https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–23
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

The Loan Fee Anomaly: A Short Seller's Best Ideas

Joseph E. Engelberg,[a] Richard B. Evans,[b] Greg Leonard,[c] Adam V. Reed,[d] Matthew C. Ringgenberg[e,*]

[a] University of California, San Diego, La Jolla, California 92093; [b] University of Virginia, Charlottesville, Virginia 22903; [c] Virginia Tech, Blacksburg, Virginia 24061; [d] University of North Carolina, Chapel Hill, North Carolina 27599; [e] University of Utah, Salt Lake City, Utah 84112
*Corresponding author

Contact: jengelberg@ucsd.edu, https://orcid.org/0000-0002-0092-441X (JEE); evansr@darden.virginia.edu, https://orcid.org/0000-0003-4631-0744 (RBE); gregory_leonard@kenan-flagler.unc.edu (GL); adam_reed@unc.edu, https://orcid.org/0000-0002-9507-8905 (AVR); matthew.ringgenberg@eccles.utah.edu, https://orcid.org/0000-0002-3013-2085 (MCR)

Received: January 13, 2023
Revised: September 24, 2023
Accepted: December 11, 2023
Published Online in Articles in Advance: October 9, 2024

https://doi.org/10.1287/mnsc.2023.00152

Copyright: © 2024 INFORMS

Abstract. We find that equity loan fees, which have been largely ignored by the anomalies literature, are the best predictor of cross-sectional returns. When compared with 102 other anomalies and other short-selling measures, the loan fee anomaly has the highest monthly long-short return (4.01%), the highest monthly Sharpe Ratio (0.66), and, unlike other anomalies, exhibits strong persistence throughout the sample. Although prior work has shown that existing anomalies reside in high loan fee stocks, we find that 42% of loan fee outperformance is due to unique information not contained in other anomalies. Future papers that examine cross-sectional predictors of returns should include the single most effective predictor: loan fees.

History: Accepted by Victoria Ivashina, finance.
Supplemental Material: The online appendix and data files are available at https://doi.org/10.1287/mnsc.2023.00152.

Keywords: asset pricing anomalies • equity loan fees • short selling

1. Introduction

Several recent papers show that equity loan fees are related to future stock returns.[1] The predictive ability of loan fees may not be surprising to those in the short-selling literature: short-sellers are well known to be informed traders, and high loan fee stocks represent investments that short-sellers are willing to pay the most for—that is, their very best ideas. Yet, despite the evidence on short selling, the literature on asset pricing anomalies has largely overlooked just how powerful loan fees are in determining the cross-section of returns. Of eight well-known meta-anomaly papers, only four examine short interest and *zero* examine loan fees.[2] We show that equity loan fees are arguably the single strongest predictor of cross-sectional returns.

When we compare the loan fee anomaly to the 102 anomalies in Green et al. (2017) (hereafter GHZ), as well as other measures of short-selling activity, we find that, during our 2013–2019 sample period, it has the highest average long-short monthly return (4.01%), the highest Sharpe Ratio (0.66), and the highest percentage of months with a positive return (74.4%). We find similar results for portfolio values: $1 invested in the top percentile of a composite anomaly portfolio, which is formed by combining the GHZ anomalies, would have finished with $1.47 by the end of our sample. The same dollar invested in the top percentile of the loan fee anomaly portfolio would have finished with $18.76.[3]

Moreover, the loan fee anomaly is not only unique in its performance, but also in its persistence. Whereas most anomalies fade over time (McLean and Pontiff 2016), the loan fee anomaly remains strong throughout our sample. When we sort all anomalies based on their past performance, the future performance of the best anomalies in GHZ is statistically indistinguishable from the worst. In contrast, the loan fee anomaly persists in its superior performance, outperforming the best anomalies in the existing literature.

Because the existing literature shows that short sellers are informed traders and that anomalies tend to be concentrated in hard-to-short stocks, it is natural to wonder whether the predictive performance of the lending fee is simply a combination of existing anomalies.[4] With this in mind, we project loan fee on all 102 anomalies and decompose it into a fitted and a residual component—the fitted value measures the portion of loan fees that are spanned by existing anomalies, while the residual measures the unique information in loan fees that is distinct from existing anomalies. When we form a long-short trading strategy based on the fitted loan fee, we find that this aggregation of anomalies returns 2.15% per month, which places it among the top 5% of the GHZ anomalies. When we examine the residual loan fee, we find it returns 1.64% per month, outperforming most of the GHZ anomalies. In other words, short sellers' best ideas appear to be, in part, ideas about how to selectively reweight existing anomalies.

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Although these results show there is unique information in the loan fee signal, a difference in levels does not tell us how much of the relation between loan fee and returns is explained by each component. Using a regression approach, we find that the fitted loan fee explains 58% of one-month-ahead return variation, while the residual explains 42%. Similarly, when we use a principal components approach, we find the fitted loan fee explains 32% of one-month-ahead return variation, while the residual explains 68%. These results suggest that a substantial component of a short seller's best ideas are *not* spanned by existing anomalies. In other words, existing anomalies are missing information about some of the most mispriced stocks, but short sellers' best ideas contain this valuable information.

As an example of the type of information that could be contained in the residual loan fee, we repeat the analysis in the context of securities fraud class-action lawsuits. The existing literature finds that short sellers are skilled at uncovering financial fraud (Karpoff and Lou 2010). Consistent with this, we find that loan fee predicts the occurrence of lawsuits related to financial misconduct. Moreover, we also find that the residual component of loan fees explains more of short sellers' best ideas about financial lawsuits than the fitted component of loan fees, suggesting again that their best ideas are not spanned by existing anomalies.

In light of the way that loan fee aggregates both existing and unique information, we then examine whether the loan fee anomaly outperforms other well-known statistical techniques for combining signals. We find that it does. When we compare loan fee to anomaly signals that were aggregated using GHZ Net, Ordinary Least Squares (OLS), Principal Components Regression (PCR), Partial Least Squares (PLS), Elastic Net (ENET), and Group Lasso with a Generalized Linear Model (GLM), we find that it consistently outperforms all six other methods. In particular, the loan fee's aggregation generates returns of 4.01% per month, and the next best aggregation method, GLM, generates returns of 1.95% per month.

The primary contribution of our study is to point out that the anomalies literature has ignored perhaps the best return predictor in the data. For those interested in models of expected returns, this omission is critical. However, the loan fee is a transaction cost. For practitioners, a natural question arises: is the loan fee anomaly a profitable trading strategy after accounting for the cost of loan fees? We find the answer is "yes." For each of the 102 GHZ anomalies, we subtract the cost of borrowing the stocks in the short leg of the anomaly portfolio. We do the same for the loan fee anomaly. After this adjustment, we find that the loan fee anomaly continues to perform well relative to the 102 GHZ anomalies. Specifically, the long-short return for the loan fee anomaly is 0.75% per month, compared with −0.15% for the average GHZ anomaly.

Finally, we examine a key question: why are loan fees so informative? A large literature shows that short sellers are, on average, informed traders. Ideally, we could view their expected return forecasts, but, unfortunately, no such database exists. We argue that equity loan fees are a better proxy for these forecasts than short interest. The intuition is simple: loan fee proxies for short sellers' willingness to pay—a high loan fee for a stock indicates that some short sellers have a low enough expected return that they expect short selling to yield a profit, even after paying the loan fee. However, short interest may not always be high when short sellers have low expected returns. Short interest could vary in the cross-section due to variation in loan supply that has nothing to do with the expected return forecasts of short sellers (e.g., Cohen et al. 2007). When we examine loan supply in the context of loan fees and short interest, we find that this is indeed the case (see Table 13). In summary, whenever loan fees are high, it is likely that the marginal short seller has a low expected return forecast because they are *choosing* to pay this cost, but this need not be the case for short interest.[5]

Although loan fees represent a cost faced by short sellers, short sellers face other limits to arbitrage, including liquidity. For example, Hou et al. (2020) document that anomalies tend to live in small and illiquid stocks. Our finding is distinct from this. In robustness tests, when we run a horse race in a Fama and MacBeth (1973) regression examining the interaction between our composite anomaly variable and size, illiquidity, and loan fee, the loan fee interaction drives out the other two. Put differently, our finding about short sellers' best ideas is distinct from the recognized association between anomaly returns, size, and liquidity.

In other robustness tests, we also examine how our results hold up throughout the cross-section and in comparison with other measures of short selling. For the cross-section, we define microcap stocks as those with market capitalization below the 20th-percentile NYSE breakpoint, as in Fama and French (2008). We then re-examine our findings when we limit the sample to exclude microcap stocks. All of our conclusions hold—the loan fee anomaly remains the number one anomaly, outperforming all 102 GHZ anomalies. As for alternative measures of short selling, we compare the loan fee anomaly to four other well-known variables: short-interest (Asquith et al. 2005), days-to-cover (Hong et al. 2021), CME Beta (Drechsler and Drechsler 2016), and short interest ratio times return volatility (Hanson and Sunderam 2014). We find that loan fee is not only the superior predictor of returns, but also drives out the return predictability of these alternatives.

Our paper makes contributions to several strands of the asset pricing anomalies literature. Although this literature began by examining individual anomalies (e.g., Fama and French 1992, Jegadeesh and Titman 1993,

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Cooper et al. 2008), it has evolved into examining large sets of anomalies (e.g., McLean and Pontiff 2016, Green et al. 2017, Hou et al. 2020, Chen and Velikov 2023). Our first contribution is to identify loan fee as arguably the best cross-sectional predictor among a comprehensive set of anomalies. To date, the best predictor of cross-sectional returns, loan fee, has been left out of the sets examined in this literature (e.g., McLean and Pontiff 2016, Green et al. 2017, Hou et al. 2020, Chen and Velikov 2023).

Our second contribution relates to the growing literature that attempts to maximize the out-of-sample return predictability by combining existing anomalies using state-of-the-art econometric techniques (e.g., Lewellen 2015, Gu et al. 2020). In particular, we compare loan fee to anomaly signals that were aggregated using GHZ Net, Ordinary Least Squares, Principal Components Regression, Partial Least Squares, Elastic Net, and Group Lasso with a Generalized Linear Model. We find that loan fee's aggregation outperforms all six other methods. The consistent outperformance of loan fee relative to these other methods suggests that the weighting inherent in loan fee, with its clear economic interpretability, outperforms other purely statistical techniques.

Moreover, as a third contribution, we show that loan fee incorporates valuable information not captured by existing anomalies. In other words, although it is true that loan fee contains the information in many existing anomalies, it is also true that short sellers' best ideas contain unique information: loan fee is able to capture hard-to-quantify, idiosyncratic information that is not contained in existing anomalies.

Fourth, our findings are also related to the recent literature on an investor's "best ideas." Most of this work has focused on mutual funds, showing that where a manager's conviction is strongest, these best ideas are strongly related to the cross-section of returns (Jiang et al. 2014, Anton et al. 2021). We extend this literature to the domain of short sellers. In so doing, we have two advantages. First, the short-selling data cover a wide range of investors, not just mutual fund managers. Second, the investors who choose to short-sell are likely among the best-informed market participants. As a result, their best ideas should be included in any discussion of the best of ideas of investors.

The rest of the paper proceeds as follows. Section 2 describes our sample and outlines our methodology. Section 3 provides a detailed discussion of the existing literature. Section 4 displays our main results. Section 5 concludes.

2. Data, Methodology. and Summary Statistics

To examine short sellers' best ideas, we combine data from Compustat, I/B/E/S, the Center for Research in Security Prices (CRSP), and equity lending data from the FIS Astec Analytics Short Lending Data.

2.1. Data and Sample Construction

Although there is a growing list of papers that examine large sets of anomalies, we use the 102 anomalies in Green et al. (2017) not only because their list includes the most widely cited anomalies (e.g., momentum (Jegadeesh and Titman 1993), asset growth (Cooper et al. 2008), book-to-market ratio (Rosenberg et al. 1985), etc.) but also because Jeremiah Green provides the code used to construct each anomaly on his website.[6] The public availability of the underlying data for 102 anomalies facilitates both replication of our results and additional exploration of anomaly findings.

From Compustat, we collect a number of accounting data items that are necessary to compute asset pricing anomalies. From CRSP, we get data on monthly returns, including dividends, trading volume, share price, and shares outstanding for all common U.S. equities (i.e., assets with CRSP share codes of 10 or 11). We filter our sample to exclude extremely small stocks—those below the 5th percentile of the NYSE market capitalization breakpoints using data provided on Kenneth French's website.[7] In robustness checks, discussed in Section 4.7, we show that our main conclusions are unchanged if we exclude all stocks below the 20th percentile of NYSE size breakpoints, following Fama and French (2008).

Finally, we add securities lending data from FIS Astec Analytics Short Lending Data to our database. FIS Astec collects the data from banks and major securities lending custodians and reports the data aggregated to the daily level. We use FIS Astec's retail average loan rate, which is the daily average fee paid by the ultimate borrower of the shares, as our measure of the loan fee.[8]

2.2. Methodology

Although we discuss our methodology in greater detail in Section 4 below, here, we introduce two key components. First, in all of our analyses, we compare the loan fee anomaly to: (i) the set of 102 anomalies individually and (ii) an aggregate of these anomalies, which we call the "GHZ Net" anomaly (as in Engelberg et al. 2018a). The GHZ Net anomaly variable aggregates the signals for all of the anomalies compiled by Green et al. (2017) in the following way: for each stock-month observation, we calculate the number of long-side (Long) and short-side (Short) anomaly portfolios that a given firm belongs to and then calculate the difference between the Long and Short variables. For example, a stock that is in 17 Long portfolios and 13 Short portfolios will have a Net = 17 − 13 = 4. In general, stocks that are in many Long (Short) anomaly portfolios will have a "Net" anomaly value that is large and positive (negative).

Second, because the performance of any anomaly depends on how the Long and Short sides of the

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

anomaly are defined (e.g., top and bottom percentile, decile, ventile, etc.), we report our results across four different definitions of the anomaly portfolios. Specifically, we present results from calculating anomalies using:

1. Top and bottom 1% (percentiles),
2. Top and bottom 2% (quinquagintiles),
3. Top and bottom 5% (ventiles), and
4. Top and bottom 10% (deciles).

In assigning stocks to portfolios, we need to address discreteness associated with any sorting variable. For the 1% example, we identify the set of stocks on each date that have a value greater than the 99th percentile and that have a value less than the 1st percentile. If there are no values greater than the 99th percentile (i.e., the 99th percentile is also the minimum value), we include all stocks that have a value equal to the 99th percentile. We employ an analogous methodology for the 98th, 95th, and 90th percentiles.

Because we are after short sellers' *best* ideas, we are most interested in how stocks with the most extreme loan fees perform. The extreme loan fees are best captured by the top/bottom 1% sample. However, the top/bottom 1% sample naturally relies on relatively few stocks to form each leg of the long/short portfolio and differs from the standard top/bottom 10% that is typically used in the literature. As such, we also present all of our results using the top/bottom 10%. We report results using the top/bottom 2% and top/bottom 5% as further evidence of the robustness of our result to how we define the long/short portfolios. Our results are not sensitive to these definitions.

2.3. Summary Statistics

Table 1 presents summary statistics for our sample, which covers the period July 2013 through December 2019. Panel A presents summary statistics for our key short-selling variables and demonstrates that short-selling variables exhibit large variation. The median *Loan Fee* is only 59 basis points per annum, yet the mean *Loan Fee* is 132 basis points per annum, and the 99th percentile is 1,710 basis points per annum. The data show that short sellers are, at times, willing to pay a large fee to act on their beliefs. The data on short interest as a percent of shares outstanding (*SIR*) and days-to-cover (*DTC*) exhibit a similar pattern. In both cases, the mean is significantly higher than the median. Panel B provides summary statistics for the *GHZ Net* variable. We also provide summary statistics for the 102 anomalies in the GHZ data set in Table A1 in the Online Appendix.

Table 1. Summary Statistics

Anomaly	Mean	St. Dev.	1st %	Median	99th %
Panel A: Short selling related anomalies					
Loan Fee	132.0	586.0	2.0	59.0	1,710.0
Days to Cover	7.2	33.2	0.7	5.1	32.4
Short Interest	5.7	6.3	0.2	3.4	29.9
Panel B: GHZ aggregate					
GHZ Net	−4.1	6.2	−21.0	−4.0	10.0

Notes. This table displays means, standard deviations (St. Dev.), and the 1st, 50th, and 99th percentiles of the anomaly sorting variables. The *GHZ Net* anomaly variable aggregates the signals for all of the anomalies compiled by Green et al. (2017); see Section 2 of the text for detailed variable definitions. The sample consist of all U.S. common equities above the 5th percentile NYSE size breakpoint over the period August 2004 to December 2019.

3. Detailed Literature Review

By showing that the loan fee anomaly predicts returns more strongly than other anomalies, we touch on several extant literatures, including the literature on asset pricing anomalies and the literature on short sellers as informed traders.

First, in the area of anomalies, we join a growing body of work. Asset pricing anomalies have been documented since at least Ball and Brown (1968). Although the literature initially focused on individual anomalies, increasingly, the literature examines large sets of anomalies at the same time in order to draw more general conclusions about asset pricing. Whereas Chen and Zimmermann (2022) find that most anomalies are replicable in-sample, a growing branch of the anomalies literature examines the extent to which anomalies are robust out-of-sample and to data mining concerns. McLean and Pontiff (2016) find that anomalies tend to decrease in profitability after academic papers examining them are first published and argue that this is due to informed trading postpublication. Harvey et al. (2016) argue that most anomalies reported in the asset pricing literature are the result of data mining. Similarly, Hou et al. (2020) find that most anomalies disappear once microcaps are excluded and returns are value-weighted.

One implication of our finding for this literature is that loan fee's predictive performance is strong, stable, and persistent, suggesting that although anomalies based on systematic information may fade or be subject to data mining, there is a consistent amount of mispricing stemming from idiosyncratic information that remains in the market. Another implication for this branch of the literature is that some of the anomalies may simply be artifacts of the relationship between loan fees and returns. If that is the case, their tendency to fade postpublication or failure to pass multiple hypotheses testing significance threshold does not reflect informed trading or data mining, but, rather, time varying and weak correlation with loan fees.

Another branch of the literature seeks to identify the characteristics that provide independent information about the cross-section of returns. Green et al. (2017) find that 12 characteristics out of 94 are reliable

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

independent predictors using linear regression. Freyberger et al. (2020) use nonparametric methods and find that most of the 62 return predictors they study do not provide additional information. Our paper has two implications for this branch of the literature. First, some of the characteristics that these authors find to have independent information may no longer have such information once loan fees are included in the analysis. Second, loan fee should be added to the list of anomalies with independent information, given that 71% of loan fees' return predictability comes from information not contained in existing anomalies. Future papers that wish to work with the consolidated set of anomalies from these papers should add loan fee to their list of anomalies.

Gu et al. (2020) use machine learning methods that allow for nonlinear interactions between signals to forecast returns. They find that variations of momentum, liquidity, and volatility are the best predictive signals. Our paper is particularly relevant to this strand of the literature that seeks to identify the strongest predictor of returns. We find that loan fee is overwhelmingly the strongest predictor of returns, especially when we focus on the top/bottom 1% of stocks for each signal. Loan fee also has an intuitive appeal as the strongest predictor of returns. Although a researcher can use sophisticated algorithms using firm characteristics to forecast expected returns, loan fee compiles the information of the market into a single price signal that reveals which stocks the market believes are overvalued.

In the short-selling literature, several papers show that measures of short-selling activity, including equity loan fees, are related to future stock returns. In particular, papers such as Asquith et al. (2005), Cohen et al. (2007), Hong et al. (2021), and Jones and Lamont (2002) show that future returns are lower when short interest is high. Moreover, several papers document evidence that short sellers are informed investors who are skilled at processing information (Boehmer et al. 2008, Karpoff and Lou 2010, Engelberg et al. 2012, Henderson et al. 2023). Of particular interest for this paper, several recent papers find evidence that equity loan fees contain information about future returns (Cohen et al. 2007, Blocher et al. 2012, Drechsler and Drechsler 2016, Blocher and Ringgenberg 2018, Engelberg et al. 2018b). Although the predictive ability of loan fees may not be surprising to those in the short-selling literature, the literature on asset pricing anomalies has largely overlooked just how powerful loan fees are in determining the cross-section of returns. For example, although there are several meta-anomaly papers that collectively examine hundreds of anomaly variables, we know of no meta-anomaly paper that examines loan fees.

A number of papers have connected short selling to various trading strategies. Dechow et al. (2001) show that firms with low ratios of fundamentals to market values (such as earnings and book values) have increased short interest. Similarly, Geczy et al. (2002) suggest that short selling can be driven by well-accepted patterns in stock prices, including size, book to market, and momentum effects. Stambaugh et al. (2012) argue that much of the profitability of anomalies comes from the short side, and Drechsler and Drechsler (2016) find that loan fees are an indication of profitability within anomaly strategies. Although these papers focus on how existing anomalies are affected by short-selling constraints, using a larger set of anomalies, we find that much of loan fee predictive ability lies *outside* existing anomalies.

4. Results

In this section, we examine loan fee as a cross-sectional predictor of returns. We start by comparing the performance of the loan fee anomaly to the performance of the 102 stock return anomalies compiled and analyzed in Green et al. (2017).

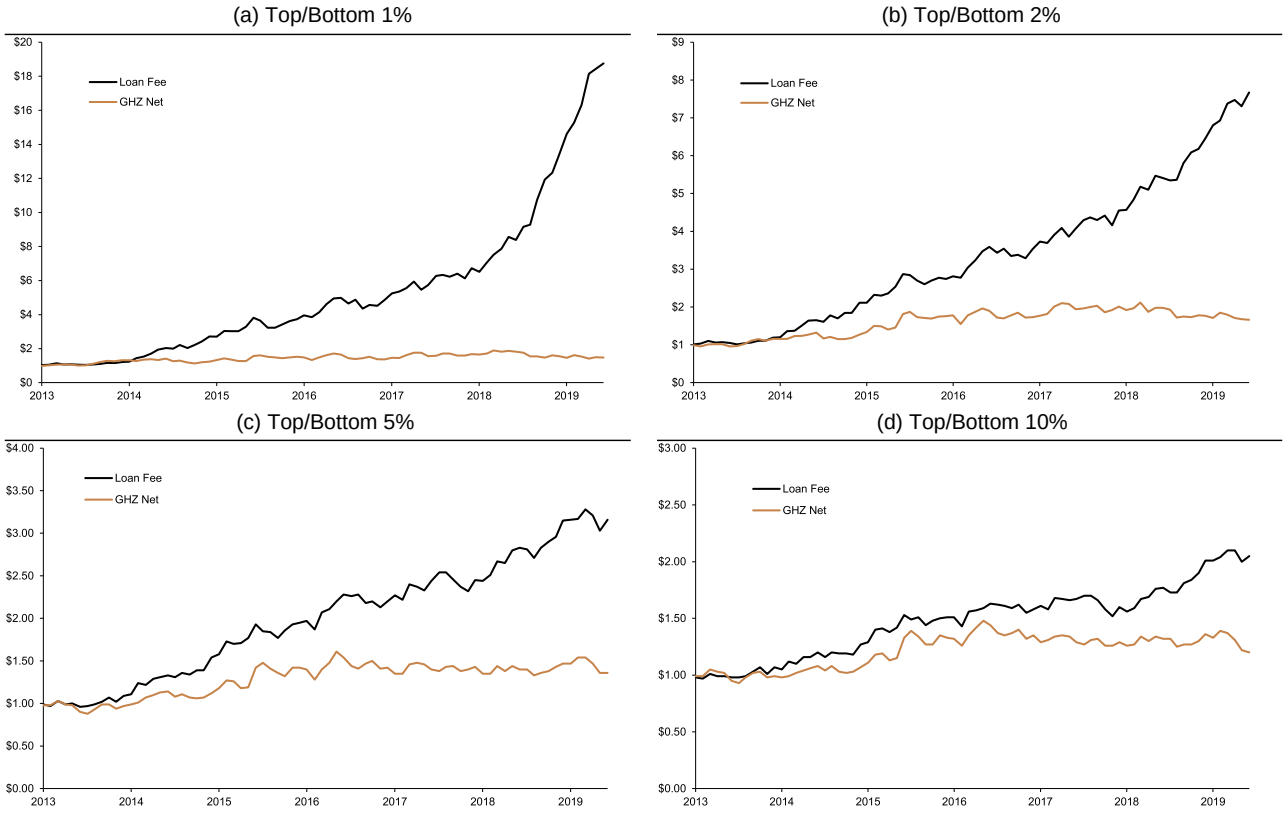
4.1. *Loan Fee* Anomaly Performance

We first examine the time-series performance of the loan fee anomaly. To provide a simple comparison of the performance of the long-short loan fee anomaly relative to the GHZ Net anomaly, Figure 1 plots the growth of $1 invested in both portfolios. The four panels of the figure compare this compound growth across our four different definitions of the long/short portfolios. In panel (a), which focuses on the most narrow definition (i.e., the long-short portfolio is formed on stocks in the top and bottom 1% of loan fee), we see that over the sample period, a dollar invested in the loan fee anomaly grows to $18.76, whereas a dollar invested in the GHZ Net anomaly grows to only $1.47. In the other three panels, the results are similar. As we repeat the analysis with the alternate bin methodologies (i.e., top and bottom 2%, top and bottom 5%, and top and bottom 10%), the performance difference shrinks, but in every case, the loan fee anomaly generates significantly higher profits than the GHZ Net anomaly variable.

In Table 2, we formally compare the loan fee anomaly to the GHZ set of anomalies. The table displays the average one-month return (shown in Panel A) and the percentage of 1-month returns that are positive (shown in Panel B), as well as the Sharpe Ratio (shown in Panel C). For *Loan Fee*, the table displays the average value ("Value") and the rank across all 102 anomalies ("Rank"). The last column provides statistics for the distribution of returns across all 102 of the GHZ anomalies (it shows the mean return, the 95th percentile, and the maximum return).

Across all four bin methodologies, it is clear that the loan fee anomaly is a top performer. In Panel A, the long-short loan fee anomaly is ranked first for average

Figure 1. (Color online) Growth of $1 Invested in Anomaly Portfolios



Notes. The figure plots the growth of $1 invested in the loan fee anomaly portfolio (black line) and the GHZ Net anomaly portfolio (orange line). The GHZ Net anomaly variable aggregates the signals for all of the anomalies compiled by Green et al. (2017). Panel (a) displays results when portfolios are formed by sorting stocks on the top and bottom 1% of anomaly variables, while panels (b), (c), and (d) display results when portfolios are formed by sorting stocks on the top and bottom 2%, 5%, and 10%, respectively.

one-month return, percentage of one-month returns that are positive, and average monthly Sharpe Ratio for the top and bottom 1% and 2% portfolios. Across the four different percentile definitions of the long/short portfolios, we see that the average return of loan fee ranges between 4.01% and 0.98%, whereas the average for the GHZ set of anomalies ranges between 0.17% and 0.06%, and the maximum return across those anomalies ranges between 2.60% and 1.15%. We also calculate the skewness and kurtosis of the loan fee anomaly relative to the GHZ set of anomalies, which is reported in Table A2 of the Online Appendix.

Instead of analyzing the anomalies individually, there is a growing literature that explores potential optimal combinations of them (e.g., Gu et al. 2020) using advanced statistical techniques to maximize the realized returns of composite anomaly signals. In a sense, loan fee is an alternative combination of market signals determined by the revealed preference of arbitrageurs. With this in mind, Table 3 compares loan fee to some of these more advanced aggregation methods. The first combination method is the simplest, *GHZ Net*, which weights each of the anomalies equally in its portfolio formation. The remaining five methods, Ordinary Least Squares, Principal Components Regression, Partial Least Squares, Elastic Net, and Group Lasso with a Generalized Linear Model, follow the estimation methodology in Gu et al. (2020).[9]

In this comparison, we show the returns, percentage of positive returns, and Sharpe Ratios for the long-short portfolio returns, as well as the rank of each combination method among the seven different methods (loan fee and the six combination methods). We find that loan fee almost always dominates the ranking of the other combination methods, with the return of the loan fee anomaly almost always larger—often much larger—than the other combination methods. For example, looking at the top and bottom 1% portfolios, we find that the loan fee's return is 4.01%, whereas the maximum of the other returns, GLM, is smaller at 1.95%. Similarly, in the rankings of combination methods, loan fee almost always comes in as number one. This broad finding is shared across all three panels of Table 3, where we compare returns, percentage of returns that are positive, and Sharpe Ratios, respectively.

The performance of loan fee raises an interesting question: to what extent does loan fee's dominant performance come from a more effective combination of

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Table 2. Performance of Long/Short Portfolio for Loan Fee and GHZ Anomalies

	Loan Fee		GHZ set of anomalies		
Anomaly definition	Value	Rank	Mean	95th %	Maximum
Panel A: Average of 1-month return					
Top/bottom 1%	4.01%	1	0.17%	1.46%	2.60%
Top/bottom 2%	2.75%	1	0.10%	1.14%	1.82%
Top/bottom 5%	1.56%	2	0.06%	0.89%	1.77%
Top/bottom 10%	0.98%	4	0.06%	0.77%	1.15%
Panel B: Percentage of 1-month returns that are positive					
Top/bottom 1%	74.4%	1	50.6%	59.0%	65.4%
Top/bottom 2%	69.2%	1	50.0%	60.3%	65.4%
Top/bottom 5%	62.8%	1	50.3%	60.3%	61.5%
Top/bottom 10%	62.8%	4	49.8%	60.3%	67.9%
Panel C: Monthly Sharpe Ratio					
Top/bottom 1%	0.66	1	0.02	0.22	0.32
Top/bottom 2%	0.58	1	0.02	0.20	0.29
Top/bottom 5%	0.40	1	0.01	0.22	0.32
Top/bottom 10%	0.30	3	0.02	0.22	0.32

Notes. The table reports the performance of the *Loan Fee* anomaly relative to the GHZ anomalies. For each anomaly, we construct a long/short portfolio, and the performance of each anomaly is evaluated using three metrics: the average one-month return (Panel A), the percentage of one-month returns that are positive (Panel B), and the monthly Sharpe Ratio (Panel C). For each of these, we also rank each anomaly variable relative to the entire set of anomalies. *Loan Fee* is measured by the average retail loan rate on the last trading day of the month. The GHZ set of anomalies is the set of 102 anomalies studied in Green et al. (2017). In each panel, we present results when portfolios are formed by sorting stocks on the top and bottom 1% of anomaly variables ("Top/Bottom 1%"), as well as when portfolios are formed by sorting stocks on the top and bottom 2%, 5%, and 10%.

existing anomalies versus information outside of existing anomalies? Although we explore this question in detail below in Section 4.3, we first explore in finer detail the relation between loan fees and existing anomalies.

In Table 4, we report the top 20 anomalies by long-short performance across our four long-short portfolio percentile definitions, including not only the average long-short performance, but also the performance of the long and short legs and the short-leg rank separately. Separating the performance of the two legs shows that the outperformance of the loan fee anomaly is driven by the short side of the anomaly. The performance of the short leg ranks in the top 10 across all four definitions, and the performance of the long leg appears to be average compared with other anomalies in the top 20. Examining the list of other anomalies across the four definitions does not identify a clear set of "superior" anomalies. Comparing the top 20 lists for all four long-short portfolio definitions, 45 different anomalies appear, and seven anomalies are consistently ranked in the top 20 across all four definitions.[10]

Table 5 and Figures 2 and 3 examine the performance of the loan fee anomaly over time. In Table 5, the average 12-month rolling performance rank of the long-short loan fee portfolio relative to the rest of the GHZ anomaly set is reported. The loan fee anomaly is consistently in the top 20 and remains first over a number of multimonth periods. Figure 2 plots the rolling 12-month average performance of the long-short *Loan Fee* and *GHZ Net* portfolios, and long and short legs of those portfolios are depicted separately in Figure 3. Figure 2 confirms this overall outperformance of the *Loan Fee* anomaly relative to the aggregation of all GHZ anomalies via the *GHZ Net* portfolio. Across the four different long/short percentile definitions, *Loan Fee* outperforms between 74% and 94% of months.

Table 5 shows one time period when the *Loan Fee* anomaly exhibits closer to average performance: late 2018/early 2019. The late 2018/early 2019 periods coincided with large negative market returns and relatively large increases in volatility. Figures 2 shows that for all three periods, the performance of both the *Loan Fee* and the *GHZ Net* portfolios declined and generated near zero returns. Additionally, Figure 3 and Table 4 show that this outperformance is generated by the short leg of the strategy. In contrast to the comparison with individual GHZ anomalies, Figure 2 shows that the *Loan Fee* long-short portfolio outperformed the *GHZ Net* portfolio in 2018 and 2019 over three of the long-short percentile definitions.

Figure 4 depicts the persistence of the *Loan Fee* anomaly as compared with the set of GHZ anomalies. To construct these figures, we calculate the average return for each anomaly over the previous two years and then form deciles based on this past performance. We then evaluate the performance of each decile over the following year. We also show the *Loan Fee* anomaly by itself.

Comparing the average GHZ anomaly which lands in the top decile to the *Loan Fee* anomaly in the top decile, the returns going forward are very different. Whereas the average top-decile GHZ anomaly mean reverts and, after a year, is statistically indistinguishable from the performance of the average bottom-decile anomaly, the *Loan Fee* anomaly persists in its superior performance, outperforming the average GHZ anomaly across all deciles.[11] In other words, the high-performing GHZ anomaly today is unlikely to be a high-performing GHZ anomaly in the following year, but the *Loan Fee* anomaly is consistently a top performer.

Another way of looking at this idea is in Figure 5, where we examine anomaly performance today conditional on being in the top decile of performers over the prior two years. In other words, we take the top-performing decile of anomalies over the prior two years, and we examine the likelihood of *remaining* in the top decile over the next year. The top left panel (Top/Bottom 1%) says that among the GHZ anomalies, a top-decile performer over the past two years has an 11% chance of being a top-decile performer over the next one year and a 17% chance of being a bottom-decile

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Table 3. Performance of Loan Fee Anomaly vs. Alternate Signal Combination Methods

	Loan Fee		*GHZ Net*		OLS		PCR		PLS		ENET		GLM	
Anomaly definition	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank
Panel A: Average of 1-month return														
Top/bottom 1%	4.01%	1	0.70%	5	0.35%	6	0.12%	7	0.98%	3	0.77%	4	1.95%	2
Top/bottom 2%	2.75%	1	0.82%	4	0.67%	5	0.33%	6	1.27%	3	0.04%	7	1.51%	2
Top/bottom 5%	1.56%	1	0.51%	5	0.69%	4	0.34%	6	0.78%	3	0.30%	7	1.08%	2
Top/bottom 10%	0.98%	1	0.30%	5	0.49%	3	0.21%	7	0.43%	4	0.25%	6	0.76%	2
Panel B: Percentage of 1-month returns that are positive														
Top/bottom 1%	74.4%	1	50.0%	7	52.6%	5	51.3%	6	59.0%	3	55.1%	4	61.5%	2
Top/bottom 2%	69.2%	1	59.0%	3.5	53.8%	5.5	51.3%	7	59.0%	3.5	53.8%	5.5	67.9%	2
Top/bottom 5%	62.8%	1	56.4%	4	59.0%	3	48.7%	7	51.3%	6	53.8%	5	60.3%	2
Top/bottom 10%	62.8%	1	55.1%	4	60.3%	3	51.3%	6	46.2%	7	52.6%	5	61.5%	2
Panel C: Monthly Sharpe Ratio														
Top/bottom 1%	0.66	1	0.11	5	0.07	6	0.02	7	0.15	3	0.13	4	0.33	2
Top/bottom 2%	0.58	1	0.14	5	0.17	4	0.05	6	0.21	3	0.01	7	0.30	2
Top/bottom 5%	0.40	1	0.11	5	0.21	3	0.06	7	0.14	4	0.08	6	0.25	2
Top/bottom 10%	0.30	1	0.08	6	0.19	3	0.05	7	0.09	4	0.08	5	0.20	2

Notes. The table reports the performance of the loan fee anomaly relative to alternate methods for combining the GHZ anomalies. For each anomaly variable, we construct a long/short portfolio, and the performance of each anomaly variable is evaluated using three metrics: the average one-month return (Panel A), the percentage of one-month returns that are positive (Panel B), and the monthly Sharpe Ratio (Panel C). For each of these, we also rank each anomaly variable relative to the others. *Loan Fee* is measured by the average retail loan rate on the last trading day of the month. The *GHZ Net* anomaly variable combines the signals for all of the anomalies compiled by Green et al. (2017); see Section 2 of the text for a detailed description of *GHZ Net*. The remaining five variables combine the GHZ signals using five different estimation methods from Gu et al. (2020): Ordinary Least Squares, Principal Components Regression, Partial Least Squares, Elastic Net, and Group Lasso with a Generalized Linear Model. See Section 4 of the text for a detailed explanation of how each of these estimation methods are implemented. In each panel, we present results when portfolios are formed by sorting stocks on the top and bottom 1% of anomaly variables ("Top/Bottom 1%"), as well as when portfolios are formed by sorting stocks on the top and bottom 2%, 5%, and 10%.

performer. Looking across all the decile bins in the four figures, the distribution is relatively flat, suggesting that conditioning on relatively strong past performance has little predictability for future performance among the GHZ anomalies.

On the other hand, when looking at the distribution of the *Loan Fee* anomaly into future performance deciles, we see a striking difference: the *Loan Fee* anomaly has an 81% chance of remaining in the top decile. Furthermore, there is a steep decline in the probability of moving into lower deciles, and there is a 0% chance of moving into the bottom half of the decile bins. Overall, this contrasts strongly with the top-performing GHZ anomalies, indicating that there is significantly more performance persistence in the *Loan Fee* anomaly than the average top-performing GHZ anomaly.

4.2. Alternative Short-Selling Predictors

The best-ideas concept began in the mutual fund literature, which proposed measures based on portfolios and concentration to capture fund managers' best ideas. There are a number of possible analogues in the short-selling literature. A natural choice is short interest, which measures the percentage of a stock's shares outstanding that is held short. When short interest is high, it is likely that many short sellers have large positions betting against the stock of interest.

Both of these approaches are quantity-based measures of best ideas, with the motivating assumption that a manager holds more of a stock that she likes. An alternative would be a price-based measure—that is, the more a manager pays to hold a position, the more she likes it. The natural price-based measure that exists on the short side is loan fee.

In Table 6, we examine the predictive power of both price- and quantity-based measures. Specifically, we run OLS panel regressions of one-month-ahead returns on *Loan Fee* (column (1)) and *Short Interest* (column (2)). We include time fixed effects in all models, and *t*-statistics calculated using Driscoll-Kraay standard errors are shown below the estimates.[12]

Two important facts emerge. First, when comparing columns (1) and (2), it is clear that *Loan Fee* is a stronger predictor of future returns than *Short Interest*, with a one-standard-deviation increase in *Loan Fee* corresponding to a 0.36% decrease in future returns (the analogous decrease in future returns for *Short Interest* is 0.22%). Second, in a multivariate, horse-race regression (column (3)), we find that *Loan Fee* reduces the importance of *Short Interest* (which becomes insignificant), whereas *Loan Fee* remains strongly statistically significant. Given these results, when we compare price- and quantity-based measures, we find that the price-based measure, *Loan Fee*, is a stronger predictor of future returns. In

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Table 4. Top 20 Anomalies by Long/Short Portfolio Performance

Rank	Anomaly	Long Leg (%)	Short Leg (%)	Port. Ret. (%)	Short Rank	Rank	Anomaly	Long Leg (%)	Short Leg (%)	Port. Ret. (%)	Short Rank
	Panel A: Top/bottom 1%						Panel B: Top/bottom 2%				
1	*Loan Fee*	1.3	−2.7	4.0	2	1	*Loan Fee*	1.2	−1.6	2.8	2
2	*Sales to cash*	0.4	−2.2	2.6	3	2	*Sales to cash*	0.7	−1.2	1.8	5
3	*Sales to price*	0.2	−2.1	2.3	4	3	*Sales to price*	0.5	−1.3	1.7	3
4	*Idiosyncratic return volatility*	1.1	−0.6	1.7	14	4	*Idiosyncratic return volatility*	1.1	−0.3	1.4	11
5	*Absolute accruals*	0.8	−0.9	1.7	9	5	*Industry adjusted book to market*	0.3	−1.1	1.4	6
6	*Accrual volatility*	0.8	−0.8	1.6	10	6	*Dispersion in forecasted EPS*	1.3	0.2	1.1	31
7	*Debt capacity/firm tangibility*	1.6	0.1	1.5	41	7	*Financial statement score (MS)*	1.3	0.2	1.1	27
8	*Dispersion in forecasted EPS*	1.3	0.0	1.3	31	8	*Industry-adjusted cash-flow-to-price ratio*	0.2	−0.9	1.1	7
9	*Abnormal earnings announcement volume*	1.0	−0.3	1.3	20	9	*Absolute accruals*	0.9	−0.2	1.0	14
10	*R&D to market capitalization*	1.4	0.1	1.2	42	10	*Maximum daily return*	1.0	0.0	1.0	16
11	*Financial statement score (PS)*	0.0	−1.3	1.2	6	11	*Debt capacity/firm tangibility*	1.3	0.3	0.9	41
12	*% change in current ratio*	0.3	−0.8	1.1	11	12	*Cash holdings*	1.0	0.1	0.9	22
13	*Financial statement score (MS)*	1.3	0.2	1.1	43	13	*R&D to market capitalization*	1.0	0.1	0.8	23
14	*Earning announcement return*	1.5	0.3	1.1	49	14	*% change in sales - % change in A/R*	0.6	−0.2	0.8	13
15	*% change in sales - % change in A/R*	0.7	−0.4	1.1	18	15	*Sales growth*	0.9	0.2	0.8	24
16	*Employee growth rate*	0.6	−0.4	1.0	16	16	*Tax income to book income*	1.2	0.4	0.7	48
17	*Industry-adjusted change in profit margin*	1.4	0.4	1.0	55	17	*Bid-ask spread*	0.9	0.2	0.7	30
18	*% change in quick ratio*	0.3	−0.7	1.0	13	18	*Return volatility*	0.9	0.3	0.7	36
19	*Industry adjusted book to market*	−0.2	−1.1	1.0	7	19	*Earning announcement return*	0.9	0.3	0.6	35
20	*Cash flow volatility*	0.7	−0.2	0.9	24	20	*Abnormal earnings announcement volume*	0.8	0.2	0.6	25
	Panel C: Top/bottom 5%						Panel D: Top/bottom 10%				
1	*Industry adjusted book to market*	0.8	−1.0	1.8	1	1	*R&D to market capitalization*	1.3	0.1	1.2	5
2	*Loan Fee*	1.2	−0.4	1.6	3	2	*Industry adjusted book to market*	0.9	−0.1	1.1	1
3	*Industry-adjusted cash-flow-to-price ratio*	0.6	−0.8	1.4	2	3	*Dispersion in forecasted EPS*	1.2	0.1	1.0	6
4	*Dispersion in forecasted EPS*	1.3	0.3	1.0	18	4	*Loan Fee*	1.1	0.2	1.0	7
5	*R&D to market capitalization*	1.2	0.1	1.0	10	5	*Industry-adjusted cash-flow-to-price ratio*	0.7	−0.1	0.8	2
6	*Idiosyncratic return volatility*	1.1	0.2	0.9	14	6	*Cash flow volatility*	0.9	0.1	0.8	4
7	*Financial statement score (MS)*	1.3	0.4	0.9	27	7	*Bid-ask spread*	1.1	0.3	0.8	14
8	*Absolute accruals*	0.9	0.3	0.7	15	8	*Idiosyncratic return volatility*	1.1	0.4	0.7	16
9	*Bid-ask spread*	1.0	0.4	0.6	25	9	*Financial statement score (MS)*	1.2	0.4	0.7	21
10	*Cash holdings*	1.0	0.4	0.6	28	10	*Accrual volatility*	0.9	0.3	0.7	9
11	*Maximum daily return*	1.0	0.4	0.6	31	11	*Absolute accruals*	1.0	0.4	0.5	20
12	*Volatility of liquidity (dollar trading volume)*	1.0	0.5	0.6	37	12	*Return volatility*	1.1	0.5	0.5	31
13	*Financial statement score (PS)*	0.7	0.2	0.6	13	13	*Cash holdings*	1.0	0.5	0.5	24
14	*Return volatility*	1.0	0.5	0.5	43	14	*Industry sales concentration*	1.2	0.6	0.5	51
15	*% change in sales - % change in A/R*	0.8	0.3	0.5	17	15	*% change in sales - % change in A/R*	0.8	0.3	0.5	12
16	*Volatility of liquidity (share turnover)*	1.0	0.5	0.5	38	16	*Maximum daily return*	1.0	0.5	0.5	28
17	*Industry sales concentration*	1.1	0.7	0.5	58	17	*6-month momentum*	0.8	0.3	0.5	13
18	*Real estate holdings*	0.5	0.0	0.5	8	18	*Sin stocks*	1.2	0.8	0.4	74
19	*Capital expenditures and inventory*	0.4	0.0	0.5	7	19	*Capital expenditures and inventory*	0.7	0.3	0.4	10
20	*Sin stocks*	1.2	0.8	0.4	76	20	*Unexpected quarterly earnings*	0.5	0.1	0.4	3

Notes. The table reports the return of the long/short portfolio and the return of each leg of the portfolio for the loan fee and the top 19 GHZ anomalies (so the top 20 overall anomalies are displayed). Anomalies are sorted from highest long/short return to lowest. "Short Rank" reports the rank of the return on the short leg of the portfolio. Because a short position is taken on the short leg of the portfolio, the lowest return has rank 1. Panel A displays results when portfolios are formed by sorting stocks on the top and bottom 1% of anomaly variables, while Panels B, C, and D display results when portfolios are formed by sorting stocks on the top and bottom 2%, 5%, and 10%.

Table 5. (Color online) The 12-Month Rolling Rank of *Loan Fee* Anomaly Among the GHZ Set of Anomalies

Anomaly definition	2014						2015												2016											
Top/bottom 1%	14	3	4	1	1	1	1	1	1	1	1	2	2	2	2	3	4	4	7	11	8	6	13	14	8	14	11	4	5	13
Top/bottom 2%	11	1	8	3	2	1	1	1	1	3	3	3	2	2	2	2	3	3	7	7	3	8	9	17	9	14	4	10	12	15
Top/bottom 5%	9	1	6	3	3	3	3	3	4	6	4	4	4	4	3	3	4	5	10	10	7	9	11	12	12	25	9	15	21	12
Top/bottom 10%	10	1	9	3	3	4	7	4	5	17	9	11	6	7	5	6	7	10	17	14	13	12	13	20	18	41	15	24	33	29

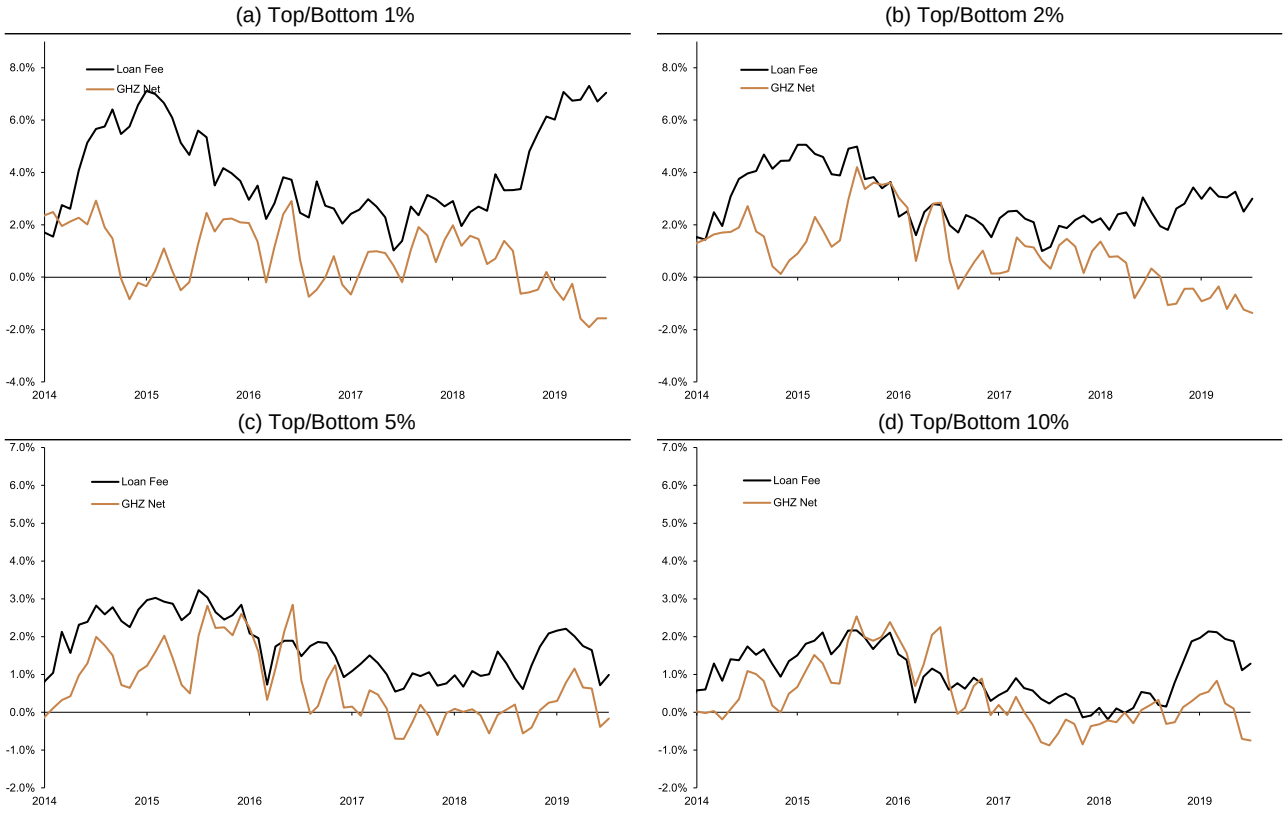
	2017												2018												2019											
Top/bottom 1%	11	3	3	5	10	7	3	3	3	12	27	19	5	5	3	3	4	4	11	5	6	5	3	4	3	3	2	1	1	2	1	1	1	1	1	1
Top/bottom 2%	16	3	5	6	15	4	3	3	3	7	26	25	6	6	3	6	6	4	10	5	6	5	3	5	12	9	3	7	3	8	6	7	5	3	5	6
Top/bottom 5%	8	3	2	5	16	11	11	7	6	16	27	27	15	10	9	16	12	9	25	10	15	11	8	13	25	31	15	11	6	6	7	12	9	5	17	16
Top/bottom 10%	18	18	8	11	30	22	24	12	20	19	28	34	28	24	23	55	55	46	66	43	57	48	29	32	48	42	19	14	6	6	7	7	6	2	8	4

Notes. The table reports the rolling rank of the average long/short return for the loan fee anomaly among the GHZ anomalies. The average long-short return is calculated over the year prior to the date that appears in the table. We present results when portfolios are formed by sorting stocks on the top and bottom 1% of anomaly variables ("Top/Bottom 1%"), as well as when portfolios are formed by sorting stocks on the top and bottom 2%, 5%, and 10%.

other words, loan fees more successfully capture short sellers' best ideas.

Of course, the literature also examines a number of other measures that could proxy for a short seller's best ideas. These include *Days-to-Cover* (Hong et al. 2021), *SIR* × *RetVol* (a measure of arbitrage risk, à la Hanson and Sunderam 2014), and *CME Beta* (Drechsler and Drechsler 2016).[13] In columns (4), (5), and (6), we examine these additional measures not only individually, but also in a horse-race regression with *Loan Fee*. Individually, each of the measures predicts future returns: the corresponding estimates are −0.13%, −0.03%, and 0.16% for *Days-to-Cover*, *SIR* × *RetVol*, and *CME Beta*, respectively. However, considering these other short-

Figure 2. (Color online) Rolling 12-Month Average of Anomaly Returns

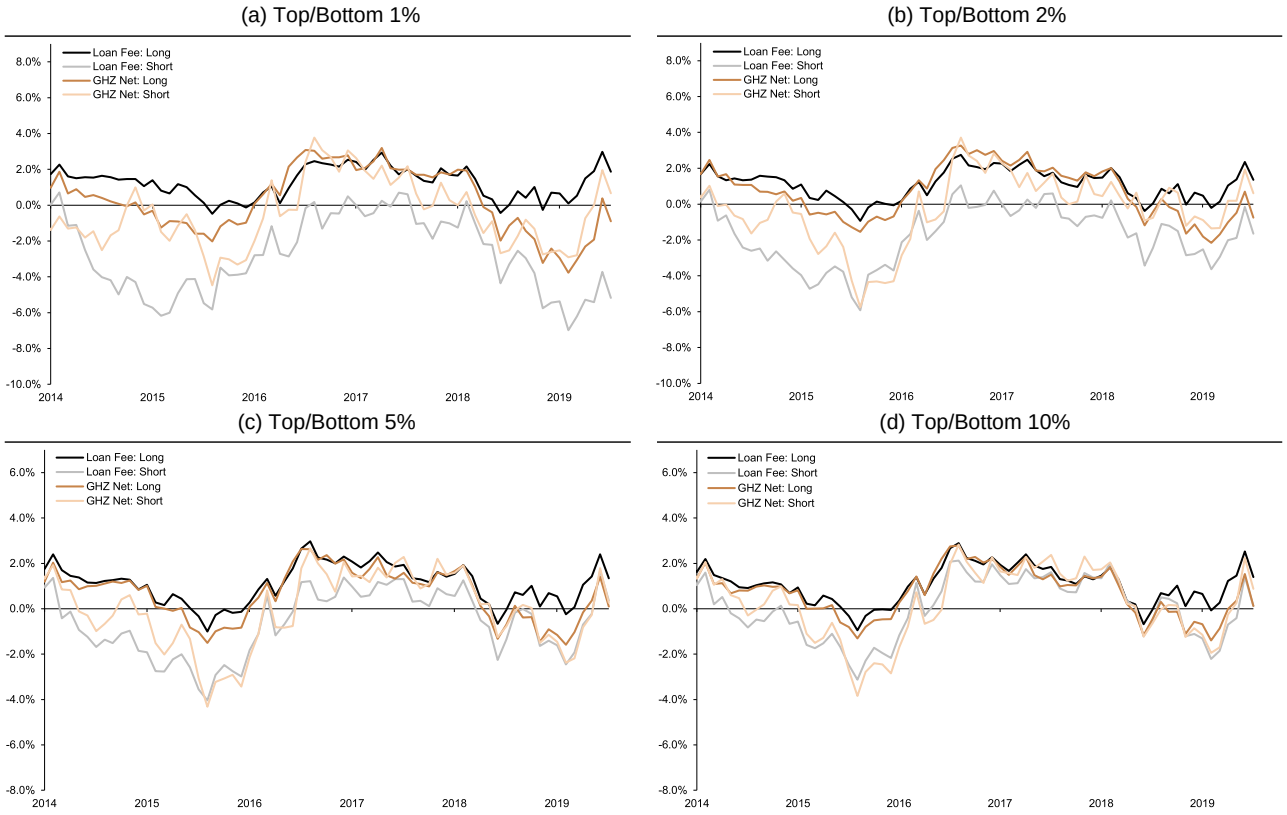


Notes. The figure plots the rolling 12-month mean performance of the loan fee anomaly portfolio (black line) and the GHZ Net anomaly portfolio (orange line) over time. The GHZ Net anomaly variable aggregates the signals for all of the anomalies compiled by Green et al. (2017). Panel (a) displays results when portfolios are formed by sorting stocks on the top and bottom 1% of anomaly variables, while panels (b), (c), and (d) display results when portfolios are formed by sorting stocks on the top and bottom 2%, 5%, and 10%, respectively.

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Figure 3. (Color online) Rolling 12-Month Average of Anomaly Long and Short Returns



Notes. The figure plots the rolling 12-month mean performance of the loan fee anomaly portfolio (black line and gray line) and the GHZ Net anomaly portfolio (orange line and light orange line) over time, and it breaks out the performance by long- and short-legs of the portfolios. The GHZ Net anomaly variable aggregates the signals for all of the anomalies compiled by Green et al. (2017). Panel (a) displays results when portfolios are formed by sorting stocks on the top and bottom 1% of anomaly variables, while panels (b), (c), and (d) display results when portfolios are formed by sorting stocks on the top and bottom 2%, 5%, and 10%, respectively.

selling variables does not change our conclusion: *Loan Fee* remains the best predictor of future returns and drives out the predictability of the alternatives.[14] Overall, loan fees capture short sellers' best ideas.

In Table 7, we repeat the analysis of Table 2 in order to compare the long-short returns of the *Loan Fee* anomaly with the four short-selling related variables (i.e., *Days-to-Cover*, *Short Interest Ratio*, *SIR* × *RetVol*, and *CME Beta*). All four of the short-selling-related variables perform well. For the Top/Bottom 10% portfolio, the long-short return for each of the short-selling-related variables is ranked no worse than 16th among the GHZ anomalies. However, the *Loan Fee* anomaly performs even better, outperforming the Top/Bottom 10% long-short return for each of the four variables by at least 0.24%. The *Loan Fee* anomaly also delivers a higher monthly Sharpe Ratio more than that of each of the four short-selling-related variables, regardless of whether the Top/Bottom 1%, 2%, 5%, or 10% is used to form the long-short portfolios.

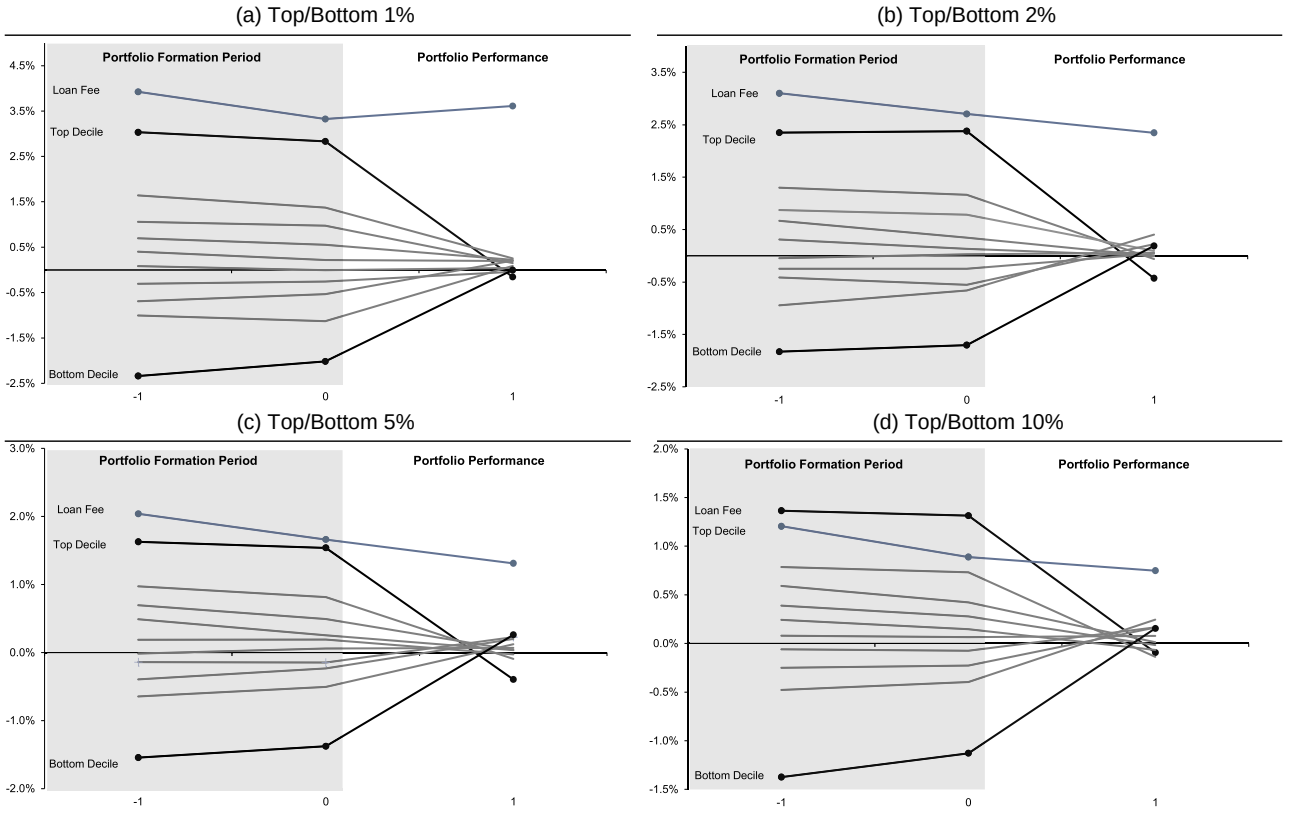
4.3. The *Loan Fee* Anomaly and Unique Information

Although the results of the two previous sections show that loan fee is a strong predictor of cross-sectional returns, other papers have made a similar point (i.e., Stambaugh et al. 2012, Drechsler and Drechsler 2016). These papers focus on whether loan fee and other short-selling constraints explain previously discovered anomalies. Unlike these papers, we ask to what extent the loan fee anomaly captures information already contained in existing anomalies versus capturing information outside of those anomalies. As a first step in addressing this question, we identify the short portfolio in each period for both the *Loan Fee* and the *GHZ Net* anomalies across our four cutoffs (the bottom 1%, 2%, 5%, and 10%). We then calculate the average performance of those stocks uniquely identified by each anomaly (i.e., "*Only Loan Fee*" and "*Only GHZ Net*") and the overlapping portfolio ("*Both*"). The results are plotted in Figure 6.

Looking first at the 1% short portfolio, we see there is only 6% overlap in the stocks identified by the *Loan Fee* and *GHZ Net* anomalies. As we increase the cutoff, we find that the overlap grows monotonically, maxing out at a 17% stock overlap for the decile cutoff. Looking at the performance, those stocks included in the short portfolios of both the *Loan Fee* and *GHZ Net* anomalies exhibit the greatest underperformance. Comparing the

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Figure 4. (Color online) Anomaly Performance Persistence



Notes. The figure displays the persistence of the loan fee anomaly relative to the set of GHZ anomalies. Each date, we look back two years and take the average return for each anomaly, and we form deciles based on these past performance levels. We then plot the mean return of each decile over the next year (black lines). We also plot the mean return for the loan fee anomaly (blue line). Panel (a) displays results when portfolios are formed by sorting stocks on the top and bottom 1% of anomaly variables, while panels (b), (c), and (d) display results when portfolios are formed by sorting stocks on the top and bottom 2%, 5%, and 10%, respectively.

Table 6. Panel Regression of One-Month-Ahead Returns on Short-Selling Variables

	Dependent variable: One-month ahead monthly returns						
Explanatory variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Loan Fee	−0.36%		−0.32%				−0.32%
	−4.52***		−5.09***				−4.62***
Short Interest Ratio		−0.22%	−0.10%		−0.20%		−0.11%
		−2.04**	−1.06		−2.00**		−1.01
Days-to-Cover				−0.13%			0.02%
				−2.67***			0.26
SIR × RetVol					−0.03%		0.02%
					−0.57		0.52
CME Beta						0.16%	0.07%
						1.44	0.73
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	191,416	191,416	191,416	191,416	191,416	191,416	191,416
R^2	14.251%	14.191%	14.258%	14.169%	14.192%	14.174%	14.261%

Notes. The table shows results from OLS panel regressions of one-month-ahead monthly returns on standardized values of *Loan Fee*, *Short Interest Ratio*, *Days-to-Cover*, *SIR × RetVol*, and *CME Beta* with time fixed effects. *Days-to-Cover* is short interest divided by average daily trading volume over the previous month. *Short Interest Ratio* (SIR) is short interest divided by market capitalization. SIR × *RetVol* is *Short Interest Ratio* times the daily return volatility over the previous month. *CME Beta* is the estimated beta on the cheap minus expensive factor from Drechsler and Drechsler (2016). The cheap minus expensive portfolio is constructed by going long stocks in the "Bottom 10%" portfolio and short stocks in the "Top 10%" portfolio for the *Loan Fee* anomaly. We proxy for a stock's beta with the CME factor by estimating its beta with the stocks in the CME portfolio over the 36 months prior to portfolio construction. Specifically, to estimate the beta for stock i in month t, we (1) identify the stocks in each leg of the CME portfolio in month t; (2) fix the set of stocks in the CME portfolio and calculate the long-short returns of the portfolio over the prior 36 months; and (3) regress the returns of stock i on the three-factor Fama-French model plus CME factor from (2). t-statistics calculated using Driscoll-Kraay standard errors are shown below the coefficient estimates.

*Statistical significance at the 10% level; **Statistical significance at the 5% level; ***statistical significance at the 1% level.

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Table 7. Performance of Long/Short Portfolio for Short-Selling-Related and GHZ Set of Anomalies

	Loan Fee		*Days to Cover*		*SIR*		*SIR × Ret. Vol.*		*CME Beta*		GHZ anomalies		
Anomaly definition	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Mean	95th %	Max.
Panel A: Average of one-month return													
Top/bottom 1%	4.01%	1	0.46%	41	1.27%	9	1.44%	7	0.43%	43	0.17%	1.46%	2.60%
Top/bottom 2%	2.75%	1	0.74%	15	1.26%	5	1.33%	5	0.80%	14	0.10%	1.14%	1.82%
Top/bottom 5%	1.56%	2	0.65%	8	0.79%	7	0.86%	7	0.52%	13	0.06%	0.89%	1.77%
Top/bottom 10%	0.98%	4	0.46%	16	0.58%	10	0.74%	9	0.53%	13	0.06%	0.77%	1.15%
Panel B: Percentage of one-month returns that are positive													
Top/bottom 1%	74.4%	1	60.3%	4	56.4%	22	57.7%	13.5	53.8%	37	50.6%	59.0%	65.4%
Top/bottom 2%	69.2%	1	64.1%	2	56.4%	16.5	56.4%	16.5	53.8%	32.5	50.0%	60.3%	65.4%
Top/bottom 5%	62.8%	1	64.1%	1	51.3%	47.5	56.4%	20	55.1%	25.5	50.3%	60.3%	61.5%
Top/bottom 10%	62.8%	4	60.3%	6	52.6%	37.5	51.3%	43	55.1%	22	49.8%	60.3%	67.9%
Panel C: Monthly Sharpe Ratio													
Top/bottom 1%	0.66	1	0.11	30	0.20	10	0.16	19	0.05	49	0.02	0.22	0.32
Top/bottom 2%	0.58	1	0.22	5	0.25	4	0.18	10	0.12	23	0.02	0.20	0.29
Top/bottom 5%	0.40	1	0.26	4	0.19	8	0.15	12	0.10	22	0.01	0.22	0.32
Top/bottom 10%	0.30	3	0.21	7	0.16	10	0.15	14	0.14	16	0.02	0.22	0.32

Notes. The table reports the performance of the *Loan Fee* anomaly relative to four other short-selling-related anomalies and the GHZ anomalies. For each anomaly, we construct a long/short portfolio and calculate its return. The performance of each anomaly is evaluated using three metrics: the average one-month returns (Panel A), the percentage of one-month returns that are positive (Panel B), and the monthly Sharpe Ratio (Panel C). For each of these, we also rank each anomaly variable relative to the entire set of anomalies. *Loan Fee* is measured by the average retail loan rate on the last trading day of the month. *Days-to-Cover* is short interest divided by average daily trading volume over the previous month. *Short Interest Ratio* (*SIR*) is short interest divided by market capitalization. *SIR* × *RetVol* is *Short Interest Ratio* times the daily return volatility over the previous month. *CME Beta* is the estimated beta on the cheap minus expensive factor from Drechsler and Drechsler (2016). The cheap minus expensive portfolio is constructed by going long stocks in the "Bottom 10%" portfolio and short stocks in the "Top 10%" portfolio for the *Loan Fee* anomaly. We proxy for a stock's beta with the CME factor by estimating its beta with the stocks in the CME portfolio over the 36 months prior to portfolio construction. Specifically, to estimate the beta for stock i in month t, we (1) identify the stocks in each leg of the CME portfolio in month t; (2) fix the set of stocks in the CME portfolio and calculate the long-short returns of the portfolio over the prior 36 months; and (3) regress the returns of stock i on the three-factor Fama-French model plus CME factor from (2). The GHZ anomalies are the 102 anomalies studied in Green et al. (2017). In each panel, we present results when portfolios are formed by sorting stocks on the top and bottom 1% of anomaly variables ("Top/Bottom 1%"), as well as when portfolios are formed by sorting stocks on the top and bottom 2%, 5%, and 10%.

stocks uniquely identified by both anomalies, we see those included in the *Only Loan Fee* short portfolio underperform the *Only GHZ Net* short portfolio by between 2.61% and 0.44% monthly, depending on the cutoff.

This set analysis is suggestive that loan fees contain information unique to short sellers. To better understand both the information contained in loan fees (that represents a selected reweighting of existing anomalies) versus information that is potentially unique to the loan fee anomaly, we begin by regressing *Loan Fee* on each anomaly variable separately and show the strongest and weakest coefficients in Table 8.

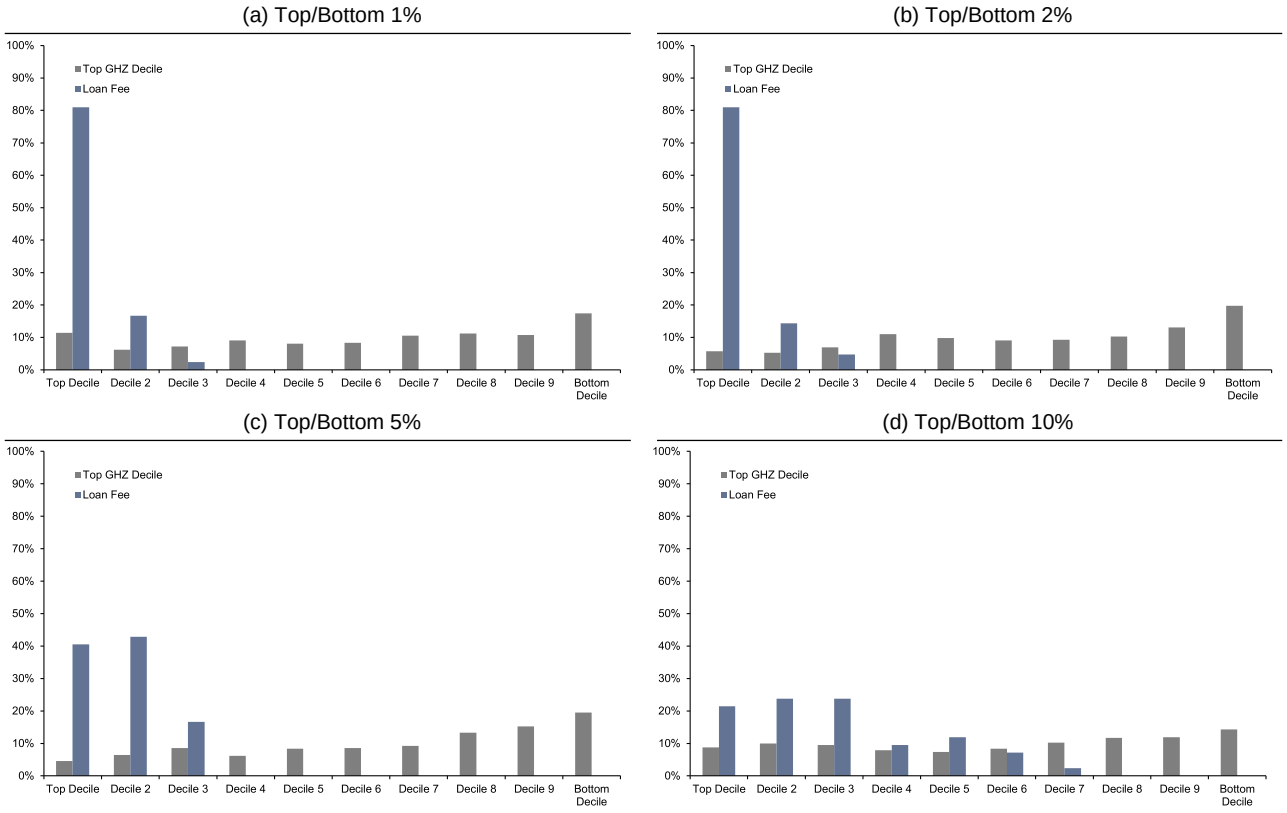
In the table, we calculate the long/short return for each anomaly using the direction indicated by the anomaly's correlation with *Loan Fee*. For example, idiosyncratic return volatility is positively correlated with *Loan Fee*. Because the *Loan Fee* anomaly goes long low loan fee stocks and short high loan fee stocks, we also go long low idiosyncratic return stocks and short high idiosyncratic return stocks, given *Loan Fee*'s positive correlation with idiosyncratic volatility. We call this the "loan fee indicated direction." As another example, *Return on Assets* (*ROA*) is negatively correlated with the *Loan Fee*, so we calculate the long/short return in the *Loan Fee* indicated direction for *ROA* by constructing a portfolio that is short low ROA stocks and long high ROA stocks.[15]

Table 8 shows that anomalies that have the strongest correlation with *Loan Fee* outperform those with the weakest correlation. For example, focusing on the last column of Panel A, we see that the anomalies with the strongest correlation have an average long/short return in the *Loan Fee* indicated direction of 0.4%, compared with an average return of 0.1% for the weakest correlated anomalies in Panel B and the overall average of 0.1% in Panel C. In other words, the anomaly variables that *Loan Fee* is highly correlated with perform better in long-short portfolios than the anomaly variables it has little correlation with. This is evidence that the *Loan Fee* anomaly represents selective exposure to the best-performing anomalies.

In order to quantify the contribution of this selective exposure to the *Loan Fee* anomaly, we next consider a multivariate setting. Whereas Table 8 examines the univariate relation between the *Loan Fee* anomaly and the GHZ set of anomalies, in Table 9, we project *Loan Fee* on all of the 102 GHZ anomalies at the same time in order

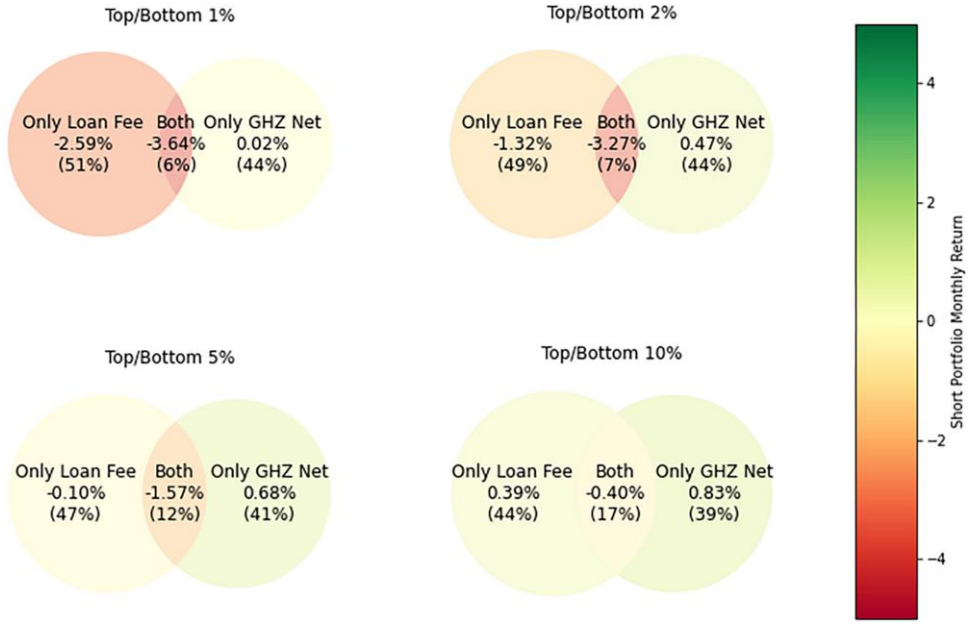
Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Figure 5. (Color online) GHZ Decile Portfolio Transition Probabilities



Notes. The figure examines anomaly performance today conditional on being in the top decile of performers over the prior four years. On each date, we take the top-performing decile of anomalies over the prior two years, and we examine the likelihood of remaining in the top decile over the next year. The figure displays probabilities for remaining in the top decile or transitioning to the second decile, third decile, etc. Panel (a) displays results when portfolios are formed by sorting stocks on the top and bottom 1% of anomaly variables, while panels (b), (c), and (d) display results when portfolios are formed by sorting stocks on the top and bottom 2%, 5%, and 10%, respectively.

Figure 6. (Color online) Set Analysis: Performance and Overlap of Loan Fee and GHZ Net Anomaly



Notes. The figure displays the performance of the unique components of the loan fee anomaly and the GHZ Net anomaly, as well as the performance when stocks are in both portfolios at the same time. We identify the short portfolio in each period for both the loan fee and the GHZ net anomalies across our four cutoffs (the bottom 1%, 2%, 5%, and 10%). We then calculate the average performance of those stocks uniquely identified by each anomaly (i.e., "Only Loan Fee" and "Only GHZ Net") and the overlapping portfolio ("Both").

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Table 8. Univariate Projection of *Loan Fee* on GHZ Set of Anomalies

				Long/Short return for portfolios from top/bottom (%):			
Rank	Anomaly	Coef.	*t*-stat	1%	2%	5%	10%
			Panel A: Anomalies that are strongly correlated with loan fee				
1	*Idiosyncratic return volatility*	0.9	25.9	1.7	1.4	0.9	0.7
2	*Bid-ask spread*	0.8	24.4	0.5	0.7	0.6	0.8
3	*Return volatility*	0.7	22.4	0.7	0.7	0.5	0.5
4	*Return on assets*	−0.7	−23.4	0.2	0.3	−0.1	0.0
5	*Earnings volatility*	0.6	26.1	1.4	1.3	0.7	0.6
6	*Scaled earnings forecast*	−0.6	−16.0	0.8	0.5	0.2	0.1
7	*Maximum daily return*	0.5	19.8	0.9	1.0	0.6	0.5
8	*Cash flow to debt*	−0.5	−25.2	0.8	0.4	0.2	0.3
9	*Earning to price*	−0.5	−20.6	0.2	0.6	0.3	0.2
10	*Size*	−0.5	−21.1	0.8	0.5	0.3	0.4
	Mean			0.8	0.7	0.4	0.4
	Median			0.8	0.6	0.4	0.5
			Panel B: Anomalies that are weakly correlated with loan fee				
1	*Industry sales concentration*	0.0	−0.9	−0.4	−0.6	−0.5	−0.5
2	*Sales to inventory*	0.0	1.3	−0.7	−0.6	−0.4	−0.3
3	*1-month momentum*	0.0	−0.4	−0.7	0.1	0.3	0.2
4	*Change in number of analysts*	0.0	−0.7	−0.6	−0.5	−0.2	0.0
5	*Sin stocks*	0.0	2.6	−0.4	−0.4	−0.4	−0.4
6	*Change in tax expense*	0.0	−2.0	−0.4	−0.4	−0.2	−0.1
7	*Change in 6-month momentum*	0.0	0.7	−0.5	−0.6	−0.3	−0.1
8	*Dividend initiation*	0.0	3.0	0.4	0.4	0.4	0.4
9	*Abnormal earnings announcement volume*	0.0	−2.4	1.3	0.6	0.3	0.1
10	*Secured debt indicator*	0.0	3.1	0.2	0.2	0.2	0.2
	Mean			−0.2	−0.2	−0.1	0.0
	Median			−0.4	−0.4	−0.2	0.0
			Panel C: Summary statistics for 102 GHZ anomalies				
	Mean			0.2	0.1	0.1	0.1
	Median			0.2	0.2	0.1	0.1

Notes. The table displays coefficient estimates and *t*-statistics for univariate regressions of *Loan Fee* on individual anomaly sorting variables. Panel A shows results for the anomalies that are most strongly related to *Loan Fee*, as given by the *t*-statistic on the univariate regression, while Panel B shows results for the anomalies that are most weakly related to *Loan Fee*, as given by the *t*-statistic on the univariate regression. Panel C displays results using all of the GHZ anomalies. In the four rightmost columns, we show long-short returns to each anomaly using our four different sorting cutoffs (1%, 2%, 5%, and 10%). At the bottom of each panel, "Mean" and "Median" display the mean and median return for each panel.

to construct a fitted loan fee and a residual loan fee. Specifically, we estimate an OLS panel regression with time fixed effects over the full sample with *Loan Fee* as the dependent variable and the standardized set of anomalies as the independent variables. The fitted value represents the component of *Loan Fee* that is spanned by existing anomalies, whereas the residual measures the component of *Loan Fee* that is unique. This decomposition of *Loan Fee* into its fitted and residual components is calculated in two different ways: (1) by projecting the loan fee on all 102 GHZ anomalies and (2) by using a principal components approach (PCA) and selecting the first 10 components to define the fitted component.[16]

Table 9 shows that long/short portfolios formed based on either the fitted value or the residual perform well relative to the set of GHZ anomalies. Focusing on the Top/Bottom 1% portfolios, we find that the fitted loan fee constructed by projecting *Loan Fee* on all 102 GHZ anomalies (PCA analysis) has a long/short return of 2.15% (1.52%), and it is ranked third (sixth) when compared with the GHZ set, which confirms our prior conclusion that *Loan Fee* is selectively correlated with high-performing anomalies. The residual from the *Loan Fee* projection (PCA decomposition) has a long/short return of 1.64% (2.97%), and when we replace the fitted value with the residual in the comparative ranking with the GHZ set, we find that it is ranked sixth (first); in other words, the unique component of the *Loan Fee* anomaly outperforms all but six of the individual GHZ anomalies in the case of the projection residual, and all of the individual GHZ anomalies in the PCA residual. This indicates that short sellers have valuable information that is not contained in the existing set of anomalies.

To get a sense of the unique component's relative contribution to the *Loan Fee* anomaly, we turn to a regression setting. In Table 10, we regress one-month-ahead monthly returns on *Predicted Loan Fee* (columns (2) and (5)), *Residual Loan Fee* (columns (3) and (6)), and both of

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Table 9. Performance of the Unique vs. Common Information in *Loan Fee* Anomaly

			All 102 GHZ anomalies				First 10 principal components						
	Loan Fee		Fitted fee		Resid. fee		Fitted fee		Resid. fee		GHZ anomalies		
Anomaly definition	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Mean	95th %	Max.
Panel A: Average of one-month return													
Top/bottom 1%	4.01%	1	2.15%	3	1.64%	6	1.52%	6	2.97%	1	0.17%	1.46%	2.60%
Top/bottom 2%	2.75%	1	2.18%	1	2.02%	1	1.08%	8	2.67%	1	0.10%	1.14%	1.82%
Top/bottom 5%	1.56%	2	1.24%	3	0.77%	7	0.72%	7	1.17%	3	0.06%	0.89%	1.77%
Top/bottom 10%	0.98%	4	0.97%	4	0.28%	25	0.58%	10	0.36%	20	0.06%	0.77%	1.15%
Panel B: Percentage of one-month returns that are positive													
Top/bottom 1%	74.4%	1	64.1%	2.5	62.8%	3	53.8%	37	60.3%	4	50.6%	59.0%	65.4%
Top/bottom 2%	69.2%	1	61.5%	2.5	65.4%	1.5	55.1%	26	65.4%	1.5	50.0%	60.3%	65.4%
Top/bottom 5%	62.8%	1	55.1%	25.5	59.0%	10	52.6%	38.5	57.7%	14.5	50.3%	60.3%	61.5%
Top/bottom 10%	62.8%	4	53.8%	30	53.8%	30	47.4%	61.5	52.6%	37.5	49.8%	60.3%	67.9%
Panel C: Monthly Sharpe Ratio													
Top/bottom 1%	0.66	1	0.34	1	0.22	7	0.16	19	0.31	2	0.02	0.22	0.32
Top/bottom 2%	0.58	1	0.36	1	0.33	1	0.12	23	0.34	1	0.02	0.20	0.29
Top/bottom 5%	0.40	1	0.23	6	0.19	7	0.10	24	0.23	6	0.01	0.22	0.32
Top/bottom 10%	0.30	3	0.21	7	0.09	26	0.09	25	0.09	25	0.02	0.22	0.32

Notes. The table shows monthly long-short returns and corresponding ranks from portfolios formed on the *Loan Fee* anomaly, *GHZ Net*, and all GHZ anomalies, as well as the *Fitted Loan Fee* and *Residual Loan Fee*. For the "All 102 GHZ Anomalies" specification, the *Fitted Loan Fee* and *Residual Loan Fee* are constructed by projecting *Loan Fee* on all of the 102 GHZ anomalies at the same time. For the "Principal Components (10)" specification, the *Fitted Loan Fee* and *Residual Loan Fee* are constructed by first conducting a principal components analysis on the 102 GHZ anomalies and then projecting loan fee on the first 10 principal components. The number of principal components was selected based on an evaluation of the scree plot, which is shown in the Online Appendix. We then examine the performance of the fitted value ("*Fitted Loan Fee*") and residual ("*Residual Loan Fee*") from this regression. Panel A displays the average of one-month returns, Panel B displays the percentage of one-month returns that are positive, and Panel C displays the monthly Sharpe Ratio. In each panel, we present results when portfolios are formed by sorting stocks on the top and bottom 1% of anomaly variables ("Top/Bottom 1%"), as well as when portfolios are formed by sorting stocks on the top and bottom 2%, 5%, and 10%.

them simultaneously (columns (4) and (7)). When we look at the R^2s in these regressions from the projection decomposition, we find that the fitted component explains 58% of the total return predictability, whereas the unique component explains 42%. Using the decomposition from our PCA, the unique component explains 68%, whereas the fitted component only explains 32%.[17] These results confirm that a substantial portion of short seller's best ideas come from valuable information that is *not* contained in the existing set of anomalies.

The previous results suggest that there is substantial information in loan fees which is not captured via traditional anomaly variables. Indeed, the majority of return

Table 10. *Loan Fee* R^2 Decomposition

		All 102 GHZ anomalies			First 10 principal components		
Explanatory variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Loan Fee	−0.36%						
	−4.52						
Predicted Loan Fee		−0.58%		−0.58%	−0.52%		−0.52%
		−2.17		−2.18	−1.21		−1.21
Residual Loan Fee			−0.29%	−0.29%		−0.33%	−0.33%
			−3.37	−3.33		−3.90	−3.88
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	191,416	191,416	191,416	191,416	191,416	191,416	191,416
R^2	0.1087%	0.0717%	0.0511%	0.1228%	0.0361%	0.0768%	0.1129%
R^2 as percent of (4)	—	58%	42%	100%	—	—	—
R^2 as percent of (7)	—	—	—	—	32%	68%	100%

Notes. The table displays results from OLS panel regressions of one-month-ahead monthly returns on *Loan Fee* (column (1)), the *Predicted Loan Fee* (columns (2) and (5)) derived from the corresponding regression in Table 9, the *Residual Loan Fee* (columns (3) and (6)) derived from the corresponding regression in Table 9, and both of them simultaneously (columns (4) and (7)). *t*-statistics calculated using Driscoll-Kraay standard errors are shown below the coefficient estimates. In the third to last row of the table, we display the within R^2 for each regression. In the last two rows, we display the within R^2 for each model as a percent of the R^2 in column (4) or (7).

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

predictability resides in *Residual Loan Fee* rather than *Fitted Loan Fee*. A natural question arises: what kind of information is it?

Of course, it would be difficult to identify *all* of the information in *Residual Loan Fee* because it includes a huge range of possible sources of idiosyncratic information. However, we can see an example of what this idiosyncratic information might look like by examining a setting where short sellers are well informed. The prior literature shows that short sellers are skilled at detecting financial misconduct (e.g., Karpoff and Lou 2010, Fang et al. 2016), and to the extent that this kind of information is orthogonal to the predicted information, it can serve as an example of a source of this idiosyncratic information.[18]

In order to examine financial misconduct, we use a database of shareholder class-action lawsuits about securities fraud (i.e., 10b-5 lawsuits) that comes from ISS Securities Class Action Services. The database covers all years in our loan fee sample period, including 940 lawsuits filed against 737 companies. We observe the date the class period began, the date the class period ended, the date the class-action lawsuit was filed, and the date it was resolved.

Table 11 shows that a meaningful fraction of short sellers' best ideas are related to class-action lawsuits.[19] Row (1) of Table 11 reports the percentage of general collateral stocks that are in any stage of a class-action lawsuit, which we define as the period from when the class period starts to when the case is resolved. Stocks in the top 10% of loan fees are twice as likely as general collateral stocks to be in some phase of a class-action lawsuit. When we focus more narrowly on short sellers' very best ideas, the results are even stronger. We find that 23.5% of the highest loan fee stocks are in some phase of a class-action lawsuit, compared with only 6.2% of general collateral stocks.

We also find more association between *Residual Loan Fee* and these class-action lawsuits than *Fitted Loan Fee*. The two nonbold rows of Table 11 answer the question: among stocks that are in any phase of a class-action lawsuit that have a high loan fee, what fraction have a high fitted loan fee, and what fraction have a high residual loan fee? When we examine stocks in the top 10% of loan fees in any phase of a class-action lawsuit, we find little difference in the propensity for those stocks to have a fitted loan fee or residual loan fee in the top 10%. However, as we narrow in on stocks in the top 1% of loan fee in any phase of a class-action lawsuit, we find that 17.1% have a fitted loan fee in the top 1%, and 87.4% have a residual loan fee in the top 1%. In other words, among short sellers' very best ideas, the unique information component contained in loan fees predicts lawsuits much better than information that overlaps with existing anomalies.

Row (2) and (3) of Table 11 repeat the analysis for two periods within the class-action lawsuit. The first period is the Class Period, which begins when shareholders allege the company first made false or misleading statements and ends when the company corrected the statements. The second period is when the lawsuit is in progress, which begins when the lawsuit is filed and ends when it is resolved. In both these periods, we find that both of our conclusions hold: (1) specials stocks are more likely to be in these two phases of a class action

Table 11. *Loan Fee* and 10b-5 Class-Action Lawsuits

	GC stocks (%)	Special stocks (%)			
	Loan Fee <	*Loan Fee* >			
Pct. stocks by lawsuit stage	90th ptile.	90th ptile.	95th ptile.	98th ptile.	99th ptile.
(1) Pct. of Stocks in Any Stage of Class-Action Lawsuit	**6.2**	**12.5**	**15.1**	**20.5**	**23.5**
Pct. of (1) that *Fitted Loan Fee* captures	—	61.2	44.6	26.9	17.1
Pct. of (1) that *Residual Loan Fee* captures	—	61.2	86.2	97.0	87.4
(2) Pct. of Stocks within Class Period of Alleged Fraud	**2.8**	**5.6**	**6.7**	**9.7**	**10.4**
Pct. of (2) that *Fitted Loan Fee* captures	—	61.5	44.3	28.5	17.6
Pct. of (2) that *Residual Loan Fee* captures	—	62.5	87.6	96.9	87.7
(3) Pct. of Stocks with Lawsuit in Progress	**3.2**	**7.2**	**8.7**	**11.7**	**14.3**
Pct. of (3) that *Fitted Loan Fee* captures	—	59.5	42.7	25.8	16.9
Pct. of (3) that *Residual Loan Fee* captures	—	62.3	88.2	97.5	87.8

Notes. The rows in bold of the table show the average percentage of stocks in a stage of a 10b-5 class-action lawsuit for general collateral and special stocks. "Any Stage of Class-Action Lawsuit" includes the time from the class period start to when the lawsuit is resolved. "Class Period of Alleged Fraud" includes the time from when the class period starts to the class period ends. "Lawsuit in Progress" includes the time from when the lawsuit is filed to when the lawsuit is resolved. The nonbold rows show what percentage of stocks that have a high *Loan Fee* and are in the indicated phase of the lawsuit also have a high *Fitted Loan Fee* or *Residual Loan Fee*. The threshold for having a high *Fitted Loan Fee* or *Residual Loan Fee* is defined using the same percentile cutoff that is used for defining a high *Loan Fee* in the corresponding column. For example, the top panel, rightmost column indicates that, on average, 23.5% of stocks with a *Loan Fee* greater than the 99th percentile are in some stage of a class-action lawsuit. Among the 23.5% of high *Loan Fee* stocks that are in some phase of a lawsuit, 17.1% have a *Fitted Loan Fee* greater than the 99th percentile, and 87.4% have a *Residual Loan Fee* greater than the 99th percentile.

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

lawsuit than general collateral stocks, and (2) the residual component of loan fee captures short sellers' best ideas about financial misconduct than fitted loan fee.

Class-action lawsuits are one setting where the prior literature has shown that short sellers have an information advantage. We find that a meaningful fraction of short sellers' best ideas stems from this informational advantage. Moreover, this advantage comes from the component of loan fees that is orthogonal to existing anomalies, suggesting that their information is unique.

4.4. The Dependence of Existing Anomalies on Loan Fees

The previous results are evidence that the majority of *Loan Fee*'s return predictability is not dependent on existing anomalies. In Table 12, we ask the opposite question (à la Drechsler and Drechsler 2016): to what extent are existing anomalies dependent on high loan fee stocks? We examine a set of double sorts in Table 12. In Panel A, we sort stocks independently into general collateral (GC) and special portfolios and by decile of the GHZ net score. In Panel B, we perform conditional sorts: we first sort stocks into GC and special portfolios, and then, within those two portfolios, sort stocks into deciles by their GHZ net score. Either double-sort procedure generates the same takeaway: although there is no difference in performance between the top and bottom GHZ deciles amongst GC stocks, there is a difference of 0.42% or 0.97%, depending on the double-sort method, between the top and bottom GHZ net deciles only among the high loan fee stocks. This suggests that the informative component of the aggregate *GHZ Net* anomaly signal resides largely within the on-special stocks. In other words, we find that anomalies live in high loan fee stocks.

4.5. The Information in Loan Fees and Short Interest

Ultimately, we are interested in short sellers' expected return forecasts for each stock. Unfortunately, to the best of our knowledge, there is no database containing this measure. However, we argue that *Loan Fee* is the best proxy for short sellers' expected return forecasts. The intuition is simple: Loan Fee proxies for short sellers' willingness to pay. In other words, a high loan fee for a given stock indicates that some short sellers have a low enough expected return so that short selling will yield a profit, even after paying the loan fee. Consider an example: if short sellers borrow shares of stock ABC at a loan fee of 15% per annum, this implies that those

Table 12. Double Sorts: One-Month Return of *Loan Fee* and *GHZ Net* Anomaly

	General collateral (GC)		Special stocks		
Decile	Mean no. of stocks	Mean return (%)	Mean no. of stocks	Mean return (%)	Return on GC – special (%)
Panel A: Independent double sort					
1	155	0.83	82	−0.22	1.05
2	199	0.75	42	0.05	0.70
3	211	0.87	28	0.14	0.73
4	207	1.03	21	0.67	0.36
5	241	0.99	23	1.17	−0.17
6	194	0.98	17	0.93	0.04
7	250	0.88	21	0.43	0.45
8	199	0.99	18	0.23	0.76
9	236	0.94	19	0.70	0.23
10	207	0.79	22	0.20	0.59
L/S	—	−0.04	—	0.42	—
Panel B: Conditional double sort					
1	211	0.78	29	−0.68	1.46
2	212	0.82	29	0.03	0.79
3	196	0.80	29	−0.10	0.90
4	229	1.07	30	0.33	0.74
5	193	0.96	29	−0.05	1.01
6	202	0.92	30	0.59	0.33
7	248	0.85	29	0.58	0.27
8	186	1.03	30	1.10	−0.07
9	217	0.94	29	0.46	0.48
10	206	0.78	29	0.29	0.49
L/S	—	0.00	—	0.97	—

Notes. The table displays monthly mean returns from dual-sorts formed on the *Loan Fee* anomaly portfolio and the *GHZ Net* anomaly portfolio. In Panel A, we independently sort stocks into GC and special portfolios, and we sort by decile of the *GHZ Net* score. In Panel B, we first sort stocks into GC and special portfolios, and then, within each of these groups, we sort by decile of the *GHZ Net* score. In both panels, L/S indicates the long-short return earned from buying stocks in decile ten and short-selling stocks in decile one.

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

short sellers have an expected return for that stock that is below −15% per annum, so that, after loan fees, their trade has a positive expected profit.

Notice that the same cannot be said for short interest. Consider a stock with high short interest, but a small loan fee—say, 0.1% per annum. It is possible that all short sellers who have taken a short position expect a return of −0.2%. In this case, it is rational for short sellers to take such a position, so many do, and short interest becomes very high, but this would not forecast a large future price drop. In summary, when loan fees are high, it likely that the marginal short seller has an exceedingly low expected return. When short interest is high, no such necessary condition exists.

We explore this idea in Table 13, which shows return data double sorted by both short interest and loan fees. The table also reports median loan supply, the median percentage of the supply of shares available to borrow that have not been lent out.[20] For both Panels A and B, the rows represent the different deciles of *Short Interest* for stocks, while the *Loan Fee* variable is used to separate the sample of stocks into two groups: general collateral and special. Looking first at the general collateral stocks, we see that sorting stocks by *Short Interest* deciles generates a return difference of 0.42% per month, whereas the same sort among special stocks generates a return difference of 1.19%. Both of these suggest that sorting stocks by *Short Interest* has predictive power for future returns. However, looking at the difference in returns between general collateral and special stocks, we see that *Loan Fee* has predictive power within *Short Interest* deciles. Although the difference in returns varies across *Short Interest* deciles, the average is 0.50%.

Looking at the median *Short Interest* and *Loan Supply* within each decile and across general collateral and special stocks gives some insight into this result, consistent with our discussion above. Although the median *Short Interest* is the same or similar across general collateral and special stocks, the median *Loan Supply* is much lower for the special stocks, suggesting that although short-selling demand for these stocks is similar, the supply is very different. Panel B shows economically similar results using a conditional double sort.

4.6. Performance Net of Loan Fees

Our *Loan Fee* anomaly analysis has found large gross profits, which is evidence that high loan fee stocks are overpriced. In this section, we examine whether the

Table 13. Double Sorts: One-Month Return of *Short Interest* and Median *Loan Supply*

	General collateral (%)			Special (%)			
Decile	Mean return	Median *Short Interest*	Median *Loan Supply*	Mean return	Median *Short Interest*	Median *Loan Supply*	L/S
Panel A: Independent double sort							
1	0.95	1	69	1.09	0	27	−0.14
2	1.14	1	92	−0.43	1	41	1.57
3	1.06	2	84	−0.63	2	52	1.69
4	0.90	2	73	0.07	2	19	0.84
5	0.82	3	61	0.59	3	14	0.23
6	0.87	4	53	1.38	4	14	−0.51
7	0.73	5	50	0.77	5	15	−0.04
8	0.73	7	47	0.28	8	14	0.45
9	0.60	11	46	0.30	11	17	0.30
10	0.53	17	48	−0.10	22	30	0.63
L/S	0.42	—	—	1.19	—	—	0.50
Panel B: Conditional double sort							
1	0.95	1	70	1.12	1	25	−0.17
2	1.15	1	92	−0.03	2	32	1.18
3	1.06	2	83	−0.03	4	19	1.09
4	0.96	2	75	1.32	6	13	−0.37
5	0.83	3	63	−0.03	9	15	0.87
6	0.93	4	54	0.05	12	18	0.88
7	0.61	5	51	−0.33	15	21	0.94
8	0.87	6	47	0.56	19	26	0.31
9	0.66	9	46	−0.60	24	30	1.26
10	0.51	14	47	−0.40	33	41	0.91
L/S	0.44	—	—	1.52	—	—	0.69

Notes. The table displays monthly mean returns from dual-sorts formed on the *Loan Fee* anomaly portfolio and *Short Interest* portfolio. *Short Interest* is calculated as a percentage of shares outstanding. In Panel A, we independently sort stocks into GC and special portfolios, and we sort by decile of the *Short Interest*. In Panel B, we first sort stocks into GC and special portfolios, and then, within each of these groups, we sort by decile of the *Short Interest*. In both panels, L/S indicates the long-short return earned from buying stocks in decile one and short-selling stocks in decile 10. For each decile, we also report the median *Loan Supply*. The *Loan Supply* for each stock is estimated by dividing the *Short Interest* by the utilization reported in the FIS Astec data for each stock.

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Loan Fee anomaly is profitable *net* of fees. As Novy-Marx and Velikov (2016) put it: "…the existence of anomalies indicates suboptimal behavior on the part of some individual traders, but does not suggest irrationality on market participants more broadly. The existence of anomalies that can be profitably traded consequently presents a stronger test of market efficiency."

Chen and Velikov (2023) find that most anomaly strategies earn abnormal returns close to zero after accounting for transaction costs. At the same time, our GHZ anomaly results are also gross of loan fees, so it is not clear, ex ante, how including these costs in the analysis will affect the performance ordering of the strategies. To address this, we repeat the performance comparison of Table 2 after subtracting borrowing costs, and the results are shown in Table 14. Although the magnitude of returns is smaller after subtracting transaction costs, it appears that the adjustment affects all of the anomalies in almost the same way. As a result, the *Loan Fee* anomaly continues to be one of the top performers across all methodologies and metrics. For example, across our four portfolio cutoffs (the bottom 1%, 2%, 5%, and 10%), we find the *Loan Fee* anomaly has a Sharpe Ratio that ranks 14th, 9th, 12th, and 11th, respectively. Similarly, when we examine the average of one-month returns, we find the *Loan Fee* anomaly ranks no worse than 15th. Considering these results, the *Loan Fee* anomaly continues to outperform the *GHZ Net* anomaly after adjusting for the fact that loan fees are a cost that must be paid by short sellers.

4.7. Robustness

In a first set of robustness tests, we examine whether the *Loan Fee* anomaly result is simply a recasting of the well-known relation between anomalies and size or illiquidity (Hou et al. 2020), which also exists in our data. In Table 15, we run Fama and MacBeth (1973) regressions with *Loan Fee*, *Size*, *Illiquidity*, *GHZ Net*, and their interaction terms. When we interact *GHZ Net* with *Size* (column (6)) and with a measure of illiquidity (*Bid-Ask Spread*; column (7)), we find the expected relation from the literature: anomalies perform better among small and illiquid stocks. However, in column (10), when we also include an interaction between *GHZ Net* and specialness, this interaction drives out the previous two. In other words, we find no statistically detectable relationship between returns and the interaction of *Size*/*Illiquidity* and the aggregate anomaly variable after accounting for its relation with *Loan Fee*.

In a second set of robustness tests, we also examine how our results hold up throughout the cross-section. Specifically, we repeat the analysis of Table 2, but first raise the NYSE market capitalization cutoff from the 5th to the 20th percentile. Table A7 in the Online Appendix reports these results. Although the average return of the *Loan Fee* anomaly long/short portfolio declines from 4.01% to 2.78% for the top/bottom 1% categorization, it is still the best-performing anomaly, as measured by average one-month returns or monthly Sharpe Ratio, and it has the highest percentage of positive one-month returns. This top ranking is maintained, in part, because

Table 14. Performance of Long/Short Portfolio Net of Short-Selling Costs

	Loan Fee		*GHZ Net*		GHZ anomalies		
Anomaly definition	Value	Rank	Value	Rank	Mean	95th %	Max.
Panel A: Average of one-month return							
Top/bottom 1%	0.75%	15	0.09%	41	−0.15%	0.99%	2.01%
Top/bottom 2%	0.55%	11	0.36%	19	−0.18%	0.87%	1.32%
Top/bottom 5%	0.37%	7	0.12%	25	−0.16%	0.53%	1.71%
Top/bottom 10%	0.30%	11	0.01%	33	−0.12%	0.43%	1.04%
Panel B: Percentage of one-month returns that are positive							
Top/bottom 1%	58.4%	4.5	48.1%	53	48.3%	57.1%	63.6%
Top/bottom 2%	49.4%	45	53.2%	15.5	47.3%	55.8%	61.0%
Top/bottom 5%	50.6%	32	50.6%	32	47.3%	55.8%	61.0%
Top/bottom 10%	49.4%	38.5	53.2%	18	47.3%	57.1%	61.0%
Panel C: Monthly Sharpe Ratio							
Top/bottom 1%	0.12	14	0.01	42	−0.04	0.16	0.24
Top/bottom 2%	0.11	9	0.06	25	−0.04	0.12	0.26
Top/bottom 5%	0.09	12	0.02	26	−0.05	0.12	0.31
Top/bottom 10%	0.09	11	0.00	34	−0.05	0.15	0.25

Notes. The table examines the performance of the *Loan Fee* anomaly and the GHZ anomalies after subtracting short-selling costs. For each anomaly, we construct a long/short portfolio, calculate its return, and subtract from the return the realized short-selling cost for the stocks in the short leg of the portfolio. The performance of each anomaly is evaluated using three metrics: the average one-month returns (Panel A), the percentage of one-month returns that are positive (Panel B), and the monthly Sharpe Ratio (Panel C). For each of these, we also rank each anomaly variable relative to the entire set of anomalies. The GHZ set of anomalies is the set of 102 anomalies studied in Green et al. (2017). In each panel, we present results when portfolios are formed by sorting stocks on the top and bottom 1% of anomaly variables ("Top/Bottom 1%"), as well as when portfolios are formed by sorting stocks on the top and bottom 2%, 5%, and 10%.

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Table 15. *Loan Fee* Anomaly and *GHZ Net*: Fama-MacBeth Regressions (Quantiles)

Explanatory variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Special	−0.64%**				−0.50%**			−0.42%**	−0.32%**	−0.32%*
	−2.29				−2.24			−1.96	−2.05	−1.89
GHZ Net		0.31%			0.08%	0.25%	0.14%	0.06%	0.04%	0.02%
		0.95			0.29	0.83	0.69	0.24	0.19	0.10
Size			0.44%			0.45%		0.28%		−0.03%
			1.14			1.15		0.69		−0.07
Bid-Ask Spread				−0.75%			−0.70%		−0.55%	−0.54%
				−1.20			−1.18		−0.92	−0.82
Special × *GHZ Net*					0.87%*			0.82%*	0.72%*	0.66%*
					1.78			1.73	1.88	1.68
Size × *GHZ Net*						−0.67%		−0.17%		−0.29%
						−1.21		−0.35		−0.53
Bid-Ask Spread × *GHZ Net*							0.74%		0.30%	0.17%
							0.81		0.34	0.18

Notes. The table reports results from Fama and MacBeth (1973) cross-sectional regressions of one-month-ahead stock returns on the indicated independent variables. "*Special*" is an indicator variable that takes a value of 1 if the stock has an average retail loan rate greater than 1% per annum, and 0 otherwise. Quantiles are used for each of the remaining independent variables (*GHZ Net*, *Size*, and *Bid-Ask Spread*). The quantiles are calculated within the month and are shifted so that the median within the month has value zero. *t*-statistics are calculated using Newey West standard errors. Note that the coefficient on the *Special* indicator variable in column (1) is not directly comparable to the coefficient in columns (5), (8), (9), and (10). The coefficient in column (1) reports the average effect of specialness on returns, whereas the other columns report the average effect of specialness on returns conditional on the stock having a median *GHZ Net* score. The average effect of specialness on returns is −0.63%, −0.55%, −0.43%, and −0.42% for columns (5), (8), (9), and (10), respectively.

*Significant at 10%; **significant at 5%; ***significant at 1%.

raising the market capitalization cutoff also affects the performance of the set of GHZ anomalies. For example, the average one-month return of the *GHZ Net* anomaly declines from 0.70% to −0.16%. The average return across the GHZ set of anomalies also declines from 0.17% to 0.05%. Overall, excluding the bottom 20% of stocks by market capitalization does not materially change our performance results.

5. Conclusion

Comprehensive loan fee data are a relatively new addition to finance academic research, beginning with D'Avolio (2002). Unlike other well-known anomalies, such as momentum or size, we do not have a century's worth of data to analyze. Nevertheless, the time-series of loan fee data are growing, and, in modern data, we find loan fee to be the single best predictor of cross-sectional returns.

Given the evidence that short sellers are informed and that loan fees represent how much those informed traders are willing to pay to hold their position, we view this predictability as evidence of a short seller's best ideas. When short sellers are willing to pay the most to bet against a stock, those stocks predictably have the largest negative future returns.

We also find that these best ideas are unique. Only 58% of the loan fee anomaly can be explained by short sellers' selective exposure to existing anomalies; the remaining 42% is unique information held by these informed traders. This result is also borne out in the setting of securities fraud class-action lawsuits, where, again, the residual component is a better predictor than the fitted.

Given the uniqueness of loan fee as both a conviction and a cost, we do not anticipate the loan fee anomaly to disappear like other anomalies have (McLean and Pontiff 2016, Calluzzo et al. 2019). When we compare the persistence of the loan fee anomaly with other high-performing anomalies, we find more evidence of mean reversion among those high-performing anomalies and less evidence with the loan fee anomaly.

Interestingly, in the debate about the robustness of cross-sectional predictors over time after accounting for data mining, the single most effective predictor has been left out of the discussion. Given the newness of these data, this is understandable. We hope this paper encourages future work to examine it.

Acknowledgments

The authors thank Victoria Ivashina (the editor), an anonymous associate editor, an anonymous referee, David Chapman, Mike Cooper, Jesse Davis, Michael Gallmeyer, Andrei Gonçalves, Charles Jones, Juhani Linnainmaa, Elena Loutskina, Christian Lundblad, Pedro Matos, Paige Ouimet, Jeff Pontiff, Jacob Sagi, Michael Schill, Daniel Schmidt, Gill Segal, Andreas Statopoplous, Davide Tomio, Selim Topaloglu, participants at the 2022 American Finance Association Annual Meeting, the 2022 European Finance Association Annual Meeting, the 2021 SFS Cavalcade 2021, the 2021 Northern Finance Association conference, and seminar participants at the University of Virginia and the University of North Carolina at Chapel Hill. All errors are our own.

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Endnotes

[1] See, for example, Cohen et al. (2007), Blocher et al. (2012), Drechsler and Drechsler (2016), Blocher and Ringgenberg (2018), Engelberg et al. (2018b), etc.

[2] For example, McLean and Pontiff (2016), Green et al. (2017), Harvey et al. (2016), Freyberger et al. (2020), Kozak et al. (2020), Chen and Zimmermann (2022), Gu et al. (2020), and Hou et al. (2020).

[3] The *Loan Fee* anomaly similarly outperforms the GHZ composite anomaly portfolio after accounting for transaction costs (see Table 14).

[4] For the former result, see, for example, Boehmer et al. (2008). For the latter result, see Stambaugh et al. (2012) and Drechsler and Drechsler (2016).

[5] To see this, consider a simple model in which the demand to short is either high or low and the supply of shares to borrow is either high or low. Thus, there are four possibilities: (i) low demand and low supply, (ii) low demand and high supply, (iii) high demand and high supply, and (iv) high demand and low supply. In cases (i) and (ii), when demand is low, both short interest and loan fees will be low. In case (iii), when the demand to short is high and supply is high, short sellers will borrow as many shares as possible, and they will continue to borrow shares until the loan fee reaches a point that further shorting is no longer profitable. Thus, both short interest and loan fees will be high. Finally, when the demand to short is high and supply is low (case (iv)), short sellers will again continue borrowing shares until shorting is no longer profitable, but this will occur at a low number of shares because of the limited supply. Thus, short interest will be low, but loan fees will be high.

[6] https://sites.google.com/site/jeremiahrgreenacctg/home.

[7] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[8] We use the FIS Astec data in our analyses because they provide the loan fee for each stock and date; however, in unreported analyses, we find that our conclusions are unchanged when we use an alternative database with a longer time period, but a coarser measure of loan fees.

[9] We estimate the parameters of each model in June of each year and fix the parameters for the next year. In June 2004, the first year in which we estimate the model, we use the prior 30 years of data and split the sample into an 18-year training sample and a 12-year validation sample. The hyperparameters of each model are selected to maximize the R^2 of the model in the validation sample based on estimation during the training sample. For the subsequent years, we roll forward the validation sample by one year and expand the training sample by one year. Our set of anomaly variables is the 94 GHZ anomalies that Gu et al. (2020) use from Green et al. (2017). We do not include the macroeconomic predictors and industry fixed effects that Gu et al. (2020) add to the list of predictors. As in Gu et al. (2020), we use the quantiles values of the anomalies when estimating the model. We implement the estimation methods using the code provided by Gu et al. (2020) on Dacheng Xiu's website, but make two adjustments. First, we use monthly demeaned returns for estimating the model and evaluating validation sample R^2, so that the procedure maximizes the cross-sectional variation in returns. Second, in some years, for elastic net, the validation sample R^2 was negative for all values of the penalty parameter, which meant that the optimal model was one with a penalty parameter so high that the model placed zero weight on every GHZ signal and resulted in a 0% long-short return for that year. To be conservative, we reran the estimation constraining the optimization so that the range over which we searched for the optimal penalty parameter was wide enough not to constrain the optimization for any of the years not affected by this issue, but no wider. The results of the constrained optimization were stronger, and they are what is reported in Table 3.

[10] Absolute accruals, Dispersion in forecasted EPS, Financial statement score (MS), Idiosyncratic return volatility, R&D to market capitalization, Percent change in sales minus percent change in accounts receivable, and Industry adjusted book to market.

[11] We conduct a *t*-test for a difference in means of the one-month return for the top decile and bottom decile. We find the difference in means is statistically insignificant at the 10% level using Newey and West (1987) standard errors.

[12] We repeat this analysis in Online Appendix Table A3 after winsorizing the variables at the 1st and 99th percentiles; the conclusions are unchanged.

[13] Drechsler and Drechsler's (2016) *CME Beta* is a covariance (estimated as a stock's covariance with a long-short loan fee portfolio), whereas Loan Fee in this paper is a *characteristic*—as noted in Daniel and Titman (1997), this is an important distinction.

[14] *Days-to-Cover* is *Short Interest* divided by average daily trading volume over the previous month. *SIR* × *RetVol* is *Short Interest* divided by market capitalization times the daily return volatility over the previous month. *CME Beta* is the estimated beta on the cheap minus expensive factor from Drechsler and Drechsler (2016). The cheap minus expensive portfolio is constructed by going long stocks in the "Bottom 10%" portfolio and short stocks in the "Top 10%" portfolio for the Loan Fee anomaly. We proxy for a stock's beta with the CME factor by estimating its beta with the stocks in the CME portfolio over the 36 months prior to portfolio construction. Specifically, to estimate the beta for stock i in month t, we (1) identify the stocks in each leg of the CME portfolio in month t; (2) fix the set of stocks in the CME portfolio and calculate the long-short returns of the portfolio over the prior 36 months; and (3) regress the returns of stock i on the three-factor Fama-French model plus CME factor from (2).

[15] In other words, for anomalies that are positively correlated with loan fee, we always go long in stocks with a low value of the anomaly variable and short stocks with a high value. For anomalies that are negatively correlated with loan fee, we always go long in the high value and short in the low value.

[16] The choice to use a cutoff of 10 components in our principal components analysis is motivated by the scree plot shown in Figure A1 of the Online Appendix. We also repeat the PCA approach using cutoffs of one and six components with similar results. This additional analysis is reported in Table A4 of the Online Appendix.

[17] We also repeat the analysis using different cutoffs for the decomposition using principal components. These results are reported in Table A5 of the online Appendix.

[18] Although the idiosyncratic nature of such information is plausible, prior work also suggests that previously documented anomalies may also help forecast class-action lawsuits. DuCharme et al. (2004), for example, find that abnormal accruals are positively associated with the incidence of shareholder class-action lawsuits.

[19] In Table A6 of the Online Appendix, we show that the results are similar in a regression framework.

[20] In our data, we find that loan supply significantly varies across stocks and within the same stock across time. For example, the median stock, on average, has a loan supply of 51% of shares outstanding, while the interquartile range is 30% to 105%. This suggests that short interest can be low for different reasons: it could be low because short sellers do not want to short, but it could also be that they want to short, but low loan supply makes it difficult.

References

Anton M, Cohen R, Polk C (2021) Best ideas. Working paper, University of Navarra, IESE Business School, Barcelona, Spain.

Asquith P, Pathak PA, Ritter JR (2005) Short interest, institutional ownership, and stock returns. *J. Financial Econom.* 78:243–276.

Downloaded from informs.org by [2607:f720:1901:4801::105] on 14 May 2025, at 11:48 . For personal use only, all rights reserved.

Ball R, Brown P (1968) An empirical evaluation of accounting income numbers. *J. Accounting Res.* 6:159–178.

Blocher J, Ringgenberg MC (2018) Converting short sale constraints. Working paper, Vanderbilt University, Nashville, TN.

Blocher J, Reed AV, Wesep EV (2012) Connecting two markets: An equilibrium framework for shorts, longs and stock loans. *J. Finance* 108:302–322.

Boehmer E, Jones CM, Zhang X (2008) Which shorts are informed? *J. Finance* 63:491–527.

Calluzzo P, Moneta F, Topaloglu S (2019) When anomalies are publicized broadly, do institutions trade accordingly? *Management Sci.* 65:4555–4574.

Chen AY, Velikov M (2023) Zeroing in on the expected returns. *J. Financial Quant. Anal*. 58(3):968–1004.

Chen AY, Zimmermann T (2022) Open source cross-sectional asset pricing. *Crit. Finance Rev.* 27(2):207–264.

Cohen L, Diether K, Malloy C (2007) Supply and demand shifts in the shorting market. *J. Finance* 62:2061–2096.

Cooper M, Gulen H, Schill MJ (2008) Asset growth and the cross-section of stock returns. *J. Finance* 63(4):1609–1651.

Daniel K, Titman S (1997) Evidence on the characteristics of cross-sectional variation in stock returns. *J. Finance* 52:1–33.

D'Avolio G (2002) The market for borrowing stock. *J. Financial Econom.* 66:271–306.

Dechow PM, Hutton AP, Meulbroek LK, Sloan RG (2001) Short sellers, fundamental analysis and stock returns. *J. Financial Econom.* 61:77–106.

Drechsler I, Drechsler QFS (2016) The shorting premium and asset pricing anomalies. Working paper, The Wharton School, University of Pennsylvania, Philadelphia.

DuCharme LL, Malatesta PH, Sefci SE (2004) Earnings management, stock issues, and shareholder lawsuits. *J. Financial Econom.* 71:27–49.

Engelberg J, McLean RD, Pontiff J (2018a) Anomalies and news. *J. Finance* 73:1971–2001.

Engelberg J, Reed AV, Ringgenberg MC (2012) How are shorts informed? Short sellers, news, and information processing. *J. Financial Econom.* 105:260–278.

Engelberg J, Reed AV, Ringgenberg MC (2018b) Short selling risk. *J. Finance* 73:755–786.

Fama E, French K (1992) The cross-section of expected stock returns. *J. Finance* 47(2):427–465.

Fama E, French K (2008) Dissecting anomalies. *J. Finance* 63: 1653–1678.

Fama E, MacBeth JD (1973) Risk, return, and equilibrium: Empirical tests. *J. Polit. Econom.* 81:607–636.

Fang VW, Huang AH, Karpoff JM (2016) Short selling and earnings management: A controlled experiment. *J. Finance* 71:1251–1294.

Freyberger J, Neuhierl A, Weber M (2020) Dissecting characteristics nonparametrically. *Rev. Financial Stud.* 33(5):2326–2377.

Geczy CC, Musto DK, Reed AV (2002) Stocks are special too: An analysis of the equity lending market. *J. Financial Econom.* 66:241–269.

Green J, Hand JR, Zhang XF (2017) The characteristics that provide independent information about average U.S. monthly stock returns. *Rev. Financial Stud.* 30(12):4389–4436.

Gu S, Kelly B, Xiu D (2020) Empirical asset pricing via machine learning. *Rev. Financial Stud.* 33(5):2223–2273.

Hanson SG, Sunderam A (2014) The growth and limits of arbitrage: Evidence from short interest. *Rev. Financial Stud.* 27:1238–1286.

Harvey CR, Liu Y, Zhu H (2016) … And the cross-section of expected returns. *Rev. Financial Stud.* 29(1), 5–68.

Henderson BJ, Jostova G, Philipov A (2023) Strategic behavior by equity lenders. Working paper, George Washington University, Washington, DC.

Hong H, Li W, Ni SX, Scheinkman J, Yan P (2021) Days to cover and stock returns. Working paper, Columbia University, New York.

Hou K, Xue C, Zhang L (2020) Replicating anomalies. *Rev. Financial Stud.* 33(5):2019–2133.

Jegadeesh N, Titman S (1993) Returns to buying winners and selling losers: Implications for stock market efficiency. *J. Finance* 48:65–91.

Jiang H, Verbeek M, Wang Y (2014) Information content when mutual funds deviate from benchmarks. *Management Sci.* 60:2038–2053.

Jones CM, Lamont OA (2002) Short-sale constraints and stock returns. *J. Financial Econom.* 66:207–239.

Karpoff JM, Lou X (2010) Short sellers and financial misconduct. *J. Finance* 65(5):1879–1913.

Kozak S, Nagel S, Santosh S (2020) Shrinking the cross-section. *J. Financial Econom.* 135(2):271–292.

Lewellen J (2015) The cross-section of expected stock returns. *Crit. Finance Rev.* 4:1–44.

McLean RD, Pontiff J (2016) Does academic publication destroy stock return predictability? *J. Finance* 71:5–32.

Newey WK, West KD (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–708.

Novy-Marx R, Velikov M (2016) A taxonomy of anomalies and their trading costs. *Rev. Financial Stud.* 29:104–147.

Rosenberg B, Reid K, Lanstein R (1985) Persuasive evidence of market inefficiency. *J. Portfolio Management* 11:9–16.

Stambaugh RF, Yu J, Yuan Y (2012) The short of it: Investor sentiment and anomalies. *J. Financial Econom.* 104:288–302.