Three Essays on Machine Learning in Empirical Finance

Jinhua Wang

Cambridge Judge Business School
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Churchill College March 2022

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed 80,000-word limit for the Degree Committee of the Faculty of Business and Management.

Chapter 1, "Survival of the Fittest? Managerial Cultural Fit and Tenure", is a sole-authored paper. Chapter 2, "Do Investors Pay Less Attention to Women (Fund Managers)?" is co-authored with my supervisor Professor Raghavendra Rau. I am responsible for half of the work. Chapter 3, "Did Trading Bots Resurrect the CAPM?" is co-authored with Professor Andreas Park from the University of Toronto. I am responsible for half of the work.

Jinhua Wang
March 2022

Three Essays on Machine Learning in Empirical Finance

Jinhua Wang

Abstract

The dissertation consists of three essays that contribute to the literature on machine learning in empirical finance.

In the first paper, I create proxies for managers' cultural fit using one of the latest machine learning technologies – the sentence embedding model - by analysing 11.5 million speeches in earnings calls. A better cultural fit is significantly and positively correlated with managerial tenure. I demonstrate that the effect of cultural fit on managerial tenure is causal using random survival forests. Firms that hire culturally disruptive managers have lower future market values and performance. The stock market reacts positively to signals that indicate low cultural dispersion within the firm.

In the second paper, we document a gender-based attention effect in the sensitivity of mutual fund flows to fund performance using individual-level fund data from a fintech platform in China. Investors increase (decrease) flows to funds following positive and strong (negative and weak) prior-month performance. However, although there is no significant difference in the performance of male and female managers, the sensitivity effect significantly weakens if the fund manager is female. The effect persists after controlling for the tone of news articles on fund managers, measured using a state-of-the-art machine learning model. Simply put, investors react less to the performance of female fund managers.

In the third paper, we document a significant, up to 10-fold increase in the synchronicity of intra-day, ultra-high frequency stock returns over the last decade. This surge in the intra-day synchronicity across stocks coincided with the advent of electronic, automated trading in U.S. markets. Using changes to the S&P500 index, we establish evidence of a causal relationship using a new machine learning tool - causal random forests. When firms are included in this major index, they enter the radar of high frequency arbitrageurs and market-making bots. These automated trading bots, who monitor prices in major securities closely and continuously, increase their quoting activities significantly and cause individual stocks' returns to synchronize at the microstructure level.

To my family and friends.

Acknowledgements

I am deeply indebted to my supervisor, Professor Raghavendra Rau, for his great support and advice during those four years. I am also thankful to Professor Andreas Park, who inspired me to embark on my PhD journey. I am grateful for Professor Melvyn Weeks, who ignited my passion for Machine Learning and Causal Inference. I would also like to thank Professor Paul Kattuman, Professor Andrei Kirilenko, Professor Oğuzhan Karakaş, Professor Bang Dang Nguyen, Professor David Chambers and Professor Pedro Saffi for their help and comments during the PhD. Additionally, I would like to thank Ms Joanna Blakeman for her administrative support and as the PhD Programme Administrator. Many thanks are also due to my fellow PhD colleagues, especially Shiqi, Elias, Adam, Eoghan, Yuxia, Vaska, Chara, and Cindy. Last but not least, I would also like to thank my family and friends in the United Kingdom, Canada and China.

Table of contents

Chapter 1

Introduction

This dissertation presents three essays on machine learning in empirical finance. First, I use machine learning to extract economic variables from high-dimensional data (such as speeches and texts). Second, I apply the extracted variables in financial datasets to study the behaviors of corporate managers in the U.S. and individual fund investors in China. Third, I use machine learning algorithms for non-parametric causal inference in financial datasets, such as high-frequency trading data.

In the first paper, I propose a machine learning approach to measure managerial cultural fit. I make an important methodological contribution to the finance literature by introducing the sentence embedding model (Devlin, Chang, Lee, and Toutanova, 2019, Reimers and Gurevych, 2019, and Grootendorst, 2020) to measure cultural fit. The sentence embedding model is a novel machine learning method to compare the semantic similarities of complete sentences. I estimate the cultural preferences of each manager by clustering speeches in earnings call transcripts on their similarity in semantic meanings. In addition, I calculate corporate cultural values based on all speeches in an earnings call session. I then construct two cultural fit measures - corporate cultural distance (between firms) and personal cultural distance (between managers and firms). The corporate cultural distance measure is related to cultural adaptation, which is considered a common process of environmental adaptation and deep-rooted in human nature (Kim, 1988, Anderson, 1994, and Kim, 2017). When a manager leaves a familiar corporate culture for a new one, the distance between the two corporate cultures is the proxy for managers' cultural adaptation costs. A key assumption is that the lower the corporate cultural distances, the lower the cultural adaptation costs and the better the cultural fit of managers at new firms. The second cultural fit measure, personal cultural distance, is the distance between a manager's cultural vector and the company's cultural vector. The personal cultural distance essentially measures how different a manager's cultural preferences are from the corporate cultural values.

Consistently, the Proportional Hazards models show that a 1 standard deviation increase in corporate cultural distances (personal cultural distances) is associated with a 7% (6%) increase in the expected hazard ratio that hurts managerial tenure. Next, I use causal survival forests (CSF) to demonstrate the causality of cultural distances on managerial tenure. CSFs are an extension to causal forests, which utilize machine learning to recover the unobserved counterfactuals and estimate the causal effects of treatment. Specifically, a large corporate cultural distance (personal cultural distance) will result in at least 7 (10) months shorter tenure in new firms. Meanwhile, the treatment effect of cultural distances on managerial tenure is not homogenous. As executive compensation increases, the treatment effects of cultural distances on tenure become more negative, indicating that cultural fit on tenure is exacerbated by higher managerial compensation.

Should companies hire managers that are more culturally aligned or disruptive? I next show that culturally disruptive hiring managers are negatively associated with firms' (future) market values. Furthermore, I show that firms with culturally aligned managers are associated with a higher future Return on Invested Capital and Earnings Per Share. Therefore, the evidence is consistent with the mechanism that cultural distance is positively correlated with managerial turnover, which is negatively correlated with firm performance.

How do investors perceive cultural fit signals sent by executives during earnings calls? While it has been documented in the literature that investors respond to earnings announcements (Da, Engelberg, and Gao, 2011, and Ben-Rephael, Da, and Israelsen, 2017), it has not been clear if investors also respond to information in earnings calls that are not directly related to firm performance. Next, I construct portfolios that go long into the companies with the lowest cultural dispersion and shorts the companies with the highest cultural dispersion. The portfolio generates an annualized excess return of 2% over the Carhart four-factor model. Although an investor is unlikely to profit from the long-short strategy after transaction costs, the results still imply that hiring managers with better cultural fit could benefit firms' financial outcomes and stock market performance.

In the second paper, we document a new different and previously unstudied type of gender bias in investor behavior, which we term gender-based attention bias. Gender-based attention bias refers to the tendency to pay less attention to women than men. Specifically, we examine whether gender-based attention bias affects the well-documented flow-performance relationship in the mutual fund literature. Sirri and Tufano (1998) and Del Guercio and Tkac (2002), among others, document that there is a positive correlation between prior mutual fund performance and subsequent fund flow, commonly termed the flow-performance sensitivity of the fund. Niessen-Ruenzi and Ruenzi (2018) document significantly lower inflows into female-managed funds than male-managed funds at the aggregate annual fund level. In contrast to the previous literature, we examine whether the flow-performance sensitivity for individual investors is affected by managerial gender, using a unique dataset provided by a large fintech platform in China.

We first document strong evidence of a differential flow-performance sensitivity between male- and female-managed funds. Interacting the flow-performance sensitivity term with a gender dummy variable, we show that flow-performance sensitivity is significantly weaker for female-managed funds. An alternative rational explanation for the muted flow-performance sensitivity for female managers is that male managers' past performance is a better predictor of their future performance. Hence investors rationally invest in what they believe will be better-performing funds. We show, however, that past performance does not predict future performance for Chinese fund managers and that there is no gender difference in the predictive ability for future performance. In addition, controlling for managerial characteristics and time fixed effects, we find that female fund managers perform significantly better (at the 10% level) than male fund managers when computing rankings at the one-month period. In other words, investors appear to believe without evidence that better-performing male managers are more likely to perform well than female managers. A second alternative explanation is that female managers are less likely to take risks. Hence, investors who do not adequately adjust for risk direct more flows to (risky) male managers. Regressing a battery of risk measures on manager gender, controlling for managerial characteristics and time fixed effects, we find no significant differences in levels of idiosyncratic or systematic risk or risk-adjusted-performance between female and male-managed mutual funds.

The Chinese sample of fund managers is similar to the data studied in other countries in that only 15% of the sample of managers are female. For example, Niessen-Ruenzi and Ruenzi (2018) analyze the gender-fund flow relationship in a sample where only 13.8% of the mutual fund manager sample is female. Hence, we investigate if the attention bias arises because investors face higher search costs when searching for female fund managers. To explicitly address the visibility problem, we use two approaches. First, we match each female manager to a similar male manager in each month using a propensity-score matching (PSM) approach using a host of managerial and fund characteristics as proxies for manager visibility. Attention bias continues to be significant in this matched fund sample, where search costs for male and female managers are likely to be approximately similar. Second, we use a Natural Language Processing (NLP) technique to extract names from 400,000 financial news articles in Chinese and count the frequency of each manager's name appearance in the article each month. We show that the level of attention bias is unaffected by the level of media coverage. We also measure the frequencies of positive media mentions and negative media mentions of each fund manager through sentiment analysis based on a variant of the Transformer model (Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, and Polosukhin, 2017) in machine learning. We show that, while the sign of the media mention does not appear to affect the level of attention bias, positive mentions of fund managers in the news strengthens the flow-performance relationship on average. However, we continue to find evidence that controlling for performance and positive media mentions, female manager earn lower fund flows than male managers.

Is there a causal relationship between the gender identity of female managers and investor fund flows? To examine this question, we employ a three-stage instrumental variable regression approach suggested by Wooldridge (2001). We use two different instrumental variables, the first being the proportion of illiterate women amongst all women in the municipal district that the investor resides in, and the second being the proportion of female new-borns amongst all new-borns in the municipal district that the investor resides in. Both instrumental variables do not instrument for fund manager per se, but for the specific investor's choice of a fund manager. The instruments do not directly drive investors' fund flow decisions but are likely to be related to investors' biases on gender identities, conditional on investors' characteristics that we control for. Our instrumental variable regressions confirm the existence of gender-based attention biases away from female managers, which cause investors to pay less attention to female-managed funds.

Finally, we formally run a regression testing the difference of individual fund flow volatilities between male- and female-managed funds. Our results show that individuals holding female-managed funds exhibit lower fund flow volatilities throughout our sample period. This may have the desirable impact of lowering the volatility of flows into the fund for mutual fund companies.

The third paper studies the intra-day synchronization of asset stock returns over a long horizon. We document a substantial increase in the fraction of intra-day stock return variations related to market-wide fluctuations. We also study whether the relationship between algorithmic, autonomous electronic trading and the substantially stronger intra-day synchronization of returns across securities is causal.

First, We assess whether bot trading or "bot trading" leads to changes in the speed of stock return synchronization across securities by studying the relationship of bot trading and the goodness-of-fit of a standard market model estimated on high-frequency, intra-day data. Our goal is to assess how the fraction in the realized return variations related to market-wide fluctuations correlates with the extent of bot trading. Motivated by Roll (1988) who, in his 1988 presidential address, discussed the extent to which security returns are captured by the R2 of an OLS regression of stock returns on the returns of the market portfolio, we focus on a related, intra-day measure. Namely, we estimate OLS regressions of intra-day stock returns on intra-day market returns, record the goodness-of-fit in the form of the regression's R^2 (and the coefficients) and perform a panel regression analysis with the R^2s as the proxy for the degree of return synchronization across securities.

We proceed by estimating, for each security in the NYSE TAQ database and day between 2003 and 2014, a standard market model for intra-day 5-second and 5-minute mid-quote returns. We use the Russell 2000 index returns represented by the exchange-traded fund IWM as our high-frequency proxy for the market. We interpret R^2 as the fraction of the variation in a stock's returns related to market-wide fluctuations at the intra-day level. We study changes to the S&P 500 constituents to establish the causal relationship. Our causal reasoning is as follows: index events are exogenous to the presence of electronic traders. When a firm gets included in an index, nothing changes for the firm itself. Still, high-frequency traders pay more attention to the stock and change prices accordingly, affecting the R^2 of the return regression (i.e., leads to a stronger return synchronization). Therefore, the exogenous variation in high-frequency bot trading allows us to establish a causal link between bot trading and the synchronization of stock returns. As firms that are included or excluded from indices tend to be large in market capitalization, we cautiously interpret our causality identification as local average treatment effects for large, highly liquid firms and have positive earnings for four consecutive quarters.

We construct a matched sample for all entry and exit events and apply a difference-in-differences panel regression approach to establish the link empirically. We follow three econometric approaches: First, we use the event in an instrumental variable (IV) approach. This approach assumes that the index event does not directly affect the R^2 (or through other channels than bot trading). Second, we perform a mediation analysis that allows the event to also have a direct effect on the R^2. The third approach uses new tools from machine learning literature, so-called causal random forests, first introduced in economics by Athey and Imbens (2016). This approach allows us to directly determine the (causal) treatment effect without a matched sample. Additionally, we expand the analysis and use an Instrumental Forest to address the endogeneity of the algorithmic trading proxy. Overall, our results indicate a causal relationship between bot trading and the fraction of the variation in firms' returns related to market-wide fluctuations.

The rest of this thesis is organized as follows. The following three chapters present my three papers, followed by a chapter concluding my main findings and implications.

Chapter 2

Survival of the fittest? Managerial Cultural Fit and Tenure

Survival of the fittest? Managerial Cultural Fit and Tenure

Jinhua Wang [1]

Abstract

I create proxies for managers' cultural fit using one of the latest machine learning technologies – the sentence embedding model - by analysing 11.5 million speeches in earnings calls. A better cultural fit is significantly and positively correlated with managerial tenure. I demonstrate that the effect of cultural fit on managerial tenure is causal using random survival forests. Firms that hire culturally disruptive managers have lower future market values and performance. The stock market reacts positively to signals that indicate low cultural dispersion within the firm.

Keywords: Corporate Culture, Manager Turnover, Natural-Language-Processing, Machine Learning, BERT, Transformer, Survival Analysis, Causal Machine Learning

[1] Cambridge Judge Business School and Churchill College, University of Cambridge, Cambridge CB3 0DS, United Kingdom, jw983@cam.ac.uk. I thank Cambridge HPC Services for providing the high-performance computation resources for the paper. I thank Raghu Rau, Oğuzhan Karakaş, Andreas Park, Marc Lipson, Adam Reed, Elias Ohneberg, Shiqi Chen and Adam Shuaib for their helpful comments. I thank Maarten Grootendorst, John Healy, James Melville, Yifan Cui and Erik Sverdrup for providing Machine Learning algorithmic references and sharing their code. I gratefully acknowledge full financial support from Cambridge Judge Business School. All remaining errors are mine.

"It is harder to help WeWork develop a customer-focused culture like Amazon's. I obviously did not do enough homework on the culture decision. Make good cultural decisions before you move from company to company."

Sebastian J. Gunningham
Co-chief executive, WeWork

1. Introduction

What is cultural fit? In the organizational behavioural literature, cultural fit refers to the difference between individuals' cultural preferences and the culture of the organization (O'Reilly, Chatman, and Caldwell 1991). In the mergers and acquisitions literature, cultural fit refers to the difference of corporate culture between two merger partners (Weber, Shenkar, and Raveh 1996, Weber 1996). Corporate culture often refers to a system of shared values or beliefs that define appropriate attitudes and behaviours for company members (O'Reilly and Chatman 1996, Guiso, Sapienza, and Zingales 2015, Li et al. 2020).

In this paper, I propose a machine learning approach to measure managerial cultural fit. I make an important methodological contribution to the finance literature by introducing the sentence embedding model (Devlin et al. 2019, Reimers and Gurevych 2019, Grootendorst 2020) to measure cultural fit. The sentence embedding model is a novel machine learning method to compare the semantic similarities of complete sentences. I estimate the cultural preferences of each individual manager by clustering speeches in earnings call transcripts on their similarity in semantic meanings. In addition, I calculate corporate cultural values based on all speeches in an earnings call session. I then construct two cultural fit measures. The first measures the cultural distance between firms and the second between individual managers and firms.

The corporate finance literature has documented various measures of corporate culture through textual mining or surveys (Fiordelisi and Ricci 2014, Guiso, Sapienza, and Zingales 2015, Graham et al. (2017), Li et al. 2020 and Li et al. 2021). In this paper, I follow Li et al. (2020) and use earnings call transcripts to score corporate culture. Specifically, I use the Q&A sessions in the earnings calls to mitigate the self-promotion of cultural values. Li et al. (2020) argue that a firm's current culture is most significantly shaped by its top managers. Furthermore, Guiso, Sapienza, and Zingales (2015) note that managers must share and follow a corporate cultural value to enforce it in the company. Therefore, speeches in earnings calls not only reflect managers' cultural preferences but also show how managers influence and shape the cultural values in firms. In other words, if most managers prefer the same cultural values, the corporate culture will plausibly be an amalgam of their shared cultural preferences. Equivalently, a manager is considered an outlier if his or her cultural preference is distant from the shared cultural preferences of other managers.

I rate each executive across the following nine cultural values: innovation, respect, integrity, quality, teamwork, control, competition, hard work and community by analyzing 11,582,429 managerial speeches during earnings call Q&A sessions over 2002-2020.[2] To measure the shared corporate culture value at the firm level, I aggregate all executives' speeches during an earnings call and rate each firm across the aforementioned nine cultural values. Next, I create nine-dimensional cultural vectors for managers and companies separately and construct two cultural fit measures.

The first cultural fit measure is related to cultural adaptation, which is considered a common process of environmental adaptation and deep rooted in human nature (Kim 1988, Anderson 1994, Kim 2017). Kim 2017 argue that the cultural adaptation process in new cultural environments typically begins with culture shock. Similarly, when a manager leaves a familiar corporate culture for a new one, the distance between the two corporate cultures is the proxy for managers' cultural adaptation costs and the level of cultural shock they may experience. Therefore, I calculate the first cultural fit measure as the Mahalanobis distance of the cultural vectors between two firms. Hereafter, I refer to the first cultural fit measure as the corporate cultural distance. The first firm is the previous firm that the manager left, and the second firm is the new firm that the manager joins after leaving the previous firm.[3] A key assumption is that the lower the corporate cultural distances, the lower the cultural adaptation costs and the better the cultural fit of managers at new firms. However, a drawback of the first cultural fit proxy is that it does not directly measure the cultural preferences of individual managers. A manager can move between two culturally close firms while being a poor cultural fit for both firms. Furthermore, the first cultural fit measure requires that a manager must have worked at

[2] The innovation, respect, integrity, quality and teamwork culture are defined and measured in Guiso, Sapienza, and Zingales (2015), as well as Li et al. (2020). They are shown to be related to mergers and acquisitions and firm performance. The hard work or community culture is only defined and measured in Guiso, Sapienza, and Zingales (2015) but not in Li et al. (2020). The control and compete culture variables are defined by Quinn and Rohrbaugh (1983) and measured in Fiordelisi and Ricci (2014) and are related to CEO turnover.

[3] In Mergers and Acquisitions (M&A) events, I consider the merged firm as the new firm that a manager joins.

two firms to have a non-zero measurement of cultural distance. The requirement restricts the sample size to less than 1795 unique managers.

Therefore, I propose a second cultural fit proxy, personal cultural distance, as the Mahalanobis distance between a manager's cultural vector and the company's cultural vector. While a manager's cultural vector reflects their personal preferences of cultural values, a company's cultural vector reflects the shared cultural values amongst other managers in the firm. The personal cultural distance essentially measures how different a manager's cultural preferences are from the corporate cultural values. Because the second cultural fit measure does not require managers to have worked at two firms, the sample size is ten times as large as what is available for the first cultural fit measure.There are 12,284 unique managers in the sample when proxying cultural fit with personal cultural distances.

Traditional literature in corporate finance has attributed CEO turnover to six main factors: performance (Weisbach 1988, Huson, Parrino, and Starks 2001, Jenter and Kanaan 2015, and Jenter and Lewellen 2021), compensation (Brookman and Thistle 2009), board monitoring (Goyal and Park 2002), risk (Bushman, Dai, and Wang 2010), asymmetric information (Dikolli, Mayew, and Nanda 2014) and corporate culture (Fiordelisi and Ricci 2014). However, the literature has yet to provide direct evidence to support or reject culture as a determinant of fit – in a survey conducted by Robert Walters, 81% of employers believe that candidates are less likely to leave when working for an organization where they are an excellent cultural fit. [4] To fill the gap in the literature, I add cultural fit as a seventh factor that affects the tenure of managers and the probability of turnover. To show that cultural fit impacts managerial tenure, I merge the culture data with Execucomp and calculate managerial tenure as the number of years until a manager exits the firm. However, the managerial tenure variable is right-censored. Therefore, I employ the Cox Proportional Hazards Model (PH) and conduct survival analyses. The PH models show that a worse cultural fit, represented by larger corporate culture distances between firms or larger personal cultural distances between firms and managers, reduces managerial tenure. To ensure that the survivorship results are not caused by formal institutions in firms, such as performance, executive compensation, or board monitoring, I control for firm fundamentals, managerial compensation, and board characteristics in all survival regressions.

Firstly, when the cultural fit is proxied by corporate cultural distances, the PH model shows that a 1 standard deviation increase in corporate cultural distance (between firms) is associated

[4] The survey can be found at https://www.robertwaltersgroup.com/content/dam/robert-walters/country/united-kingdom/files/whitepapers/Robert-Walters-Cultural-Fit-Whitepaper.pdf

with a 7% increase in the expected hazard ratio that hurts managerial tenure. To investigate if performance has an interaction effect with corporate cultural distances, I interact cultural distances with ROA and find that cultural distances between firms weaken the positive impact of ROA on managerial tenure. In other words, the negative relationship between cultural distances and tenure is exacerbated by better firm performance.

Secondly, when personal cultural distances (between managers and firms) represent the cultural fit, the PH model shows that a 1 standard deviation increase in cultural distances is associated with a 6% increase in the expected hazard ratio that hurts managerial tenure. Consistent with the organizational behaviour literature (O'Reilly, Chatman, and Caldwell 1991, and Vandenberghe 1999), the results suggest that the worse the cultural fit, the shorter the manager's tenure at the company. While ROA positively affects managerial tenure, there is no statistically significant interaction effect between personal cultural distances and ROA. Therefore, the interaction effect between firm performance and cultural fit disappears when I apply the second cultural fit measure to a larger sample of managers and firms.

Graham et al. (2017) show in their survey that 54% of executives would walk away from culturally misaligned targets during Mergers and Acquisitions (M&A) deals. Therefore, cultural misfits caused by M&As may be the main drivers that shorten managerial tenure in firms. As a robustness check, I obtain the M&A data from the SDC database and remove all executives that work for companies that have participated in M&A deals during 2006 – 2020. By employing the personal cultural fit measure, I show that the negative correlation between personal cultural distance and managerial tenure persists after removing all firms that have participated in M&A.

Next, I use causal survival forests (CSFs) to demonstrate the causality of cultural distances on managerial tenure. CSFs are an extension to causal forests, utilising machine learning to recover the unobserved counterfactual and estimate the causal effects of treatment. CSFs adjust for right-censorship in the data and, unlike PH models, do not make parametric assumptions of the covariates' functional forms. I create a treatment variable that equals one if the Mahalanobis distance between culture vectors exceeds its median value. Consistent with the PH models, CSFs estimate an economically significant negative impact of large cultural distances on managerial tenure. Specifically, a large corporate cultural distance (between firms) will result in at least 7 months shorter tenure in new firms. Similarly, a large personal cultural distance (between managers and firms) will lead to over 10 months shorter tenure in firms. Meanwhile, the treatment effect of cultural distance on managerial tenure is not homogenous. As executive compensation increases, the treatment effects of cultural distances on tenure become more negative, indicating that cultural fit on tenure is exacerbated by higher managerial compensation.

Should companies hire managers that are more culturally aligned or disruptive? The literature has documented a negative correlation between manager turnover and firm performance (Kaplan 1993, Brookman and Thistle 2009, Dikolli, Mayew, and Nanda, 2014, Li et al. 2021). Since cultural distance is positively correlated with managerial turnover, it is possible that cultural distance negatively impacts firm performance and values. To explore the lead-lag relationship between cultural distance and firm performance and values, I create a measure of firm-level cultural dispersion as the average personal cultural distances within the firm. I then run a Panel-OLS regression with next year's Tobin's Q as the dependent variable and the lagged cultural dispersion as the main independent variable. The results show that hiring managers that are culturally disruptive is negatively associated with firms' (future) market values. Furthermore, I show that firms with culturally aligned managers are associated with a higher future Return on Invested Capital and Earnings Per Share. Therefore, the evidence is consistent with the mechanism that cultural distance is positively correlated with managerial turnover, which is negatively correlated with firm performance.

How do investors perceive cultural fit signals sent by executives during earnings calls? While it has been documented in the literature that investors respond to earnings announcements (Da, Engelberg, and Gao 2011, Ben-Rephael, Da, and Israelsen 2017), it has not been clear if investors also respond to information in earnings calls that are not directly related to firm performance. Therefore, I construct portfolios that go long into the companies with the lowest cultural dispersion and short the companies with the highest cultural dispersion. The portfolio generates an annualized excess return of 2% over the Carhart four-factor model. Although an investor is unlikely to profit from the long-short strategy after transaction costs, the results still imply that hiring managers with better cultural fit could benefit firms' financial outcomes and stock market performance.

While I find that larger cultural distances have an economically significant and negative effect on managerial tenure, it is not clear how cultural fit relates to the traditional literature on the probability of executive turnover (Weisbach 1988, Huson, Parrino, and Starks 2001, Goyal and Park 2002, Brookman and Thistle 2009, Dikolli, Mayew, and Nanda 2014, Jenter and Kanaan 2015). Using logit regressions, I conduct a robustness check and show an increased probability of managerial turnover correlated with large cultural distances. In addition, I show that higher community culture can significantly reduce the probability of managerial turnover, while a higher integrity culture significantly increases the probability of managerial turnover.

Inconsistent with Fiordelisi and Ricci (2014), I do not find any significant impacts of the innovation or competition culture. There are two main factors that could lead to these differences in our results. First, I control for the cultural distances, or cultural fit, an omitted variable in Fiordelisi and Ricci (2014)'s analysis. Second, my corporate culture measure has higher precision and accuracy than the bag-of-words approach that Fiordelisi and Ricci (2014) use in their research.

The rest of the paper is organized as follows. Section 2 provides a literature review. Section 3 describes my data, including measuring corporate culture with sentence embeddings. Section 4 shows how corporate cultural distances affect managerial tenure. Section 5-6 show how personal cultural distances affect managerial tenure. Section 7 shows the causality of cultural distances on managerial survival time with causal survival forests. Section 8 explores the lead-lag relationship between average personal cultural distances and firm values. Section 9 shows the performance of long-short portfolios sorted on average personal cultural distances. Section 10 conducts robustness checks. Section 11 concludes.

2. Literature Review

Prior literature in finance employs three main methodologies to measure corporate culture. The first methodology relies on the bag-of-words approach (Loughran and Mcdonald 2011, Loughran and Mcdonald 2016, Fiordelisi and Ricci 2014, Guiso, Sapienza, and Zingales 2015, and Li et al. 2020). The approach creates dictionaries of words that are synonyms for cultural values. The approach then counts the frequencies of words in the dictionaries that also appear in the 10-K reports or earnings call transcripts to obtain the cultural measure. The second methodology relies on the survey approach (Graham et al. 2017), which survey a group of employees or executives on their opinions of corporate cultural values. The third approach uses the Hofstede (2001) framework to measure corporate culture based on the culture of nations (Beugelsdijk and Frijns 2010, Dodd, Frijns, and Gilbert 2015, Frijns, Dodd, and Cimerova 2016).

Early organizational behaviour literature has explored the relationship between person-organization-fit and employee turnover through surveys. O'Reilly, Chatman, and Caldwell (1991) surveyed approximately 200 MBA students and 93 accountants for their cultural preferences and measured their actual turnover 24 months after the survey. Using a set of key informants to assess the culture of firms, they show that person-organization-fit is positively correlated with an individual's staying with an organization. Vandenberghe (1999) finds consistent results to O'Reilly, Chatman, and Caldwell (1991) in the healthcare industry by surveying 630 health care professionals in 28 health care organizations and measuring their actual turnover 12 months after the survey. In comparison, the literature in finance is comparatively scarce. Fiordelisi and Ricci (2014) study the impact of corporate culture on CEO turnover and find that the probability of CEO change is positively influenced by competition- and creation-oriented cultures. They also find that the negative correlation between firm performance and CEO turnover is reinforced by control-oriented cultures and reduced by creation-oriented cultures. Companies with a teamwork-oriented (or collaboration-oriented) culture have a stronger inclination to change CEOs.

My paper is related to the strand of literature on corporate culture and firm behaviours. Guiso, Sapienza, and Zingales (2015) study the link between different dimensions of corporate culture (innovation, integrity, quality, respect, teamwork, safety, community, communication and hard work) and a firm's performance, as well as how different governance structures impact the ability to sustain integrity as a corporate value. They stress a positive and statistically significant correlation between the level of managerial integrity perceived by the employees and firm performance proxied by Tobin's q and return on sales. They also show a pronounced drop in firm integrity after going public and a positive correlation between CEO compensation and the integrity culture. However, unlike my paper, they do not find significant results related to teamwork and community cultures. Grieser et al. (2017) use a unique dataset from ashleymadison.com as a proxy for corporate culture and show that the integrity culture at the corporate level predicts firm-level unethical behaviour as well as innovation. Ji, Rozenbaum, and Welch (2017) find that firms with lower levels of culture and values are more likely to be subjected to SEC fraud enforcement actions and securities class action lawsuits. Li et al. (2020) analyse 209,480 earnings call transcripts and use Natural-Language-Processing to create an alternative corporate culture value measure that includes five dimensions: innovation, integrity, quality, respect and teamwork. They also show the correlation between corporate culture and business outcomes, risk-taking, earnings management, executive compensation design, firm value and deal-making.

In addition, my paper is the first to apply causal survival forests (CSF) for survival analysis in the corporate culture domain. Traditional models assume strong parametric assumptions on the underlying hazard and survival mechanisms. However, corporate culture is often subtle, non-linear and measured with lots of noise. CSFs deliver unbiased results without making strong parametric assumptions on the underlying hazard and survival mechanisms, an essential advantage in the corporate culture domain. While no clearly defined theory specifies the underlying survival or hazard mechanisms' functional forms regarding corporate culture, it has been common in the financial literature to assume a linear survival mechanism (Brookman and Thistle 2009). Any misspecifications of the functional forms of the linear survival mechanism will therefore bias the estimated coefficients. CSF alleviates such using a separate-hold-out dataset to estimate the effect of treatment on survival time.

My paper contributes to a relatively new strand of literature on corporate social responsibility (CSR). My measurement of the community culture, which is proposed by Guiso, Sapienza, and Zingales (2015) but not measured in Li et al. (2020), is similar to the concept of CSR, as the seed words include "community", "society", "environmental", "philanthropy" and "sustainability". Gaudencio, Coelho, and Ribeiro (2020) show that perceptions of CSR contribute to reducing turnover intentions, which may increase the survival time of executives. Furthermore, Orij et al. (2021) show that CSR reduces the likelihood of CEO turnover in general, increases the likelihood of CEO turnover in case of poor financial performance and greatly reduces the likelihood of CEO turnover in case of better financial performance. Therefore, consistent with the existing literature that managerial turnover is linked to CSR, I show that CSR is also an important contributing factor to managerial cultural fit which affects managerial tenure.

3. Data

My data consists of four primary sources: earnings call transcripts spanning from 2002 – 2020, manager characteristics from ExecuComp, board characteristics from BoardEx, and firm characteristics from Compustat.

3.1 Earnings call transcripts data

The primary component of my data is earnings call transcripts, which document the date, speakers' names, and their conversations during each earnings call. Each row in the data contains a single speech of an executive during the earnings call Q&A section. For example, on October 29, 2020, the first four speeches in the Q&A section of Apple Inc. contain the following excerpts: [5]

[1] Shannon Cross: "*Tim, can you talk a bit more about China …*"

[2] Tim Cook: "*Thanks, Shannon. If you look at China and look at last quarters – I'll talk about both last quarter and this quarter a bit …*"

[5] I only show the first few sentences of each speech for illustration purposes.

[3] Shannon Cross: "*Okay, great. And then can you talk a bit about just overall in the world -- the cadence that you see sort of for the 5G adoption launch ...*"

[4] Tim Cook: "*Yes. We're working hard to provide the best experience for our iPhone users. To do so, we've been collaborating closely with carriers all around the world to ensure iPhone has great throughput and coverage and battery and call quality …*"

There are 11,582,429 earnings call transcript speeches from 7859 unique firms from 2002 – 2020. A speech is typically short and contains either a question from the audience or an answer from the executive. On average, eight unique people speak at an earnings call. While Li et al. (2020) concatenate all questions and answers of everybody in the same earnings call into bulks of texts and match cultural values through a dictionary generated by Word2Vec, I allow SBERT to comprehend the semantic meanings of speeches from each executive without breaking sentences into words. To ensure that my SBERT model has as many examples as possible to learn the speech topics, I only merge the transcripts data with Compustat, Execucomp and BoardEx after obtaining the cultural values from the full transcripts data.

Since SBERT relies on the original sentences' semantic meanings, unlike Li et al. (2020), no word-pre-processing is necessary before applying SBERT. There is no need for Named Entity Recognition and Removal or lemmatisation. There is also no need to remove punctuation marks, stop words or single-lettered words from the sentences. I keep everything in its original format, preserving as much information as possible regarding the semantic meaning of the texts.

3.2 Sentence-BERT embeddings

SBERT is a variant of a new Machine Learning model called BERT proposed by Field Devlin et al. (2019). BERT is a pre-trained neural network architecture designed to handle sequential input sentences with an attention mechanism that provides context for any position in the input sentence. However, the original BERT model is unsuitable for unsupervised topic modelling. It requires each pair of sentences to be fed into the network to compute sentence similarity. Instead, I employ a variant of BERT called Sentence-BERT proposed by Reimers and Gurevych (2019), modifying the pre-trained BERT network that uses Siamese and triplet network structures to derive semantically meaningful sentence embeddings that can be compared using cosine similarity. The idea of sentence-BERT is to embed each sentence into a vector. Semantically similar sentences have vectors pointing approximately in the same direction and hence a high cosine similarity in the vector space.

The sentence-BERT model I use can digest input texts with up to 256 words, usually enough for the main body of Q&A conversational speeches in earnings call transcripts.[6] Texts longer than 256 words are truncated. Each conversational speech is converted to a vector of 384 dimensions, and the resulting earnings call data has a matrix of (11,582,429, 384) dimensions.

3.3 Supervised UMAP Manifold Learning and HDBSCAN clustering

The dimensionality of the earnings call matrix is too large for any modern computing cluster – there are 11,582,429 vectors of 384 dimensions. Therefore, I employ an algorithm in Machine Learning called Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP), a dimensional reduction mechanism proposed by McInnes, Healy, and Melville (2020), to reduce the dimensionality of the matrix from 384 to 5. UMAP searches for a low-dimensional projection with the closest possible fuzzy topological structure. The process is firmly supported by mathematical proofs in topology (McInnes, Healy, and Melville, 2020).

The next step is to cluster conversational speeches of similar semantic meanings into the same cluster. Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) offers significant performance advantages over K-meanings clustering, as it does not require ex-ante specifications of clusters and offers outlier detection that detects clusters of varying densities. The HDBSCAN algorithm returned 8558 unique topics containing 6,032,058 transcript conversations, with the other transcript conversations classified as noise topics. The largest topic cluster contains about 7000 speeches, while the smallest cluster contains about 100 speeches. [7]

3.4 TF-IDF topic extraction

After obtaining the topic clusters, the next step is to extract topic representations from each topic cluster. A commonly used methodology for topic extraction is TF-IDF. TF represents Term Frequency, the word frequency in a document, and IDF represents Inverse Document Frequency, which is the inverse frequency of documents with the word in its corpus. TF-IDF multiplies term frequency by inverse document frequency, and the advantage for doing so is that it up-weights the importance of rare words and down-weights the importance of frequently appearing words. I concatenate all conversations within a culture topic cluster and view each

[6] The details of the pre-trained sentence-BERT model can be found in Internet Appendix 1. The Internet Appendix is available online at: https://ginward.github.io/culture_appendix.pdf.
[7] The details of the UMAP and HDBSCAN models can be found in Internet Appendix 1.

cluster as a single document. I then count the number of words and bi-gram occurrences within a topic cluster, compute the TF-IDF values across all topic clusters and extract words or bi-grams with the top TF-IDF scores as the topic representation. In addition, to ensure that the topic representation is related to cultural topics as much as possible, I follow Grootendorst (2020) and up-weight the TF-IDF score by 1.2 if the word appears as a seed word in Internet Appendix A.2. Finally, I obtain topic representations for each of the 8558 unique topic clusters.

3.5 Search for topics related to corporate cultural values

Although I nudge both the UMAP and TF-IDF algorithms towards our cultural values, they do not strictly enforce that the generated topics are cultural. Therefore, I follow Grootendorst (2020) by searching through the topics for the seed words shown in Internet Appendix A2. For each culture topic, I concatenate the seed word into a sentence, such as "innovation innovative innovate invention inventive invent". I then encode the sentence into a vector using sentence-BERT and compare the seed vector with the topic vector. The topic vectors are created by concatenating topic representations and encoding the sentence with sentence-BERT. I extract the top 20 topics most similar to each culture value. If a topic appears to be similar to more than one cultural value, I assign the culture value with the highest similarity to the topic. The topic representations for each cultural value are shown in Internet Appendix A3.

3.6 Measuring cultural values and distances

In this paper, I measure corporate cultural values at both the firm and manager levels. To resolve any ambiguity in terms, I refer to the cultural values measured at the firm level as *corporate cultural values* and cultural values measured at the manager level as *personal cultural preferences*. When measuring corporate cultural values, I aggregate all managers' speeches at the firm level for each year and count the frequency of speeches that belong to a specific cultural topic category. I then normalise the frequency by the total number of speeches at the company during the same year and obtain the cultural values for the topic category. In Internet Appendix A4, I create lists of top-ranked S&P500 firms by various corporate cultural values during 2002 – 2006, 2007 – 2012 and 2013 – 2020. For example, Amazon showed a top quality culture during 2013-2020. When measuring personal cultural preferences, I aggregate all speeches at the manager level for each year and count the frequency of speeches belonging to each manager's cultural topic category. In Internet Appendix A5, I rank the top managers affiliated with S&P500 firms by personal cultural preferences. For example, Andrew Mcnellis from TripAdvisor Inc. was the leading executive prioritising quality culture from 2013-2020.

For both corporate cultural values and personal cultural preferences, I construct culture vectors consisting of nine dimensions:

$$u_{i,t}=\begin{bmatrix} Innovation_{i,t} \\ Integrity_{i,t} \\ Quality_{i,t} \\ Respect_{i,t} \\ Teamwork_{i.t} \\ Community_{i,t} \\ Hardwork_{i,t} \\ Control_{i,t} \\ Compete_{i,t} \end{bmatrix}$$

where *i* indexes firms or managers, and *t* indexes the time.

There are two cultural distance measures in this paper. The first cultural distance measure approximates the cultural distance between companies, while the second one approximates between managers' cultural preferences and corporate culture. Hereafter, I denote the cultural distance between companies as *Corporate cultural distance* and between managers' cultural preferences and corporate culture as *Personal cultural distance*. In Internet Appendix A6, I rank top executives by personal cultural distances. For example, Brandy Burkhalter from Centene Corp is the top executive whose cultural values differ the most from the firm's cultural values from 2013-2020.

While the traditional literature (Beugelsdijk and Frijns 2010, Dodd, Frijns, and Gilbert 2015, Frijns, Dodd, and Cimerova 2016, Karolyi 2016, Li et al. 2020) commonly use standardised Euclidean distance as the metric for cultural distances between individual board members in a company, standardised Euclidean distance assumes zero correlation between cultural dimensions. The strong assumption on zero correlation between cultural dimensions may overweight cultural dimensions with high correlations and underweight cultural dimensions with low correlations. To relax the assumption of zero correlation between cultural dimensions and make sure my results are robust, I employ the Mahalanobis Distance, which is widely used in the econometrics and machine learning literature (Mitchell and Krzanowski 1985, Neudecker 1997, Xiang, Nie, and Zhang 2008).

The first cultural distance measure, corporate cultural distance, is calculated as follows:

$$Corporate\ cultural\ distance = \sqrt{(u-v)V^{-1}(u-v)^T}$$

where u is the cultural vector of the previous company, and v is the cultural vector of the new company. I only include managers who have switched firms at least once. V^{-1} is the inverse of the 9×9 dimensional covariance matrix of the cultural vectors. Figure 1 Panel A shows the time-series variation of average cultural distance against managerial tenure. As managerial tenure increases, the volatility and the magnitude of cultural distances (cultural fit) both decrease (increase). This evidence is consistent with Kim (2017), who argue that cultural adaptation improves over time.

In Figure 2, the graph on the left compares two firms for David Anderson, who left Nielson Holdings PLC in 2019 for Alexion Pharm Inc. in 2017. Nielson Holdings PLC's culture emphasizes mostly teamwork and community, while Alexion Pharm Inc.'s culture emphasizes mostly innovation and respect. The corporate cultural distance between the two firms is 5.98. Similarly, the graph on the right compares two firms for Oscar Munoz who left United Airlines Holdings Inc. in 2019 for CSX Corporation in 2015. United Airlines Holdings Inc.'s culture is strongest in quality and community, while CSX corporation's culture is strongest in hard work and control. The corporate cultural distance between the two firms is 2.72.

The second cultural distance measure, personal cultural distance, is calculated as follows:

$$Personal\ cultural\ distance = \sqrt{(u-v)V^{-1}(u-v)^T}$$

where u is the manager's 9-dimensional cultural preferences vector, and v is the average 9-dimensional corporate cultural vector. V^{-1} is the inverse of the 9×9 dimensional covariance matrix of the cultural vectors. I calculate the cultural distance for each manager at the annual frequency. Consistent with the previous evidence that cultural fit increases with managerial tenure, Figure 1 Panel B shows that the volatility and magnitude of average cultural distance both decrease over time.

In Figure 3, the graph on the left compares the corporate culture of Columbia Sportswear Company to the cultural preferences of Timothy Boyle in 2019. Columbia Sportswear Company's culture places the highest emphasis on quality, while Timothy Boyle emphasized teamwork. The personal cultural distance between the manager and the firm is 1.06. The graph on the right compares the corporate culture of Salesforce.com Inc. to the cultural preferences of Marc Benioff in 2019. The corporate culture and personal cultural preferences are similar, and therefore, the personal cultural distance is 0.31, which is relatively smaller.

3.7 Match corporate cultural values with Compustat, BoardEx and Execucomp

Next, I calculate each company's annual asset growth, sales growth, R&D/sales ratio and sales/invested capital ratio with the data supplied by Compustat. I follow Li et al. (2020) and map the SIC industry codes supplied by Compustat to Fama-French 12-industry classification. Crucially, Execucomp provides the compensation and employment data on the five highest-paid executives in a firm per fiscal year. Execucomp provides four different variables on executive compensation: salary pay, bonus pay, stock awards and option awards. I classify those four managerial compensations into two subcategories – incentive pay and non-incentive pay. Incentive pay includes salary and bonus pay, while non-incentive pay includes stock awards and option awards. Execucomp also provides other managerial characteristics that determine the managerial tenure (survival time), such as age and gender. To control for forced turnovers, I obtain databases of forced CEO turnover from Gentry et al. (2021) and Peters and Wagner (2014). Finally, BoardEx provides board characteristics data, such as board attrition rate, the board male-gender ratio, the average number of board directors' qualifications, and the average network size of board directors.

I create two separate datasets by merging corporate cultural values and personal cultural preferences separately with Compustat, BoardEx and Execucomp. The first dataset measures culture at the firm level and includes only managers who have switched firms. In contrast, the second measures culture preferences at the manager level and consists of all managers that have mentioned cultural values in their speeches in earnings calls. The first dataset includes 1791 unique managers, while the second consists of 12,284 unique managers.

To create the first dataset, I match corporate cultural distances (between firms) with Compustat and Execucomp by gvkey and year. As the corporate cultural distance measure requires managers to have worked in at least two firms that both have culture measurements, to increase the number of observations and better measure corporate cultural distances between firms, I fill empty corporate cultural values with their closest non-zero values. I first fill missing values with companies' most recent past cultural values. If there are still missing values left, I further fill missing values with companies' closest future cultural values. Next, I match the data with BoardEx by ticker symbol and year and obtain the final data for analysis. Table 1 Panels A and B report the summary statistics. The average age of a manager in my sample is 52 years old. 88% of managers are male, and 12% are female. The first dataset spans from 2002 to 2020.

To create the second dataset, I match personal cultural distances (between managers and firms) with Execucomp by the executive's full name and the fiscal year. First, I extract each executive's name from earnings call transcripts and match with Execucomp by the executive's name, the firm's gvkey and the year. To avoid ambiguity in names caused by titles, prefixes, middle names and letter cases, I follow Gulbranson (2021) and extract the first and last names from managerial names using a set of regex rules and the U.S. Census Surname Data, the Naming Practice Guide U.K., Wikipedia Anthroponymy, Wikipedia Naming Conventions and Wikipedia List of Titles. Second, I calculate each executive's personal cultural preferences. I then match the Execucomp data with Compustat by "gvkey" and year. When calculating personal cultural distances, I do not fill in missing cultural values, given that there is a sufficient amount of data available. After calculating the cultural distance, I keep only the managers who at least mentioned one of the cultural values during the year. Finally, I match with the BoardEx data by company ticker and year. The second dataset spans from 2006 to 2020.

Table 1 report the summary statistics. The number of managers in the second dataset is approximately 10 times the number in the first dataset – there are a total of 12,284 unique managers, of which 1086 are females. The average age in the second dataset is 55 years, which is slightly older than the average age in the first dataset.

3.8 Calculating managerial tenure (survival time) in firms

Although Execucomp has "joined_co" and "left_co" variables that indicate the date that an executive joins and leaves the firm, many of the dates are empty. Therefore, some imputation is necessary to infer how many years a manager has served in a firm and if a manager has left the firm.

For those executives with non-empty "joined_co" and "left_co", I calculate executives' tenure as the number of years between "joined_co" and "left_co" and set the censoring indicator $exit = 1$. For those executives with empty "joined_co" but non-empty "left_co", I calculate executives' tenure as the number of years between the earliest year that the executive appeared in the company's annual proxy and "left_co", and I set the censoring indicator $exit = 1$. For executives with empty "left_co", I calculate executives' tenure as the number of years between the earliest and last dates the executive appears in the company's annual proxy. I set the censoring indicator $exit = 1$ if the maximum year that a manager appears in the firm's proxy statement is less than the latest year that the firm files proxy statements. To ensure that I only capture managerial survival but not firm survival, I only keep the managers who have exited the firm before 2020 or who serve at a firm that still files annual proxy statements until 2020. I do not consider executives that serve at two companies simultaneously or executives that left a firm but re-joined later. Table 1 Panel A shows that the average survival time of managers who switched firms is 4.35 years. Panel B shows that the average survival time of all managers who have mentioned cultural values during earnings calls is 8.52 years.

3.9 Why measure corporate cultural values with SBERT but not Word2Vec or Bag-of-words?

There are two reasons I use SBERT instead of the traditional Word2Vec or bag-of-words approaches.

First, a word's context can completely change its meanings. In the Word2Vec approach that Li et al. (2020) use to measure corporate culture, every word is mapped to a unique vector. A naïve search of similar words will completely ignore the semantic meanings of sentences and introduce ambiguity in the resulting bag-of-words dictionary for each cultural value. Compared to the word vector approach by Li et al. (2020), which embed each word into the vector space, sentence vectors can better reflect the semantic meanings of speeches. For example, with a word vector approach, the word “excellent” in the following two phrases will have the same embedding: “an excellent technology” and “an excellent person”, as each word can only correspond to one unique vector in the vector space. The two phrases have very different semantic meanings even though they contain similar words. “An excellent technology” delivers information on innovation culture, while “an excellent person” delivers information on respect culture. Suppose one compares the cosine similarity of the word “excellent” in both phrases, the model will output a cosine similarity of 1 for both words in both phrases. However, the semantic meanings are dramatically different. Meanwhile, sentence-embeddings of “an excellent technology” and “an excellent person” only have a low cosine similarity of 0.60. 28,018 transcripts include the word “excellent”, and it is difficult to identify which “excellent” words deliver information on the company's innovativeness.

Second, Word2Vec’s search of similar words sometimes returns frequent words that co-occur with but are not directly linked to the seed word. For example, in Li et al. (2020)’s dictionary for innovation culture, the word “customer_experience”, “restaurant_experience”, “organisations”, “workspace” and many more ambiguous words not directly related to innovation are returned by the Word2Vec model and categorised into the innovation culture category. However, both “customer_experience” and “restaurant_experience” are more related to the quality culture, while “organisations” is a little ambiguous but could relate to teamwork more closely than innovation. Most importantly, the word “organisation” appeared frequently in 83,500 earnings call transcripts 192,000 times, which increases the measure of innovation culture significantly for companies that mention the word "organisation" in the earnings call. Since the final regression output relies heavily on the machine learning measures of cultural values, if the measurement of cultural values itself is severely biased, the OLS efficiency and unbiasedness properties are no longer guaranteed.

4. Do corporate cultural distances influence managerial tenure?

As the first step in exploring the relationship between cultural fit and managerial turnover, I proxy managerial cultural fit with corporate cultural distances between companies. Specifically, I employ corporate cultural distance as the primary independent variable and manager tenure as the dependent variable of interest. In this section, I restrict my sample to those managers who have had at least one turnover, so that there are non-empty measurements of corporate cultural distances. As corporate culture is dynamic and evolving, I next create a dynamic corporate cultural distance measure by calculating the distance between the corporate culture at the new and old firms.

I follow Therneau, Crowson, and Atkinson (2022) and employ the Cox Proportional Hazard Model that allows for time-dependent covariates.[8] First, I create time-dependent cultural distances by calculating the difference between the corporate cultural values at the current company in the current year and the corporate cultural values at the previous company in the year that the executive left. Time-dependent firm, manager and board characteristics are created similarly. The PH model with time-dependent coefficients can be written as follows:

$$\log h(T_{i,n,t}) = \log h_0(T_{i,n,t}) + \beta_0 + \beta_1\, Corporate\ cultural\ distance_{n,m,t-1}$$
$$+ \omega C_{i,n,t-1} + \mu M_{i,n,t-1} + \psi F_{i,n,t-1} + \lambda B_{i,n,t-1} + \delta_{i,n,t-1} + W_{i,n,t}$$

where $h(T)$ is the hazard function determined by covariates, $h_0(T)$ is the non-parametric baseline hazard function, *i* indexes managers, *n* indexes the firm that the manager moved to, m indexes the firm that the manager exited, and *t* indexes time. *C* is a vector of average cultural values at the new firm. *M* is a vector of average manager control variables, including compensation. *F* is a vector of firm control variables, including the firm characteristics. *B* is a

[8] While there are some earlier applications of the Accelerated Failure Time (AFT) model with dynamic covariates in the management literature (Brookman and Thistle 2009), the medical sciences literature has gradually switched to Cox Proportional-Hazard (PH) model when it comes to dynamic covariates. The main reason is that the AFT model makes strong assumptions on the underlying likelihood function when the covariates are dynamic (see Tseng, Hsieh, and Wang 2005, Broström 2012), while the likelihood function in the PH model extends naturally to dynamic covariates under normal conditions (Therneau, Crowson, and Atkinson 2021). Therefore, when measuring dynamic cultural distances between firms, I employ the PH model with time-varying coefficients in this paper.

vector of average board control variables, including the board characteristics. δ is the industry fixed effects. W is the error term. I cluster standard errors at the manager level.

Next, I code time-dependent covariates using time intervals based on their updating frequency. I follow Therneau, Crowson, and Atkinson (2022) and delay cultural distances, managerial characteristics, firm characteristics and board characteristics by one year to avoid reverse causality. Finally, I apply the equivalent Cox partial likelihood in estimating the model coefficients. [9]

Figure 4 plots the baseline survival probabilities against time in the PH models. The red line shows the average survival probabilities when the corporate cultural distances are less than the median. The black line indicates the average survival probabilities when the corporate cultural distances are more than the median. As the red line is constantly below the black line, the survival probabilities are systematically reduced if the corporate cultural distances are larger than the median value.

The regression results are shown in Table 2. In a PH model, a hazard ratio can be computed as e^{β} for each coefficient. Suppose the hazard ratio is greater (less) than 1, or equivalently if β is greater (less) than 0, the hazard increases (decreases) and the length of survival decreases (increases). Table 2 Columns 1-3 show the PH model regressions results. Table 2 Model (1) shows the results after controlling for the logarithmic managerial total compensation, including incentive and non-incentive pay. Table 2 Model (2) and (3) show the results if I control for the logarithmic incentive and non-incentive compensation respectively. All three models show significant and positive coefficients of cultural distance, controlling for managerial compensation, firm and board characteristics differences. Model (1) suggests that a 1 standard deviation (2.5) increase in the cultural distance between companies can lead to 1.07 ($= e^{0.0274\times 2.5}$,) times higher risk of tenure termination. Model (2) and (3) show approximately equal coefficients of the same sign and significance. Therefore, a higher cultural distance significantly increases the hazard by a large positive factor, reducing managers' tenure (survival time) at the new firm. Thus, the Cox PH model with time-varying covariates confirms my earlier hypothesis that large cultural distances are associated with shorter managerial tenure.

Previous literature has shown that firm performance is linked to executive turnover (Coughlan and Schmidt 1985, Warner, Watts, and Wruck 1988). Therefore, I interact the ROA difference between the new and previous firms with cultural distances and the results are shown

[9] A fundamental assumption of the PH model is that the hazards are proportional. To show that assumption is not violated, I plot the time-dependent coefficient, $beta(t)$, for each covariate in Internet Appendix A7.

in Table 2 Columns 4-6. Consistently, I find that a higher ROA reduces managers' risk of termination. The results also show that the interaction effect between firm performance and cultural distances is statistically significant. In all models, the positive impact of performance on managerial tenure is weakened by cultural distances. Equivalently, the negative effect of cultural distances on managerial tenure is strengthened by stronger firm performance.

5. Do executives' personal cultural distances affect managerial tenure?

So far, I have explored the effect of cultural distances between firms on managerial survival time at the new firm. However, given that the number of managers who have switched firms at least once is small, the previous results are based on a small sample. This section measures an executive's personal cultural fit as the distance between a manager's cultural preferences and corporate culture. The sample size is approximately ten times as large as the previous sample, where I measure cultural fit as corporate cultural distances between firms.

I apply the Cox Proportional Hazard model with time-varying covariates. I follow Therneau, Crowson, and Atkinson (2022) and delay cultural distances, managerial characteristics, firm characteristics and board characteristics by one year to avoid reverse causality. The PH model can be written as follows:

$$\log h\left(T_{i,n,t}\right) = \log h_0\left(T_{i,n,t}\right) + \beta_0 + \beta_1 Personal\ cultural\ distance_{i,\ n,t-1}$$
$$+ \omega C_{i,n,t-1} + \mu M_{i,n,t-1} + \psi F_{i,n,t-1} + \lambda B_{i,n,t-1} + \delta_{i,n,t-1} + W_{i,n,t}$$

where $h(T)$ is the hazard function determined by covariates, $h_0(T)$ is the non-parametric baseline hazard function, *i* indexes managers, *n* indexes the firm, and *t* indexes time. All other variables are as defined in the previous section. I cluster standard errors at the manager level.

The Cox PH coefficients are shown in Table 3 Model 1-3. The results suggest that the larger the distance between a manager's cultural preferences and the corporate culture, the lower the managerial survival time. Model 1-3 show a hazard ratio of $e^{0.0221}$. If the Mahalanobis cultural distance increase by 1 standard deviation (2.5), which is the sample average, the expected hazard is 1.06 times higher. In other words, the larger a manager's personal cultural distance from the company, the higher the expected hazard of the manager when staying in the firm.[10] Furthermore, Figure 4 shows that the survival probabilities when the personal cultural distances are higher than the median value are systematically lower than when the personal cultural distances are lower than the median value. This evidence is

[10] In untabulated regressions, I test the significance of *cultural distance*^{2} in the survival regressions. The squared term is insignificant.

consistent with the organizational behaviour literature (O'Reilly, Chatman, and Caldwell 1991, and Vandenberghe 1999), which posit that person-organization-fit matters in employee turnover. Additionally, the results show that a higher community culture or a higher hard work culture increases managerial survival time. Consistent with Fiordelisi and Ricci (2014) who find that higher competition increases the probability of executive turnover, my results show that a higher competition culture decreases managerial survival time.

Next, I explore if cultural distances interact with firm performance when affecting managerial survival. Table 3 Models 4-6 show that the interaction term between cultural distances and firm performance is insignificant, suggesting that managers' personal cultural distances do not interact with firm performance. Although the interaction effect is significant for the corporate cultural distance measure, the statistical significance disappears when I apply the personal cultural distance measure to a much larger sample of firms and managers. Therefore, it is likely that firm performance and cultural fit are two independent contributing factors that affect managerial tenure.

6. Is the relationship between cultural distances and managerial tenure driven by mergers and acquisitions?

According to the survey by Graham et al. (2017), 54% of executives would walk away from culturally misaligned targets after mergers and acquisitions (M&A). I do not exclude turnovers induced by M&A in my main results for two reasons. The first reason is that M&A also cause cultural misfits, and therefore, their effect on executive tenure also falls within the scope of this paper. The second reason is that the number of firms that were involved in M&A during 2006-2020 is large and removing executives from those companies will reduce the sample size to a great extent. Nevertheless, I next employ a robustness check by filtering out all companies that have participated in any M&A deals during 2006-2020. First, I obtain the acquirer and target companies' CUSIPs from the Refinitiv SDC database. Second, I delete all executives that work for firms whose CUSIPs have matches in either the acquirer or the target database. This removed 2048 unique companies and 11296 managers. The final sample contains 176 unique firms and 988 managers. I employ the second cultural fit measure – personal cultural distance – and re-run the PH model with time-varying coefficients.[11] The results are shown in Table 4. Consistently, the coefficient of personal cultural distance is

[11] I do not employ the first cultural fit measure – corporate cultural distance – because the number of executives that satisfy the following two criteria at the same time is tiny: 1) they have switched firms for at least once in the past; 2) for all the firms that the executives worked for, none of the firms have ever participated in a M&A deal.

positive and significant, which indicates that an increase in personal cultural distances is associated with a decrease in managerial tenure, excluding all companies that have participated in M&A deals. Therefore, the relationship between cultural distances and managerial tenure is not driven by Mergers and Acquisitions alone.

7. Is the relationship between cultural distance and survival time causal?

Given the large number of covariates that I control for in the PH models and a lack of theoretical support of the additive linear functional form inside the survival function, the log-linear PH models are prone to functional form misspecification, which might lead to biased estimates of the effects of cultural distances on manager survival time. To complement the PH models and to establish causality between cultural distance and managerial survival time, I next apply a new econometrics tool in the literature, causal survival forests, which estimate the conditional average treatment effects of a binary treatment on a right-censored dependent variable without making parametric assumptions on the functional form.

7.1 What is causal survival forest?

Athey and Imbens (2016) and Athey, Tibshirani, and Wager (2019) firstly propose causal forests, an ensemble of decision trees that adds proven consistency and asymptotic normality to the random forests in the Machine Learning literature. Causal forests utilise a technique called "honest estimation", which distinguishes itself from traditional random forests by using only the first half of the subsample for splitting and using only the second half of the subsample for populating the leaf nodes and estimating the treatment effects, which effectively reduces bias in tree predictions. This methodology is akin to letting the machine construct a non-parametric model with half of the data and estimating the treatment effects with the constructed model on the other half of the data. A causal forest then recursively partitions the data into its leaf nodes by splitting on the covariates with the objective to maximise the treatment effect heterogeneity in its leaf nodes. The process can be metaphorically seen as matching each subject in the control group to a similar subject in the treatment group and calculating their outcome differences as the main treatment effect.

Cui et al. (2020) build on the causal forests literature and propose a new methodology called causal survival forests, which allows the dependent variable to be right censored by adjusting causal forests estimations with censoring probabilities. Causal survival forests estimate the following estimator:

$$\tau(X)=E[Survival\ Time(\omega=1) - Survival\ Time(\omega=0)\ |\ X=x]$$

where $Survival\,Time(\omega = 1)$ is the survival time for the treated group and $Survival\ Time(\omega = 0)$ is the survival time for the control group. I define the cultural distance between the new and old firms when the manager moves as large if the distance is larger than its median value. I consider those managers who move between firms with large cultural distances as the treated group ($\omega = 1$) and those managers who move between firms with small cultural distances as the control group ($\omega = 0$):

$$\omega_{culture}=1,\ when\ cultural\ distance > Median\ (cultural\ distance)$$

$$\omega_{culture}=0,\ when\ cultural\ distance \leq Median\ (cultural\ distance)$$

X is the list of covariates that I control so that the following conditions are satisfied:

$$\{Survival\ Time(\omega=1)\ ,\ Survival\ Time(\omega=0)\}\ \perp \omega\ |\ X$$

The above assumption is the unconfoundedness assumption, which states that conditioning on covariate vector X, treatment ω is independent of the outcome variable, which is survival time in my case. Cui et al. (2020) prove that the assumption is satisfied if X includes all prognostic factors used to determine the treatment ω.

Furthermore, as pointed out by Cui et al. (2020), when there are a large amount of near end-time censored subjects, the following positivity assumption should apply:

The survival time T_i is bounded from above by $T_{max} \in R_+$ almost surely, and $P[C_i < T_{max}|X_i, W_i] \leq 1 - \eta_c$ for come $\eta_c > 0$.

The above assumption is required for the nonparametric identification of conditional expectations of T_i. In the case when the treatment is corporate cultural distance, this means that I would need to redefine the maximum survival time to 12 years, as Figure 5 shows that most managerial survival time is censored after 12 years (i.e., most managers are not observed to leave the firm after 12 years). Therefore, I set T_{max} to 12 years and redefine my treatment effect estimator as the treatment effect on survival time conditional on a maximum survival time of 12 years so that the estimated censoring probabilities all lie in a reasonable range. Similarly, when the treatment is personal cultural distance, I redefine the maximum survival time to 18 years to ensure that the above assumption is satisfied, which is shown in Figure 5.

7.2 Is the relationship between cultural distances and managerial tenure causal?

First, I explore the causal impact of cultural distances on tenure. Following previous sections, I employ two measures of cultural fit – corporate cultural distances (between firms) and personal cultural distances (between managers and firms). To convert my data from the panel structure to the cross-sectional structure for causal effects estimations, I calculate average corporate cultural values and average personal cultural preferences during managers' tenure in firms. I then follow the same procedures as in previous sections and calculate corporate cultural distances between firms and personal cultural distances between managers and firms. I control for the average firm characteristics, managerial characteristics, and board characteristics. Next, I build 5000 trees in each causal survival forest.

Figure 6 plots the predicted survival probabilities against time in the causal survival forests models. As tenure increases, the survival probabilities decrease. The red line shows the average survival probabilities when cultural distances are less than the median. The black line indicates the average survival probabilities when cultural distances are more than the median. The red line is constantly below the black line in both cultural fit measures, suggesting that the survival probabilities are systematically reduced if cultural distances are larger than the median value.

Table 5 shows the average treatment effects (ATE). Table 5 panel A Columns 1-2 show the ATE when using corporate cultural distances and personal cultural distances as the proxies for cultural fit, respectively. Consistent with my PH models, cultural distances have a negative and significant causal effect on managerial tenure. Therefore, the relationship between cultural fit and tenure is correlative and causal. The effects are also economically significant. Having large corporate cultural distances will lead to a significant decrease of tenure by 7 months (12×0.6). Similarly, having large personal cultural distances will lead to a significant decrease of tenure by 10 months (12×0.8).

7.3 Is the impact of cultural fit heterogeneous across managers?

To understand which covariates matter in determining the heterogeneity of treatment effects, I show the variable importance measure in Table 5 panel B Columns 1-2. The variable importance measure is the weighted sum of the number of times each feature was split.[12] Managerial compensation, among other variables, is an important determinant of the treatment effect heterogeneity. Next, I plot the average treatment effects conditional on executive

[12] To further understand what the decision trees look like, I select the decision with the lowest error after pruning, also called the best tree, and show the tree in Internet Appendix A8.

compensation in Figure 7. The plots suggest that the causal impact of cultural distances on managerial tenure gets more negative as managerial compensation increases. In other words, higher compensated executives' tenure will be more adversely impacted by a bad cultural fit.

8. How are culture distances related to firm values?

I have established that cultural fit is positively associated with managerial tenure. Meanwhile, the literature has documented a negative correlation between manager turnover and firm performance (Kaplan 1993, Brookman and Thistle 2009, Dikolli, Mayew, and Nanda, 2014, Li et al. 2021). Therefore, I hypothesize that cultural fit is also positively associated with firm performance. To test the hypothesis, I explore the relationship between the previous year's cultural distances and the next year's firm performance and market values. My main independent variable of interest, cultural dispersion, is calculated as the average personal cultural distance among all managers in the firm each year:

$$Dispersion_{n,t}=\sum_{i \in firm\ n\ in\ year\ t} \frac{Personal\ cultural\ distance_{n,i,t}}{N_{n,t}}$$

where i indexes managers, n indexes firms and t indexes the year. N is the number of managers in each firm. The higher the cultural dispersion, the lower the average cultural fit within the firm.

First, I explore if hiring culturally disruptive managers hurt firm values. I run a panel-OLS regression where all independent variables are lagged by one year:

$$Q_{n,t}=+\beta_0+\beta_1 Dispersion_{n,t-1}+\omega C_{n,t-1}+\psi F_{n,t-1}+\lambda B_{n,t-1}+\delta_n+\gamma_{t-1}+W_{n,t}$$

where i indexes managers, n indexes the firm, and t indexes year. I follow Brookman and Thistle (2009) by using the annual Tobin's Q as the dependent variable:

$$Q=\frac{Total\ Asset+Common\ Shares\times Annual\ Closing\ Price\text{-}Total\ Common\ Equity}{Total\ Asset}$$

δ and γ are firm and year fixed effects, respectively. All other variables are as defined in the previous section. W is the error term. I cluster standard errors at the manager level.

The results are shown in Table 6 Column (1). An increase of 1 standard deviation (2.5) in the cultural dispersion is associated with a 0.1 decrease in next year's Tobin's Q. The effect is statistically significant. In economic terms, for an average company with an average book value of \$21 billion, an increase of 1 standard deviation in the cultural distance will lead to an over \$2.1 billion decrease in the market value. Therefore, the evidence suggests that hiring managers with bad cultural fits is associated with lower future firm values. In addition, a higher community culture is significantly and positively correlated with future firm values. This evidence is consistent with the literature on corporate social responsibility (CSR) that CSR creates firm value through various channels (Servaes and Tamayo 2013, Buchanan, Cao, and Chen 2018, Bardos, Ertugrul, and Gao 2020).

Second, I explore if hiring culturally aligned managers benefit firm performance. I run a panel-OLS regression where firm performance is the dependent variable and the lagged cultural dispersion is the main independent variable:

$$Performance_{n,t}=+\beta_0+\beta_1 Dispersion_{n,t-1}+\omega C_{n,t-1}+\psi F_{n,t-1}+\lambda B_{n,t-1}+\delta_n+\gamma_{t-1}+W_{n,t}$$

where i indexes managers, n indexes the firm, and t indexes year. Performance is proxied by Return on Invested Capital (ROI) or Earnings Per Share (EPS):

$$ROI=\frac{Net\ Income}{Invested\ Capital}$$

$$EPS=\frac{Net\ Income}{Common\ Shares}$$

All other variables are as defined in the previous section. I cluster standard errors at the manager level.

The results are shown in Table 6 Columns 2-3. In Column (2), the coefficient of cultural dispersion is negative and significant. An increase of 1 standard deviation (2.5) in the cultural dispersion is associated with a 0.75 decrease in next year's ROI. In economic terms, for an average company with an average invested capital of $13 million, an increase of 1 standard deviation in cultural dispersion will lead to an over $10 million decrease in the net income. Similarly, in Column (3), an increase of 1 standard deviation (2.5) in the cultural dispersion is associated with a 0.13 decrease in next year's EPS.

Therefore, I conclude that firms that hire more culturally aligned managers have higher market values and performance. This is consistent with the literature that managerial turnover is negatively correlated with firm performance and my previous evidence that cultural fit is positively correlated with managerial tenure.

9. How does the stock market respond to cultural distances?

How do investors react to the signals related to cultural fit in earnings calls? While it has been documented in the literature that investors respond to earnings announcements (Da, Engelberg, and Gao 2011, Ben-Rephael, Da, and Israelsen 2017), it has not been clear if investors also respond to information in earnings calls that are not directly related to firm performance. To answer the question, I calculate the excess returns of long-short portfolios sorted on cultural dispersion.[13] Because companies make earnings call every quarter, to obtain monthly measurements of cultural distances, I remove observations with zero cultural vectors and fill firms' monthly cultural distances with their latest available quarterly values. I only use the cultural values up to the month-end earnings call to calculate the Mahalanobis Distance's covariance matrix. Given that the number of companies with available earnings call transcripts before 2006 is relatively small when constructing portfolios, I only include companies from February 2006 to November 2020.

Before sorting the stock into portfolios, I follow Moussawi (2019) to standardise, winsorise and neutralise the negative cultural distances into clean signal values. I first standardise the negative cultural distances into a distribution of 0 mean and 1 standard deviation by month. Second, I winsorise the standardized negative cultural distances to a maximum of 3 standard deviations. Third, I neutralise the negative cultural dispersion against the market beta, the size-factor beta, and the Fama-French 48 industry dummy variables for each month:

$$Negative\ cultural\ dispersion_i = \alpha_i + \omega\beta_{market,i} + \mu\beta_{SMB,i} + \psi Industry_i + \epsilon_i$$

where i indexes the stock. β_{market} is the coefficient of the market factor and β_{SMB} is the coefficient of the size factor in the Fama-French 4-factor model. *Industry* is a dummy variable created from the Fama-French 48 industries. β_{market}, β_{SMB}, and *industry* are supplied by Wharton Research Data Services.

The neutralised signal is calculated as the sum of residual and intercept of the above regression:

$$Signal_i = \alpha_i + \epsilon_i$$

The neutralisation procedure orthogonalises cultural dispersion against the market beta, the size-factor beta and industries and removes the effects of those signals on portfolio formation. Hereafter, I denote the neutralised signal as the clean signal.

I sort stocks into quintiles for each month based on their clean signals. The portfolios sorted on the clean signals preserve the monotonic order of negative cultural distances from portfolios 1 to 5. Portfolio-1, which corresponds to the portfolio with the lowest clean signal, contains the companies with the lowest negative cultural distances (or equivalently, the highest cultural distances) and portfolio-5, which corresponds to the portfolio with the highest clean signal,

[13] Because the number of companies with corporate cultural distances measures is too small, I only use personal cultural distances to sort companies into portfolios.

contains the companies with the highest negative cultural distances (or equivalently, the lowest cultural distances).

Next, I obtain the Carhart 4-factor model (Fama and French 1993, Carhart 1997) excess returns for each stock in my portfolio from Wharton Research Data Services. I then calculate the cumulative excess returns of portfolios 1-5 by equally weighting or value weighting individual stocks in each portfolio respectively. While my main goal is not to show a new asset-pricing factor based on executives' cultural dispersion, Figure 8 Panel A and B show that stocks with the lowest cultural dispersion outperform stocks with the highest cultural dispersion in terms of cumulative excess returns over the Carhart 4-factor model. Figure 8 Panel A shows the excess returns for equally weighted portfolios, while Figure 8 Panel B shows the excess returns for value-weighted portfolios. Both graphs show consistent evidence that portfolio 5, which goes long in stocks with the highest clean signals (lowest cultural dispersion), has higher cumulative excess returns than portfolio 1, which goes long in stocks with the lowest clean signals (highest cultural dispersion). To further show that stocks with lower cultural dispersion perform better, I construct equally weighted and value-weighted long-short portfolios by going long in stocks with the lowest cultural dispersion and shorting the stocks with the highest cultural dispersion. Figure 8 Panel C and Panel D show that the cumulative excess returns of the equally weighted and the value-weighted long-short portfolios are positive from 2007 onwards, even during the financial crisis periods in 2008. The evidence consistently points out that companies with lower cultural dispersion outperform those with higher cultural dispersion.

While I update the cultural signal for those companies who make earnings call announcements in each month, a potential concern is that the group of companies who choose to make earnings call announcements may self-select into announcing good news on average, and the group of companies who delay earnings calls may be also delaying announcing any bad news until the situation improves. Therefore, the selected companies may announce the good news, leading to positive excess returns over the Carhart 4-factor model. To mitigate this concern, I follow Moussawi (2019) and compare the performance of a signal weighted long-only portfolio against the CRSP U.S. Common Index weighted portfolio for the same set of stocks each month. A bias of selecting stocks with good news will boost the returns in both portfolios, as both portfolios include the same stock universe. I construct the following market index conditional on my portfolio holdings as follows:

$$market_t = \sum_{i \,\epsilon\, portfolio} w_{it} \times RET_{it}$$

where i indexes the stock, t indexes the month, w is the weight of the stock in the CRSP U.S. Common Stock Index and RET is the return of the stock. I only include stock i in calculating the market index if the stock is also in my portfolio sorted on clean signals. I then calculate the cumulative excess returns of the long-only signal-weighted portfolio as follows:

$$Long\ RET = \frac{\sum_{i \in portfolio} Signal_{it} \times |Signal_{it} > 0| \times RET_{it}}{\sum_{j \in portfolio} Signal_{jt} \times |Signal_{jt} > 0|}$$

where i and j indexes the stock, t indexes the month, *signal* is the standardised and neutralised signal. $|Signal > 0|$ is a dummy variable that equals 1 if the value of signal is greater than 0 and equals 0 if the value of signal is less than or equal to 0. A higher signal value is associated with a lower cultural distance. Therefore, the portfolio places higher weights on stocks with lower cultural distances and lower weights on stocks with higher cultural distances. Finally, I plot the cumulative returns of the long-only signal-weighted portfolio against the CRSP U.S. Common Stock Index conditional on my portfolio holdings in Figure 9. The figure shows that the signal-weighted long portfolio consistently outperforms the market since 2009, while there are no significant differences in performance before 2009. However, Figure 9 Panel B shows that the turnover ratios of my portfolios are high. Therefore, the high excess returns do not necessarily translate to tradable profits in the stock market.

Finally, Table 7 shows the annual portfolio returns for portfolios in each sorted quintile. Equally weighted and value-weighted portfolio 5 have the highest Sharpe ratios of 0.736 and 0.776, respectively, across all portfolios. To fully reflect the strength of the signal in my portfolio, I construct a signal weighted portfolio that longs stocks with positive signals and shorts stocks with negative signals. The holding positions of each stock is weighted by their signal strengths, and the overall portfolio return is calculated as follows:

$$Signal\ RET = \frac{\sum_{i \in portfolio} Signal_{it} \times |Signal_{it} > 0| \times RET_{it}}{\sum_{j \in portfolio} Signal_{jt} \times |Signal_{jt} > 0|} - \frac{\sum_{i \in portfolio} Signal_{it} \times |Signal_{it} < 0| \times RET_{it}}{\sum_{j \in portfolio} Signal_{jt} \times |Signal_{jt} < 0|}$$

where $|Signal < 0|$ is a dummy variable that equals 1 if the value of signal is less than 0 and equals 0 if the value of signal is greater than or equal to 0. All other variables have been defined previously. The signal-weighted portfolio return equals 1.25%, which is greater than 0. This piece of evidence further confirms our earlier hypothesis that companies with smaller cultural distances outperform companies with larger cultural distances, and this phenomenon is not constrained to the top and bottom portfolios. By longing companies with smaller cultural distances and shorting companies with larger cultural distances, and weighting the stock positions with their signal strengths, the portfolio can produce a positive annual return.

Therefore, hiring managers with a better cultural fit (or, equivalently, lower personal cultural distances) is perceived as a good signal by the investors. Although I do not claim that I have discovered a new asset-pricing factor, I show that companies with lower average cultural distances amongst executives are related to higher adjusted future stock returns and outperform the market portfolios.

10. Is cultural fit related to managerial probabilities of turnover?

The traditional corporate culture and CEO turnover literature (Fiordelisi and Ricci 2014) have shown that the competition and control-oriented culture positively influence the probability of CEO turnover. As a robustness check, I employ the following logit regression model:

$$C_{i,n,t}=Logit(\beta_0+\beta_1 Cultural\ Diff_{i,n,t}+\mu M_{i,n,t}+\psi F_{i,n,t}+\lambda B_{i,n,t}+\epsilon_{i,n,t})$$

where i indexes executives, n indexes firms executives move to, and t indexes the current year at the new firm. ϵ is the error term. The dependent variable, $C_{i,n,t}$, equals to 1 if the manager has terminated their position and equals to 0 if the manager has not terminated their position at the new company until the end of my sample period. All other variables are as defined in the previous section.

I first explore the impact of corporate cultural distances on the probability of managerial turnover. Table 8 Panel A Columns 1-3 show the logit regression coefficients. Consistent with my PH models, larger cultural distance is associated with a statistically significant and larger probability of managers exiting the firms. I find that the competition, control or innovation culture does not matter in predicting the probability of executives' turnover. In contrast, Fiordelisi and Ricci (2014) find that the competition culture contributes positively to the probability of executives' turnover. In addition, having a strong community culture can significantly reduce the probability of executives' turnover, and an emphasis on the integrity culture can significantly increase the probability of executives' turnover.

Next, I explore the impact of personal cultural distances on the probability of managerial turnover. The logit regression results are shown in Table 8 Columns 4-6. Consistently, the worse the cultural fit (the larger the personal cultural distances), the higher the probability of managerial turnover. Also consistent with the PH regression results, the higher the community culture, the lower the probability of managerial turnover.

11. Conclusions

My paper makes several contributions. First, I create a new measure of corporate cultural fit by measuring corporate culture with a state-of-the-art machine learning model, Sentence-BERT. My methodology differentiates from the traditional dictionary-based approach by taking into consideration the semantic meaning of complete sentences and avoiding ambiguous out-of-context terms in dictionaries.

Second, I document a positive (negative) and economically significant impact of cultural fit (cultural distances) on managerial tenure using survival models. The effect exists in both proxies for cultural fit – corporate cultural distances (between firms) and personal cultural distances (between managers and firms). Simply put, managers tend to stay longer in firms where they better fit into the corporate culture. Although M&A is one of the important drivers of corporate cultural change, I show that the relationship between cultural fit and managerial tenure is not driven solely by M&A. Cultural fit is related to cultural adaptation, which is deeply rooted in human nature.

Third, I employ causal survival forests to show that the effect of cultural fit on managerial tenure is causal. Causal survival forest is a new econometrics tool that allows non-parametric estimations of causal effects when the data is right-censored. My results imply that better (worse) cultural fit is one of the reasons that cause managers to stay longer (shorter) in firms. Specifically, I find that the negative causal effect of bad cultural fit on managerial tenure is exacerbated when managerial pay is higher. Therefore, it is important for companies who desire stabilities in the management team to hire managers who fit better culturally.

Fourth, I show evidence that firms that hire managers with good cultural fit have higher future market values and performance. Simply put, a good cultural fit is beneficial for firms' future operations and performance. Furthermore, Investors perceive lower cultural dispersion within the firm as a positive signal, as a long-short strategy that goes long in the stocks with lower cultural dispersion and shorts the stocks with higher cultural dispersion generates positive returns over the Carhart four-factor model.

References

Anderson, Linda E., 1994, A new look at an old construct: Cross-cultural adaptation, *International Journal of Intercultural Relations* 18, 293–328.

Athey, Susan, and Guido Imbens, 2016, Recursive partitioning for heterogeneous causal effects, *Proceedings of the National Academy of Sciences* 113, 7353–7360.

Athey, Susan, Julie Tibshirani, and Stefan Wager, 2019, Generalized random forests, *The Annals of Statistics* 47, 1148–1178.

Bardos, Katsiaryna Salavei, Mine Ertugrul, and Lucia Silva Gao, 2020, Corporate social responsibility, product market perception, and firm value, *Journal of Corporate Finance* 62, 101588.

Ben-Rephael, Azi, Zhi Da, and Ryan D. Israelsen, 2017, It Depends on Where You Search: Institutional Investor Attention and Underreaction to News, *The Review of Financial Studies* 30, 3009–3047.

Beugelsdijk, Sjoerd, and Bart Frijns, 2010, A cultural explanation of the foreign bias in international asset allocation, *Journal of Banking & Finance* 34, 2121–2131.

Brookman, Jeff, and Paul D. Thistle, 2009, CEO tenure, the risk of termination and firm value, *Journal of Corporate Finance* 15, 331–344.

Broström, Göran, 2012, *Event History Analysis with R* (Chapman and Hall/CRC, Boca Raton).

Buchanan, Bonnie, Cathy Xuying Cao, and Chongyang Chen, 2018, Corporate social responsibility, firm value, and influential institutional ownership, *Journal of Corporate Finance* 52, 73–95.

Bushman, Robert, Zhonglan Dai, and Xue Wang, 2010, Risk and CEO turnover, *Journal of Financial Economics* 96, 381–398.

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

Coughlan, Anne T., and Ronald M. Schmidt, 1985, Executive compensation, management turnover, and firm performance, *Journal of Accounting and Economics* 7, 43–66.

Cui, Yifan, Michael R. Kosorok, Stefan Wager, and Ruoqing Zhu, 2020, Estimating heterogeneous treatment effects with right-censored data via causal survival forests, *arXiv:2001.09887 [cs, stat]*.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2011, In Search of Attention, *The Journal of Finance* 66, 1461–1499.

Dikolli, Shane S., William J. Mayew, and Dhananjay Nanda, 2014, CEO tenure and the performance-turnover relation, *Review of Accounting Studies* 19, 281–327.

Dodd, Olga, Bart Frijns, and Aaron Gilbert, 2015, On the Role of Cultural Distance in the Decision to Cross-List, *European Financial Management* 21, 706–741.

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

Fiordelisi, Franco, and Ornella Ricci, 2014, Corporate culture and CEO turnover, *Journal of Corporate Finance* 28. Inside the Board Room, 66–82.

Frijns, Bart, Olga Dodd, and Helena Cimerova, 2016, The impact of cultural diversity in corporate boards on firm performance, *Journal of Corporate Finance* 41, 521–541.

Gentry, Richard J., Joseph S. Harrison, Timothy J. Quigley, and Steven Boivie, 2021, A database of CEO turnover and dismissal in S&P 1500 firms, 2000–2018, *Strategic Management Journal* 42, 968–991.

Goyal, Vidhan K, and Chul W Park, 2002, Board leadership structure and CEO turnover, *Journal of Corporate Finance* 8, 49–66.

Graham, John R., Campbell R. Harvey, Jillian Popadak, and Shivaram Rajgopal, 2017, Corporate Culture: Evidence from the Field. Working Paper Series, National Bureau of Economic Research.

Grootendorst, Maarten, 2020, BERTopic: Leveraging BERT and c-TF-IDF to create easily interpretable topics., *Zenodo*.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, 2015, Corporate Culture, Societal Culture, and Institutions, *American Economic Review* 105, 336–339.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, 2015, The value of corporate culture, *Journal of Financial Economics* 117, 60–76.

Gulbranson, Derek, 2021, *Name Parser*. GitHub. A simple Python module for parsing human names into their individual components.

Hofstede, Geert, 2001, *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations*. 2nd ed. (Sage, Thousand Oaks, CA).

Huson, Mark R., Robert Parrino, and Laura T. Starks, 2001, Internal Monitoring Mechanisms and CEO Turnover: A Long-Term Perspective, *The Journal of Finance* 56, 2265–2297.

Jenter, Dirk, and Fadi Kanaan, 2015, CEO Turnover and Relative Performance Evaluation, *The Journal of Finance* 70, 2155–2184.

Jenter, Dirk, and Katharina Lewellen, 2020, Performance-Induced CEO Turnover, *The Review of Financial Studies* 34, 569–617.

Kaplan, Steven N., 1993, Top Executives, Turnover and Firm Performance in Germany. Working Paper. Working Paper Series, National Bureau of Economic Research.

Karolyi, G Andrew, 2016, The gravity of culture for finance, *Journal of Corporate Finance* 41, 610–625.

Kim, Young Yun, 1988, *Communication and Cross-Cultural Adaptation: An Integrative Theory.* Communication and cross-cultural adaptation: An integrative theory. (Multilingual Matters, Clevedon, England).

Kim, Young Yun, 2017, Cross-Cultural Adaptation, by by Young Yun Kim *Oxford Research Encyclopedia of Communication* (Oxford University Press).

Li, Kai, Xing Liu, Feng Mai, and Tengfei Zhang, 2021, The Role of Corporate Culture in Bad Times: Evidence from the COVID-19 Pandemic, *Journal of Financial and Quantitative Analysis* 56, 2545–2583.

Li, Kai, Feng Mai, Rui Shen, and Xinyan Yan, 2020, Measuring Corporate Culture Using Machine Learning, *The Review of Financial Studies* 34, 3265–3315.

Li, Qin, Ben Lourie, Alexander Nekrasov, and Terry Shevlin, 2021, Employee Turnover and Firm Performance: Large-Sample Archival Evidence, *Management Science*, mnsc.2021.4199.

Loughran, Tim, and Bill Mcdonald, 2011, When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks, *The Journal of Finance* 66, 35–65.

Loughran, Tim, and Bill Mcdonald, 2016, Textual Analysis in Accounting and Finance: A Survey: Textual Analysis in Accounting and Finance, *Journal of Accounting Research* 54, 1187–1230.

McInnes, Leland, John Healy, and James Melville, 2020, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, *arXiv:1802.03426 [cs, stat]*.

Mitchell, Ann FS, and Wojtek J Krzanowski, 1985, The Mahalanobis distance and elliptic distributions, *Biometrika* 72, 464–467.

Moussawi, Rabih, 2019, WRDS Factors SAS Code, *Wharton Research Data Services*.

Neudecker, Heinz, 1997, Mahalanobis Distance for Multinomial Data, *Econometric Theory* 13, 890–891.

O'Reilly, Charles A., and Jennifer A. Chatman, 1996, Culture as social control: Corporations, cults, and commitment., *Research in organizational behavior: An annual series of analytical essays and critical reviews, Vol. 18.* (Elsevier Science/JAI Press, US).

O'Reilly, Charles A., Jennifer Chatman, and David F. Caldwell, 1991, People and Organizational Culture: A Profile Comparison Approach to Assessing Person-Organization Fit, *The Academy of Management Journal* 34, 487–516.

Peters, Florian S., and Alexander F. Wagner, 2014, The Executive Turnover Risk Premium: The Executive Turnover Risk Premium, *The Journal of Finance* 69, 1529–1563.

Quinn, Robert E., and John Rohrbaugh, 1983, A Spatial Model of Effectiveness Criteria: Towards a Competing Values Approach to Organizational Analysis, *Management Science* 29, 273–393.

Reimers, Nils, and Iryna Gurevych, 2019, Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, *arXiv:1908.10084 [cs]*.

Servaes, Henri, and Ane Tamayo, 2013, The Impact of Corporate Social Responsibility on Firm Value: The Role of Customer Awareness, *Management Science* 59, 1045–1061.

Therneau, Terry, Cynthia Crowson, and Elizabeth Atkinson, 2022, Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model, *CRAN*.

Tseng, Yi-Kuan, Fushing Hsieh, and Jane-Ling Wang, 2005, Joint modelling of accelerated failure time and longitudinal data, *Biometrika* 92, 587–603.

Vandenberghe, Christian, 1999, Organizational culture, person-culture fit, and turnover: a replication in the health care industry, *Journal of Organizational Behavior* 20, 175–184.

Warner, Jerold B., Ross L. Watts, and Karen H. Wruck, 1988, Stock prices and top management changes, *Journal of Financial Economics* 20, 461–492.

Weber, Yaakov, 1996, Corporate cultural fit and performance in mergers and acquisitions, *Human relations* 49, 1181–1202.

Weber, Yaakov, Oded Shenkar, and Adi Raveh, 1996, National and Corporate Cultural Fit in Mergers/Acquisitions: An Exploratory Study, *Management Science* 42, 1215–1227.

Weisbach, Michael S., 1988, Outside directors and CEO turnover, *Journal of Financial Economics* 20, 431–460.

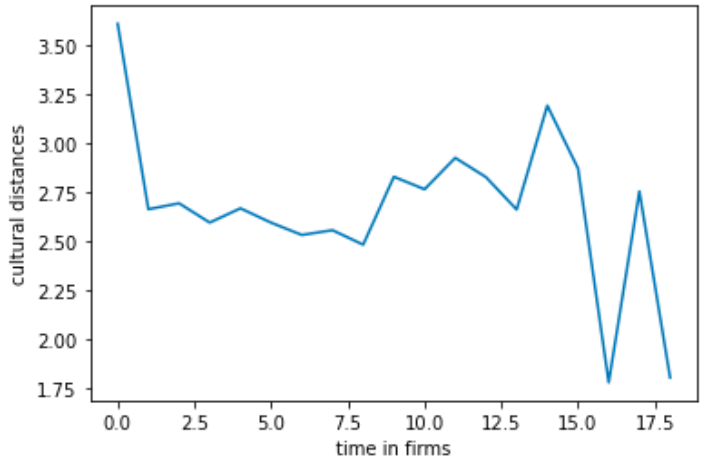
Xiang, Shiming, Feiping Nie, and Changshui Zhang, 2008, Learning a Mahalanobis distance metric for data clustering and classification, *Pattern Recognition* 41, 3600–3612.

Appendix. Variable Definitions

Variable	Description	Data Source
	Corporate Culture Variables	
Innovation culture	Innovation culture score measured with machine learning analysing earnings call transcripts.	Capital IQ Earnings Call Transcript
Integrity culture	Integrity culture score measured with machine learning analysing earnings call transcripts.	Capital IQ Earnings Call Transcript
Quality culture	Quality culture score measured with machine learning analysing earnings call transcripts.	Capital IQ Earnings Call Transcript
Respect culture	Respect culture score measured with machine learning analysing earnings call transcripts.	Capital IQ Earnings Call Transcript
Teamwork culture	Teamwork culture score measured with machine learning analysing earnings call transcripts.	Capital IQ Earnings Call Transcript
Control culture	Control culture score measured with machine learning analysing earnigs call transcripts.	Capital IQ Earnings Call Transcript
Competition culture	Competition culture score measured with machine learning analysing earnings call transcripts.	Capital IQ Earnings Call Transcript
Hard work culture	Hard work culture score measured with machine learning analysing earnings call transcripts.	Capital IQ Earnings Call Transcript
Community culture	Community culture score measured with machine learning analysing earnings call transcripts.	Capital IQ Earnings Call Transcript
Corporate cultural distance	Euclidean distance of the culture vectors between companies.	Capital IQ Earnings Call Transcript
Personal cultural distance	Euclidean distance of the culture vectors between managers and companies.	Capital IQ Earnings Call Transcript
	Manager Variables	
Manager gender	The gender of the manager.	ExecuComp
Manager age	The age of the manager.	ExecuComp
Manager tenure (survival time in years)	The number of years that a manager spends in a company before leaving (or until the end of the sample, if the manager did not leave).	ExecuComp
Manager salary pay (Thousands $)	The dollar value of the base salary (cash and non-cash) earned by the named executive officer during the fiscal year.	ExecuComp

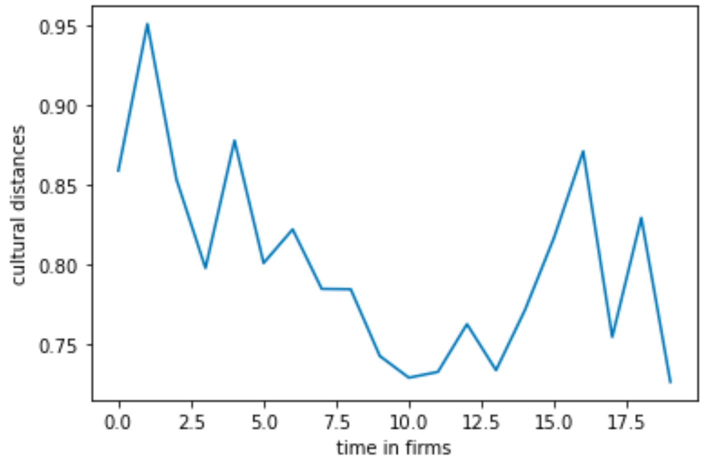
Manager option awards (Thousands $)	The value of option awards - FAS 123R.	ExecuComp
Manager stock awards (Thousands $)	The value of stock awards - FAS 123R.	ExecuComp
Manager bonus (Thousands $)	The dollar value of a bonus (cash and non-cash) earned by the named executive officer during the fiscal year.	ExecuComp
Incentive pay	log(option awards pay + stock awards pay)	ExecuComp
Non-incentive pay	log(salary pay + bonus pay)	ExecuComp
Total pay	log(salary pay + option awards pay + stock awards pay + bonus pay)	ExecuComp
Forced turnover	A dummy variable indicating if a manager's leaving is forced.	ExecuComp
	Board Variables	
Board attrition	Number of Directors that have left a role as a proportion of average number of directors for the preceding reporting period at the annual report date selected.	BoardEx
Board gender ratio	The proportion of male directors at the annual report date selected.	BoardEx
Board number of qualifications	The average number of qualifications at undergraduate level and above for all the directors at the annual report date selected.	BoardEx
Board network size	Network size of selected individual (number of overlaps through employment, other activities, and education).	BoardEx
Board no. of directors	Number of executive directors, supervisory directors or all of the directors at the annual report date selected.	BoardEx
	Firm Variables	
R&D	Research and development expense	Compustat
Sales	Gross sales	Compustat
Invested capital	Sum of long-term debt, preferred stock (carrying value), non-controlling interests and common equity.	Compustat
R&D/sale ratio	Research and development expense / Sales	Compustat
Sales/invested capital ratio	Sales / Invested Capital	Compustat
Sales growth	Percentage growth of sale	Compustat
Asset growth	Percentage growth of total asset	Compustat
Size	Market value for single issue companies is common shares outstanding multiplied by the month-end price that corresponds to the period end date.	Compustat

Book value	Book Value Per Share is based on fiscal year-end data and represents Common Equity Liquidation Value (CEQL) divided by Common Shares Outstanding (CSHO).	Compustat
Book to market ratio	Book value divided by Size.	Compustat
Net income	The fiscal period income or loss reported by a company after subtracting expenses and losses from all revenues and gains.	Compustat
Total asset	The total assets/liabilities of a company at a point in time.	Compustat
ROA	Net Income divided by Total Assets	Compustat
Leverage ratio	Total debt / stockholder’s equity	Compustat
Tobin's Q	Size divided by Book value	Compustat
Idiosyncratic volatility	The idiosyncratic volatility of a rolling-12-month Fama-French-Carhart factor model.	CRSP
Total volatility	The total volatility of a rolling-12-month Fama-French-Carhart factor model.	CRSP



A

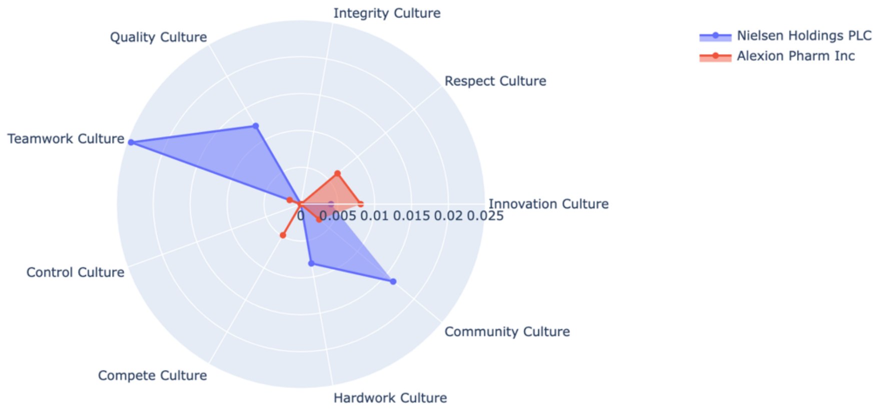
Corporate Cultural Distances (Between Firms)



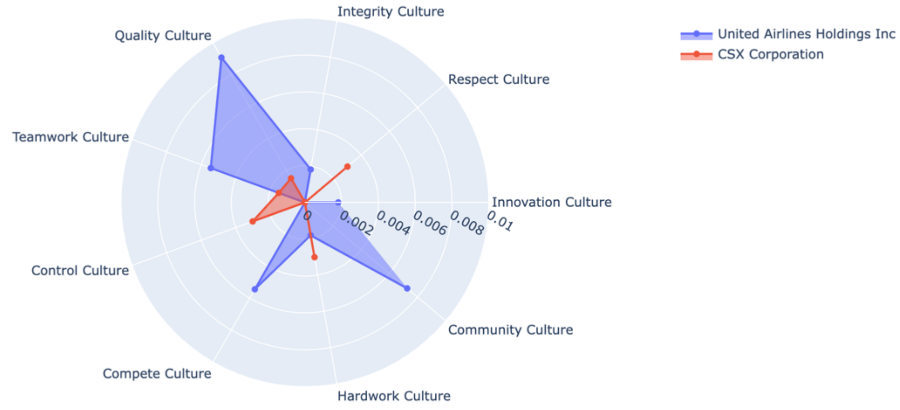
B

Personal Cultural Distances (Between Managers and Firms)

Figure 1
This figure shows the time-series variation of cultural distances in firms. The horizontal axis shows managerial tenure, and the vertical axis shows the average cultural distances at the specific managerial tenure. The graph on the left shows the time-series variation of corporate cultural distances (between firms) and the graph on the right shows the time-series variation of personal cultural distances (between managers and firms).



Nielsen Holdings PLC (2019) and Alexion Pharm Inc (2017)
Executive: David John Anderson
Cultural Distance: 5.98



United Airlines Holdings Inc. (2019) and CSX Corporation (2015)
Executive: Oscar Munoz
Cultural Distance: 2.72

Figure 2
This figure compares the corporate culture between two companies. The graph on the left compares Nielson Holdings PLC's culture in 2019 to Alexion Pharm Inc's culture in 2017 as David Anderson leaves for the former from the latter. The graph on the right compares United Airlines Holdings Inc.'s culture in 2019 to CSX Corporation's culture in 2015 as Oscar Munoz leaves for the former from the latter.

Timothy Boyle and Columbia Sportswear Company (2019)

Cultural Distance: 1.06

Marc Benioff and Salesforce.com, inc. (2019)

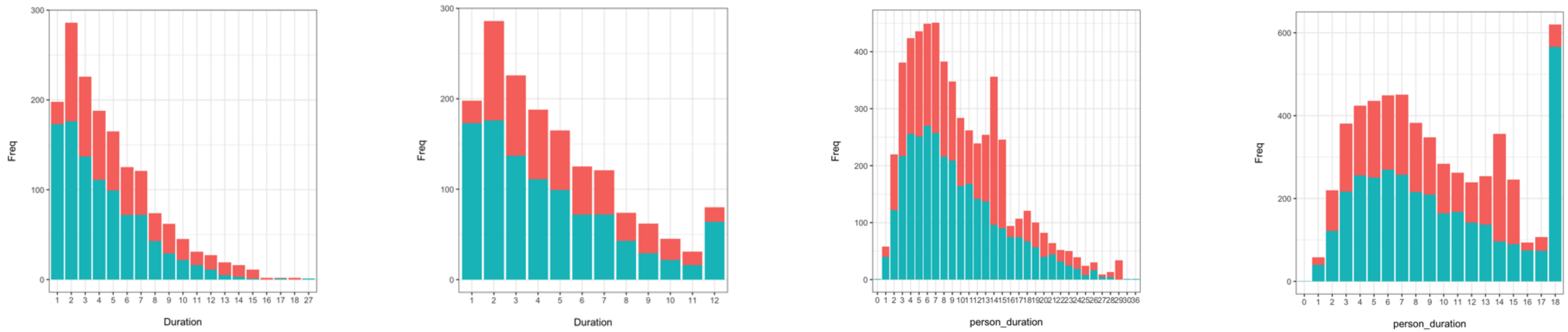
Cultural Distance: 0.31

Figure 3
This figure compares the corporate culture of a company and the cultural preferences of a manager. The graph on the left compares the corporate culture of Columbia Sportswear Company to the cultural preferences of Timothy Boyle in 2019. The graph on the right compares the corporate culture of Salesforce.com Inc. to the cultural preferences of Marc Benioff in 2019.

A B

Figure 4 Baseline plot of survival function from Proportional-Hazard model

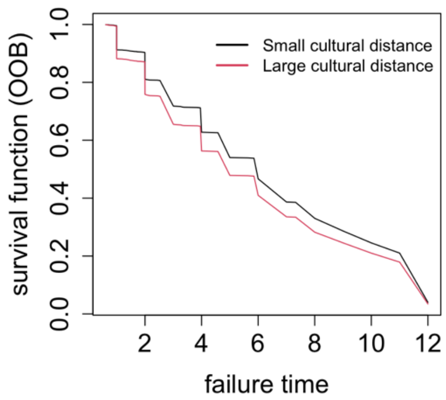
Plots A shows the baseline survival function from the Proportional-Hazard model for Columns (1) in Table 2, and Plots B shows the baseline survival function from the Proportional-Hazard model for Columns (1) in Table 3. The red line shows the survival probability plot against time when the personal cultural distances are less than the median. The red line shows the survival probabilities when cultural distances are large, while the black line in shows the survival probabilities when cultural distances are small.



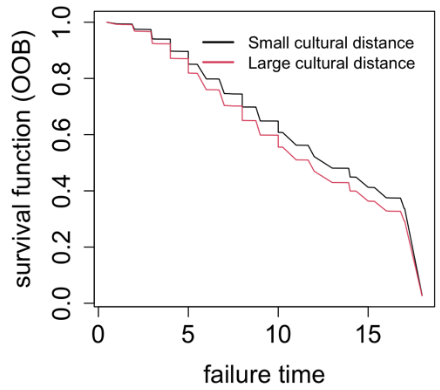
A B C D

Figure 5 Histogram of survival time for censored/non-censored subjects.

The treatment variable in Panel A and B is corporate cultural distance, while the treatment variable in Panel C and D is personal cultural distance. Panel A and C show the survival time histogram for censored/non-censored subjects. The red colour indicates the censored objects (i.e., those managers who do not exit firms). Panel B and D show the histogram of survival time for censored/non-censored subjects after truncation

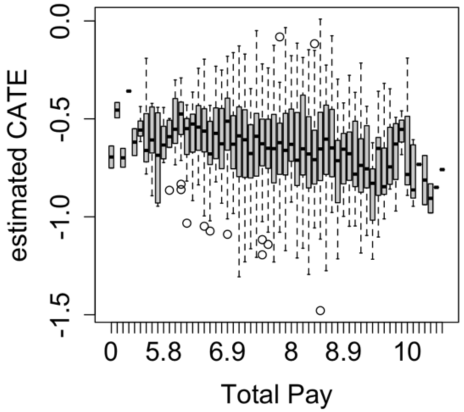


A

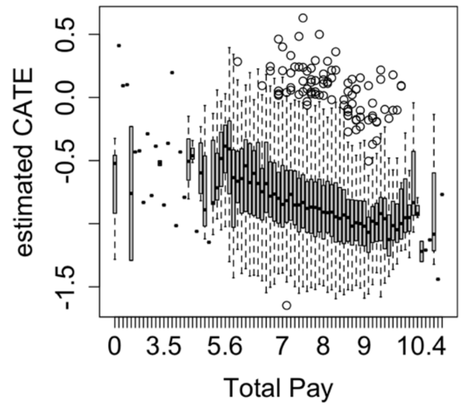


B

Figure 6 Survival forests prediction plot.
Panel A and B shows the survival probability predictions when the treatments are corporate and personal cultural distance, respectively. The vertical axis shows the survival probabilities, while the horizontal axis shows the failure time. The red line shows the survival probabilities when cultural distances are large, while the black line in shows the survival probabilities when cultural distances are small.

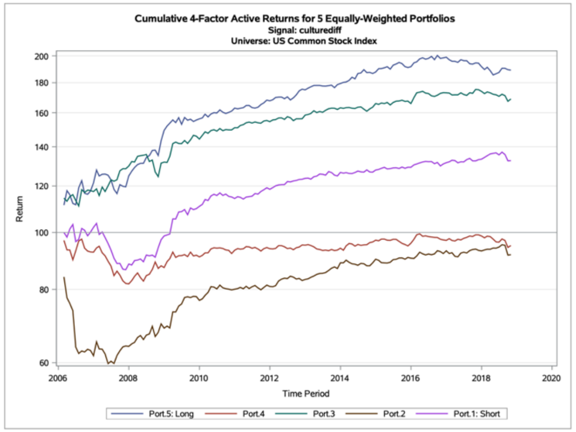


A

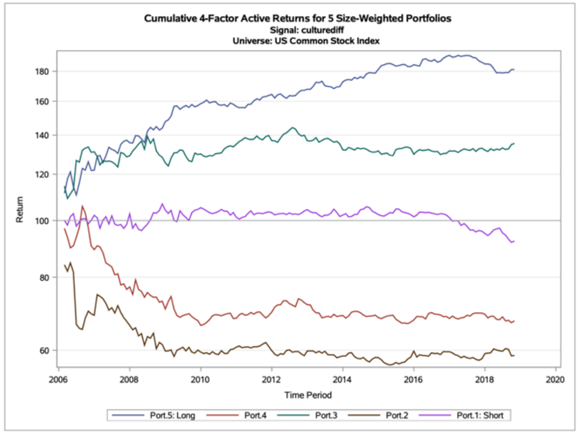


B

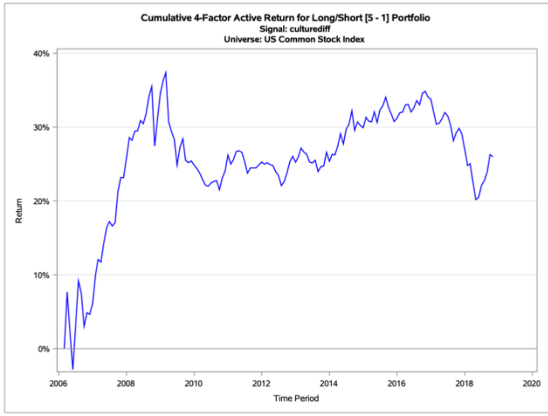
Figure 7 Heterogeneity of the effect of cultural distances on managerial survival time across managerial pay
Panel A shows the box plot of treatment effect heterogeneity across total pay when cultural fit is proxied by corporate cultural distances. Panel B shows the box plot of treatment effect heterogeneity across total pay when the cultural fit is proxied by personal cultural distances.



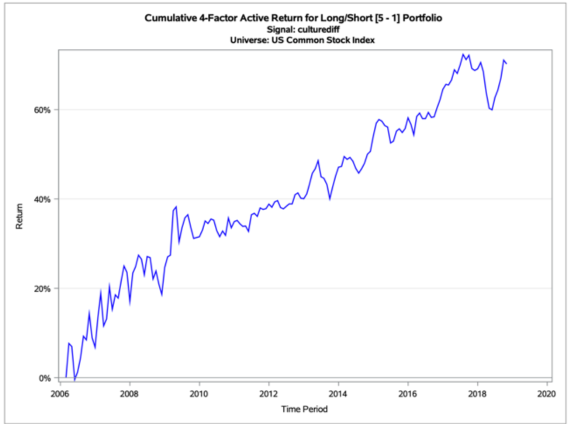
A



B



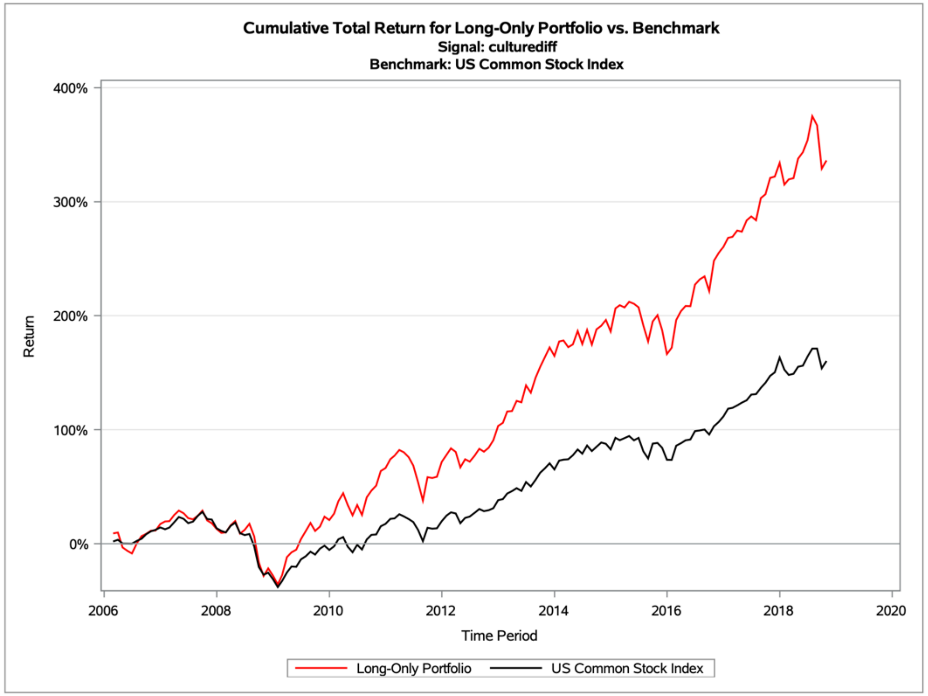
C



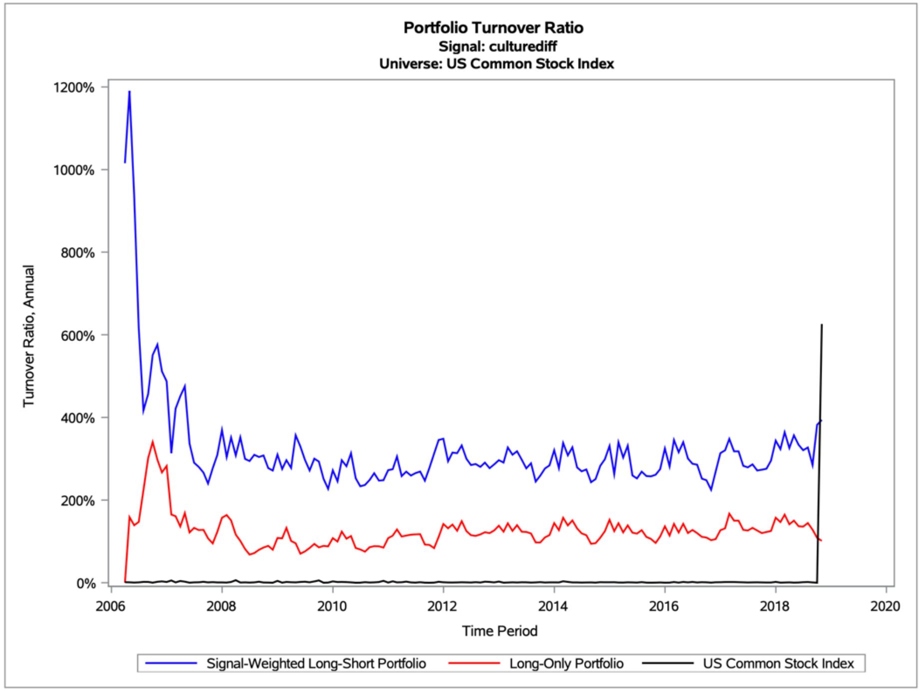
D

Figure 8

Figure A and B show the active returns of five portfolios sorted on clean signal while the benchmark is the Carhart 4-factor model. Figure A shows the cumulative returns of the equally weighted portfolios, while Figure B shows the cumulative returns of the value-weighted portfolios. Figure C and D show the active returns of the Long-Short portfolios sorted on clean signal while the benchmark is the Carhart 4-factor model. Figure C shows the cumulative returns of the equally weighted portfolios, while Figure D shows the cumulative returns of the value-weighted portfolios. Portfolio 5 contains the stocks with the lowest cultural distances (or equivalently, highest negative cultural distances), and portfolio 1 includes the stocks with the highest cultural distances (or equivalently, lowest negative cultural distances).



A



B

Figure 9
This figure shows the cumulative returns of the long portfolios weighted by clean signal. A lower cultural distance is related to a higher weight in the long portfolio. The benchmark is the CRSP U.S. Common Stock Index. Figure A shows the cumulative returns, while Figure B shows the turnover ratios of the signal-weighted portfolios.

Table 1

This table reports summary statistics. Panel A shows the summary statistics when measuring cultural fit with corporate cultural distances (between firms) and personal cultural distances (between managers and firms). Panel B shows the manager distribution by gender and forced turnovers. There are 1081 unique firms and 1795 unique managers in the corporate cultural distance sample from 2002 to 2020. There are 2,224 unique firms and 12,284 unique managers in the personal cultural distance sample from 2006 to 2020.

Panel A

	Corporate cultural distance sample		Personal cultural distance sample	
	mean	std	mean	std
Managerial pay				
Salary (USD Thousands)	534.3258	334.7555	684.8939	524.9042
Option awards (USD Thousands)	669.0448	2918.3551	670.1189	2650.0152
Stock awards (USD Thousands)	1952.9330	3143.1210	2877.6256	6194.2487
Bonus (USD Thousands)	222.9624	588.1731	154.8668	706.3619
Corporate cultural values				
Corporate cultural distance	2.5275	2.4590		
Innovation culture	0.0017	0.0040		
Integrity culture	0.0005	0.0016		
Quality culture	0.0036	0.0056		
Respect culture	0.0009	0.0021		
Teamwork culture	0.0013	0.0028		
Community culture	0.0011	0.0052		
Hardwork culture	0.0007	0.0019		
Control culture	0.0007	0.0020		
Compete culture value	0.0032	0.0046		
Personal cultural preferences				
Personal culture distance			2.4056	4.6307
Innovation culture			0.0088	0.0559
Integrity culture			0.0033	0.0261
Quality culture			0.0138	0.0449
Respect culture			0.0032	0.0173
Teamwork culture			0.0062	0.0435
Community culture			0.0046	0.0333
Hardwork culture			0.0037	0.0287
Control culture			0.0029	0.0167
Compete culture value			0.0117	0.0503
Board attrition	0.0369	0.0533	0.0350	0.0582
Board gender ratio	0.8231	0.1034	0.7984	0.1154
Board no. of qualifications	2.1604	0.3739	2.1600	0.4102

Board network size	1975.0879	948.9088	1930.2387	910.9729
Board number of directors	9.5303	1.9908	9.6463	2.3738
Managerial characteristics				
Manager age	54.4321	5.9398	55.4045	7.4152
Manager tenure	4.3510	3.2659	8.5211	5.7687
Firm characteristics				
ROA	0.0204	0.1144	0.0171	0.1736
Leverage ratio	2.5590	62.6191	2.1478	44.2098
R&D/sale	0.0442	0.1872	0.8494	30.8695
Size (USD millions)	15365.0372	47270.5418	21312.0416	69153.8074
Book/market ratio	-0.0078	0.2637	0.0042	0.1198
Sale/invested capital	1.7208	3.9953	3.1789	66.8369
Sales growth	0.1493	3.3667	0.1333	5.5342
Asset growth	0.2758	6.5196	0.0934	0.3405
Idiosyncratic volatility	0.0854	0.1784	0.0725	0.1037
Total volatility	0.1155	0.0669	0.1172	0.0721

Panel B

	Corporate cultural distance sample	Personal cultural distance sample
Categorical Variables	**Frequency**	**Frequency**
Manager gender (male)	1583	11198
Manager gender (female)	212	1086
Forced turnover	4	421

Table 2
This table reports coefficients from the Cox proportional-hazards model with time-dependent covariates for the effect of corporate cultural distance on the tenure (survival time) of the manager in the new firm. The dependent variable is the tenure (survival time) of the manager in the new firm after leaving the old firm. The tenure (survival) time is a censored variable. I calculate it as the number of years a manager has spent in a firm before leaving (or until the end of my sample period, if the manager has not left). Columns (1) and (4) control for the logarithmic total pay (salary pay + stock awards pay + option awards pay + bonus pay) difference that the manager gets between the old firm and the new firm. Columns (2) and (5) control for the logarithmic incentive pay (stock awards pay + option awards pay). Columns (3) and (6) controls for the logarithmic non-incentive pay (salary pay + bonus pay). Industry fixed effects are based on the Fama-French 12-industry classification. All other variables are defined in the Appendix. Standard errors are reported in parentheses and are clustered at the manager level. *** significant at 1%, ** significant at 5%, *significant at 10%.

	(1)	(2)	(3)	(4)	(5)	(6)
	All pay	Incentive pay	Non-incentive pay	All pay	Incentive pay	Non-incentive pay
Corporate Culture						
Cultural distance	0.0274 *	0.0300 **	0.0296 *	0.0222	0.0251	0.0247
	(0.0155)	(0.0153)	(0.0159)	(0.0188)	(0.0185)	(0.0191)
Cultural distance x ROA				0.2509 **	0.2374 **	0.2392 **
				(0.1067)	(0.1062)	(0.1070)
Innovation culture	4.3050	5.2740	4.1720	6.6850	7.5980	6.4290
	(6.3750)	(6.2860)	(6.2830)	(6.2770)	(6.2210)	(6.2220)
Integrity culture	24.2200	22.7200	23.0100	27.9400	26.1300	26.3800
	(17.9800)	(17.8500)	(18.3900)	(18.1500)	(17.9900)	(18.5700)
Quality culture	2.3890	3.4160	2.9930	2.4910	3.5890	3.1800
	(5.2880)	(5.2180)	(5.3770)	(5.2020)	(5.1420)	(5.3050)
Respect culture	-5.5860	-4.1200	-3.4770	-6.6010	-5.0180	-4.3680
	(12.8200)	(12.5400)	(12.7500)	(12.8400)	(12.5400)	(12.7600)
Teamwork culture	7.3980	5.5620	5.0890	7.9080	5.9480	5.6540
	(10.0100)	(10.1300)	(10.3300)	(9.9450)	(10.0900)	(10.2800)
Community culture	-20.4500 **	-20.7000 ***	-20.8900 **	-21.1900 **	-21.3500 ***	-21.5300 **
	(7.9750)	(7.9950)	(8.1150)	(8.2480)	(8.2710)	(8.3950)

Hardwork culture	-24.1800	-24.0200 *	-24.3000 *	-23.1900	-23.7100	-24.3300
	(15.0200)	(14.3900)	(14.5700)	(15.1000)	(14.7200)	(14.8800)
Control culture	-3.7480	-6.0460	-6.6660	-7.3570	-9.7930	-10.4400
	(13.7800)	(13.4900)	(13.6400)	(13.6500)	(13.5200)	(13.7000)
Compete culture	1.3240	2.1680	2.5000	0.6953	1.5860	1.8270
	(6.4200)	(6.2320)	(6.1070)	(6.4280)	(6.2440)	(6.1230)
Managerial Characteristics						
Forced leave	2.1700 ***	2.0140 ***	2.0550 ***	2.1500 ***	1.9960 ***	2.0390 ***
	(0.1464)	(0.1458)	(0.1534)	(0.1474)	(0.1463)	(0.1545)
Manager age	0.0223 ***	0.0226 ***	0.0244 ***	0.0222 ***	0.0224 ***	0.0243 ***
	(0.0069)	(0.0068)	(0.0066)	(0.0069)	(0.0068)	(0.0066)
Manager gender (female)	0.2240 **	0.2686 **	0.2667 **	0.2225 **	0.2679 **	0.2653 **
	(0.1133)	(0.1122)	(0.1123)	(0.1132)	(0.1119)	(0.1120)
Managerial Pay						
All pay (Log)	-0.2171 ***			-0.2194 ***		
	(0.0294)			(0.0295)		
Incentive pay (Log)		-0.0571 ***			-0.0581 ***	
		(0.0134)			(0.0134)	
Nonincentive pay (Log)			-0.1127 ***			-0.1172 ***
			(0.0375)			(0.0380)
Board Characteristics						
Board number of directors	0.0041	-0.0076	-0.0035	0.0049	-0.0069	-0.0026
	(0.0189)	(0.0188)	(0.0190)	(0.0189)	(0.0188)	(0.0191)
Board gender ratio	0.8081 **	0.8254 **	0.9341 ***	0.8287 **	0.846 **	0.9575 ***
	(0.3488)	(0.3494)	(0.3411)	(0.3480)	(0.3487)	(0.3405)
Board no. of qualifications	0.0305	0.0474	0.0381	0.0241	0.0421	0.0321
	(0.0858)	(0.0862)	(0.0861)	(0.0857)	(0.0860)	(0.0860)

Board network size	0.0002 ***	0.0001 ***	0.0001 ***	0.0002 ***	0.0001 ***	0.0001 ***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Board attrition rate	1.003 *	0.8148	0.9081	0.9645	0.7826	0.8777
	(0.6008)	(0.5932)	(0.5981)	(0.6020)	(0.5938)	(0.5993)
Firm Characteristics						
Total volatility	0.5608	0.8454 *	0.8069 *	0.609	0.8947 *	0.8525 *
	(0.4564)	(0.4553)	(0.4457)	(0.4558)	(0.4564)	(0.4480)
Idiosyncratic volatility	-0.8604 *	-0.9953 **	-0.8972 **	-0.8262 *	-0.9608 **	-0.8637 **
	(0.4450)	(0.4852)	(0.4487)	(0.4318)	(0.4742)	(0.4363)
Leverage ratio	-0.0014	-0.0011	-0.0009	-0.0014	-0.0011	-0.0009
	(0.0009)	(0.0008)	(0.0008)	(0.0009)	(0.0008)	(0.0008)
R&D/sales ratio	-0.0633	-0.0591	-0.0708	-0.0553	-0.0516	-0.0619
	(0.0662)	(0.0614)	(0.0707)	(0.0607)	(0.0570)	(0.0648)
ROA	-0.4359	-0.501 *	-0.4779	-1.03 ***	-1.052 ***	-1.035 ***
	(0.2991)	(0.2876)	(0.2936)	(0.3562)	(0.3440)	(0.3513)
Size (USD trillions)	-0.3175	-0.7911	-0.8617	-0.3052	-0.7837	-0.8529
	(0.8518)	(0.8697)	(0.8639)	(0.8673)	(0.8848)	(0.8794)
Book/market ratio	-0.3299	-0.0271	-0.0897	-0.2253	0.0782	0.0127
	(0.7579)	(0.8795)	(0.8932)	(0.7618)	(0.8867)	(0.8977)
Sales/invested capital ratio	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006
	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Sales growth	-0.1131	-0.137	-0.1439	-0.1383	-0.1609	-0.1677
	(0.1108)	(0.1166)	(0.1146)	(0.1199)	(0.1243)	(0.1222)
Asset growth	0.051	0.0488	0.0634	0.0546	0.0521	0.0665
	(0.0629)	(0.0632)	(0.0603)	(0.0625)	(0.0627)	(0.0598)
Fixed effects						
Industry	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Concordance	0.627	0.608	0.608	0.63	0.611	0.612

Table 3
This table reports coefficients from the effect of personal cultural distance on executives' tenure. The dependent variable is the tenure (survival time) of the manager at the firm. Column (1) controls for the logarithmic total pay. Column (2) controls for the logarithmic incentive pay. Column (3) controls for the logarithmic non-incentive pay. Industry fixed effects are based on the Fama-French 12-industry classification. All other variables are defined in the Appendix. Standard errors are reported in parentheses and are clustered at the manager level. *** significant at 1%, ** significant at 5%, *significant at 10%.

	(1)	(2)	(3)	(4)	(5)	(6)
	All pay	Incentive pay	Non-incentive pay	All pay	Incentive pay	Non-incentive pay
Corporate Culture						
Personal Culture distance	0.0221 ***	0.0244 ***	0.024 ***	0.0218 ***	0.024 ***	0.0236 ***
	(0.0051)	(0.0051)	(0.0052)	(0.0051)	(0.0052)	(0.0052)
Personal Culture distance x ROA				0.0458	0.0518	0.05
				(0.0394)	(0.0399)	(0.0394)
Corporate Innovation culture	2.378	2.78 *	2.747 *	2.242	2.627	2.613
	(1.7100)	(1.6820)	(1.6350)	(1.7590)	(1.7420)	(1.6870)
Corporate Integrity culture	-5.295	-5.504	-8.37	-5.386	-5.618	-8.369
	(7.7970)	(7.7980)	(7.8610)	(7.8040)	(7.8070)	(7.8620)
Corporate Quality culture	2.162	2.454	2.542	2.132	2.42	2.505
	(2.8280)	(2.8090)	(2.7850)	(2.8360)	(2.8180)	(2.7940)
Corporate Respect culture	3.235	2.677	2.617	3.252	2.699	2.648
	(4.3730)	(4.5340)	(4.5610)	(4.3590)	(4.5170)	(4.5430)
Corporate Teamwork culture	6.783	6.467	6.954	6.563	6.226	6.696
	(4.4790)	(4.5030)	(4.3630)	(4.4920)	(4.5120)	(4.3850)
Corporate Community culture	-6.72 **	-6.483 **	-6.493 **	-6.681 **	-6.456 **	-6.463 **
	(3.1100)	(3.2010)	(3.1900)	(3.1130)	(3.2070)	(3.1960)
Corporate Hardwork culture	1.39	1.915	0.6759	1.186	1.679	0.4704
	(4.7700)	(4.9060)	(5.3640)	(4.7820)	(4.9220)	(5.3670)
Corporate Control culture	2.175	1.403	1.294	2.098	1.29	1.191
	(6.4090)	(6.4850)	(6.5050)	(6.4430)	(6.5310)	(6.5490)

Corporate Compete culture	3.939	4.33	4.371	3.913	4.299	4.335
	(3.0260)	(2.9740)	(2.9930)	(3.0310)	(2.9790)	(2.9980)
Managerial Characteristics						
Forced leave	1.703 ***	1.674 ***	1.714 ***	1.699 ***	1.67 ***	1.709 ***
	(0.1540)	(0.1492)	(0.1470)	(0.1542)	(0.1493)	(0.1473)
Manager age	0.0232 ***	0.0213 ***	0.0226 ***	0.0238 ***	0.022 ***	0.0233 ***
	(0.0044)	(0.0043)	(0.0044)	(0.0045)	(0.0044)	(0.0044)
Manager gender (female)	-0.0132	0.0142	0.0109	-0.0154	0.0115	0.0083
	(0.1265)	(0.1273)	(0.1269)	(0.1264)	(0.1272)	(0.1268)
Managerial Pay						
All pay (Log)	-0.1426 ***			-0.1409 ***		
	(0.0293)			(0.0293)		
Incentive pay (Log)		-0.025 *			-0.0239 *	
		(0.0130)			(0.0131)	
Nonincentive pay (Log)			-0.087 ***			-0.0845 ***
			(0.0275)			(0.0277)
Board Characteristics						
Board number of directors	-0.003	-0.0095	-0.0075	-0.0024	-0.0087	-0.0068
	(0.0158)	(0.0159)	(0.0158)	(0.0159)	(0.0159)	(0.0158)
Board gender ratio	0.1976	0.3483	0.4141	0.1991	0.3503	0.4124
	(0.3100)	(0.3103)	(0.3094)	(0.3097)	(0.3100)	(0.3092)
Board no. of qualifications	0.1464 *	0.1463 *	0.1384 *	0.1399 *	0.1384 *	0.1314 *
	(0.0783)	(0.0781)	(0.0784)	(0.0785)	(0.0783)	(0.0786)
Board network size	0.0001 **	0.0001 *	0.0001 *	0.0001 **	0.0001 **	0.0001 **
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Board attrition rate	1.327 **	1.396 **	1.394 **	1.323 **	1.392 **	1.389 **
	(0.5853)	(0.5733)	(0.5738)	(0.5856)	(0.5732)	(0.5741)

Firm Characteristics						
Total volatility	-0.493	-0.3746	-0.3035	-0.5088	-0.3926	-0.3218
	(0.4799)	(0.4661)	(0.4582)	(0.4802)	(0.4665)	(0.4589)
Idiosyncratic volatility	0.6424 ***	0.6415 ***	0.6695 ***	0.6413 ***	0.6412 ***	0.6678 ***
	(0.0966)	(0.0996)	(0.0984)	(0.0966)	(0.0996)	(0.0984)
Leverage ratio	-0.0007	-0.0007	-0.0005	-0.0007	-0.0007	-0.0005
	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0017)	(0.0017)
R&D/sales ratio	-0.0005	-0.0004	-0.0004	-0.0005	-0.0004	-0.0004
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
ROA	-0.5735 **	-0.5815 **	-0.5617 **	-0.7052 **	-0.7298 ***	-0.7052 **
	(0.2366)	(0.2357)	(0.2356)	(0.2749)	(0.2739)	(0.2745)
Size (USD trillions)	0.1353	-0.1926	-0.0967	0.0861	-0.2554	-0.1559
	(0.7838)	(0.8475)	(0.8584)	(0.7910)	(0.8573)	(0.8670)
Book/market ratio	-2.269 ***	-2.073 ***	-2.071 ***	-2.261 ***	-2.067 ***	-2.065 ***
	(0.5949)	(0.5751)	(0.5741)	(0.5961)	(0.5757)	(0.5748)
Sales/invested capital ratio	-0.015	-0.0152	-0.0158	-0.0151	-0.0155	-0.016
	(0.0115)	(0.0115)	(0.0117)	(0.0115)	(0.0115)	(0.0118)
Sales growth	0.0053 ***	0.0052 ***	0.0056 ***	0.0054 ***	0.0053 ***	0.0057 ***
	(0.0012)	(0.0012)	(0.0011)	(0.0012)	(0.0012)	(0.0012)
Asset growth	-0.0293	-0.0285	-0.0371	-0.0291	-0.0285	-0.0368
	(0.1130)	(0.1109)	(0.1128)	(0.1132)	(0.1111)	(0.1130)
Fixed effects						
Industry	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Concordance	0.648	0.635	0.641	0.649	0.635	0.642

Table 4
This table reports coefficients from the effect of personal cultural distance on executives' tenure removing all firms with any M&A activities in my sample. The dependent variable is the tenure (survival time) of the manager at the firm. Column (1) controls for the logarithmic total pay. Column (2) controls for the logarithmic incentive. Column (3) controls for the logarithmic non-incentive pay. Industry fixed effects are based on the Fama-French 12-industry classification. All other variables are defined in the Appendix. Standard errors are reported in parentheses and are clustered at the manager level. *** significant at 1%, ** significant at 5%, *significant at 10%.

	(1)	(2)	(3)
	All pay	Incentive pay	Non-incentive pay
Corporate Culture			
Cultural distance	0.0205 **	0.0228 ***	0.0203 **
	(0.0090)	(0.0088)	(0.0093)
Innovation culture	1.086	2.926	1.935
	(5.8650)	(5.0280)	(5.4030)
Integrity culture	-72.87	-86.97	-79.84
	(113.1000)	(117.6000)	(123.6000)
Quality culture	-11.51	-2.897	-5.863
	(19.8000)	(18.9400)	(17.8500)
Respect culture	-14.95	-19.48	-14.88
	(26.0400)	(25.5700)	(25.1400)
Teamwork culture	16.42	20.38 *	13.19
	(10.8800)	(10.4100)	(11.9000)
Community culture	18.89	16.03	18.15
	(15.4600)	(15.7500)	(16.1700)
Hardwork culture	20.19	17.65	16.04
	(27.0700)	(26.7900)	(27.0600)
Control culture	122.2 ***	114.6 ***	116.1 ***
	(33.0900)	(33.9300)	(32.8500)
Compete culture	20.64 *	21.85 *	22.05 *
	(11.4000)	(12.2800)	(11.7500)
Managerial Characteristics			
Forced leave	2.392 ***	2.264 ***	2.231 ***
	(0.5675)	(0.5675)	(0.5884)
Manager age	0.0591 ***	0.0612 ***	0.0607 ***
	(0.0154)	(0.0155)	(0.0151)
Manager gender (female)	-0.1456	-0.1477	-0.2536
	(0.3809)	(0.3958)	(0.3856)
Managerial Pay			
All pay (Log)	-0.2205		
	(0.1385)		
Incentive pay (Log)		-0.001	
		(0.0601)	
Nonincentive pay (Log)			-0.1822 **

			(0.0913)
Board Characteristics			
Board number of directors	-0.1017	-0.116	-0.0869
	(0.0813)	(0.0830)	(0.0825)
Board gender ratio	0.088	0.5125	0.2617
	(1.0920)	(1.0660)	(1.0700)
Board no. of qualifications	0.6716 *	0.6519 *	0.6667 *
	(0.3807)	(0.3683)	(0.3740)
Board network size	0.0002	0.0002	0.0002
	(0.0002)	(0.0002)	(0.0002)
Board attrition rate	1.308	1.143	1.308
	(1.9690)	(1.9540)	(1.9950)
Firm Characteristics			
Total volatility	-5.781 *	-5.306 *	-5.248 *
	(2.9840)	(2.8170)	(2.9590)
Idiosyncratic volatility	0.0266	0.0068	0.1993
	(2.9160)	(2.6870)	(2.7280)
Leverage ratio	0.0085	0.0085	0.0088
	(0.0097)	(0.0093)	(0.0095)
R&D/sales ratio	-0.0017	-0.0013	-0.0014
	(0.0019)	(0.0018)	(0.0019)
ROA	-1.47 *	-1.368 *	-1.364 *
	(0.8264)	(0.7849)	(0.8245)
Size (USD trillions)	-2.632	-7.081	-7.265
	(11.5800)	(12.5400)	(12.0000)
Book/market ratio	-3.676 **	-4.075 ***	-3.794 **
	(1.4810)	(1.5010)	(1.4840)
Sales/invested capital ratio	-0.0018	-0.0008	-0.0023
	(0.0634)	(0.0639)	(0.0634)
Sales growth	0.0049	0.0055 *	0.0055 *
	(0.0030)	(0.0031)	(0.0031)
Asset growth	0.3476	0.3138	0.2974
	(0.3606)	(0.3561)	(0.3502)
Fixed effects			
Industry	TRUE	TRUE	TRUE
Concordance	0.762	0.759	0.768

Table 5

This table reports the causal survival forests average treatment effect estimates for the effect of the culture distances on the tenure (survival time) of the manager in the new firm. Panel A reports the treatment effect, which is estimated as $\tau(x) = E[Y(T = 1) - Y(T = 0)|X = x]$, where $Y(T = 1)$ is the survival time for the treatment group and $Y(T = 0)$ is the survival time for the control group. In Column (1), I define $T = 1$ when the corporate cultural distance is greater than its median value in the sample, indicating a strong cultural distance. I define $T = 0$ when the corporate cultural distance is smaller than its median value in the sample, indicating a weak cultural distance. In Column (2), I define $T = 1$ when the personal cultural distance is greater than its median value in the sample, indicating a strong cultural distance. I define $T = 0$ when the personal cultural distance is smaller than its median value in the sample, indicating a weak cultural distance. X is the list of covariates with which two managers are matched into the same bin in the causal survival forests. Panel B reports the variable importance of the covariates when matching the managers into the same bin. The survival function is estimated with survival forests as $S(t, x) = P[Y > t \,|X = x]$. I convert Fama-French 12-industry classification to one-hot encodings, whose variable importance outputs are suppressed. All other variables are defined in the Appendix. Standard errors are reported in parentheses and are clustered at the manager level. *** significant at 1%, ** significant at 5%, *significant at 10%.

Panel A. Average treatment effects

	Corporate cultural difference	Personal cultural difference
	(1)	(2)
Average Treatment Effect	-0.6217 ***	-1.7734 ***
	(0.18)	(0.25)

Panel B. Variable importance measures

	Importance	
	Corporate cultural difference	Personal cultural difference
Corporate Culture		
Corporate Innovation culture	3.65%	1.63%
Corporate Integrity culture	1.74%	0.17%
Corporate Quality culture	1.76%	3.94%
Corporate Respect culture	3.83%	0.80%
Corporate Teamwork culture	3.39%	0.94%
Corporate Community culture	1.54%	0.73%
Corporate Hardwork culture	1.27%	1.50%
Corporate Control culture	5.11%	0.75%
Corporate Competition culture	8.47%	2.56%
Managerial Characteristics		

Forced leave	0.00%	0.02%
Manager age	3.29%	2.49%
Manager gender (female)	0.52%	0.02%
Managerial Pay		
All pay (Log)	2.32%	3.34%
Incentive pay (Log)	2.28%	3.19%
Nonincentive pay (Log)	5.19%	2.99%
Board Characteristics		
Board number of directors	2.34%	3.18%
Board gender ratio	6.61%	3.52%
Board no. of qualifications	3.09%	8.51%
Board network size	4.27%	11.84%
Board attrition rate	4.17%	1.09%
Firm Characteristics		
Total volatility	3.80%	3.58%
Idiosyncratic volatility	2.91%	4.01%
Leverage ratio	2.64%	5.32%
R&D/sales ratio	2.25%	1.67%
ROA	3.19%	4.35%
Size (USD trillions)	3.19%	3.78%
Book/market ratio	5.06%	7.60%
Sales/invested capital ratio	3.62%	5.84%
Sales growth	3.72%	7.04%
Asset growth	4.00%	2.80%

Table 6
This table reports the coefficients from Panel-OLS regressions for the lead-lag relationship between average personal cultural distance and firm value or performance. The dependent variables in Column 1-3 are the Tobin's Q, Return on Invested Capital, and Earnings per Share at time t respectively. All independent variables are at time t-1. All other variables are defined in the Appendix. Standard errors are reported in parentheses and are clustered at the manager level. *** significant at 1%, ** significant at 5%, *significant at 10%.

	(1)	(2)	(3)
	Dependent variable: Tobin's Q_t	Dependent variable: ROI_t	Dependent variable: EPS_t
Corporate Culture			
Cultural dispersion $_{t-1}$	-0.0415 **	-0.2999 *	-0.0529 *
	(0.0208)	(0.1801)	(0.0316)
Innovation culture $_{t-1}$	1.8116	-2.9225	3.0597
	(1.5994)	(8.2157)	(1.9502)
Integrity culture $_{t-1}$	4.2293	33.741	-0.0587
	(3.1106)	(21.2130)	(5.4327)
Quality culture $_{t-1}$	1.7884	6.3351	1.3725
	(1.2177)	(7.0584)	(1.6663)
Respect culture $_{t-1}$	1.4672	15.238	2.2295
	(2.2062)	(19.0640)	(3.3560)
Teamwork culture $_{t-1}$	0.9944	15.438	3.5187 *
	(1.4396)	(9.6640)	(2.0041)
Community culture $_{t-1}$	2.8807 *	22.006 *	5.9403 **
	(1.4975)	(12.9440)	(2.9649)
Hardwork culture $_{t-1}$	3.8359	33.74	11.8752 **
	(2.7344)	(20.8680)	(4.8720)
Control culture $_{t-1}$	7.134 **	2.0567	8.1606
	(3.2618)	(29.9380)	(5.3334)
Compete culture $_{t-1}$	2.1574 *	51.217	2.2498
	(1.1525)	(44.6240)	(1.6593)
Firm Characteristics			
Total volatility $_{t-1}$	-0.3621	1.2193	-0.3595
	(0.2981)	(2.9597)	(0.4261)
Idiosyncratic volatility $_{t-1}$	-0.0391 ***	-0.0611	-0.0074
	(0.0040)	(0.0574)	(0.0081)
Leverage ratio $_{t-1}$	-0.0002	0.0019	-0.0034
	(0.0002)	(0.0017)	(0.0021)
R&D/sales ratio $_{t-1}$	-0.0002 ***	-0.0001	-0.0002
	(0.0000)	(0.0001)	(0.0001)

Book/market ratio $_{t-1}$	-0.3008	-0.4348	-0.402
	(0.2189)	(0.4959)	(0.3972)
Sales/invested capital ratio $_{t-1}$	0.0002	0.0307	0.0002
	(0.0007)	(0.0289)	(0.0013)
Sales growth $_{t-1}$	-0.0014 ***	0.0000	-0.0006
	(0.0004)	(0.0009)	(0.0006)
Asset growth $_{t-1}$	0.0064	0.0544	0.0791 *
	(0.0386)	(0.1829)	(0.0459)
Board Characteristics			
Board number of directors $_{t-1}$	-0.0209	-0.3631	-0.0335
	(0.0157)	(0.3193)	(0.0483)
Board gender ratio $_{t-1}$	0.2328	16.138	-0.5102
	(0.3077)	(17.5430)	(0.5967)
Board no. of qualifications $_{t-1}$	-0.0764	-0.9255	-0.3451 **
	(0.0979)	(0.7976)	(0.1572)
Board network size $_{t-1}$	-0.0001	-0.0009	0.0000
	(0.0000)	(0.0010)	(0.0001)
Board attrition rate $_{t-1}$	-0.4134	17.73	-0.0289
	(0.2967)	(19.5480)	(0.5320)
Fixed effects			
Firm	TRUE	TRUE	TRUE
Year	TRUE	TRUE	TRUE
R2	0.0032	0.0021	0.0030

Table 7
This table reports the total unadjusted portfolio returns based on stocks sorted on the clean signal calculated from the negative of cultural dispersion in companies. Portfolio 1 includes the stocks with the lowest negative cultural dispersion (or equivalently, the highest cultural dispersion), and portfolio 5 consists of the stocks with the highest negative cultural dispersion (or equivalently, the lowest cultural dispersion). The benchmark portfolio return is the average unadjusted returns of stocks in my portfolio weighted by their weight in the CRSP U.S market index.

	Equally-weighted portfolio			Value-weighted portfolio		
	Return, Annual	Risk, Annual	Sharpe Ratio	Return, Annual	Risk, Annual	Sharpe Ratio
Portfolio 1 (Short)	12.01%	19.90%	0.603	8.13%	14.78%	0.550
Portfolio 2	8.37%	21.11%	0.397	4.53%	16.29%	0.278
Portfolio 3	14.37%	20.87%	0.689	10.43%	17.51%	0.596
Portfolio 4	9.31%	18.84%	0.494	7.04%	16.72%	0.421
Portfolio 5 (Long)	14.24%	19.36%	0.736	12.65%	16.30%	0.776
Portolio [5-1]	2.23%	5.87%	0.380	4.53%	7.34%	0.616
Long-only Portfolio	12.25%	19.32%	0.115	12.25%	19.32%	0.234
Benchmark Portfolio	7.79%	14.63%	0.152	7.79%	14.63%	0.309
Signal-weighted Portfolio	1.25%	9.31%	0.135	1.25%	9.31%	0.135

Table 8

This table reports coefficients from Logistic regressions for the effect of cultural distances between firms on the probability of managerial turnovers. Columns 1-3 report the effect of corporate cultural distances, and Columns 4-6 report the effect of personal cultural distances. The dependent variable is a dummy variable that indicates if the manager has left the firm before the sample period ends. Columns (1) and (4) include logarithmic total pay. Columns (2) and (6) include logarithmic incentive pay. Columns (3) and (7) include logarithmic non-incentive pay. All other variables are defined in the Appendix. Standard errors are reported in parentheses and are clustered at the manager level. *** significant at 1%, ** significant at 5%, *significant at 10%.

	(1)	(2)	(3)	(4)	(5)	(6)
	Independent variable: Corporate cultural distance			Independent variable: Personal cultural distance		
	All pay	Incentive pay	Non-incentive pay	All pay	Incentive pay	Non-incentive pay
Corporate Culture						
Cultural distance	0.0421 *	0.0484 **	0.0487 **	0.1069 ***	0.1176 ***	0.1158 ***
	(0.0233)	(0.0232)	(0.0234)	(0.0190)	(0.0210)	(0.0202)
Innovation culture	-0.135	1.1255	-1.1021	-0.2757	0.158	-0.0483
	(8.1921)	(8.1557)	(8.1579)	(3.1686)	(3.1905)	(3.0820)
Integrity culture	42.823 *	37.873 *	38.974 *	-13.456	-13.646	-22.059 *
	(23.1090)	(22.8750)	(23.3360)	(12.2950)	(12.5040)	(13.2690)
Quality culture	3.3989	3.9848	3.9036	-0.5979	0.2649	0.3427
	(6.4960)	(6.4655)	(6.5395)	(3.3128)	(3.2726)	(3.2580)
Respect culture	-7.6899	-6.3675	-4.6116	4.6039	3.9788	4.8801
	(15.0480)	(14.7430)	(15.0460)	(6.4767)	(6.5540)	(6.3397)
Teamwork culture	8.4649	7.5391	4.259	3.3373	2.5751	3.1463
	(13.1010)	(13.1600)	(13.5620)	(4.7587)	(4.8600)	(4.7503)
Community culture	-20.734 **	-21.161 ***	-20.887 **	-13.242 ***	-12.265 ***	-12.502 ***
	(8.0575)	(8.1563)	(8.1993)	(3.5909)	(3.7486)	(3.6990)
Hardwork culture	-20.884	-22.13	-23.789	-3.7652	-2.9192	-7.5042
	(16.5390)	(16.0800)	(16.5440)	(7.4829)	(7.2015)	(8.5732)
Control culture	-4.2469	-8.7319	-9.5896	-4.9579	-6.3719	-6.9038
	(14.7780)	(14.6240)	(14.5960)	(7.5612)	(7.5716)	(7.6001)

Compete culture	3.4719	4.0974	4.1245	2.1099	2.1003	2.2219
	(7.4330)	(7.2345)	(7.1890)	(3.5685)	(3.6436)	(3.4533)
Managerial Characteristics						
Forced leave	3.0633 ***	2.8262 ***	2.8838 ***	5.1149 ***	5.0023 ***	5.0273 ***
	(0.3841)	(0.3862)	(0.3938)	(0.5917)	(0.5926)	(0.5963)
Manager age	0.0288 ***	0.0262 ***	0.0279 ***	0.0354 ***	0.0286 ***	0.033 ***
	(0.0075)	(0.0075)	(0.0075)	(0.0052)	(0.0051)	(0.0052)
Manager gender (female)	0.2646 **	0.3221 **	0.3173 **	-0.06	0.0009	-0.0222
	(0.1293)	(0.1281)	(0.1279)	(0.1557)	(0.1567)	(0.1575)
Managerial Pay						
All pay (Log)	-0.3164 ***			-0.3431 ***		
	(0.0402)			(0.0408)		
Incentive pay (Log)		-0.0866 ***			-0.0972 ***	
		(0.0158)			(0.0135)	
Nonincentive pay (Log)			-0.2161 ***			-0.237 ***
			(0.0446)			(0.0375)
Board Characteristics						
Board number of directors	0.0086	-0.0063	0.0011	0.0373 **	0.027	0.0312 *
	(0.0220)	(0.0218)	(0.0220)	(0.0185)	(0.0183)	(0.0184)
Board gender ratio	0.9494 **	0.9716 **	1.1857 ***	0.5867 *	0.7955 **	0.9903 ***
	(0.3976)	(0.3985)	(0.3975)	(0.3509)	(0.3507)	(0.3559)
Board no. of qualifications	0.0413	0.0677	0.0511	0.1684 *	0.1943 **	0.171 *
	(0.1061)	(0.1063)	(0.1061)	(0.0894)	(0.0900)	(0.0900)
Board network size	0.0002 ***	0.0002 ***	0.0002 ***	0.0002 ***	0.0001 ***	0.0001 **
	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
Board attrition rate	1.9535 ***	1.6968 **	1.8154 **	3.047 ***	3.0195 ***	2.9575 ***

	(0.7066)	(0.6990)	(0.7086)	(0.6796)	(0.6752)	(0.6790)
Firm Characteristics						
Total volatility	0.7569	1.1527	1.2451 *	-0.9385	-0.695	-0.4672
	(0.7353)	(0.7171)	(0.7191)	(0.6871)	(0.6804)	(0.6707)
Idiosyncratic volatility	-1.0818 *	-1.2537 **	-1.1092 **	0.6753 *	0.6075 *	0.8067 **
	(0.5599)	(0.6078)	(0.5521)	(0.3543)	(0.3544)	(0.3751)
Leverage ratio	-0.0009	-0.0007	-0.0005	-0.0002	-0.0003	0.0007
	(0.0012)	(0.0011)	(0.0010)	(0.0017)	(0.0016)	(0.0020)
R&D/sales ratio	-0.0362	-0.0318	-0.0407	0.0001	0.0002	0.0002
	(0.0429)	(0.0414)	(0.0479)	(0.0013)	(0.0013)	(0.0013)
ROA	-0.412	-0.4938	-0.4436	-1.1063 ***	-1.1542 ***	-1.0789 ***
	(0.3696)	(0.3564)	(0.3611)	(0.3448)	(0.3557)	(0.3540)
Size (USD trillions)	-0.4295	-1.1121	-1.0878	-0.5541	-1.0427	-0.6595
	(1.0479)	(1.1165)	(1.1004)	(0.9955)	(1.0587)	(1.1132)
Book/market ratio	-0.6528	-0.1843	-0.2763	-1.8653 *	-1.2116	-0.8941
	(0.9414)	(0.9341)	(0.9889)	(1.0747)	(1.0579)	(1.1238)
Sales/invested capital ratio	0.0009	0.0009	0.001	0.0074	0.0062	0.0073
	(0.0021)	(0.0021)	(0.0021)	(0.0160)	(0.0159)	(0.0164)
Sales growth	-0.1525	-0.173	-0.1844	0.0121 ***	0.0113 **	0.0123 ***
	(0.1463)	(0.1532)	(0.1467)	(0.0046)	(0.0045)	(0.0041)
Asset growth	0.0657	0.0629	0.071	-0.0549	-0.0497	-0.0745
	(0.0731)	(0.0729)	(0.0677)	(0.1433)	(0.1384)	(0.1424)
AIC	4446.9	4483.9	4485.1	5363.9	5416.2	5420.7

Chapter 3

Do Investors Pay Less Attention to Women (Fund Managers)?

Do Investors Pay Less Attention to Women (Fund Managers)?

P. Raghavendra Rau and Jinhua Wang *

Abstract

We document a gender-based attention effect in the sensitivity of mutual fund flows to fund performance using individual-level fund data from a fintech platform in China. Investors increase (decrease) flows to funds following positive and strong (negative and weak) prior-month performance. However, although there is no significant difference in the performance of male and female managers, the sensitivity effect significantly weakens if the fund manager is female. The effect persists after controlling for managerial characteristics and fund objectives, as well as individual investor fixed effects. Simply put, investors react less to the performance of female fund managers.

Keywords: Mutual funds, flow-performance relationship, attention bias, gender bias, fintech, inclusive finance, behavioral finance, psychology, Natural Language Processing

* Rau and Wang: Cambridge Judge Business School, University of Cambridge, Cambridge CB2 1AG, UK, r.rau@jbs.cam.ac.uk and jw983@jbs.cam.ac.uk. All user data in our research is proprietary and provided by an anonymous fintech company. No raw data was provided. The randomized data provided by the company are anonymised and strictly protected on their servers. We thank David Chambers, Naveen Daniel, Sue Denim, Oguzhan Karakas, Lalitha Naveen, Alexandra Niessen-Ruenzi, Lin Peng, Devaki Rau, Stefan Ruenzi, Pedro Saffi, Jason Sandvik, and seminar participants at ESCP Paris, Korea University Business School, and the University of Cambridge for their helpful comments and suggestions.

1. Introduction

We document the existence of a new different and previously unstudied type of gender bias in investor behavior, that we term gender-based attention bias. Gender-based attention bias refers to the tendency to pay less attention to women than men. The literature in psychology and the social sciences on gender issues argues that this type of attention bias is manifested at both the personal and professional levels. At the personal level, the literature documents that boys and girls receive differential attention in families, especially in developing countries. For example, Barcellos, Carvalho, and Lleras-Muney (2014) find that boys receive more childcare time than girls, they are breastfed longer, and they get more vitamin supplements in India. At the professional level, Cortina (2008), for example, argues that women are more likely to be ignored or interrupted or experience their contributions being belittled than men within organizations. This type of bias has been shown to exist among attorneys (Cortina et al., 2002), university faculty (Richman et al., 1999), and court employees (Cortina et al., 2001). Bigelow et al. (2014) report that attention bias also seems to exist in women-led initial public offerings. They show that female CEOs are disproportionately disadvantaged in their ability to attract growth capital, perceived as less capable than their male counterparts, and their firms are considered less attractive.

Examining gender bias in corporate finance is challenging because it is difficult to establish how much performance depends on an individual executive's skill and effort – especially when so many executives contribute to a firm's value. In contrast, in mutual funds with a single manager, it is relatively straightforward to attribute performance to managerial effort and to relate investment flows (rewards) to performance (effort).

In this paper, we examine whether gender-based attention bias affects the well-documented flow-performance relationship in the mutual fund literature. Sirri and Tufano (1998) and Del Guercio and Tkac (2002), among others, document that there is a positive correlation between prior mutual fund performance and subsequent fund flow, commonly termed the flow-performance sensitivity of the fund. Money chases after the best performing funds in the previous month, leaving funds with poor performance. Unfortunately, it has been difficult to examine whether investor fund flows are affected by gender bias because of a lack of data on investor-level fund flows. Consequently, the flow-performance literature typically focuses on aggregate level fund-level flows, making it difficult to isolate individual investors' differential responses towards fund managers' gender.

In contrast to the previous literature, we examine whether the flow-performance sensitivity for *individual investors* is affected by managerial gender, using a unique dataset provided by a large fintech platform in China. The fintech platform allows users to make and receive payments to and from other people and businesses, and also provides in-app access to a variety of saving and investment products with different risk levels. The investment products offered on the platform include open-ended mutual funds managed by fund managers affiliated to fund management companies independent of the platform. The app allows the investors to rank funds on the basis of raw returns over the prior one-, three-, six,- and 12-month return horizons. Importantly, it provides clear information on the gender of the fund managers (including a photograph in a large number of cases) to the investors at the time of investment. Prior papers that have examined the effect of gender on fund flows are unable to establish that fund investors are even aware of who is managing their fund. In China, this information is almost the first piece of information investors receive when choosing to invest on the app.

Our data are based on a random sample of investors drawn from the platform and consisting of 172,672 retail investors' monthly investment positions in 253 domestic stock funds over the period from August 2017 to July 2019, for a total of 2.35 million user-fund-month observations. The database contains monthly data on each individual investor's fund investment and redemption amounts and details of capital gains or losses for every fund owned by the investor in that month. It also contains individual characteristics of each investor on the platform, such as their gender, ages, monthly payment amounts, places of residency, and risk aversion levels (surveyed through a questionnaire upon users' registration on the platform).

We first document that the flow-performance sensitivity for individual investors in China is similar to patterns documented in the prior literature for funds elsewhere in the world. Our fund-flow and performance measures are similar to those defined by Sirri and Tufano (1998), except that our measures are defined at the investor-month level, while their measures are defined at the fund-year level. Our findings are consistent with those documented by Hong, Lu, and Pan (2020), who use public Chinese fund data to document that the Chinese fund market is also characterized by performance chasing behavior which markedly increases after the introduction of fintech platforms in China.

We next document strong evidence of a differential flow-performance sensitivity between male- and female-managed funds. Interacting the flow-performance sensitivity term with a gender dummy variable, we show that flow-performance sensitivity is significantly weaker for female-managed funds. We document decreased flow sensitivity to performance for women across all the raw return horizons, from one- to 12-months, the fund allows the investor to sort over. Simply put, when female-managed funds do well, they experience significantly lower fund inflows than male-managed funds. However, when these funds perform poorly, they experience relatively lower fund outflows than male-managed funds. The attention bias is reduced when performance drops, suggesting that investors pay less attention to female than male managers when performance increases but punish both sets of managers when performance drops. Using piecewise regressions, we show that the decrease in flow sensitivity appears to exist across all levels of performance for female managers except for the very top managers.

Our results are robust to controlling for fixed effects at the investor level, allowing us to address potential omitted variable concerns arising from differing investor backgrounds or personalities. The results are also robust to including time fixed effects, reducing a potential omitted variable bias caused by common economic shocks. All the regressions also control for manager characteristics, such as their educational backgrounds and their length of tenures as managers for the funds, and for fund characteristics such as fund objectives, fund size, and fees.

An alternative rational explanation for the muted flow-performance sensitivity for female managers is that male managers' past performance is a better predictor of their future performance. Hence investors rationally invest in what they believe will be better-performing funds. We show, however, that past performance does not predict future performance for Chinese fund managers and, moreover, that there is no gender difference in the predictive ability for future performance. In addition, controlling for managerial characteristics and time fixed effects, we find that female fund managers perform significantly better (at the 10% level) than male fund managers when computing rankings at the one month period. In other words, investors appear to believe without evidence that better-performing male managers are more likely to perform well than female managers.

A second alternative explanation is that female managers are less likely to take risks and hence investors who do not adequately adjust for risk direct more flows to (risky) male managers. Regressing a battery of risk measures on manager gender, controlling for managerial characteristics and time fixed effects, we find no significant differences in levels of idiosyncratic or systematic risk or risk-adjusted performance between female and male-managed mutual funds. Evans and Sun (2020) show that retail investors use simple risk-adjustment heuristics to direct fund flows. We show that the differential flow-performance relationship continues to exist when we rank funds based on plausible heuristics such as risk-adjusted returns measured by Jensen's Alpha (Jensen, 1968).

The Chinese sample of fund managers is similar to the data studied in other countries in that only 15% of the sample of managers are female. For example, Niessen-Ruenzi and Ruenzi (2018) analyze the gender-fund flow relationship in a sample where only 13.8% of the mutual fund manager sample is female. Hence, we investigate if the attention bias arises because investors face higher search costs when searching for female fund managers. For example, an investor who is actually gender-neutral towards the choice of fund manager might appear biased because of the high search costs involved in finding female managers. We note that, in univariate two-sample p-tests, female-managed funds consistently outperform male-managed funds over 1-, 3-, 6-, and 12-month horizons, suggesting that female-managed funds tend to be ranked higher on the app.

To explicitly address the visibility problem, we use two approaches. First, we match each female manager to a similar male manager in each month using a propensity-score matching (PSM) approach using a host of managerial and fund characteristics as proxies for manager visibility. Although PSM does not allow us to establish causality, the covariates of managerial and fund characteristics are well balanced in the matched sample. Attention bias continues to be significant in this matched fund sample, where search costs for male and female managers are likely to be approximately similar.

Second, we use a Natural Language Processing (NLP) technique to extract names from 400,000 financial news articles in Chinese and count the frequency of each manager's name appearance in the article each month. We show that the level of attention bias is unaffected by the level of media coverage. We also measure the frequencies of positive media mentions and negative media mentions of each fund manager through sentiment analysis based on a variant of the Transformer model (Vaswani et al. 2017) in machine learning. We show that, while the sign of the media mention does not appear to affect the level of attention bias, positive mentions of fund managers in the news strengthens the flow-performance relationship on average. However, we continue to find evidence that controlling for performance and positive media mentions, female manager earn lower fund flows than male managers.

One might reasonably assume that female investors may be less subject to this attention bias than male investors, as the prior literature (Lovén, Herlitz, and Rehnman, 2011) suggests that female investors will more naturally identify with and be biased towards female fund managers. We find weak evidence that female investors exhibit less gender bias towards female managers than male investors. Similarly, users from smaller cities appear more subject to the gender attention bias than users from larger cities. In contrast, gender bias seems to be unrelated to user age or risk aversion. Furthermore, we show that attention bias appears to be innate to investors. In a sub-sample of all *first*-time investors, we show that at the time of the first investment, investors are less likely to invest in female-managed funds for the same level of performance.

Is there a causal relationship between the gender identity of female managers and investor fund flows? To examine this question, we employ a three-stage instrumental variable regression approach suggested by Wooldridge (2001). We use two different instrumental variables, the first being the proportion of illiterate women amongst all women in the municipal district that the investor resides in, and the second being the proportion of female new-borns amongst all new-borns in the municipal district that the investor resides in. Both instrumental variables do not instrument for fund manager per se, but for the specific investor's choice of a fund manager. The instruments do not directly drive investors' fund flow decisions but are likely to be related to investors' biases on gender identities, conditional on investors' characteristics that we control for.

Specifically, in the first stage, we estimate a logit regression where we model the choice of fund manager gender using the instrumental variables, the proportion of illiterate women and the proportion of female new-borns as explanatory variables. In the second stage, we compute the fitted probability of choosing a female manager from the first-stage logit. In the third stage, we use the fitted probability to instrument for manager gender and interact with performance of the fund. Our instrumental variable regressions confirm the existence of gender-based attention biases away from female managers, which cause investors to pay less attention to female-managed funds.

Finally, we formally run a regression testing the difference of individual fund flow volatilities between male- and female-managed funds. Our results show that individuals holding female-managed funds exhibit lower fund flow volatilities throughout our sample period. For mutual fund companies, this may have the desirable impact of lowering the volatility of flows into the fund.

The literature that documents the existence of gender bias in executive performance (such as CEOs) suffers from the handicap that it is impossible to clearly attribute firm-level performance to individual executive efforts. In contrast, in the mutual fund industry, the performance of a sole-managed mutual fund is clearly attributable to the manager. However, in the absence of data on fund flows from specific individuals to specific funds, it is again not possible to relate fund flow at the user level to performance at the fund level. With its unique dataset matching user-level flows to specific funds, this paper is the first to document the existence of a gender-based attention bias away from women in the professional finance industry.

Our study makes four additional contributions to the finance literature. First, our paper contributes to a rich literature on gender differences in investment behaviour. Early research in this area focuses mainly on the difference in performance between male and female investors. We add to this literature by documenting differences between male and female investors in their levels of attention bias on the basis of gender.

Second, this study complements the literature on mutual fund flows associated with search costs and manager heterogeneities. Chevalier and Ellison (1999) show that mutual fund managers who attended higher SAT undergraduate institutions have systematically higher risk-adjusted excess returns. Huang and Wang (2015) show that manager fixed effects predict future fund performance, and investors reward managers with higher fixed effects by directing flows to the funds they manage. Our research complements their study by showing that manager gender, tenure, and education backgrounds, which are included in manager fixed effects, also have significant impacts on fund flows in our sample. However, those managerial factors are independent of the attention bias documented in this paper.

Third, this study complements the existing literature on the mutual fund performance-flow relationship. Berk and Green (2004) show that investors learn from fund managers' past performance and allocate funds accordingly. Sirri and Tufano (1998) show that investors chase after funds with higher relative performance in the previous year. Bailey, Kumar, and Ng (2011) show that mutual fund investors are subject to behavioral biases. We add to this literature and show that the fund performance-flow relationship is affected by gender-driven attention bias. Atkinson, Baird, and Frye (2003) find that although male- and female-managed funds do not differ significantly in terms of performance, risk, and other fund characteristics, net asset flows into fixed income funds managed by females are lower than for males. Similarly, Niessen-Ruenzi and Ruenzi (2018) document significantly lower inflows into female-managed funds than male-managed funds at the aggregate annual fund level. However, both these papers are unable to directly establish that investors are even aware of who is managing their funds, let alone that investors focus on the gender of these managers. In our setting, in contrast, manager identity and gender are extremely salient when the investor is making the investment decision. In addition, given that the gender composition in our sample is similar to that in other settings, it seems reasonable to believe these results would extend elsewhere, with the caveat that this would assume that there are no significant differences in gender biases across countries and cultures.

Most important, this study adds to the existing literature on gender issues in the finance industry. Adams and Kirchmaier (2016) document that there is a lower fraction of women on the board for firms in the STEM and Finance sectors than in the non-STEM sector. Rau, Sandvik, and Vermaelen (2021) show that initial public offerings by firms with gender diverse boards suffer significantly greater underpricing at the offering than firms with only male boards. Adams and Funk (2012) show that, unlike the well-documented fact that women are more risk-averse in the general population, women in the boardroom are more risk-loving and less security-oriented than their male counterparts. The gender bias appears particularly strong in the investment fund industry. In 2019, for example, women accounted for 37.5% of all lawyers, 49% of judges, 34.5% of economists, 19% of surgeons, and 26% of chief executives, according to the U.S. Census Bureau.[1] In contrast, the percentage of funds managed by women has barely changed. It was 10.3% in 2016 and 11% in 2020.[2] While there are several explanations for the employment gap between men and women in various industries[3], Niessen-Ruenzi and Ruenzi (2018) propose a customer-based discrimination explanation specifically for the mutual fund industry. Because mutual fund investors appear to direct significantly lower flows to female-managed mutual funds than to male-managed funds, they argue that, in response, rational fund companies might choose to hire fewer women since fund companies generate their profits from fees charged on assets under management. In contrast, we show that the attention bias works both ways. Though investors appear more sensitive to fund performance when the fund manager is male, the sensitivity is bi-directional. Investors are less sensitive to underperforming female managers. For mutual fund companies, this may have the potential to lower the volatility of flows into the fund.

The remainder of the paper is organized as follows. Section 2 documents the literature in psychology and the social sciences on gender-based attention biases. Section 3 describes our data and the measure of fund flows and fund performance. Section 4 presents our main empirical analyses. Section 5 concludes.

2. Literature on attention bias

There are three strands of literature in the psychology and social sciences on gender issues that are related to attention bias.

---

[1] Data available at the US Census Bureau.

[2] Data available at Citywire Alpha Female Report, 2020.

[3] Examples include including hiring discrimination against women (Goldin and Rouse, 2000), occupational choice by women into other professions (Polachek, 1981) gender differences in the willingness to compete (Sutter and Gätzle-Rützler, 2014), or career discontinuities (Bertrand, Goldin, and Katz, 2010).

The first strand examines whether boys and girls in families receive different levels of attention, especially in developing countries. Barcellos, Carvalho, and Lleras-Muney (2014) find that boys receive more childcare time than girls, they are breastfed longer, and they get more vitamin supplements in India. Park and Rukumnuaykit (2004) use nutrient intake data from the China Health and Nutrition Survey to show that rural fathers, especially less educated men, favour sons while rural mothers do not. These findings suggest that there are geographic differences in the level of attention bias away from women.

The second strand examines gender stereotypes and biases in households and corporations. Hannum, Kong, and Zhang (2009) use survey data to show that the vast majority of mothers in their sample expect to reply on sons for support in their old age, and nearly one in five mothers do not expect girls to go to school in rural China, suggesting one reason why more attention is paid to boys than girls. They also show that parents view boys as having greater talent than girls. In a random experiment on judgments of fame, Banaji and Greenwald (1995) show that subjects were more likely to assign fame to male than female names. At the professional level, Cortina (2008) argues that women are more likely to be being ignored or interrupted or experience their contributions being belittled than men within organizations. Neumark and Bank (1996) show that men and women are treated differently in job applications and women are less likely to be hired. Newton and Simutin (2015) show that CEOs pay executives of the opposite gender less than executives of their own gender, and older and male CEOs exhibit the greatest propensity to differentiate based on gender.

The final strand examines gender-based double standards. Botelho and Abraham (2017) use lab-based evidence to show that double standards disadvantage women when evaluators face heightened search costs related to the number of candidates being compared to or higher levels of uncertainty stemming from variation in the amount of pertinent information available. Botelho and Gertsberg (2020) use a quasi-natural experiment to show that women are disadvantaged in the evaluative process and are given lower ratings on Yelp. Given the low number of female to male managers in the mutual fund space, the search costs for female managers are likely to be higher than those for male managers. Hence, when evaluating female fund managers' performance, investors may believe the lack of female managers in the profession is a sign that female managers are less competent than male managers.

3. Data

Our research is based on a random sample of user investments into stock funds supplied by a large anonymous fintech platform based in China. The fintech platform allows users to make and receive payments to and from other people and businesses though a smartphone application (app) interface, and also provides in-app access to a variety of saving and investment products with different risk levels.

The fintech platform does not have in-house fund managers itself. It only serves as a portal to fund investments with significantly lower (typically a tenth of) transaction fees than traditional brokers. Investors can choose from a variety of fund types, including stock funds, currency funds, index funds, hybrid funds, and Qualified Domestic Institutional Investor (QDII) funds. The lower risk investment products available on the app include zero-interest and risk-free savings as well as low-interest currency funds, while the higher risk investment products include open-ended mutual funds managed by fund managers affiliated to fund management companies independent of the platform.

In this paper, we focus on investments by actively managed stock mutual funds because of the wealth of extant research on the fund-flow relationship in actively managed stock funds and to avoid biases caused by differential liquidities and risks among the different types of underlying assets. The app provides information on the fund managers (including a photograph in a large number of cases) to the investors and allows the investors to rank funds on the basis of raw returns over the prior one-, three-, six- and 12-month raw return horizons.

Figures 1 and 2 provide screenshots of the typical user experience when they access the smartphone app. When investors open the app to invest, the app presents to them a page listing funds ranked by past performance. Investors can choose to rank fund performance by their objectives and over the past 1-, 3-, and 6-month horizons, as shown in Figure 1 panels A, B, and C, respectively. When investors scroll down to the bottom of the list, a second page is automatically loaded by the platform and presented to investors immediately, an experience termed an “infinite scroll”. While the platform does not alter fund rankings through fund advertisements or promotions, investors can search for a fund’s name and bypass the infinite scroll list if they learn the fund’s name through advertising elsewhere on the internet. If investors click on a specific fund on the list, they will be further shown a second fund profile page, where they can read a short description of each fund manager listing the name, gender, education background, tenure at the fund, and company. Most managers also have photographs on their profile pages, as shown in Figure 2 Panels A and B. It is also fairly easy for investors to infer the fund manager’s gender from Chinese names in the extremely rare cases where both the fund manager description and profile images are not available. We note that prior literature on gender biases on mutual fund flows are unable to directly establish that investors are aware of who is managing their funds, let alone focus on the gender of these managers. For example, Niessen-Ruenzi and Ruenzi (2018) rely on a controlled laboratory experiment to establish that gender bias exists in the experimental setting and extrapolate the results to the general population. Figure 2 Panels A and B show that, in our setting, manager identity and gender are extremely salient when the investor is making the investment decision.

The fund profile page also presents the detailed ranking of a fund among funds within the same objective or category over past 1-, 3-, 6-, and 12-month horizons, as shown in Figure 2 Panel C.

The random sample we acquire from the fintech platform is largely representative of the mainland Chinese population. Figure 3 Panels A and B depict the geographic distributions of the sample and the Chinese population in mainland China. Most investors in our sample are concentrated in the three major economic regions in mainland China: The Beijing-Tianjin-Hebei Economic Zone, the Yangtze River Delta region and the Pearl River Delta region, which is consistent with the Chinese population distribution. Panel C depicts the geographic distribution of average stock fund investment amounts, while Panel D shows the geographic distribution of the users' monthly average profits. The province (city) with the highest average monthly capital gain is Beijing, and the province (city) with the highest average fund amount invested is Shanghai.

3.1 Sample construction

Our sample consists of two main databases supplied by the platform. The first database documents monthly investment positions in 253 stock funds for 172,672 retail investors on the platform over the period from August 2017 to July 2019. Over this period, the Shanghai stock index rose by 6.3% between July 2017-January 2018, dropped by 25.7% over the year 2018, and rose by 11.7% from January 2019 to the end of our sample period.

The investment position database contains each individual investor's invested amount in each fund, capital gains or losses experienced over the month, and investment and redemption amount for each fund at the end of each month. We exclude hybrid funds, index funds, and other fund types from the sample, focusing only on actively managed stock funds. We also exclude funds that are co-managed by multiple fund managers. Finally, we eliminate funds where there is only a single female manager across the sample in that fund objective. Fund objectives identify the core stocks that a fund manager targets when forming the portfolio. For example, income funds target stocks that pay high dividends, while growth funds target stocks that are likely to increase in value over time. Appreciation funds target stocks that both pay high dividends and increase in value over time. Our final sample consists of funds with the following objectives: Appreciation, Stable Growth, Growth, and Income.

The second database documents individual characteristics of each investor on the platform, such as their gender, age, monthly payment amount, place of residency and risk tolerance levels.[4] We match the two databases by investors' unique (anonymized) IDs as well as fund codes.

Next, we match this sample to three China Stock Market & Accounting Research (CSMAR) Databases: the fund finance database, the fund manager database, and the fund evaluation database. The fund finance database documents the balance sheets and income statements of funds, including management fees, sales fees, and transaction fees at the fund level. The fund manager database documents the start and end dates of each manager's tenure at each fund. It also includes managers' characteristics, such as their gender and degree of education. The fund evaluation database documents the monthly Net Asset Value (NAV) for each fund, adjusted for dividends, splits and reinvestments. The fund evaluation database also provides CAPM risk-adjusted returns of funds, also known as alphas. We merge the platform database to the fund finance and the fund evaluation databases through fund codes and trading months. We merge the platform database to the fund manager database based on fund code if the trading month falls between the start and end dates of the manager's tenure at the fund.

Finally, to create our instrumental variables, we merge our data with the National Bureau of Statistics of China (NBSC)'s Census Data (2011) on the proportion of illiterate women amongst all women and the proportion of female new-borns amongst all new-borns in different municipal districts. We merge the census data with the primary municipal districts of residence of the platform users in our sample. Figure 4 illustrates the census data at the province level.

3.2 Measures of fund flow and fund performance

We construct our measure of fund flow using individual-level data provided by the platform. Our definition of fund flow is similar to the definition by Sirri and Tufano (1998), except that our fund-flow is defined at the individual level:

$$Flow_{i,f,t} = \frac{Fund\ Amount_{i,f,t} - Fund\ Amount_{i,f,t\text{-}1} - Capital\ Gain\ or\ Loss_{i,f,t}}{Fund\ Amount_{i,f,t\text{-}1}} \quad (1)$$

[4] The randomized raw data sample is only accessible through the fintech platform and cannot be downloaded by researchers. It is impossible for researchers to identify the true identity of any specific investor from the data.

where i indexes investors, f indexes funds, and t indexes time. $Fund\ Amount_{i,f,t}$ represents investor i's position in fund f at the end of the current month t, while $Fund\ Amount_{i,f,t\text{-}1}$ is the same variable lagged by one month. $Capital\ Gain\ or\ Loss_{i,f,t}$ is the capital gain or loss that investor i incurred in fund f and month t at the end of the current month.[5] To remove outliers arising from fund conversions, we winsorize fund flows at the 99.9% level and the 0.1% level.

Because investors are able to make investments and redemptions frequently during the month, and we only have month-end data on individual fund holdings, it is impossible to calculate their actual return on investments using their month-end capital gains or losses. Therefore, we follow Bollen and Busse (2005) and compute funds' adjusted NAV returns as proxies for investors' prior returns on investments:

$$NAV\ Return_{f,t\text{-}1,\text{1-month}} = \frac{NAV_{f,t\text{-}1} - NAV_{f,t\text{-}2}}{NAV_{f,t\text{-}2}} \tag{2}$$

where f indexes funds, and t indexes time. The adjusted NAV represents the fund's Net Asset Value, the unit price of one share of the fund, after adjusting the fund NAV for dividends paid over the month.

Fund rankings are displayed on the platform as "infinite scrolls", as shown in Figure 1. As users scroll down the list, more funds are loaded on the screen instantly. According to their preferences, users can choose to display fund rankings by past 1-month, 3-month or 6-month returns (though not 12-month returns). Over the period of our analysis, the default ranking was one month. However, users can also view individual fund rankings amongst funds of the same category or objective by past 1-month, 3-month, 6-month, or 12-month returns.

We follow Sirri and Tufano (1998) in ranking fund performance within each fund objective. To check the robustness of our results, we also create three performance metrics over longer investment horizons for our performance-flow panel regressions, including 3-, 6-, and 12-month fund returns ranked among funds with the same objectives, calculated as follows:

$$Rank_{f,t\text{-}1,\text{3-month}} = Percentage\ Rank\left(\frac{NAV_{f,t\text{-}1} - NAV_{f,t\text{-}4}}{NAV_{f,t\text{-}4}}\right) \tag{3}$$

[5] Most researchers follow Sirri and Tufano (1998) and compute flow as the percentage growth of the fund in excess of the growth that would have occurred had no new funds flowed in and had all dividends been reinvested. To compute the growth had no new funds flowed in, the literature has typically used the fund return over the previous year, assuming that the flow occurs over the end of the period. In our case, since our data is at the investor-month level, we use the actual capital gain or loss incurred by the investor over the month and assume that the investor flow occurs at the end of the month.

$$Rank_{f,t\text{-}1,6\text{-month}} = Percentage\ Rank\left(\frac{NAV_{f,t\text{-}1} - NAV_{f,t\text{-}7}}{NAV_{f,t\text{-}7}}\right) \tag{4}$$

$$Rank_{f,t\text{-}1,12\text{-month}} = Percentage\ Rank\left(\frac{NAV_{f,t\text{-}1} - NAV_{f,t\text{-}13}}{NAV_{f,t\text{-}13}}\right) \tag{5}$$

where f indexes funds, and t indexes time. *Percentage Rank* is a function that ranks each fund's performance into percentiles, with 0 being the worst-performing fund and 1 being the best performing fund.

We then follow Sirri and Tufano (1998) and create three variables based on each fund's performance percentile: the bottom performance quintile is defined as $Min(RANK_{f,t\text{-}1,1\text{-month}}, 0.2)$, while the combined middle three performance quintiles are defined as $Min(RANK_{f,t\text{-}1,1\text{-}month} - BOTPERF_{f,t\text{-}1,1\text{-}month}, 0.6)$, and the top performance quintile is defined as $RANK_{f,t\text{-}1,1\text{-}month} - BOTPERF_{f,t\text{-}1,1\text{-}month} - MIDPERF_{f,t\text{-}1,1\text{-}month}$. Funds that fall into the top performance quintile appear first when users start to scroll down the fund list in the platform, and funds that fall into the bottom performance quintile appear last when users reach the end of the list in the platform. Funds that appear in the middle three quintiles are mediocre funds that appear in an intermediate position. We also carry out robustness checks in several alternative piecewise regressions, progressively shrinking the top and bottom sections to decile, vingtile, and percentile rankings. All these fund rankings are dynamic in the sense that they are regenerated every month.

3.3 Descriptive statistics

Table 1 Panel A.1 reports summary statistics for fund variables. There are 3,515 manager-fund-month observations, with 591 manager-fund-month observations (16.8%) being funds managed by female managers. The proportion is higher than the sample in Niessen-Ruenzi and Ruenzi (2018) (13.8%). Female-managed funds have significantly higher monthly relative performance ranks than male-managed funds over all return horizons (1-, 3-, 6-, and 12-months), implying that female-managed funds are displayed higher in the app rankings on average than male-managed funds. Standard deviations of funds' returns are approximately similar between male- and female-managed funds across all these return horizon periods. Consistent with Niessen-Ruenzi and Ruenzi (2018), fund flow is significantly higher for male- than for female-managed funds. The standard deviation of fund flow is, however, lower for female than male-managed funds.

Female-managed funds are also smaller (Total Net Asset values) than their male counterparts though the funds have similar ages. Relative to U.S. fund managers, both male and female managers have relatively low tenures in our Chinese sample on average, as the mutual fund industry in China is relatively young. Male managers, however, have longer tenures than female managers on average.

Fund sales, fund management, and fund transaction fees are obtained from CSMAR's annual fund balance sheet data. Female-managed funds have significantly lower management fees and transaction fees than do male-managed funds. This suggests that female managers have lower salary expenses but perform at least as well as male managers in terms of overall fund NAV returns and slightly better than male managers in terms of fund monthly relative performance. Female-managed funds are associated with higher sales expense fees than their male counterparts, suggesting that female managers require higher promotion efforts to increase their visibility to investors. The relative lack of visibility is also seen in the media mention frequency ranks. Female managers are mentioned less frequently in the media than male managers, and the difference is statistically significant. When mentioned, both the number of positive and negative mentions are lower than for male managers.

Table 1 Panel A.2 shows the number of funds with different fund objectives. For funds managed by male managers, the top fund objectives are Appreciation Funds (1484) and Income Funds (605). Similarly, for funds managed by female managers, the top fund objectives are Appreciation Funds (220) and Growth Funds (195). Panel A.3 reports summary statistics for our manager variables. Both male and female managers are highly educated on average, with most male managers and all female managers reporting at least a master's degree or above.

Table 1 Panel B reports summary statistics for our user variables. There are 2,345,875 user-fund-month observations, with 255,718 observations consisting of funds held by female users. The average overall fund holding is 5,893 CNY (approximately US$910), showing that most investors on the platform are micro-investors with relatively small investments into stock funds. Male investors have significantly higher average fund holdings than female investors do, and male investors also have higher average fund inflows on average. Although both male and female investors exhibit negative average monthly capital gains, female investors appear to perform slightly better than male investors, though the difference is not statistically significant. Male investors spend significantly more than female investors on the platform per month on average. Male investors also appear to be significantly older than female investors on average. Female investors tend to reside in smaller cities than male investors on average. Finally, female investors report higher levels of risk tolerance (or equivalently, lower levels of risk aversion) than male investors.

4. Results

4.1 Do investors react to prior fund performance?

We first examine if the flow-performance sensitivity for individual investors in China is similar to patterns documented in the prior literature for funds elsewhere in the world. We employ the following panel OLS regression model:

$$Flow_{i,f,t}=\alpha+\beta_1\ Rank_{f,t\text{-}1}+\beta_2 Manager\ Gender_{i,f,t}+\mu M_{f,t}+\psi F_{f,t}+\lambda U_{i,t}+\theta O_f+\gamma_i+\delta_t+\epsilon_{i,f,t} \quad (6)$$

where i indexes investors, f indexes funds, and t indexes time. γ and δ denote individual and time fixed effects. ϵ is the error term. We follow Niessen-Ruenzi and Ruenzi (2018) and cluster our standard errors at the entity (user) level.

The dependent variable, *Flow*, is the percentage change in fund amount for a particular user's investment in a specific fund. *Rank* is the relative performance rank of funds in the previous month, with 0 being the worst-performing fund and 1 being the best performing fund. *Manager Gender* is a dummy variable that equals 1 if the fund manager is female and 0 if the fund manager is male. *M* is a vector of fund manager control variables, including a fund manager's education background and tenure duration. *F* is a vector of fund control variables that have been found to affect fund flows in the previous literature (Sirri and Tufano, 1998, Barber, Odean, and Zheng, 2005), including fund management fees, fund transaction fees and fund sales fees, which are normalized by fund total assets, respectively. We control for the fund size, measured by the logarithm of Total Net Assets in the previous quarter. We also control for fund age (the logarithm of the number of months since the fund's inception), fund risk (the standard deviation of fund's daily returns over the past month), and the aggregate level fund flow in the previous quarter. *U* is a vector of user control variables, including a user's average spending and standard deviation of spending for the previous 6 months, which are proxies for a user's income level and stability, respectively. *O* is a vector of fund objective dummy variables.

Table 2 Panel A presents the results for our baseline models. The models use various relative fund performance rankings within the same fund objective as the main regressor. As documented in the prior literature, performance matters. Model (1) does not include managerial gender. Consistent with the prior literature, there is a significant and positive association between a fund's relative performance ranking in the previous month and the current month fund in-flow. An increase in relative performance ranking of a fund by 1% over the previous month is associated with an increase in current month fund flow of 0.36% at the individual investor level.

Model (2) shows that the positive relationship between a fund's previous month return ranking and fund in-flow persists with similar significance and magnitude, across all the return horizons, even after adding manager gender as an explanatory variable. However, in contrast to Niessen-Ruenzi and Ruenzi (2018), we do not find significant gender biases at the monthly individual investor level, as the manager gender term is insignificant across all the ranking measures.

Turning to fund characteristics, we divide fund sales, fund management, and fund transaction fees by fund total assets to obtain the fund sales expense ratio, fund management expense ratio, and fund transaction fee ratio, respectively. The sales expense ratio is positively and significantly correlated with fund flows in all our model specifications in Table 2. This is consistent with the results on search cost and fund flows in Sirri and Tufano (1998). While the platform does not adjust rankings based on as the fund payments to the platform, investors can locate funds directly by searching for them. If the sales fee is a proxy for fund expenditure on advertising, it is plausible that investors become aware of the fund and search for the fund name directly, bypassing the list of ranked funds on the platform.

A higher management fee expense ratio is associated with significantly lower fund-inflows. The negative correlation between fund management fee expense ratio and fund flow is consistent with Christoffersen (2001), who finds that fund managers voluntarily waive their management fees in order to improve the net performance of their funds, which is strongly and positively correlated with fund-inflows. Managerial tenure at the current fund is also significantly positively related to fund inflows. Our findings are consistent with Christoffersen and Sarkissian (2009), who show that funds managed by more experienced managers deliver high returns, and hence have higher fund inflows. Finally, the negative sign on the relationship between transaction fee expense ratio and fund flow is consistent with the literature on transaction costs and fund underperformance (Rakowski, 2010, Grinblatt and Titman, 1989, Chalmers, Edelen, and Kadlec, 2001, Edelen, 1999, and Wermers, 2000) though in our sample, the effect is statistically insignificant.

In Table 2 Panel B, we follow Niessen-Ruenzi and Ruenzi (2018) and examine if there is a gender bias at the aggregate fund level. As before, past performance is significantly positively related to fund flows. Although the coefficient on manager gender is negative, suggesting that female managers receive lower aggregate fund flows on average, the coefficient is statistically insignificant. Therefore, inconsistent with Niessen-Ruenzi and Ruenzi (2018), in China at least, there does not appear to be a gender bias at the quarterly aggregate fund level. One explanation is that the gender bias does not show up at shorter horizons than the annual aggregate levels studied by Niessen-Ruenzi and Ruenzi (2018).

Overall, however, our results suggest that Chinese investors display broadly similar behavior to investors as documented elsewhere in the world.

4.2 Is the flow-performance relationship affected by gender-based attention bias?

In this section, we examine whether the flow-performance sensitivity differs by gender. We employ the following panel-OLS regression model:

$$Flow_{i,f,t}=\alpha+\beta_0\,Rank_{f,t-1}+\beta_1\,Rank_{f,t-1}\times Manager\ Gender_{i,f,t}+\beta_2 Manager\ Gender_{i,f,t}+\mu M_{f,t} +\psi F_{f,t}+\lambda U_{i,t}+\theta O_f+\gamma_i+\delta_t+\epsilon_{i,f,t} \quad (7)$$

where i indexes investors, f indexes funds, and t indexes time. γ and δ denote individual and time fixed effects. ϵ is the error term. As before, we follow Niessen-Ruenzi and Ruenzi (2018) and cluster our standard errors at the entity (user) level. Our variable of interest, *Rank × Manager Gender* is the interaction term between fund relative performance ranking and manager gender. M, F, U, and O are the vectors of fund manager control variables, fund control variables, user control variables, and fund objective dummy variables, respectively, described in the previous section.

Table 3 presents the results. In model (1), we interact the manager gender dummy variable with the relative performance ranking of funds in the previous month, controlling for user, month, manager education, and fund objective fixed effects. Interestingly, the coefficient on managerial gender is significantly positive, suggesting that, without conditioning on performance, female managers enjoy significantly larger fund inflows than male managers. As before, the current month's fund flow is significantly positively correlated with the prior month's relative performance of the fund. The coefficient on the fund relative performance is 0.39 and statistically significant at the 1% level, suggesting that the greater the prior relative performance, the higher the current month's fund inflow. The sensitivity of fund flows to short horizon performance has typically been documented at annual horizons.[6] This paper is the first to document evidence of performance-flow sensitivity at the monthly level on an individual investor basis.

However, the effect is significantly smaller for a female fund manager. The interaction term between manager gender and relative performance of funds has a statistically significant

[6] One of the few exceptions is Ferreira, et al. (2012) who examine the flow-performance relationship using aggregate quarterly data.

(at the 1% level) coefficient of -0.33. This suggests that, for a male manager, if the relative performance ranking of the fund in the previous month increases by 1%, the fund flow increases subsequently by 0.39%. However, for a female manager, the same relative performance ranking increase increases monthly fund inflows by only 0.06% (=0.39% – 0.33%).

In economic terms, the monthly difference in the monthly performance-flow relationship between male and female managers for the average user is 19.45 CNY (approximately US$3.01), or US$36.12 in annual terms when the performance ranking of the fund increases by 1% each month for 12 months. While this effect may seem small at the investor level, we note that the number of subscribers to each fund is substantial. Table 1 Panel A.1 shows that the average TNA for a male-managed fund is 1.02 billion CNY (approximately US$158.14 million), while the average TNA for a female-managed fund is 0.78 billion CNY (approximately US$254.8 million). Hence, the attention bias at the individual investor level translates to a difference in the monthly fund-level performance-flow relationship between male and female managers of approximately 2.57 million CNY (approximately US$0.40 million) for an average female-managed fund per month, or US$ 4.8 million in annual terms when the performance ranking of the fund increases by 1% each month over a 12-month period.

Table 3 Models 2-4 show that the attention bias continues to exist for all the other rankings available on the platform including the 3-, 6- and 12-month return horizons. In each case, the interaction term between manager gender (female) and 3-, 6-, and 12-month return ranking is negative and significant. The magnitude of the (negative) coefficient on the interaction term is considerably larger than the (positive) coefficient on managerial gender across all the models, turning the overall effect of gender on flow negative.[7]

Our results therefore suggest the presence of a significant attention bias in the flow–performance sensitivity away from female funds. If a male-managed fund performs well over the previous month, the fund inflows are higher than at a female-managed fund. However, if a female-managed fund performs poorly in the previous month, the monthly fund outflow will also be lower than at a male-managed fund. Simply put, investors appear to pay less attention to female-managed funds.

If investors are rational and their goals are to maximize their returns on investments, there should not be any systematic difference in the flow-performance relationship based on gender. One explanation is that people have double standards toward females (Botelho and Abraham,

[7] In untabulated regressions, we subtract automatic fund investments from the numerator of our fund flow measure, and our results remain unchanged.

2017, and Botelho and Gertsberg, 2020). To receive the same attention as male managers, female managers must perform better than mediocre male managers. Another explanation is the deeply rooted cultural norm whereby girls get less attention in their families than boys when growing up, which leads to less attention being paid to females in general (Barcellos, Carvalho, and Lleras-Muney, 2014, and Park and Rukumnuaykit, 2004). It is also possible that people project stereotypes onto female managers which bias them against trusting their funds to female managers if the performance is short of excellent (Neumark and Bank, 1996, Newton and Simutin, 2015, Hannum, Kong, and Zhang, 2009, Banaji and Greenwald, 1995).

An alternative, rational, explanation to the muted flow-performance sensitivity for female managers is that male managers' past performance is a better predictor of their future performance. We therefore examine whether past performance predicts future performance for Chinese fund managers and whether there is a gender difference in the predictive ability for future performance. This is related to the hot hands effect, first documented by Hendricks, Patel, and Zeckhauser (1993), who find that the relative performance of no-load and growth-oriented mutual funds persists in the near term. Carhart (1997) argues that persistence in mutual fund returns is mostly driven by the one-year momentum effect of Jegadeesh and Titman (1993) and finds no evidence for the hot hands effect. Nevertheless, if investors believe that male fund managers are more likely to have hot hands, then an increase in short-term performance might be consistent with rational investor behavior in directing flows to these funds. In particular, if male fund managers are more likely to have hot hands than female managers, the differential gender sensitivity of flow to performance might be an outcome of rational choices by investors.

We regress the current month fund return ranking on the future 3-, 6-, and 12-month return rankings, respectively. Table 3 Panel B shows the coefficients of the regression estimates. Current month fund return ranking appears to be negatively and significantly correlated with the future 3-, 6- and 12-month return rankings, inconsistent with the hypothesis that Chinese mutual fund managers have hot hands. In addition, since the coefficient of the manager gender dummy variable is insignificant, there is no evidence that male managers are more likely to have hot hands than female managers.

We then examine whether the gender attention effect is also subject to the tendency of fund investors, documented using US data (see for example, Sirri and Tufano, 1998, or Del Guercio and Tkac, 2002) to buy past winners more intensely than they sell past losers. Huang et al. (2007) show that the magnitude of this relationship has declined over time for US mutual funds. Ferreira et al. (2012) find marked differences in in the flow-performance relationship across countries, suggesting that US findings do not apply directly to other countries. In particular, for less developed countries, they find little evidence of convexity at the individual country level. China is not included in their sample.

We test if attention bias affects the tendency of Chinese investors to preferentially buy winning funds while avoiding sell losing funds based on the gender of the fund manager. For example, investors might react asymmetrically to winners and losers on the basis of performance, directing flows preferentially towards high-performing male fund manager while being faster to direct flows away from poorly-performing female fund managers.

Specifically, we divide our sample into two subsamples where the first includes only funds whose returns have increased in the previous month compared to two months ago (winners), and the second includes funds whose returns have decreased in the previous month relative to two months ago (losers). We then create a dummy variable that indicates if the fund is a winner or a loser and formally test whether there is a gender-based differential performance chasing effect between male- and female-managed funds. Table 3 Panel C reports the coefficients for these regressions.

Columns 1-2 show that, while both subsamples exhibit significant attention bias, the magnitude of the attention bias coefficient is around five times the size for the winners than the losers, suggesting that funds pay significantly less attention to women when the funds are performing particularly well and more attention to them when the funds are performing poorly. Column (3) formally tests the difference in magnitude of the attention bias between the two subsamples, using a triple interaction term (Return increase dummy × Manager gender × Fund performance ranking). The triple interaction term is significantly negative, suggesting that the well-documented asymmetric flow-performance relationship is also affected by attention bias.

Finally, we follow Sirri and Tufano (1998) in using piecewise regression approaches in Table 3 Panel D. The first specification uses the Sirri-Tufano specification, cutting the funds on the top quintile, the mid-three quintiles and the bottom quintile of performance. The gender attention bias appears to be concentrated in the mid- and bottom quintiles of performance. In the top quintile, there is no evidence of gender bias in the interaction term. However, it is possible that the top quintile is too coarse a specification if the very top female managers do not experience a gender bias. Hence, in the subsequent specifications, we progressively shrink the size of the top section from quintiles to deciles to percentiles. The attention bias continues to exist in the mid and bottom sections across all our cuts, suggesting that progressively refining the definition of top performing fund managers does not change the results. A gender bias appears in the top 1% and 5% sections as well, though the extremely small numbers of female managers in these sections implies that these findings are noisy.

4.3 Alternative Performance Metrics

So far, our performance metrics are based on the objective-adjusted rankings reported by the platform. However, it could be argued that investors do not use these platform-provided rankings but adjust for risks in other ways. Evans and Sun (2020) show that U.S. retail investors use simple risk-adjustment heuristics provided by Morningstar to direct fund flows. Specifically, using Morningstar's 2002 rating methodology change, they show that before the change, flows are strongly correlated with CAPM alphas. After the change, when funds are ranked by size and book-to-market groups, flows become more sensitive to 3-factor alphas (FF3).

To check the robustness of our results, we use two simple alternative heuristics that investors might consider for our performance-flow panel regressions. We compute the Jensen's Alpha (Jensen, 1968), calculated from funds' daily returns over the past month. Jensen's Alpha assumes that the CAPM (Sharpe, 1964) model is correct and calculates the risk-adjusted return for investors. We obtain Jensen's Alpha for each fund from CSMAR which computes it from the following regression equation with funds' daily returns during the past month:

$$\alpha_{i,t}=R_{i,t}-R_{f,t}+\beta_{i,t}\times\left(R_{m,t}-R_{f,t}\right) \tag{8}$$

where i indexes for funds, t indexes for days, R_i is the daily return of fund i, R_f is the risk-free rate and R_m is the market return. We then rank the alphas into percentiles by month.

The second performance metric that we use is the arithmetic average of daily returns for the fund over the past month, which is calculated as follows:

$$\overline{R}_i=\frac{1}{T}\sum_{t=1}^{T}R_{i,t} \tag{9}$$

where i indexes for funds, t indexes for days, T is the number of days in the past month, and R is the daily fund return. We then rank the daily average returns into percentiles by month. We then use these performance metrics in the same panel-OLS regression model as in equation 7. Both these heuristics are relatively simple to calculate and it is plausible that an investor might use them as alternatives to the rankings provided by the platform.

Table 4 presents the results. Model (1) and (2) use the Jensen's Alpha rank for the past month as the performance metric, while Model (3) and (4) use the daily fund performance rank for the past month as the performance metric. Our results across all four models are consistent with our previous results in Table 2 and Table 3.

Models (1) and (2) show a positive and significant correlation between prior month's risk-adjusted return and current month's fund flow. In addition, the interaction term between risk-adjusted return and fund manager gender is negative and statistically significant. If the relative Jensen's Alpha ranking of a male-managed fund increases by 1% in the previous month, the current month fund flow increases by 0.37%. In economic terms, for an average user with a fund amount of 5,893 CNY (approximately US$910), the fund flow increases by 21.80 CNY (approximately US$3.37) per month. In contrast, the fund flow to female-managed funds increases only by 0.06% (0.40% – 0.34%). This is equivalent to a difference of 2.65 million CNY (approximately US$ 410,141) in the performance-flow relationship between male- and female-managed funds at the monthly fund level.. A similar pattern exists when we use the average daily returns for funds in the past month in models (3) and (4).

4.4 Do female managers perform systematically worse than male managers?

One explanation for our results might be that female managers perform systematically worse than male managers. In this section, we examine whether there is a difference in performance between male and female managers using a multiple regression approach on all stock funds covered by CSMAR. We employ the following regression model:

$$\begin{aligned} Performance_{f,t} = \alpha + \beta_1 Manager\ Gender_{f,t} + \mu M_{f,t} \\ + \psi F_{f,t} + \theta O_f + \gamma_i + \delta_t + \epsilon_{f,t} \end{aligned} \tag{10}$$

where *f* indexes funds, and *t* indexes time. γ and δ denote individual and time fixed effects. ϵ is the error term. We cluster our standard errors at the fund level. The dependent variable, *Performance* is drawn from a battery of fund performance, idiosyncratic risk, systematic risk, stock selection ability and market timing ability measures.

Table 5 reports the results. Models 1-4 report regression coefficients for regressions where the dependent variables are the 1-, 3-, 6-, and 12-month return rankings, respectively. Model (5) reports coefficients for a regression where the dependent variable is the daily return standard deviation, to proxy for the idiosyncratic risk that the fund bears. Model (6) reports coefficients for a regression where the dependent variable is the Sharpe (1965) ratio, which measures the excess return per unit of idiosyncratic risk that the fund bears. Model 7-8 report coefficients where dependent variables are betas and alphas of the Sharpe (1964) CAPM model. The CAPM beta proxies for the systematic risk that the fund bears against the market, while the CAPM alpha proxies for the ability of the fund to beat the market portfolio. Model (9) reports the coefficients where the dependent variable is the Treynor (1965) Index, which measures the excess return per unit of systematic risk that the fund bears. Model 10-11 use the Treynor and Mazuy (1966) model's measure of market timing and stock selection abilities as dependent variables, respectively. Model 12-15 use the Chang and Lewellen (1984) model's measure of bear- and bull-market timing as well as stock selection abilities as dependent variables.

Across all fund performance specifications (except for the 1-month return ranking measure), manager gender is not significantly related to fund return rankings after controlling for manager tenure, education, the previous quarter fund size and aggregate-level flow, fund sales fee, fund transaction fee, fund management fee and fund objectives. When measuring fund performance in terms of 1-month fund performance, female managers perform 4.24 percentage points better than male managers, which is statistically significant at the 1% level. We do not detect significant differences in idiosyncratic or systematic risks or excess returns per unit of systematic or idiosyncratic risk between funds managed by male and female managers. Furthermore, male-managed funds do not exhibit superior stock selection or market timing abilities (other than the TM model). Overall, gender appears to be unrelated to the fund's performance, risk, market timing ability or stock selection ability. It is therefore difficult to explain the differential performance chasing behaviour between male- and female-managed funds using a rational asset pricing framework. It appears more likely that this is due to investor preferences for male-managed funds.

4.5 Is attention bias driven by the lack of female fund managers in the sample?

Table 1 shows that there are more male managers than female managers in our sample. Therefore, a natural question to ask is if the attention bias exists because of the sheer number of male managers, which captures most of the investors' attention. In other words, do investors exhibit attention bias away from female managers simply because there are fewer female managers and they are difficult to find? We note that univariate two-sample t-tests in Table 1 show that female managers perform significantly better in terms of 1-, 3-, 6-, and 12-month returns, suggesting that users see female-managed funds higher in the app rankings, on average, than male-managed funds, implying that female-managed funds are more visible than male-managed funds.

Nevertheless, to explicitly eliminate the effect of the difference in the number of male and female managers in our sample, we use two approaches. In this section, we report results from a propensity score matching (PSM) approach where we show that attention bias still exists even after matching the male and female managers on a host of managerial and fund characteristics.

While the PSM approach does not address causality, it balances manager and fund covariates in our sample and mitigates concerns that higher search costs for female managers affect the attention bias of investors.

We use a logistic regression to calculate the propensity score of a fund choosing a female manager and control for variables that affects a fund's visibility to the investors. Panel A of Table 6 shows the results of the logistic regression where we regress a fund's choice of manager gender on past 1-, 3-, 6-, and 12-month returns ranked within each fund objective, fund fees, fund size, fund age, fund return standard deviation, aggregate level fund flow, manager tenure, and manager education (whether the manager has at least an undergraduate degree).[8] We then estimate propensity scores at the fund level and match each female manager to a male manager by month using a PSM with nearest neighbour matching. Panel B of Table 6 shows the difference-in-means of the independent variables for male managers versus female managers for both the unmatched and matched samples, respectively. T–statistics for the difference–in–means test indicate that all variables differ significantly for the unmatched sample. In contrast, the corresponding difference-in-means tests indicate that all variables do not differ significantly for the matched sample, and there is a good covariate balance across the matched variables.

Using the matched fund sample, we merge individual investor data to the matched funds and re-run the regression with the same control variables and fixed effects as equation (7). The results are reported in Table 6 Panel C. Model (1) shows the regression coefficients when we use the prior 1-month fund returns as the performance measure. The current month's fund flow continues to be significantly positively correlated with the prior month's relative fund performance. The relative performance coefficient is 0.37 and statistically significant at the 1% level, suggesting that higher relative performance increases the current month's fund flow. However, as before, the effect is significantly smaller for a female fund manager. The interaction term between manager gender and relative performance of funds has a coefficient of -0.20, which is also statistically significant at the 10% level. Hence, investors still appear to exhibit an attention bias even in a sample matched on performance and visibility. Models 2-4 report regression coefficients when we use the 3-, 6-, and 12-month fund returns as the performance measure, respectively. In all models, the interaction term is negative and significant, with model 2-3 significant at the 1% level and model 4 significant at the 10% level.

[8] Table 1 Panel C.1. shows that all the female managers have at least an undergraduate degree while some male managers stop at the undergraduate degree level.

Overall, our PSM results suggest that attention bias does not appear to be driven by higher search costs of finding female managers.

4.6 Do investors pay less attention to women fund managers because of a lack of media attention?

In the second approach, we examine if attention bias is related to difference in media coverage or sentiment between male and female fund managers. Da, Engelberg, and Gao (2011) propose a new measure of retail investor attention using search frequency in Google and find that investor attention predicts higher stock prices in the next 2 weeks and an eventual price reversal within the year. Ben-Rephael, Da and Israelsen (2017) measure institutional investor attention using news searching and news reading activity on Bloomberg terminals and find that institutional attention responds more quickly to major news events, leads retail attention, and facilitates permanent price adjustment. It is possible, therefore, that more frequent coverage or more positive coverage of male managers than female managers in the news may lead to investors paying more attention to male managers than to female managers.

To control for media coverage on fund managers, we collect approximately 1.2 million Chinese news articles from CSMAR and filter out approximately 400,000 news articles within the financial news category. We do not conduct a plain search of managers' names in the news articles, as some Chinese names are also common phrases (such as the word "trillion", which is both a typical Chinese name and an expression of numerical count in Chinese) and may lead to large levels of noise in the frequency count of media coverage on managers' names. Instead, we use spaCy, a state-of-the art natural-language-processing model based on convolutional neural networks, to extract people's names from the news articles through part-of-speech tagging and named-entity-recognition. We then count the frequency of each manager's name in news articles each month and rank the frequency by fund objectives into percentiles. The resulting variable, media mention frequency rank, proxies for media's coverage on fund managers in each month. Table 1 Panel A shows that female managers are mentioned less frequently in the media than male managers, and the difference is statistically significant.

To further distinguish between positive mentions and negative mentions of each manager in the news, we apply sentiment analysis to each sentence that includes a manager's name. We employ the SKEP (Sentiment Knowledge Enhanced Pre-training for Sentiment Analysis) model proposed by Tian et al. (2020) for sentiment analysis. SKEP is a variant of the Transformer model (Vaswani et al. 2017) that incorporates sentiment knowledge by self-supervised learning. Specifically, we apply the state-of-the-art SKEP model pre-trained based on ERNIE (Enhanced Language Representation with Informative Entities) proposed by Zhang et al. (2019).[9] Each sentence is classified as either positive sentiment or negative sentiment by the SKEP model. We then count the frequencies of sentences of positive sentiments and negative sentiments, respectively, for each manager and each month. Next, we normalize the frequencies by the total number of sentences that mention each manager in each month. Finally, we rank the normalized frequencies by month and fund objective and obtain our measure of positive and negative mentions of managers in the news. Table 1 Panel A shows that there are more positive mentions than negative mentions in our sample. In addition, whether in terms of positive or negative mentions, male managers are mentioned more frequently than female managers.

We then add media mention frequency, positive media mentions and negative mentions as control variables in Table 7. Table 7 Column (1) includes the overall media mention frequency as the control variable and shows that the gender bias remains negative and significant after controlling for media coverage. Table 7 Column (2) disentangles the overall media mention frequency into positive and negative media mention frequencies based on sentiment analysis and show that, while gender bias remains significant, negative (positive) mentions in the media reduces (increases) fund flows. Table 7 Column (3) interacts positive and negative media mention frequencies with fund performance and shows that positive mentions in the media strengthen the flow-performance sensitivity, though the gender bias still stays significant. Table 7 Column (4) triple interacts positive and negative media mention frequencies with fund performance and fund manager gender and shows that the gender bias exists only for the female fund managers who are mentioned positively in the news. In other words, relative to another male fund manager with identical performance and positive mentions in the news, the female manager earns lower fund flows. The results from Columns 1-3 suggest that the gender bias is unlikely to be the result of different levels of media coverage or sentiments between male and female managers. Although the evidence in Column (4) is relatively weak (there are very few negative mentions of female managers in our sample), it suggests that the bias is stronger when a female manager is given positive media coverage in the news, as investors still pay less attention to them than to an equivalent male manager with identical performance and positive mentions in the news.

[9] We run the SKEP model on 2,409,963 sentences with 8 Nvidia-Ampere GPUs, a process that takes approximately six hours to finish.

4.7 What type of users are prone to attention bias?

We next examine cross-sectional evidence on attention bias based on four user characteristics: user gender, age, city of residency and risk tolerance. We pick these four characteristics based on evidence in the literature that shows different levels of gender biases across these characteristics. For example, considering gender, experimental evidence (see for example, Lovén, Herlitz, and Rehnman, 2011) shows that women remember more female than male faces, whereas men do not seem to display an own-gender bias in face recognition memory. Similarly, for age, Das Gupta and Shuzhuo (1999), among others, argue that the wars and famine experienced in China over the last century led to the prioritization of female children over male children in terms of nutrition and education. We conjecture therefore, that older investors are more likely to be affected by attention bias. For city of residency, the literature on gender biases argues that in smaller cities or rural areas in developing countries, female children are given less resources and paid less attention in families (e.g., Barcellos, Carvalho, and Lleras-Muney, 2014, Hannum, Kong, and Zhang, 2009, and Park and Rukumnuaykit, 2004). We conjecture that attention bias is likely to be higher in smaller cities. We define big cities as tier-1 cities in China, which include Beijing, Shanghai, Guangzhou and Shenzhen. The rest of the cities with higher tiers fall into our definition of small cities. Finally, the platform assesses users' risk tolerance based on a questionnaire, and each user is classified into a risk band, with values ranging from 0 to 6. The smaller the value, the less risk-tolerant the user. We define users with a risk band value below or equal to 2 as risk-averse, and users with a risk band value greater than 2 as risk-tolerant.

We interact each of the four user characteristics with manager gender and fund performance. Table 8 presents the results for each of the four triple-interaction regressions where fund performance is measured as fund returns over the past 1-month horizon. Gender bias remains significant in all four regression specifications.

Table 8 Column (1) suggests that female users tend to direct lower flows to funds as well as displaying a weaker flow-performance relationship than to male users. However, the triple interaction term of user gender, manager gender, and fund performance is positive and statistically significant, suggesting that female users have a lesser degree of gender bias than male users. Column (2) suggests that older users tend to have higher fund flows and a stronger flow-performance relationship than younger users, but the degree of gender bias is insignificant among users of different ages. Column (3) suggests that users from smaller cities tend to have higher fund flows, a stronger flow-performance relationship, and a stronger gender bias than users from larger cities. Finally, Column (4) does not show significant results relating the degree of user risk aversion to flow-performance sensitivity.

4.8 Is attention bias innate to investors?

So far, our models have focused on investors who have had non-zero fund holdings in the previous month. We next investigate if attention bias develops over time or if it is innate to the investor by examining investors who are investing for the *first* time on the platform. Specifically, we run a logistic regression with the following specification:

$$\begin{aligned} \mathit{Manager\ Gender}_{i,f,t} = \mathit{Logit}(\alpha + \beta_1\ \mathit{Rank}_{f,t\text{-}1} + \mu M_{f,t} \\ + \psi F_{f,t} + \lambda U_{i,t} + \theta O_f + \epsilon_{i,f,t}) \end{aligned} \quad (11)$$

where i indexes investors, f indexes funds, and t indexes time. ϵ is the error term.

Table 9 presents the results. Models 1-2 report regression coefficients including only user characteristics (Model 1), and user, fund, and manager characteristics (Model 2). In model (2), managerial tenure is negatively related to the likelihood of investing in a female-managed fund. Across both Models (1) and (2) however, controlling for managerial, fund, and user characteristics, performance is negatively related to the propensity to invest in female-managed funds.

4.9 Is there a causal relationship between gender bias and fund flows?

Our main variable of interest, manager gender, could be subject to endogenous and unobserved factors in the error term. Therefore, we next use two instrumental variables to establish a causal relationship between manager gender and fund flows. These instruments do not instrument for fund manager per se, but for the specific investor's choice of a fund manager. We first discuss the economic arguments supporting the validity of the two different instrumental variables.

The first instrumental variable is the proportion of illiterate women in the entire female population at the municipal district level in 2010 based on Census data released by the National Bureau of Statistics of China. Since the Song Dynasty (960), the Chinese imperial examinations, or *Keju*, have been used as a civil service examination system for selecting candidates for the state bureaucracy. However, the examination system did not allow female candidates to participate, and for almost a thousand years, a typical social norm was that females should not receive formal education and should remain illiterate. The situation has been greatly improved since the establishment of the PRC, who introduced a compulsory education law which mandates free education for both male children and female children below the 9th grade since 1986. Before the introduction of the compulsory education law, families could choose to send their children to school for basic education by paying tuition fees. Families who did not possess enough resources to send all their children to school might prioritize boys over girls, leaving a proportion of illiterate women.

The motivation for this instrument is that people from districts with a higher proportion of illiterate women might exhibit a stronger gender bias towards men, as their parents, grandparents or friends might have directed educational resources towards boys over girls before 1986. Since the literature shows that culture (values, knowledge and practices) that are prevalent in one generation are transferred to the next generation, it is likely that some gender biases are transferred intergenerationally as well. However, it is unlikely that the proportion of illiterate women in a local district will influence investors' decisions on fund flows through channels other than gender biases, after controlling for investor income and other characteristics, implying that the proportion of illiterate women is a suitable instrumental variable.

The second instrumental variable is the proportion of female new-borns amongst all new-borns at the municipal district level in 2010. In 2003, the regulations in China banned foetal gender identification for non-medical needs and any artificial termination of pregnancy for gender selection purposes. Though the act is a legal requirement to increase gender equality in all provinces, some illegal enterprises continued to conduct foetal gender identification and artificial termination of pregnancy in the first few years after the law was introduced. Those activities were commonly known as "the two-illegal activities" in China.[10] While gender identification and artificial termination have both been significantly reduced by law enforcement officials today, as of the 2010 census, some provinces continue to display a degree of gender imbalance relative to world averages. Therefore, it is plausible that the proportion of female new-borns at the local district level reflects the level of local gender bias in the area. However, the proportion of female new-borns is unlikely to be directly related to people's fund flow controlling for people's income, as there is no statistical evidence that there is a significant

[10] The phrase is used commonly in government issued (Chinese) news releases (see for example, http://www.gov.cn/xinwen/2015-05/07/content_2857935.htm). The government has repeatedly tried to crack down on these gender discrimination activities (see Hou, Liqiang and Shan, Juan, 2014, Joint forces to curb illegal abortions, *China Daily*, 4 September 2014).

price difference in medical costs when giving birth to a male or female child. Therefore, the proportion of female new-borns is an ideal candidate for an exogenous instrument.

Our endogenous variable, manager gender, is a binary variable. Although the traditional 2SLS estimator is still consistent for binary endogenous variables, it is not necessarily efficient. Therefore, we follow Adams, Almeida, and Ferreira (2009) in carrying out a three-stage procedure in identifying the causal effects. In the first stage, we estimate the following Logit model:

$$Manager\ Gender_{i,f,t} = Logit(\alpha + \beta_0 Instrumental\ Variable_i + \beta_1\ Rank_{f,t-1} + \mu M_{f,t} + \psi F_{f,t} + \lambda U_{i,t} + \theta O_f + \epsilon_{i,f,t}) \tag{12}$$

where i indexes investors, f indexes funds, and t indexes time. ϵ is the error term. *Instrumental Variable$_i$* is the proportion of illiterate women in the female population or the proportion of female new-borns among all newly-born children at the local district where the investor resides. *M*, *F*, *U*, and *O* are the vectors of fund manager control variables, fund control variables, user control variables, and fund objective dummy variables described in the previous section.

In the second stage of the procedure, we then compute the fitted probability of choosing female managers, $\widehat{Manager\ Gender}$, from the Logit regression above. In the third stage of the procedure, we use $\widehat{Manager\ Gender}$ to instrument for *Manager gender*, and $\widehat{Manager\ Gender} \times Rank$ to instrument for Manager gender × Rank, respectively, in the following equation using a standard 2SLS procedure:

$$Flow_{i,f,t} = \alpha + \beta_0\ Rank_{f,t-1} + \beta_1\ Rank_{f,t-1} \times Manager\ Gender_{i,f,t} + \beta_2 Manager\ Gender_{i,f,t} + \mu M_{f,t} + \psi F_{f,t} + \lambda U_{i,t} + \theta O_f + \gamma_i + \delta_t + \epsilon_{i,f,t} \tag{13}$$

where i indexes investors, f indexes funds, and t indexes time. ϵ is the error term. We cluster our standard errors at the user level.

As Wooldridge (2001) notes, the advantage of the above procedure is that it delivers consistent estimates in the third stage while allowing for the presence of non-linearities in the first stage. Furthermore, the consistency guarantee of the procedure does not require a correct specification of the functional form in the first stage regression, and although fitted values from the first stage are used in the third stage as inputs to the standard 2SLS procedure, the standard IV standard errors are still asymptotically valid.

Table 10 shows the results of our three-stage instrumental variable estimation procedures.[11] Table 10 Panel A shows the coefficient estimates of the first stage Logit model. Model (1) shows that the proportion of female new-borns at the district level is positively and significantly correlated with the probability of an investor choosing a female manager. This is consistent with our earlier hypothesis that the larger the proportion of female new-borns versus male new-borns, the lower the gender bias in the local district and the more likely that investors from that local district are going to invest in a female manager. Model (2) shows that the proportion of illiterate females amongst all females at the district level is negatively and significantly correlated with the probability of an investor choosing a female manager. Again, this is consistent with our earlier hypothesis that the larger the proportion of illiterate females at the district level, the higher the gender bias in the local district and the less likely that investors from that local district are going to invest in a female manager. Both of our instruments are highly statistically significant in both models, suggesting that they are strong predictors of the probability of choosing a female manager. We note that these results are also consistent with Niessen-Ruenzi and Ruenzi (2018) in that investors in areas that are more biased towards males are less likely to invest in female-managed funds.

Table 10 Panel B shows the coefficient estimates of the third stage 2SLS model. Consistent with our previous OLS estimates, both model (1) and (2) show a significant and positive correlation between past-month fund performance and current month fund flows. Most importantly, the coefficient of manager gender, our instrumented variable, is positive and significant, while the coefficient of the interaction term between manager gender and previous-month fund performance is negative and significant. The coefficients in our instrumental variable regressions therefore confirm our earlier hypothesis that gender bias towards female managers has a causal influence on investors' fund flow decisions.

4.10 Do female-managed funds have lower individual fund flow volatilities?

Table 1 Panel A shows that female-managed funds have lower aggregate-level monthly volatilities on the fintech platform. Our main results also show that investors respond less to female managers' performance than male managers' performance due to an attention induced gender bias. Do individuals have lower fund flow volatilities while holding female-managed funds?

[11] We reject the null hypothesis that the model is over-identified using several over-identification tests including the Anderson-Rubin Test, the Sargan test, the Basmann test, and the Wooldridge test. In addition, the first (or second in the three-stage IV process) stage have F-statistics of over 360, rejecting the null hypothesis that our instrument is weak.

In the final part of the paper, we compute individual fund flow volatilities by measuring the standard deviation of fund flows for each individual and each fund during our entire sample period. We use individual fund flow volatilities as the dependent variable, and managerial, fund and user characteristics as the explanatory variables. Table 11 reports the results. The manager gender term has a negative and significant (at the 10% level) coefficient, implying that, at the individual user level, fund volatilities are lower for female-managed funds than for male-managed funds. For mutual fund companies, this may have the desirable impact of lowering the volatility of flows into the fund.

5. Conclusions

The prior literature has argued that investors discriminate against women, with the consequence that there are relatively few female mutual fund managers in the industry. In this paper, we document the existence of a new and previously unstudied type of gender bias - an attention bias away from female fund managers. Prior literature shows that there is a positive correlation between prior mutual fund performance and the subsequent fund flow. We show that this flow-performance sensitivity is affected by a differential gender effect. Using a unique sample of individual investor flows into individual funds in China, we provide robust evidence that the investors are more sensitive to the performance of male managed-funds than for female-managed funds.

The bias exists across all the return horizons where the platform app allows sorting of fund returns, as well in simple heuristics for performance such as Jensen's alpha and daily average returns. There are also significant cross-sectional differences between investors. Female users appear to display lower levels of gender bias towards female-manged funds. Similarly, users living in smaller cities display stronger levels of gender bias away from female-managed funds. The level of gender bias appears to be innate to investors – an attention bias manifests even in the first set of investments made by a user on the platform. The attention bias uncovered in the sample appears to be irrational and cannot be explained by the difference in performance between male and female managers or the difference in media coverage between male and female managers.

Niessen-Ruenzi and Ruenzi (2018) argue that because mutual fund investors appear to direct significantly lower flows to female-managed mutual funds than to male-managed funds, rational fund companies might choose to hire fewer women since fund companies generate their profits from fees charged on assets under management. Our paper shows that the attention bias works both ways. Though investors appear more sensitive to fund performance when the fund manager is male, the sensitivity is bi-directional. Investors are also less flow-sensitive to underperforming female managers. For mutual fund companies, this may have the desirable impact of lowering the volatility of flows into the fund.

References

Adams, Renée, Heitor Almeida, and Daniel Ferreira, 2009, Understanding the relationship between founder–CEOs and firm performance, *Journal of Empirical Finance*, 16, 136-150.

Adams, Renée B., and Patricia Funk, 2012, Beyond the glass ceiling: Does gender matter?, *Management Science* 58, 219–235.

Adams, Renée B., and Tom Kirchmaier, 2016, Women on boards in finance and STEM Industries, *American Economic Review* 106, 277–281.

Atkinson, Stanley M., Samantha Boyce Baird, and Melissa B. Frye, 2003, Do female mutual fund managers manage differently?, *Journal of Financial Research* 26, 1–18.

Bailey, Warren, Alok Kumar, and David Ng, 2011, Behavioral biases of mutual fund investors, *Journal of Financial Economics* 102, 1–27.

Banaji, Mahzarin R., and Anthony G. Greenwald, 1995, Implicit gender stereotyping in judgments of fame, *Journal of Personality and Social Psychology* 68, 181–198.

Barber, Brad M., Terrance Odean, and Lu Zheng, 2005, Out of sight, out of mind: The effects of expenses on mutual fund flows, *The Journal of Business* 78, 2095–2120.

Barcellos, Silvia Helena, Leandro S. Carvalho, and Adriana Lleras-Muney, 2014, Child gender and parental investments in India: Are boys and girls treated differently?, *American Economic Journal: Applied Economics* 6, 157–189.

Ben-Rephael, Azi, Zhi Da, and Ryan D. Israelsen, 2017, It depends on where you search: Institutional investor attention and underreaction to news, *Review of Financial Studies* 30, 3009–3047.

Berk, Jonathan B., and Richard C. Green, 2004, Mutual fund flows and performance in rational markets, *Journal of political economy* 112, 1269–1295.

Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz, 2010, Dynamics of the gender gap for young professionals in the financial and corporate sectors, *American Economic Journal: Applied Economics* 2, 228–255.

Bigelow, Lyda, Leif Lundmark, Judi McLean Parks, and Robert Wuebker, 2014, Skirting the Issues: Experimental evidence of gender bias in IPO prospectus evaluations, *Journal of Management* 40, 1732–1759.

Bollen, Nicolas P. B., and Jeffrey A. Busse, 2005, Short-term persistence in mutual fund performance, *Review of Financial Studies* 18, 569-597.

Botelho, Tristan L., and Mabel Abraham, 2017, Pursuing quality: How search costs and uncertainty magnify gender-based double standards in a multistage evaluation process, *Administrative Science Quarterly* 62, 698–730.

Botelho, Tristan L., and Marina Gertsberg, 2020, Evaluator status and its effect on gender bias in evaluations, *Academy of Management Proceedings* 2020, 13989.

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

Chalmers, John M.R., Roger M. Edelen, and Gregory B. Kadlec, 2001, On the perils of financial intermediaries setting security prices: The mutual fund wild card option, *The Journal of Finance* 56, 2209–2236.

Chang, Eric C., and Wilbur G. Lewellen, 1984, Market timing and mutual fund investment performance, *The Journal of Business* 57, 57.

Chevalier, Judith, and Glenn Ellison, 1999, Are some mutual fund managers better than others? Cross-sectional patterns in behavior and performance, *Journal of Finance* 54, 875–899.

Christoffersen, Susan E. K., 2001, Why do money fund managers voluntarily waive their fees?, *The Journal of Finance* 56, 1117–1140.

Christoffersen, Susan E. K., and Sergei Sarkissian, 2009, City size and fund performance, *Journal of Financial Economics* 92, 252–275.

Cortina, Lilia M., Vicki J. Magley, Jill Hunter Williams, and Regina Day Langhout, 2001, Incivility in the workplace: Incidence and impact, *Journal of Occupational Health Psychology* 6, 64-80.

Cortina, Lilia M., Kimberly A. Lonsway, Vicki J. Magley, Leslie V. Freeman, Linda L. Collinsworth, Mary Hunter, and Louise F. Fitzgerald, 2002, What's gender got to do with it? Incivility in the federal courts, *Law & Social Inquiry* 27, 235-270.

Cortina, Lilia M., 2008, Unseen injustice: Incivility as modern discrimination in organizations, *Academy of Management Review* 33, 55-75.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2011, In search of attention, *Journal of Finance* 66, 1461–1499.

Das Gupta, Monica, and Li Shuzhuo, 1999, Gender bias in China, South Korea and India 1920-1990: Effects of war, famine and fertility decline, *Development and Change* 30, 619–652.

Del Guercio, Diane, and Paula A. Tkac, 2002, The determinants of the flow of funds of managed portfolios: Mutual funds versus pension funds, *Journal of Financial and Quantitative Analysis* 37, 523-557.

Edelen, Roger M,, 1999, Investor flows and the assessed performance of open-end mutual funds, *Journal of Financial Economics* 53, 439–466.

Evans, Richard B, and Yang Sun, 2020, Models or stars: The role of asset pricing models and heuristics in investor risk adjustment, *The Review of Financial Studies* 34, 67–107.

Ferreira, Miguel A., Aneel Keswani, Antonio F. Miguel, and Sofia B. Ramos, 2012, The flow-performance relationship around the world, *Journal of Banking & Finance* 36, 1759-1780.

Goldin Claudia and Cecilia Rouse, 2000, Orchestrating impartiality: The impact of “blind” auditions on female musicians, *American Economic Review* 90, 715–741.

Grinblatt, Mark, and Sheridan Titman, 1989, Mutual fund performance: An analysis of quarterly portfolio holdings, *Journal of business*, 393–416.

Hannum, Emily, Peggy Kong, and Yuping Zhang, 2009, Family sources of educational gender inequality in rural China: A critical assessment, *International Journal of Educational Development* 29, 474–486.

Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser, 1993, Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988, *Journal of Finance* 48, 93-130.

Hong, Claire Yurong, Xiaomeng Lu, and Jun Pan, 2020, FinTech platforms and mutual fund distribution. NBER working paper series no. w26576 (National Bureau of Economic Research, Cambridge, Mass).

Huang, Jennifer, Kelsey D. Wei, and Hong Yan, 2007, Participation costs and the sensitivity of fund flows to past performance, *Journal of Finance* 62, 1273-1311.

Huang, Jun, and Albert Y. Wang, 2015, The predictability of managerial heterogeneities in mutual funds, *Financial Management* 44, 947–979.

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

Jensen, Michael C., 1968, The performance of mutual funds in the period 1945-1964, *Journal of Finance* 23, 389–416.

Lovén, Johanna, Agneta Herlitz, and Jenny Rehnman, 2011, Women’s own-gender bias in face recognition memory, *Experimental Psychology* 58, 333-340.

Neumark, David, Roy J. Bank, and Kyle D. Van Nort, 1996, Sex discrimination in restaurant hiring: An audit study, *The Quarterly Journal of Economics* 111, 915-941.

Newton, David, and Mikhail Simutin, 2015, Of age, sex, and money: Insights from corporate officer compensation on the wage inequality between genders, *Management Science* 61, 2355–2375.

Niessen-Ruenzi, Alexandra, and Stefan Ruenzi, 2018, Sex matters: Gender bias in the mutual fund industry, *Management Science* 65, 3001–3025.

Park, Albert, and Pungpond Rukumnuaykit, 2004, Eat drink man woman: Testing for gender bias in China using individual nutrient intake data, unpublished working paper, University of Michigan.

Polachek, Solomon W., 1981, Occupational self-selection: A human capital approach to sex differences in occupational structure, *Review of Economic Statistics* 63, 60–69.

Rakowski, David, 2010, Fund flow volatility and performance, *Journal of Financial and Quantitative Analysis* 45, 223–237.

Rau, P. Raghavendra, Jason Sandvik, and Theo Vermaelen, 2021, Are women undervalued? Board gender diversity and IPO underpricing, working paper, University of Cambridge.

Richman, Judith A., Kathleen M. Rospenda, Stephanie J. Nawyn, Joseph A. Flaherty, and et al., 1999, Sexual harassment and generalized workplace abuse among university employees: Prevalence and mental health correlates, *American Journal of Public Health* 89, 358-63.

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

Sharpe, William F., 1965, Risk-Aversion in the Stock Market: Some Empirical Evidence, *Journal of Finance* 20, 416–422.

Sirri, Erik R., and Peter Tufano, 1998, Costly search and mutual fund flows, *Journal of Finance* 53, 1589–1622.

Sutter, Matthias, and Daniela Gätzle-Rützler, 2014, Gender differences in the willingness to compete emerge early in life and persist, *Management Science* 61, 2339–2354.

Tian, Hao, Can Gao, Xinyan Xiao, Hao Liu, Bolei He, Hua Wu, Haifeng Wang, and Feng Wu, 2020, SKEP: Sentiment Knowledge Enhanced Pre-training for Sentiment Analysis, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (Association for Computational Linguistics, Online)*.

Treynor, Jack L., 1965, How to rate management of investment funds, *Harvard Business Review* 43, 63-75.

Treynor, Jack, and Kay Mazuy, 1966, Can mutual funds outguess the market, *Harvard Business Review* 44, 131–136.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin, 2017, Attention is All you Need, *Advances in Neural Information Processing Systems (Curran Associates, Inc.)*.

Wermers, Russ, 2000, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses, *The Journal of Finance* 55, 1655–1695.

Wooldridge, Jeffrey M., 2001, *Econometric analysis of cross section and panel data* (MIT Press, Cambridge, Mass).

Zhang, Zhengyan, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu, 2019, ERNIE: Enhanced Language Representation with Informative Entities, *arXiv:1905.07129 [cs]*.

Appendix. Variable Definitions

Variable	Description	Source
User Variables		
Current month spending	Total amount paid out from the platform account in the current month (in thousands CNY).	Platform User Table
User age	User age.	Platform User Table
User gender	User gender.	Platform User Table
User city size	The tier of the user's city of residency. 1 is the biggest city (Beijing, Shanghai, etc.), and 6 is the smallest city.	Platform User Table
User risk band	User risk band, from 0 - 5, where 0 represents the highest level of risk aversion and 5 represents the lowest.	Platform User Table
User fund amount	Fund holding amount (in CNY).	Platform Fund Table
User monthly capital gain or loss	Monthly capital gain or loss of the user's current fund holding relative to the value of the position as of the previous month (in CNY).	Platform Fund Table
User fund flow	The fund flow for non-first-time investors.	Platform Fund Table
Female new-born ratio	The proportion of female new-borns amongst all new-borns in 2010 at the local district level.	China Census 2010
Female illiteracy ratio	The proportion of illiterate females in the female population at the local district level.	China Census 2010
Manager Variables		
Manager gender	Fund manager gender (0 for male and 1 for female).	CSMAR Fund Manager Database
Manager degree	Manager's education background.	CSMAR Fund Manager Database
Manager tenure	The number of years the manager has been managing the current fund.	CSMAR Fund Manager Database
Media mention frequency rank	The ranking of the number of times each manager's name is mentioned in the news in each month.	CSMAR News Database
Positive (negative) mention frequency rank	The ranking of the number of times each manager's name is positively (negatively) mentioned in the news in each month.	CSMAR News Database
Fund Variables		
Fund risk	Fund's standard deviation of daily returns.	CSMAR Fund Finance Database
Log(TNA)	Logarithm of Total Net Assets	CSMAR Fund Finance Database
Log(Fund age)	Logarithm of the number of months since fund's inception.	CSMAR Fund Finance Database
Aggregate fund flow	The aggregate level fund flow at the quarterly level.	CSMAR Fund Finance Database
Fund sales fee (%)	Fund's annual selling service fee (standardized by dividing by the fund's annual total asset value).	CSMAR Fund Finance Database
Fund management fee (%)	Fund's annual remuneration of managers (standardized by dividing by the fund's annual total asset value).	CSMAR Fund Finance Database
Fund transaction fee (%)	Fund's annual transaction fee (standardized by dividing by the fund's annual total asset value).	CSMAR Fund Finance Database
Fund 1-month NAV return	Previous-month fund-level NAV return (adjusted for splits and dividends) in decimals.	CSMAR Fund Evaluation Database

Fund 1-month NAV return rank	Previous-month fund level NAV return ranked on a percentile basis, with 0 being the worst performing fund and 1 being the best.	CSMAR Fund Evaluation Database
Fund 3-month NAV return rank	Previous 3-month fund level NAV return ranked on a percentile basis, with 0 being the worst performing fund and 1 being the best.	CSMAR Fund Evaluation Database
Fund 6-month NAV return rank	Previous 6-month fund level NAV return ranked on a percentile basis, with 0 being the worst performing fund and 1 being the best.	CSMAR Fund Evaluation Database
Fund 12-month NAV return rank	Previous 12-month fund level NAV return ranked on a percentile basis, with 0 being the worst performing fund and 1 being the best.	CSMAR Fund Evaluation Database
Alpha rank	Fund's alpha calculated using the CAPM on past-month daily returns ranked into percentiles.	CSMAR Fund Evaluation Database
Daily return rank	Fund's average daily return for the past month ranked into percentiles.	CSMAR Fund Evaluation Database
Sharpe ratio	Fund's Sharpe ratio calculated using daily returns in the current month.	CSMAR Fund Evaluation Database
CAPM beta	The beta coefficient from the CAPM model calculated using daily returns in the current month.	CSMAR Fund Evaluation Database
CAPM alpha	The alpha coefficient from the CAPM model calculated using daily returns in the current month.	CSMAR Fund Evaluation Database
Treynor index	Fund's Treynor index calculated using daily returns in the current month.	CSMAR Fund Evaluation Database
CL-Model bear market timing	Chang and Lewellen (1984)'s measure of the fund's market timing ability in bear markets.	CSMAR Fund Evaluation Database
CL-Model bull market timing	Chang and Lewellen (1984)'s measure of the fund's market timing ability in bull markets.	CSMAR Fund Evaluation Database
CL-Model stock selection	Chang and Lewellen (1984)'s measure of stock selection ability.	CSMAR Fund Evaluation Database

Figure 1. Illustration of Fund Ranking Information Available to Investors on the platform

Fund rankings on the platform are displayed as “infinite scrolls”. As users scroll down the list, more funds instantly appear at the bottom. Users can choose to rank funds by their past 1-month, 3-month or 6-month returns, shown in panels A, B, and C, respectively.

Figure 2. Illustration of Fund Manager Information Available to Investors on the platform

Panels A and B show the profile pages for a female and a male fund manager, respectively, on the platform. Each panel shows the manager's profile image, tenure, and reviews. Panel C shows the ranking of the historical performance of a fund on the platform by its past 1-, 3-, 6-, and 12-month returns.

A

B

历史业绩 历史净值

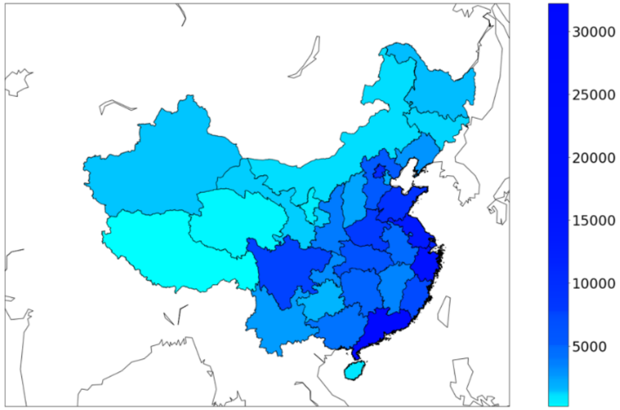
时间区间	涨跌幅	同类业绩排名ⓘ
近1周	+7.59%	14 / 3587
近1月	+26.08%	2 / 3534
近3月	+61.97%	1 / 3392
近6月	+42.11%	6 / 3059
近1年	+102.39%	18 / 2565

更多数据 >

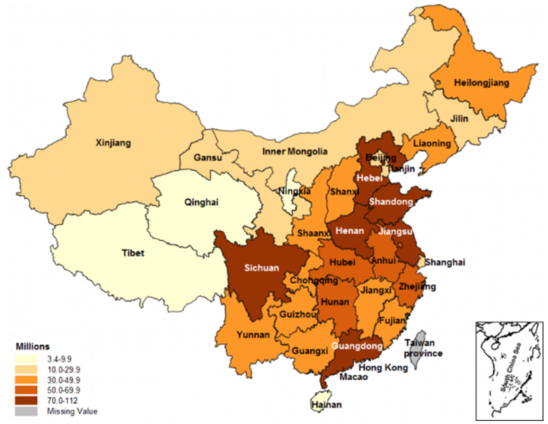
C

Figure 3. Geographic Distribution of Stock Fund Users, Amounts and Profits

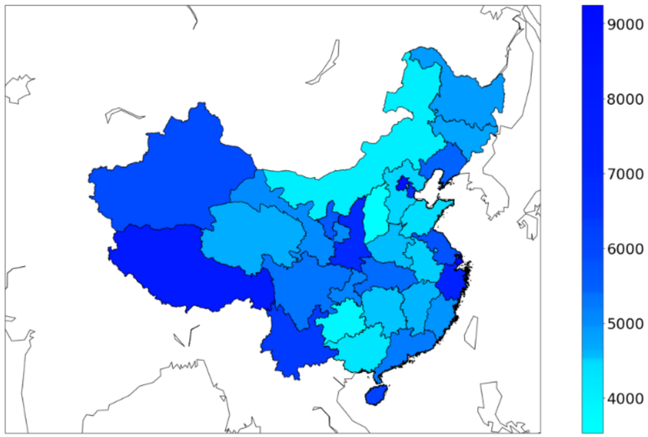
Panel A shows the geographic distribution of the number of stock fund users in mainland China. The deeper the color, the more users there are in the province relative to other provinces. Panel B shows the geographic distribution of the population in China.[12] The deeper the color, the higher the number of people who reside in the province relative to other provinces. Panel C shows the geographic distribution of average stock fund amounts in CNY. The deeper the color, the larger the average fund amount in the province relative to other provinces. Panel D shows the geographic distribution of user's monthly average profits in CNY. The deeper the color, the higher the average profits in the province relative to other provinces.



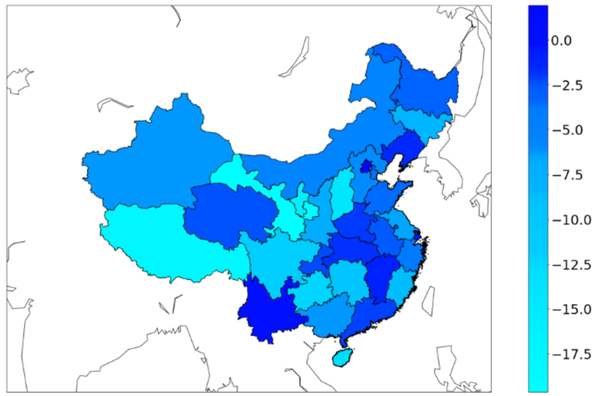
A



B



C



D

[12] Source: United Nations Children's Emergency Fund. URL: https://www.unicef.cn/en/figure-13-population-density-province-2017

Figure 4 Geographic Distribution of Instrumental Variables

Panel A shows the geographic distribution of the proportions of illiterate women amongst all women in the respective provinces in mainland China in 2010. The deeper the color on the map, the higher the proportion of illiterate women in the province. Panel B shows the distribution of the proportions of female newborns amongst all newborns in the respective provinces in mainland China in 2010. The deeper the color on the map, the higher the proportion of female newborns in the province.

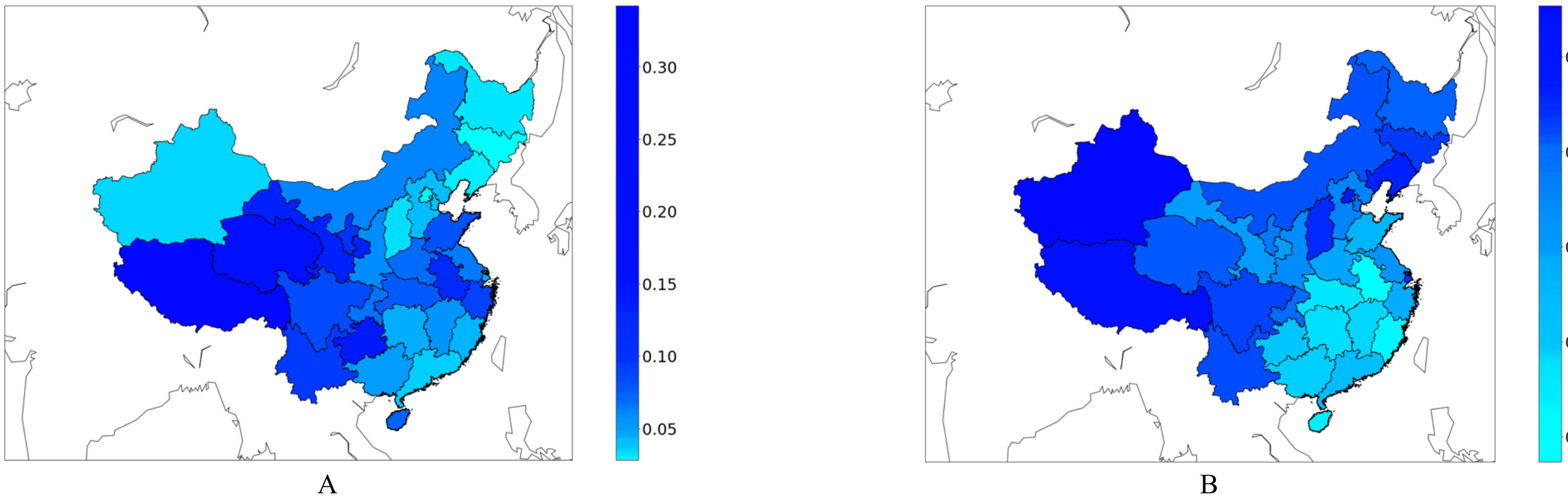


Table 1. Sample Descriptive Statistics

This table reports summary statistics for the funds, users, and managers. Panel A reports the mean, standard deviation, and the number of observations for fund and manager characteristics. Panel B reports the mean, standard deviation, and the number of observations for the platform users' fund investments and characteristics, respectively. All fees, fund amounts, user monthly capital gain/loss, and user monthly spending amount are in CNY. Return ranks are in percentiles.

Panel A.1 Fund characteristics

	Entire sample		Male Managers		Female Managers		Two Sample T-test
	N (Fund-Months) = 3,515		N (Fund-Months) = 2,924		N (Fund-Month) = 591		
Variable	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**p-value**
Fund 1-month NAV return rank	0.51	0.29	0.50	0.29	0.53	0.27	0.02
Fund 3-month NAV return rank	0.51	0.29	0.50	0.29	0.54	0.28	0.01
Fund 6-month NAV return rank	0.51	0.29	0.50	0.29	0.54	0.29	0.00
Fund 12-month NAV return rank	0.51	0.29	0.49	0.29	0.56	0.29	0.00
Fund Flow	0.67	10.64	0.69	10.81	0.50	9.12	0.00
Fund TNA (thousands CNY)	981,601.40	1,388,371.00	1,022,803.00	1,462,619.00	777,752.60	912,182.50	0.00
Fund age (Months)	41.15	20.26	40.98	18.59	41.99	27.05	0.27
Manager tenure (in years)	2.12	1.35	2.19	1.38	1.70	1.08	0.00
Sales fee	0.01%	0.00	0.01%	0.00	0.02%	0.00	0.00
Management fee	1.69%	0.01	1.71%	0.01	1.57%	0.00	0.00
Transaction fee	1.04%	0.01	1.07%	0.01	0.89%	0.01	0.00
Media mention frequency (rank)	0.11	0.28	0.11	0.28	0.08	0.23	0.00
Positive mention frequency rank	0.10	0.27	0.11	0.28	0.08	0.23	0.01
Negative mention frequency rank	0.03	0.14	0.04	0.15	0.02	0.09	0.01

Panel A.2 Fund objectives

Variable	Entire sample Fund-Month Count	Male Managers Fund-Month Count	Female Managers Fund-Month Count	Female Managers Percentage
Fund Objective - Appreciation	1,704	1,484	220	12.91%
Fund Objective - Stable Growth	307	263	44	14.33%
Fund Objective - Growth	767	572	195	25.42%
Fund Objective - Income	737	605	132	17.91%

Panel A.3. Manager education (fund-month counts)

	Entire sample	Male Managers	Female Managers
Manager Degree - Undergraduate	86	86	0
Manager Degree - Master	3,459	2,924	535
Manager Degree - MBA / EMBA	36	36	0
Manager Degree - Doctoral	331	265	66

Panel B. User data

Variable	Entire sample N (User-Months) = 2,345,875 Mean	Std. Dev.	Male Users N (User-Months) = 2,090,157 Mean	Std. Dev.	Female Users N (User-Months) = 255,718 Mean	Std. Dev.	Two Sample T-test p-value
Fund Amount	5,893.31	24,928.60	5,963.80	25,308.92	5,317.15	21,561.50	0.00
Fund Flow	0.67	10.66	0.69	10.83	0.51	9.13	0.00
User Monthly Capital Gain/Loss (CNY)	-3.94	1,546.76	-4.01	1,575.02	-3.36	1,292.82	0.84
User Monthly Spending Amount (thousands CNY)	5.30	20.07	5.31	20.39	5.27	17.20	0.33
User Age	32.40	8.31	32.41	8.32	32.36	8.22	0.01
User City Size	2.43	1.25	2.43	1.25	2.44	1.26	0.00
User Risk Tolerance	3.16	1.06	3.16	1.06	3.19	1.06	0.00

Table 2

This table reports coefficients from panel-OLS regressions for the effect of funds' prior months' performance on users' current month fund flow for the period between August 2017 to July 2019. In Panel A, the dependent variable is each user's fund flow, defined as the difference between the user's current month fund amount and the prior month fund amount (adjusting for capital gains or losses), divided by previous month fund amount. The 1-, 3-, 6-, and 12-month return rankings are the fund's NAV return over the respective 1-, 3-, 6- and 12-months ranked into percentiles. Columns 1-2 use the funds' prior month NAV return ranked within funds with the same fund objective as the measure of fund performance. Columns 3-5 use the funds' past 3-, 6- and 12- NAV returns ranked within funds with the same fund objective as the measure of fund performance, respectively. The rolling average spending over the past 6 months is computed as the rolling average of each user's spending on the fintech platform over the past 6 months and the rolling spending deviation over the prior 6 months is the standard deviation of each user's spending on the fintech platform over the past 6 months. Both the rolling average and rolling standard deviations are standardized to avoid extremely large or extremely small coefficients. Panel B reports coefficients for the effects of previous-quarter fund performance on quarterly level fund flows. Fund flows are aggregated at the quarterly level and are winsorized at the 99% level. All other variables are defined in the Appendix. The panel-OLS regressions in Panel A include user, month, fund objective and manager degree fixed effects, whose coefficients are suppressed. The panel-OLS regressions in Panel B include fund, month, fund objective and manager degree fixed effects, whose coefficients are suppressed. The base gender dummy variable is male. Standard errors, clustered at the user level in Panel A and at the fund level in Panel B, are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Fund flow sensitivity to performance at the individual investor level

	Lagged alternative performance measures (Independent Variables) Return ranking over				
	1-month	1-month	3-month	6-month	12-month
	(1)	(2)	(3)	(4)	(5)
Fund performance					
Fund performance ranking	0.3562 ***	0.3559 ***	0.4589 ***	0.4881 ***	0.2652 ***
	(0.0251)	(0.0251)	(0.0262)	(0.0286)	(0.0302)
Managerial characteristics					
Manager gender (Female)		0.0296	0.0142	-0.0078	0.0186
		(0.0284)	(0.0284)	(0.0285)	(0.0294)
Manager tenure	0.0447 ***	0.0464 ***	0.0443 ***	0.0386 ***	0.0406 ***
	(0.0066)	(0.0070)	(0.0070)	(0.0070)	(0.0071)
Fund characteristics					
Fund risk (Previous month)	0.3588 ***	0.3596 ***	0.3714 ***	0.3567 ***	0.4490 ***
	(0.0314)	(0.0314)	(0.0312)	(0.0314)	(0.0329)

Log(Fund age)	0.0245	0.0182	0.0257	0.0388	0.0355
	(0.0282)	(0.0292)	(0.0292)	(0.0295)	(0.0334)
Log(TNA) (previous quarter)	0.0504 ***	0.0521 ***	0.0530 ***	0.0386 ***	0.0303 ***
	(0.0092)	(0.0092)	(0.0092)	(0.0092)	(0.0095)
Aggregate fund flow (previous quarter)	-0.0214 ***	-0.0214 ***	-0.0244 ***	-0.0296 ***	0.0133
	(0.0038)	(0.0038)	(0.0037)	(0.0038)	(0.0120)
Fund sales fee (%)	0.7368 ***	0.7478 ***	0.8423 ***	0.7666 ***	0.8080 ***
	(0.2529)	(0.2529)	(0.2523)	(0.2539)	(0.2689)
Fund management fee (%)	-0.1928 ***	-0.1944 ***	-0.1535 ***	-0.1338 ***	-0.1659 ***
	(0.0199)	(0.0199)	(0.0202)	(0.0202)	(0.0207)
Fund transaction fee (%)	-0.0220	-0.0208	-0.0246	-0.0179	-0.0186
	(0.0166)	(0.0167)	(0.0167)	(0.0167)	(0.0171)
User income					
Rolling average spending in prior 6 months (thousands CNY)	0.0027	0.0027	0.0027	0.0026	0.0028
	(0.0030)	(0.0030)	(0.0030)	(0.0030)	(0.0031)
Std dev of rolling spending levels in prior 6 months (thousands CNY)	-0.0024	-0.0024	-0.0024	-0.0024	-0.0024
	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0021)
Current month spending (thousands CNY)	-0.0013 ***	-0.0013 ***	-0.0013 ***	-0.0013 ***	-0.0011 ***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Fixed Effects					
Month	TRUE	TRUE	TRUE	TRUE	TRUE
User	TRUE	TRUE	TRUE	TRUE	TRUE
Manager Degree	TRUE	TRUE	TRUE	TRUE	TRUE
Fund Objective	TRUE	TRUE	TRUE	TRUE	TRUE
Standard Error Clustering					
User	TRUE	TRUE	TRUE	TRUE	TRUE
Base Dummy Variables					
Gender	Male	Male	Male	Male	Male
R^2 (Within)	0.0001	0.0002	0.0003	0.0002	-0.0001
N	2,272,010	2,272,010	2,272,010	2,268,451	2,197,949

Panel B. Fund flow sensitivity to performance at the aggregate fund level

	Dependent Variable
	Quarterly Level Fund Flow
	(1)
Fund performance	
Fund previous quarter return ranking	0.1301***
	(0.0303)
Managerial characteristics	
Manager gender (Female)	-0.0680
	(0.0989)
Manager tenure	0.0122
	(0.0117)
Fund characteristics	
Fund risk (Previous month)	0.1951***
	(0.0700)
Log(TNA) (previous quarter)	-0.3369***
	(0.0722)
Aggregate fund flow (previous quarter)	-0.0017
	(0.0575)
Fund sales fee (%)	-1.8038***
	(0.6228)
Fund management fee (%)	-0.0228***
	(0.0077)
Fund transaction fee (%)	-0.0456***
	(0.0173)
Fixed Effects	
Month	TRUE
Fund	TRUE
Manager Degree	TRUE
Standard Error Clustering	
Fund	TRUE
Base Dummy Variables	
Gender	Male
R^2 (Within)	0.1258
N	2,005

Table 3

This table reports coefficients from panel-OLS regressions for the period between August 2017 to July 2019. In Panel A, the dependent variable is the user's fund flow, defined as the difference between the user's current month fund amount and the prior month fund amount (adjusting for capital gains or losses), divided by previous month fund amount. The 1-, 3-, 6-, and 12-month return rankings are the fund's NAV return over the respective 1-, 3-, 6- and 12-months ranked into percentiles. Column (1) uses funds' prior month NAV returns ranked within funds with the same fund objective as the measure of fund performance. Columns 2-4 use each funds' past 3-, 6- and 12-month NAV returns ranked within funds with the same fund objective as the measure of fund performance, respectively. Panel B Columns 1-3 report coefficients for the effect of current month's fund performance on the future 3-, 6- and 12-month performance, respectively. In Panel C, the dependent variable is each user's fund flow and we split our sample into two subsamples. The winner subsample in Column (1) consists of funds whose returns have increased in the previous month compared to two months ago, and the loser subsample in Column (2) consists of funds whose returns have decreased in the previous month compared to two months previously. Column (3) includes all funds in the sample and adds a dummy variable that indicates if the fund's performance has increased in the previous month compared to two months previously. In Panel D, the dependent variable is each user's fund flow and we apply a piecewise regression approach. Column (1) split funds' previous month NAV returns ranked within funds with the same fund objective into bottom 20%, middle 60% and top 20% quintiles and apply a piecewise regression approach. Column (2) split funds' previous month NAV returns ranked within funds with the same fund objective into bottom 15%, middle 70% and top 15% quintiles and apply a piecewise regression approach. Column (3) split funds' previous month NAV returns ranked within funds with the same fund objective into bottom 10%, middle 80% and top 10% quintiles and apply a piecewise regression approach. Column (4) split funds' previous month NAV returns ranked within funds with the same fund objective into bottom 5%, middle 90% and top 5% quintiles and apply a piecewise regression approach. Column (5) split funds' previous month NAV returns ranked within funds with the same fund objective into bottom 1%, middle 98% and top 1% quintiles and apply a piecewise regression approach. All other variables are defined in the Appendix. The panel-OLS regressions in Panel A and C include user, month, fund objective, and manager degree fixed effects, whose coefficients are suppressed. The panel-OLS regressions in Panel B include fund, month and manager degree fixed effects, whose coefficients are suppressed. The base gender dummy variable is male. Standard errors, clustered at the user level in Panel A and C and at the fund level in Panel B, are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Fund flow-performance sensitivity over different ranking horizons

	Lagged alternative performance measures (Independent Variables) Return ranking over			
	1-month	3-month	6-month	12-month
	(1)	(2)	(3)	(4)
Fund performance				
Fund performance ranking	0.3860 ***	0.4816 ***	0.5178 ***	0.2872 ***
	(0.0263)	(0.0274)	(0.0302)	(0.0317)
Fund performance × Manager gender (Female)	-0.3258 ***	-0.2520 ***	-0.2642 ***	-0.2013 ***
	(0.0781)	(0.0711)	(0.0703)	(0.0724)
Managerial characteristics				
Manager gender (Female)	0.2062 ***	0.1617 ***	0.1626 ***	0.1587 ***
	(0.0472)	(0.0459)	(0.0463)	(0.0520)
Manager tenure	0.0465 ***	0.0441 ***	0.0383 ***	0.0403 ***
	(0.0070)	(0.0070)	(0.0070)	(0.0071)
Fund characteristics				
Fund risk (Previous month)	0.3543 ***	0.3693 ***	0.3560 ***	0.4482 ***
	(0.0313)	(0.0312)	(0.0314)	(0.0329)
Log(Fund age)	0.0173	0.0229	0.0301	0.0234
	(0.0292)	(0.0292)	(0.0294)	(0.0335)

Log(TNA) (previous quarter)	0.0523 ***	0.0533 ***	0.0399 ***	0.0315 ***
	(0.0092)	(0.0092)	(0.0093)	(0.0095)
Aggregate fund flow (previous quarter)	-0.0213 ***	-0.0247 ***	-0.0302 ***	0.0126
	(0.0038)	(0.0037)	(0.0038)	(0.0120)
Fund sales fee (%)	0.7621 ***	0.8627 ***	0.7912 ***	0.8129 ***
	(0.2530)	(0.2521)	(0.2539)	(0.2688)
Fund management fee (%)	-0.1963 ***	-0.1562 ***	-0.1379 ***	-0.1642 ***
	(0.0199)	(0.0201)	(0.0203)	(0.0207)
Fund transaction fee (%)	-0.0258	-0.0294 *	-2.1647	-2.2535
	(0.0167)	(0.0167)	(0.0167)	(0.0171)
User income				
Rolling average spending in prior 6 months	0.0027	0.0027	0.0027	0.0028
(thousands CNY)	(0.0030)	(0.0030)	(0.0030)	(0.0031)
Std dev of rolling spending levels in prior 6	-0.0024	-0.0024	-0.0024	-0.0024
months (thousands CNY)	(0.0020)	(0.0020)	(0.0020)	(0.0021)
Current month spending (thousands CNY)	-0.0013 ***	-0.0013 ***	-0.0013 ***	-0.0011 ***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
R^2 (Within)	0.0002	0.0003	0.0003	-0.0001
N	2,272,010	2,272,010	2,268,451	2,197,949

Panel B. Predictability of future returns based on one month return rankings

	Dependent Variables		
	Future 3-month return rank	Future 6-month return rank	Future 12-month return rank
	(1)	(2)	(3)
Manager gender (Female)	0.0435	0.0256	-0.0184
	(0.0464)	(0.0496)	(0.0625)
Fund current 1-month NAV return rank	-0.0303**	-0.0268**	-0.0496***
	(0.0126)	(0.0111)	(0.0096)
Fund current 1-month NAV return rank × Manager gender (Female)	0.0213	0.0068	0.0055
	(0.0307)	(0.0278)	(0.0209)
Fund Level Flow (Current Quarter)	0.0245*	0.0279*	0.038***
	(0.0142)	(0.0161)	(0.0142)
Manager tenure	-0.0235***	-0.02*	-0.0053
	(0.0078)	(0.0118)	(0.0161)
Fund risk (Current month)	-0.1372***	-0.1141***	-0.1035***
	(0.0232)	(0.0256)	(0.0252)
Log(TNA)	-0.0594***	-0.1066***	-0.1329***
	(0.0196)	(0.0230)	(0.0234)
Fund sales fee (%)	0.0172	-0.3700	-0.2334
	(0.3206)	(0.3185)	(0.2418)
Fund management fee (%)	-0.0146***	-0.0154***	-0.006491
	(0.0050)	(0.0059)	(0.0050)
Fund transaction fee (%)	-0.0294***	-0.0362***	-0.0291**
	(0.0098)	(0.0125)	(0.0117)
R^2 (Within)	0.0061	0.0423	0.0628
N	7,252	7,250	7,029

Panel C. Symmetry in attention bias to increases and decreases in performance

	Subsamples		
	Winners	Losers	All Funds
	(1)	(2)	(3)
Fund performance			
Fund performance ranking	0.8146 ***	0.2798 ***	0.2872 ***
	(0.0783)	(0.0305)	(0.0299)
Fund Performance × Manager gender (Female)	-0.7842 ***	-0.1926 **	-0.2014 **
	(0.1822)	(0.0952)	(0.0937)
Return increase dummy			-0.1406 ***
			(0.0327)
Return increase dummy × Manager gender (Female)			0.2157 **
			(0.0957)
Return increase dummy × Fund performance ranking			0.4819 ***
			(0.0663)
Return increase dummy × Manager gender (Female) ×			-0.5453 ***
Fund performance ranking			(0.1761)
Managerial characteristics			
Manager gender (Female)	0.5110 ***	0.1331 **	0.1560 ***
	(0.1076)	(0.0559)	(0.0543)
Manager tenure	0.0638 ***	0.0365 ***	0.0458 ***
	(0.0169)	(0.0083)	(0.0070)
Fund characteristics			
Fund risk (Previous month)	0.2717 ***	0.4365 ***	0.3609 ***
	(0.0747)	(0.0382)	(0.0314)
Log(Fund age)	-0.1323 *	0.0397	0.0061
	(0.0736)	(0.0337)	(0.0291)
Log(TNA) (previous quarter)	0.0869 ***	0.0516 ***	0.0569 ***
	(0.0232)	(0.0107)	(0.0092)
Aggregate fund flow (previous quarter)	-0.0255 **	-0.0109 **	-0.0181 ***
	(0.0099)	(0.0043)	(0.0037)
Fund sales fee (%)	0.91315	0.8585 ***	0.7888 ***
	(0.5708)	(0.3087)	(0.2525)
Fund management fee (%)	-0.2671 ***	-0.1473 ***	-0.1814 ***
	(0.0528)	(0.0227)	(0.0198)
Fund transaction fee (%)	-0.027151	-0.014163	-0.0362 **
	(0.0402)	(0.0199)	(0.0168)
User income			
Rolling average spending in prior 6 months	0.0072	0.0017	0.0027
(thousands CNY)	(0.0097)	(0.0024)	(0.0030)
Std dev of rolling spending levels in prior 6	-0.0065	-0.0015	-0.0024
months (thousands CNY)	(0.0068)	(0.0015)	(0.0020)
Current month spending (thousands CNY)	-0.0028 **	-0.0008 *	-0.0013 ***
	(0.0012)	(0.0004)	(0.0004)
R^2 (Within)	0.0012	0.0002	0.0003
N	588,136	1,752,974	2,272,010

Panel D. OLS piece-wise regression results, cut on differing levels of fund performance rankings

	Top 20% Middle 60% Bottom 20%	Top 15% Middle 70% Bottom 15%	Top 10% Middle 80% Bottom 10%	Top 5% Middle 90% Bottom 5%	Top 1% Middle 98% Bottom 1%
	1-month	1-month	1-month	1-month	1-month
	(1)	(2)	(3)	(4)	(5)
Fund performance					
Top section fund performance ranking	1.1181 ***	1.7533 ***	3.5568 ***	9.8232 ***	58.571 ***
	(0.1993)	(0.2808)	(0.4539)	(1.0383)	(6.0587)
Top section fund performance × Manager gender (Female)	-0.4203	-0.758	-1.8545	-7.3135 *	-87.979 ***
	(0.5825)	(0.8465)	(1.4706)	(3.8518)	(21.5930)
Mid-section fund performance ranking	0.2727 ***	0.2565 ***	0.2290 ***	0.2323 ***	0.2844 ***
	(0.0497)	(0.0414)	(0.0356)	(0.0305)	(0.0266)
Mid-section fund performance × Manager gender (Female)	-0.2239 *	-0.1826 *	-0.1647 *	-0.1454 *	-0.1516 *
	(0.1182)	(0.1004)	(0.0885)	(0.0813)	(0.0788)
Bottom section fund performance ranking	0.3364 *	0.3857	0.6760 *	1.7871	76.406 *
	(0.1717)	(0.2412)	(0.4085)	(1.1122)	(46.4110)
Bottom section fund performance × Manager gender (Female)	-1.0481 *	-1.8730 **	-3.0886 **	-6.7705 **	-160.36
	(0.5517)	(0.7901)	(1.3219)	(3.1512)	(203.9500)
Managerial characteristics					
Manager gender (Female)	0.3140 ***	0.3818 ***	0.4162 ***	0.4558 ***	1.7391
	(0.0932)	(0.1054)	(0.1234)	(0.1510)	(2.0327)
Manager tenure	0.0452 ***	0.0441 ***	0.0422 ***	0.0413 ***	0.0421 ***
	(0.0069)	(0.0070)	(0.0070)	(0.0070)	(0.0070)
Fund characteristics					
Return standard deviation (Previous month)	0.3287 ***	0.3224 ***	0.3190 ***	0.3133 ***	0.3170 ***
	(0.0313)	(0.0314)	(0.0315)	(0.0315)	(0.0313)
Log(Fund age)	0.0147	0.0143	0.0162	0.0176	0.0074
	(0.0290)	(0.0290)	(0.0291)	(0.0292)	(0.0292)
Log(Total Net Asset) (previous quarter)	0.0550 ***	0.0550 ***	0.0539 ***	0.0525 ***	0.0505 ***
	(0.0092)	(0.0092)	(0.0092)	(0.0092)	(0.0093)
Aggregate fund flow (previous quarter)	-0.0220 ***	-0.0222 ***	-0.0226 ***	-0.0233 ***	-0.0234 ***
	(0.0038)	(0.0038)	(0.0038)	(0.0038)	(0.0038)
Standardized fund sales fee	78.527 ***	78.794 ***	78.354 ***	80.097 ***	78.454 ***
	(25.2880)	(25.2840)	(25.2750)	(25.2850)	(25.2990)
Standardized fund management fee	-20.409 ***	-20.519 ***	-20.415 ***	-20.890 ***	-21.516 ***
	(2.0020)	(2.0034)	(2.0040)	(2.0003)	(1.9935)
Standardized fund transaction fee	-2.4102	-2.4935	-2.6424	-2.1544	-1.166
	(1.6706)	(1.6725)	(1.6738)	(1.6730)	(1.6661)
User income					
Rolling average spending in prior 6 months (thousands CNY)	0.0027	0.0027	0.0027	0.0027	0.0027
	(0.0030)	(0.0030)	(0.0030)	(0.0030)	(0.0030)
Std dev of rolling spending levels in prior 6 months (thousands CNY)	-0.0024	-0.0024	-0.0024	-0.0024	-0.0024
	(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0020)

Current month spending (thousands CNY)	-0.0013 ***	-0.0013 ***	-0.0013 ***	-0.0013 ***	-0.0013 ***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
R^2 (Within)	0.0003	0.0003	0.0003	0.0004	0.0004
N	2,272,010	2,272,010	2,272,010	2,272,010	2,272,010

Table 4

This table reports coefficients from panel-OLS regressions for the effect of funds' prior months' Jensen's alpha ranks or average daily return ranks on users' current month fund flow for the period between August 2017 to July 2019. The dependent variable is each user's fund flow, defined as the difference between the user's current month fund amount and the prior month fund amount, divided by previous month fund amount. Column (1) and (2) use funds' prior month Jensen's alpha ranked within funds with the same fund objective as the measure of fund performance. Column (3) and (4) use funds' previous month daily returns ranked within funds with the same fund objective as the measure of fund performance. The rolling average spending in the past six months is computed as the rolling average of each user's spending on the fintech platform over the past 6 months and the rolling spending deviation over the prior 6 months is the standard deviation of each user's spending on the fintech platform over the past 6 months. Both the rolling average and rolling standard deviations are standardized to avoid extremely large or extremely small coefficients. All other variables are defined in the Appendix. All models include user, month, fund objective, and manager degree fixed effects, whose coefficients are suppressed. The base gender dummy variable is male while the base manager degree level is an undergraduate degree. Standard errors, which are clustered at the user level, are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	Lagged alternative performance measures (Independent Variables)			
	Jensen's Alpha ranking		Average return ranking	
	(1)	(2)	(3)	(4)
Fund performance				
Fund performance	0.3717 ***	0.4031 ***	0.3320 ***	0.3626 ***
	(0.0250)	(0.0262)	(0.0244)	(0.0257)
Fund Performance × Manager gender (Female)		-0.3365 ***		-0.3051 ***
		(0.0758)		(0.0714)
Managerial characteristics				
Manager gender (Female)	0.0294	0.2126 ***	0.0329	0.2006 ***
	(0.0284)	(0.0461)	(0.0284)	(0.0438)
Manager tenure	0.0468 ***	0.0468 ***	0.0467 ***	0.0471 ***
	(0.0070)	(0.0070)	(0.0070)	(0.0070)
Fund characteristics				
Fund risk (Previous month)	0.3474 ***	0.3420 ***	0.3299 ***	0.3254 ***
	(0.0314)	(0.0314)	(0.0315)	(0.0314)
Log(Fund age)	0.0168	0.0153	0.0197	0.0169
	(0.0292)	(0.0292)	(0.0293)	(0.0292)

Log(TNA) (previous quarter)	0.0539 ***	0.0542 ***	0.0517 ***	0.0524 ***
	(0.0092)	(0.0092)	(0.0092)	(0.0092)
Aggregate fund flow (previous quarter)	-0.0232 ***	-0.0234 ***	-0.0231 ***	-0.0234 ***
	(0.0037)	(0.0037)	(0.0037)	(0.0037)
Fund sales fee (%)	0.7483 ***	0.7629 ***	0.7260 ***	0.7407 ***
	(0.2529)	(0.2529)	(0.2529)	(0.2530)
Fund management fee (%)	-0.1908 ***	-0.1925 ***	-0.1932 ***	-0.1948 ***
	(0.0199)	(0.0199)	(0.0199)	(0.0199)
Fund transaction fee (%)	-0.021807	-0.026806	-0.023738	-0.0279 *
	(0.0167)	(0.0167)	(0.0167)	(0.0167)
User income				
Rolling average spending in prior 6 months (thousands CNY)	0.0027	0.0027	0.0027	0.0027
	(0.0030)	(0.0030)	(0.0030)	(0.0030)
Std dev of rolling spending levels in prior 6 months (thousands CNY)	-0.0024	-0.0024	-0.0024	-0.0024
	(0.0020)	(0.0020)	(0.0020)	(0.0020)
Current month spending (thousands CNY)	-0.0013 ***	-0.0013 ***	-0.0013 ***	-0.0013 ***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
R^2 (Within)	0.0002	0.0002	0.0002	0.0002
N	2,272,010	2,272,010	2,272,010	2,272,010

Table 5

This table reports coefficients for the effect of managerial gender on fund performance for the period between August 2017 to July 2019. The dependent variables are various measures of fund performance. Column (1) uses the funds' current month NAV returns ranked within funds with the same fund objective as the dependent variable. Column 2-4 use the funds' current 3-, 6-, and 12-month NAV returns ranked within funds with the same fund objective as dependent variables, respectively. Column (5) uses funds' current-month daily return standard deviation as the dependent variable. Column (6) uses funds' current-month Sharpe ratio as the dependent variable. Column 7-8 uses funds' current-month CAPM beta and alpha as dependent variables. Column (9) uses funds' Treynor index as the dependent variable. Column 10-11 use funds' TM model market timing ability and stock selection ability as dependent variables. Column 12-15 use funds' CL model market timing and stock selection ability as dependent variables. All other variables are defined in the Appendix. All models include fund, month, and manager degree fixed effects, whose coefficients are suppressed. The base gender dummy variable is male. Standard errors, which are clustered at the fund level, are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Fund manager performance regressions

	Dependent Variables							
	1-month return ranking	3-month return ranking	6-month return ranking	12-month return ranking	Return Std. Dev.	Sharpe Ratio	CAPM Beta	CAPM Alpha
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Manager gender (Female)	0.0424*	0.0767	0.0449	-0.0351	-0.0701	0.0099	-0.0212	-0.0017
	(0.0239)	(0.0493)	(0.0647)	(0.0722)	(0.0812)	(0.0161)	(0.0721)	(0.0158)
Log(TNA) (previous quarter)	-0.0517***	-0.0796***	-0.0476***	0.035	0.0262	-0.0222***	0.0364	-0.0314***
	(0.0143)	(0.0178)	(0.0166)	(0.0260)	(0.0319)	(0.0085)	(0.0288)	(0.0080)
Aggregate fund flow (previous quarter)	-0.0014	0.0255*	0.0913***	0.0618***	0.0187	0.0018	-0.0196	-0.0036
	(0.0118)	(0.0155)	(0.0176)	(0.0174)	(0.0177)	(0.0065)	(0.0160)	(0.0060)
Manager tenure	-0.0182***	-0.0252***	-0.0322***	-0.0384***	0.0081	-0.0115***	-0.0052	-0.0099***
	(0.0051)	(0.0072)	(0.0094)	(0.0114)	(0.0091)	(0.0022)	(0.0081)	(0.0026)
Fund risk (Previous Month)	-0.0653***	0.0314	0.0995***	0.1686***				
	(0.0189)	(0.0247)	(0.0235)	(0.0267)				
Fund sales fee (%)	0.338597	0.010095	-0.422233	0.247107	0.316774	0.158524	0.1126	0.26023
	(0.2839)	(0.3673)	(0.3607)	(0.5159)	(0.6907)	(0.2572)	(0.4702)	(0.2181)
Fund management fee (%)	-0.0083*	-0.0165***	-0.0185***	-0.0758**	-0.003582	0.0069**	-0.0112***	-0.003043
	(0.0050)	(0.0064)	(0.0068)	(0.0325)	(0.0027)	(0.0032)	(0.0029)	(0.0024)
Fund transaction fee (%)	-0.0214***	-0.0388***	-0.0584***	-0.023901	0.01552	-0.0183***	0.0275***	-0.0167***
	(0.0082)	(0.0133)	(0.0146)	(0.0208)	(0.0116)	(0.0031)	(0.0104)	(0.0041)
R^2 (Within)	0.0094	0.0338	0.0472	0.0203	-0.0006	0.0127	0.0174	0.0123
N	6,424	6,424	6,405	5,913	6,424	6,424	6,424	6,424

Fund manager performance regressions (continued)

	Dependent Variables						
	Treynor Index	TM-Model Market Timing	TM-Model Stock Selection	CL-Model Bear Market Timing	CL-Model Bull Market Timing	CL-Model Market Timing	CL-Model Stock Selection
	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Manager gender (Female)	0.6975	0.0449	0.0108	-0.0325	-0.0074	0.0251	0.0124
	(0.6593)	(0.0273)	(0.0147)	(0.0913)	(0.0679)	(0.0755)	(0.0202)
Log(TNA) (previous quarter)	-0.1875	0.0073	-0.0332***	0.0577*	0.032	-0.0257	-0.0279**
	(0.1957)	(0.0164)	(0.0091)	(0.0349)	(0.0275)	(0.0326)	(0.0118)
Aggregate fund flow (previous quarter)	-0.0094	-0.0063	-0.0117*	-0.0261	-0.0023	0.0238	-0.0165*
	(0.0720)	(0.0109)	(0.0070)	(0.0203)	(0.0206)	(0.0250)	(0.0092)
Manager tenure	0.0594	-0.0025	-0.0098***	-0.0006	-0.0049	-0.0043	-0.0101**
	(0.0417)	(0.0059)	(0.0028)	(0.0121)	(0.0087)	(0.0134)	(0.0040)
Standardized fund sales fee (%)	0.940788	0.327851	0.2527*	-0.09152	0.194437	0.285952	0.268855
	(2.7052)	(0.2369)	(0.1515)	(0.3778)	(0.6505)	(0.3991)	(0.1636)
Standardized fund management fee (%)	-0.2279***	-0.006065	-0.002134	-0.007477	-0.0119***	-0.004459	-0.0026*
	(0.0326)	(0.0052)	(0.0016)	(0.0079)	(0.0042)	(0.0115)	(0.0014)
Standardized fund transaction fee (%)	0.04102	0.000956	-0.0167***	0.0430***	0.018296	-0.02473	-0.0134***
	(0.0296)	(0.0108)	(0.0041)	(0.0138)	(0.0139)	(0.0174)	(0.0051)
R^2 (Within)	0.0339	0.0020	0.0060	0.0153	0.0026	0.0044	0.0010
N	6,424	6,424	6,424	6,424	6,424	6,424	6,424

Table 6

This table reports coefficients from Propensity-Score-Matching (PSM) regressions for the effects of funds' prior month performance and manager gender on users' current month fund flow for the period between August 2017 to July 2019. Panel A reports the propensity score matching logistic regression pre-matching, where the dependent variable equals 1 if the fund manager is female and 0 if the fund manager is male. Fund objective fixed effects and an intercept are included. Panel B reports pre- and post-match sample covariate balance tests. Panel C reports the post-match panel-OLS regression results, where each female manager is matched to a male manager. Panel C Columns 1-4 report coefficient results when 1-, 3-, 6-, and 12-month fund return ranks are used as fund performance measures, respectively. Manager degree (Undergraduate) is a dummy variable that equals 1 if the fund manager only has an undergraduate degree (or equivalently, do not have a degree equivalent to a Master's degree or higher). All other variables are defined in the Appendix. The base gender dummy variable is male. All models in Panel C include fund, month, and manager degree fixed effects, whose coefficients are suppressed. Standard errors in panel C, which are clustered at the user level, are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Pre-match propensity score regression

	Manager gender
Fund performance	
Fund prior 1-month NAV return rank	0.0606
	(0.1900)
Fund prior 3-month NAV return rank	0.0578
	(0.2470)
Fund prior 6-month NAV return rank	-0.2484
	(0.2820)
Fund prior 12-month NAV return rank	0.6896 ***
	(0.2280)
Fund characteristics	
Fund risk (Previous month)	-0.4415 ***
	(0.1030)
Log(Fund age)	0.8659 ***
	(0.1420)
Log(TNA) (previous quarter)	-0.1300 ***
	(0.0370)
Aggregate fund flow (previous quarter)	0.0405
	(0.0540)
Fund sales fee (%)	1.272255
	(0.8695)
Fund management fee (%)	-0.034297
	(0.0691)
Fund transaction fee (%)	-0.3559 ***
	(0.0827)
Managerial characteristics	
Manager tenure	-0.2716 ***
	(0.0360)
Manager degree (Undergraduate)	-17.5518 ***
	(0.2220)
N	3,512
Pseudo R^2	0.0724

Panel B: T-tests: Covariate balance

	Pre				Post			
	Male	Female	Difference	Two-sample t-test P-value	Male	Female	Difference	Two-sample t-test P-value
Fund performance								
Fund prior 1-month NAV return rank	0.50	0.53	-0.03	0.02	0.55	0.53	0.02	0.31
Fund prior 3-month NAV return rank	0.50	0.53	-0.03	0.01	0.52	0.53	-0.01	0.48
Fund prior 6-month NAV return rank	0.50	0.54	-0.04	0.01	0.53	0.54	0.00	0.80
Fund prior 12-month NAV return rank	0.49	0.56	-0.06	0.00	0.56	0.56	0.00	0.79
Fund characteristics								
Fund risk (Previous month)	1.43	1.36	0.07	0.00	1.37	1.36	0.02	0.47
Log(Fund age)	3.65	3.62	0.03	0.09	3.58	3.62	-0.04	0.16
Log(TNA) (previous quarter)	19.89	19.60	0.30	0.00	19.48	19.60	-0.12	0.24
Aggregate fund flow (previous quarter)	0.03	0.08	-0.05	0.25	0.06	0.08	-0.02	0.60
Fund sales fee	0.01%	0.02%	0.00	0.00	0.02%	0.02%	0.00	0.92
Fund management fee	1.71%	1.57%	0.00	0.00	1.55%	1.57%	0.00	0.66
Fund transaction fee	1.07%	0.89%	0.00	0.00	0.85%	0.89%	0.00	0.35
Managerial characteristics								
Manager tenure	2.44	1.99	0.45	0.00	2.0199	1.9889	0.03	0.66
Undergraduate degree (Count)	69	0	69.00		590	590	0.00	

Panel C. Regression estimates, post-matching

	Lagged alternative performance measures (Independent Variables)			
	Fund prior 1-month NAV return rank	Fund prior 3-month NAV return rank	Fund prior 6-month NAV return rank	Fund prior 12-month NAV return rank
	(1)	(2)	(3)	(4)
Fund performance				
Fund performance ranking	0.3721 ***	0.6278 ***	0.9451 ***	0.3145 ***
	(0.0749)	(0.0877)	(0.1109)	(0.0979)
Fund Performance × Manager gender (Female)	-0.1952 *	-0.3143 ***	-0.5170 ***	-0.2114 *
	(0.1108)	(0.1164)	(0.1242)	(0.1189)
Managerial characteristics				
Manager gender (Female)	0.0954	0.1622 **	0.2470 ***	0.1275
	(0.0724)	(0.0783)	(0.0806)	(0.0881)
Manager tenure	0.0815 ***	0.0784 ***	0.0621 ***	0.0824 ***
	(0.0171)	(0.0173)	(0.0177)	(0.0172)
Fund characteristics				
Fund risk (Previous month)	0.2197 ***	0.1531 **	0.1454 *	0.2639 ***
	(0.0780)	(0.0771)	(0.0777)	(0.0790)
Log(Fund age)	-0.2046 ***	-0.1972 ***	-0.1765 ***	-0.1575 **
	(0.0643)	(0.0638)	(0.0636)	(0.0638)
Log(TNA) (previous quarter)	0.0698 ***	0.0745 ***	0.0659 ***	0.0544 **
	(0.0226)	(0.0225)	(0.0226)	(0.0227)
Aggregate fund flow (previous quarter)	-0.0469	-0.0431	-0.1032 **	-0.043
	(0.0416)	(0.0407)	(0.0426)	(0.0410)
Fund sales fee (%)	-0.16025	0.001715	-0.12596	-0.22429
	(0.5620)	(0.5637)	(0.5640)	(0.5650)
Fund management fee (%)	-0.1896 ***	-0.1321 **	-0.0861	-0.1767 ***
	(0.0547)	(0.0548)	(0.0554)	(0.0540)
Fund transaction fee (%)	0.01188	-0.012067	-0.000244	0.002672
	(0.0508)	(0.0508)	(0.0499)	(0.0500)
User income				
Rolling average spending in prior 6 months (thousands CNY)	-0.002	-0.002	-0.002	-0.0019
	(0.0028)	(0.0028)	(0.0028)	(0.0028)
Std dev of rolling spending levels in prior 6 months (thousands CNY)	0.0024	0.0025	0.0025	0.0024
	(0.0021)	(0.0021)	(0.0021)	(0.0021)
Current month spending (thousands CNY)	-0.0013	-0.0013 *	-0.0014 *	-0.0013
	(0.0008)	(0.0008)	(0.0008)	(0.0008)
R^2 (Within)	0.0004	0.0008	0.0010	0.0003
N	676,270	676,270	676,270	676,270

Table 7

This table reports coefficients from panel-OLS regressions for the period between August 2017 to July 2019, controlling for the frequency of media mentions of fund managers' names. The dependent variable is the user's fund flow. Media mention frequency rank is the ranking of the number of times each manager's name is mentioned in the news in each month. Positive mention frequency rank is the ranking of the number of times each manager's name is positively mentioned in the news in each month. Negative mention frequency rank is the ranking of the number of times each manager's name is negatively mentioned in the news in each month. Media mention frequency × manager gender (Female) is the interaction term between media mention frequency rank and the gender of the fund manager. Negative (positive) mention frequency rank × manager gender (Female) × Fund Performance is the triple interaction term of negative (positive) mention frequency, the gender of the fund manager and fund performance. All other variables are defined in the Appendix. The panel-OLS regressions include user, month, fund objective, and manager degree fixed effects, whose coefficients are suppressed. The base gender dummy variable is male. Standard errors, clustered at the user level in Panel A and C and at the fund level in Panel B, are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	Lagged alternative performance measures (Independent Variables) Return ranking over			
	1-month	1-month	1-month	1-month
	(1)	(2)	(3)	(4)
Media				
Media mention frequency rank	0.2537 ***			
	(0.0269)			
Media mention frequency rank × Manager gender (Female)	-0.3727 ***			
	(0.0721)			
Negative mention frequency rank		-0.1867 ***	-0.0871	-0.0906
		(0.0445)	(0.0608)	(0.0631)
Positive mention frequency rank		0.2366 ***	-0.1292 ***	-0.1431 ***
		(0.0269)	(0.0452)	(0.0466)
Negative mention frequency rank × Fund Performance			0.0279	-0.0133
			(0.1519)	(0.1576)
Positive mention frequency rank × Fund Performance			0.6780 ***	0.7766 ***
			(0.0824)	(0.0868)
Manager gender (Female) × Negative mention frequency rank × Fund Performance				0.539
				(0.3970)
Manager gender (Female) × Positive mention frequency rank × Fund Performance				-1.2285 ***
				(0.3128)
Fund performance				
Fund performance ranking	0.3965 ***	0.3797 ***	0.2384 ***	0.2169 ***
	(0.0265)	(0.0268)	(0.0290)	(0.0292)
Fund Performance × Manager gender (Female)	-0.2920 ***	-0.3206 ***	-0.2700 ***	-0.0367
	(0.0779)	(0.0781)	(0.0779)	(0.0831)
Managerial characteristics				
Manager gender (Female)	0.2525 ***	0.2012 ***	0.1691 ***	0.1112 **
	(0.0480)	(0.0471)	(0.0471)	(0.0494)
Manager tenure	0.0499 ***	0.0505 ***	0.0508 ***	0.0529 ***
	(0.0070)	(0.0070)	(0.0070)	(0.0070)

Fund characteristics				
Fund risk (Previous month)	0.3364 ***	0.3272 ***	0.3162 ***	0.3116 ***
	(0.0312)	(0.0311)	(0.0312)	(0.0314)
Log(Fund age)	0.0287	0.0343	0.0271	0.0234
	(0.0293)	(0.0293)	(0.0293)	(0.0293)
Log(TNA) (previous quarter)	0.0389 ***	0.0415 ***	0.0411 ***	0.0408 ***
	(0.0093)	(0.0093)	(0.0093)	(0.0093)
Aggregate fund flow (previous quarter)	-0.0198 ***	-0.0199 ***	-0.0208 ***	-0.0209 ***
	(0.0038)	(0.0038)	(0.0038)	(0.0038)
Fund sales fee (%)	0.7884 ***	0.7971 ***	0.7277 ***	0.7173 ***
	(0.2529)	(0.2529)	(0.2532)	(0.2530)
Fund management fee (%)	-0.1855 ***	-0.1834 ***	-0.1810 ***	-0.1825 ***
	(0.0200)	(0.0200)	(0.0200)	(0.0201)
Fund transaction fee (%)	-0.022198	-0.02312	-0.016431	-0.019658
	(0.0167)	(0.0167)	(0.0167)	(0.0167)
User income				
Rolling average spending in prior 6 months (thousands CNY)	0.0027	0.0027	0.0027	0.0027
	(0.0030)	(0.0030)	(0.0030)	(0.0030)
Std dev of rolling spending levels in prior 6 months (thousands CNY)	-0.0024	-0.0024	-0.0024	-0.0024
	(0.0020)	(0.0020)	(0.0020)	(0.0020)
Current month spending (thousands CNY)	-0.0013 ***	-0.0013 ***	-0.0013 ***	-0.0013 ***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
R^2 (Within)	0.0004	0.0004	0.0005	0.0006
N	2,272,010	2,272,010	2,272,010	2,272,010

Table 8

This table reports coefficients from panel-OLS regressions for the effects of funds' prior month performance and manager gender on users' current month fund flow for different subsamples. Column (1) reports coefficients where the triple-interaction term is fund performance × manager gender (female) × investor gender. Column (2) reports coefficients where the triple-interaction term is fund performance × manager gender (female) × investor age. Column (3) reports coefficients where the triple-interaction term is fund performance × manager gender (female) × investor city size. Column (4) reports coefficients where the triple-interaction term is fund performance × manager gender (female) × investor risk aversion. All other variables are defined in the Appendix. All models include user, month, fund objective and manager degree fixed effects, whose coefficients are suppressed. Standard errors, which are clustered at the user level, are reported in parentheses. The base gender dummy variable is male. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	User Characteristics (Triple Interaction Variables)			
	(1)	(2)	(3)	(4)
User characteristics				
User gender	-0.1966 ***	-0.3200 ***	-0.3200 ***	-0.3222 ***
	(0.0234)	(0.0162)	(0.0162)	(0.0163)
Fund performance × User gender	-0.2495 ***			
	(0.0458)			
Fund performance × User gender × Manager gender (Female)	0.1319 *			
	(0.0696)			
User age	0.0097 ***	0.0044 ***	0.0097 ***	0.0093 ***
	(0.0011)	(0.0015)	(0.0011)	(0.0011)
Fund performance × User age		0.0105 ***		
		(0.0029)		
Fund performance × User age × Manager gender (Female)		-0.0031		
		(0.0043)		
User city tier	0.0371 ***	0.0371 ***	0.0204 *	0.0363 ***
	(0.0070)	(0.0070)	(0.0105)	(0.0070)
Fund performance × User city tier			0.0382 *	
			(0.0199)	
Fund performance × User city tier × Manager gender (Female)			-0.0581 **	
			(0.0285)	
User risk band	0.0075	0.0077	0.0075	-0.0044
	(0.0087)	(0.0087)	(0.0087)	(0.0132)
Fund performance × User risk band				0.0081
				(0.0240)
Fund performance × User risk band × Manager gender (Female)				0.0116
				(0.0354)
Fund performance				
Fund performance	0.4897 ***	0.0603	0.3094 ***	0.3749 ***
	(0.0335)	(0.0952)	(0.0525)	(0.0798)
Fund performance × Manager gender (Female)	-0.4651 ***	-0.3151 **	-0.2764 ***	-0.4555 ***
	(0.0802)	(0.1537)	(0.1009)	(0.1313)
Managerial characteristics				
Manager gender (Female)	0.2861 ***	0.2837 ***	0.2847 ***	0.2861 ***
	(0.0404)	(0.0404)	(0.0404)	(0.0406)

Manager tenure	0.0690 ***	0.0688 ***	0.0689 ***	0.0688 ***
	(0.0057)	(0.0057)	(0.0057)	(0.0057)
Fund characteristics				
Fund risk (Previous month)	0.4809 ***	0.4811 ***	0.4807 ***	0.4826 ***
	(0.0274)	(0.0274)	(0.0274)	(0.0275)
Log(Fund age)	-0.0996 ***	-0.0997 ***	-0.1004 ***	-0.1009 ***
	(0.0240)	(0.0240)	(0.0240)	(0.0241)
Log(TNA) (previous quarter)	0.1009 ***	0.1012 ***	0.1012 ***	0.1014 ***
	(0.0072)	(0.0072)	(0.0072)	(0.0072)
Aggregate fund flow (previous quarter)	-0.0203 ***	-0.0204 ***	-0.0204 ***	-0.0203 ***
	(0.0033)	(0.0033)	(0.0033)	(0.0033)
Fund sales fee (%)	1.3562 ***	1.3548 ***	1.3574 ***	1.3660 ***
	(0.2062)	(0.2062)	(0.2062)	(0.2077)
Fund management fee (%)	-0.4018 ***	-0.4012 ***	-0.4019 ***	-0.4034 ***
	(0.0196)	(0.0196)	(0.0196)	(0.0197)
Fund transaction fee (%)	0.004128	0.003945	0.004291	0.003225
	(0.0146)	(0.0146)	(0.0146)	(0.0146)
User income				
Rolling average spending in prior 6 months (thousands CNY)	0.0038	0.0038	0.0038	0.0039
	(0.0024)	(0.0024)	(0.0024)	(0.0024)
Std dev of rolling spending levels in prior 6 months (thousands CNY)	-0.0017	-0.0017	-0.0017	-0.0017
	(0.0015)	(0.0015)	(0.0015)	(0.0015)
Current month spending (thousands CNY)	-0.0005	-0.0005	-0.0005	-0.0005
	(0.0004)	(0.0004)	(0.0004)	(0.0004)
R^2 (Within)	-0.0001	-0.0001	-0.0001	-0.0001
N	2,268,870	2,268,870	2,268,870	2,258,747

Table 9

This table reports coefficients from logistic regressions for first-time fund investors on the fintech platform. The dependent variable is the gender of the fund manager when the investor invests for the first time. Column (1) includes only regressors that describe user characteristics. Column (2) includes only regressors that describe fund characteristics. Column (3) includes regressors that both describe fund and user characteristics. All variables are defined in the Appendix. All models include fund objective fixed effects, whose coefficients are suppressed. Model 2 also includes manger degree fixed effects. Standard errors, which are heteroskedasticity robust, are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)
Fund performance		
Fund prior 1-month NAV return rank	-0.3347 ***	-0.4617 ***
	(0.0290)	(0.0380)
Managerial characteristics		
Manager tenure		-0.7833 ***
		(0.0110)
Fund characteristics		
Fund risk (Previous month)		-0.0722 ***
		(0.0220)
Log(Fund age)		0.4525 ***
		(0.0400)
Log(TNA) (previous quarter)		-0.2684 ***
		(0.0090)
Aggregate fund flow (previous quarter)		-0.2868 ***
		(0.0130)
Fund sales fee (%)		-1.1532 **
		(0.5496)
Fund management fee (%)		0.3031 ***
		(0.0228)
Fund transaction fee (%)		-1.6371 ***
		(0.0505)
User characteristics		
User age	-0.0038 ***	-0.0034 **
	(0.0010)	(0.0010)
User city tier	0.0186 **	0.0300 ***
	(0.0070)	(0.0080)
User risk band	-0.0139	-0.0166
	(0.0090)	(0.0100)
User gender	-0.0342 *	-0.0218
	(0.0200)	(0.0220)
Fintech platform income (High)	-0.0227	-0.031
	(0.0290)	(0.0320)
Fintech platform income (Low)	0.1041	0.1135
	(0.0780)	(0.0830)
Rolling average spending in prior 6 months (thousands CNY)	-0.0023	-0.0018
	(0.0020)	(0.0020)
Std dev of rolling spending levels in prior 6 months (thousands CNY)	0.0014	0.001
	(0.0010)	(0.0020)
Current month spending (thousands CNY)	-0.0017 *	-0.0016 *
	(0.0010)	(0.0010)
Pseudo R^2	0.07929	0.2045
N	142,587	139,565

Table 10

This table reports coefficients from three-stage instrumental variable regressions for the effects of funds' prior month performance and manager gender on users' current month fund flow. Panel A reports the coefficients of the first-stage Logit regressions, while Panel B reports the coefficients of the third-stage instrumental variable regressions. The female new-born ratio is the proportion of new-born female babies amongst all babies in the year 2010 at the local district level. The female illiteracy ratio is the proportion of females that are illiterate amongst all females that are 15 years or order, in the year 2010 at the local district level. All other variables are defined in the Appendix. The Logit regressions in Panel A include fund objective and manager degree fixed effects, whose coefficients are suppressed. The panel-OLS regressions in Panel B include user, month, fund objective and manager degree fixed effects, whose coefficients are suppressed. The base gender dummy variable is male. Standard errors, which are heteroskedasticity robust in Panel A and clustered at the user level in Panel B, are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Panel A. First stage (Logit) regressions

	(1)	(2)
Instrumental Variable		
Female new-born ratio	0.3031 *	
	(0.1580)	
Female illiteracy ratio		-0.4627 ***
		(0.0900)
Fund performance		
Fund prior 1-month NAV return rank	0.2151 ***	0.2150 ***
	(0.0080)	(0.0080)
Managerial characteristics		
Manager tenure	-0.7416 ***	-0.7416 ***
	(0.0020)	(0.0020)
Fund characteristics		
Fund risk (Previous month)	-0.4690 ***	-0.4689 ***
	(0.0060)	(0.0060)
Log(Fund age)	1.5881 ***	1.5879 ***
	(0.0080)	(0.0080)
Log(TNA) (previous quarter)	-0.2790 ***	-0.2790 ***
	(0.0030)	(0.0030)
Aggregate fund flow (previous quarter)	-0.0781 ***	-0.0780 ***
	(0.0010)	(0.0010)
Fund sales fee (%)	-4.3926 ***	-4.3916 ***
	(0.0616)	(0.0616)
Fund management fee (%)	0.5739 ***	0.5739 ***
	(0.0078)	(0.0078)
Fund transaction fee (%)	-0.2937 ***	-0.2938 ***
	(0.0061)	(0.0061)
User characteristics		
User age	0.0011 ***	0.0012 ***
	(0.0000)	(0.0000)
User city tier	0.0080 ***	0.0123 ***
	(0.0020)	(0.0020)
User risk band	0.0282 ***	0.0282 ***
	(0.0030)	(0.0030)

User gender	-0.0166 ***	-0.0174 ***
	(0.0060)	(0.0060)
Fintech platform income (High)	-0.0318 ***	-0.0302 ***
	(0.0090)	(0.0090)
Fintech platform income (Low)	0.0246	0.0249
	(0.0250)	(0.0250)
Rolling average spending in prior 6 months (thousands CNY)	0.0009 **	0.0009 **
	(0.0000)	(0.0000)
Std dev of rolling spending levels in prior 6 months (thousands CNY)	-0.0028 ***	-0.0028 ***
	(0.0000)	(0.0000)
Current month spending (thousands CNY)	0.0001	0.0001
	(0.0000)	(0.0000)
Pseudo R^2	0.2030	0.2030
N	1,928,212	1,928,212

Panel B. Third-stage regressions

	Instruments	
	Female new-born ratio	Female illiteracy ratio
	(1)	(2)
Fund performance		
Fund prior 1-month NAV return rank	0.8013 ***	0.8029 ***
	(0.0387)	(0.0387)
Fund prior 1-month NAV return rank × Manager gender (Female)	-2.5646 ***	-2.5737 ***
	(0.2700)	(0.2697)
Managerial characteristics		
Manager gender (Female)	4.9886 ***	4.9461 ***
	(0.4860)	(0.4831)
Manager tenure	0.3692 ***	0.3651 ***
	(0.0447)	(0.0444)
Fund characteristics		
Fund risk (Previous month)	0.0728 ***	0.0711 ***
	(0.0255)	(0.0254)
Log(Fund age)	-1.2178 ***	-1.2066 ***
	(0.1249)	(0.1242)
Log(TNA) (previous quarter)	0.2336 ***	0.2325 ***
	(0.0143)	(0.0143)
Aggregate fund flow (previous quarter)	0.0024	0.0024
	(0.0034)	(0.0034)
Fund sales fee (%)	2.8238 ***	2.8094 ***
	(0.2859)	(0.2852)
Fund management fee (%)	-0.4931 ***	-0.4910 ***
	(0.0306)	(0.0305)
Fund transaction fee (%)	0.0439 **	0.0427 **
	(0.0212)	(0.0211)
User characteristics		
User age	0.0090 ***	0.0090 ***
	(0.0012)	(0.0012)
User city tier	0.0361 ***	0.0362 ***
	(0.0075)	(0.0075)
User risk band	0.0099	0.0099
	(0.0095)	(0.0095)
User gender	-0.3012 ***	-0.3013 ***
	(0.0179)	(0.0179)
Fintech platform income (High)	0.2207 ***	0.2206 ***
	(0.0253)	(0.0253)
Fintech platform income (Low)	-0.0408	-0.0407
	(0.0769)	(0.0769)
Rolling average spending in prior 6 months (thousands CNY)	0.0044	0.0044
	(0.0027)	(0.0027)
Std dev of rolling spending levels in prior 6 months (thousands CNY)	-0.0018	-0.0018
	(0.0017)	(0.0017)
Current month spending (thousands CNY)	-0.0007	-0.0007
	(0.0005)	(0.0005)
R^2 (Within)	-0.0072	-0.0070
N	1,928,212	1,928,212

Table 11

This table reports coefficients from cross-sectional OLS regressions for the effects of managerial gender on users' fund flow volatilities during our sample period. The dependent variable, fund volatility, is defined as the standard deviation of each user's fund flow for each fund during the sample period. All other variables are defined in the Appendix. The model includes user, fund objective, and manager degree fixed effects, whose coefficients are suppressed. The base gender dummy variable is male. Standard errors, which are clustered at the user level, are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Fund performance	
Fund (Previous) 1-month NAV Return Rank	2.2510 ***
	(0.2377)
Managerial characteristics	
Manager Gender (Female)	-0.1163 *
	(0.0685)
Fund characteristics	
Fund Risk (Previous Month)	0.4364 ***
	(0.1083)
Log(Fund age)	-0.1671 **
	(0.0829)
Log(TNA) (previous quarter)	0.3735 ***
	(0.0279)
Fund sales fee (%)	4.0561 ***
	(0.8354)
Fund management fee (%)	-0.4638 ***
	(0.0730)
Fund transaction fee (%)	0.030346
	(0.0577)
User income	
Standardized Rolling Average Spending in the Past 6 Months	-0.0031
	(0.0130)
Standardized Rolling Spending Standard Deviation in the Past 6 Months	0.0016
	(0.0083)
Current Month Spending	-0.0068
	(0.0043)
R^2 (Within)	0.0048
N	284,753

Chapter 4

Bots Synchronize Stock Returns

Bots Synchronize Stock Returns *

Andreas Park †　　　　Jinhua Wang ‡

Abstract

We document a significant, up to 10-fold increase in the synchronicity of intra-day, ultra-high frequency stock returns over the last decade. This surge in the intra-day synchronicity across stocks coincided with the advent of electronic, automated trading in U.S. markets. Using changes to the S&P500 index, we establish evidence of a causal relationship. When firms are included in this major index, they enter the radar of high frequency arbitrageurs and market-making bots. These automated trading bots, who monitor prices in major securities closely and continuously, increase their quoting activities significantly and cause individual stocks' returns to synchronize at the microstructure level.

*We thank Lasse Pedersen, Erik Theissen, Mark Lipson, Wing Wah Tam, Vincent van Kervel, Raghu Rau, and Eoghan O'Neill for very helpful comments and Michelle Dathan and Christoph Schiller for sharing some data. Andreas Park gratefully acknowledges financial support from SSHRC and Rotman's FinHub. We thank Compute Canada and Cambridge HPC for providing high-performance computing support.

†University of Toronto andreas.park@rotman.utoronto.ca.

‡University of Cambridge, Judge Business School, jw983@jbs.cam.ac.uk.

1. Introduction

Over the last two decades, there has been a well-documented rise in algorithmic and autonomous trading. The broad consensus in the market microstructure literature is that the digitization and automation of trading has lowered trading costs and removed frictions.[1] It is, however, an open question whether the rise in "robotic" or "bot" trading has broader implications for stock returns, costs of capital, and investment.

Computer algorithms are faster, cheaper, and more reliable than humans in implementing mechanical trading strategies such as inter-market arbitrage. Robots can also implement complex portfolio trading strategies, statistical arbitrage, and relative pricing strategies faster than human. While humans typically need minutes to process information, automated quote submissions and trading should cause a faster synchronization of stock prices across the universe of stocks.[2] In this paper, we study the intra-day synchronization of asset stock returns over a long horizon, and we document a substantial increase in the fraction of intra-day stock return variations that are related to market-wide fluctuations. We also study whether relationship between algorithmic, autonomous electronic trading and the substantially stronger intra-day synchronization of returns across securities is causal. Such a relationship would be a pre-requiste for the broader implications on costs of capital.

It is challenging to identify this relationship with the standard, stock by stock microstructure measures because asset pricing measures intrinsically require a multi-asset view. We assess whether bot trading or "bot trading" leads to changes in the speed of stock return synchronization *across* securities by studying the relationship of bot trading and the goodness-of-fit of a standard market model estimated on high frequency, intra-day data. Our goal is to assess how the fraction in the realized return variations that is related to market-wide fluctuations correlates with the extent of bot trading. Motivated by Roll (1988) who, in his 1988 presidential address, discussed the extent to which security returns are captured by the R^2 of an OLS regression of stock returns on the returns of the market portfolio, we focus on a related, intra-day measure. Namely, we estimate OLS regressions of intra-day stock returns on intra-day market returns, record the goodness-of-fit in the form of the regression's R^2 (and the coefficients) and perform a panel regression analysis with the R^2s as the proxy for the degree of return synchronization across securities.

Taking a step back, stock prices move for a number of reasons: stock-specific news releases, shifts in the firm's fundamental value, news about related firms, industry or macro-related news, and so on. We're interested in synchronicity across assets. Asset pricing models yield a particu-

[1] See, for instance, Menkveld (2016) for a comprehensive review of the literature.

[2] We are unaware of research that documents this particular effect. There is work that has found evidence that on a stock-by-stock basis, pricing discrepancies that formerly persisted over days or weeks disappear rapidly in today's markets. For instance, Grégoire and Martineau (2022) show that the post-earnings announcement drift, a topic of much debate in the corporate finance and accounting literature, has all but disappeared since the advent of electronic trading.

lar form of co-movement, namely, one that captures the systematic risks. Under the CAPM, for instance, a well-diversified portfolio should synchronize with the market perfectly in the sense that the R^2 of a regression of average (well-diversified) portfolio returns on market returns is 1. In theory, the price of the respective securities in the portfolio should adjust to changes in market returns instantaneously. That is, whether we measure returns on the millisecond or daily level, the R^2 should be the same. In practice, human market participants need a longer time to digest information and to translate them into justifiable price movements. Computer algorithms, on the other hand, can easily and cheaply identify instances when the adjustment has not yet occurred at high frequencies and act automatically. Moreover, their work scales across arbitrary numbers of securities. We aim to capture this synchronicity, or, rather, the change in synchronicity as induced by the prevalence of bots.

We proceed by estimating, for each security in the NYSE TAQ database and day between 2003 and 2014, a standard market model for intra-day 5-second and 5-minute mid-quote returns. We use the returns for the Russell 2000 index represented by the exchange traded fund IWM as our high-frequency proxy for the market.[3] We interpret R^2 as the fraction of the variation in a stock's returns related to market-wide fluctuations at the intra-day level. We find some striking stylized facts, which can be easily gleaned from Figure 1. From 2003 to early 2007, the R^2 is very small and "flat"; even for the 1,000 largest firms the average R^2 is below 1%. From 2007 to 2010, however, the R^2 increases substantially, from 1% to, on average, 10% for the 1,000 largest firms and it stays at high levels for the remainder of the sample. Notably, although the average intra-day βs also increase over our entire sample horizon, most of the increase in the βs occurs right at the beginning of the sample, between 2003 to early 2007, whereas for the time span when the R^2 increases dramatically, there is no equivalent increase in βs. In other words there is no indication that the higher R^2 is the result of changes in systematic risk; rather, it appears that intra-day returns become more precisely synchronized, or, more formally, that market-wide fluctuations constitute a larger fraction of intra-day stock return variations.

The most significant change in the R^2 occurs between 2006 and 2009. This is a crucial period in the history of equity market structure because the Regulation National Market System (Reg. NMS) came into effect. One feature of Reg. NMS is that marketplaces are formally required to respect one another's quotes which requires these markets to be electronically linked. A second critical change between 2006 and 2009 was the advent of a new type of trader, the so-called high frequency traders which rely on algorithmic, automated, and autonomous order submission strategies.[4] One feature of this type of trader is that they submit and subsequently cancel a very large number of orders, and their presence in the data is reflected, for instance, by

[3]Our results are similar if we use the S&P500, represented by the ETF SPY.

[4]All markets had provided electronic access long before 2007, and traders had the ability to run algorithms for years. The NYSE, for instance, introduced its "auto-quote" system in 2003; see Hendershott, Jones, and Menkveld (2011).

the surge in the number of (top-of-the-book) quote changes over the same time horizon. The visual co-movement of the R^2 and, for instance, the average number of quote changes, documented in Figure 1, is striking — for average monthly figures the correlation of the time series is 56%.

This correlation does not imply causation, as, crucially, there were other developments at the time, such as the financial crisis, the European debt crisis, or shutdowns of the U.S. government, all of which may have increased the importance of systematic risk and raised the R^2s. To establish the causal relationship, we study changes to the S&P 500 constituents. In principle, index membership has no bearing on a firm's fundamental or systematic risks and index in- or exclusions, and if we believe that returns reflect fundamental and systematic risks, then a change in index membership should not affect the R^2 of our regression.

However, when a stock enters or drops out of an index multiple things happen. In our view the most critical consequence is that the stocks that get included in a major index formally become part of many high frequency trading strategies. One straightforward consequence is that funds have to trade in-and-out of positions, and this activity will cause (temporary) volume spikes, which we do observe in the data. For our work, the important change is that these stocks now become relevant for arbitrage strategies on the relationship of index constituents to traded index products (such as futures or ETFs). High frequency traders therefore change their monitoring of prices, and this should affect the speed and precision at which returns synchronize across assets.

Our causal reasoning is therefore as follows: index events are exogenous to the presence of electronic traders. When a firm gets included in an index, nothing changes for the firm itself, but high frequency traders pay more attention to the stock, change prices accordingly, and this affects the R^2 of the return regression (i.e., leads to a stronger return synchronization). The exogenous variation in high-frequency bot trading therefore allows us to establish a causal link between bot trading and the synchronization of stock returns. As firms that are included or excluded from indices tend to be large in market capitalization, we cautiously interpret our causality identification as local average treatment effects for firms that are large, highly liquid and have positive earnings for four consecutive quarters.

To establish the link empirically, we construct a matched sample for all entry and exit events and apply a difference-in-differences panel regression approach. We follow three econometric approaches: First, we use the event in an instrumental variable (IV) approach. This approach assumes that the index event does not affect the R^2 directly (or through other channels than bot trading). Second, we perform a mediation analysis that allows the event to also have a direct effect on the R^2. The third approach uses new tools from the machine learning literature, so-called causal random forests, first introduced in the economics literature by Athey and Imbens (2016). This approach allows us to determine the (causal) treatment effect directly without a matched sample.[5]

The main objective of this part of the analysis is to establish the existence of a causal link from bot trading to the synchronization of returns, and we observe that all three approaches establish a statistically significant effect of bot trading on the intra-day return synchronicity across securities. Namely, after a firm is included in the S&P 500, its R^2 increases significantly, on average by 34%. Concurrently, the number of top-of-the-book quotes increases by about 10%, bid-ask spreads decline by about 1 basis point, and there is a significant increase in order fragmentation. These latter three effects together indicate that there is a substantial increase in bot trading activity. Increased order fragmentation, meaning that liquidity and trading activity spreads across multiple orders, in particular is an effect of high frequency bot trading. Further to this point, the changes occur predominantly in the later part of the sample, precisely when bot trading became prevalent in markets. These findings establish the causal link between (high-frequency) bot trading and the ability of stock returns to reflect changes relative to the market.

The results for index exclusion events are less conclusive: although there is a decline in R^2 of a similar magnitude to that of the increase for index inclusion events, there is no decrease in the many of the activities that we attribute to bot trading after the event (such as the quoting activities). It is possible, for instance, that the activities by algorithmic traders have been (temporarily) substituted with activities by mutual and index funds that trade out of their positions as a result of the index change. We observe that there is an increase in bid-ask spreads and a decrease in order fragmentation, which would be consistent with the departure of high-frequency bot trading. Altogether the case of index exclusions is simply not conclusive to make assertions about the relationship of bot trading and return synchronicity.

We emphasize that we are not testing the intra-day validity of the CAPM akin to the many analyses in the tradition of Fama-French; neither is our goal to find the best possible intra-day model of returns. Rather, our goal is to highlight that the advent of bot trading increased the speed to which variation in intra-day returns is reflected by changes in systematic risk.

We do believe that overall our results indicate a causal relationship between bot trading and the fraction of the variation in firms' returns that are related to market-wide fluctuations.

Related Literature. Our paper contributes to several strands of the literature. First, our work relates to the literature on idiosyncratic volatility and the market model at the monthly or yearly level. Pástor and Pietro (2003) and Fama and French (2004) find consistent evidence that newer firms have more volatile fundamentals and have higher idiosyncratic volatilities. Brown and Kapadia (2007) discover that idiosyncratic risks are higher when newer and riskier firms are getting listed. In addition, Bartram, Brown, and Stulz (2012) find that the market model R^2 increase in firm size, age, cash holdings and liquidity across various countries at the annual

[5] The tree-based classification that is part of a random forest estimation implicitly delivers the "apples-to-apples" comparison that we strive for by constructing matching samples in the classic econometric approaches.

level. Sequentially, Bartram, Brown, and Stulz (2019) discover that as firms listed in the U.S. are getting older, larger and more liquid in the recent years, the idiosyncratic volatility declines and the market model R^2 increases from 2000–2017 at the daily level. Bartram, Brown, and Stulz (2019)'s results are consistent with our findings of an increase in the market model R^2 at the high-frequency intra-day level from 2003–2014. In untabulated graphs, we show that the intra-day R^2 increase is much higher relative to the daily R^2 increase from 2001 to 2014. While firm fundamentals could be a potential factor that leads to a higher daily R^2 in recent years, we posit that bot trading is essential in accelerating the synchronization of returns across securities at the intra-day level.

Second, our work relates to the extensive literature on the impact of algorithmic and high frequency trading on markets that is too expansive to review here. Menkveld (2016) provides a recent, extensive review of the literature. The general consensus is that, at the very least, the increase in electronic trading over the past decade has led to a decline in transaction costs. For instance, using the switch from manual to automated quotes on the New York Stock Exchange in 2003, Hendershott, Jones, and Menkveld (2011) show that increases in algorithmic trading cause an improvement in liquidity; in our work, we will pay special attention to their measure of algorithmic trading. There is also some work in microstructure that studies the impact of high frequency traders' activities on the trading costs for related securities. Namely, Shkilko and Sokolov (2020) studies the relationship of trading costs between futures and ETFs for the same index when high frequency traders do and do not have a speed advantage. Our work is qualitatively different as we are interested in the extend to which returns of securities correlate with the market.

Third, our work relates to the extensive literature on index changes. Vijh (1994) find that S&P500 trading strategies reduce the non-synchronicity of S&P500 constituent stock prices and increase the beta of S&P500 constituents. Denis, McConnell, Ovtchinnikov, and Yu (2003) tracked the realized earnings and earnings-per-share forecasts of stocks that are newly added to the S&P500 index and they find significant improvements in both. They conclude that being included in the S&P500 index is not an information-free event. Chen, Noronha, and Singal (2004) document a permanent increase in the price of firms that are added to an index but they also show that there is no concurrent permanent decline in the price of deleted firms from an index. They attribute the cause of asymmetric price effects to increased investor awareness for firms that at least have been part of an index. Elliott, Van Ness, Walker, and Warr (2006) found evidence that investor awareness and price pressure hypothesis are the factors behind the increase in stock value after inclusion to the S&P500 index. Cai (2007) highlight a significantly positive price reaction for the firms added to an index compared to that of the industry and size matched firms. They conclude that index addition conveys positive information. However, Hrazdil and Scott (2009) found new evidence that index inclusions to S&P500 are information-free events. They attribute the earning increase after index inclusion to larger discretionary accruals instead of increase in cash flows. They also found no relationship between the unexpected earnings and the inclusion date abnormal returns.

Finally, our work relates to the nascent literature on the exchange-traded-funds (ETFs), a financial product that has increased tremendously in popularity over the last two decades. Lettau and Madhavan (2018) provides a comprehensive overview of the functions of ETFs. ETFs are issued by so-called fund manufacturers such as Blackrock or State Street, and the creation and redemption process is administered by a market maker. In many cases, this market maker also takes an active role in the provision of intra-day liquidity, and in modern, high-speed markets, this role requires high frequency trading capabilities. In terms of total value, the share of ETFs of listed capital on U.S. markets rose from 2% in 2003 to 6% in 2014; the number of listed ETFs rose from about 270 to over 1,600. Over the same time span, the number of traded corporate securities fell by 18-20%; see Doidge, Karolyi, and Stulz (2017). What is striking, is that over the same time, ETFs account for a out-sized portion of trading: based on our data, in 2003-2006, ETFs accounted for only 12% of the daily value of trading whereas from 2010 to 2014 this fraction shifted to 27%; notably, 11% of the 15% increase stem from the pre-2006 ETFs. In other words, the modest increase in its share of ETF market value was accompanied by a drastic increase in ETF trading activity.

The question arises whether there are economic implications from the rise of this investment product. Bhattacharya and O'Hara (2018) develop a model with learning and information linkages with the ETF being a readily tradeable composite asset. They find that when ETF market makers cannot synchronize prices quickly, markets can become fragile when speculators herd due to cross-market information externalities. In Cong and Xu (2019)'s model, the composite nature of ETFs can cause greater price volatility and co-movements for underlying assets. Glosten, Nallareddy, and Zou (2021) provide empirical evidence on the informational efficiency impact of ETFs on its underlying securities using Russell 2000 index changes. Being in a heavily traded ETF (by virtue of being included in the underlying index), increases short-run informational efficiency for stocks with weak information environments. We follow Glosten, Nallareddy, and Zou (2021)'s methodology and study the impact of S&P500 index in- and exclusions on bot trading and through it on the fraction of a stock's return return that is explained by market returns.

A number of authors study the correlation among firms that are part of ETFs. For instance, Da and Shive (2018) argue, that speculators' activities to profit from arbitrage between ETFs and their underlying securities can lead to excess noise. Israeli, Lee, and Sridharan (2017) study the impact of the ETF ownership on underlying securities and they conclude that higher ETF ownership reduces the extent to which stock prices reflect firm-specific information. Notably, much of this literature studies the impact of the introduction of exchange traded products. Our causal analysis relies on events that have no bearing on the existence of ETFs — tradeable products for the S&P500, which we rely upon, have been around since the 1990s.

2. Data and Methodology

We use the monthly quote files for the TAQ database via WRDS from Jan 1, 2003 to Dec 31, 2014, the WRDS-generated daily "indicators" for daily trading measures (trades, quotes, spreads, etc.), CRSP for cross-sectional information for individual securities, and COMPUSTAT for index constituents.

2.1. Generation of the Variable of Interest

For each security and each day we compute the midquote of the end-of-second NBBOs, and from this data we compute the 5-second and 5-minute returns for all securities except for ETFs that have quotations in the TAQ database. We do not restrict attention to firms that exist for the entire sample. For the market return, we use the returns for ticker symbol IWM, which is iShares' exchange traded fund for the Russell 2000 index.[6]

To appreciate the extent of the data processing task: our panel has 21M stock-day observations, each 5-second return file has 4,680 observations per day, and returns files were created based on the TAQ quote files, which often exceed 100,000 observations per security and day (though some less frequently traded securities only have very few quote observations). Overall, the exercise involved the processing of several trillion data points.

In our analysis, we focus on two return horizons: 5-second and 5-minute returns. Entries in the monthly TAQ data are recorded at second granularity and taking time-stamp rounding into account, it makes little sense to use anything less than 2-second granularity. Furthermore, computer algorithms need time to process changes in the market portfolio in line with the market model, we therefore need to take a sufficiently coarse look. We recognize, however, that especially in the later years of the sample, computer algorithms may react at much faster speeds. The 5-minute granularity is a time horizon that is sufficiently fast such that a human can react to a piece of news with computer-aided fundamental analysis.

For each security and day, we then estimate the market model as

$$R_{it} = \alpha_i + \beta_i \cdot R_{\mathrm{M},t} + \epsilon_{it} \tag{1}$$

where R_{it} is the rate of return for security i in 5-second/5-minute interval t, $R_{\mathrm{M},t}$ is the contemporaneous return of the ETF IWM, our proxy for the market return. For each regression, we record the residual sum of squares, the R-squared R^2. We follow Roll (1988) and Morck, Yeung, and Yu (2013) in interpreting R^2s as the degree that a stock synchronizes with the market. These R^2s are our main variable of interest; in our regressions, we usually multiply them by factor 100. We also winsorize all variables in our sample (except for prices and market

[6]We also performed the entire analysis in the paper using the State Street exchange traded fund SPY which tracks the S&P500 index. The returns for these two ETFs are over 95% correlated and our results are qualitatively unchanged if we use SPY.

capitalization) at the 1% level to ensure that our findings aren't driven by outliers.[7] We also run the regression in (1) for daily returns, using CRSP closing prices and a 20-day rolling window.

There are several frictions that may impede a high R^2 in our regressions. First, the smallest price change per security is determined by the price level of the stock and the market proxies. For instance, suppose our proxy for the index trades at \$100. A 1 cent increase translates into a return of 1bps. The smallest change for a, say, \$10 is 10bps, and so the stock price may not react to small index changes, even if its fundamentals are 100% captured by the market index. Second, prices are changing rapidly and not entirely in lock-step, and it is possible that our intra-day snapshots do not fully capture asynchronous price movements. For all these reasons, we would expect that longer-term horizons have higher R^2s than shorter horizons — and that is indeed what we observe.

2.2. Bot trading

Our overall interest are trading strategies that involve some form of market monitoring and regular posting of quotes, akin to what a market maker would be doing. A common characteristic of such strategies is that they involve the submission of many orders. Another strategy is cross-asset and cross-market arbitrage which involves checking quotes on various venues regularly and taking advantage of any mis-priced quotes, both for individual securities as well as for portfolios. These strategies create activities on their own, and their presence also requires that market-making strategies are equally vigilant in monitoring prices to prevent stale quotes. These strategies are often associated with a particular type of trading firm, the so-called high-frequency traders. These firms use a large variety of other strategies, which are not of interest to our analysis. There are many other algorithmic, order-execution strategies, such as order-splitting algorithms for large orders, but these are not of interest for our analysis and they are not easily detectable in public data. Rather, our argument rests on the idea that there are "robots" that automatically, autonomously, and somewhat mechanically adjust quotes in response to market developments. In this paper, we will use the term bot trading to signify this type of robotic quote-submission and quote alteration behavior.

Panel A in Figure 2, which uses monthly averages of the daily measures, shows that there has been a significant increase in the number of quotes per security and day over our study period. In the early part of our sample, there were approximately 2,000-3,000 thousand quotes per security and day, towards the end of our sample, this number is closer to 60,000. Over the same time horizon, both transactions per day and dollar-volumes have also increased.

We use various measures as proxies for bot trading. The most basic one is the number of quotes, because electronic market making with bot trading requires the submission (and cancellation) of many quotes.[8] The second is the quote-to-trade ratio which loosely measures

[7]Our results are unchanged if we use the non-winsorized numbers.

[8]See, for instance, Malinova, Park, and Riordan (2013).

how many quotes are necessary for one trade to occur. One concern is that over time, order size declined as execution algorithms split large orders into ever-smaller limit order sizes so that this ratio increases simply by virtue of declining trade sizes. Hendershott, Jones, and Menkveld (2011) therefore introduced a third, related measure: the number of quotes per $100 traded which is impervious to smaller trade sizes. Panel B in Figure 2 shows, however, that there has been a substantial increase in both these two ratios, by about a factor of 3. Bot trading activities are associated with improvements in transaction costs, which are commonly measured by the bid-ask-spread. Panel C of Figure 2 plots the average monthly quoted spread for the 1,000 largest firms against the average number of daily quotes,[9] and it highlights the strong negative relationship between these two variables, i.e., more quotes (which are indicative of more high frequency market making) are associated with lower spreads.

A fourth measure that the literature attributes to the presence of high-frequency market makers (who engage in bot trading) is the fragmentation of orders. We proxy order fragmentation using the inverse of the Hirschman-Herfindahl Index of Market Concentration, as provided by WRDS data services. This measure, loosely, proxies for the the dominance of large orders as a fraction of all volume: the higher the number, the more large orders fragment into small orders during the day. Finally, tight bid-ask spreads are also associated with the presence of high-frequency market making.

In summary, we proxy for algorithmic trades with the (log of) the number of quotes, the quote-to-trade ratio, the Hendershott-Jones-Menkveld ratio (quotes per $100 of volume), and the order fragmentation index, and we also document the dollar-volume and the time-weighted bid-ask spreads. Table 2 shows that all these measures are all highly correlated.

2.3. Consideration for Special Events

Our sample covers several major developments in markets as well as times of great economic and political turmoil worldwide: our sample covers the 2008-9 financial crisis, the Eurozone debt crisis in 2011, the shutdown of the U.S. government in 2013, the downgrade of U.S. sovereign debt, and the 2010 Flash Crash. There were also major regulatory initiatives, such as the introduction of Reg NMS, the short-sale ban of stocks of financial firms during the financial crisis, and the S.E.C. ban on "naked" market access in November 2011.[10] We use several approaches in our empirical analysis to account for these events, such as by eliminating the episodes from the data as well as by using dummy variables to estimate differential effects. Since our results are consistent for all specifications, we leave all these episodes in the data.

[9]The plot shows a dotted line for the period around the financial crisis during which spreads spiked up for reasons other than bot trading.

[10]"Naked" access refers to brokers letting firms use their systems without the broker performing any control functions of the clients' activities.

3. Long Run Relationships

3.1. Overview

As a first step, we simply plot the monthly averages for the R^2 for the daily estimates of (1) for 5-second and 5-minute returns. Panel A in Figure 1 displays these plots and it demonstrates the substantial increase in R^2 over the sample horizon. Visually, there are (at least) three phases:

- 2003-end 2006: low and flat R^2;
- 2007-2009: strong increase in R^2;
- 2010-2014: higher levels of R^2, with some significant volatility.

The most significant change occurs between 2006 and 2008. This is a crucial period in the world of equity market structure because Regulation National Market System (Reg. NMS) came into effect. Among other things, Reg. NMS mandated the electronic linking of markets by virtue of the trade-through prohibition, which required that a trade on one market could not occur at a price worse then the best displayed quote on another, "protected" market.

Panel B in Figure 1 plots the daily R^2 (where we compute the latter using closing prices over 20-day rolling windows) as well as the ratio of the 5-second to the daily R^2. Although the R^2 for daily estimates increases over the same horizon, the increase in the ratio indicates that the relative increase in the intra-day measure is more pronounced. One may wonder if the increase in the R^2 stems merely from an increase in the estimate β. Panel C shows that, although there is an increase in the 5-second estimates in the first two years of the sample, there is no major change in βs when the R^2 surges.

Panel A in Figure 3 plots the R^2 against the number of quotes, Panel B plots the 5-second-to-daily ratio against the number of quotes. Panel A in particular indicates a striking co-movement of quotes and R^2s; Table 3 confirms the high correlation (0.56) for the monthly panel. The table also indicates significant correlations with other variables that capture bot trading, such as order fragmentation and the size of bid-ask spreads. The correlation with the quote-to-trade ratio as well as the HJM-automated quoting measure is much lower. As well, the correlation to the 5-second-to-daily ratio is lower, too.

In a regression analysis, we regress the R^2 on our various measures of interest using

$$\mathit{DV}_{it} = \beta_1 \times \mathit{EV}_{it} + \sum_{j=1}^{3} \beta_{j+1} \mathit{controls}_{jit} + \delta_i + \epsilon_t, \tag{2}$$

where DV_{it} is the average month t of the 5-second/5-minute/daily/ratio R^2 for security i; EV_{it} is the bot trading related variable of interest, where we use quotes, trades, and dollar-volume (and their respective logs), the quote-to-trade ratio, the Hendershott-Jones-Menkveld ratio (quotes per \$100 of volume), bid-ask spreads, and the order fragmentation index; $\mathit{controls}_{jit}$ are the monthly average volatility index VIX (which does not vary by security i, as well as the security's average monthly log-closing price and market capitalization; δ_i are security fixed effects. In untabulated regressions we also included year and industry fixed effects; the results are robust. We include price and market cap as controls as these two variables are typically associated with trading costs and levels of trading activity; as we argued before, the price level in particular may affect the extent to which a stock's returns can react to movements in the market. Standard errors are double-clustered by month and security to control for cross-sectional and time-series correlations.

Results. Table 4 contains our estimation results; the table is compressed and displays only the estimates for the variables of interest; each cell in the rows for "full sample" is the result of a single regression. Overall there are about 863,000 month-security observations and the R^2 of the regressions ranges from .54 to .76. The estimates confirm the observations from the basic correlation tables, namely, that intra-day R^2s co-move strongly and significantly with movements in bot trading-related variables. More trades, more dollar-volume, and more quotes are positively associated with higher relation of returns to the market. Similarly for the quote to trade ratio and the HJM-algo measure. For spreads and order fragmentation, the relationship is negative. For spreads, the relationship is as expected, because lower spreads are commonly attributed to more bot trading. We note that the HJM algo measure is only significant for intra-day R^2s, not for the daily one. We also note that quotes per se can have no direct effect on the daily measure and that the positive estimate is likely a spurious correlation.

To further tease out the effects of the different phases, we run a regression where we split the sample into the three phases using dummies for the respective time horizons. Specifically, we estimated the following regression equation

$$\mathit{DV}_{it} = \sum_{j=1}^{3} \beta_j \cdot \mathit{EV}_{it} \times \mathit{Phase}\ j_t + \sum_{j=1}^{3} \beta_{j+3} \cdot \mathit{controls}_{jit} + \sum_{j=1}^{2} \alpha_j \cdot \mathit{Phase}\ j_t + \delta_i + \epsilon_t, \quad (3)$$

where *Phase* 1_t is a dummy that is 1 if t is in 2003-2006 and 0 otherwise, *Phase* 2_t is a dummy that is 1 if t is in 2007-2009 and 0 otherwise, and *Phase* 3_t is a dummy that is 1 if t is in 2010-2014 and 0 otherwise.

Results. Table 4 contains our estimation results, where the lines with 2003-2006, 2007-2009, and 2010-2014 together are a single regression. We generally observe that for the 5-second R^2, estimated coefficients are larger for the 2007-2009 and 2010-2014 time frames, and that early estimates are smaller and sometimes of opposing sign (e.g., for fragmentation). As for the full sample, the HJM measure has low explanatory power.

4. Causal Impact of Bot Trading

4.1. Background

For the time horizon of our study, there were many concurrent events. Most prominently, the 2008-9 financial led to dramatic movements on markets, and it is possible that investors chose to re-assess their position towards market risk, which would contribute to the increase in the R^2. It is plausible that the increase in the R^2 for daily estimations is partly driven by changes in investor attitude towards systematic risk.

The upward shift in the R^2 also coincided with the increasing proliferation of exchange traded funds. In terms of total value, the number of listed ETFs rose from about 270 to over 1,600 and their share of listed capital, based on CRSP data, in U.S. markets rose from 2% in 2003 to 6% in 2014. Over the same time span, the number of traded corporate securities fell by 18-20%. What is striking, is that over the same time, ETFs accounted for a substantially larger fraction of trading: in 2003-2006, ETFs accounted for only 12% of the daily value of trading whereas from 2010 to 2014 this fraction shifted to 27%; notably, 11% of the 15% increase stem from the pre-2006 ETFs.[11] In other words, the modest increase in its share of ETF market value was accompanied by a drastic increase in ETF trading activity.

A consequence of the proliferation of ETFs listings is that during the 2003-2014 time period investors increasingly invested in index products. It is thus imaginable that this shift caused markets to align more with systematic risk. We believe, however, that this argument is weak: the assets under management of the ETF sector have been increasing steadily and not in shifts. Instead, the main shift here, too, relates to automated trading: ETF market makers (which, again, are computer algorithms, not people) have an incentive to ensure that prices of their fund and the underlying securities are aligned, and they therefore use dynamic strategies to eliminate pricing errors by either adjusting quotes across securities or by trading against mis-priced, "stale" orders. Moreover, ETF market makers have an incentive to align prices irrespective of its assets under management or level of trading activity: even if an ETF trades only once per day, an automated ETF market maker would still adjust quotes continuously. In other words, here, too, bot trading is the root cause for the shift, not investor preferences.

Finally, since 2003, many firms have left public markets, either through mergers and acquisitions or delistings, as discussed in Doidge, Karolyi, and Stulz (2017). One can argue, of course, that the remaining firms have become more diversified, but this development would affect the regression coefficients, not the explained fit in the R^2.

[11] This trading activity is highly concentrated: the top 31 ETFs account for 10.8% of the 12%, and the top 10 funds account for 9.9%.

4.2. Overview

In this section, we attempt to establish the causal relationship between bot trading and the relationship of stock returns to market returns. Our identification strategy is straightforward: we look at changes to the S&P 500 constituents. When a stock enters or drops out of an index multiple things happen. In our view the most critical ones are that investment funds change their holdings, and that high frequency traders that arbitrage on the relationship of index constituents to traded index products (such as futures or ETFs) increase or reduce their activities. For entry events, both the activities of funds and automated traders “go in the same direction”, whereas for exits, they go in opposing directions, at least temporarily. Crucially, none of these activities affect to the fundamentals of a security which should ultimately drive how much of intra-day stock return variance is explain by market returns. As S&P index only includes firms that are large, highly liquid and have positive earnings for four consecutive quarters, we cautiously interpret our causality identification as local average treatment effects for firms that satisfy the above criteria. However, we are optimistic about the external validity of our identification results, as bots do not only watch stocks with large sizes and high liquidity. In general any firm that catches (drops) the bots attention could be subject to the same synchronous causal effect of a rise (drop) in the intra-day R^2.

We follow two approaches: the first is a standard regression approach where we identify all entry and exit events, find a matched firm for each affected firm, and then study whether subsequent to the event there have been differences in the key variables. The second approach, which we present in detail in the next section, uses a new empirical tool from the machine learning literature, so called causal random forests.

4.3. Events and Matching

We obtain changes of index constituents for the S&P500 index from the COMPUSTAT database and we match these back to the data that we created from TAQ. Overall, we find 419 applicable events in our sample. For each affected security, we find a matching security using average measures for the month prior to the index change. Common practice in the microstructure literature is to match based on stock price, market capitalization, and possibly some trading characteristics; see Davies and Kim (2009). In the case of our analysis, there are some concerns. First, firms that enter or exit the S&P500 are usually around the threshold of being one of the largest 500 publicly traded firms in the U.S. Investment funds may attempt to predict changes in index composition and trade in the firms that are close to the market capitalization threshold, and these activities around index change events could affect both treatment and control firms. Matching on trading characteristics is equally tricky because good matches prior to an event may not be good matches after the event. In the end, we performed the analysis using a variety of techniques so as to establish that the results do not hinge upon a particular choice of matching variables. We will therefore first present the results of our main estimation for a variety of matching techniques to highlight that the findings vary little with the matching approach. We then narrow our focus and use the standard approach where we match based on log-closing prices and log-market capitalizations.

In our matching, we employ *nearest-neighbor matching* in the sense that for each index-changing security i we find the security j that minimizes the scaled matching error as follows:

$$matcherror_{ij} = \sum_{k=1}^{M} \left(\frac{C_k^i - C_k^j}{C_k^i + C_k^j} \right)^2 , \tag{4}$$

where C_k is one of the above-mentioned matching characteristics, e.g., for corporate equities: log of firm size and log of price. On days with multiple changes we match with replacement, and we exclude all 419 index-changing securities from the sample of possible matches.

We then construct a panel of trading characteristics for the 20 trading days before and after (including the event date) the change. The basis of our statistical approach is a conventional difference-in-differences analysis of this panel dataset, and we following the methodology in Hendershott, Jones, and Menkveld (2011) and Malinova and Park (2015). The dependent variable ΔDV_{it} is the value of the daily realizations of the 5-second and 5-minute R^2 as well as the bot trading related variables for the "treated" (i.e., affected) security i at time t less the value for the matched security. Using this dependent variable, we will estimate the following regression:

$$\Delta DV_{it} = \alpha \cdot \mathit{change}_t + \mathit{controls}_{it} + \delta_i + \epsilon_{it}, \tag{5}$$

where change_t is an indicator variable set to 1 on the index change start date, $\mathit{controls}_t$ are time series controls such as the VIX, which controls for the level of market-wide volatility, as well as panel controls such as the daily price and market capitalization; δ_i are entry-exit event fixed effects. The coefficient of interest α captures the effect of the index inclusion/exclusion for treated securities, and we run the regressions separately for exit and entry events.

We conduct inference in all regressions using double-clustered Cameron, Gelbach, and Miller (2011) standard errors, which are robust to cross-sectional correlation and idiosyncratic time-series persistence.[12]

Results. We begin with the visual examination of the data, where we focus on the 5-second R^2 and the daily quotes for the case where we match by price and market capitalization. Panels A and B in Figure 4 plot the number of quotes for entry and exit events respectively for treatments and their matches, Panels C and D plot the R^2s. It is straightforward to see that the quoting activities increase after index inclusions, and the R^2 concurrently increase markedly for entry firms. For index exclusions, the situation is more difficult. Namely, quotes first drop notably, but then they pick up again to rise to pre-change levels. The same, holds for R^2s, but the R^2 does not reach the pre-exclusion levels. Overall, we interpret these graphs as evidence

[12]Cameron, Gelbach, and Miller (2011) and Thompson (2011) developed the double-clustering approach simultaneously. See also Petersen (2009) for a detailed discussion of (double-)clustering techniques.

that the effects work as we predict.

Table 6 contains our estimation results, where we present the results for entry and exit events for various versions of the matched samples and where we display only the estimate for α (in other words, each entry is for a single regression). The estimations confirm the observations from the plots. Following the index inclusion, the intra-day 5-second and 5-minute R^2s significantly increases, by 2 and 4.5 percentage points respectively. Alongside this increase, we observe a significant increase in quotes, as well as trades and overall dollar-volume. Of course, an index change usually triggers an increase in volume because index funds need to adjust their holdings, and in the short run, this may lead to some excess volume. It is therefore not surprising that this increase is so strong, that the quote to trade ratio and the HJM algo measure of bot trading decline. Consistent with higher bot trading and quoting activity, we also find that spreads decline and that order fragmentation increases. Since an index inclusion event has no bearing on a firm's fundamental, and since there are strong indications that bot trading increased because of the index inclusion, we attribute the increase in the R^2 to an increase in algorithmic quoting activities.

We note, however, that the daily R^2 for index inclusion events declines. This decline does not affect our conclusions regarding the impact of bot trading, but, at first blush, it is surprising. The decline is, however, consistent with Baruch and Zhang (2021)'s theoretical model. They predict that if there is a shift of investors from being non-indexers to indexers, then the statistical fit (measured by the R^2) of the CAPM regression would decrease.

Our results for index deletions are weaker. We do observe a significant decline in the intra-day R^2s, but the evidence on changes in quoting activities is weak, and for some matching configurations, the number of quote updates actually increases. For our preferred matching approach (price and market capitalization), we find an increase in dollar-volume traded. For many matching configurations, the quote to trade ratios and the HJM algo measure decline. The concern with all these measures is that they are potentially related to trading activity, many funds need to adjust their holdings after and index deletion. Therefore, heightened activities by these funds may give the impression that there is not decline in bot trading. Once exception is the order fragmentation, which declines in most specification, consistent with the view that bots stop quoting in excluded stocks.

Overall, we conclude that index deletions and inclusions lead to drops and increases respectively in the type of market making activity that allows the alignment of prices and that drives the R^2.

Similarly to the full panel regressions, we also want to assess whether there are changes in the estimated effects across time. We therefore ran a regression specification of (5) that resembles (3):

$$\Delta DV_{it} = \sum_{j=1}^{3} \beta_j \cdot \textit{change}_t \times \textit{Phase } j_t + \textit{controls}_{it} + \sum_{j=1}^{2} \alpha_j \times \textit{Phase } j_t + \delta_i + \epsilon_{it}. \qquad (6)$$

Results. Table 7 contains our estimation results, where we present our findings only for the case where we match treatment and control by price and market capitalization. Compared to Table 6, the table is compressed and only displays the estimates for β_1, β_2, and β_3. We note that the change in the 5-second intra-day R^2 is significantly larger in the second and third phases (post 2007) compared to the early sample (2.4% to 3% as opposed to 0.3%); likewise, changes in quoting activity only occur in the later part of the sample, consistent with our long-run observations of overall market activity.

Although not all estimates point in the same direction, overall we believe that these findings support the notion that bot trading has a significant impact on the portion of the variance of intra-day stock returns that is explained by market returns.

4.4. Instrumental Variable Regression

Our working hypothesis is that an index inclusion or exclusion does not affect the relation of a stock's return with the market directly and that instead such an event affects algorithmic/high frequency trading directly. In this case, such an event is an instrument that is correlated with high frequency traders' activities and uncorrelated, directly, with the return correlations. There are some caveats. First, the inclusion event may affect the R^2 directly or, rather, through other channels than those captured by our proxies for high frequency trading. Moreover, these proxies may reflect these traders' activities only imperfectly. For these reasons, we will interpret the findings of this part of our analysis only with caution. We perform our regression analysis using the inclusion/exclusion events as a binary instrument for bot trading in a two-stage least square instrumental variable regression.

$$\begin{aligned} QA_{it} &= \beta_1 event_t + \beta_2 VIX_t + controls_{it} + \delta_i + \epsilon_{it} \\ DV_{it} &= \alpha_1 \widehat{QA_{it}} + \alpha_2 VIX_t + controls_{it} + \delta_i + \epsilon_{it}, \end{aligned} \tag{7}$$

where the QA_{it}, are the respective proxies for bot trading that we instrument by its estimated value from the first stage regression, $\widehat{QA}_{it}$ (as in the previous analysis, we are using the difference of the value of the treatment and the control). Variable $event_t$ is 0 before an index inclusion/exclusion and 1 thereafter. Estimate $\hat{\beta}_1$ is the average effect on quoting activity after the index event. As before, DV_{it} is the difference of the R^2 for treatment and control (at the 5-second and 5-minute horizon); VIX_t is the daily realization of the U.S. volatility index VIX; $controls_{it}$ are log price and log marketcap for the treated firm; and δ_i are firm fixed effects. Estimate $\hat{\alpha}_1$ measures the impact of a 1-unit increase in the bot trading activity proxy on the dependent variable. Based on the analysis in the subsequent subsection, we choose algorithmic activity proxies for which there is an effect of the event, namely, for entry events, we use the logarithm of the number of quotes and the level of order fragmentation, for exit events we use only order fragmentation (because there is no evidence for a change in the logarithm of the number of quotes).

Results. We omit the first stage results because these are equivalent to those covered in Table 6. We include diagnostics tests for under or weak identification and note that these raise no concerns. Table 8 reports the second stage results of the instrumental variable analysis. This table is consistent with our prior analysis and indicates that an increase (decrease) in bot trading causally increases (decreases) the alignment of stock and market returns as measured by the R^2s of the intro-day regressions.

4.5. Mediation Analysis

In the discussion of our instrumental variable analysis from the previous subsection we recognize that an index inclusion and exclusion event is not a perfect instrument because the R^2, our measure of interest, might be affected by this event through channels other than bot trading. Indeed, one possibility is that our proxies themselves are imperfect and capture only a portion of the effect of the event on bot trading (so that the "direct" effect stems from the "uncaptured" effect on bot trading). A weaker form to assess the effect is a mediation analysis as established by Baron and Kenny (1986). We use the method developed by Imai, Keele, and Yamamoto (2010) for causal mediation analysis. The analysis requires the estimation of a structural model of equations with

$$\begin{aligned} QA_{it} &= \beta_1 event_t + \beta_2 VIX_t + controls_{it} + \epsilon_{it} \\ DV_{it} &= \alpha_1 event_t + \alpha_2 QA_{it} + \alpha_3 VIX_t + controls_{it} + \epsilon_{it}, \end{aligned} \tag{8}$$

where all variables are as before. We cluster standard errors by index event.

Results. We do not tabulate the direct estimations of (8) because the estimates mimic the findings of the analysis thus far. Table 9 contains the results for (8); the signs of the effects are consistent with the analysis thus far. A key set of estimates of interest in mediation analysis are the Average Causal Mediation Effect (ACME), the direct effect (of the index event), and the indirect (mediated) effect. Table 10 contains the values for the two bot trading proxies that we used in the instrumental variable analysis: the logarithm of the number of quotes (for inclusion events only) and the level of order fragmentation (for inclusion and exclusion events). The findings are consistent with our results so far, namely, that the change in bot trading, triggered by the index events, lead to a increase/decrease in the respective R^2s. As with the previous analysis, the estimated effect for index exclusions is not significant, but the effect for inclusions is.

5. Causal and Instrumental Random Forests

In additional to the standard econometric methods used in Section 4., we also re-examine our results from Table 6 using so-called instrumental random forests, a tool from the nascent literature on machine learning in finance. Although the findings are very clear for index inclusion events (a jump in R^2 and an increase in bot trading proxied, for instance, by the number of quotes), there are methodological challenges in our econometric analysis. One concern is the matched sample approach because it is difficult to exclude the possibility that the control securities themselves have been affected by the index event. The random forest approach is a possible way around this problem. The tool has features of a classification algorithm and therefore implicitly performs its analysis of the treatment effect (i.e., the index inclusion event) by comparing across "similar" securities. In this section, we will briefly outline the tool and present our findings. We include the technical details in the Internet Appendix. [13]

5.1. What is a Causal Forest?

A random forest is an estimation and classification tool that is commonly used in the machine learning literature, introduced first by Breiman (2001). The basic idea is to build an ensemble of decision trees to predict a variable of interest by the rule of majority voting.

We first explain the idea of a decision tree using the example of predicting the price for a house based on various features such as location, number of levels, bathrooms, fireplaces, quality of the built and finishes, and so on. A linear regression that uses all these covariates as explanatory variables may ignore the importance of interaction terms, as well as non-linear higher-order relationships between covariates. A misspecified functional form in a linear regression model might result in biased coefficients as well as non-robust standard errors. The basis of a random forest is the construction of a series of decision trees, where using the latter allows us to relax the linear functional form assumption and to systematically identify interaction terms. One way to think about a decision tree is that a tree is a systematic way to run through the various variables as a series of questions (e.g., "was the house built before 1950", "is it in the suburbs", "does it have more than one level", "does it have a garage", "does it have more than 3 bedrooms", etc.) The goal then is to form a prediction for the appropriate price (range) ("given the answers to the series of questions, our prediction is a price between \$1.5M and \$1.7M").

A random forest is a systematic way to build a decision tree (or "split" of the data) that minimizes the mean-squared errors. The intuition that in building the forest, the algorithm asks different questions and in different orders, and then to assess the quality of the predictions using a training data set for cross-validation. For this procedure, one repeatedly divides the data into three subsets: a training set, a cross-validation set, and a hold-out set. The training set is further divided into subsets from which we build trees and we then evaluate the estimates using the cross-validation set. Once we identify the optimal trees, we apply the forest to the held-out test set. The cross-validation encourages the external validity of a non-linear random forests model by awarding models that also have low mean-squared errors out-of-sample.

[13] The Internet Appendix is available online at: `https://ginward.github.io/CAPM_appendix.pdf`

5.2. The Purpose of the Causal Forest Methodology

The heart of causal effect estimation is the estimation of an effect relative to an unobservable counterfactual. In our case, we like to assess the impact of the presence of high-frequency bot trading on the R^2 of the stock's return. In an ideal world, we would be able to estimate the R^2 under the exact same market conditions, information releases etc. once with and once without bot trading, but the financial market is not a laboratory environment where we can set up such a perfect experiment. Instead, we look for a situation where there is a shock to bot trading and then estimate the before-to-after effect relative to a security that did not experience a shock but that is otherwise very similar. The matched, control security provides us with the presumed counterfactual, and the index inclusion/exclusion provides us with the shock.

Even though this kind of setup, in particular index changes, has often been used in the finance literature, there are some significant conceptual concerns. If the securities that are added to or excluded from ETFs are randomly (unpredictably) selected, then the dependent variable will be independent from the characteristics of the securities and we can assume it to be an exogenous, unpredictable shock. However, the stocks that are added or excluded from ETFs are not random selected, but there is a reason for a switch, most commonly because the stock experienced a prolonged drop in market capitalization (or other firms became bigger). Therefore, the estimator is no longer unbiased.

The traditional literature has two ways to resolve this bias. One way is the method of nearest-neighbor matching based on several hand-picked characteristics, as discussed in Section 4.3.4. However, it is often not clear why some characteristics are selected for matching while others are not. Another common approach is propensity score matching. However, propensity matching requires strong parametric assumptions on the propensity scores that can be hard to justify.

Causal forests alleviate some of the above two issues: first, causal forests systematically select characteristics to match securities based on the objective of maximizing the heterogeneity of treatment effects; second, causal forests do not impose any parametric assumptions on the propensity scores. We outline the details in the Internet Appendix.

5.3. Presentation of Estimation Results

One challenge with using machine learning tools is that the model outputs often do not lend themselves to an explanation that is on par with that of an OLS regression. We present our findings using two approaches. The first is by listing the so-called "variable importance." Causal forests partition the data by permuting all explanatory variables and selecting the one that results in the strongest drop in mean-squared errors. The specific procedure in a tree is to "split" the data along a variable, and the "variable importance" is essentially the frequency of splits for a variable. The second approach is that of the "best-pruned tree" as described in Wager (2019b). This approach selects the most representative tree from the forest with the minimum post-pruned mean-squared error. This best pruned tree provides the reader with an idea of the decision making process and it informs the reader of the (possibly non-linear) rules that affect the treatment effects.

5.4. Causal Random Forest Estimation of the Change in R^2 and bot trading

We construct a difference-in-difference estimator by employing the above described "causal forest method." Our task is to assess the effect of the treatment on the the dependent variables DV_i, the R^2 and the bot trading measures. In contrast to the panel estimation of the preceding section, here the dependent variables are computed as the difference of the 20-day average before to after the event.

We apply a causal forest to time-differenced data to obtain individual specific i.i.d. observations, and we cluster our standard errors by index events. When we estimate the causal impact on R^2, for 5-second, 5-minute and daily, we use one of the following variables as the dependent variable: $\Delta(R^2$ 5sec), $\Delta(R^2$ 5min), and $\Delta(R^2$ daily). When we estimate the causal impact on bot trading proxies, we use one of the following variables as the dependent variable: Δ(Number of Quotes) [14] or Δ(Fragmentation). As covariates we use the the 20-day average before the event of the following variables: Price, Market Cap, $\$-$Volume, qspread cents, Number of Analysts in the Preceding Quarter Intra-day Volatility and Year. We use both binary and continuous treatment variables. As binary treatment variables we use dummy variables that indicate if a security was included or excluded from the S&P500 index. As continuous treatment variables we use the difference of the 20-day average before to after the event of the bot trading proxies: Δ(Number of Quotes) or Δ(Fragmentation).

Results. Our causal forests estimation results support our findings from Tables 6 and 7. Table 11 shows the effect of index inclusion events and index exclusion events on the two main bot trading proxies, the number of quotes and fragmentation, and then their effect on R^2. Our main estimators of interest are the average treatment effects for the treated securities as well as the average treatment effects for the overlap-weighted securities. We observe a significant positive effect of inclusion events on the intra-day R^2 (5-second and 5-minute), and a significant negative effect of exclusion events on the intra-day R^2 (5-second and 5-minute). We also observe a positive and significant effect of the number of quotes on both intra-day and daily R^2s. Likewise, the effect of liquidity fragmentation on the intra-day R^2s is positive and significant. We also observe a negative and significant effect of exclusion events on daily R^2 and a negative and significant effect of fragmentation on the daily R^2. These findings are consistent with our observations from the panel estimation. As in our OLS approach, the number of quotes and fragmentation may be endogenous, and for now the effect of number of quotes and fragmentation R^2 captures only correlation and not causality. However, we expand the analysis in the next subsection and use an *Instrumental Forest* to address the endogeneity

[14] The number of quotes is scaled down by 10,000 when applied in Causal Forests.

of the algorithmic trading proxy.

Table 12 shows the effect of index inclusion and exclusion on bot trading proxies: fragmentation and number of quotes. Our main estimators of interest are the average treatment effects for the treated securities as well as the average treatment effects for the overlap-weighted securities. We observe that the inclusion events have a positive and significant effect on both bot trading proxies. As with the OLS approach, we do not observe a significant effect of index exclusion events.

5.5. Explanatory Causal Forests

Tables 14 and 15 show the variable importance in the causal impact estimation. The measure counts the frequency of splits on each of the variables in causal forests at each depth and the variable with the highest frequency count is the variable with the highest influence on the causal effect. We restrict the maximum depth when calculating the variable importance measure to be four, as it is likely that earlier splits have higher variable importance than later splits that are deeper in the tree. Unlike traditional machine learning, when making tree splits, we do not observe the true treatment effects, which are required to construct the true loss function. Therefore, we follow Athey, Tibshirani, and Wager (2019) and construct a proxy loss function that maximises the heterogeneity of treatment effects in various leaves of random forests. Athey, Tibshirani, and Wager (2019) have shown that, maximizing treatment effect heterogeneity is equivalent to minimizing the true loss function. Consequently, the variable importance measure shows us the importance ranking of variables that determine the treatment effects heterogeneity. For example, the forest will split on market capitalization if the effect of bots on R^2 is very different across firms with differnet market capitalization.

The purpose of understanding the variable importance of causal forests is two fold. First, it allows us to obtain insights as to how the propensity score of the treatment and can intuitively be used to hand-pick the best matching variable (our approach in Section 4.3. instead was based on common practice in the literature). Second, the methods allows treatment effects to be heterogeneous across securities conditional on model features. Understanding the most representative tree in the forest provides insights on how the impact varies with different interactions of model features. However, one needs to be cautious not to over-interpret the displayed best-pruned tree because it is merely a single statistical representation of the method and treatment effects estimates in small leaves of the tree can be unstable (see Wager (2019a)).

Results. When we use index inclusion as treatments and the R^2s as the dependent variable, the most important covariates are Price and $\$-$volume. When we use index inclusion as treatment and Δ(Fragmentation) or Δ(Numberof Quotes) as the dependent variable, the most important covariates are also Price and $\$-$volume. The non-linear matching criteria by causal forest complements our matching criteria by price and market capitalization in Section 4.4.3., which uses both price and market capitalization for matching. The causal forest also suggests that the heterogeneity of the impacts on R^2 can be largely explained by price and $\$-$volume. A caveat of the variable importance measure is that it is not weighted by the depth of the trees. Usually, the earlier a variable appears in splitting the data, the more important it is in predicting the causal effects.

The best pruned tree provides more insights as to at which level a variable is commonly used in terms of splitting the data. Figure 5 shows the best-pruned tree when we use index inclusion as treatment and 5-minute R^2 as the dependent variable. We omit the best-pruned tree plots for 5-second and daily R^2 to save space.

Figure 5 shows that the best tree chooses to split on $\$-$volume twice before splitting on any other variables, and then it splits on price and repeatedly on market cap. Indeed, average $-volume is the most important criterion when making accurate predictions of the change in R^2 in index inclusion events — if it is sufficiently large, no other covariates have no impact on predicting the effect of an inclusion events on R^2. Lower in the tree, the data also splits along the Number of Analysts in the Preceding Quarter and by time horizon (before/after 2005). This latter observation is consistent with our earlier observation that the most important change in the R^2 occurs between 2006 and 2009. At the lowest displayed level, volatility matters when determining the impact on R^2.

The best-pruned trees in index exclusion events are much shallower than the best-pruned trees in the index inclusion events and we omit them from this version of the paper. An implication from the shallower tree is that there is less heterogeneity in index exclusion events compared to index inclusion events.

5.6. Instrumental Random Forests

Our random forest analysis thus far has focused on the treatment effect of index changes on variables of interest similarly to our OLS approach in Section 4.4.3. In the final part of our analysis, we use Instrumental Random Forests to establish causality. The procedure here is akin to the instrumental variable estimation in Section 4.4.4. and we outline the technical details in the Internet Appendix. As in the standard IV estimation, we use index inclusion or index exclusion as the instrumental variables and bot trading proxies as treatment variables (instrumented variables).

Results. Table 16 and Table 17 show the variable importance measure of the variables in the instrumental forest estimation. When we use the index inclusion as the instrument, we find the most important covariates are Price, $\$-$volume, Number of Analysts in the Preceding Quarter and qspread cents. This complements our analysis with the "preferred" matching criteria in Section 4.4.4., where we match by price and market capitalization.

Figures 6 and 7 show the best-pruned trees when we use the change in 5-minute R^2 in inclusion events as the dependent variable and Δ(Fragmentation) or Δ(Number of Quotes) as the treatment, respectively. We omit other best-pruned trees to save space.

Figure 6 shows that when Δ(Fragmentation) is used as the treatment, the best tree chooses to split on $\$-$volume first. As before, this suggests that $\$-$volume is the most important criterion when making accurate predictions of the impact of the change in Fragmentation on R^2 in the inclusion events due to bot trading. We also observe that there is a region for $\$-$volume such that other covariates don't matter much in predicting the causal effects of inclusion events on the change in R^2 via bot trading proxies if the value for $\$-$volume is outside the region. If $\$-$volume is in the specific range, the Price matters in predicting the impact of Fragmentation on R^2 in the inclusion events due to bot trading. High priced securities usually require larger capital commitments for market making activities and it is therefore likely that high frequency quoting for such securities is less prevalent.

The best tree further splits on the years 2003, 2005, and 2011. This confirms our earlier observation that there are (at least) three distinct phases of the substantial increase in R^2 over the sample horizon, with 2003-end 2006 with low and flat R^2; 2007-2009 with strong increase in R^2; and 2010-2014 with higher levels of R^2 a strong fluctuations.

Figure 7 shows that, when Δ(Number of Quotes) is used as the treatment, the best tree choose to split on the Number of Analysts in the Preceding Quarter before splitting on any other variables. This suggests that the Number of Analysts in the Preceding Quarter is an most important criterion when making accurate predictions of the impact of the change in \$-volume on R^2 in the inclusion events. The Number of Analysts in the Preceding Quarter is a proxy for the amount of attention a security gets in the market. We observe that there is a range for which the Number of Analysts in the Preceding Quarter is such that other covariates do not affect the heterogeneity of algorithmic trading's impact on R^2. Here, too, the best tree splits on the years 2003, 2005, and 2011.

Most importantly, we observe a positive and significant effect of both the of our high frequency bot trading proxies (the number of quotes and fragmentation) on the 5-second R^2 and daily R^2 for the inclusion events. For the 5-minute events, we observe a positive and significant effect for the number of quotes, but not fragmentation. These findings are similar to our results from the OLS IV estimation.

6. Discussion and Conclusion

Our paper makes several contributions. First, we document the significant increase in the fraction of intra-day stock return variations related to market-wide fluctuations following the rise in bot trading in U.S. markets. Much of the market microstructure literature studies changes in price efficiency and price discovery on a stock-by-stock basis. We propose a way to expand microstructure research to the relationship of returns of stocks and the market as a whole. The latter, arguably, is a core component of asset pricing research (but we emphasize that our paper is not intended as an asset-pricing study).

Second, using changes to index compositions, we provide evidence of a causal relationship between the activities of bot trading and the of synchronization of stock returns across assets. Arguably, index membership is arbitrary, occurs only on paper, has no impact on the operation of a firm, and, therefore, firm-specific return synchronization should not be affected by index inclusion or exclusion. As firms get included in an index, they enter the radar of high frequency traders who closely monitor prices in major securities on an ongoing basis. Their monitoring manifests itself in quoting activities that, in turn, cause individual stocks' returns to synchronize more closely with the market. Although our work does not fall into standard asset pricing, we believe that we identify an important channel for the impact of market microstructure changes (through the advent of bot trading) on asset pricing.

Third, we provide a methodological innovation by applying new tools from the machine learning literature that help lend further credibility to our analysis and that provide a roadmap for future applications.

References

Athey, Susan, and Guido Imbens, 2016, Recursive partitioning for heterogeneous causal effects, *Proceedings of the National Academy of Sciences* 113, 7353–7360.

Athey, Susan, Julie Tibshirani, and Stefan Wager, 2019, Generalized random forests, *The Annals of Statistics* 47, 1148–1178.

Baron, Reubeti M., and David A. Kenny, 1986, Moderator-mediator variables distinction in social psychological research: Conceptual, strategic, and statistical considerations, *Journal of Personality and Social Psychology* 51, 1173–82.

Bartram, Söhnke M, Gregory Brown, and René M Stulz, 2012, Why are us stocks more volatile?, *The Journal of Finance* 67, 1329–1370.

Bartram, Söhnke M, Gregory W Brown, and René M Stulz, 2019, Why is there a secular decline in idiosyncratic risk in the 2000s?, *Fisher College of Business Working Paper* p. 019.

Baruch, Shmuel, and Xiaodi Zhang, 2021, The distortion in prices due to passive investing, *Management Science* Articles in Advance, 1–16.

Bhattacharya, Ayan, and Maureen O'Hara, 2018, Can ETFs increase market fragility? effect of information linkages in ETF markets, working paper Cornell University https://ssrn.com/abstract=2740699.

Breiman, Leo, 2001, Random Forests, *Machine Learning* 45, 5–32.

Brown, Gregory, and Nishad Kapadia, 2007, Firm-specific risk and equity market development, *Journal of Financial Economics* 84, 358–388.

Cai, Jie, 2007, What's in the news? Information content of S&P 500 additions, *Financial Management* 36, 113–124.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, 2011, Robust inference with multiway clustering, *Journal of Business & Economic Statistics* 29, 238–249.

Chen, Honghui, Gregory Noronha, and Vijay Singal, 2004, The price response to S&P 500 index additions and deletions: Evidence of asymmetry and a new explanation, *The Journal of Finance* 59, 1901–1930.

Cong, Lin William, and Douglas Xu, 2019, Rise of factor investing: Asset prices, informational efficiency, and security design, working paper Cornell University https://ssrn.com/abstract_id=2800590.

Da, Zhi, and Sophie Shive, 2018, Exchange traded funds and asset return correlations, *European Financial Management* 24, 136–168.

Davies, Ryan J., and Sang Soo Kim, 2009, Using matched samples to test for differences in trade execution costs, *Journal of Financial Markets* 12, 173 – 202.

Denis, Diane K, John J McConnell, Alexei V Ovtchinnikov, and Yun Yu, 2003, S&P 500 index additions and earnings expectations, *The Journal of Finance* 58, 1821–1840.

Doidge, Craig, Andrew Karolyi, and Rene Stulz, 2017, The U.S. listing gap, *Journal of Financial Economics* 123, 464–487.

Elliott, William B, Bonnie F Van Ness, Mark D Walker, and Richard S Warr, 2006, What drives the S&P 500 inclusion effect? An analytical survey, *Financial Management* 35, 31–48.

Fama, Eugene F, and Kenneth R French, 2004, New lists: Fundamentals and survival rates, *Journal of Financial Economics* 73, 229–269.

Glosten, Lawrence, Suresh Nallareddy, and Yuan Zou, 2021, ETF activity and informational efficiency of underlying securities, *Management Science* 67, 22–47.

Grégoire, Vincent, and Charles Martineau, 2022, How is earnings news transmitted to stock prices?, *Journal of Accounting Research* 60, 261–297.

Hendershott, Terrence, Charles Jones, and Albert Menkveld, 2011, Does algorithmic trading improve liquidity?, *The Journal of Finance* 66, 1–33.

Hrazdil, Karel, and Thomas Scott, 2009, S&P 500 index revisited: Do index inclusion announcements convey information about firms' future performance?, *Quarterly Journal of Finance and Accounting* 48, 79–113.

Imai, Kosuke, Luke Keele, and Teppei Yamamoto, 2010, Identification, inference, and sensitivity analysis for causal mediation effects, *Statistical Sciences* 25, 51–71.

Israeli, Doron, Charles M.C. Lee, and Suhas A. Sridharan, 2017, Is there a dark side to exchange traded funds? An information perspective, *Review of Accounting Studies* 22, 1048–1083.

Lettau, Martin, and Ananth Madhavan, 2018, Exchange-traded funds 101 for economists, *Journal of Economic Perspectives* 32, 135–154.

Malinova, Katya, and Andreas Park, 2015, Subsidizing liquidity: The impact of make/take fees on market quality, *The Journal of Finance* 70, 509–536.

———, and Ryan Riordan, 2013, Do retail traders suffer from high frequency traders?, Working paper University of Toronto.

Menkveld, Albert J., 2016, The economics of high-frequency trading: Taking stock, *Annual Review of Financial Economics* 8, 1–24.

Morck, Randall, Bernard Yeung, and Wayne Yu, 2013, R^2 and the economy, *Annu. Rev. Financ. Econ.* 5, 143–166.

Pástor, L'uboš, and Veronesi Pietro, 2003, Stock valuation and learning about profitability, *The Journal of Finance* 58, 1749–1789.

Petersen, Mitchell A., 2009, Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches, *Review of Financial Studies* 22, 435–480.

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

Shkilko, Andriy, and Konstantin Sokolov, 2020, Every cloud has a silver lining: Fast trading, microwave connectivity, and trading costs, *The Journal of Finance* 75, 2899–2927.

Thompson, Samuel B., 2011, Simple formulas for standard errors that cluster by both firm and time, *Journal of Financial Economics* 99, 1–10.

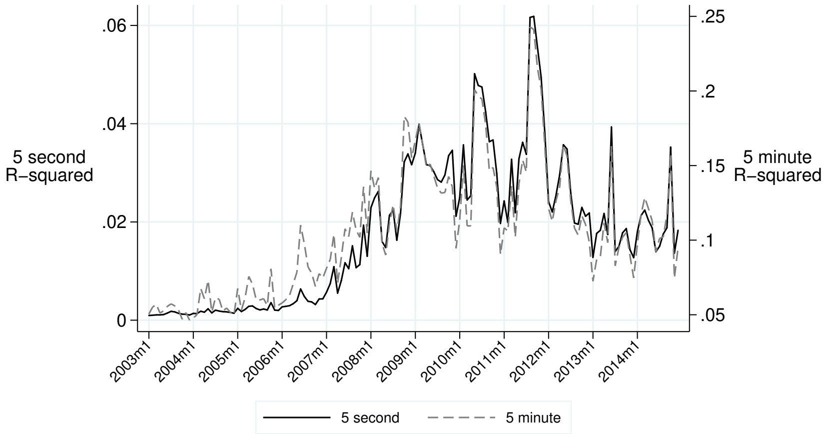
Vijh, Anand M, 1994, S&P 500 trading strategies and stock betas, *The Review of Financial Studies* 7, 215–251.

Wager, Stefan, 2019a, Comments on pull request 502, `https://github.com/grf-labs/grf/pull/502`.

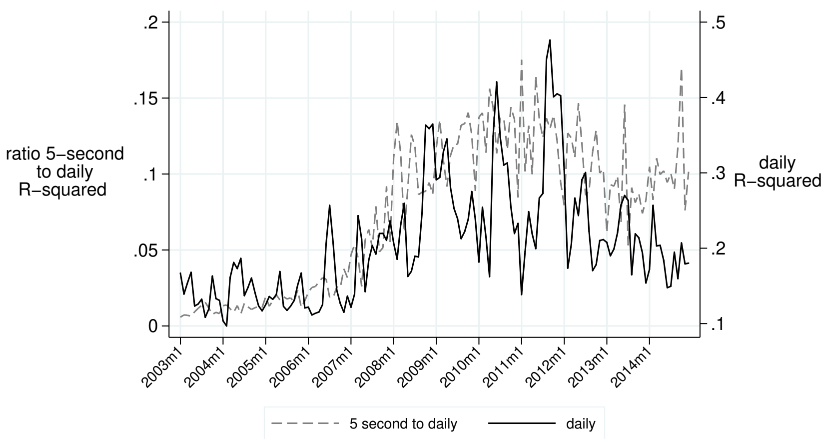
——— , 2019b, Find the best tree in the random forest, `https://github.com/grf-labs/grf/issues/281`.

Figure 1
Monthly Averages of Key Variables

Panel A: 5-second and 5-minute R^2



Panel B: daily R^2 and ratio of 5-second to daily R^2



Panel C: daily R^2 and β estimates

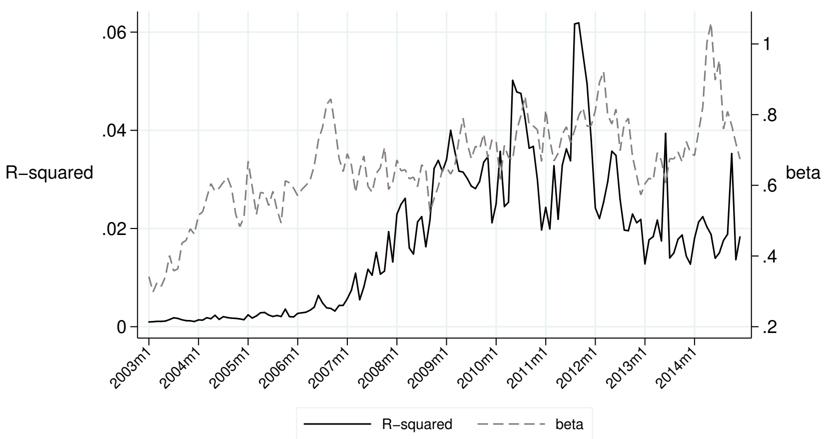
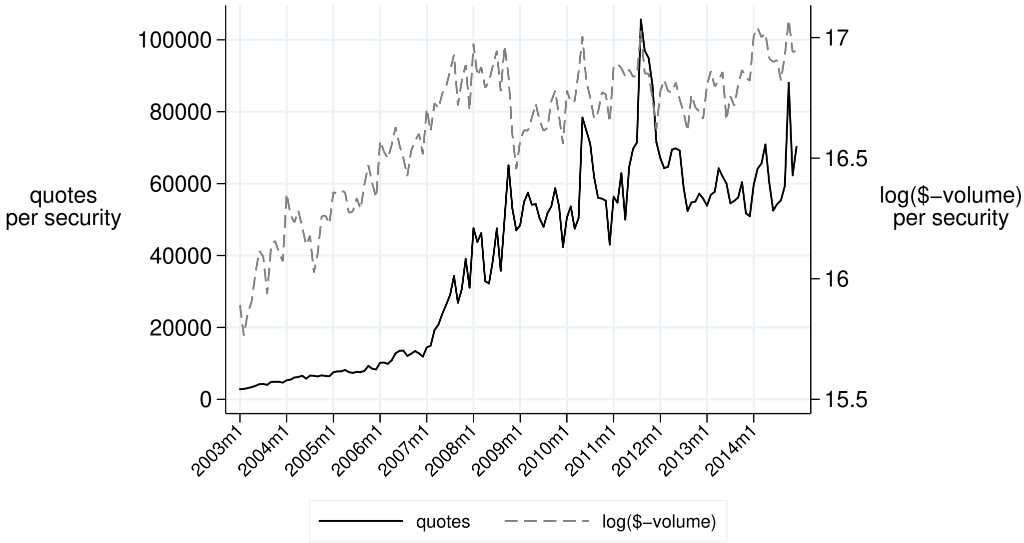
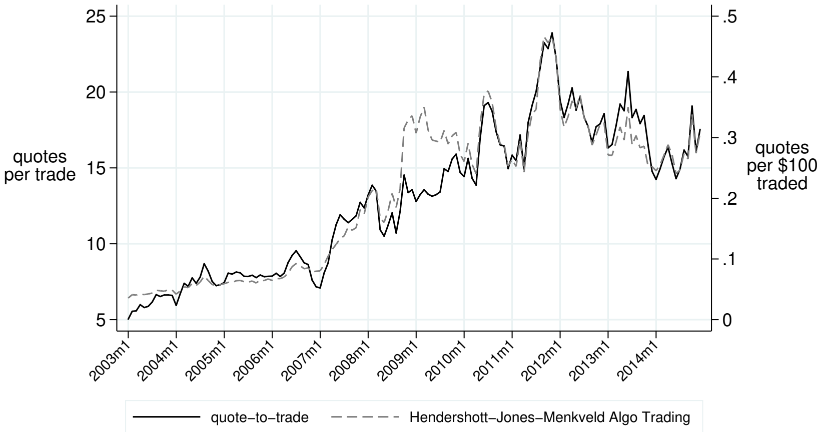


Figure 2
Monthly Averages of Key Variables

Panel A: quotes and $-volume per security



Panel B: quote-to-trade and quotes per $100 of volume



Panel C: quotes and bid-ask-spreads

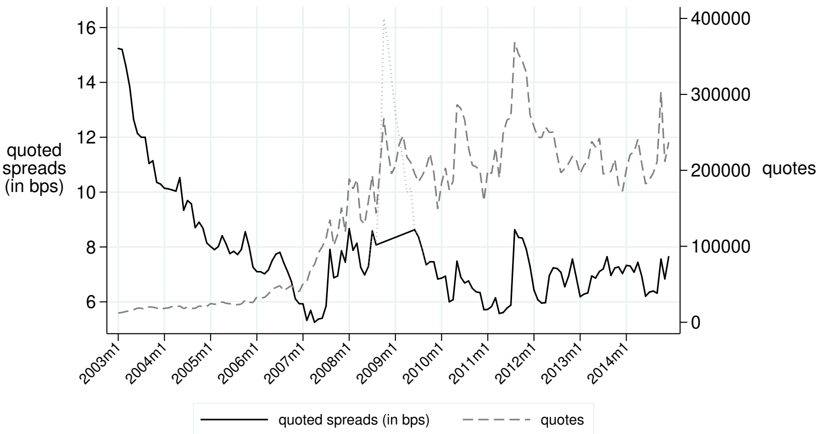
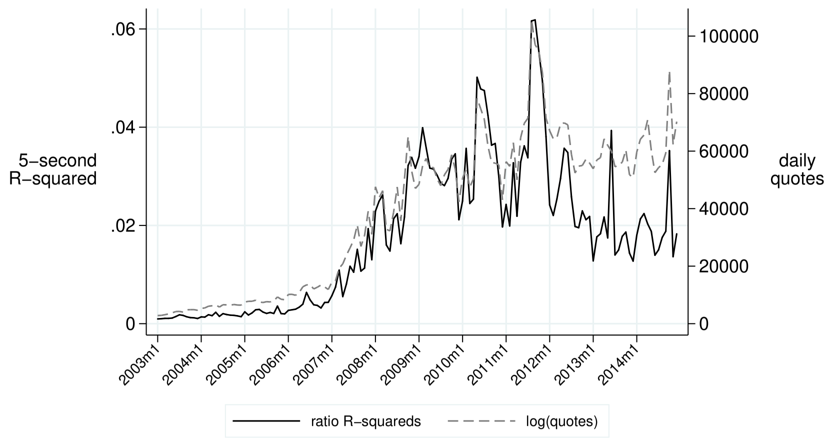
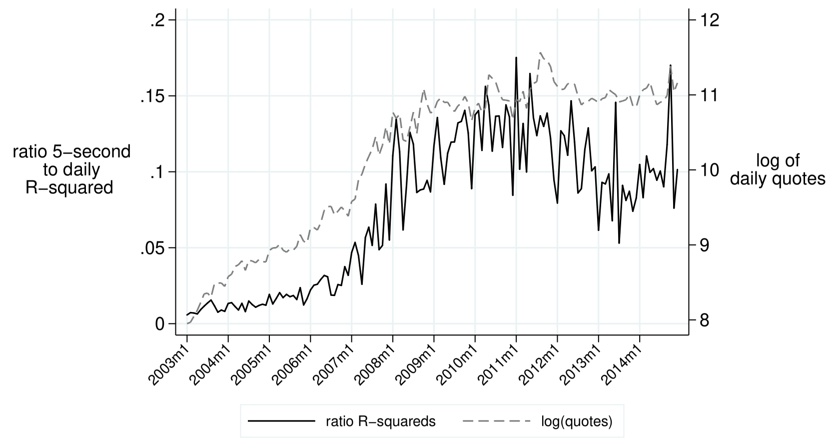


Figure 3
Monthly Averages of Key Variables

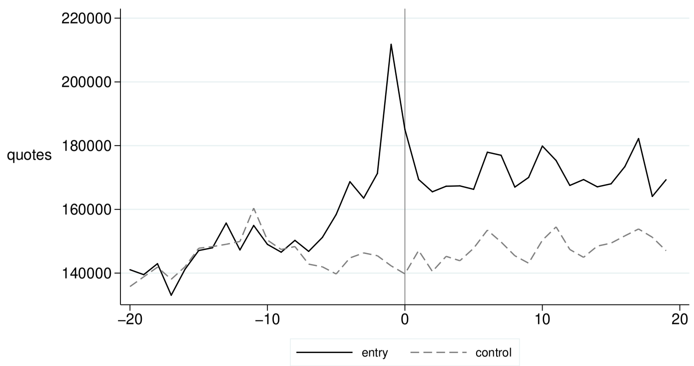


Panel A: 5-second R^2 vs. quotes

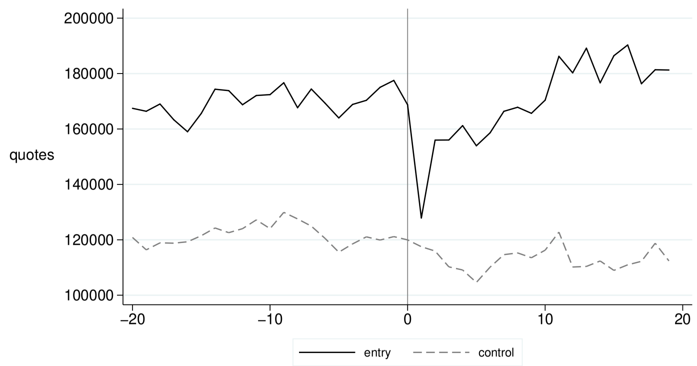


Panel B: ratio of 5-second to daily R^2 and quotes

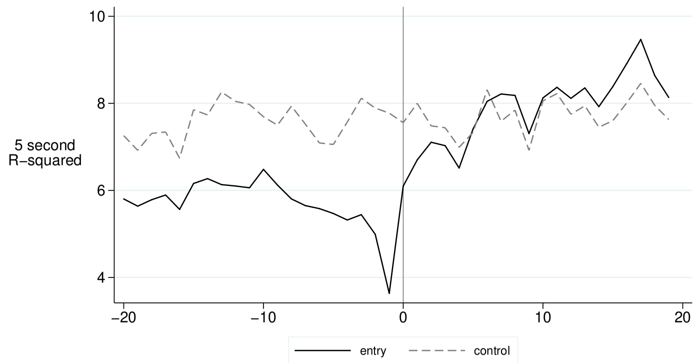
Figure 4
Time Series of R^2 and Quotes for Index Constituents



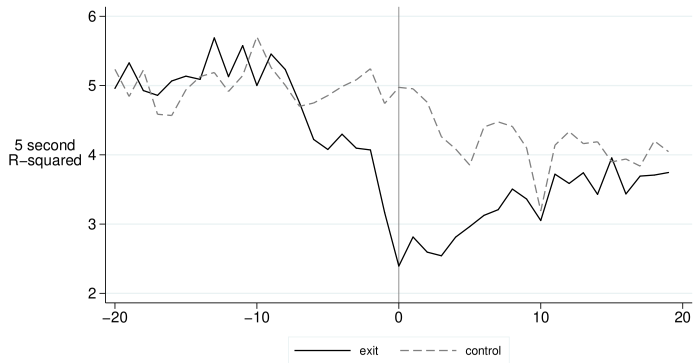
Panel A: quotes treated vs. control for entries



Panel B: quotes treated vs. control for exits



Panel C: 5-second R^2 treated vs. control for entries



Panel D: 5-second R^2 treated vs. control for exits

Figure 5
Best Tree in R-squared 5-minute Inclusion Events

This figure shows the best-pruned tree in the causal forest estimation of treatment effects on 5-minute R-squared in index inclusion events. Each node shows a condition to classify the securities into the next child node. The terminal nodes are referred to as leaves. When the unconfoundedness and overlap assumptions hold (see the Appendix), securities in the leaf nodes can be seen as conditionally randomly assigned to treatment and control groups. size shows the number of securities in the leaf nodes, avg_Y shows the average R-squared of all the securities in the leaf nodes and W shows the average value of treatment in the leaf nodes. W is very close to 0 as our data-set contains more controlled securities than treated securities.

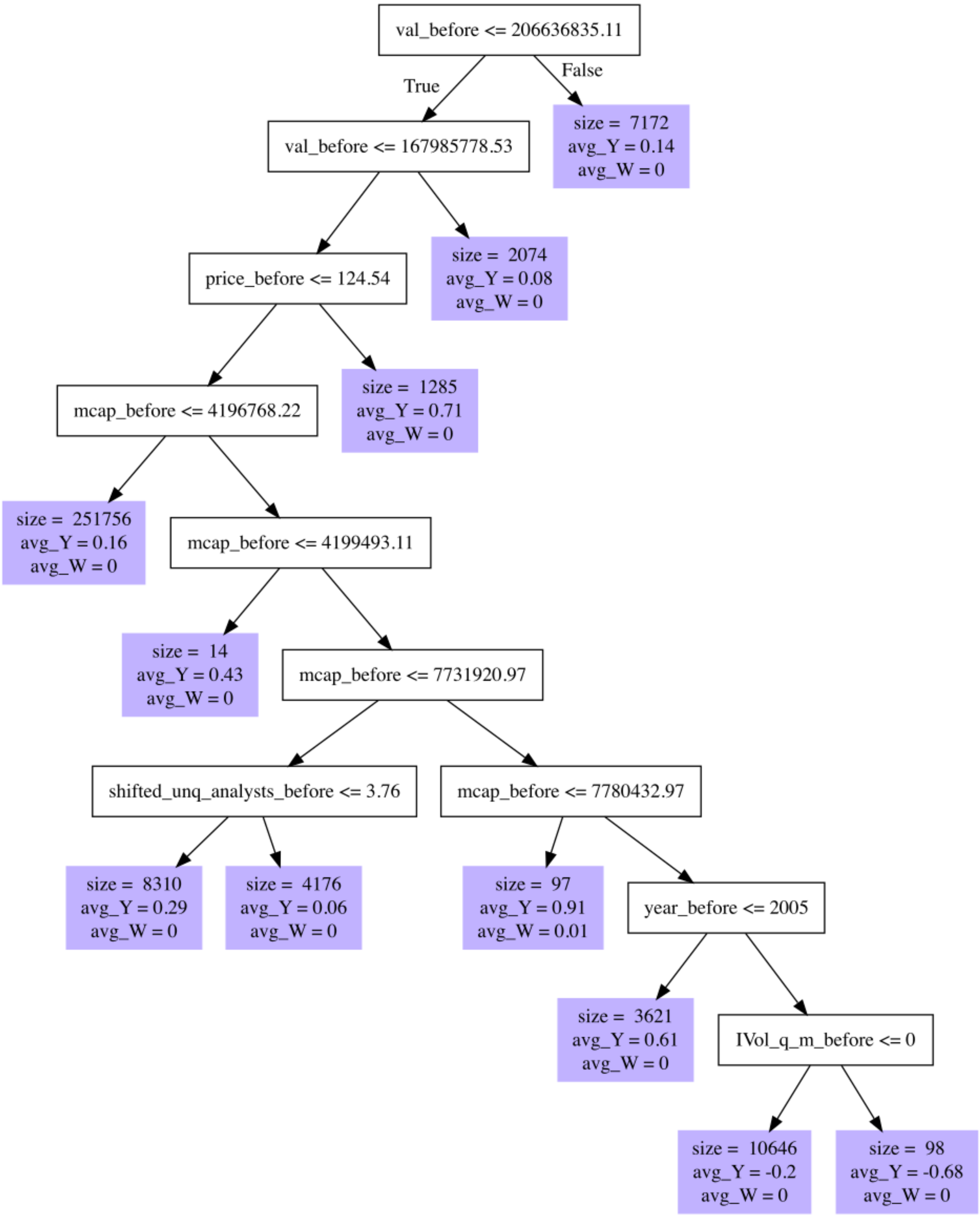


Figure 6
Best Tree: Inclusion Events, 5-minute R^2, Treatment=Fragmentation

This figure shows the best-pruned tree in the instrumental forest estimation of treatment effects on 5-minute R-squared in index inclusion events when Δ(Fragmentation) is used as the treatment. The figures structures is the same as Figure 5.

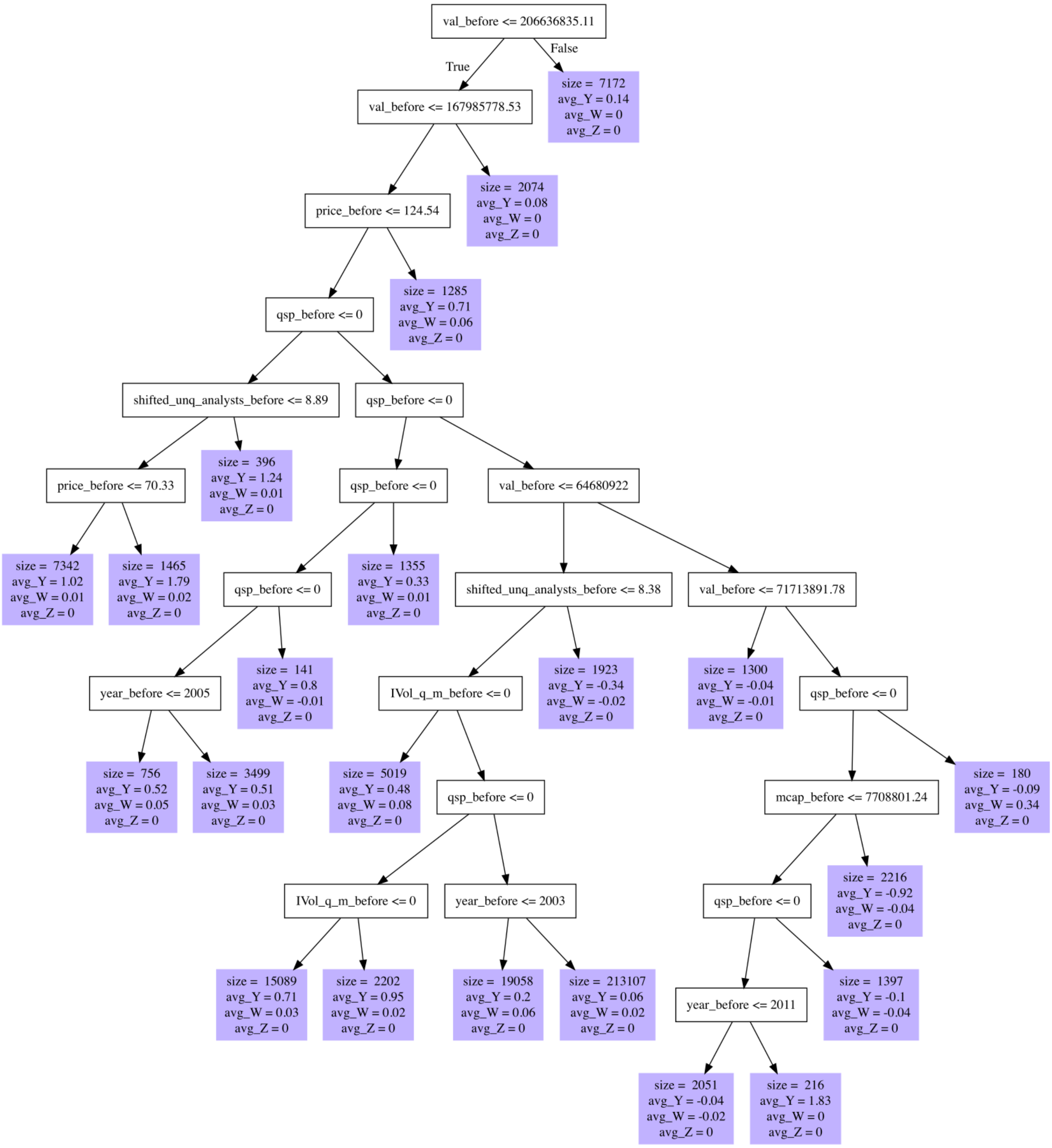


Figure 7
Best Tree: Inclusion Events, 5-minute R^2, Treatment=Number of Quotes

This figure shows the best-pruned tree in the instrumental forest estimation of treatment effects on 5-minute R-squared in index inclusion events when Δ(Number of Quotes) is used as the treatment. The figures structures is the same as Figure 5.

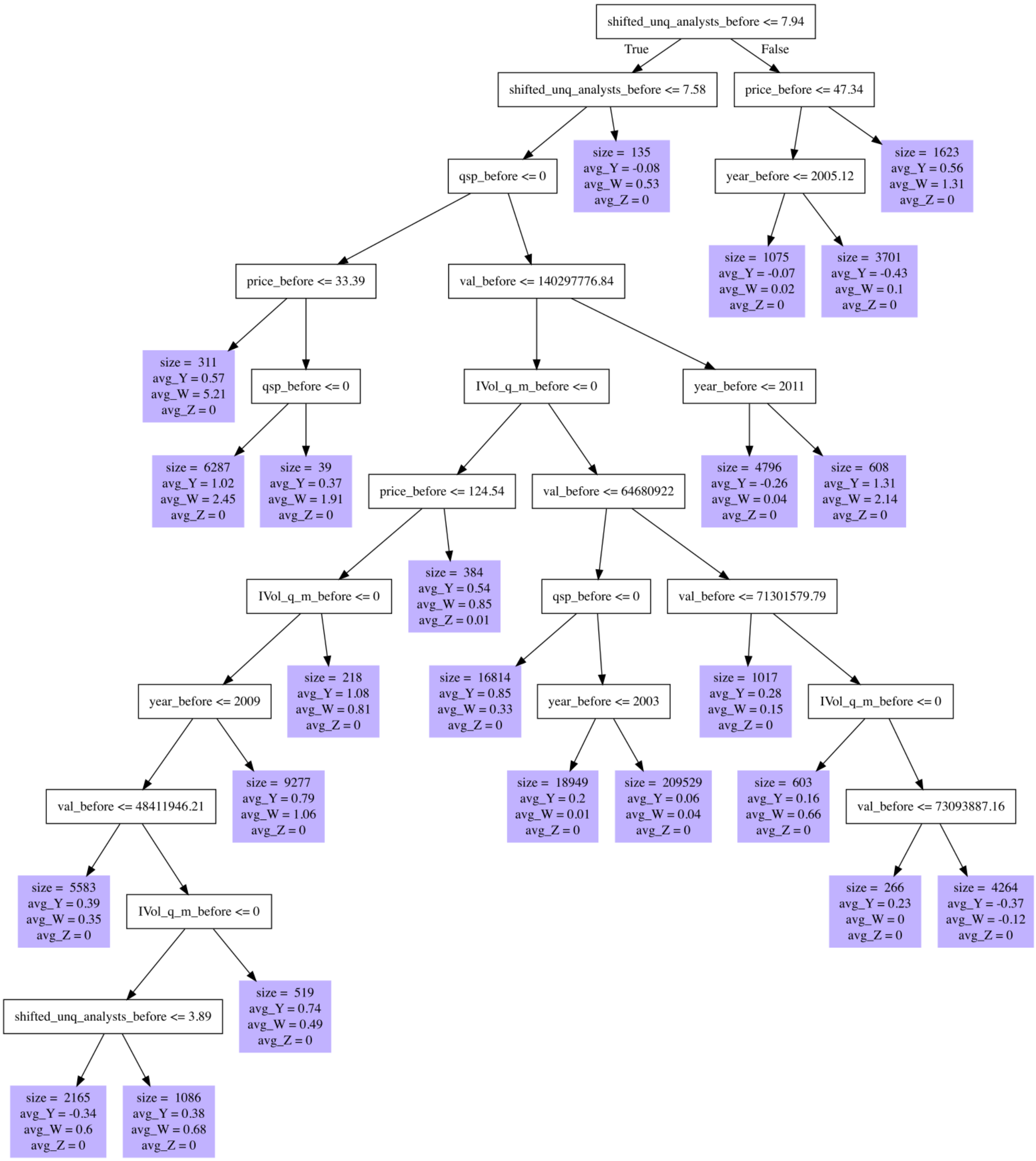


Table 1
Summary Statistics

This table displays summary statistics for the variables that we use in the paper. We present the numbers for the full sample as well as for the smallest and largest quartile by market capitalization. The figures are based on the monthly averages and we split by the three phases in the time series (2003-2006, 2007-2009, and 2010-2014)

	all stocks			smallest quartile			largest quartile		
	2003 –2006	2007 –2009	2010 –2014	2003 –2006	2007 –2009	2010 –2014	2003 –2006	2007 –2009	2010 –2014
5-second (in %)	0.2	1.7	2.1	0.1	0.4	0.2	0.4	5.1	5.9
5-minute	6.4	12.5	12.4	2.0	3.9	2.5	12.6	24.7	22.8
daily	15.0	23.6	23.8	5.1	9.1	8.6	24.1	35.4	34.2
transactions (in K)	1.0	3.3	3.8	0.1	0.1	0.2	2.8	10.5	11.2
$-volume (in $M)	12.1	19.3	21.5	0.2	0.2	0.3	42.8	69.7	75.5
quotes (in K)	7.6	41.5	66.5	1.4	3.0	4.7	20.0	126.9	194.5
bid-ask spread (in bps)	114.6	128.4	88.0	307.2	368.0	259.0	12.6	11.2	8.5
fragmentation	6.5	6.9	6.8	3.3	3.5	4.0	9.4	9.6	9.0
price	19.8	19.4	23.9	5.8	5.3	5.6	39.0	41.2	52.6
market cap (in $B)	2.4	2.7	3.7	0.0	0.0	0.0	8.9	10.0	13.5
quote-to-trade	79.7	98.9	85.7	218.2	280.6	198.8	10.4	18.4	29.7
HJM algo	4.5	13.1	12.9	15.4	46.6	46.4	0.1	0.4	0.7
beta 5 second	0.04	0.15	0.16	0.02	0.09	0.03	0.08	0.29	0.30
beta 5 minute	0.31	0.42	0.46	0.07	0.10	0.12	0.50	0.68	0.64
beta daily	0.54	0.66	0.69	0.23	0.29	0.37	0.69	0.82	0.78

Table 2
Correlation of Variables for Automated Quoting Activity

This table displays summary statistics and the cross-sectional correlation of the sample averages among key trading variables. All estimates are significant at the 1% level.

	log(transactions)	log($-volume)	log(quotes)	quote-to-trade	HJM algo	spread	fragmentation
log(transactions)	1.00						
log($-volume)	0.94	1.00					
log(quotes)	0.90	0.86	1.00				
quote-to-trade	-0.19	-0.16	-0.01	1.00			
HJM algo	-0.04	-0.05	-0.01	0.25	1.00		
spread	-0.61	-0.70	-0.59	0.09	0.06	1.00	
fragmentation	0.91	0.91	0.78	-0.20	-0.05	-0.66	1.00

Table 3
Correlation of R^2 with Automated Quoting Measures

This table displays summary statistics and the cross-sectional correlation of the sample averages among key trading variables. All estimates are significant at the 1% level.

	5 second	5 minute	daily	ratio 5-second to daily
log(transactions)	0.49	0.60	0.45	0.04
log($-volume)	0.48	0.60	0.46	0.03
log(quotes)	0.56	0.67	0.53	0.04
quote-to-trade	0.03	0.02	0.02	0.00
HJM algo	-0.01	-0.02	-0.02	0.00
spread	-0.21	-0.36	-0.35	0.01
fragmentation	0.34	0.53	0.43	0.02

Table 4
Long-Run Regressions

This table displays summary statistics and the cross-sectional correlation of the sample averages among key trading variables. Standard errors are double-clustered by security and date; we present t-statistics in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels.

	log(transactions)	log($-volume)	log(quotes)	quote-to-trade	HJM-algo	spreads	fragmentation
Panel A: 5-second							
full sample	0.73***	0.34***	0.80***	0.00***	0.00**	-0.00**	-0.17***
	(16.710)	(9.879)	(17.953)	(4.926)	(2.479)	(-2.604)	(-6.114)
2003–2006	0.13***	0.15***	0.41***	0.00***	0.00	0.00**	-0.20***
	(3.660)	(4.528)	(10.011)	(4.182)	(1.573)	(2.360)	(-8.347)
2007-2009	0.34***	0.22***	0.49***	0.00	-0.00*	-0.00***	-0.02
	(8.825)	(6.931)	(13.229)	(0.880)	(-1.884)	(-5.877)	(-0.668)
2010-2014	0.49***	0.30***	0.58***	0.00***	0.00*	0.00***	0.14***
	(12.241)	(8.648)	(14.741)	(4.745)	(1.957)	(3.030)	(5.866)
Panel B: 5-minute							
full sample	2.59***	1.55***	2.70***	0.00***	0.00**	-0.00***	0.72***
	(17.708)	(12.600)	(20.278)	(5.363)	(2.515)	(-4.090)	(6.982)
2003–2006	1.39***	1.04***	2.26***	0.00***	0.00	-0.00*	0.60***
	(10.995)	(9.684)	(16.266)	(4.911)	(1.499)	(-1.943)	(7.441)
2007-2009	1.86***	1.27***	2.36***	0.00	-0.00**	-0.01***	1.14***
	(14.690)	(11.414)	(20.928)	(0.641)	(-2.201)	(-6.158)	(11.766)
2010-2014	2.10***	1.41***	2.45***	0.00***	0.00**	0.00	1.47***
	(15.646)	(11.978)	(19.212)	(5.743)	(1.984)	(0.939)	(15.735)
Panel C: Daily							
full sample	2.41***	0.97***	3.17***	0.00***	-0.00	-0.01***	0.74***
	(10.189)	(4.999)	(12.617)	(3.990)	(-0.515)	(-7.705)	(3.959)
2003–2006	1.36***	0.51***	2.91***	0.00	0.01***	-0.01***	0.84***
	(5.532)	(2.824)	(11.575)	(1.365)	(2.797)	(-7.602)	(5.564)
2007-2009	1.54***	0.63***	2.74***	-0.00	-0.00***	-0.02***	1.14***
	(6.768)	(3.406)	(14.668)	(-1.045)	(-2.786)	(-8.409)	(6.254)
2010-2014	2.10***	0.93***	3.07***	0.00***	0.00	-0.00	1.78***
	(9.308)	(4.900)	(13.489)	(3.837)	(1.319)	(-1.512)	(10.258)

Table 5
Summary Statistics

This table displays summary statistics for the variables that we use in the analysis of entry and exit events. The figures are based on plain averages for the 20 trading days before and after an index inclusion/exclusion.

	exit						*entry*					
	before		after		difference after-before		before		after		difference after-before	
	treatment	control	treatment	control	treatment	control	treatment	control	treatment	control	treatment	control
5-second (in %)	3.8	3.5	3.7	2.6	-0.1	-0.9	5.6	4.6	5.7	5.9	0.1	1.2
5-minute	17.9	15.3	17.7	14.8	-0.2	-0.5	21.8	19.0	22.0	21.3	0.2	2.3
daily	30.6	26.2	29.4	24.3	-1.2	-1.9	31.4	29.9	33.3	28.2	1.9	-1.7
transactions (in K)	1.0	1.4	1.0	1.7	0.0	0.3	1.0	1.3	1.0	1.4	0.0	0.1
$-volume (in $M)	74.0	117.9	74.4	91.3	0.4	-26.6	96.7	147.5	96.6	143.6	-0.1	-3.9
quotes (in 100K)	1.2	1.7	1.2	1.8	0.0	0.1	1.5	1.5	1.5	1.7	0.0	0.2
bid-ask spread (in bps)	13.0	10.1	13.1	15.7	0.1	5.6	7.8	7.1	7.9	6.3	0.1	-0.8
fragnentation	9.2	9.1	9.3	9.1	0.1	0.0	9.6	9.6	9.6	10.0	0.0	0.4
price	38.5	36.1	38.6	26.2	0.1	-9.9	63.0	62.4	62.5	62.4	-0.5	0.0
market cap (in $B)	9.3	9.4	9.4	4.7	0.1	-4.7	9.8	9.9	9.8	10.1	0.1	0.2
quote-to-trade	16.1	13.3	16.3	11.5	0.1	-1.9	17.9	13.4	18.4	12.7	0.5	-0.6
HJM algo	0.4	0.3	0.4	0.4	0.0	0.1	0.2	0.2	0.3	0.2	0.0	0.0
beta 5 second	0.2	0.2	0.2	0.2	0.0	0.0	0.3	0.3	0.3	0.3	0.0	0.0
beta 5 minute	0.6	0.5	0.6	0.7	0.0	0.2	0.6	0.6	0.6	0.6	0.0	0.0
beta daily	0.8	0.6	0.7	0.8	-0.1	0.2	0.7	0.7	0.8	0.7	0.1	0.0

Table 6
Impact of Index Inclusion and Exclusion Events

This table displays our estimation of (5) where we test whether there are changes in key variables of interest subsequent to a stock being included in (Panel A) and excluded (Panel B) of the S&P 500. Each row refers to a different set of matching characteristics, where we match based on the monthly averages in the calendar month prior to the index change (e.g., June if the change happened on July 5th). Variables of interest are the 5-second, 5-minute and daily R^2 of the market regressions (where we proxy the market return by that of the Russell 2000 index), the natural logarithms of the trades, $-volume, and quote changes, the quote-to-trade ratio, the Hendershott, Jones, and Menkveld measure of automated quoting, the time-weighted daily quoted spread (measured in basis points of the prevailing midquote), and the inverse of the Hirschman-Herfindahl index of market concentration as a measure of order fragmentation. In the regression, the dependent variables are the difference of the value for the treatment and the control. All regressions control for the treatment stocks log-price and log-market capitalization as well as the daily value of the volatility index VIX; we display only the estimate for the change. Each entry is the result of a single regression. Standard errors are double-clustered by security and date; we present t-statistics in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels.

	5 second	5 minute	daily	log(trades)	log($-vol)	log(quotes)	quote-to -trade	HJM algo	spreads	fragmentation
Panel A: inclusion events										
price & m-cap	0.01***	0.02***	-0.05***	0.13***	0.12***	0.07***	-1.15***	-0.02**	-0.87***	0.42***
	(5.070)	(4.120)	(-2.938)	(6.313)	(4.965)	(3.783)	(-2.579)	(-2.217)	(-6.442)	(8.229)
price & quotes	0.01***	0.03***	-0.01	0.11***	0.09***	0.12***	1.39*	0.01	-0.51***	0.42***
	(5.421)	(5.397)	(-0.896)	(4.782)	(3.415)	(5.238)	(1.786)	(0.801)	(-3.381)	(8.223)
price, m-cap, & quotes	0.02***	0.03***	-0.03**	0.11***	0.09***	0.09***	-0.23	-0.00	-0.86***	0.44***
	(7.194)	(5.061)	(-2.377)	(5.140)	(3.537)	(4.999)	(-0.612)	(-0.329)	(-8.611)	(8.960)
price, quotes, & transactions	0.01***	0.02***	-0.06***	0.15***	0.14***	0.08***	-1.27***	-0.02***	-0.94***	0.38***
	(5.377)	(3.280)	(-4.559)	(6.678)	(5.280)	(3.483)	(-3.381)	(-2.845)	(-8.039)	(7.411)
price & $-vol	0.01***	0.02***	-0.05***	0.13***	0.13***	0.08***	-1.00**	-0.02**	-0.81***	0.46***
	(5.459)	(4.269)	(-3.571)	(5.693)	(4.456)	(3.701)	(-2.400)	(-2.572)	(-6.906)	(8.748)
price & transactions	0.01***	0.02***	-0.06***	0.15***	0.15***	0.10***	-1.04**	-0.02**	-0.89***	0.40***
	(5.595)	(4.044)	(-4.381)	(6.551)	(5.269)	(4.195)	(-2.436)	(-2.437)	(-6.654)	(7.759)
price, m-cap, & transactions	0.01***	0.02***	-0.04**	0.14***	0.12***	0.09***	-0.72*	-0.01	-0.84***	0.35***
	(5.770)	(4.392)	(-2.551)	(6.044)	(4.674)	(4.248)	(-1.664)	(-1.484)	(-7.874)	(6.810)
Panel B: exclusion events										
price & m-cap	-0.01***	-0.02***	-0.02	0.05	0.13***	0.03	-0.52	-0.02	-0.34	-0.29***
	(-2.578)	(-2.869)	(-0.961)	(1.131)	(2.604)	(0.952)	(-0.515)	(-0.607)	(-1.050)	(-3.122)
price & quotes	-0.01***	-0.02***	-0.02	0.10***	0.20***	0.08**	2.26	0.04	-0.73	-0.27***
	(-3.144)	(-2.597)	(-0.950)	(2.660)	(4.565)	(2.341)	(1.115)	(0.753)	(-1.389)	(-2.810)
price, m-cap, & quotes	-0.01**	-0.01*	-0.04*	0.05	0.10**	0.04	0.30	0.02	0.35	-0.26***
	(-2.325)	(-1.747)	(-1.934)	(1.169)	(2.164)	(1.282)	(0.432)	(0.869)	(1.046)	(-3.029)
price, quotes, & transactions	-0.01***	-0.02***	-0.05***	0.11**	0.18***	0.01	-1.93***	-0.05***	-0.47	-0.07
	(-3.078)	(-2.803)	(-3.007)	(2.487)	(3.638)	(0.209)	(-3.157)	(-2.879)	(-1.229)	(-0.673)
price & $-vol	-0.01***	-0.02***	-0.04**	0.15***	0.25***	0.07**	-1.29***	-0.04**	-0.47	-0.09
	(-2.907)	(-3.671)	(-2.275)	(3.543)	(5.202)	(2.153)	(-2.824)	(-2.429)	(-1.230)	(-0.860)
price & transactions	-0.01***	-0.02***	-0.04**	0.09**	0.16***	0.00	-1.87***	-0.06***	-0.09	-0.12
	(-2.973)	(-2.869)	(-2.432)	(2.215)	(3.749)	(0.052)	(-3.210)	(-3.094)	(-0.176)	(-1.236)
price, m-cap, & transactions	-0.01***	-0.03***	-0.06***	0.10**	0.19***	-0.00	-1.96***	-0.05***	-0.37	-0.12
	(-3.323)	(-3.465)	(-3.252)	(2.319)	(3.870)	(-0.142)	(-3.402)	(-3.146)	(-0.837)	(-1.297)

Table 7
Impact of Index Inclusion and Exclusion Events by Years

This table displays our estimation of (6) where we test whether there are changes in key variables of interest subsequent to a stock being included (Panel A) and excluded (Panel B) of the S&P 500, split up by three phases: the years 2003-2006 (the early part of the sample), 2007-2009 (the middle), and 2010-2014 (the late part of the sample). The estimates are based on the matching treatment and control based on price and market capitalization. Variables of interest and controls are the same as in Table 6. We display only the estimates for the change for the respective phases as well as a t-test for equality of coefficients. Standard errors are double-clustered by security and date; we present t-statistics in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels.

	5 second	5 minute	daily	log(trades)	log($-vol)	log(quotes)	quote-to -trade	HJM algo	spreads	fragmentation
Panel A: inclusion events										
2003-2006	0.00**	0.02***	0.03	0.11***	0.11**	-0.01	-1.12*	-0.00	-1.02***	0.44***
	(1.967)	(2.920)	(0.753)	(3.270)	(2.208)	(-0.215)	(-1.687)	(-0.634)	(-2.643)	(4.097)
2007-2009	0.01***	0.02**	-0.09***	0.16***	0.16***	0.10***	-1.89***	-0.04***	-0.85***	0.28***
	(2.965)	(2.028)	(-4.625)	(5.122)	(4.390)	(3.049)	(-2.613)	(-2.657)	(-4.693)	(3.509)
2010-2014	0.02***	0.03***	-0.04	0.11***	0.10**	0.13***	-0.25	-0.00	-0.80***	0.55***
	(4.785)	(2.922)	(-1.366)	(2.666)	(2.124)	(3.377)	(-0.271)	(-0.129)	(-4.226)	(7.197)
early $\neq$ middle	Yes***		Yes***			Yes**		Yes**		
early $\neq$ late	Yes***					Yes***				
middle $\neq$ late								Yes*		Yes**
Panel B: exclusion events										
2003-2006	0.00**	-0.02**	0.02	0.06	0.25**	0.03	0.59	-0.01	-0.54	-0.46**
	(2.042)	(-2.119)	(0.556)	(0.763)	(2.448)	(0.422)	(0.534)	(-0.270)	(-0.476)	(-2.382)
2007-2009	-0.01	-0.02	0.01	-0.01	0.07	0.03	1.85	0.04	-0.20	-0.41***
	(-1.320)	(-1.299)	(0.265)	(-0.154)	(0.832)	(0.410)	(0.855)	(0.604)	(-0.317)	(-2.693)
2010-2014	-0.01**	-0.02*	-0.07*	0.07	0.10	0.04	-2.27*	-0.04	-0.27	-0.12
	(-2.296)	(-1.807)	(-1.955)	(1.157)	(1.447)	(0.900)	(-1.900)	(-0.992)	(-0.694)	(-0.892)
early $\neq$ middle	Yes*									
early $\neq$ late	Yes***		Yes*				Yes*			
middle $\neq$ late										

Table 8
Impact of Index Inclusion and Exclusion: Instrumental Variable Approach (2nd stage)

This table displays our estimation of the two step estimation of (7) where we use the inclusion/exclusion events as an instrument for our proxies of automated quoting: the level of order fragmentation (for exclusion and inclusion events, Panels A and B) and the natural logarithm of the number of quotes (for entry events, Panel B). The estimates are based on the matching treatment and control based on price and market capitalization. Variables of interest and controls are the same as in Table 6, where we exclude the estimates for all controls. Standard errors are double-clustered by security and date; we present t-statistics in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels. We further present the Kleinbergen-Papp statistics for weak- and under-identification.

	5-second	5-minute	daily	5-second	5-minute	daily
Panel A: Exclusion events						
fragmentation				0.02**	0.07**	0.06
				(2.054)	(2.127)	(0.882)
under-identification				9.708	9.917	8.928
weakid (χ^2)				9.419	9.605	8.669
Panel B: inclusion events						
log(quotes)	0.16***	0.29***	-0.58**			
	(3.587)	(3.027)	(-2.385)			
fragmentation				0.03***	0.05***	-0.11***
				(4.238)	(3.883)	(-2.722)
under-identification	13.23	13.23	12.84	49.36	49.36	41.11
weakid (χ^2)	14.23	14.23	14.20	67.34	67.34	56.42

Table 9
Mediation Analysis

This table displays our estimation of our mediation analysis, where we present only the results from the estimation of the joint effect, (8). The estimates are based on the matching treatment and control based on price and market capitalization. Variables of interest and controls are the same as in Table 6. We display only the estimates for the change for the respective phases as well as a t-test for equality of coefficients. Standard errors are clustered by security; we present t-statistics in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels.

	5-second	5-minute	daily	5-second	5-minute	daily
Panel A: Exclusion events						
event	-0.01***	-0.02***	-0.06**	-0.01***	-0.02**	-0.05**
	(0.00)	(0.01)	(0.02)	(0.00)	(0.01)	(0.02)
quotes	0.02***	0.05***	0.04**			
	(0.00)	(0.01)	(0.02)			
fragmentation				0.00***	0.01***	0.00
				(0.00)	(0.00)	(0.00)
Observations	6,193	6,179	5,387	6,192	6,178	5,386
R-squared	0.152	0.135	0.042	0.051	0.049	0.017
Panel B: inclusion events						
event	0.01***	0.02***	-0.05***	0.01***	0.02***	-0.05***
	(0.00)	(0.01)	(0.02)	(0.00)	(0.01)	(0.02)
quotes	0.03***	0.03***	0.03			
	(0.00)	(0.01)	(0.02)			
fragmentation				0.00***	0.01***	0.00
				(0.00)	(0.00)	(0.00)
Observations	7,855	7,855	6,104	7,855	7,855	6,104
R-squared	0.139	0.035	0.026	0.024	0.014	0.020

Table 10
Mediation Analysis: direct and indirect effects

This table displays our estimation of our mediation analysis, where we present the Average Causal Mediation Effect (ACME), the direct effect (of the index event), and the indirect (mediated) effect. Our statistical package provides only the 95% confidence interval for the estimate; we indicate by ** whether this interval is entirely positive or negative.

	exclusions						*inclusions*					
	5-second		5-minute		daily		5-second		5-minute		daily	
Panel A: Mediator = log(number of quotes)												
ACME							0.002	**	0.003	**	0.002	
direct							0.010	**	0.019	**	-0.053	**
indirect							0.013	**	0.022	**	-0.051	**
Panel B: Mediator = order fragmentation												
ACME	0.000		0.001		0.000		0.001	**	0.003	**	0.001	
direct	-0.012	**	-0.020	**	-0.053	**	0.011	**	0.019	**	-0.052	**
indirect	-0.011	**	-0.020	**	-0.053	**	0.013	**	0.022	**	-0.051	**

Table 11
Causal Forest Estimation of Average Treatment Effects

This table displays our estimation of the average treatment effects where we test whether there are changes in intra-day R-squared subsequent to 1) a stock being included (Panel A) and excluded (Panel B) of the S&P 500; 2) changes in the automated quoting proxies from before to after the inclusion/exclusion events. We use one of the following variables as the dependent variable: $\Delta(R^2$ 5sec), $\Delta(R^2$ 5min), and $\Delta(R^2$ daily). As covariates we use the the 20-day average before the event of the following variables: Price, Market Cap, $\$-$Volume, qspread cents, Number of Analysts in the Preceding Quarter Intra-day Volatility and Year. We use both binary and continuous treatment variables. As binary treatment variables we use dummy variables that indicates if a security was included or excluded from an index during the event. As continuous treatment variables we use the difference of the 20-day average before to after the event of the automated quoting proxies: Δ(Number of Quotes) or Δ(Fragmentation). Standard errors are clustered by events; we present standard errors in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels.

	$\tau_{AllSecurities}$			$\tau_{TreatedSecurities}$			$\tau_{Overlap-weightedSecurities}$		
	5-second	5-minute	daily	5-second	5-minute	daily	5-second	5-minute	daily
Panel A: Inclusion events									
event	0.34	0.84	−0.55	1.29***	2.57***	−2.77***	1.29***	2.57***	−2.78***
	(0.58)	(1.66)	(6.44)	(0.24)	(0.50)	(1.04)	(0.25)	(0.49)	(1.05)
quotes	0.17***	1.30***	−0.43				0.20***	0.37***	0.12***
	(0.01)	(0.16)	(0.22)				(0.01)	(0.03)	(0.05)
fragmentation	0.17***	0.65***	−0.95***				0.10***	0.49***	−0.84***
	(0.04)	(0.12)	(0.16)				0.03	(0.10)	(0.12)
Panel B: Exclusion events									
event	−0.84**	−4.50***	−5.21*	−2.04***	−5.33***	−9.07***	−2.04***	−5.33***	−9.07***
	(0.48)	(1.19)	(3.57)	(0.75)	(1.06)	(2.86)	(0.75)	(1.03)	(2.85)
quotes	0.17***	1.30***	−0.43**				0.20***	0.37***	0.15***
	(0.02)	0.19	0.27				(0.02)	0.04	(0.06)
fragmentation	0.15***	0.62***	−1.06***				0.09***	0.48***	−0.93***
	(0.05)	(0.16)	(0.18)				(0.03)	(0.12)	(0.14)

Table 12
Causal Forests Estimation on Automated Quoting Indicators

This table displays our estimation of the average treatment effects where we test whether there are changes in automated quoting proxies subsequent to a stock being included (Panel A) and excluded (Panel B) of the S&P 500. We use one of the following variables as the dependent variable: Δ(Number of Quotes) or Δ(Fragmentation). As covariates we use the the 20-day average before the event of the following variables: Price, Market Cap, $\$-$Volume, qspread cents, Number of Analysts in the Preceding Quarter Intra-day Volatility and Year. As treatment variables we use dummy variables that indicates if a security was included or excluded from an index during the event. Standard errors are clustered by events; we present standard errors in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels.

	$\tau_{AllSecurities}$		$\tau_{TreatedSecurities}$		$\tau_{Overlap-weightedSecurities}$	
	fragmentation	quote	fragmentation	quote	fragmentation	quote
Panel A: Inclusion events						
event	0.51***	2955.08	0.51***	17722.55***	0.51***	17719.64***
	(0.20)	(4913.27)	(0.04)	(5831.26)	(0.04)	(5716.56)
Panel B: Exclusion events						
event	0.04	−2285.00	0.10	−6040.58	0.10	−6041.21
	(0.08)	(6500.92)	(0.09)	(11490.68)	(0.09)	(11060.44)

Table 13
Instrumental Forests Estimation of Average Treatment Effects

This table displays our estimation of the average treatment effects using instrumental forests. As instrumental variables we use dummy variables that indicates if a security was included or excluded from an index during the event. As instrumented variables we use one of the following automated quoting proxy variables: Δ(Number of Quotes) or Δ(Fragmentation). As covariates we use the the 20-day average before the event of the following variables: Price, Market Cap, $\$-$Volume, qspread cents, Number of Analysts in the Preceding Quarter Intra-day Volatility and Year. Standard errors are clustered by events; we present standard errors in parentheses. *, **, *** indicate significance at the 10%, 5%, and 1% levels.

	5-second R-squared		5-minute R-squared		daily R-squared	
	Instrumented Variables					
	fragmentation	quote	fragmentation	quote	fragmentation	quote
Panel A: Inclusion events						
event	1.88**	1.12**	3.30	4.68***	−12.55***	−4.88***
(instrumental variable)	(0.84)	(0.51)	(2.76)	(1.73)	(3.48)	(1.30)
Panel B: Exclusion events						
event	218.31*	17661.05	2566.77	−187.04	−5253.265	−431.15*
(instrumental variable)	(165.40)	(17571.38)	(3516.71)	(181.44)	(8882.328)	(301.64)

Table 14
Causal Forest Index Inclusion Variable Importance

This table displays our variable importance measure for causal forest estimation of average treatment effects in the index inclusion events. The horizontal axis list the covariates X_i and the vertical axis list the dependent variables. The numbers in the table are the percentage splits on a particular covariate. The higher the percentage, the deeper the gradient in the cell, and the more important the variable is in the causal forest.

Covariates / Dependent Variables	price_before	val_before	mcap_before	qsp_before	shifted_unq_analysts_before	lVol_q_m_before	year_before
fragmentation	0.33	0.32	0.08	0.07	0.13	0.06	0.01
frag_rsq5min	0.01	0.31	0.04	0.11	0	0.1	0.42
frag_rsq5sec	0.01	0.45	0.03	0.19	0	0.05	0.27
frag_rsq_daily	0.01	0.2	0.13	0.04	0.01	0.17	0.43
quotes	0.23	0.38	0.07	0.09	0.15	0.06	0.02
quotes_rsq_5min	0.01	0.78	0.08	0.07	0.01	0.01	0.04
quotes_rsq5sec	0.02	0.77	0.09	0.06	0.01	0.01	0.04
quotes_rsqdaily	0.02	0.38	0.19	0.06	0.01	0.09	0.25
rsq_5min	0.45	0.26	0.08	0.07	0.07	0.06	0.01
rsq_5sec	0.23	0.34	0.08	0.12	0.1	0.1	0.02
rsq_daily	0.4	0.25	0.16	0.06	0.07	0.05	0.01

Table 15
Causal Forest Index Exclusion Variable Importance

This table displays our variable importance measure for causal forest estimation of average treatment effects in the index exclusion events. The horizontal axis list the covariates X_i and the vertical axis list the dependent variables. The numbers in the table are the percentage splits on a particular covariate. The higher the percentage, the deeper the gradient in the cell, and the more important the variable is in the causal forest.

Covariates / Dependent Variables	val_before	lVol_q_m_before	qsp_before	mcap_before	shifted_unq_analysts_before	price_before	year_before
fragmentation	0.23	0.14	0.18	0.1	0.18	0.1	0.06
frag_rsq5min	0.25	0.13	0.08	0.05	0.01	0.01	0.47
frag_rsq5sec	0.39	0.08	0.16	0.03	0	0.01	0.32
frag_rsq_daily	0.24	0.21	0.04	0.15	0.02	0.01	0.32
quotes	0.24	0.15	0.19	0.13	0.13	0.06	0.08
quotes_rsq_5min	0.78	0.01	0.08	0.08	0.01	0.01	0.04
quotes_rsq5sec	0.78	0.01	0.07	0.08	0.01	0.02	0.03
quotes_rsqdaily	0.41	0.13	0.07	0.12	0.01	0.02	0.24
rsq_5min	0.22	0.18	0.17	0.14	0.14	0.08	0.08
rsq_5sec	0.28	0.13	0.18	0.13	0.15	0.07	0.05
rsq_daily	0.2	0.13	0.24	0.12	0.13	0.11	0.06

Table 16
Instrumental Forest Index Inclusion Variable Importance

This table displays our variable importance measure for instrumental forest estimation of average treatment effects in the index inclusion events. The horizontal axis list the covariates X_i and the vertical axis list the dependent variables. The numbers in the table are the percentage splits on a particular covariate. The higher the percentage, the deeper the gradient in the cell, and the more important the variable is in the instrumental forest.

Covariates / Dependent Variables	price_before	val_before	shifted_unq_analysts_before	mcap_before	qsp_before	lVol_q_m_before	year_before
instru_frag_5min	0.43	0.26	0.09	0.07	0.07	0.06	0.01
instru_quotes_5min	0.23	0.36	0.14	0.07	0.13	0.06	0.01
instru_frag_5sec	0.24	0.33	0.13	0.08	0.11	0.09	0.02
instru_quotes_5sec	0.25	0.39	0.14	0.06	0.09	0.05	0.01
instru_frag_daily	0.41	0.23	0.07	0.17	0.06	0.04	0.01
instru_quotes_daily	0.37	0.34	0.07	0.11	0.06	0.04	0.01

Table 17
Instrumental Forest Index Exclusion Variable Importance

This table displays our variable importance measure for instrumental forest estimation of average treatment effects in the index exclusion events. The horizontal axis list the covariates X_i and the vertical axis list the dependent variables. The numbers in the table are the percentage splits on a particular covariate. The higher the percentage, the deeper the gradient in the cell, and the more important the variable is in the instrumental forest.

Covariates / Dependent Variables	val_before	qsp_before	shifted_unq_analysts_before	lVol_q_m_before	price_before	mcap_before	year_before
instru_frag_5min	0.22	0.19	0.18	0.17	0.09	0.09	0.06
instru_quotes_5min	0.23	0.19	0.16	0.17	0.05	0.13	0.07
instru_frag_5sec	0.21	0.21	0.19	0.16	0.09	0.09	0.05
instru_quotes_5sec	0.18	0.21	0.16	0.19	0.07	0.13	0.07
instru_frag_daily	0.23	0.19	0.17	0.14	0.1	0.1	0.07
instru_quotes_daily	0.23	0.17	0.16	0.17	0.08	0.11	0.07

Chapter 5

Conclusion

In this dissertation, I present three essays in empirical finance, focusing on the applications of machine learning in finance. I contribute to the literature by extracting economic variables with machine learning from high-dimensional data (such as speeches and texts). Next, empowered by the economic variables measured by machine learning, I uncover new insights on the behaviors of corporate managers in the U.S. and individual fund investors in China. Finally, I apply machine learning algorithms for non-parametric causal inference in financial datasets, such as high-frequency trading data.

In the first paper, I make several contributions. First, I create a new measure of corporate cultural fit by measuring corporate culture with a state-of-the-art machine learning model, Sentence-BERT. My methodology differentiates from the traditional dictionary-based approach by considering the semantic meaning of complete sentences and avoiding ambiguous out-of-context terms in dictionaries. Second, using survival models, I document a positive (negative) and economically significant impact of cultural fit (cultural distances) on managerial tenure. The effect exists in both proxies for cultural fit – corporate cultural distances (between firms) and personal cultural distances (between managers and firms). Simply put, managers tend to stay longer in firms where they better fit into the corporate culture. Although M&A is one of the important drivers of corporate cultural change, I show that the relationship between cultural fit and managerial tenure is not driven solely by M&A. Cultural fit is related to cultural adaptation, which is deeply rooted in human nature. Third, I employ causal survival forests to show that the effect of cultural fit on managerial tenure is causal. Causal survival forest is a new econometrics tool that allows non-parametric estimations of causal effects when the data is right-censored. My results imply that better (worse) cultural fit is one of the reasons that cause managers to stay longer (shorter) in firms. Specifically, I find that the negative causal effect of bad cultural fit on managerial tenure is exacerbated when executive pay is higher. Therefore, it is important for companies who desire stabilities in the management team to hire managers who fit better culturally. Fourth, I show evidence that firms that hire managers with good cultural fit have higher future market values and performance. Simply put, a good cultural fit is beneficial for firms' future operations and performance. Furthermore, Investors perceive lower cultural dispersion within the firm as a positive signal. A long-short strategy that goes long in the stocks with lower cultural dispersion and shorts the stocks with higher cultural dispersion generates positive returns over the Carhart four-factor model.

In the second paper, we document the existence of a new and previously unstudied type of gender bias - an attention bias away from female fund managers. We show that this flow-performance sensitivity is affected by a differential gender effect. Using a unique sample of individual investor flows into individual funds in China, we provide robust evidence that the investors are more sensitive to the performance of male-managed funds than for female-managed funds. The bias exists across all the return horizons where the platform app allows sorting of fund returns, as well in simple heuristics for performance such as Jensen's alpha and daily average returns. There are also significant cross-sectional differences between investors. Female users appear to display lower levels of gender bias towards female-manged funds. Similarly, users living in smaller cities display stronger levels of gender bias away from female-managed funds. The level of gender bias appears to be innate to investors – an attention bias manifests even in the first set of investments made by a user on the platform. The attention bias uncovered in the sample appears to be irrational and cannot be explained by the difference in performance between male and female managers or the difference in media coverage between male and female managers. Our paper shows that the attention bias works both ways. Though investors appear more sensitive to fund performance when the fund manager is male, the sensitivity is bi-directional. Investors are also less flow-sensitive to underperforming female managers. This may have the desirable impact of lowering the volatility of flows into the fund for mutual fund companies.

In the third paper, we document the significant increase in the fraction of intra-day stock return variations related to market-wide fluctuations following the rise in bot trading in U.S. markets. Much of the market microstructure literature studies changes in price efficiency and price discovery on a stock-by-stock basis. We propose a way to expand microstructure research to the relationship of returns of stocks and the market as a whole. Using changes to index compositions, we provide evidence of a causal relationship between bot trading activities and the synchronization of stock returns across assets. Arguably, index membership is arbitrary, occurs only on paper, has no impact on a firm's operation, and, therefore, firm-specific return synchronization should not be affected by index inclusion or exclusion. As firms get included in an index, they enter the radar of high frequency traders who closely monitor prices in major securities on an ongoing basis. Their monitoring manifests itself in quoting activities that, in turn, cause individual stocks' returns to synchronize more closely with the market. Finally, we provide a methodological innovation by applying new tools from the machine learning literature that help lend further credibility to our analysis and provide a roadmap for future applications.

References

Anderson, Linda E., 1994, A new look at an old construct: Cross-cultural adaptation, *International Journal of Intercultural Relations* 18, 293–328.

Athey, Susan, and Guido Imbens, 2016, Recursive partitioning for heterogeneous causal effects, *Proceedings of the National Academy of Sciences* 113, 7353–7360.

Ben-Rephael, Azi, Zhi Da, and Ryan D. Israelsen, 2017, It Depends on Where You Search: Institutional Investor Attention and Underreaction to News, *The Review of Financial Studies* 30, 3009–3047.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2011, In Search of Attention, *The Journal of Finance* 66, 1461–1499.

Del Guercio, Diane, and Paula A Tkac, 2002, The determinants of the flow of funds of managed portfolios: Mutual funds vs. pension funds, *Journal of Financial and Quantitative Analysis* 37, 523–557.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, 2019, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *arXiv:1810.04805 [cs]* arXiv: 1810.04805.

Grootendorst, Maarten, 2020, BERTopic: Leveraging BERT and c-TF-IDF to create easily interpretable topics., *Zenodo* Version Number: v0.7.0.

Kim, Young Yun, 1988, *Communication and cross-cultural adaptation: An integrative theory.* . Communication and cross-cultural adaptation: An integrative theory. (Multilingual Matters: Clevedon, England) Pages: 223.

———, 2017, Cross-Cultural Adaptation, in *Oxford Research Encyclopedia of Communication* (Oxford University Press).

Niessen-Ruenzi, Alexandra, and Stefan Ruenzi, 2018, Sex Matters: Gender Bias in the Mutual Fund Industry, *Management Science* 65, 3001–3025 Publisher: INFORMS.

Reimers, Nils, and Iryna Gurevych, 2019, Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, *arXiv:1908.10084 [cs]* arXiv: 1908.10084.

Roll, Richard, 1988, R2, *The Journal of Finance* 43, 541–566.

Sirri, Erik R., and Peter Tufano, 1998, Costly Search and Mutual Fund Flows, *The Journal of Finance* 53, 1589–1622 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/0022-1082.00066.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin, 2017, Attention is all you need, *Advances in neural information processing systems* 30.

Wooldridge, Jeffrey M., 2001, *Econometric analysis of cross section and panel data* (MIT Press: Cambridge, Mass).