts following, however, suggests that it is the other way around (see Elgers et al. (2001)). The predictive power of the fundamental coverage is therefore not immediately obvious.

Table X shows profitability of a strategy that buys stocks in CRSP that have the fundamental coverage in Datastream and shorts those that do not. We define stocks

Electronic copy available at: https://ssrn.com/abstract=3150654

without the fundamental coverage as those for which we cannot construct book-to-market ratio as defined in Fama and French (1992). The sample spans July 1990 to December 2016. The strategy yields significantly positive returns for both equal-weighted and value-weighted returns. The significance is even higher once the returns are adjusted for the five Fama-French factors (Fama and French (2015)). The increase in significance is due to SMB factor capturing the effect of size that goes in the opposite direction than the effect of the fundamental coverage. The significantly positive mean returns remain even for a large cap universe of stocks, defined as stocks with capitalization larger than that of bottom decile of the NYSE. The table also shows minimum and average number of stocks in CRSP without the fundamental coverage in each month. The average number of stocks is over 500 even for the large cap universe and the results are therefore based on a large sample of stocks.

The relative underperformance of stocks without fundamental coverage is in line with underperformance of stocks with small number of analysts following. The similarity is hardly surprising since we have previously shown that the number of analysts is one of the criterion for the decision whether to provide the fundamental coverage in Datastream. The similarity is also strengthened by the fact that Thompson Reuters owns both the database for analysts' forecasts (IBES) and the fundamental database (Worldscope in Datastream). The decision whether to cover a given firm is therefore interconnected in both databases. The theoretical reasoning for no coverage can be very similar across the databases as well. The analysts are less likely to cover stocks that are underperforming and have small growth potential since they have only limited resources at their disposal. They therefore try to channel these resources at firms that attract the most investor's attention. The coverage in Datastream is also likely to prioritize stocks with large investors attention to successfully compete with other data vendors.

[Place Table X about here.]

Electronic copy available at: https://ssrn.com/abstract=3150654

A. Low profitability firms without fundamental coverage in Datastream

We have previously documented large differences in returns on profitability anomalies in Compustat relative to Datastream. The discrepancy is mainly due to stocks in low profitability category and we study them in more detail here. The stocks that are among the least profitable in Compustat and have no fundamental coverage in Datastream have severely underperformed since 2000. This underperformance could be connected to the low interest of the investor since they were not worth following by one of the main data vendors. It could also be because they are difficult to short, which introduces limits to arbitrating and allows only a slow adjustment. We will now study the low profitability stocks without the fundamental coverage in Datastream in more detail.

Table XI presents the average monthly returns on a strategy that buys all stocks without fundamental coverage in Datastream that are in the bottom decile or quintile of operational profitability in Compustat. We measure profitability by operating profits to assets as in Ball et al. (2016). Our sample either includes all non-financial stocks or we further discard all stocks with sizes smaller than bottom decile on the NYSE in the previous June. The portfolios are either value-weighted (VW) or equal-weighted (EW). A value-weighted strategy in which shorts stocks without fundamental coverage in Datastream that are in the lowest profitability decile in Compustat yields 28% annually over the 2000 to 2016 period. This is also significant for equal-weighted returns. The returns remain significant on a large cap universe. Alphas with respect to the Fama-French five factor model are even more significant with t-statistics of approximately 6. There are, on average, 132 stocks in the portfolio for the full sample but fewer for the large cap sample. The evidence is thus based only on few data points. We have tried to look at individual instances of these stocks. The stocks are often facing bankruptcy and have management problems.

There are several possible explanations for this anomaly. First, it could be the case that the fundamental data have been backfilled in Compustat only after some time. The stocks have been in CRSP for 72 months on average, so the late addition of fundamental information on new issues cannot fully explain the difference. It is also possible that the

Electronic copy available at: https://ssrn.com/abstract=3150654

difference is due to the inattention of investors. We can proxy for the attention by the number of analysts following them. Elgers et al. (2001) show that the number of financial analysts covering the stocks can predict the future return. The stocks in the portfolios have, on average, 3.19 analysts covering them, which is lower than the 7.35 analysts for all the other stocks. This is in line with our previous analysis that the stocks would have fundamental coverage if they had more than 4 analysts coverings them.

[Place Table XI about here.]

VII. Robustness

Here, we provide robustness to our findings. Our previous analysis focused on quantile portfolios with return weighting following the original studies. We will now show that our conclusions remain unchanged for a different construction of the portfolios. Table XII presents the differences in the portfolios sorted on anomalies for different constructions of the portfolios. We extend our previous analysis to decile and tercile breakpoints in portfolio sorts and value-weighting. It is apparent that there is only a slight difference for the various breakpoints on equal-weighted portfolios. Value-weighted portfolios have lower average returns and t-statistics, but some differences among the databases still remain.

[Place Table XII about here.]

Panel D captures the number of significant signals with t-statistics larger than 2 for the various portfolio constructions. The number of significant anomalies is very similar across the two databases, but it is generally smaller in Datastream. The number of signals that are significant across both the databases is always lower than for Compustat alone by at least one fourth. The previous conclusions thus carry over to other portfolio constructions and are very robust.

Electronic copy available at: https://ssrn.com/abstract=3150654

VIII. Conclusion

We have compared fundamental data from two sources, and we have shown that measurement error in the fundamental data can be large. There are substantial differences in the raw financial statements caused by different methodologies for the construction of statements in the databases. These are less pronounced for portfolios created from sorts on fundamental signals. The findings on the significance of anomalies constructed with Compustat are thus robust to measurement error. We have documented several problems with Datastream. We have managed to correct some, but others have no clear solution. The strong message of this paper is that Datastream is a good source of data only after approximately 2000, and its use in an earlier period could be connected to significant bias. This is true for both the US and the International samples. We have also revisited the role of delisting returns and have not found any serious bias introduced by completely ignoring it, unlike in the previous studies.

Appendix A. List of fundamental anomalies

[Place Table XIII about here.]

References

Abarbanell, Jeffery S, and Brian J Bushee, 1998, Abnormal returns to a fundamental analysis strategy, *Accounting Review* 19–45.

Alwathainani, Abdulaziz M, 2009, Consistency of firms' past financial performance measures and future returns, *The British Accounting Review* 41, 184–196.

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2009, High idiosyncratic volatility and low returns: International and further us evidence, *Journal of Financial Economics* 91, 1–23.

Ball, Ray, Joseph Gerakos, Juhani T Linnainmaa, and Valeri Nikolaev, 2016, Accruals, cash

Electronic copy available at: https://ssrn.com/abstract=3150654

flows, and operating profitability in the cross section of stock returns, *Journal of Financial Economics* 121, 28–45.

Banz, Rolf W, 1981, The relationship between return and market value of common stocks, *Journal of financial economics* 9, 3–18.

Barbee Jr, William C, Sandip Mukherji, and Gary A Raines, 1996, Do sales–price and debt–equity explain stock returns better than book–market and firm size?, *Financial Analysts Journal* 52, 56–60.

Bartov, Eli, and Myungsun Kim, 2004, Risk, mispricing, and value investing, *Review of Quantitative Finance and Accounting* 23, 353–376.

Basu, Sanjoy, 1977, Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis, *The journal of Finance* 32, 663–682.

Belo, Frederico, and Xiaoji Lin, 2011, The inventory growth spread, *The Review of Financial Studies* 25, 278–313.

Belo, Frederico, Xiaoji Lin, and Santiago Bazdresch, 2014, Labor hiring, investment, and stock return predictability in the cross section, *Journal of Political Economy* 122, 129–177.

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

Bhandari, Laxmi Chand, 1988, Debt/equity ratio and expected common stock returns: Empirical evidence, *The journal of finance* 43, 507–528.

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

Bradshaw, Mark T, Scott A Richardson, and Richard G Sloan, 2006, The relation between corporate financing activities, analysts' forecasts and stock returns, *Journal of Accounting and Economics* 42, 53–85.

Chan, Louis KC, Josef Lakonishok, and Theodore Sougiannis, 2001, The stock market valuation of research and development expenditures, *The Journal of Finance* 56, 2431–2456.

Electronic copy available at: https://ssrn.com/abstract=3150654

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

Dai, Rui, 2012, International accounting databases on wrds: Comparative analysis .

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

Dechow, Patricia M, Richard G Sloan, and Mark T Soliman, 2004, Implied equity duration: A new measure of equity risk, *Review of Accounting Studies* 9, 197–228.

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

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

Eisfeldt, Andrea L, and Dimitris Papanikolaou, 2013, Organization capital and the cross-section of expected returns, *The Journal of Finance* 68, 1365–1406.

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

Fairfield, Patricia M, J Scott Whisenant, and Teri Lombardi Yohn, 2003, Accrued earnings and growth: Implications for future profitability and market mispricing, *The accounting review* 78, 353–371.

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

Fama, Eugene F, and Kenneth R French, 2012, Size, value, and momentum in international stock returns, *Journal of financial economics* 105, 457–472.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1 – 22.

Fama, Eugene F, and Kenneth R French, 2017, International tests of a five-factor asset pricing model, *Journal of Financial Economics* 123, 441–463.

Electronic copy available at: https://ssrn.com/abstract=3150654

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of political economy* 81, 607–636.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2017, Taming the factor zoo .

Francis, Jennifer, Ryan LaFond, Per M Olsson, and Katherine Schipper, 2004, Costs of equity and earnings attributes, *The accounting review* 79, 967–1010.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2017, Dissecting characteristics nonparametrically, Technical report, National Bureau of Economic Research.

Green, Jeremiah, John RM Hand, and X Frank Zhang, 2017, The characteristics that provide independent information about average us monthly stock returns, *The Review of Financial Studies* hhx019.

Griffin, John M, Patrick J Kelly, and Federico Nardari, 2010, Do market efficiency measures yield correct inferences? a comparison of developed and emerging markets, *The Review of Financial Studies* 23, 3225–3277.

Hafzalla, Nader, Russell Lundholm, and E Matthew Van Winkle, 2011, Percent accruals, *The Accounting Review* 86, 209–236.

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

Harvey, Campbell R, Yan Liu, and Heqing Zhu, 2016, ... and the cross-section of expected returns, *The Review of Financial Studies* 29, 5–68.

Haugen, Robert A, and Nardin L Baker, 1996, Commonality in the determinants of expected stock returns, *Journal of Financial Economics* 41, 401–439.

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

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

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

Electronic copy available at: https://ssrn.com/abstract=3150654

Hou, Kewei, Chen Xue, and Lu Zhang, 2017, Replicating anomalies, Technical report, National Bureau of Economic Research.

Ikenberry, David, Josef Lakonishok, and Theo Vermaelen, 1995, Market underreaction to open market share repurchases, *Journal of financial economics* 39, 181–208.

Ince, Ozgur S, and R Burt Porter, 2006, Individual equity return data from thomson datastream: Handle with care!, *Journal of Financial Research* 29, 463–479.

Jacobs, Heiko, 2016, Market maturity and mispricing, *Journal of Financial Economics* 122, 270–287.

Jacobs, Heiko, and Sebastian Müller, 2017, ... and nothing else matters? on the dimensionality and predictability of international stock returns .

Lakonishok, Josef, Andrei Shleifer, and Robert W Vishny, 1994, Contrarian investment, extrapolation, and risk, *The journal of finance* 49, 1541–1578.

Lee, Kuan-Hui, 2011, The world price of liquidity risk, *Journal of Financial Economics* 99, 136–161.

Lewellen, Jonathan, et al., 2015, The cross-section of expected stock returns, *Critical Finance Review* 4, 1–44.

Lockwood, Larry, and Wikrom Prombutr, 2010, Sustainable growth and stock returns, *Journal of Financial Research* 33, 519–538.

Loughran, Tim, and Jay W Wellman, 2011, New evidence on the relation between the enterprise multiple and average stock returns, *Journal of Financial and Quantitative Analysis* 46, 1629–1650.

Lyandres, Evgeny, Le Sun, and Lu Zhang, 2007, The new issues puzzle: Testing the investment-based explanation, *The Review of Financial Studies* 21, 2825–2855.

McLean, R David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *The Journal of Finance* 71, 5–32.

McLean, R David, Jeffrey Pontiff, and Akiko Watanabe, 2009, Share issuance and cross-sectional returns: International evidence, *Journal of Financial Economics* 94, 1–17.

Electronic copy available at: https://ssrn.com/abstract=3150654

Nagelkerke, Nico JD, et al., 1991, A note on a general definition of the coefficient of determination, *Biometrika* 78, 691–692.

Newey, Whitney K, and Kenneth D West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica: Journal of the Econometric Society* 703–708.

Novy-Marx, Robert, 2010, Operating leverage, *Review of Finance* 15, 103–134.

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

Ortiz-Molina, Hernán, and Gordon M Phillips, 2014, Real asset illiquidity and the cost of capital, *Journal of Financial and Quantitative Analysis* 49, 1–32.

Penman, Stephen H, Scott A Richardson, and Irem Tuna, 2007, The book-to-price effect in stock returns: accounting for leverage, *Journal of Accounting Research* 45, 427–467.

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

Richardson, Scott A, Richard G Sloan, Mark T Soliman, and Irem Tuna, 2006, The implications of accounting distortions and growth for accruals and profitability, *The Accounting Review* 81, 713–743.

Shumway, Tyler, 1997, The delisting bias in crsp data, *The Journal of Finance* 52, 327–340.

Sloan, Allan, 1996, Create an account or log in, *Accounting review* 71, 289–315.

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

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

Thomas, Jacob K, and Huai Zhang, 2002, Inventory changes and future returns, *Review of Accounting Studies* 7, 163–187.

Electronic copy available at: https://ssrn.com/abstract=3150654

Titman, Sheridan, KC John Wei, and Feixue Xie, 2004, Capital investments and stock returns, *Journal of financial and Quantitative Analysis* 39, 677–700.

Titman, Sheridan, KC John Wei, and Feixue Xie, 2013, Market development and the asset growth effect: International evidence, *Journal of Financial and Quantitative Analysis* 48, 1405–1432.

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

Ulbricht, Niels, and Christian Weiner, 2005, Worldscope meets compustat: A comparison of financial databases .

Watanabe, Akiko, Yan Xu, Tong Yao, and Tong Yu, 2013, The asset growth effect: Insights from international equity markets, *Journal of Financial Economics* 108, 529–563.

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

Electronic copy available at: https://ssrn.com/abstract=3150654

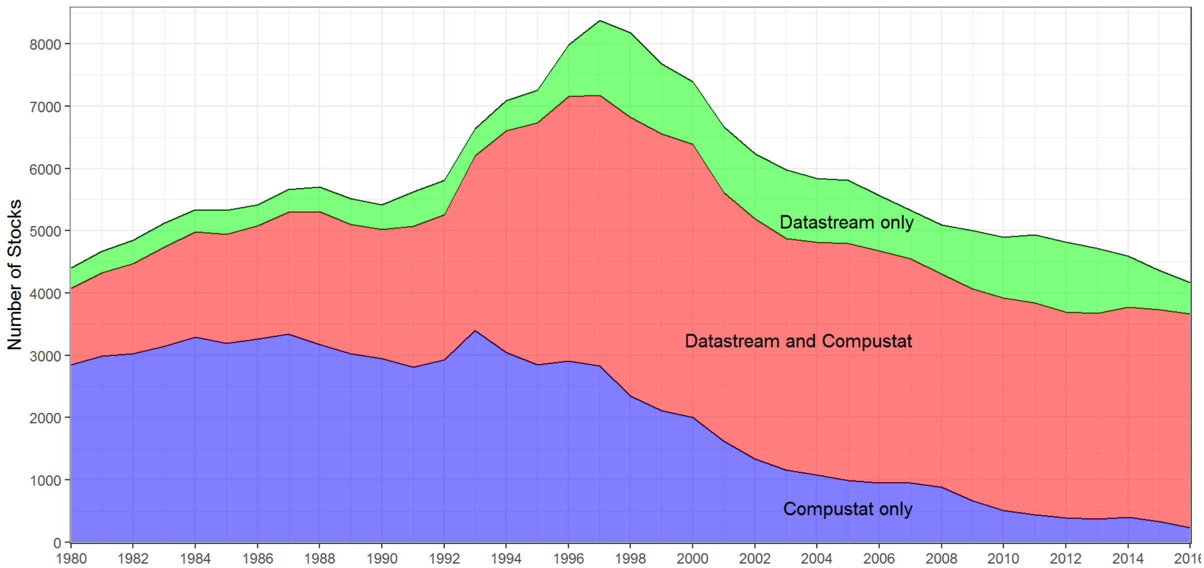


Figure 1. Number of stocks with fundamental coverage in Compustat and Datastream over time.

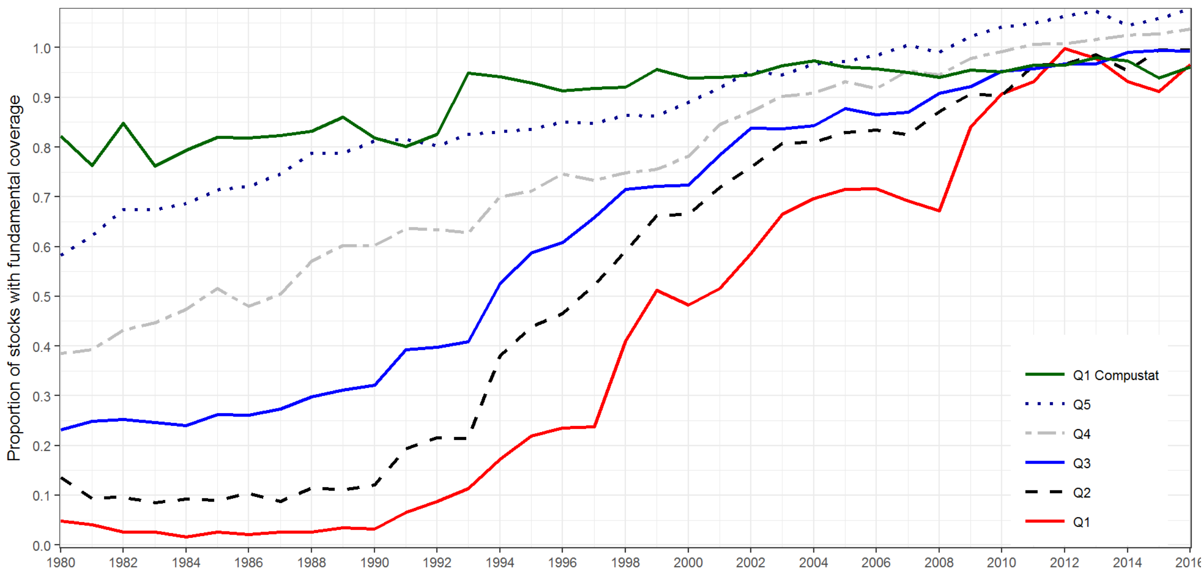


Figure 2. Fraction of stocks in CRSP with fundamental coverage in Compustat or Datastream in a given size quintile.

Electronic copy available at: https://ssrn.com/abstract=3150654

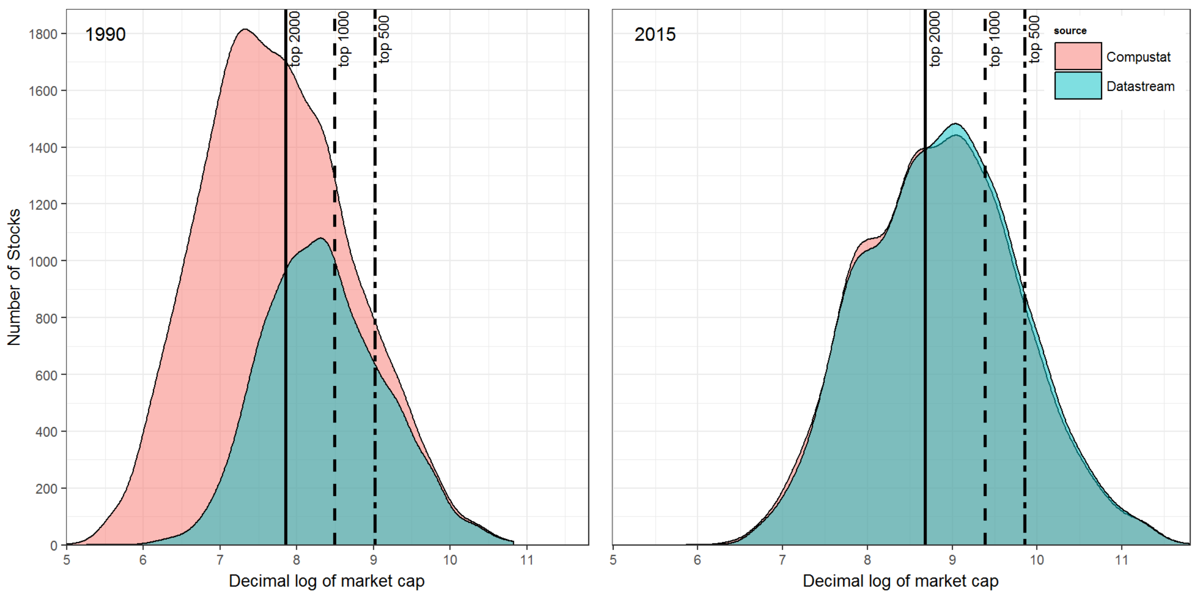
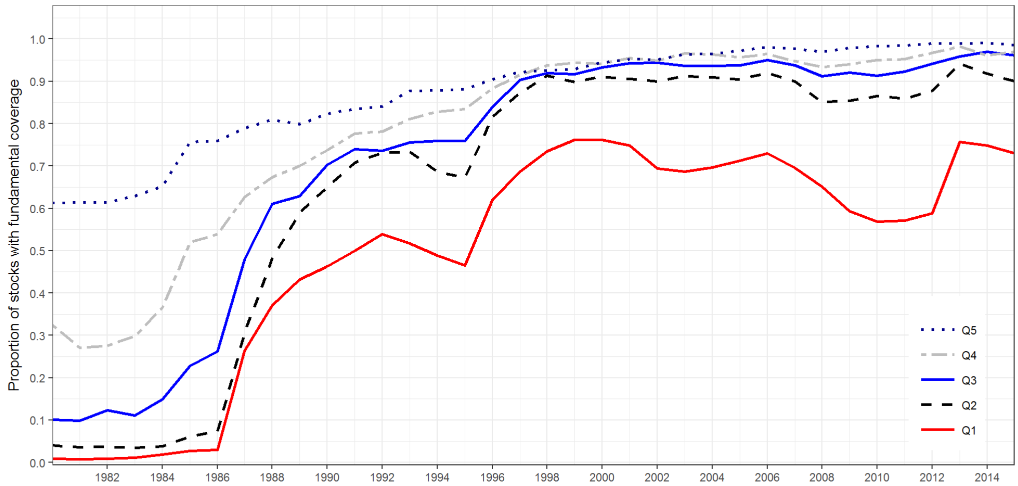


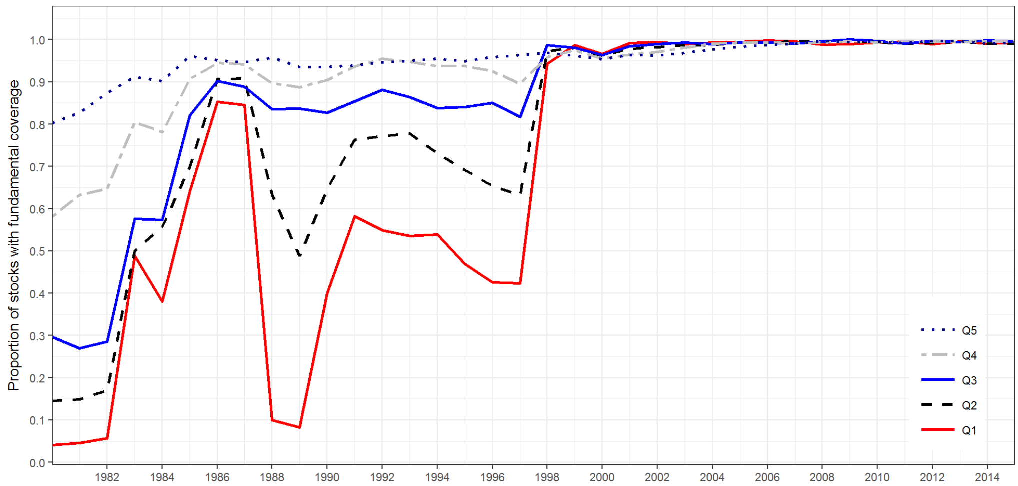
Figure 3. Histogram of market cap of stocks with fundamental coverage in Compustat and Datastream.

Electronic copy available at: https://ssrn.com/abstract=3150654

Panel A: Europe.



Panel B: Japan.



Panel C: Asia Pacific.

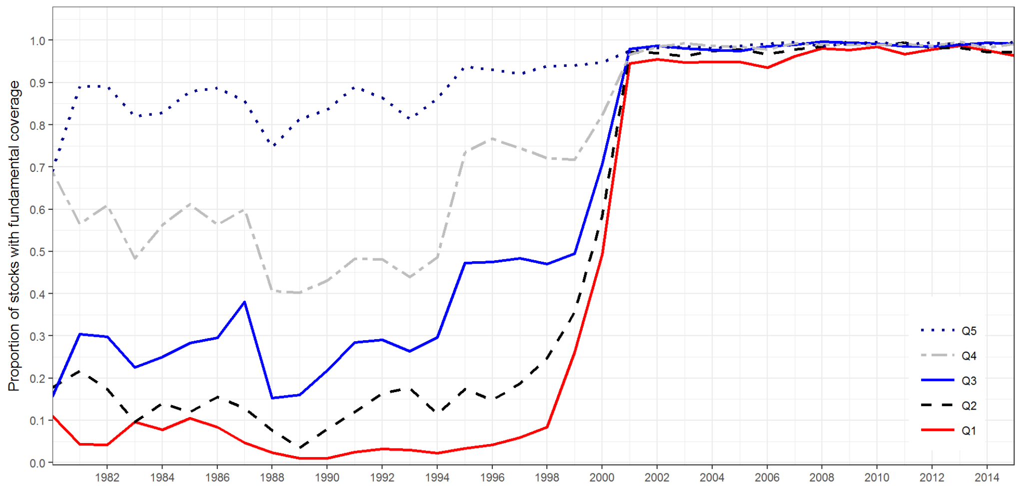


Figure 4. Fraction of stocks with fundamental coverage in Datastream in a given size quintile.

Electronic copy available at: https://ssrn.com/abstract=3150654

Table I
Quality of returns in Datastream

The table shows the correlation between returns in Datastream and CRSP in the US depending on the stock price at the end of the previous month and the fraction of returns that are winsorized every month. We separately focus on 3 periods: 1980 to 1989, 1990 to 1999, and 2000 to 2016.

	1980 - 1989				1990 - 1999				2000 - 2016			
Winsorize	All	\$.25+	\$1+	\$5+	All	\$.25+	\$1+	\$5+	All	\$.25+	\$1+	\$5+
None	0.930	0.946	0.961	0.970	0.966	0.972	0.987	0.992	0.996	0.996	0.996	0.996
.01%	0.937	0.950	0.962	0.971	0.973	0.978	0.989	0.992	0.995	0.995	0.996	0.996
.1%	0.953	0.960	0.968	0.976	0.961	0.976	0.987	0.991	0.977	0.979	0.984	0.994
1%	0.935	0.943	0.958	0.974	0.927	0.947	0.967	0.981	0.937	0.942	0.953	0.978

Table II
Impact of omitting delisting returns in CRSP

The tables show returns and their corresponding t-statistics among portfolios created from sorts on fundamental anomalies. We compare two ways of adjusting for delisting returns with respect to our adjustment. The first one is with all delisting returns set equal to zero and the second one follows Shumway (1997). We also show the correlation between portfolios in the two comparisons. The list of anomalies is provided in Appendix A. The source of fundamental data is Compustat. The portfolios are constructed by buying stocks in the top quintile of the signal and shorting stocks in the bottom quintile of the signal. The sample period is July 1963 to December 2016. The standard errors in t-statistics are adjusted for autocorrelation and heteroskedasticity, as in Newey and West (1987), with 12 lags.

	Our delisting adjustment versus no delisting							Our delisting adjustment versus Shumway (1997)						
		Return			t-stat				Return			t-stat		
	Corr	Our	No Delist	Diff	Our	No Delist	Diff	Corr	Our	Shumway	Diff	Our	Shumway	Diff
Accruals	0.998	0.50	0.51	0.01	4.73	4.83	0.10	0.999	0.50	0.50	-0.00	4.73	4.68	-0.05
	(0.000)	(0.06)	(0.06)	(0.00)	(0.39)	(0.38)	(0.02)	(0.000)	(0.06)	(0.06)	(0.00)	(0.39)	(0.40)	(0.01)
Profitability	0.999	0.35	0.34	-0.01	2.63	2.54	-0.09	1.000	0.35	0.36	0.01	2.63	2.66	0.03
	(0.000)	(0.05)	(0.05)	(0.01)	(0.35)	(0.35)	(0.03)	(0.000)	(0.05)	(0.05)	(0.00)	(0.35)	(0.35)	(0.01)
Value	1.000	0.62	0.63	0.00	3.41	3.45	0.03	1.000	0.62	0.62	0.00	3.41	3.41	-0.00
	(0.000)	(0.08)	(0.08)	(0.01)	(0.35)	(0.34)	(0.03)	(0.000)	(0.08)	(0.08)	(0.00)	(0.35)	(0.35)	(0.01)
Investment	0.999	0.45	0.45	0.00	3.89	3.90	0.00	0.999	0.45	0.45	-0.00	3.89	3.89	-0.01
	(0.000)	(0.06)	(0.06)	(0.00)	(0.24)	(0.26)	(0.03)	(0.000)	(0.06)	(0.06)	(0.00)	(0.24)	(0.24)	(0.01)
Intangibles	0.999	0.40	0.40	0.01	2.68	2.70	0.02	1.000	0.40	0.40	-0.00	2.68	2.66	-0.02
	(0.000)	(0.07)	(0.07)	(0.00)	(0.39)	(0.38)	(0.03)	(0.000)	(0.07)	(0.07)	(0.00)	(0.39)	(0.39)	(0.01)
All	0.999	0.47	0.47	0.00	3.58	3.61	0.02	0.999	0.47	0.47	-0.00	3.58	3.57	-0.01
	(0.000)	(0.03)	(0.03)	(0.00)	(0.18)	(0.18)	(0.01)	(0.000)	(0.03)	(0.03)	(0.00)	(0.18)	(0.18)	(0.01)

Electronic copy available at: https://ssrn.com/abstract=3150654

Table III
Variables from Compustat mapped onto Datastream

The table shows all fundamental variables that were required for construction of our fundamental anomalies. We first specify their name in Compustat and then document how we construct them in Datastream. We also show Pearson's and Spearman's correlation coefficients between variables in the two databases in our merged sample. The sample spans from January 1989 to December 2016.

		Pearson	Spearman	
BALANCE SHEET				
ASSETS				
Current Assets				
Cash and Short-Term Investments	CHE	0.619	0.990	WC02001
Short-Term Investments	IVST	0.556	0.764	WC02008
Receivables - Total	RECT	0.770	0.984	WC02051
Inventories - Total	INVT	0.824	0.972	WC02101
Current Assets - Other - Total	ACO	0.804	0.964	WC02149 + WC02140
Prepaid Expenses	XPP	0.912	0.911	WC02140
Current Assets - Total	ACT	1.000	1.000	WC02201
Non-Current Assets				
Long-Term Investments	IVAO	0.866	0.745	WC02258 + WC02250
Property Plant and Equipment - Total (Net)	PPENT	0.993	0.997	WC02501
Property Plant and Equipment - Total (Gross)	PPEGT	0.997	0.998	WC02301
Property Plant and Equipment Buildings at Cost	FATB	0.997	0.993	WC18376
Property Plant and Equipment Leases at Cost	FATL	0.771	0.754	WC18381
Investment and Advances - Equity	IVAEQ	0.941	0.846	WC02256
Intangible Assets - Total	INTAN	0.994	0.966	WC02649
Goodwill	GDWL			Set equal to 0
Assets - Total	AT	0.982	1.000	WC02999
LIABILITIES AND SHAREHOLDERS' EQUITY				
Current Liabilities				
Debt in Current Liabilities	DLC	0.961	0.953	WC03051
Account Payable/Creditors - Trade	AP	0.884	0.993	WC03040
Current Liabilities - Other - Total	LCO	0.952	0.991	WC03066 + WC03054 + WC03063 + WC03061
Accrued Expenses	XACC			Set equal to 0
Income Taxes Payable	TXP	0.937	0.860	WC03063
Current Liabilities - Total	LCT	1.000	0.999	WC03101
Long-Term Liabilities				
Long-Term Debt - Total	DLTT	0.985	0.988	WC03251
Liabilities - Other	LO	0.633	0.892	WC03273 + WC03262
Liabilities - Total	LT	0.998	0.998	WC03351
Minority Interest - Balance Sheet	MIB	0.763	0.791	WC03426
Shareholders' Equity				
Preferred/Preference Stock (Capital) - Total	PSTK	0.816	0.898	WC03451
Retained Earnings	RE	0.994	0.990	WC03495
Shareholders' Equity - Total	SEQ	0.995	0.999	WC03501 + WC03451
Common/Ordinary Equity - Total	CEQ	0.995	0.998	WC03501
Deffered Revenue Current	DRC			Set equal to 0
Deffered Revenue Long-Term	DRLT	0.307	0.683	WC03262
Preferred Stock Redemption Value	PSTKRV	0.877	0.914	Set equal to PSTK
Preferred Stock Liquidating Value	PSTKL	0.878	0.914	Set equal to PSTK

Electronic copy available at: https://ssrn.com/abstract=3150654

		Pearson	Spearman	
INCOME STATEMENT				
Revenue - Total	REVT			Set equal to SALE
Sales/Turnover (Net)	SALE	0.999	0.999	WC01001
Cost of Goods Sold	COGS	0.990	0.969	WC01051
Selling, General and Administrative Expenses	XSGA	0.989	0.982	WC01101
Research and Development Expense	XRD	0.986	0.983	WC01201
Earnings Before Interest, Taxes & Depreciation	OIBDP	0.963	0.983	WC01151 + WC01250
Depreciation and Amortization - Total	DP	0.989	0.992	WC01151
Earnings Before Interest and Taxes	OIADP	0.925	0.971	WC01250
Interest and Related Expense	XINT	0.885	0.993	WC01251
Pretax Income	PI	0.994	0.992	WC01401
Income Taxes - Total	TXT	0.997	0.995	WC01451
Income Before Extraordinary Items	IB	0.995	0.990	WC01551
CASH FLOW STATEMENT				
Indirect Operating Activities				
Operating Activities - Net Cash Flow	OANCF	0.990	0.996	WC04860
Investing Activities				
Capital Expenditures	CAPX	0.976	0.992	WC04601
Investing Activities - Net Cash Flow	IVNCF	0.990	0.994	- WC04870
Financing Activities				
Purchase of Common and Preferred Stock	PRSTKC	0.981	0.967	WC04751
Sale of Common and Preferred Stock	SSTK	0.928	0.960	WC04251
Cash Dividends	DV	0.998	0.992	WC04551
Dividends on Common Stock	DVC	0.987	0.985	WC05376
Long-Term Debt - Issuance	DLTIS	0.946	0.944	WC04401
Long-Term Debt - Reduction	DLTR	0.915	0.948	WC04701
Net Changes in Current Debt	DLCCH			WC04821
Financing Activities - Net Cash Flow	FINCF	0.987	0.991	WC04890
OTHER ITEMS				
Book Value per Share	BKVLPS	0.921	0.982	WC05476
SIC Industry Classification	SIC			WC07023
Earnings per Share	EPSPX	0.956	0.983	WC05210
Earnings per Share after Extraordinary Items	EPSPI	0.942	0.987	WC05230
Employees	EMP	0.937	0.992	WC07011
Net Income	NI			Set equal to IB
Preffered Dividends in Arrears	DVPA			Set equal to 0
Treasury Stock - Preferred	TSTKP			Set equal to 0

Electronic copy available at: https://ssrn.com/abstract=3150654

Table IV
Datastream vs Compustat in the common sample

The table shows the average coefficients form Fama-Mecbeth regressions of fundamental signals for cross-sectional regressions of fundamental signals from Compustat on returns with either OLS or instrumental variable regressions where we instrument with signals from Datastream. We regress the cross-sectional quantiles of fundamental signals rather than their raw values. The tables also shows returns and their corresponding t-statistics among portfolios created from sorts on fundamental anomalies. We consider 3 cases for the comparison. First, we compare portfolios created with CRSP & Compustat or with just Datastream in Panel A. We then decompose the overall difference in Panel B by using CRSP returns for both sources of fundamental data or Compustat fundamental signals for both sources of data on returns. We also show the correlation between the two cases. The list of anomalies is provided in Appendix A. The source of fundamental data is either Compustat (CS) or Datastream (DS). The portfolios are constructed by buying stocks in the top quintile of the signal and shorting stocks in the bottom quintile of the signal. The sample period is July 1990 to December 2016. The standard errors in t-statistics are adjusted for autocorrelation and heteroskedasticity, as in Newey and West (1987), with 12 lags.

Panel A

	Signals								Portfolios						
	Correlation		Fama-Mecbeth regressions						CS + CRSP or full DS						
			Coefficients			t-stat				Return			t-stat		
	Pears	Spear	CS	IV	Diff	CS	IV	Diff	Corr	CS	DS	Diff	CS	DS	Diff
Accruals	0.917	0.934	0.56	0.57	0.01	3.25	3.21	-0.04	0.950	0.60	0.59	-0.01	3.25	3.24	-0.00
	(0.011)	(0.018)	(0.07)	(0.07)	(0.01)	(0.26)	(0.28)	(0.06)	(0.018)	(0.10)	(0.09)	(0.01)	(0.26)	(0.24)	(0.07)
Profitability	0.837	0.951	-0.10	-0.10	0.00	-0.05	-0.01	0.04	0.964	0.01	0.00	-0.01	0.22	0.22	0.01
	(0.032)	(0.012)	(0.13)	(0.13)	(0.01)	(0.47)	(0.47)	(0.04)	(0.015)	(0.13)	(0.12)	(0.02)	(0.41)	(0.40)	(0.06)
Value	0.707	0.972	0.50	0.51	0.01	1.74	1.75	0.01	0.994	0.64	0.62	-0.02	1.97	1.91	-0.06
	(0.056)	(0.006)	(0.15)	(0.16)	(0.01)	(0.51)	(0.52)	(0.02)	(0.002)	(0.14)	(0.14)	(0.01)	(0.42)	(0.42)	(0.03)
Investment	0.905	0.960	0.38	0.40	0.01	2.24	2.30	0.06	0.977	0.37	0.36	-0.01	2.02	2.06	0.04
	(0.028)	(0.007)	(0.12)	(0.12)	(0.01)	(0.60)	(0.61)	(0.03)	(0.007)	(0.09)	(0.09)	(0.01)	(0.48)	(0.49)	(0.07)
Intangibles	0.809	0.881	0.53	0.56	0.03	1.62	1.61	-0.01	0.909	0.51	0.51	-0.00	1.50	1.45	-0.05
	(0.050)	(0.034)	(0.17)	(0.18)	(0.03)	(0.50)	(0.50)	(0.08)	(0.036)	(0.15)	(0.15)	(0.03)	(0.42)	(0.42)	(0.07)
All	0.841	0.939	0.40	0.41	0.01	1.92	1.92	0.01	0.959	0.45	0.44	-0.01	1.93	1.92	-0.01
	(0.018)	(0.009)	(0.06)	(0.06)	(0.01)	(0.24)	(0.24)	(0.02)	(0.009)	(0.06)	(0.06)	(0.01)	(0.21)	(0.21)	(0.03)

Panel B

	CRSP returns							Compustat signals						
		Return			t-stat				Return			t-stat		
	Corr	CS	DS	Diff	CS	DS	Diff	Corr	CS	DS	Diff	CS	DS	Diff
Accruals	0.957	0.60	0.60	-0.01	3.25	3.24	-0.00	0.990	0.60	0.60	-0.01	3.25	3.24	-0.00
	(0.018)	(0.10)	(0.10)	(0.01)	(0.26)	(0.24)	(0.07)	(0.002)	(0.10)	(0.09)	(0.01)	(0.26)	(0.26)	(0.02)
Profitability	0.969	0.01	0.00	-0.01	0.22	0.22	0.01	0.995	0.01	-0.00	-0.01	0.21	0.18	-0.03
	(0.015)	(0.13)	(0.12)	(0.02)	(0.41)	(0.39)	(0.07)	(0.002)	(0.13)	(0.13)	(0.01)	(0.42)	(0.43)	(0.02)
Value	0.995	0.64	0.62	-0.01	1.97	1.93	-0.04	0.997	0.64	0.63	-0.01	1.96	1.95	-0.01
	(0.001)	(0.14)	(0.14)	(0.01)	(0.42)	(0.42)	(0.02)	(0.001)	(0.14)	(0.14)	(0.00)	(0.42)	(0.42)	(0.02)
Investment	0.982	0.37	0.36	-0.01	2.02	2.05	0.03	0.994	0.36	0.36	-0.00	2.04	2.06	0.01
	(0.005)	(0.09)	(0.10)	(0.01)	(0.48)	(0.50)	(0.06)	(0.002)	(0.09)	(0.09)	(0.00)	(0.49)	(0.49)	(0.02)
Intangibles	0.912	0.51	0.51	0.00	1.50	1.44	-0.06	0.998	0.51	0.50	-0.00	1.51	1.51	-0.00
	(0.036)	(0.15)	(0.15)	(0.03)	(0.42)	(0.42)	(0.07)	(0.001)	(0.15)	(0.15)	(0.00)	(0.42)	(0.42)	(0.01)
All	0.963	0.45	0.44	-0.01	1.93	1.92	-0.01	0.994	0.45	0.44	-0.01	1.94	1.93	-0.01
	(0.009)	(0.06)	(0.06)	(0.01)	(0.21)	(0.20)	(0.03)	(0.001)	(0.06)	(0.06)	(0.00)	(0.21)	(0.21)	(0.01)

Electronic copy available at: https://ssrn.com/abstract=3150654

Table V
Datastream vs Compustat in their own full samples

The table shows the returns and their corresponding t-statistics among portfolios created from sorts on fundamental anomalies. We compare portfolios created with CRSP & Compustat or with just Datastream for either all available stocks or for a reduced sample. The full sample starts in July 1990 and ends in December 2016. The reduced sample begins in July 2010 and omits all financial stocks or those with capitalization under $100 million. We also show correlation between the two cases. The list of anomalies is provided in Appendix A. The portfolios are constructed by buying stocks in the top quintile of the signal and shorting stocks in the bottom quintile of the signal. The standard errors in t-statistics are adjusted for autocorrelation and heteroskedasticity, as in Newey and West (1987), with 12 lags.

	Full samples							Cap over $100 million & no financial & 2010+						
		Return			t-stat				Return			t-stat		
	Corr	CS	DS	Diff	CS	DS	Diff	Corr	CS	DS	Diff	CS	DS	Diff
Accruals	0.762	0.56	0.62	0.06	3.27	3.17	-0.10	0.874	0.11	0.11	-0.00	0.58	0.65	0.08
	(0.031)	(0.08)	(0.10)	(0.06)	(0.26)	(0.28)	(0.36)	(0.022)	(0.03)	(0.03)	(0.02)	(0.15)	(0.19)	(0.10)
Profitability	0.841	0.36	-0.01	-0.37	1.41	0.35	-1.06	0.911	0.26	0.31	0.05	0.80	1.24	0.44
	(0.028)	(0.08)	(0.14)	(0.16)	(0.23)	(0.47)	(0.43)	(0.022)	(0.07)	(0.08)	(0.04)	(0.33)	(0.33)	(0.16)
Value	0.899	0.64	0.63	-0.01	2.04	1.98	-0.06	0.960	0.19	0.22	0.03	0.63	0.80	0.16
	(0.022)	(0.10)	(0.15)	(0.15)	(0.26)	(0.46)	(0.46)	(0.008)	(0.07)	(0.06)	(0.02)	(0.29)	(0.25)	(0.11)
Investment	0.815	0.49	0.46	-0.03	2.70	2.55	-0.16	0.912	0.22	0.22	-0.00	1.13	1.24	0.11
	(0.035)	(0.07)	(0.11)	(0.10)	(0.26)	(0.53)	(0.46)	(0.023)	(0.05)	(0.04)	(0.02)	(0.24)	(0.25)	(0.13)
Intangibles	0.713	0.41	0.58	0.16	1.63	1.67	0.04	0.845	0.12	0.12	-0.01	0.15	0.10	-0.05
	(0.084)	(0.09)	(0.16)	(0.12)	(0.31)	(0.46)	(0.48)	(0.063)	(0.08)	(0.06)	(0.03)	(0.43)	(0.42)	(0.11)
All	0.802	0.50	0.48	-0.02	2.31	2.08	-0.23	0.898	0.17	0.19	0.01	0.66	0.79	0.13
	(0.021)	(0.04)	(0.06)	(0.05)	(0.14)	(0.22)	(0.19)	(0.015)	(0.03)	(0.02)	(0.01)	(0.13)	(0.13)	(0.06)

Table VI
Explaining the difference in returns across Datastream and Compustat

The table shows the results from regressions of differences in the returns of portfolios from alternative databases. The portfolios are created from sorts on fundamental anomalies constructed with data from either CRSP and Compustat or with just Datastream. We then regress the monthly returns from Datastream minus the returns from Compustat on size in Compustat or the difference in size across the two databases. The size is measured as the mean cross-sectional quantile of the size of stocks in the portfolio with respect to the full universe of US stocks at the beginning of each month. The list of anomalies is provided in Appendix A. The portfolios are constructed by buying stocks in the top quintile of the signal and shorting stocks in the bottom quintile of the signal. The sample period is July 1990 to December 2016. The standard errors in regressions are clustered at time and anomaly effects.

	I	II	III
Intercept	-1.48	-0.06	-1.44
	(-3.44)	(-0.67)	(-3.25)
Size	2.69		2.79
	(3.35)		(3.42)
Difference in Size		-1.26	-1.37
		(-2.19)	(-2.50)
R^2	0.0038	0.0016	0.0056

Electronic copy available at: https://ssrn.com/abstract=3150654

Table VII
Portfolio constructions reducing the discrepancy between databases

The table shows the returns and their corresponding t-statistics among portfolios created from sorts on fundamental anomalies. We compare the portfolios created with CRSP & Compustat or with just Datastream for either the large cap universe of stocks or for the full sample of stocks with breakpoints from the largest 1000 stocks. The full sample includes all available stocks while the large cap universe is restricted to stocks with capitalization larger than that of bottom decile at NYSE. We also show the correlation between the two cases. The sample starts in July 1990 and ends in December 2016. The list of anomalies is provided in Appendix A. The value-weighted portfolios are constructed by buying stocks in the top quintile of the signal and shorting stocks in the bottom quintile of the signal. The standard errors in t-statistics are adjusted for autocorrelation and heteroskedasticity, as in Newey and West (1987), with 12 lags.

	Large VW							Breakpoints from 1000 largest stocks VW						
		Return			t-stat				Return			t-stat		
	Corr	CS	DS	Diff	CS	DS	Diff	Corr	CS	DS	Diff	CS	DS	Diff
Accruals	0.858	0.23	0.22	-0.01	1.34	1.23	-0.11	0.868	0.21	0.22	0.01	1.21	1.21	0.00
	(0.018)	(0.04)	(0.04)	(0.02)	(0.24)	(0.21)	(0.12)	(0.017)	(0.05)	(0.05)	(0.02)	(0.21)	(0.19)	(0.12)
Profitability	0.869	0.24	0.17	-0.07	1.08	0.86	-0.21	0.875	0.22	0.15	-0.07	1.09	0.81	-0.27
	(0.030)	(0.06)	(0.06)	(0.04)	(0.23)	(0.26)	(0.16)	(0.030)	(0.05)	(0.07)	(0.04)	(0.22)	(0.28)	(0.16)
Value	0.948	0.19	0.23	0.03	0.69	0.84	0.14	0.942	0.19	0.20	0.01	0.73	0.80	0.07
	(0.008)	(0.03)	(0.03)	(0.03)	(0.09)	(0.13)	(0.10)	(0.017)	(0.02)	(0.03)	(0.03)	(0.09)	(0.15)	(0.09)
Investment	0.851	0.26	0.23	-0.02	1.70	1.53	-0.17	0.854	0.22	0.21	-0.02	1.62	1.49	-0.12
	(0.026)	(0.03)	(0.04)	(0.03)	(0.22)	(0.28)	(0.16)	(0.028)	(0.03)	(0.03)	(0.01)	(0.24)	(0.29)	(0.12)
Intangibles	0.762	0.17	0.26	0.09	0.82	1.13	0.31	0.778	0.21	0.23	0.02	1.09	1.02	-0.07
	(0.069)	(0.05)	(0.06)	(0.04)	(0.21)	(0.22)	(0.19)	(0.068)	(0.04)	(0.05)	(0.03)	(0.19)	(0.22)	(0.15)
All	0.857	0.22	0.22	0.01	1.15	1.14	-0.01	0.863	0.21	0.20	-0.01	1.16	1.09	-0.07
	(0.017)	(0.02)	(0.02)	(0.01)	(0.10)	(0.10)	(0.07)	(0.017)	(0.02)	(0.02)	(0.01)	(0.10)	(0.10)	(0.06)

Electronic copy available at: https://ssrn.com/abstract=3150654

Table VIII
Predicting fundamental coverage

The table reports the estimated coefficients and corresponding t-statistics for the stock level regression of fundamental coverage on characteristics

$$Fundamental\ Coverage_{it} = \beta_0 + \beta_1 \mathbb{1}\{Size_{it} > \$100M\} + \beta_2 \mathbb{1}\{Analysts\ Following_{it} \geq 5\}$$
$$+ \beta_3 Size\ Quantile_{it} + \beta_4 (Size\ Quantile_{it} - Size\ Quantile_t^{\$100M}) \mathbb{1}\{Size_{it} > \$100M\} + \epsilon_{it}$$

We also report the Nagelkerke et al. (1991) R^2 index to measure goodness of fit. The standard errors in the reported t-statistics are HAC robust.

	1990 - 2002						2003 - 2016					
	β_0	β_1	β_2	β_3	$\beta_4+\beta_3$	R^2	β_0	β_1	β_2	β_3	$\beta_4+\beta_3$	R^2
Australia	-1.87	0.21	2.12	2.92	4.01	0.40	1.35	-0.22	1.72	3.12	-1.69	0.08
	(-26.38)	(1.16)	(9.74)	(14.69)	(2.82)		(23.64)	(-1.10)	(4.20)	(13.77)	(-1.49)	
Austria	0.71	0.45	0.59	-0.15	1.12	0.06	-0.69	-0.09	2.74	5.06	0.71	0.31
	(2.00)	(1.12)	(1.77)	(-0.14)	(0.50)		(-1.54)	(-0.17)	(2.75)	(3.57)	(0.33)	
Belgium	-1.00	-0.61	2.44	4.69	0.02	0.31	-1.40	-0.30	0.41	7.90	3.09	0.43
	(-3.37)	(-1.39)	(3.96)	(4.95)	(0.01)		(-4.14)	(-0.55)	(0.64)	(7.01)	(1.58)	
Canada	-2.11	0.35	1.24	4.25	3.99	0.45	0.60	-0.10	1.15	6.53	1.18	0.17
	(-28.16)	(2.33)	(8.74)	(14.56)	(4.94)		(6.50)	(-0.68)	(4.63)	(9.68)	(2.21)	
Denmark	0.06	-0.55	1.76	4.04	-1.70	0.16	0.96	-1.04	2.13	7.31	-1.83	0.15
	(0.26)	(-1.12)	(4.39)	(5.07)	(-0.68)		(3.10)	(-1.61)	(3.11)	(6.02)	(-0.83)	
Finland	0.38	-0.95	1.20	3.13	-0.05	0.10	0.69	-0.86	1.94	6.54	-1.77	0.15
	(1.05)	(-2.27)	(3.33)	(2.61)	(-0.02)		(1.39)	(-1.34)	(2.78)	(3.64)	(-0.81)	
France	-0.36	-0.59	2.09	2.73	-0.26	0.16	-0.52	-0.23	1.26	5.46	1.45	0.27
	(-2.88)	(-3.65)	(9.74)	(7.48)	(-0.42)		(-4.79)	(-1.18)	(3.54)	(14.06)	(1.71)	
Germany	0.50	-0.43	1.58	1.92	0.82	0.10	-1.26	-1.18	1.71	7.80	3.07	0.45
	(2.79)	(-2.21)	(6.73)	(3.65)	(1.28)		(-14.00)	(-5.64)	(3.86)	(19.87)	(2.98)	
Greece	-0.46	-0.24	1.82	3.36	1.25	0.17	1.19	0.03	0.56	3.78	1.51	0.09
	(-2.11)	(-1.00)	(5.08)	(5.87)	(1.12)		(4.63)	(0.06)	(1.43)	(3.73)	(0.72)	
Hong Kong	-1.85	-0.25	0.27	3.31	6.27	0.19	1.00	-0.10	2.01	3.02	-0.36	0.07
	(-9.02)	(-1.89)	(1.86)	(9.09)	(6.71)		(5.59)	(-0.69)	(6.37)	(7.88)	(-0.57)	
Ireland	-1.45	-0.63	1.00	8.41	-0.17	0.44	1.40	-1.14	0.18	3.45	2.20	0.04
	(-4.52)	(-1.01)	(1.56)	(7.12)	(-0.08)		(2.00)	(-1.96)	(0.26)	(1.93)	(1.14)	
Italy	-0.08	-0.06	2.17	2.83	1.15	0.19	-0.98	-0.21	1.49	7.67	0.67	0.28
	(-0.22)	(-0.24)	(6.04)	(3.28)	(1.12)		(-3.24)	(-0.61)	(2.13)	(8.87)	(0.37)	
Japan	0.03	-0.73	2.20	5.41	3.14	0.20	3.39	0.15	1.51	0.75	-1.53	0.01
	(0.39)	(-10.46)	(6.48)	(20.97)	(12.98)		(33.64)	(0.96)	(6.66)	(2.06)	(-3.12)	
Netherlands	-0.24	-1.06	2.74	3.31	0.04	0.32	-0.90	-2.75	1.61	12.05	-1.91	0.34
	(-0.79)	(-2.38)	(9.24)	(3.45)	(0.04)		(-2.75)	(-3.39)	(3.11)	(7.32)	(-1.32)	
New Zealand	-2.91	0.14	1.92	4.86	-4.26	0.48	0.69	0.04	0.76	2.41	5.47	0.16
	(-9.30)	(0.30)	(5.87)	(7.10)	(-1.82)		(3.00)	(0.07)	(0.93)	(3.71)	(2.07)	
Norway	0.34	-0.15	2.18	2.40	-0.45	0.14	0.99	-0.49	1.97	3.46	-0.75	0.10
	(1.59)	(-0.52)	(4.82)	(3.78)	(-0.28)		(3.22)	(-1.56)	(4.84)	(3.73)	(-0.67)	
Portugal	-1.48	-0.12	2.44	5.02	-1.54	0.36	-0.42	0.74	-0.11	5.10	3.64	0.38
	(-8.52)	(-0.29)	(4.79)	(7.05)	(-0.88)		(-1.28)	(0.90)	(-0.07)	(3.88)	(0.64)	
Singapore	-0.69	-0.55	0.71	1.95	6.14	0.13	2.02	-0.42	1.79	1.58	3.38	0.04
	(-2.10)	(-3.00)	(3.31)	(3.51)	(3.92)		(6.51)	(-1.78)	(2.70)	(2.40)	(2.28)	
Spain	-0.06	-0.72	1.83	2.43	1.81	0.23	-1.02	1.36	2.27	6.06	-4.59	0.33
	(-0.14)	(-1.78)	(4.74)	(2.26)	(1.45)		(-2.00)	(2.22)	(3.47)	(4.24)	(-2.20)	
Sweden	-0.36	-0.36	3.01	3.43	0.04	0.23	0.38	-0.56	1.03	6.31	-1.21	0.19
	(-2.52)	(-1.39)	(6.44)	(7.08)	(0.04)		(2.92)	(-1.98)	(2.05)	(11.78)	(-1.18)	
Switzerland	-0.37	-0.75	1.66	3.51	0.76	0.19	1.71	0.37	1.23	2.81	-1.49	0.05
	(-1.34)	(-2.52)	(5.76)	(4.75)	(0.67)		(3.43)	(0.62)	(2.12)	(1.95)	(-0.60)	
UK	0.17	0.16	0.68	2.88	-0.31	0.11	0.68	-0.48	0.53	3.71	0.87	0.09
	(2.31)	(1.28)	(4.56)	(11.10)	(-0.61)		(9.88)	(-3.94)	(3.33)	(14.81)	(1.96)	
USA	-0.33	0.57	1.09	1.87	1.35	0.22	1.22	-0.40	0.56	4.57	0.74	0.05
	(-6.83)	(10.68)	(15.11)	(12.93)	(5.25)		(15.66)	(-5.23)	(7.09)	(12.00)	(4.00)	

Electronic copy available at: https://ssrn.com/abstract=3150654

Table IX
Independently significant signals

The table shows signals that independently predict the returns on individual stocks in different regions. We measure predictability by significance of coefficients in the Fama and MacBeth (1973) regressions. We regress the returns on past quantiles of fundamental signals across all stocks in the given region and month. We then focus on the t-statistics on the time-series mean of these coefficients. We report all signals with t-statistics larger than 2 (U) and those with p-values smaller than 5% after adjusting the original p-values for FDR (**A**). The regressions are either equal-weighted (E, standard OLS) or value-weighted (V, WLS with weights given by market cap). We compare the selected signals for CRSP & Compustat with those for Datastream for either the large cap universe of stocks or for the full sample of stocks. The full sample (All) includes all available stocks, while the large cap universe (Large) is restricted to stocks with capitalizations larger than that of bottom decile of the NYSE. The sample starts in July 1990 and ends in December 2016. The list of anomalies is provided in Appendix A.

	Compustat				Datastream															
	USA				USA				Europe				Japan				Asia Pacific			
	All		Large		All		Large		All		Large		All		Large		All		Large	
	E	V	E	V	E	V	E	V	E	V	E	V	E	V	E	V	E	V	E	V
EPr	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	**A**	-	-	U	-	U	-	U	-
CBOP	**A**	U	U	U	U	U	**A**	U	**A**	-	**A**	U	-	-	U	-	-	-	-	-
NOA	-	U	U	U	**A**	U	**A**	U	**A**	-	-	-	-	-	U	-	**A**	-	-	-
SP	U	U	**A**	U	**A**	-	U	-	**A**	-	-	-	-	-	U	-	**A**	-	-	-
RDM	**A**	-	**A**	-	U	-	U	-	**A**	-	U	-	**A**	-	U	-	-	-	-	-
ChNOA	**A**	-	**A**	-	-	-	U	-	**A**	U	U	U	-	-	-	-	-	-	-	-
PY	U	-	**A**	-	**A**	-	-	-	U	-	U	-	-	-	-	-	-	-	-	-
BM	**A**	-	-	-	U	-	-	-	**A**	**A**	**A**	**A**	-	-	-	-	**A**	U	U	U
WWI	U	-	-	-	**A**	-	-	-	U	-	-	-	-	U	-	U	**A**	-	-	-
CM	-	-	-	-	**A**	-	-	-	-	-	U	-	-	-	-	-	**A**	-	-	-
OL	**A**	U	U	U	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SaGr	**A**	-	U	-	-	-	U	-	-	-	-	-	U	-	-	-	-	-	-	-
GriI	U	-	-	-	**A**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GrLTNOA	**A**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	U	-	-	-
AT	-	-	-	-	**A**	-	-	-	-	-	-	-	-	U	U	U	-	-	-	-
CEI5Y	-	-	-	-	-	-	U	-	**A**	U	**A**	-	-	-	U	-	U	U	U	U
ChGMChS	U	-	U	-	-	-	-	-	-	-	U	-	-	-	-	-	**A**	-	-	-
EP	-	-	-	-	-	U	-	U	**A**	-	U	-	-	-	-	-	U	-	-	-
NEF	-	-	-	-	U	-	-	-	U	-	-	-	-	-	-	-	**A**	-	-	-
SuGr	-	-	-	-	-	-	-	-	**A**	-	U	-	U	-	-	-	-	-	-	-
Acc	U	U	-	-	-	U	-	U	U	-	-	-	-	-	-	-	U	-	-	-
ChNNCOA	U	U	U	U	-	-	-	-	U	-	-	-	-	-	-	-	-	U	U	U
POA	U	-	U	-	-	-	U	-	-	-	-	-	-	-	-	-	-	-	U	U
NPY	U	-	-	-	U	-	U	-	-	-	-	-	-	-	-	-	-	-	-	-
AGr	U	-	-	-	U	-	-	-	U	-	-	-	-	-	-	-	-	-	-	-
ICh	U	-	-	-	U	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ES	-	-	-	-	-	U	-	U	-	-	-	-	-	-	-	-	-	-	-	-
OC	U	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TAN	U	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ChNCOL	U	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ChFL	U	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FSc	-	-	-	-	U	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HR	-	-	-	-	U	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Lvrg	-	-	-	-	U	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ChCOL	-	-	-	-	U	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ChPPEIA	-	-	-	-	-	U	-	-	U	-	-	-	-	-	-	-	-	-	-	-
EM	-	-	U	-	-	-	-	-	-	-	-	-	U	-	-	-	-	-	-	-
ChiAT	-	-	-	-	-	-	-	-	U	U	U	U	-	-	-	-	-	-	-	-
NOACh	-	-	-	-	-	-	-	-	U	-	-	-	U	-	-	-	-	-	-	-
TXFIN	-	-	-	-	-	-	-	-	U	-	-	-	-	-	-	-	-	-	-	-
AL	-	-	-	-	-	-	-	-	U	U	-	-	-	-	-	-	-	-	-	-
EC	-	-	-	-	-	-	-	-	-	-	-	-	U	-	-	-	-	-	-	-
IR	-	-	-	-	-	-	-	-	-	-	-	-	U	-	-	-	-	-	-	-
OPtE	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	U	U	U
ChNNCWC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	U	-	U
ICBE	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	U	U	U
CDI	-	-	-	-	-	-	-	-	-	-	-	U	-	-	-	-	-	-	-	-
HI	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	U
NDF	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	U

Electronic copy available at: https://ssrn.com/abstract=3150654

Table X
Firms without fundamental coverage in Datastream

The table shows returns and alphas with their corresponding t-statistics on long-short portfolios created from stocks in CRSP by buying those that have fundamental coverage in Datastream and shorting those that do not. The portfolios are either equal-weighted (EW) or value-weighted (VW). The full sample includes all available non-financial stocks while the large cap universe is restricted to stocks with capitalization larger than that of bottom decile of the NYSE. The sample spans from July 1990 to December 2016. The alpha is estimated with respect to the Fama-French five factor model, and the factor loadings are also provided. The reported returns are in percent per month. The standard errors in t-statistics are adjusted for autocorrelation and heteroskedasticity, as in Newey and West (1987), with 24 lags.

	Full sample		Large cap	
	EW	VW	EW	VW
Mean Return	1.45	0.44	0.89	0.40
	(7.95)	(3.71)	(6.35)	(3.43)
Alpha FF5	1.20	0.42	0.80	0.39
	(7.44)	(6.41)	(9.47)	(5.87)
Mkt	0.17	-0.03	-0.02	-0.04
	(7.94)	(-0.85)	(-0.90)	(-1.07)
SMB	-0.08	-0.30	-0.22	-0.27
	(-1.52)	(-7.58)	(-5.72)	(-6.86)
HML	0.01	-0.04	0.18	-0.05
	(0.20)	(-0.78)	(3.81)	(-0.85)
RMW	0.41	0.30	0.24	0.29
	(8.73)	(4.68)	(5.17)	(4.36)
CMA	0.06	0.00	0.08	0.01
	(0.78)	(0.06)	(1.33)	(0.22)
Avg # of Stocks	1581	1581	551	551
Min # of Stocks	342	342	178	178

Electronic copy available at: https://ssrn.com/abstract=3150654

Table XI
Low profitability firms without fundamental coverage in Datastream

The table shows returns and alphas with their corresponding t-statistics on portfolios created from stocks that are within the bottom decile (quintile) of profitability stocks in Compustat but do not have fundamental coverage in Datastream. The portfolios are either equal-weighted (EW) or value-weighted (VW). We measure profitability by operating profits to assets as in Ball et al. (2016). The full sample includes all available non-financial stocks while the large cap universe is restricted to stocks with capitalization larger than that of bottom decile of the NYSE. The sample spans from July 2000 to December 2016. The alpha is estimated with respect to the Fama-French five factor model, and the factor loadings are also provided. The reported returns are in percent per month. The standard errors in t-statistics are adjusted for autocorrelation and heteroskedasticity, as in Newey and West (1987), with 24 lags.

	Full sample				Large cap			
	Decile		Quintile		Decile		Quintile	
	EW	VW	EW	VW	EW	VW	EW	VW
Mean Return	-2.03	-2.32	-1.86	-1.42	-2.13	-2.37	-1.76	-1.28
	(-2.88)	(-3.21)	(-2.88)	(-2.00)	(-2.97)	(-2.86)	(-2.73)	(-1.74)
Alpha FF5	-2.30	-2.52	-2.11	-1.61	-2.46	-2.57	-2.03	-1.43
	(-5.69)	(-6.61)	(-5.99)	(-4.78)	(-5.00)	(-4.83)	(-6.55)	(-3.74)
Mkt	0.73	0.76	0.69	0.91	0.89	0.87	0.86	0.95
	(7.26)	(6.90)	(10.30)	(12.40)	(9.79)	(11.90)	(11.80)	(11.40)
SMB	0.99	1.06	0.90	0.69	1.22	1.06	0.91	0.60
	(7.33)	(9.51)	(9.73)	(9.39)	(8.99)	(8.13)	(10.80)	(5.97)
HML	-0.22	-0.40	-0.21	-0.36	-0.49	-0.28	-0.26	-0.34
	(-0.92)	(-2.25)	(-0.96)	(-1.91)	(-2.82)	(-1.19)	(-1.32)	(-1.65)
RMW	-1.03	-1.03	-0.90	-0.70	-0.72	-0.87	-0.67	-0.68
	(-5.94)	(-4.70)	(-7.36)	(-6.74)	(-4.98)	(-6.25)	(-6.45)	(-5.35)
CMA	0.57	0.49	0.48	0.22	0.38	0.21	0.08	0.09
	(2.22)	(2.61)	(2.14)	(1.15)	(2.08)	(0.99)	(0.46)	(0.50)
Avg # of Stocks	132	132	234	234	16	16	40	40
Min # of Stocks	24	24	49	49	3	3	5	5

Electronic copy available at: https://ssrn.com/abstract=3150654

Table XII
Robustness - different portfolio construction

The table shows the returns and their corresponding t-statistics among portfolios created from sorts on fundamental anomalies. We compare the portfolios created with CRSP & Compustat or with just Datastream. The portfolios are either value-weighted or equal-weighted with decile, quintile, or tercile breakpoints in sorts. We also show correlation between the two cases. The sample starts in July 1990 and ends in December 2016. The list of anomalies is provided in Appendix A. The portfolios are constructed by buying stocks in the top decile, quintile, or tercile of the signal and shorting stocks in the bottom decile, quintile, or tercile of the signal. The standard errors in t-statistics are adjusted for autocorrelation and heteroskedasticity, as in Newey and West (1987), with 12 lags.

	Equal-weighted portfolios							Value-weighted portfolios						
		Return			t-stat				Return			t-stat		
	Corr	CS	DS	Diff	CS	DS	Diff	Corr	CS	DS	Diff	CS	DS	Diff
Panel A: Decile Portfolios														
Accruals	0.716	0.67	0.73	0.06	3.15	2.83	-0.32	0.782	0.37	0.35	-0.01	1.54	1.42	-0.12
	(0.029)	(0.08)	(0.12)	(0.08)	(0.28)	(0.31)	(0.40)	(0.020)	(0.06)	(0.06)	(0.03)	(0.23)	(0.22)	(0.13)
Profitability	0.741	0.52	-0.01	-0.52	1.66	0.35	-1.31	0.824	0.48	0.24	-0.24	1.46	0.91	-0.55
	(0.049)	(0.13)	(0.16)	(0.21)	(0.35)	(0.46)	(0.55)	(0.036)	(0.11)	(0.08)	(0.12)	(0.32)	(0.30)	(0.28)
Value	0.846	0.80	0.78	-0.02	2.18	1.93	-0.25	0.859	0.26	0.31	0.06	0.78	0.92	0.14
	(0.023)	(0.13)	(0.22)	(0.20)	(0.31)	(0.50)	(0.57)	(0.017)	(0.07)	(0.08)	(0.07)	(0.23)	(0.22)	(0.19)
Investment	0.770	0.61	0.60	-0.00	2.72	2.61	-0.11	0.758	0.39	0.31	-0.08	1.86	1.45	-0.41
	(0.044)	(0.10)	(0.15)	(0.12)	(0.28)	(0.52)	(0.46)	(0.041)	(0.07)	(0.07)	(0.06)	(0.34)	(0.29)	(0.23)
Intangibles	0.604	0.49	0.80	0.31	1.62	1.75	0.14	0.618	0.12	0.34	0.22	0.58	0.91	0.34
	(0.107)	(0.12)	(0.22)	(0.19)	(0.33)	(0.45)	(0.57)	(0.083)	(0.14)	(0.11)	(0.14)	(0.39)	(0.23)	(0.41)
All	0.734	0.62	0.61	-0.01	2.35	2.01	-0.34	0.767	0.32	0.32	-0.00	1.26	1.15	-0.11
	(0.026)	(0.05)	(0.08)	(0.07)	(0.15)	(0.22)	(0.22)	(0.021)	(0.04)	(0.04)	(0.04)	(0.14)	(0.12)	(0.11)
Panel B: Quintile Portfolios														
Accruals	0.762	0.56	0.62	0.06	3.27	3.17	-0.10	0.834	0.27	0.25	-0.01	1.52	1.37	-0.15
	(0.031)	(0.08)	(0.10)	(0.06)	(0.26)	(0.28)	(0.36)	(0.015)	(0.05)	(0.05)	(0.02)	(0.24)	(0.20)	(0.12)
Profitability	0.827	0.38	-0.03	-0.41	1.51	0.28	-1.24	0.868	0.30	0.13	-0.16	1.13	0.70	-0.43
	(0.032)	(0.09)	(0.14)	(0.16)	(0.30)	(0.46)	(0.46)	(0.028)	(0.07)	(0.06)	(0.08)	(0.22)	(0.25)	(0.23)
Value	0.899	0.64	0.63	-0.01	2.04	1.98	-0.06	0.908	0.25	0.28	0.02	0.85	0.98	0.13
	(0.022)	(0.10)	(0.15)	(0.15)	(0.26)	(0.46)	(0.46)	(0.011)	(0.06)	(0.05)	(0.06)	(0.17)	(0.18)	(0.18)
Investment	0.815	0.49	0.46	-0.03	2.70	2.55	-0.16	0.835	0.23	0.23	-0.01	1.55	1.43	-0.12
	(0.035)	(0.07)	(0.11)	(0.10)	(0.26)	(0.53)	(0.46)	(0.027)	(0.04)	(0.05)	(0.04)	(0.27)	(0.28)	(0.19)
Intangibles	0.713	0.41	0.58	0.16	1.63	1.67	0.04	0.733	0.16	0.30	0.14	0.74	1.11	0.37
	(0.084)	(0.09)	(0.16)	(0.12)	(0.31)	(0.46)	(0.48)	(0.068)	(0.08)	(0.07)	(0.09)	(0.25)	(0.22)	(0.29)
All	0.800	0.50	0.48	-0.03	2.33	2.07	-0.26	0.834	0.24	0.24	0.00	1.19	1.15	-0.04
	(0.021)	(0.04)	(0.06)	(0.05)	(0.14)	(0.22)	(0.20)	(0.016)	(0.03)	(0.03)	(0.03)	(0.11)	(0.10)	(0.09)
Panel C: Tercile Portfolios														
Accruals	0.754	0.47	0.51	0.04	2.99	2.98	-0.01	0.853	0.17	0.19	0.02	1.05	1.12	0.06
	(0.053)	(0.08)	(0.09)	(0.06)	(0.29)	(0.29)	(0.37)	(0.021)	(0.05)	(0.04)	(0.02)	(0.20)	(0.13)	(0.13)
Profitability	0.835	0.32	-0.01	-0.33	1.60	0.34	-1.26	0.886	0.22	0.16	-0.07	1.13	0.91	-0.22
	(0.034)	(0.07)	(0.11)	(0.12)	(0.27)	(0.44)	(0.41)	(0.028)	(0.05)	(0.05)	(0.04)	(0.23)	(0.25)	(0.16)
Value	0.914	0.54	0.50	-0.03	1.97	1.96	-0.01	0.935	0.18	0.19	0.00	0.73	0.76	0.02
	(0.018)	(0.08)	(0.13)	(0.12)	(0.24)	(0.46)	(0.42)	(0.010)	(0.03)	(0.04)	(0.03)	(0.11)	(0.16)	(0.10)
Investment	0.831	0.39	0.36	-0.03	2.43	2.40	-0.04	0.867	0.21	0.19	-0.01	1.53	1.48	-0.04
	(0.031)	(0.05)	(0.10)	(0.08)	(0.24)	(0.54)	(0.44)	(0.025)	(0.03)	(0.04)	(0.02)	(0.22)	(0.31)	(0.17)
Intangibles	0.772	0.35	0.44	0.09	1.55	1.53	-0.02	0.766	0.12	0.23	0.11	0.69	1.07	0.37
	(0.064)	(0.08)	(0.13)	(0.08)	(0.34)	(0.44)	(0.42)	(0.062)	(0.05)	(0.05)	(0.05)	(0.22)	(0.19)	(0.16)
All	0.817	0.42	0.38	-0.04	2.19	1.97	-0.22	0.860	0.18	0.19	0.01	1.03	1.08	0.05
	(0.021)	(0.03)	(0.05)	(0.04)	(0.14)	(0.21)	(0.19)	(0.016)	(0.02)	(0.02)	(0.01)	(0.09)	(0.10)	(0.07)

Panel D: Number of significant signals

	Equal-weighted			Value-weighted		
	CT	DS	both	CT	DS	both
Decile portfolios	44	39	30	14	13	7
Quintile portfolios	41	39	29	11	12	7
Tercile portfolios	38	38	26	9	5	3

Electronic copy available at: https://ssrn.com/abstract=3150654

Table XIII
List of Published Fundamental Anomalies

Accruals		
Acc	Accruals	Sloan (1996)
ChCE	Change in Common Equity	Richardson, Sloan, Soliman, and Tuna (2006)
ChCOA	Change in Current Operating Assets	Richardson et al. (2006)
ChCOL	Change in Current Operating Liabilities	Richardson et al. (2006)
ChFL	Change in Financial Liabilities	Richardson et al. (2006)
ChLTI	Change in Long-Term Investments	Richardson et al. (2006)
ChNFA	Change in Net Financial Assets	Richardson et al. (2006)
ChNNCWC	Change in Net Non-Cash Working Capital	Richardson et al. (2006)
ChNNCOA	Change in Net Non-Current Operating Assets	Richardson et al. (2006)
ChNCOA	Change in Non-Current Operating Assets	Richardson et al. (2006)
ChNCOL	Change in Non-Current Operating Liabilities	Richardson et al. (2006)
GriI	Growth in Inventory	Thomas and Zhang (2002)
ICh	Inventory Change	Thomas and Zhang (2002)
IGr	Inventory Growth	Belo and Lin (2011)
MBaAC	M/B and Accruals	Bartov and Kim (2004)
NWCCh	Net Working Capital Changes	Soliman (2008)
POA	Percent Operating Accrual	Hafzalla, Lundholm, and Matthew Van Winkle (2011)
PTA	Percent Total Accrual	Hafzalla et al. (2011)
TA	Total Accruals	Richardson et al. (2006)
Intangibles		
ChGMChS	△ Gross Marging - △ Sales	Abarbanell and Bushee (1998)
SmI	△ Sales - △ Inventory	Abarbanell and Bushee (1998)
AL	Asset Liquidity	Ortiz-Molina and Phillips (2014)
EPr	Earnings Predictability	Francis et al. (2004)
ES	Earnings Smoothness	Francis et al. (2004)
HI	Herfindahl Index	Hou and Robinson (2006)
HR	Hiring rate	Belo, Lin, and Bazdresch (2014)
ICBE	Industry Concentration Book Equity	Hou and Robinson (2006)
IARER	Industry-adjusted Real Estate Ratio	Tuzel (2010)
OC	Org. Capital	Eisfeldt and Papanikolaou (2013)
RDM	RD / Market Equity	Chan et al. (2001)
TAN	Tangibility	Hahn and Lee (2009)
URDI	Unexpected RD Increases	Eberhart, Maxwell, and Siddique (2004)
WWI	Whited-Wu Index	Whited and Wu (2006)
Investment		
CAPEX	△ CAPEX - △ Industry CAPEX	Abarbanell and Bushee (1998)
AGr	Asset Growth	Cooper, Gulen, and Schill (2008)
ChNOA	Change Net Operating Assets	Hirshleifer, Hou, Teoh, and Zhang (2004)
ChPPEIA	Changes in PPE and Inventory-to-Assets	Lyandres, Sun, and Zhang (2007)
CDI	Composite Debt Issuance	Lyandres et al. (2007)
CEI5Y	Composite Equity Issuance (5-Year)	Daniel and Titman (2006)
DI	Debt Issuance	Spiess and Affleck-Graves (1995)
GrLTNOA	Growth in LTNOA	Fairfield, Whisenant, and Yohn (2003)
INV	Investment	Titman, Wei, and Xie (2004)
NDF	Net Debt Finance	Bradshaw, Richardson, and Sloan (2006)
NEF	Net Equity Finance	Bradshaw et al. (2006)
NOA	Net Operating Assets	Hirshleifer et al. (2004)
NOACh	Noncurrent Operating Assets Changes	Soliman (2008)
SR	Share Repurchases	Ikenberry, Lakonishok, and Vermaelen (1995)
TXFIN	Total XFIN	Bradshaw et al. (2006)
Profitability		
AT	Asset Turnover	Soliman (2008)
CT	Capital Turnover	Haugen and Baker (1996)
CBOP	Cash-based Operating Profitability	Ball et al. (2016)
ChiAT	Change in Asset Turnover	Soliman (2008)
EP	Earnings / Price	Basu (1977)
EC	Earnings Consistency	Alwathainani (2009)
FSc	F-Score	Piotroski (2000)
GP	Gross Profitability	Novy-Marx (2013)
Lvrg	Leverage	Bhandari (1988)
OSc	O-Score (More Financial Distress)	Dichev (1998)
OPtA	Operating Profits to Assets	Ball et al. (2016)
OPtE	Operating Profits to Equity	Fama and French (2015)
Value		
AM	Assets-to-Market	Fama and French (1992)
BM	Book Equity / Market Equity	Fama and French (1992)
CM	Cash Flow / Market Equity	Lakonishok, Shleifer, and Vishny (1994)
DurE	Duration of Equity	Dechow, Sloan, and Soliman (2004)
ECoBP	Enterprise Component of Book/Price	Penman, Richardson, and Tuna (2007)
EM	Enterprise Multiple	Loughran and Wellman (2011)
IR	Intangible Return	Daniel and Titman (2006)
LCoBP	Leverage Component of Book/Price	Penman et al. (2007)
NPY	Net Payout Yield	Boudoukh, Michaely, Richardson, and Roberts (2007)
OL	Operating Leverage	Novy-Marx (2010)
PY	Payout Yield	Boudoukh et al. (2007)
SaGr	Sales Growth	Lakonishok et al. (1994)
SP	Sales/Price	Barbee Jr, Mukherji, and Raines (1996)
SuGr	Sustainable Growth	Lockwood and Prombutr (2010)

Electronic copy available at: https://ssrn.com/abstract=3150654

Online Appendix for Does the Source of Fundamental Data Matter?

July 2018

Electronic copy available at: https://ssrn.com/abstract=3150654

A. Construction of the anomalies

Anomalies are grouped into 5 categories: accruals, profitability, value, investment, and intangibles. Construction of individual anomalies follows Harvey et al. (2016), McLean and Pontiff (2016) and Hou et al. (2017), with the exception of selecting a subset of exchanges and frequency of rebalancing. When these exceptions apply, they are described in the individual anomalies' definitions.

Accruals

Accruals (Acc)

Based on Sloan (1996), accruals are defined as

$$Acc = \frac{(\Delta act_t - \Delta che_t) - (\Delta lct_t - \Delta dlc_t - \Delta tp_t) - dp_t}{(at_t + at_{t-1})/2}$$

where Δact_t is change in current assets, Δche_t is change in cash and cash equivalents, Δlct_t is annual change in current liabilities, Δdlc_t is annual change in debt included in current liabilities, Δtp_t is annual change in income taxes payable and dp is depreciation and amortization expense.

Change in Current Operating Assets (ChCOA)

Based on Richardson et al. (2006), change in current operating assets is defined as

$$ChCOA = \frac{COA_t - COA_{t-1}}{at_{t-1}}$$

where COA_t are current operating assets, $COA_t = act_t - che_t$ in which act_t are current assets, che_t are cash and short-term investment and at_{t-1} are one-year lagged total assets

Change in Current Operating Liabilities (ChCOL)

Based on Richardson et al. (2006), change in current operating liabilities is defined as

$$ChCOL = \frac{COL_t - COL_{t-1}}{at_{t-1}}$$

where COL_t are current operating liabilities, $COL_t = lct_t - dlc_t$ in which lct_t are current liabilities, dlc_t is debt in current liabilities and at_{t-1} are one-year lagged total assets.

Electronic copy available at: https://ssrn.com/abstract=3150654

Change in Net Non-Cash Working Capital (ChNNCWC)

Based on Richardson et al. (2006), Change in Net Non-Cash Working Capital is defined as

$$ChNNCWC = \frac{WC_t - WC_{t-1}}{at_{t-1}}$$

where WC_t is working capital, $WC_t = COA_t - COL_t$ in which COA_t are current operating assets defined above in Change in Current Operating Assets anomaly and COL_t are current operating liabilities defined above in Change in Current Operating Liabilities anomaly.

Change in Net Non-Current Operating Assets (ChNNCOA)

Based on Richardson et al. (2006), Change in Net Non-Current Operating Assets is defined as

$$ChNNCOA = \frac{NCOA_t - NCOA_{t-1}}{at_{t-1}}$$

where NCO_t are non-current operating asset, $NCOA_t = NCA_t - NCL_t$ in which NCA_t are non-current assets defined in Change in Non-Current Operating Assets anomaly and NCL_t are non-current operating liabilities defined in Change in Non-Current Operating Liabilities anomaly.

Change in Non-Current Operating Assets (ChNCOA)

Based on Richardson et al. (2006), Change in Non-Current Operating Assets is defined as

$$ChNCOA = \frac{NCA_t - NCA_{t-1}}{at_{t-1}}$$

where NCA_t are non-current assets defined as $NCA_t = at_t - act_t - ivao_t$ where at_t are total assets, act_t are current assets, $ivao_t$ is investment and advances (0 if missing).

Change in Non-Current Operating Liabilities (ChNCOL)

Based on Richardson et al. (2006), Change in Non-Current Operating Liabilities is defined as

$$ChNCOL = \frac{NCL_t - NCL_{t-1}}{at_{t-1}}$$

where $NCL_t = lt_t - lct_t - dltt_t$ in which lt_t are total liabilities, lct_t are current liabilities and $dltt_t$ is long-term debt (0 if missing).

Electronic copy available at: https://ssrn.com/abstract=3150654

Change in Net Financial Assets (ChNFA)

Based on Richardson et al. (2006), Change in Net Financial Assets is defined as

$$ChNFA = \frac{NFNA_t - NFNA_{t-1}}{at_{t-1}}$$

where

$$NFNA_t = FNA_t - FNL_t$$

are net financial assets. FNA_t are financial assets, $FNA_t = ivst_t + ivao_t$. Where $ivst_t$ are short-term investments, $ivao_t$ are long-term investments. FNL_t are financial liabilities, $FNL_t = dltt_t + dlc_t + pstk_t$. Where $dltt_t$ is long-term debt, dlc_t is debt in current liabilities, and $pstk_t$ is preferred stock.

Change in Long-Term Investments (ChLTI)

Based on Richardson et al. (2006), Change in Long-Term Investments is defined as

$$ChLTI = \frac{ivao_t - ivao_{t-1}}{at_{t-1}}$$

where $ivao_t$ are long-term investments and at_{t-1} are one-year lagged total assets.

Change in Common Equity (ChCE)

Based on Richardson et al. (2006), Change in Common Equity is defined as

$$ChCE = \frac{ceq_t - ceq_{t-1}}{at_{t-1}}$$

where ceq_t is common equity and at_{t-1} are one-year lagged total assets.

Change in Financial Liabilities (ChFL)

Based on Richardson et al. (2006), Change in Financial Liabilities is defined as

$$ChFL = \frac{FNL_t - FNL_{t-1}}{at_{t-1}}$$

where FNL_t are net financial liabilities defined in anomaly Change in Net Financial Assets and at_{t-1} are one-year lagged total assets.

Electronic copy available at: https://ssrn.com/abstract=3150654

Growth in Inventory (GriI)

Based on Thomas and Zhang (2002), Growth in Inventor is defined as

$$GriI = \frac{invt_t - invt_{t-1}}{(at_t + at_{t-1})/2}$$

where $invt_t$ are inventories and at_t are total assets.

Inventory Change (ICh)

Based on Thomas and Zhang (2002), inventory change is defined as

$$ICh = \frac{invt_t - invt_{t-1}}{at_{t-1}}$$

where $invt_t$ are inventories and at_{t-1} are one-year lagged total assets. Only firms with positive inventories in this or previous year are included.

Inventory Growth (IGr)

Based on Belo and Lin (2011), inventory growth is defined as

$$IGr = \frac{invt_t - invt_{t-1}}{invt_{t-1}}$$

where $invt_t$ are inventories.

M/B and Accruals (MBaAC)

Based on Bartov and Kim (2004), M/B and Accruals is defined as

$$MBaAC = \begin{cases} 1 & \text{if stock is in low book-to-market } (BM_t) \text{ and high accrual } (Accr_t) \text{ quintiles} \\ -1 & \text{if stock is in high book-to-market } (BM_t) \text{ and low accrual } (Accr_t) \text{ quintiles} \\ 0 & \text{otherwise} \end{cases}$$

Accruals (Acc_t) are defined above, and book-to-market (BM_t) - book equity divided by market equity - is defined in category *Value*.

Net Working Capital Changes (NWCCh)

Based on Soliman (2008), net working capital changes are defined as

$$NWCCh = \frac{NWC_t - NWC_{t-1}}{at_{t-1}}$$

Electronic copy available at: https://ssrn.com/abstract=3150654

$NWC_t = (act_t - che_t) - (lct_t - dlc_t)$ is net working capital, where act_t are current assets, che_t is cash and cash equivalents, clt_t are current liabilities and dlc_t is debt in current liabilities.

Percent Operating Accruals (POA)

Based on Hafzalla et al. (2011), percent operating accruals are defined as

$$POA = \frac{ni_t - oancf_t}{|ni_t|}$$

where ni_t is net income and $oancf_t$ is cash flow from operations.

Percent Total Accruals (PTA)

Based on Hafzalla et al. (2011), percent total accruals are defined as

$$PTA = \frac{ni_t - (-sstk_t + prstkc_t + dv_t + oancf_t + ivncf_t + fincf_t)}{|ni_t|}$$

where ni_t is net income, $sstk_t$ sale of common and preferred stock, $prstkc_t$ is purchase of common and preferred stock, dv_t is total dividends, $oancf_t$ is cash flow from financing, $ivncf_t$ is cash flow from investment and $fincf_t$ is cash from from financing.

Total Accruals (TA)

Based on Richardson et al. (2006), total accruals are defined as

$$TA = \frac{TACCR_t - TACCR_{t-1}}{at_{t-1}}$$

where $TACCR_t = NCO_t + WC_t + NFNA_t$ NCO_t are net non-current operating assets defined in anomaly Change in Net Non-Current Operating Assets, WC_t is working capital defined in anomaly Change in Net Non-Cash Working Capital and $NFNA_t$ are net financial assets defined in anomaly Change in Net Financial Assets.

Intangibles

Asset Liquidity (AL)

Based on Ortiz-Molina and Phillips (2014), asset liquidity is defined as

$$AL = \frac{che_t + 0.75(act_t - che_t) + 0.5(at_t - act_t - gdwl_t - intan_t)}{at_{t-1}}$$

Electronic copy available at: https://ssrn.com/abstract=3150654

where at_{t-1} are one-year lagged total assets, act_t are current assets, che_t is cash and short-term investments, $gdwl_t$ is goodwill (0 if missing) and $intan_t$ are intangibles (0 if missing).

ΔGross Margin - ΔSales (ChGMChS)

Based on Abarbanell and Bushee (1998), ΔGross Margin - ΔSales is defined as

$$ChSChAR = \frac{GM_t - \frac{GMs_{t-1}+GM_{t-2}}{2}}{\frac{GM_{t-1}+GM_{t-2}}{2}} - \frac{sale_t - \frac{sale_{t-1}+sale_{t-2}}{2}}{\frac{sale_{t-1}+sale_{t-2}}{2}}$$

where $sale_t$ is net sales and GM_t is gross margin, defined as $GM_t = sale_t - cogs_t$, where $cogs_t$ is cost of goods sold.
Only firms with positive two-year sales and two-year gross margin averages are included.

Earnings Predictability (EPr)

Based on Francis et al. (2004), Earnings Predictability is defined as volatility of residuals from the first-order autoregressive model using the ten-year rolling window for split-adjusted earnings per share. Split-adjusted earnings per share are defined as $EPS_t = \frac{epspx_t}{ajex_t}$.
Only firms with no missing required data over the ten-year rolling window are included.

Earning Smoothness (ES)

Based on Ortiz-Molina and Phillips (2014), earnings smoothness is defined as

$$ES = \frac{std(ELA_t)}{std(CFOA_t)}$$

where the standard deviation is calculated over the ten-year rolling window and only firms with no missing required data over the ten-year history are included. Further

$$ELA_t = \frac{ib_t}{at_{t-1}}$$

and

$$CFOA_t = ib_t - (DCA_t - DCL_t - DCHE_t + DSTD_t - dp_t))$$

where ib_t are earnings and at_{t-1} is lagged total assets. DCA_t is one-year change in current assets, DCL_t is the one-year change in current liabilities, $DCHE_t$ is the one-year change in cash and short-term investments, $DSTD_t$ is the one-year change in debt in current liabilities, and dp_t is depreciation and amortization.

Electronic copy available at: https://ssrn.com/abstract=3150654

Herfindahl Index (HI)

Based on Hou and Robinson (2006), Herfindahl index as a measure of industry concentration defined as

$$HI = \frac{H_t + H_{t-1} + H_{t-2}}{3}$$

$H_t = \sum_{i=1}^{N_j} sale_{i,j}$, where $sale_{ij}$ is the sale of firm i in industry j and N_j is the total number of firms in the 3-digit SIC code defined industry.

Hiring rate (HR)

Based on Belo et al. (2014), hiring rate is defined as

$$HR = \frac{emp_{t-1} - empt - 2}{0.5emp_{t-1} + 0.5emp_{t-2}}$$

where emp_t is the number of employees. Stocks with $HR = 0$, often a consequence of a stale information, are exluded.

Industry-adjusted Real Estate Ratio (IARER)

Based on Tuzel (2010),industry-adjusted real estate ratio is defined as

$$IARER = RER_t - \frac{\sum_{j=1}^{N_j} RER_{ij}}{N_j}$$

i.e. the real estate ratio minus its, 2-digit SIC code defined, industry average. Real estate ratio is defined as

$$RER_t = (fatb_t + fatl_t)/ppent_t$$

where $fatb_t$ is the sum of buildings at cost, $fatl_t$ is leases at cost and $ppegt_t$ is gross property, plant, and equipment.
Industries with less than five firms are excluded.

Industry Concentration Assets (ICA)

Based on Hou and Robinson (2006), Industry Concentration Assets is Herfindahl index (HI), defined above, with total assets at_t as a measure of market share instead of sales $sale_t$.

Industry Concentration Book Equity (ICBE)

Based on Hou and Robinson (2006), Industry Concentration Book Equity is Herfindahl index (HI), defined above, with book equity BE_t defined in anomaly Book Equity / Market Equity.

Electronic copy available at: https://ssrn.com/abstract=3150654

Org. Capital (OC)

Based on Eisfeldt and Papanikolaou (2013), organizational capital is defined recursively. For the first year of stocks appearance in data, organizational capital is set equal to 4 times selling, general and administrative expense (0 if missing), i.e.

$$OC_{t_0} = 4 * xsga_{t_0}$$

All next years, organizational capital is defined as

$$OC_t = \frac{\frac{0.85*OC_{t-1}+xsga_t}{cpi_t}}{at_t}$$

where cpi_t is and at_t are total assets.

R&D / Market Value of Equity (RDM)

Based on Chan et al. (2001), R&D-to-market value of equity is defined as

$$RDM = \frac{xrd_t}{ME_t}$$

where xrd is research and development expense and $ME_t = prc_t * shrout_t$ is the market equity defined as price times shares outstanding, at the end of the previous year.

Tangibility (TAN)

Based on Hahn and Lee (2009), tangibility is defined as

$$TAN = \frac{che_t + 0.715rect_t + 0.547invt_t + 0.535ppegt_t}{at_t}$$

where che_t are cash holdings, $rect_t$ are accounts receivable, $invt_t$ is inventory and $ppegt_t$ is property, plant and equipment.

Unexpected R&D Increases (URDI)

Based on Eberhart et al. (2004), unexpected R&D increases is a binary variable defined as

$$URDI = \begin{cases} 1 & \text{if } (\frac{xrd_t}{revt_t} > 0.05) \ \& \ (\frac{xrd_t}{at_t} > 0.05) \ \& \ (\frac{xrd_t}{xrd_{t-1}} > 1.05) \ \& \ (\frac{\frac{xrd_t}{at_t}}{\frac{xrd_{t-1}}{at_{t-1}}} > 1.05) \\ 0 & \text{otherwise} \end{cases}$$

Electronic copy available at: https://ssrn.com/abstract=3150654

where xrd_t are R&D expenditures, $revt_t$ is total revenue and at_t is total assets. $URDI = 1$ if R&D scaled by assets and revenue is greater than 5%, the yearly percentage change in R&D expenditures is greater than 5%; and R&D scaled by assets increased by more than 5%.

Whited-Wu Index (WWI)

Based on Whited and Wu (2006), Whited-Wu index is defined as

$$WWI_{it} = -0.091CF_t - 0.062DIVP_t + 0.021LDA_t - 0.044log(at_t) + 0.102ISG_t - 0.035(SG_t)$$

where

$$CF_T = \sqrt[4]{1 + \frac{\frac{ib_t + dp_t}{at_t}}{4}} - 1$$

where ib_t is income before extraordinary items, dp_t is depreciation and amortization, at_t are total assets, $DIVP_t$ is a binary variable equal to one if firm pays cash dividends ($dvpsx_t > 0$) and 0 otherwise, and $LDA_t = \frac{dltt_t}{at_t}$ is the long-term debt to total assets.

$$ISG_t = \frac{(\sum_{i=1}^{N_j} sale_{i,j})_t}{(\sum_{i=1}^{N_j} sale_{i,j})_t}$$

where $sale_{ij}$ is the sale of firm i in industry j and N_j is the total number of firms in the 3-digit SIC code defined industry including at least 3 firms.

$$SG_t = \sqrt[4]{1 + \frac{\frac{sale_t}{sale_{t-1}}}{4}} - 1$$

Investment

Asset Growth (AGr)

Based on Cooper et al. (2008), asset growth is defined as

$$AGr = \frac{at_t}{at_{t-1}}$$

where at_t are total assets.

Electronic copy available at: https://ssrn.com/abstract=3150654

Change in Net Operating Assets (ChNOA)

Based on Hirshleifer et al. (2004), Change in Net Operating Assets is defined as

$$ChNOA = \frac{NOA_t - NOA_{t-1}}{at_{t-1}}$$

where NOA_t are net operating assets defined below and at_{t-1} are lagged total assets.

Changes in PPE and Inventory-to-Assets (ChPPEIA)

Based on Lyandres et al. (2007), Changes in PPE and Inventory-to-Assets is defined as

$$ChPPEIA_t = \frac{(ppegt_t - ppegt_{t-1}) + (invt_t - invt_{t-1})}{at_{t-1}}$$

wehere $ppegt_t$ is gross property, plant and equipment, $invt_t$ is total inventories and at_{t-1} are lagged total assets.

Composite Debt Issuance (CDI)

Based on Lyandres et al. (2007), Composite Debt Issuance is defined as

$$CDI = log(\frac{dltt_t + dlc_t}{dltt_{t-5} + dlc_{t-5}})$$

where $dltt_t$ is total long-term debt and dlc_t is debt in current liabilities.

∆CAPEX - ∆Industry CAPEX (CAPEX)

Based on Abarbanell and Bushee (1998), change in investment minus the change in industry investment (∆CAPEX - ∆Industry CAPEX). Where

$$\Delta CAPEX = \frac{capx_{it} - \frac{capx_{i,t-1}+capx_{i,t-2}}{2}}{\frac{capx_{i,t-1}+capx_{i,t-2}}{2}}$$

and ∆Industry CAPEX is defined analogously for aggregated industry CAPEX. $capx_t$ is capital expenditure.
Stocks in industries with less than 3 firms are excluded.

Debt Issuance (DI)

Based on Spiess and Affleck-Graves (1995), debt issuance is defined as

$$DI = \begin{cases} 1 & \text{if } dltis_t > 0 \\ 0 & \text{otherwise} \end{cases}$$

Electronic copy available at: https://ssrn.com/abstract=3150654

where $dltis_t$ is long-term debt/issuance.

Growth in LTNOA (GriLTNOA)

Based on Fairfield et al. (2003), growth in long-term net operating assets is defined as

$$GriLTNOA = NOA_t - NOA_{t-1} - ACCR_t$$

, where NOA_t are net operating assets, defined below and $ACCR_t$ are accruals defined above in category *Accruals*.

Investment (INV)

Based on Titman et al. (2004), investment is defined as

$$INV = \frac{capx_t/revt_t}{avg_{3t}(\frac{capx}{revt})}$$

where $capx_t$ is capital expenditures, $revt_x$ is total revenue and $avg_{3t}()$ is average from the previous three years.
Stocks with revenue < \$10m are excluded.

Net Debt Finance (NDF)

Based on Bradshaw et al. (2006), Net Debt Finance is defined as

$$NDF_t = \frac{dltis_t - dltr_t + dlcch_t}{(at_t + at_{t-1})/2}$$

where $dltis_t$ is long-term debt issuance, $dltr_t$ is long-term debt reduction , $dlcch_t$ are current debt changes and at_t are total assets.

Net Equity Finance (NEF)

Based on Bradshaw et al. (2006), Net Equity Finance is defined as

$$NEF_t = \frac{sstk_t - prstkc_t - dv_t}{(at_t + at_{t-1})/2}$$

where $sstk_t$ is sale of common and preferred stock (0 if missing), $prstkc_t$ is purchase of common and preferred stock (0 if missing) , dv_t are cash dividend, and at_t are total assets.

Electronic copy available at: https://ssrn.com/abstract=3150654

Net Operating Asset (NOA)

Based on Hirshleifer et al. (2004), net operating assets are defined as

$$NOA = \frac{OA_t - OL_t}{at_{t-1}}$$

OA_t and OL_t are operating assets and operating liabilities defined as $OA_t = at_t - che_t$ and $OL_t = at_t - dlc_t - dltt_t - mib_t - pstkrv_t - ceq_t$, where at_t is total assets, che_t is cash and short-term investment, dlc_t is current portion of long-term debt, $dltt$ is long-term debt, mib_t is minority interest, $pstkrv$ is preferred stock and ceq is common equity.

Noncurrent Operating Assets Changes (NOACh)

Based on Soliman (2008), noncurrent operating assets changes are defined as

$$NOACh = \frac{NCOA_t - NCOA_{t-1}}{at_t}$$

where $NCOA_t$ is noncurrent operating assets. Noncurrent operating assets are defined as

$$NCOA_t = (at_t - act_t - ivaeq_t) - (lt_t - lct_t - dltt_t)$$

, where at_t are total asssets, act_t are current assets, $ivaeq_t$ are investment and advances (0 if missing), lt_t are total liabilities, lct_t are current liabilities and $dltt_t$ is long-term debt.

Share Repurchases (SR)

Based on Ikenberry et al. (1995), share repurchases are defined as binary variable

$$SR = \begin{cases} 1 & \text{if } prstkc_t > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $prstkc_t$ is purchase of common and preferred stock.

Total XFIN (TXFIN)

Based on Bradshaw et al. (2006), total net external financing is defined as

$$TXFIN = \frac{sstk_t - dv_t - prstkc_t + dltis_t - dltr_t}{at_t}$$

where at_t are total assets, $sstk_t$ is sale of common and preferred stock (0 if missing), dv_t are cash dividends, $prstkc_t$ is purchase of common and preferred stock (0 if missing), $dltis_t$ is sale of long-term debt and $dltr_t$ is purchase of long-term debt.

Electronic copy available at: https://ssrn.com/abstract=3150654

Profitability

Asset Turnover (AT)

Based on Soliman (2008), asset turnover is defined as

$$AT = \frac{sale_t}{avg_{2t}(NOA)}$$

where NOA are net operating assets defined as $NOA = (at_t - che_t) - (lt_t - dltt_t - dlc_t - mib_t)$ and $avg_{2t}(NOA)$ is average NOA from the previous two years. at_t are total assets, che_t is cash and cash equivalents, lt_t are total liabilities, $dltt_t$ is long-term debt, dlc_t is debt in current liabilities, and mib_t is minority interest (0 if missing). Firms with negative NOA and negative operating income ($oiadp$) are exluded.

Capital Turnover (CT)

Based on (Haugen and Baker, 1996), capital turnover is defined as

$$CT = \frac{sale_t}{at_{t-1}}$$

where $sale_t$ is sales and at_{t-1} are one-year lagged total assets.

Cash-based Operating Profitability (CBOP)

Based on Ball et al. (2016), cash-based operating profitability is defined as

$$\begin{gathered} CBOP = \big(revt_t - cogs_t - xsga_t + xrd_t - (rect_t - rect_{t-1}) - (invt_t - invt_{t-1}) - \\ -(xpp_t - xpp_{t-1}) + (drc_t + drlt_t - drc_t - drlt_t) + (rect_t - rect_{t-1}) + (ap_t - ap_{t-1}) \\ +(xacc_t - xacc_{t-1})\big)/at_t \end{gathered}$$

where at_t are total assets, $revt_t$ is total revenue, $cogs_t$ is cost of goods sold, $xsga_t$ are selling, general, and administrative expenses, xrd_t are research and development expenditures (0 if missing), $rect_t$ are accounts receivables, $invt_t$ is inventory, xpp_t are prepaid expenses, drc_t is current deferred revenue, $drlt_t$ is long-term deferred revenue, ap_t are accounts payable and $xacc_t$ are accrued expenses. Changes (in brackets) are all equal to 0 if missing.

Change in Asset Turnover (ChiAT)

Based on Soliman (2008), change in asset turnover is defined as

$$ChiAT = AT_t - AT_{t-1}$$

Electronic copy available at: https://ssrn.com/abstract=3150654

where AT_t is asset turnover defined above.

Earnings Consistency (EC)

Based on Alwathainani (2009), earnings consistency is defined as

$$EC = \sqrt[5]{\Pi_{i=1}^{5}(1 + eg_i)} - 1$$

where eg_i is earnings growth is defined as

$$eg_t = \frac{epspx_t - epspx_{t-1}}{\frac{|epsx_t| + |epsx_{t-1}|}{2}}$$

where $epspx_t$ are earnings per share excluding extraordinary items. Stocks with $|eg_t| > 6$ are deleted. Also stocks with the previous two earnings growths with opposite signs are excluded $(eg_t * eg_{t-1})$

Earnings / Price (EP)

Based on (Basu, 1977), earnings-to-price is defined as

$$EP = \frac{ib_t}{ME_t}$$

where ib_t is income before extraordinary items and $ME_t = prc_t * shrout_t$ is market equity, i.e. price times shares outstanding.
Firms with $ib_t \leq 0$ are excluded.

F-Score (FSc)

Based on Piotroski (2000), F-score is defined as the sum of nine binary variables (F1-F9) and is further limited only to firms in the highest quintile with respect to book-to-market

$$F = \sum_{i=1}^{9} F_i$$

Electronic copy available at: https://ssrn.com/abstract=3150654

Binary variables are defined as

$$
\begin{aligned}
F1 = &\ \ 1 \text{ if } ni_t > 0;\ \ 0 \text{ otherwise} \\
F2 = &\ \ 1 \text{ if } oancf_t > 0\ ; 0 \text{ otherwise} \\
F3 = &\ \ 1 \text{ if } \tfrac{ni_t}{at_t} > \tfrac{ni_{t-1}}{at_{t-1}}\ ; 0 \text{ otherwise} \\
F4 = &\ \ 1 \text{ if } oancf_t > ni_t\ ; 0 \text{ otherwise} \\
F5 = &\ \ 1 \text{ if } \tfrac{dltt_t}{at_t} < \tfrac{dltt_{t-1}}{at_{t-1}}\ ; 0 \text{ otherwise} \\
F6 = &\ \ 1 \text{ if } \tfrac{act_t}{lct_t} > \tfrac{act_{t-1}}{lct_{t-1}}\ ; 0 \text{ otherwise} \\
F7 = &\ \ 1 \text{ if } sstk_t - (pstk_t - pstk_{t-1}) \leq 0\ ; 0 \text{ otherwise} \\
F8 = &\ \ 1 \text{ if } \tfrac{oiadp_t}{sale_t} > \tfrac{oiadp_{t-1}}{sale_{t-1}}\ ; 0 \text{ otherwise} \\
F9 = &\ \ 1 \text{ if } \tfrac{sale_t}{at_t} > \tfrac{sale_{t-1}}{at_{t-1}}\ ; 0 \text{ otherwise}
\end{aligned}
$$

where ni_t is net income, $oancf_t$ is cash-flow from operating activities, at_t are total assets, $dltt_t$ is long term debt, act_t is current assets, lct_t are current liabilities, $sstk_t$ is sale of common and preferred stock, $pstk_t$ is total preferred stock, $oiadp_t$ is operating income after depreciation, and $sale_t$ is net sales.

Gross Profitability (GP)

Based on Novy-Marx (2013), gross profitability is defined as

$$GP = \frac{revt_t - cogs_t}{at_{t-1}}$$

where $revt_t$ is total revenue, $cogs_t$ is cost of goods sold, and at_{t-1} are total assets lagged by one year.

Operating Profits to Assets (OPtA)

Based on Ball et al. (2016), operating profits to assets are defined as

$$OPtA = \frac{revt_t - cogs_t - xsga_t + xrd_t}{at_t}$$

where $revt_t$ is total revenue, $cogs_t$ is cost of goods sold, $xsga_t$ is SG&A, xrd_t are research and development expenditures, and at_t are total assets.

Operating Profits to Assets (OPtE)

Based on Fama and French (2015), operating profits to equity are defined as

$$OPtE = \frac{revt_t - cogs_t - xsga_t + xint_t}{be_t}$$

Electronic copy available at: https://ssrn.com/abstract=3150654

where $revt_t$ is total revenue, $cogs_t$ is cost of goods sold, $xsga_t$ is SG&A, $xint_t$ is interest and related expense (total), and be_t is book equity defined in Book Equity / Market Equity variable. At least one from xint, cogs, xsga cannot be missing and the missing values are filled with zeros.

Leverage (Lvrg)

Based on Bhandari (1988), leverage is defined as

$$Lvrg = \frac{dltt_t + dlc_t}{ME_t}$$

where $dltt_t$ is long-term debt, dlc_t is debt in current liabilities and $ME_t = prc_t * shrout_t$ is market equity defined in anomaly of earnings/price.

O-Score (OSc)

Based on Dichev (1998), O-score is defined as

$$OSc = -1.32 - 0.4078log(\frac{at_t}{cpi_t}) + 6.03 * (\frac{dltt_t + dlc_t}{at_t}) - 1.43 * (\frac{act_t - lct_t}{at_t}) + 0.076 * (\frac{lct_t}{act_t}) -$$
$$-1.72 * (OENEG_t) - 2.37 * (\frac{ni_t}{at_t}) - 1.83 * (\frac{pi_t}{dp_t}) + 0.285 * (INTWO_t) - 0.521 * (\frac{ni_t - ni_{t-1}}{|ni_t| + |ni_{t-1}|})$$

where at_t are total assets, cpi_t is inflation, $dltt_t$ are long-term liabilities, dlc_t are short-term liabilities, act_t are current assets, lct_t are current liabilities, $OENEG_t$ is binary variable equal to one if $lt_t > at_t$ and 0 otherwise, ni_t is net income, $INTWO_t$ is binary variable equal to one if stock has negative net income in both previous years and 0 otherwise. Only stocks with SIC codes from 1 to 3999 and from 5000 to 5999 are included.

Value

Assets-to-Market (AM)

Based on Fama and French (1992), assets-to-market is defined as

$$AM = \frac{at_t}{ME_t}$$

where at_t are assets total and ME_t is market equity.

Electronic copy available at: https://ssrn.com/abstract=3150654

Book Equity / Market Equity (BM)

Based on Fama and French (1992), book-to-market equity is defined as

$$BM = log(\frac{BE_t}{ME_t})$$

Market equity is price times shares outstanding, $ME_t = prc_t * shrout_t$. Book equity is defined conditional on missing items as

$$BE_t = seq_t - PS_t$$

where seq_t is total stockholders' equity, if missing then $seq_t = ceq_t + pstk_t$, or $seq_t = at_t - lt_t$, where ceq_t is tangible common equity, $pstk_t$ is preferred stock using liquidating value, at_t are total assets, lt_t are total liabilities,and PS_t is preferred stock measured using (ordered on availability) redemption, liquidating or par value, i.e. $pstkrv_t, pstkl_t, pstk_t$.

Cash Flow / Market Value of Equity (CM)

Based on Lakonishok et al. (1994), cash flow to market value of equity is defined as

$$CM = \frac{ib_t + dp_t}{ME_t}$$

where ib_t is net income, dp_t is depreciation and amortization and ME_t is market equity defined above in book-to-market equity anomaly.

Duration of Equity (DurE)

Based on Dechow et al. (2004), duration of equity is defined as

$$DurE_t = \frac{58}{3} + \frac{1}{MC_t} \sum_{j=1}^{10} \frac{cd_j j(j - 58/3)}{1.12}$$

where cd_j is defined recursively from the following equations: $g_{j+1} = 0.06 + 0.24g_j$, $be_j = be_0(1 + g_j)$, $roe_{j+1} = 0.12 + 0.57roe_j$, and $cd_j = roe_j be_{j-1}$. The starting values are $be_0 = ceq_t$, $roe_0 = \frac{ib_t}{ceq_{t-1}}$, and $g_0 = \frac{sale_t}{sale_{t-1}} - 1$. be_t is the book equity, ceq_{t-1} is a lag of common equity, ib_t are earnings, and $sale_t$ are net sales.

Enterprise Component of Book/Price (ECoBP)

Based on Penman et al. (2007), enterprise component of book/price is defined as

$$ECoBP = \frac{BE_t + ND_t}{ND_t + ME_t}$$

Electronic copy available at: https://ssrn.com/abstract=3150654

where BE_t and ME_t are book value of equity and market equiy, defined above in book-to-market equity anomaly. $ND_t = dltt_t + dlc_t + pstk_t + dvpa_t - tstkp_t - che_t$ is net debt, where che_t is cash and short-term investments, $dltt_t$ is long-term debt, dlc_t is debt in current liabilities, $pstk_t$ is preferred stock, $dvpa_t$ is preferred dividends in arrears and $tstkp_t$ is preferred treasury stock.

Enterprise Multiple (EM)

Based on Loughran and Wellman (2011), enterprise multiple is defined as

$$EM = \frac{EV_t}{oibdp_t}$$

where $oibdp_t$ is operating cash flow and EV_t is enterprise value defined as $EV_t = ME_t + dltt_t + dlc_t + pstk_t + dvpa_t - tstkp_t - che_t$. ME_t is market equity defined above in book-to-market equity anomaly, $dltt_t$ is long-term debt, dlc_t is debt in current liabilities, $pstk_t$ is preferred stock, $dvpa_t$ is preferred dividends in arrears, $tstkp_t$ is preferred treasury stock and che_t is cash and short-term investments.

Intangible Return (IR)

Based on Daniel and Titman (2006), intangible return is defined as residual from the following cross-sectional regression

$$log(r_{t-5,t}) = \beta_0 + \beta_1 BM_{t-5} + \beta_2 log(RB_{t-5,t}) + \epsilon_t$$

where $r_{t-5,t}$ is 5- year stock return, BM_{t-5} is 5-year-lagged book-to-market defined in anomaly Book Equity / Market Equity and $RB_{t-5,t} = log(\frac{BE_t}{BE_{t-5}) - \sum_{p=t-5}^{t-1}(r_p - log(\frac{P_p}{P_{p-1}}))})$ in which BE_t is the book equity defined in anomaly Book Equity / Market Equity , r_p is the stock return for year p and P_p is the price at the end of year p.

Leverage Component of Book/Price (LCoBP)

Based on Penman et al. (2007), leverage component of book/price is defined as

$$LCoBP = BE_t - ECoBP_t$$

where BE_t is book value of equity defined above in book-to-market equity anomaly, and $ECoBP_t$ is enterprise component of book/price defined above.

Electronic copy available at: https://ssrn.com/abstract=3150654

Net Payout Yield (NPY)

Based on Boudoukh et al. (2007), net payout yield is defined as

$$NPY = \frac{dvc_t + prstkc_t - sstk_t}{ME_t}$$

where dvc_t are dividends common/ordinary, $prstkc_t$ is purchase of common and preferred stock, $sstk_t$ is sale of common and preferred stock, and ME_t is market equity.

Operating Leverage (OL)

Based on Novy-Marx (2010), operating leverage is defined

$$OL = \frac{xsga_t + cogs_t}{at_t}$$

where $xsga_t$ is SG&A, $cogs_t$ is cost of goods sold, and at_t are total assets.

Payout Yield (PY)

Based on Boudoukh et al. (2007), payout yield is defined as

$$PY = \frac{dvc_t + prstkc_t - (pstkrv_t + pstkrv_{t-1})}{ME_t}$$

where dvc_t are dividends common/ordinary, $prstkc_t$ is purchase of common and preferred stock, $pstkrv_t$ is preferred stock/redemption, and ME_t is market equity.

Sales Growth (SaGr)

Based on Lakonishok et al. (1994), sales growth is defined as

$$SaGr = \frac{5SGR_t + 4SGR_{t-1} + 3SGR_{t-2} + 2SGR_{t-3} + 1SGR_{t-4}}{15}$$

where SGR_t is the rank of firm in year t based on the simple sales growth defined as $SG = sale_t/sale_{t-1}$.

Sustainable Growth (SuGr)

Based on Lockwood and Prombutr (2010), sustainable growth is defined as $SuGr = BE_t/BE_{t-1}$, where BE_t is book equity defined above in book-to-market equity anomaly.

Electronic copy available at: https://ssrn.com/abstract=3150654

Sales/Price (SP)

Based on Barbee Jr et al. (1996), sales-to-price is defined as $SP = revt_t/ME_t$, where $revt_t$ is total revenue and ME_t is the market equity defined above in the book-to-market equity anomaly.

Table IA1
Industries in the Datastream level 3 classification and corresponding four digit SIC

Datastream lvl 3 industry	SIC codes
Automobiles & Parts	3011, 3510, 3714, 3751, 5013
Basic Resources	800, 1000, 1040, 1090, 1220, 1221, 2421, 2600, 2611, 2621, 2631, 3310, 3312, 3317, 3330, 3334, 3350, 3360, 3444, 3460, 3720, 5050, 5051
Chemicals	2810, 2820, 2821, 2833, 2851, 2860, 2870, 2890, 2891, 2990, 3080, 3081, 3341, 5160
Construct. & Material	1400, 1540, 1600, 1623, 1731, 2400, 2430, 2950, 3211, 3231, 3241, 3250, 3270, 3272, 3281, 3290, 3430, 3440, 3442, 3448, 5031, 5070, 5072
Financial Services(3)	6111, 6141, 6153, 6159, 6162, 6163, 6172, 6189, 6200, 6211, 6221, 6282, 6361, 6500, 6510, 6770, 6795, 6798, 6799, 8880, 8888, 9995
Food & Beverage	100, 200, 900, 2000, 2011, 2013, 2015, 2020, 2024, 2030, 2033, 2040, 2050, 2052, 2060, 2070, 2080, 2082, 2086, 2090, 2092
Healthcare	2590, 2800, 2834, 2835, 2836, 3060, 3821, 3826, 3841, 3842, 3843, 3844, 3845, 3851, 4100, 5047, 6324, 8000, 8011, 8050, 8051, 8060, 8062, 8071, 8082, 8090, 8093, 8300, 8731
Ind. Goods & Services	1700, 2390, 2650, 2670, 2673, 2750, 2761, 3050, 3086, 3089, 3221, 3320, 3357, 3390, 3411, 3412, 3443, 3451, 3452, 3470, 3480, 3490, 3523, 3524, 3530, 3531, 3532, 3537, 3540, 3541, 3550, 3555, 3560, 3561, 3562, 3564, 3567, 3569, 3575, 3580, 3585, 3590, 3600, 3612, 3613, 3620, 3621, 3634, 3640, 3669, 3670, 3672, 3677, 3678, 3679, 3690, 3711, 3713, 3715, 3721, 3724, 3728, 3730, 3743, 3760, 3812, 3822, 3823, 3824, 3825, 3827, 3829, 3861, 3910, 4011, 4013, 4210, 4213, 4231, 4400, 4412, 4513, 4700, 4731, 4950, 4953, 4955, 4961, 5000, 5063, 5065, 5080, 5082, 5084, 5090, 5099, 6099, 6794, 7320, 7350, 7359, 7361, 7363, 7374, 7377, 7380, 7381, 7384, 7385, 7389, 7829, 8111, 8200, 8351, 8600, 8700, 8711, 8734, 8741, 8742, 8744, 9721
Insurance	6311, 6321, 6331, 6351, 6411
Media	2711, 2721, 2731, 2732, 2741, 2780, 4832, 4833, 4841, 7310, 7311, 7330, 7331, 7819, 7822, 8900
Oil & Gas	1311, 1381, 1382, 1389, 2911, 3533, 4522, 4610, 4900, 5171, 5172, 6792
Pers & Househld Goods	1531, 2100, 2111, 2200, 2211, 2221, 2250, 2253, 2273, 2300, 2320, 2330, 2340, 2451, 2452, 2510, 2511, 2520, 2522, 2531, 2540, 2771, 2840, 2842, 2844, 3021, 3100, 3220, 3260, 3420, 3433, 3630, 3651, 3716, 3790, 3873, 3911, 3931, 3942, 3944, 3949, 3950, 3960, 5020, 5030, 5064, 5130, 5150, 5190, 6552
Real Estate	6519, 6531
Retail	700, 2790, 3140, 4220, 5094, 5010, 5110, 5122, 5140, 5141, 5180, 5200, 5211, 5271, 5311, 5331, 5399, 5400, 5411, 5412, 5500, 5531, 5600, 5621, 5651, 5661, 5700, 5712, 5731, 5734, 5735, 5912, 5940, 5944, 5945, 5960, 5961, 5990, 6399, 7200, 7340, 7500, 7600, 7841
Technology	3559, 3570, 3571, 3572, 3576, 3577, 3578, 3579, 3661, 3663, 3674, 3695, 4899, 5040, 5045, 7370, 7371, 7372, 7373
Telecommunications	4812, 4813, 4822
Travel & Leisure	1520, 3652, 3990, 4512, 4581, 5810, 5812, 6512, 6513, 6532, 7000, 7011, 7510, 7812, 7830, 7900, 7948, 7990, 7997
Utilities	4911, 4922, 4923, 4924, 4931, 4932, 4941, 4991, 5900
Banks	6021, 6022, 6029, 6035, 6036, 6199

Electronic copy available at: https://ssrn.com/abstract=3150654

Table IA2
Impact of delisting in Compustat - detailed

	Our delisting vs no delisting				Our delisting vs Shumway (1997)			
	Corr	Our	No delisting	Diff	Corr	Our	Shumway	Diff
Accruals								
Acc	1.000	0.29 (2.26)	0.31 (2.42)	(6.33)	1.000	0.29 (2.26)	0.29 (2.24)	(-3.60)
ChCE	0.999	0.38 (2.17)	0.42 (2.41)	(5.45)	1.000	0.38 (2.17)	0.36 (2.08)	(-2.81)
ChCOA	0.998	0.53 (5.04)	0.55 (5.17)	(2.48)	0.999	0.53 (5.04)	0.53 (4.96)	(-1.77)
ChCOL	0.999	0.36 (4.00)	0.37 (4.07)	(2.51)	1.000	0.36 (4.00)	0.36 (3.98)	(-1.60)
ChFL	0.999	0.56 (7.88)	0.55 (7.83)	(-2.21)	1.000	0.56 (7.88)	0.56 (7.91)	(2.59)
ChLTI	0.992	0.13 (2.91)	0.14 (3.03)	(1.06)	0.995	0.13 (2.91)	0.12 (2.73)	(-1.39)
ChNCOA	1.000	0.68 (5.46)	0.70 (5.60)	(4.20)	1.000	0.68 (5.46)	0.68 (5.45)	(-1.95)
ChNCOL	0.997	0.17 (2.01)	0.18 (2.15)	(2.04)	0.998	0.17 (2.01)	0.17 (1.91)	(-1.48)
ChNFA	0.997	0.42 (6.10)	0.41 (5.99)	(-2.71)	1.000	0.42 (6.10)	0.42 (6.13)	(2.82)
ChNNCOA	0.999	0.71 (5.97)	0.72 (6.04)	(1.99)	0.999	0.71 (5.97)	0.71 (6.00)	(0.57)
ChNNCWC	0.997	0.35 (4.19)	0.36 (4.33)	(2.00)	0.998	0.35 (4.19)	0.34 (4.09)	(-1.49)
GriI	0.998	0.48 (5.34)	0.50 (5.49)	(2.23)	0.998	0.48 (5.34)	0.48 (5.27)	(-1.45)
ICh	0.998	0.50 (5.26)	0.52 (5.43)	(2.56)	0.999	0.50 (5.26)	0.50 (5.18)	(-1.61)
IGr	0.998	0.55 (5.22)	0.56 (5.38)	(2.47)	0.999	0.55 (5.22)	0.54 (5.15)	(-1.50)
MBaAC	0.996	1.43 (7.02)	1.43 (6.97)	(-0.01)	1.000	1.43 (7.02)	1.42 (7.00)	(-1.62)
NWCCh	0.994	0.49 (6.76)	0.48 (6.77)	(-0.42)	0.997	0.49 (6.76)	0.48 (6.66)	(-0.84)
POA	0.999	0.70 (5.51)	0.69 (5.50)	(-1.58)	1.000	0.70 (5.51)	0.70 (5.52)	(1.73)
PTA	0.999	0.33 (3.20)	0.35 (3.37)	(4.26)	1.000	0.33 (3.20)	0.33 (3.19)	(-0.41)
TA	0.999	0.44 (3.56)	0.48 (3.84)	(5.15)	0.999	0.44 (3.56)	0.44 (3.46)	(-2.17)
Intangibles								
AL	0.999	0.44 (2.74)	0.45 (2.81)	(1.33)	0.999	0.44 (2.74)	0.44 (2.69)	(-1.22)
ChGMChS	0.999	0.24 (3.55)	0.22 (3.34)	(-5.01)	1.000	0.24 (3.55)	0.24 (3.56)	(2.13)
EPr	0.998	0.66 (4.60)	0.65 (4.42)	(-2.13)	1.000	0.66 (4.60)	0.67 (4.61)	(2.29)
ES	1.000	0.21 (1.00)	0.23 (1.10)	(4.40)	1.000	0.21 (1.00)	0.21 (0.99)	(-0.77)
HI	0.999	0.06 (0.50)	0.07 (0.52)	(0.14)	0.999	0.06 (0.50)	0.06 (0.47)	(-0.88)
HR	0.999	0.42 (3.75)	0.43 (3.88)	(3.47)	1.000	0.42 (3.75)	0.41 (3.72)	(-3.42)
IARER	0.999	0.31 (2.43)	0.31 (2.47)	(0.45)	1.000	0.31 (2.43)	0.31 (2.45)	(1.02)
ICBE	0.999	0.12 (0.86)	0.13 (0.90)	(0.84)	0.999	0.12 (0.86)	0.12 (0.83)	(-1.00)
OC	1.000	0.45 (2.57)	0.46 (2.65)	(3.66)	1.000	0.45 (2.57)	0.45 (2.56)	(-1.38)
RDM	1.000	1.19 (4.16)	1.21 (4.21)	(3.82)	1.000	1.19 (4.16)	1.19 (4.14)	(-2.64)
SmI	0.996	0.37 (5.39)	0.37 (5.42)	(0.10)	0.997	0.37 (5.39)	0.36 (5.35)	(-0.56)
TAN	0.999	0.29 (1.81)	0.28 (1.77)	(-1.37)	1.000	0.29 (1.81)	0.28 (1.79)	(-0.50)
URDI	1.000	0.47 (2.42)	0.46 (2.37)	(-2.54)	1.000	0.47 (2.42)	0.47 (2.41)	(-1.33)
WWI	0.999	0.36 (1.76)	0.41 (1.98)	(4.51)	1.000	0.36 (1.76)	0.35 (1.71)	(-1.67)
Investment								
AGr	0.999	0.63 (3.90)	0.66 (4.05)	(3.97)	1.000	0.63 (3.90)	0.62 (3.83)	(-2.36)
CAPEX	0.998	0.37 (4.67)	0.38 (4.74)	(2.15)	1.000	0.37 (4.67)	0.37 (4.65)	(-1.96)
CDI	1.000	0.21 (2.22)	0.21 (2.18)	(-1.11)	1.000	0.21 (2.22)	0.21 (2.24)	(1.92)
CEI5Y	0.999	0.28 (2.30)	0.27 (2.30)	(-0.33)	1.000	0.28 (2.30)	0.28 (2.30)	(0.80)
ChNOA	0.997	0.27 (4.09)	0.27 (4.10)	(0.86)	1.000	0.27 (4.09)	0.27 (4.08)	(-1.49)
ChPPEIA	1.000	0.63 (5.32)	0.65 (5.44)	(3.51)	1.000	0.63 (5.32)	0.63 (5.31)	(-2.89)
DI	0.998	0.25 (3.78)	0.25 (3.74)	(-1.62)	0.999	0.25 (3.78)	0.25 (3.75)	(0.01)
GrLTNOA	0.999	0.61 (4.41)	0.62 (4.49)	(2.94)	1.000	0.61 (4.41)	0.60 (4.39)	(-3.44)
INV	0.996	0.27 (3.80)	0.28 (3.85)	(0.53)	0.997	0.27 (3.80)	0.28 (3.82)	(0.94)
NDF	0.998	0.34 (4.29)	0.34 (4.31)	(-0.04)	1.000	0.34 (4.29)	0.34 (4.28)	(0.02)
NEF	1.000	0.72 (3.16)	0.69 (3.07)	(-6.04)	1.000	0.72 (3.16)	0.72 (3.17)	(2.81)
NOA	0.998	0.53 (4.99)	0.54 (5.12)	(2.52)	0.999	0.53 (4.99)	0.52 (4.91)	(-1.60)
NOACh	1.000	0.55 (4.03)	0.55 (4.03)	(-0.05)	1.000	0.55 (4.03)	0.55 (4.03)	(-0.50)
SR	0.997	0.20 (2.66)	0.17 (2.31)	(-4.80)	0.999	0.20 (2.66)	0.20 (2.76)	(2.31)
TXFIN	1.000	0.89 (4.80)	0.88 (4.71)	(-3.86)	1.000	0.89 (4.80)	0.90 (4.81)	(1.49)
Profitability								
AT	1.000	0.26 (2.25)	0.26 (2.25)	(-0.26)	1.000	0.26 (2.25)	0.26 (2.25)	(0.34)
CBOP	1.000	0.53 (3.30)	0.53 (3.30)	(-0.19)	1.000	0.53 (3.30)	0.53 (3.31)	(1.79)
CT	1.000	0.28 (1.97)	0.27 (1.96)	(-0.77)	1.000	0.28 (1.97)	0.28 (1.98)	(1.41)
ChiAT	0.999	0.21 (3.60)	0.21 (3.54)	(-1.46)	1.000	0.21 (3.60)	0.21 (3.63)	(1.13)
EC	1.000	0.20 (2.69)	0.20 (2.65)	(-1.63)	1.000	0.20 (2.69)	0.20 (2.70)	(1.01)
EP	0.998	0.72 (5.32)	0.71 (5.28)	(-1.42)	0.999	0.72 (5.32)	0.72 (5.38)	(1.15)
FSc	0.999	0.45 (2.99)	0.41 (2.74)	(-5.17)	1.000	0.45 (2.99)	0.45 (3.05)	(3.22)
GP	0.999	0.34 (2.49)	0.32 (2.32)	(-4.46)	0.999	0.34 (2.49)	0.34 (2.55)	(1.93)
Lvrg	1.000	0.30 (1.81)	0.32 (1.95)	(6.01)	1.000	0.30 (1.81)	0.29 (1.80)	(-2.69)
OPtA	0.999	0.56 (2.93)	0.51 (2.70)	(-5.88)	0.999	0.56 (2.93)	0.58 (3.00)	(3.13)
OPtE	1.000	0.34 (1.78)	0.31 (1.63)	(-4.68)	1.000	0.34 (1.78)	0.35 (1.83)	(2.64)
OSc	1.000	0.08 (0.39)	0.03 (0.15)	(-5.71)	1.000	0.08 (0.39)	0.09 (0.44)	(4.25)
Value								
AM	1.000	0.88 (4.61)	0.91 (4.71)	(4.60)	1.000	0.88 (4.61)	0.88 (4.60)	(-2.60)
BM	1.000	0.98 (5.75)	0.99 (5.77)	(1.36)	1.000	0.98 (5.75)	0.98 (5.74)	(-1.78)
CM	1.000	0.87 (4.05)	0.83 (3.85)	(-5.10)	1.000	0.87 (4.05)	0.88 (4.11)	(3.11)
DurE	1.000	0.94 (4.47)	0.94 (4.45)	(-1.24)	1.000	0.94 (4.47)	0.94 (4.47)	(0.40)
ECoBP	1.000	0.79 (4.36)	0.79 (4.37)	(0.12)	1.000	0.79 (4.36)	0.79 (4.36)	(-0.25)
EM	0.999	0.26 (1.74)	0.28 (1.84)	(2.31)	1.000	0.26 (1.74)	0.26 (1.72)	(-0.45)
IR	1.000	0.49 (2.67)	0.52 (2.84)	(5.70)	1.000	0.49 (2.67)	0.48 (2.64)	(-3.20)
LCoBP	1.000	0.39 (3.14)	0.39 (3.14)	(-0.32)	1.000	0.39 (3.14)	0.39 (3.14)	(0.67)
NPY	1.000	0.90 (4.15)	0.88 (4.05)	(-4.85)	1.000	0.90 (4.15)	0.91 (4.18)	(3.23)
OL	1.000	0.46 (2.90)	0.47 (2.97)	(3.29)	1.000	0.46 (2.90)	0.46 (2.90)	(-0.27)
PY	1.000	0.34 (1.94)	0.33 (1.85)	(-4.75)	1.000	0.34 (1.94)	0.35 (1.99)	(3.83)
SP	1.000	1.01 (4.42)	1.03 (4.49)	(3.08)	1.000	1.01 (4.42)	1.01 (4.42)	(-0.92)
SaGr	0.998	0.24 (2.30)	0.25 (2.39)	(2.06)	1.000	0.24 (2.30)	0.24 (2.28)	(-1.77)
SuGr	0.999	0.17 (1.26)	0.20 (1.52)	(5.09)	0.999	0.17 (1.26)	0.16 (1.19)	(-2.07)

Electronic copy available at: https://ssrn.com/abstract=3150654

Table IA3
Datastream vs Compustat in the common sample - Panel A - detailed

	Signals					Portfolios			
	Correlation		Fama-Mecbeth regressions			CS + CRSP or full DS			
	Pears	Spear	CS	IV from DS	Diff	corr	CS	DS	CS - DS
Accruals									
Acc	0.993	0.987	0.82 (3.61)	0.83 (3.62)	(0.40)	0.997	0.55 (2.63)	0.54 (2.61)	(-0.19)
ChCE	0.966	0.981	1.07 (4.48)	1.07 (4.50)	(-0.10)	0.995	1.16 (3.88)	1.14 (3.86)	(-1.30)
ChCOA	0.951	0.981	0.66 (3.36)	0.69 (3.45)	(1.83)	0.989	0.68 (3.69)	0.66 (3.61)	(-0.92)
ChCOL	0.943	0.972	0.42 (2.65)	0.41 (2.55)	(-0.45)	0.979	0.49 (2.87)	0.46 (2.75)	(-1.13)
ChFL	0.933	0.957	0.40 (3.83)	0.42 (3.94)	(1.33)	0.939	0.26 (2.98)	0.27 (3.21)	(0.13)
ChLTI	0.787	0.640	0.40 (2.71)	0.37 (1.95)	(-0.28)	0.659	0.24 (2.48)	0.18 (1.88)	(-1.02)
ChNCOA	0.966	0.946	1.02 (5.13)	1.10 (5.25)	(4.40)	0.987	0.93 (4.12)	0.94 (4.36)	(0.17)
ChNCOL	0.847	0.843	0.45 (2.87)	0.42 (2.44)	(-0.52)	0.864	0.33 (2.72)	0.26 (2.39)	(-1.06)
ChNFA	0.882	0.884	0.02 (0.23)	0.04 (0.29)	(0.36)	0.883	-0.05 (-0.47)	0.02 (0.23)	(1.73)
ChNNCOA	0.967	0.960	0.97 (5.40)	1.03 (5.64)	(4.04)	0.988	0.86 (4.21)	0.89 (4.44)	(0.65)
ChNNCWC	0.944	0.969	0.35 (2.85)	0.37 (2.98)	(1.01)	0.967	0.45 (3.52)	0.49 (3.77)	(1.63)
GriI	0.917	0.960	0.48 (3.22)	0.47 (2.99)	(-1.19)	0.979	0.42 (2.87)	0.38 (2.69)	(-1.75)
ICh	0.935	0.970	0.47 (3.18)	0.45 (3.03)	(-1.27)	0.981	0.51 (3.15)	0.48 (3.04)	(-1.21)
IGr	0.888	0.966	0.48 (3.08)	0.46 (2.93)	(-1.52)	0.978	0.52 (3.04)	0.44 (2.70)	(-2.45)
MBaAC	0.895	0.896	0.50 (2.56)	0.51 (2.55)	(0.84)	0.972	1.85 (4.87)	1.81 (4.69)	(-0.79)
NWCCh	0.938	0.969	0.19 (1.99)	0.22 (2.20)	(1.11)	0.968	0.31 (3.65)	0.33 (3.67)	(1.03)
POA	0.868	0.981	0.37 (2.70)	0.40 (2.88)	(2.73)	0.978	0.42 (3.30)	0.48 (3.88)	(2.65)
PTA	0.883	0.965	0.54 (3.51)	0.56 (3.43)	(0.86)	0.971	0.52 (4.62)	0.48 (4.27)	(-1.58)
TA	0.913	0.924	1.05 (4.40)	1.12 (4.35)	(1.83)	0.979	1.02 (3.55)	0.97 (3.58)	(-1.39)
Intangibles									
AL	0.820	0.874	0.52 (2.00)	0.89 (2.23)	(2.27)	0.969	0.45 (1.56)	0.64 (1.72)	(1.51)
ChGMChS	0.318	0.804	-0.16 (-1.73)	-0.19 (-1.81)	(-0.52)	0.841	-0.15 (-1.59)	-0.16 (-1.62)	(-0.07)
EPr	0.934	0.961	0.11 (0.46)	0.15 (0.63)	(2.36)	0.976	0.21 (1.12)	0.23 (1.13)	(0.27)
ES	0.994	0.997	0.84 (3.18)	0.85 (3.19)	(0.85)	0.999	0.83 (2.64)	0.82 (2.64)	(-0.99)
HI	0.721	0.747	-0.19 (-1.10)	-0.24 (-0.99)	(-0.51)	0.718	-0.13 (-0.69)	-0.11 (-0.68)	(0.22)
HR	0.858	0.925	0.73 (4.06)	0.78 (4.27)	(2.65)	0.972	0.73 (3.63)	0.72 (3.79)	(-0.33)
IARER	0.660	0.633	0.49 (1.75)	0.37 (1.01)	(-0.43)	0.681	0.75 (2.33)	0.45 (1.52)	(-1.43)
ICBE	0.615	0.644	-0.28 (-0.87)	-0.13 (-0.35)	(0.61)	0.631	-0.20 (-0.52)	-0.07 (-0.34)	(0.45)
OC	0.738	0.963	0.84 (3.36)	0.83 (3.25)	(-0.33)	0.986	0.62 (2.40)	0.58 (2.33)	(-0.82)
RDM	0.889	0.977	2.18 (4.06)	2.21 (4.05)	(1.24)	0.998	1.87 (3.09)	1.86 (3.14)	(-0.15)
SmI	0.931	0.958	0.11 (1.26)	0.07 (0.72)	(-2.64)	0.976	0.09 (0.91)	0.04 (0.40)	(-2.54)
TAN	0.964	0.976	0.46 (1.58)	0.49 (1.63)	(1.98)	0.997	0.40 (1.17)	0.39 (1.15)	(-0.29)
URDI	0.882	0.882	0.42 (1.55)	0.46 (1.60)	(1.97)	0.991	0.54 (1.77)	0.62 (1.89)	(1.94)
WWI	0.998	0.998	1.30 (3.10)	1.30 (3.09)	(-0.25)	0.998	1.11 (3.12)	1.13 (3.24)	(0.62)
Investment									
AGr	0.986	0.985	1.11 (4.84)	1.13 (4.86)	(2.26)	0.995	1.16 (3.84)	1.12 (3.80)	(-1.12)
CAPEX	0.723	0.943	0.65 (5.03)	0.68 (5.39)	(0.90)	0.969	0.59 (4.14)	0.58 (4.42)	(-0.08)
CDI	0.980	0.982	-0.03 (-0.23)	-0.02 (-0.15)	(0.80)	0.963	-0.08 (-0.57)	-0.11 (-0.74)	(-0.75)
CEI5Y	0.893	0.954	-0.20 (-0.68)	-0.23 (-0.74)	(-1.02)	0.995	0.19 (0.94)	0.17 (0.86)	(-1.01)
ChNOA	0.976	0.957	0.40 (2.98)	0.43 (3.25)	(1.41)	0.954	0.32 (2.70)	0.33 (2.87)	(0.25)
ChPPEIA	0.894	0.969	0.72 (4.41)	0.71 (4.29)	(-0.61)	0.990	0.64 (3.52)	0.62 (3.53)	(-0.88)
DI	0.920	0.920	0.38 (2.07)	0.41 (2.10)	(1.22)	0.981	0.18 (2.29)	0.18 (2.36)	(0.40)
GrLTNOA	0.784	0.957	1.01 (4.75)	1.06 (4.81)	(3.58)	0.995	0.77 (4.29)	0.78 (4.29)	(0.74)
INV	0.636	0.895	0.49 (4.51)	0.53 (4.55)	(1.10)	0.918	0.44 (4.02)	0.41 (3.73)	(-0.65)
NDF	0.939	0.963	0.30 (2.60)	0.33 (2.79)	(1.27)	0.923	0.12 (1.35)	0.18 (2.31)	(1.42)
NEF	0.975	0.979	-0.43 (-1.19)	-0.43 (-1.20)	(-0.64)	0.999	-0.05 (-0.13)	-0.06 (-0.17)	(-0.59)
NOA	0.972	0.980	0.87 (2.71)	0.89 (2.72)	(1.33)	0.995	0.73 (2.58)	0.72 (2.60)	(-0.21)
NOACh	0.981	0.990	0.53 (2.99)	0.53 (3.00)	(0.12)	0.995	0.46 (2.32)	0.46 (2.27)	(-0.27)
SR	0.956	0.956	-0.14 (-1.33)	-0.15 (-1.41)	(-0.82)	0.993	-0.17 (-1.74)	-0.19 (-1.88)	(-1.22)
TXFIN	0.954	0.973	0.05 (0.21)	0.07 (0.26)	(0.74)	0.994	0.22 (0.78)	0.18 (0.68)	(-1.20)
Profitability									
AT	0.931	0.991	0.36 (2.58)	0.37 (2.60)	(0.64)	0.993	0.20 (1.35)	0.21 (1.44)	(0.35)
CBOP	0.822	0.899	0.02 (0.08)	0.09 (0.38)	(1.28)	0.809	0.50 (1.75)	0.32 (1.23)	(-1.29)
CT	0.988	0.994	-0.02 (-0.10)	-0.03 (-0.10)	(-0.45)	0.998	0.05 (0.22)	0.06 (0.22)	(0.11)
ChiAT	0.859	0.961	0.16 (1.64)	0.19 (1.92)	(2.24)	0.955	0.10 (0.87)	0.14 (1.24)	(1.30)
EC	0.952	0.961	0.00 (0.03)	-0.02 (-0.18)	(-1.08)	0.952	0.03 (0.25)	0.03 (0.28)	(0.19)
EP	0.696	0.972	0.40 (1.75)	0.39 (1.67)	(-1.12)	0.994	0.51 (2.12)	0.50 (2.12)	(-0.27)
FSc	0.962	0.960	-0.31 (-1.28)	-0.35 (-1.36)	(-1.18)	0.951	-0.36 (-1.29)	-0.36 (-1.30)	(0.05)
GP	0.818	0.904	-0.00 (-0.03)	-0.01 (-0.06)	(-0.14)	0.962	-0.02 (-0.10)	-0.03 (-0.15)	(-0.17)
Lvrg	0.627	0.991	0.30 (0.84)	0.30 (0.83)	(0.02)	0.998	0.49 (1.29)	0.48 (1.25)	(-0.55)
OPtA	0.854	0.934	-0.34 (-1.18)	-0.29 (-1.00)	(1.33)	0.974	-0.03 (-0.10)	-0.00 (-0.01)	(0.53)
OPtE	0.769	0.870	-0.64 (-1.67)	-0.68 (-1.60)	(-0.51)	0.990	-0.33 (-0.78)	-0.31 (-0.73)	(0.36)
OSc	0.770	0.978	-1.14 (-3.25)	-1.12 (-3.19)	(1.05)	0.996	-1.04 (-2.99)	-1.01 (-2.91)	(1.32)
Value									
AM	0.641	0.993	0.65 (1.54)	0.64 (1.53)	(-0.76)	0.998	1.28 (3.03)	1.26 (3.01)	(-0.99)
BM	0.962	0.985	0.97 (3.03)	0.97 (3.04)	(0.59)	0.996	1.18 (3.46)	1.14 (3.28)	(-2.49)
CM	0.655	0.968	-0.31 (-0.68)	-0.30 (-0.64)	(0.29)	0.999	-0.09 (-0.17)	-0.06 (-0.12)	(0.89)
DurE	0.653	0.982	0.74 (2.39)	0.74 (2.41)	(0.48)	0.997	0.75 (2.25)	0.74 (2.26)	(-0.55)
ECoBP	0.598	0.980	0.58 (1.64)	0.60 (1.65)	(1.08)	0.997	0.80 (2.10)	0.74 (1.96)	(-2.58)
EM	0.347	0.907	0.90 (3.27)	1.00 (3.38)	(2.87)	0.991	0.79 (2.64)	0.76 (2.52)	(-0.89)
IR	0.985	0.989	1.20 (4.12)	1.21 (4.15)	(0.46)	0.997	1.12 (3.68)	1.12 (3.72)	(0.16)
LCoBP	0.394	0.955	0.54 (1.96)	0.52 (1.86)	(-0.80)	0.994	0.41 (1.36)	0.35 (1.25)	(-1.51)
NPY	0.683	0.965	-0.40 (-1.22)	-0.41 (-1.25)	(-0.90)	0.996	-0.01 (-0.03)	-0.02 (-0.06)	(-0.32)
OL	0.987	0.988	0.56 (2.51)	0.55 (2.46)	(-0.97)	0.977	0.64 (3.06)	0.59 (2.72)	(-1.52)
PY	0.627	0.947	-0.60 (-2.27)	-0.62 (-2.29)	(-0.70)	0.987	-0.34 (-1.50)	-0.34 (-1.57)	(0.13)
SP	0.663	0.992	1.10 (2.69)	1.10 (2.68)	(0.01)	0.999	1.20 (2.56)	1.18 (2.52)	(-1.36)
SaGr	0.991	0.991	0.24 (1.29)	0.25 (1.34)	(1.54)	0.995	0.26 (1.48)	0.26 (1.53)	(0.37)
SuGr	0.716	0.970	0.89 (4.04)	0.93 (4.16)	(2.77)	0.991	0.96 (3.61)	0.96 (3.73)	(-0.04)

Electronic copy available at: https://ssrn.com/abstract=3150654

Table IA4
Datastream vs Compustat in the common sample - Panel B - detailed

	CRSP returns				Compustat signals			
	Corr	CS	DS	CS - DS	Corr	CS	DS	CS - DS
Accruals								
Acc	0.999	0.55 (2.63)	0.54 (2.58)	(-0.96)	0.997	0.55 (2.65)	0.55 (2.65)	(0.05)
ChCE	0.998	1.16 (3.88)	1.13 (3.84)	(-2.07)	0.997	1.17 (3.88)	1.17 (3.90)	(-0.08)
ChCOA	0.994	0.68 (3.69)	0.68 (3.65)	(0.04)	0.994	0.68 (3.69)	0.66 (3.65)	(-1.41)
ChCOL	0.991	0.49 (2.87)	0.49 (2.84)	(0.10)	0.991	0.49 (2.87)	0.44 (2.70)	(-2.38)
ChFL	0.952	0.26 (2.98)	0.29 (3.35)	(0.95)	0.980	0.26 (2.91)	0.25 (2.90)	(-0.26)
ChLTI	0.658	0.24 (2.48)	0.17 (1.78)	(-1.18)	0.984	0.24 (2.47)	0.25 (2.59)	(0.97)
ChNCOA	0.989	0.93 (4.12)	0.95 (4.28)	(0.62)	0.996	0.93 (4.11)	0.91 (4.16)	(-0.96)
ChNCOL	0.879	0.33 (2.72)	0.27 (2.31)	(-1.08)	0.988	0.33 (2.73)	0.35 (2.88)	(1.09)
ChNFA	0.893	-0.05 (-0.47)	0.03 (0.31)	(1.88)	0.972	-0.04 (-0.43)	-0.06 (-0.54)	(-0.70)
ChNNCOA	0.991	0.86 (4.21)	0.90 (4.42)	(1.33)	0.996	0.86 (4.19)	0.86 (4.24)	(-0.21)
ChNNCWC	0.977	0.45 (3.52)	0.48 (3.67)	(1.19)	0.988	0.45 (3.55)	0.46 (3.62)	(0.37)
GriI	0.984	0.42 (2.87)	0.39 (2.75)	(-1.80)	0.994	0.43 (2.90)	0.42 (2.87)	(-0.82)
ICh	0.988	0.51 (3.15)	0.49 (3.07)	(-0.94)	0.994	0.51 (3.20)	0.51 (3.19)	(-0.20)
IGr	0.988	0.52 (3.04)	0.47 (2.82)	(-2.18)	0.989	0.52 (2.99)	0.49 (2.90)	(-0.97)
MBaAC	0.977	1.85 (4.87)	1.84 (4.74)	(-0.25)	0.986	1.85 (4.87)	1.77 (4.78)	(-1.69)
NWCCh	0.972	0.31 (3.65)	0.33 (3.81)	(1.07)	0.985	0.32 (3.67)	0.34 (3.62)	(1.00)
POA	0.981	0.42 (3.30)	0.46 (3.62)	(1.41)	0.993	0.42 (3.28)	0.44 (3.38)	(1.31)
PTA	0.981	0.52 (4.62)	0.50 (4.31)	(-1.04)	0.988	0.52 (4.64)	0.49 (4.56)	(-1.56)
TA	0.986	1.02 (3.55)	0.97 (3.50)	(-1.55)	0.996	1.02 (3.55)	1.02 (3.60)	(-0.19)
Intangibles								
AL	0.969	0.45 (1.56)	0.62 (1.60)	(1.22)	0.997	0.45 (1.58)	0.43 (1.58)	(-0.65)
ChGMChS	0.857	-0.15 (-1.59)	-0.16 (-1.77)	(-0.15)	0.991	-0.14 (-1.54)	-0.15 (-1.53)	(-0.25)
EPr	0.982	0.21 (1.12)	0.24 (1.19)	(0.57)	0.994	0.22 (1.16)	0.21 (1.08)	(-0.88)
ES	0.999	0.83 (2.64)	0.81 (2.63)	(-1.20)	1.000	0.81 (2.62)	0.82 (2.63)	(1.42)
HI	0.718	-0.13 (-0.69)	-0.10 (-0.63)	(0.30)	0.999	-0.14 (-0.74)	-0.16 (-0.82)	(-1.97)
HR	0.980	0.73 (3.63)	0.74 (3.86)	(0.16)	0.994	0.74 (3.72)	0.73 (3.69)	(-0.31)
IARER	0.680	0.75 (2.33)	0.45 (1.53)	(-1.43)	0.999	0.77 (2.37)	0.78 (2.41)	(0.37)
ICBE	0.630	-0.20 (-0.52)	-0.07 (-0.32)	(0.46)	1.000	-0.20 (-0.52)	-0.21 (-0.54)	(-0.86)
OC	0.989	0.62 (2.40)	0.58 (2.28)	(-1.03)	0.998	0.63 (2.42)	0.62 (2.42)	(-0.44)
RDM	0.998	1.87 (3.09)	1.87 (3.14)	(0.09)	0.999	1.86 (3.09)	1.85 (3.08)	(-0.47)
SmI	0.978	0.09 (0.91)	0.04 (0.45)	(-2.31)	0.998	0.09 (0.97)	0.10 (1.02)	(0.61)
TAN	0.999	0.40 (1.17)	0.38 (1.14)	(-0.95)	0.999	0.39 (1.18)	0.39 (1.16)	(-0.15)
URDI	0.992	0.54 (1.77)	0.62 (1.89)	(2.05)	1.000	0.54 (1.76)	0.53 (1.72)	(-1.71)
WWI	1.000	1.11 (3.12)	1.11 (3.15)	(-0.15)	0.999	1.11 (3.12)	1.13 (3.21)	(0.84)
Investment								
AGr	0.998	1.16 (3.84)	1.14 (3.82)	(-0.66)	0.996	1.15 (3.86)	1.14 (3.84)	(-0.74)
CAPEX	0.971	0.59 (4.14)	0.60 (4.44)	(0.28)	0.996	0.59 (4.22)	0.59 (4.29)	(-0.16)
CDI	0.971	-0.08 (-0.57)	-0.13 (-0.83)	(-1.29)	0.991	-0.12 (-0.80)	-0.10 (-0.74)	(0.64)
CEI5Y	0.995	0.19 (0.94)	0.16 (0.82)	(-1.33)	1.000	0.19 (0.97)	0.19 (0.96)	(-0.38)
ChNOA	0.964	0.32 (2.70)	0.34 (2.98)	(0.62)	0.991	0.32 (2.64)	0.32 (2.61)	(-0.01)
ChPPEIA	0.993	0.64 (3.52)	0.62 (3.53)	(-0.77)	0.997	0.64 (3.53)	0.63 (3.55)	(-0.51)
DI	0.983	0.18 (2.29)	0.18 (2.34)	(0.34)	0.998	0.18 (2.29)	0.18 (2.28)	(0.41)
GrLTNOA	0.996	0.77 (4.29)	0.78 (4.25)	(0.61)	0.999	0.77 (4.33)	0.78 (4.37)	(0.91)
INV	0.926	0.44 (4.02)	0.42 (3.87)	(-0.43)	0.985	0.43 (4.04)	0.41 (3.90)	(-1.23)
NDF	0.953	0.12 (1.35)	0.15 (2.02)	(0.87)	0.965	0.14 (1.60)	0.16 (1.88)	(1.20)
NEF	0.999	-0.05 (-0.13)	-0.06 (-0.17)	(-0.65)	1.000	-0.05 (-0.14)	-0.05 (-0.16)	(-1.15)
NOA	0.997	0.73 (2.58)	0.72 (2.56)	(-0.58)	0.998	0.72 (2.60)	0.73 (2.64)	(0.43)
NOACh	0.997	0.46 (2.32)	0.46 (2.28)	(-0.18)	0.998	0.47 (2.43)	0.47 (2.44)	(-0.54)
SR	0.994	-0.17 (-1.74)	-0.18 (-1.81)	(-0.64)	0.999	-0.17 (-1.74)	-0.18 (-1.81)	(-1.49)
TXFIN	0.995	0.22 (0.78)	0.18 (0.65)	(-1.55)	0.998	0.21 (0.77)	0.21 (0.78)	(0.21)
Profitability								
AT	0.997	0.20 (1.35)	0.20 (1.37)	(-0.13)	0.995	0.21 (1.45)	0.21 (1.46)	(-0.15)
CBOP	0.808	0.50 (1.75)	0.29 (1.13)	(-1.55)	0.997	0.50 (1.75)	0.50 (1.76)	(0.12)
CT	0.999	0.05 (0.22)	0.06 (0.24)	(0.46)	0.999	0.06 (0.24)	0.05 (0.19)	(-0.82)
ChiAT	0.979	0.10 (0.87)	0.15 (1.27)	(2.35)	0.975	0.12 (0.98)	0.10 (0.91)	(-0.92)
EC	0.953	0.03 (0.25)	0.02 (0.19)	(-0.24)	0.999	-0.01 (-0.05)	0.00 (0.00)	(1.99)
EP	0.994	0.51 (2.12)	0.50 (2.09)	(-0.63)	1.000	0.51 (2.10)	0.52 (2.14)	(2.21)
FSc	0.968	-0.36 (-1.29)	-0.32 (-1.16)	(0.66)	0.983	-0.35 (-1.29)	-0.40 (-1.45)	(-1.05)
GP	0.962	-0.02 (-0.10)	-0.03 (-0.12)	(-0.07)	0.997	-0.02 (-0.08)	-0.02 (-0.09)	(-0.13)
Lvrg	0.999	0.49 (1.29)	0.46 (1.21)	(-1.72)	0.999	0.49 (1.29)	0.51 (1.35)	(2.12)
OPtA	0.977	-0.03 (-0.10)	0.02 (0.06)	(1.02)	0.994	-0.03 (-0.10)	-0.08 (-0.25)	(-2.00)
OPtE	0.991	-0.33 (-0.78)	-0.31 (-0.75)	(0.28)	1.000	-0.33 (-0.77)	-0.35 (-0.81)	(-1.36)
OSc	0.998	-1.04 (-2.99)	-1.00 (-2.86)	(1.88)	0.998	-1.04 (-2.98)	-1.06 (-3.02)	(-1.36)
Value								
AM	0.999	1.28 (3.03)	1.27 (3.01)	(-1.03)	0.999	1.28 (3.01)	1.27 (3.06)	(-0.32)
BM	0.997	1.18 (3.46)	1.16 (3.36)	(-1.85)	0.998	1.18 (3.46)	1.16 (3.37)	(-1.80)
CM	0.999	-0.09 (-0.17)	-0.07 (-0.14)	(0.57)	1.000	-0.09 (-0.17)	-0.09 (-0.19)	(-0.64)
DurE	0.998	0.75 (2.25)	0.73 (2.20)	(-1.74)	0.999	0.74 (2.22)	0.76 (2.26)	(1.06)
ECoBP	0.999	0.80 (2.10)	0.76 (2.02)	(-1.99)	0.998	0.80 (2.09)	0.77 (2.04)	(-1.49)
EM	0.993	0.79 (2.64)	0.75 (2.48)	(-1.12)	0.998	0.79 (2.67)	0.79 (2.75)	(-0.19)
IR	0.997	1.12 (3.68)	1.11 (3.72)	(-0.15)	0.999	1.10 (3.66)	1.11 (3.67)	(0.74)
LCoBP	0.997	0.41 (1.36)	0.36 (1.27)	(-1.55)	0.998	0.41 (1.37)	0.39 (1.34)	(-1.07)
NPY	0.997	-0.01 (-0.03)	-0.01 (-0.03)	(0.03)	0.999	-0.02 (-0.05)	-0.03 (-0.11)	(-1.26)
OL	0.981	0.64 (3.06)	0.61 (2.85)	(-0.96)	0.983	0.64 (3.08)	0.64 (2.96)	(-0.34)
PY	0.988	-0.34 (-1.50)	-0.33 (-1.50)	(0.52)	0.998	-0.34 (-1.51)	-0.35 (-1.56)	(-0.41)
SP	1.000	1.20 (2.56)	1.18 (2.51)	(-2.48)	0.999	1.20 (2.55)	1.21 (2.59)	(0.49)
SaGr	0.996	0.26 (1.48)	0.26 (1.52)	(0.45)	0.997	0.26 (1.47)	0.27 (1.56)	(1.08)
SuGr	0.995	0.96 (3.61)	0.98 (3.72)	(0.72)	0.997	0.96 (3.62)	0.94 (3.57)	(-1.64)

Electronic copy available at: https://ssrn.com/abstract=3150654

Table IA5
Datastream vs Compustat in their own full samples - detailed

	Full samples				Cap over $100 million & no financial & 2001+			
	Corr	CS	DS	CS - DS	Corr	CS	DS	CS - DS
Accruals								
Acc	0.953	0.31 (1.39)	0.58 (2.56)	(2.65)	0.956	-0.05 (-0.34)	0.04 (0.27)	(2.82)
ChCE	0.896	0.64 (2.27)	1.40 (4.16)	(3.87)	0.910	0.11 (0.53)	0.26 (1.16)	(2.13)
ChCOA	0.692	0.64 (3.28)	0.73 (3.77)	(0.59)	0.933	0.09 (0.40)	0.04 (0.18)	(-0.85)
ChCOL	0.710	0.55 (3.49)	0.43 (2.07)	(-0.82)	0.946	0.18 (0.90)	0.21 (0.95)	(0.49)
ChFL	0.612	0.50 (5.46)	0.29 (3.41)	(-3.24)	0.909	0.04 (0.28)	0.09 (0.80)	(0.60)
ChLTI	0.400	0.21 (2.71)	0.19 (1.83)	(-0.16)	0.597	0.21 (1.54)	0.26 (1.28)	(0.42)
ChNCOA	0.876	0.82 (3.55)	1.12 (4.70)	(2.11)	0.911	0.24 (1.13)	0.31 (1.61)	(1.49)
ChNCOL	0.802	0.20 (1.74)	0.30 (2.73)	(1.24)	0.874	-0.02 (-0.08)	0.04 (0.23)	(0.56)
ChNFA	0.503	0.29 (2.72)	-0.04 (-0.39)	(-3.04)	0.667	0.01 (0.05)	0.02 (0.17)	(0.15)
ChNNCOA	0.870	0.84 (3.86)	1.05 (4.87)	(1.42)	0.891	0.31 (1.54)	0.36 (2.46)	(0.71)
ChNNCWC	0.777	0.29 (1.97)	0.44 (3.10)	(2.00)	0.867	-0.05 (-0.43)	-0.09 (-0.99)	(-0.65)
GriI	0.771	0.49 (3.26)	0.42 (3.19)	(-0.70)	0.945	0.20 (0.84)	0.07 (0.36)	(-2.27)
ICh	0.760	0.52 (3.09)	0.54 (3.63)	(0.16)	0.913	0.28 (1.12)	0.17 (0.73)	(-2.05)
IGr	0.776	0.50 (2.72)	0.62 (3.85)	(1.06)	0.921	0.02 (0.10)	0.04 (0.18)	(0.31)
MBaAC	0.842	1.67 (4.96)	1.47 (3.69)	(-1.13)	0.939	-0.02 (-0.04)	-0.23 (-0.41)	(-1.98)
NWCCh	0.778	0.43 (4.44)	0.30 (3.16)	(-1.98)	0.841	-0.01 (-0.05)	-0.01 (-0.07)	(-0.02)
POA	0.825	0.68 (5.18)	0.26 (1.96)	(-4.57)	0.941	0.15 (0.65)	0.08 (0.46)	(-0.85)
PTA	0.783	0.35 (3.30)	0.61 (4.06)	(2.35)	0.861	0.13 (1.20)	0.13 (1.21)	(0.03)
TA	0.851	0.64 (2.67)	1.07 (3.87)	(2.61)	0.785	0.28 (1.61)	0.31 (1.86)	(0.38)
Intangibles								
AL	0.927	0.38 (1.35)	0.73 (1.88)	(1.75)	0.808	0.36 (1.02)	0.09 (0.27)	(-2.82)
ChGMChS	0.482	0.16 (1.57)	-0.19 (-1.59)	(-2.90)	0.841	0.01 (0.03)	-0.05 (-0.31)	(-0.51)
EPr	0.835	0.72 (3.82)	0.14 (0.70)	(-4.20)	0.971	0.43 (0.99)	0.34 (0.94)	(-0.91)
ES	0.904	0.07 (0.20)	0.81 (2.83)	(3.67)	0.982	0.06 (0.22)	0.12 (0.43)	(1.08)
HI	0.702	0.00 (0.03)	-0.19 (-0.62)	(-0.88)	0.890	0.26 (1.34)	0.23 (0.82)	(-0.27)
HR	0.821	0.59 (3.09)	0.99 (4.50)	(2.68)	0.942	0.39 (1.69)	0.36 (1.73)	(-0.45)
IARER	0.148	0.31 (2.11)	0.56 (2.41)	(0.89)	0.084	-0.17 (-0.49)	-0.14 (-0.81)	(0.08)
ICBE	0.087	-0.03 (-0.14)	0.07 (0.31)	(0.34)	0.682	0.13 (0.72)	0.25 (1.12)	(0.72)
OC	0.913	0.46 (1.66)	0.97 (3.05)	(3.04)	0.930	0.32 (0.95)	0.30 (0.90)	(-0.23)
RDM	0.978	1.37 (2.31)	1.99 (3.11)	(4.05)	0.970	0.54 (1.61)	0.47 (1.29)	(-0.79)
SmI	0.345	0.33 (2.78)	0.07 (0.74)	(-2.05)	0.805	-0.36 (-4.41)	-0.36 (-4.65)	(0.08)
TAN	0.972	0.45 (1.34)	0.37 (1.03)	(-0.73)	0.981	-0.33 (-1.63)	-0.20 (-1.16)	(2.73)
URDI	0.944	0.52 (1.56)	0.48 (1.45)	(-0.39)	0.955	0.22 (0.90)	0.22 (0.90)	(-0.06)
WWI	0.929	0.44 (1.19)	1.26 (3.59)	(5.49)	0.986	-0.15 (-0.83)	-0.02 (-0.11)	(2.62)
Investment								
AGr	0.880	0.96 (3.01)	1.42 (4.30)	(2.13)	0.951	0.25 (0.97)	0.25 (0.86)	(-0.02)
CAPEX	0.660	0.41 (2.77)	0.60 (4.83)	(1.78)	0.890	0.18 (1.12)	0.15 (0.96)	(-0.60)
CDI	0.822	0.17 (1.05)	-0.06 (-0.37)	(-2.40)	0.893	0.18 (1.06)	0.11 (0.78)	(-0.77)
CEI5Y	0.984	0.41 (2.10)	0.22 (1.13)	(-3.84)	0.978	0.42 (1.73)	0.48 (2.04)	(0.89)
ChNOA	0.586	0.28 (2.66)	0.51 (3.99)	(2.11)	0.838	-0.11 (-0.63)	-0.12 (-0.83)	(-0.32)
ChPPEIA	0.849	0.65 (3.38)	0.81 (4.38)	(1.29)	0.903	0.41 (2.22)	0.38 (2.46)	(-0.44)
DI	0.782	0.35 (3.53)	0.25 (2.25)	(-1.23)	0.955	-0.02 (-0.13)	-0.02 (-0.24)	(-0.20)
GrLTNOA	0.930	0.42 (2.56)	0.86 (4.61)	(6.77)	0.959	0.22 (1.15)	0.33 (2.14)	(2.00)
INV	0.790	0.21 (1.77)	0.45 (3.91)	(3.51)	0.875	0.04 (0.27)	0.10 (0.73)	(0.82)
NDF	0.553	0.42 (4.90)	0.30 (3.34)	(-1.71)	0.630	0.02 (0.12)	0.15 (1.33)	(0.93)
NEF	0.968	0.68 (2.03)	0.01 (0.02)	(-6.08)	0.990	0.33 (1.31)	0.28 (1.05)	(-0.97)
NOA	0.709	0.70 (3.19)	0.91 (3.22)	(1.07)	0.950	0.19 (1.07)	0.23 (1.51)	(0.63)
NOACh	0.936	0.55 (2.59)	0.59 (3.29)	(0.69)	0.936	0.38 (2.48)	0.43 (2.73)	(0.83)
SR	0.892	0.13 (1.16)	-0.27 (-1.76)	(-4.97)	0.976	0.31 (2.51)	0.23 (2.01)	(-2.90)
TXFIN	0.884	1.04 (3.86)	0.29 (1.04)	(-6.44)	0.949	0.54 (1.73)	0.32 (1.08)	(-1.68)
Profitability								
AT	0.946	0.16 (1.12)	0.30 (1.95)	(2.20)	0.967	0.39 (2.03)	0.55 (2.75)	(1.80)
CBOP	0.815	0.80 (2.56)	0.43 (1.61)	(-2.65)	0.826	0.42 (0.83)	0.60 (1.67)	(0.74)
CT	0.726	0.09 (0.42)	0.01 (0.04)	(-0.29)	0.975	0.52 (2.48)	0.52 (2.81)	(0.10)
ChiAT	0.743	0.18 (1.86)	0.12 (1.14)	(-0.96)	0.890	-0.19 (-1.64)	-0.21 (-1.26)	(-0.25)
EC	0.725	0.09 (0.85)	0.11 (1.03)	(0.33)	0.839	-0.04 (-0.41)	0.14 (1.42)	(2.37)
EP	0.914	0.56 (2.63)	0.64 (2.84)	(1.16)	0.950	-0.01 (-0.07)	0.10 (0.51)	(1.33)
FSc	0.749	0.29 (1.03)	-0.71 (-1.83)	(-4.14)	0.739	0.22 (0.59)	0.43 (1.07)	(1.16)
GP	0.835	0.28 (1.17)	-0.01 (-0.07)	(-2.05)	0.923	0.37 (1.32)	0.39 (1.33)	(0.13)
Lvrg	0.967	0.25 (0.72)	0.55 (1.44)	(3.00)	0.973	0.44 (1.28)	0.46 (1.36)	(0.35)
OPtA	0.772	0.93 (2.69)	-0.10 (-0.29)	(-5.84)	0.890	0.51 (1.33)	0.44 (1.56)	(-0.47)
OPtE	0.954	0.44 (1.08)	-0.38 (-0.83)	(-4.75)	0.974	0.39 (1.62)	0.38 (1.88)	(-0.11)
OSc	0.942	0.28 (0.75)	-1.05 (-2.83)	(-8.17)	0.982	0.10 (0.26)	-0.09 (-0.25)	(-2.55)
Value								
AM	0.946	1.09 (2.86)	1.20 (2.89)	(0.71)	0.988	0.43 (0.94)	0.46 (1.03)	(0.62)
BM	0.932	1.20 (3.79)	1.19 (3.37)	(-0.10)	0.978	0.14 (0.40)	0.13 (0.38)	(-0.32)
CM	0.961	0.71 (1.53)	-0.26 (-0.51)	(-6.79)	0.983	0.58 (2.34)	0.44 (1.88)	(-2.54)
DurE	0.914	0.90 (2.65)	0.81 (2.35)	(-0.60)	0.979	0.04 (0.11)	0.07 (0.19)	(0.55)
ECoBP	0.951	0.82 (2.22)	0.52 (1.36)	(-2.29)	0.985	0.16 (0.41)	0.20 (0.50)	(0.98)
EM	0.942	-0.05 (-0.15)	0.76 (2.45)	(5.19)	0.946	-0.31 (-1.44)	-0.09 (-0.45)	(3.15)
IR	0.902	0.58 (1.85)	1.22 (3.83)	(4.16)	0.959	-0.00 (-0.01)	0.08 (0.27)	(0.90)
LCoBP	0.949	0.39 (1.48)	0.22 (0.82)	(-1.78)	0.977	-0.20 (-1.09)	-0.23 (-1.27)	(-0.41)
NPY	0.950	0.87 (2.81)	-0.03 (-0.08)	(-7.80)	0.980	0.49 (1.87)	0.37 (1.32)	(-1.42)
OL	0.761	0.49 (2.73)	0.78 (3.92)	(2.36)	0.908	0.47 (1.99)	0.48 (2.39)	(0.19)
PY	0.864	0.25 (1.16)	-0.40 (-1.42)	(-4.23)	0.934	0.14 (0.59)	0.23 (0.99)	(1.59)
SP	0.973	1.13 (2.55)	1.39 (2.92)	(1.91)	0.987	0.42 (1.15)	0.53 (1.47)	(1.90)
SaGr	0.682	0.21 (1.45)	0.36 (1.91)	(1.05)	0.939	0.18 (1.00)	0.24 (1.30)	(1.09)
SuGr	0.863	0.40 (1.68)	1.04 (3.94)	(4.37)	0.891	0.12 (0.64)	0.23 (1.19)	(1.19)

Electronic copy available at: https://ssrn.com/abstract=3150654

Table IA6
Portfolio constructions reducing discrepancy between databases - detailed

	Large VW				Breakpoints from 1000 largest stocks VW			
	Corr	CS	DS	CS - DS	Corr	CS	DS	CS - DS
Accruals								
Acc	0.940	0.09 (0.96)	0.17 (1.54)	(2.47)	0.953	0.07 (0.56)	0.16 (1.20)	(3.51)
ChCE	0.957	0.21 (0.94)	0.29 (1.26)	(1.57)	0.963	0.27 (1.27)	0.31 (1.39)	(0.82)
ChCOA	0.928	0.12 (0.63)	0.15 (0.66)	(0.42)	0.953	0.10 (0.56)	0.16 (0.75)	(0.96)
ChCOL	0.958	-0.02 (-0.11)	-0.00 (-0.02)	(0.29)	0.973	0.01 (0.03)	0.01 (0.06)	(0.11)
ChFL	0.760	0.29 (2.68)	0.16 (1.73)	(-1.55)	0.809	0.30 (2.72)	0.19 (1.90)	(-1.64)
ChLTI	0.678	-0.02 (-0.17)	-0.04 (-0.23)	(-0.19)	0.731	-0.03 (-0.19)	-0.06 (-0.33)	(-0.40)
ChNCOA	0.819	0.32 (2.25)	0.34 (2.07)	(0.36)	0.748	0.27 (1.89)	0.33 (2.03)	(0.72)
ChNCOL	0.733	-0.08 (-0.56)	-0.07 (-0.60)	(0.05)	0.765	-0.00 (-0.03)	-0.04 (-0.35)	(-0.46)
ChNFA	0.820	0.27 (1.55)	0.20 (0.98)	(-0.76)	0.844	0.25 (1.50)	0.16 (0.79)	(-1.01)
ChNNCOA	0.829	0.42 (3.07)	0.34 (2.23)	(-1.10)	0.789	0.33 (2.52)	0.35 (2.27)	(0.25)
ChNNCWC	0.911	0.29 (1.78)	0.33 (2.08)	(0.52)	0.884	0.16 (1.11)	0.26 (1.86)	(1.64)
GriI	0.896	0.38 (2.52)	0.22 (1.55)	(-2.45)	0.855	0.36 (2.46)	0.23 (1.69)	(-1.56)
ICh	0.828	0.41 (2.50)	0.26 (1.62)	(-1.66)	0.870	0.44 (2.76)	0.32 (2.13)	(-1.84)
IGr	0.911	0.06 (0.32)	0.05 (0.29)	(-0.12)	0.926	0.07 (0.45)	0.08 (0.49)	(0.12)
MBaAC	0.805	0.75 (1.46)	0.72 (1.28)	(-0.07)	0.829	0.90 (1.85)	0.89 (1.60)	(-0.02)
NWCCh	0.895	0.24 (1.71)	0.23 (1.65)	(-0.17)	0.898	0.14 (1.04)	0.20 (1.49)	(0.97)
POA	0.914	0.24 (2.05)	0.36 (2.96)	(1.77)	0.921	0.18 (1.21)	0.27 (2.23)	(1.62)
PTA	0.888	0.16 (0.93)	0.21 (1.18)	(0.60)	0.914	0.09 (0.55)	0.17 (0.98)	(1.38)
TA	0.825	0.17 (0.92)	0.23 (1.11)	(0.66)	0.876	0.15 (0.81)	0.19 (0.90)	(0.44)
Intangibles								
AL	0.664	0.10 (0.56)	0.48 (2.82)	(2.58)	0.766	0.24 (1.35)	0.32 (1.85)	(0.67)
ChGMChS	0.670	0.02 (0.20)	0.04 (0.25)	(0.10)	0.640	0.06 (0.73)	-0.05 (-0.38)	(-0.96)
EPr	0.940	0.42 (2.04)	0.30 (1.63)	(-1.40)	0.931	0.44 (2.20)	0.30 (1.62)	(-1.67)
ES	0.955	0.05 (0.19)	0.26 (1.04)	(2.00)	0.961	0.13 (0.67)	0.25 (1.05)	(1.32)
HI	0.528	-0.05 (-0.40)	0.09 (0.55)	(0.89)	0.558	0.07 (0.46)	0.08 (0.42)	(0.01)
HR	0.959	0.04 (0.17)	-0.01 (-0.05)	(-0.77)	0.965	-0.05 (-0.22)	-0.07 (-0.29)	(-0.38)
IARER	0.067	0.31 (1.34)	0.44 (1.20)	(0.34)	0.079	0.32 (1.49)	0.13 (0.34)	(-0.34)
ICBE	0.537	0.14 (0.94)	0.13 (0.92)	(-0.08)	0.525	0.27 (1.26)	0.20 (1.32)	(-0.36)
OC	0.916	0.28 (1.55)	0.27 (1.38)	(-0.23)	0.931	0.28 (1.53)	0.28 (1.46)	(0.11)
RDM	0.946	0.54 (1.93)	0.71 (2.48)	(1.89)	0.967	0.52 (2.21)	0.69 (2.59)	(2.54)
SmI	0.713	0.13 (0.92)	0.12 (0.72)	(-0.07)	0.764	0.12 (0.85)	0.11 (0.63)	(-0.10)
TAN	0.942	0.02 (0.17)	0.04 (0.23)	(0.19)	0.957	0.06 (0.44)	0.11 (0.72)	(1.07)
URDI	0.865	0.34 (1.76)	0.45 (1.71)	(0.90)	0.868	0.34 (1.76)	0.45 (1.72)	(0.93)
WWI	0.968	0.02 (0.07)	0.35 (0.95)	(2.69)	0.980	0.13 (0.48)	0.37 (1.27)	(3.06)
Investment								
AGr	0.960	0.31 (1.43)	0.32 (1.39)	(0.15)	0.965	0.25 (1.08)	0.27 (1.16)	(0.53)
CAPEX	0.731	0.22 (1.29)	0.11 (0.61)	(-0.93)	0.786	0.13 (0.81)	0.07 (0.46)	(-0.60)
CDI	0.918	0.08 (0.51)	-0.05 (-0.31)	(-1.44)	0.898	0.08 (0.56)	-0.04 (-0.24)	(-1.12)
CEI5Y	0.967	0.28 (1.76)	0.32 (1.83)	(0.68)	0.963	0.23 (1.44)	0.21 (1.23)	(-0.34)
ChNOA	0.825	0.28 (2.12)	0.40 (2.41)	(1.00)	0.660	0.32 (2.45)	0.24 (2.16)	(-0.63)
ChPPEIA	0.909	0.25 (1.60)	0.27 (1.87)	(0.26)	0.920	0.26 (1.80)	0.27 (1.86)	(0.03)
DI	0.884	0.26 (2.59)	0.21 (1.52)	(-0.78)	0.887	0.26 (2.61)	0.22 (1.59)	(-0.66)
GrLTNOA	0.809	0.19 (1.84)	0.20 (2.15)	(0.11)	0.809	0.21 (1.68)	0.20 (1.88)	(-0.21)
INV	0.886	0.17 (1.29)	0.15 (0.94)	(-0.45)	0.914	0.18 (1.72)	0.16 (1.29)	(-0.32)
NDF	0.726	0.18 (1.70)	0.20 (2.23)	(0.23)	0.787	0.16 (1.29)	0.19 (2.01)	(0.43)
NEF	0.950	0.22 (0.77)	0.05 (0.17)	(-2.21)	0.962	0.18 (0.71)	0.13 (0.48)	(-0.77)
NOA	0.614	0.45 (3.56)	0.44 (2.83)	(-0.09)	0.598	0.41 (3.33)	0.46 (3.51)	(0.35)
NOACh	0.816	0.40 (2.58)	0.56 (3.73)	(1.79)	0.865	0.38 (3.21)	0.46 (3.75)	(1.53)
SR	0.914	0.03 (0.24)	0.05 (0.32)	(0.33)	0.918	0.03 (0.26)	0.04 (0.24)	(0.11)
TXFIN	0.862	0.54 (2.17)	0.31 (1.25)	(-2.39)	0.878	0.26 (1.30)	0.22 (1.04)	(-0.50)
Profitability								
AT	0.981	0.21 (1.30)	0.31 (1.65)	(1.80)	0.975	0.25 (1.70)	0.34 (1.92)	(1.45)
CBOP	0.778	0.64 (2.54)	0.49 (2.05)	(-1.13)	0.801	0.57 (2.56)	0.48 (2.27)	(-0.88)
CT	0.794	0.06 (0.29)	0.16 (1.01)	(0.73)	0.830	0.05 (0.23)	0.16 (1.03)	(0.92)
ChiAT	0.860	0.24 (1.49)	0.22 (1.47)	(-0.28)	0.895	0.15 (1.01)	0.11 (0.75)	(-0.84)
EC	0.905	0.24 (1.88)	0.12 (1.08)	(-2.20)	0.925	0.21 (1.60)	0.15 (1.18)	(-1.17)
EP	0.965	0.32 (1.15)	0.36 (1.29)	(0.76)	0.978	0.37 (1.30)	0.37 (1.36)	(0.03)
FSc	0.620	-0.02 (-0.07)	-0.30 (-0.81)	(-0.83)	0.635	0.09 (0.34)	-0.39 (-1.07)	(-1.52)
GP	0.928	0.13 (0.64)	0.22 (0.96)	(1.15)	0.927	0.14 (0.71)	0.15 (0.66)	(0.05)
Lvrg	0.983	0.03 (0.07)	0.01 (0.03)	(-0.24)	0.984	0.10 (0.29)	0.08 (0.26)	(-0.32)
OPtA	0.820	0.48 (1.85)	0.41 (1.68)	(-0.56)	0.790	0.43 (1.82)	0.28 (1.41)	(-1.33)
OPtE	0.864	0.39 (1.36)	0.12 (0.36)	(-1.67)	0.829	0.31 (1.44)	0.24 (0.84)	(-0.43)
OSc	0.931	0.11 (0.44)	-0.10 (-0.41)	(-2.59)	0.935	0.01 (0.04)	-0.16 (-0.83)	(-2.36)
Value								
AM	0.986	0.15 (0.50)	0.19 (0.63)	(1.00)	0.985	0.15 (0.49)	0.22 (0.71)	(2.00)
BM	0.970	0.20 (0.73)	0.29 (1.06)	(1.66)	0.975	0.12 (0.44)	0.19 (0.73)	(1.57)
CM	0.968	0.39 (1.12)	0.30 (0.78)	(-1.06)	0.976	0.38 (1.17)	0.24 (0.71)	(-2.10)
DurE	0.903	0.23 (0.78)	0.13 (0.48)	(-0.97)	0.918	0.21 (0.73)	0.24 (0.83)	(0.33)
ECoBP	0.969	0.11 (0.34)	0.14 (0.41)	(0.41)	0.984	0.08 (0.24)	0.10 (0.30)	(0.44)
EM	0.923	0.10 (0.44)	0.27 (1.16)	(1.52)	0.948	0.20 (1.00)	0.23 (1.30)	(0.61)
IR	0.965	0.15 (0.54)	0.16 (0.59)	(0.10)	0.967	0.17 (0.60)	0.25 (0.94)	(1.78)
LCoBP	0.964	0.27 (1.02)	0.46 (1.73)	(2.00)	0.983	0.25 (0.96)	0.36 (1.26)	(1.61)
NPY	0.950	0.37 (1.11)	0.23 (0.77)	(-1.84)	0.938	0.17 (0.61)	0.12 (0.40)	(-0.51)
OL	0.904	0.21 (1.07)	0.27 (1.47)	(0.93)	0.860	0.23 (1.36)	0.29 (1.90)	(0.77)
PY	0.900	-0.01 (-0.03)	0.06 (0.15)	(0.42)	0.751	0.13 (0.41)	-0.09 (-0.29)	(-1.24)
SP	0.973	0.26 (0.81)	0.42 (1.28)	(3.39)	0.985	0.28 (0.90)	0.40 (1.19)	(2.23)
SaGr	0.960	0.09 (0.40)	0.02 (0.08)	(-0.95)	0.961	0.06 (0.27)	0.01 (0.04)	(-0.73)
SuGr	0.938	0.16 (0.90)	0.23 (1.15)	(0.92)	0.952	0.20 (1.04)	0.24 (1.18)	(0.81)

Electronic copy available at: https://ssrn.com/abstract=3150654