The London School of Economics and Political Science

Essays in Financial Economics

Olga A. Obizhaeva

Thesis submitted to the Department of Finance of the London School of Economics and Political Science for the degree of *Doctor of Philosophy*

September 2017

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

Statement of conjoint work

I confirm that chapter 2 is jointly co-authored with Albert S. Kyle and Anna A. Obizhaeva. Albert S. Kyle has worked as a consultant for various companies, exchanges, and government agencies. He is a non-executive director of a U.S.-based asset management company. I contributed 33% of the work for chapter 2.

I declare that my thesis consists of 45,943 words.

Acknowledgement

I would like to thank my supervisors Dong Lou and Igor Makarov for their guidance and support. I am grateful to Amil Dasgupta and Ulf Axelson for supervision in my early years at LSE. I benefited from discussions with the faculty of Finance department at LSE. I would also like to acknowledge the financial support from the London School of Economics and the Systemic Risk Centre. I would not be where I am now without support of Pete Kyle, Mark Kritzman, and Anna Obizhaeva who have been inspiring examples of researchers.

The five years at LSE were very special thanks to my cohort: Jesus Gorrin, Michael Punz, Una Savic, and Seyed Seyedan. We outstay the hurrying flight of years, bearing with each other through laughter and through tears. I have been lucky to be surrounded by my fellow PhD students: Lorenzo Bretscher, Svetlana Bryzgalova, Fabrizio Core, Sergei Glebkin, Inna Grinis, James Guo, Andrea Englezu, Lukas Kremens, John Kuong, Dimitris Papadimitriou, Paola Pederzoli, Bernardo Ricca, Gosia Ryduchowska, Petar Sabtchevsky, Su Wang, Yue Yuan, Zizi Zeng.

I am indebted to my parents, Galina and Alexander, my sister Anya, my brother Nikola, and my grandmother Alexandra. You have been there for me, always. My time in London would not be so memorable without all my friends, especially Jacqueline Li.

Abstract

This thesis consists of three essays in financial economics.

The first chapter analyses the fundraising process in the hedge fund industry and the role financial intermediaries play in this process. Using the SEC form D filings, I document that broker-sold funds underperform directly-sold funds by 2% (1.6%) per year on a risk-adjusted basis before (after) fees. Also directly-sold funds, on average, have larger average investor's size, larger minimum investment size, and charge higher performance fees comparing to broker-sold ones. Empirical results are consistent with a stylized model of fundraising. I estimate the model implied average broker's compensation to be $1.5 million per year.

The second chapter (co-authored with Albert S. Kyle and Anna A. Obizhaeva) introduces a new structural model of stock returns generating process. The model assumes that stock prices change in response to buy and sell bets arriving to the market place as predicted by market microstructure invariance. These bets are shredded by traders into sequences of transactions according to some bet-shredding algorithms. Arbitrageurs take advantage of any noticeable returns predictability, and market makers clear the market. This structural model is calibrated to match empirical time-series and cross-sectional patterns of higher moments of returns. We calibrate hard-to-observe parameters of bet-shredding using the method of simulated moments, analyse its properties, and show how much shredding has increased over time.

The third chapter studies cross-sectional and time series variation in the size of repurchase programs. I find that this variation is explained by the variables motivated by market microstructure invariance theory. My results suggest that when determining the size of repurchase programs, managers may target percentage impact costs of these programs or target inventory levels sufficient to allocate their future bets about their companies.

Contents

List of Tables

List of Figures

Chapter 1

Fundraising in the Hedge Fund Industry

High search and due diligence costs due to the opacity of the hedge fund industry make the fundraising process challenging even for hedge funds with a good reputation and a strong track record. Financial intermediaries, such as brokers, consultants, and placement agents, help funds and investors to find one another and to overcome barriers to transact. This paper studies, empirically and theoretically, the role of intermediaries in the fundraising process of hedge funds.

There is yet no consensus about the role and social value of intermediaries. Some people think that intermediation is socially useful. This view is usually justified with several arguments. First, intermediaries may help counterparties find one another and transact, by exploiting their positional advantage and industry knowledge, as per Rubinstein and Wolinsky (1987). Second, intermediaries may help alleviate adverse selection problems, as per Booth and Smith (1986) and Garella (1989). Third, intermediaries may add value by decreasing the costs of making decisions and executions, as per Spulber (2001).

Others think that intermediaries impose unnecessary costs on society. Judge (2014) argues that intermediaries often promote institutional arrangements to maximize their economic rents, and illustrates her point using examples of real estate agents, stock brokers, mutual funds, and exchanges. Warren Buffett opposed and publicly criticized intermediaries on numerous occasions. For example, in 1996 class B shares of Berkshire Hathaway were issued as a response to unit trusts that sold fractional units of Berkshire's shares to small investors.

To analyze empirically the role that financial intermediaries play in the fundraising process of hedge funds, I download and process the entire collection of form D filings that hedge funds report to the U.S. Securities and Exchange Commission ("the SEC") under Regulation D. These filings have information on all third parties involved in the fundraising process. It allows one to identify the hedge funds offered to investors directly and those sold to investors through intermediary brokers.

I match this dataset with the Morningstar hedge funds database using a fuzzy match algorithm. My final dataset combines information on fundraising process, contract characteristics, and performance of hedge funds.

First, I find that, on average, broker-sold funds underperform the directly-sold funds by a substantial margin. Following Fung and Hsieh (2004), I find that broker-sold funds again consistently underperform directly-sold funds by 1.6% on a risk-adjusted basis after accounting for fees. As suggested by Berk and van Binsbergen (2013), the measure of the dollar value added of broker-sold funds is, on average, $210,000 per month lower than that of directly-sold funds.

Second, I construct gross returns series using the modified methodology developed by Brooks, Clare and Motson (2007), Hodder, Jackwerth and Kolokolova (2012), and Kolokolova (2010), and document that broker-sold funds underperform directly-sold funds by 2% per year before fees as well. The pre-fee dollar value added by broker-sold funds is, on average, $190,000 per month lower than that of directly-sold funds. Since pre-fee risk-adjusted performance is a likely indication of skill, this evidence contradicts the view that intermediaries help to identify skillful funds.

Third, I find that, on average, funds sold by brokers charge lower incentive fees compared to funds sold directly, whereas there is no significant difference in terms of management fees.

Fourth, I find that funds sold directly have a larger minimum and average investment size than funds sold by brokers. Regulators define investors who qualify for the accredited investor status based on their income or net worth, suggesting that size is correlated with sophistication of investor. Therefore, this evidence implies that broker-sold funds and directly-sold funds may target different clienteles; directly-sold funds attract, on average, more sophisticated investors than broker-sold funds.

Finally, I analyze heterogeneity of brokers, classifying brokers into in-house and outside brokers based on the similarity of names of a fund and a broker. I find that funds sold by in-house brokers underperform directly-sold funds by 2.1% per year on a risk-adjusted basis after accounting for fees, while funds sold by outside brokers underperform directly-sold funds by 1.4% per year. Funds sold by in-house and outside brokers underperform directly-sold funds by 2% per year on a risk-adjusted basis before accounting for fees. Moreover, funds sold by outside brokers have lower incentive fees than funds sold directly, while the incentive fees of funds sold by in-house brokers do not differ from those of funds sold directly. Funds that are sold through outside brokers have a lower minimum investment sizes than that of directly-sold funds, while the minimum investment sizes of funds sold through in-house brokers do not differ from that of directly-sold funds.

The choice of fundraising channels is an equilibrium outcome; therefore these empirical findings have no causal interpretation, but rather provide an empirical description of an equilibrium. I present a stylized theoretical model of fundraising in the hedge fund industry and show that the implications of the model are consistent with documented empirical findings. The model builds on the work of Nanda, Narayanan and Warther (2000) and Stoughton, Wu and Zechner (2011).

There are two funds that differ in skill: a good fund and a bad fund. Hedge funds do not have their own capital and have to raise funds from outside investors. Since the hedge fund industry is opaque, the process of finding and vetting a suitable fund is costly. To assist with fundraising, a hedge fund may hire an intermediary broker, who will certify the type of the fund and persuade investors to allocate their capital into the fund. Investors differ in their search and due diligence costs.

Sophisticated investors have low search and due diligence costs, others have no industry connections and face high search and due diligence costs. I solve for a separating equilibrium, in which funds endogenously choose portfolio management fees and capital-raising channels, whereas investors decide to invest into hedge funds on their own or based on recommendation of an intermediary.

This equilibrium has a simple intuition. The existence of both types of funds is socially optimal, since both funds generate positive returns, which are greater than the outside option. Some investors, however, are not able to invest in the hedge fund industry without financial advice. Only sophisticated investors can find the good fund, while other investors with high due-diligence costs are not able to do so. The broker steps in to resolve this inefficiency. The broker is able to lower the costs of investors by internalizing the due-diligence process and this allows the high-cost investors to allocate their endowments into the hedge fund industry. In return, the broker requires compensation. The bargaining power of the broker and the relative outperformance of the good fund over the bad fund are crucial for the existence of a separating equilibrium. The good fund separates from the bad fund when it generates a sufficiently high return that is enough to compensate for investors' search and due diligence costs. Investors in the good directly-sold fund get higher after-fee returns compared to the after-fee returns of the investors in the bad broker-sold fund, regardless of the fee that the bad broker-sold fund charges.

I calibrate the model and estimate the implied average compensation that brokers receive for their capital introduction services. I assume that the compensation of a broker is proportional to the total dollar fees that a hedge fund collects from its investors. I estimate the total dollar fees using data on the assets under management, the performance, and the compensation structure of the hedge fund. Assuming that the bargaining power of the broker equals to 1/3, which corresponds to the equilateral division of the surplus among the fund, the investors, and the broker, I find the average annual compensation of the broker to be equal to \$1.5 millions. This is roughly consistent with the annualized estimated difference between the dollar value added by broker-sold funds and directly-sold funds in the data.

The paper is related to several strands of literature. It contributes to the literature on capital formation. Duffie (2010) discusses the problem of slow movement of investment capital to trading opportunities and its implications for asset price dynamics. Berk and Green (2004), Garleanu and Pedersen (2016), Vayanos (2004), Pastor and Stambaugh (2012), and Vayanos and Woolley (2013) model the asset management industry theoretically. There is also an extensive empirical literature that studies capital formation in the asset management industry. Chevalier and Ellison (1997) and Sirri and Tufano (1998) find that investors allocate their capital into mutual funds with a positive past performance and flee from mutual funds with negative past returns. The hedge fund literature also finds that the performance of funds is an important factor that affects capital flows. For example, Goetzmann, Ingersoll and Ross (2003) and Fung et al. (2008) find that alpha generating hedge funds experience larger capital inflows than funds that do not have alpha. Horst and Salganik-Shoshan (2014) find that capital flows to the highest performing strategies and to the better performing funds within the strategy. Baquero and Verbeek (2015) document that funds with a longer positive track record get more capital. Lu, Musto and Ray (2013) study the indirect advertising of hedge funds and find that it helps to attract capital. Baquero and Verbeek (2009) use a regime-switching model, while Jorion and Schwarz (2015) use form D filings to separate fund inflows and outflows and analyze flows to performance relationship. Getmansky (2002) studies the life-cycles of hedge funds at the individual and strategy level and finds that age, assets under management and the standard deviation of returns negatively affects fund flows. Joenväärä, Kosowski and Tolonen (2013), Getmansky et al. (2015), and Aiken, Clifford and Ellis (2015) analyze the effect of liquidity restrictions on capital flows. My paper contributes to this literature by analyzing capital formation in the hedge fund industry and the role that intermediaries play in this process.

This paper is also related to studies on distribution channels and marketing in the asset management industry. Investors pay substantial amounts of money in the form of sales loads and broker commissions. This raises the questions of why they pay such high fees to intermediaries and what benefits these investors get in return. Bergstresser, Chalmers and Tufano (2009) and Del Guercio and Reuter (2014) find that mutual funds sold by brokers significantly underperform funds sold directly (both before and after fees). Possible explanations include the substantial intangible benefits that brokers provide and the partition of mutual fund clientele into sophisticated and disadvantaged investors. As opposed to mutual fund retail investors, hedge funds investors are usually sophisticated financial institutions and individuals qualified for accredited investor status. It may be understandable to find evidence of underperformance in broker-sold mutual funds, but it is more surprising to find the same result in a hedge funds setting. The authors also document that directly-sold mutual funds charge lower fees than mutual funds sold through brokers. I find the opposite result for the incentive fees of hedge funds, while I find no difference in hedge funds' management fees across different fundraising channels. Furthermore, Christoffersen, Evans and Musto (2013) establish that underperformance of broker-sold funds mostly arises in mutual funds that are sold through outside brokers rather than in-house brokers. The authors also document that in-house brokers receive a higher front load comparing to outside brokers. In contrast, I find that hedge funds offered through in-house brokers underperform both directly sold funds and funds sold through outside brokers. Also, hedge funds sold through in-house brokers charge higher incentive fees than funds sold through outside brokers.

The empirical analysis of this paper is closely related to that of Agarwal, Nanda and Ray (2013). The authors find that hedge funds that are selected by institutions investing directly outperform hedge funds that are selected by institutions that use advisory services. They analyze raw and style-adjusted after-fee performance of hedge fund investments aggregated at the level of hedge fund family, while granularity of data in my study allows to perform analysis at the individual fund level.

The theoretical part of the paper is related to the work of Stoughton, Wu and Zechner (2011), who model the interaction of active portfolio manager, financial adviser, and investors under various settings. Similar to their model, investors' choice of performing due diligence on their own or delegating it to the broker depends on their due diligence costs, but I emphasize the endogeneity of the choice of capital raising channels by hedge funds.

The rest of the paper is organized as follows. Section I describes the data and outlines the key economic variables that are used in the analysis. Section II documents the empirical findings on the fundraising process of hedge funds. Section III presents a simple model of fundraising that reconciles the empirical findings and estimates the model-implied compensation that intermediaries receive for capital introduction services. Finally, section IV concludes the discussion.

1.1 Data

I use a combination of two databases. The first database is constructed from form D filings. The second is a Morningstar hedge funds database. Additional data is downloaded from Thomson Reuters and the David A. Hsieh Data Library.

1.1.1 Form D filings

Although hedge funds qualify for exemptions to formal registration of fundraising offerings, the Securities Act of 1933 requires all funds that raise capital from investors (with at least one U.S. investor) to notify regulators about the fundraising process by filing a form D with the SEC. A fund is required to file a notice no later than 15 calendar days after the *date of the first sale* of the fund's offering. As long as the fund remains open, it is required to update filings on an annual basis as well as in the case of detected mistakes in the previous filings.[1]

Table 1.1 summarizes all the data fields in the form D. Fund reports administrative information and information about its fundraising process: its name, the address of its principal place of business, the names and addresses of the executive officers, the amount of capital raised, the number and types of investors, and each

[1]See detailed information about offering exemptions in Rules 504, 505, and 506 of Regulation D. Source: Sections 230.501 through 230.506 appear at 47 FR 11262, Mar. 16, 1982. Note that amendment to form D filing is denoted as D/A. Hereto, I refer to both initial form D notice and its amendments as form D filings. Compliance guide about filing and amending a Form D notice may be found at https://www.sec.gov/info/smallbus/secg/formdguide.htm.

person who is paid directly or indirectly in connection with the fundraising process. The information that funds disclose in Form D filings must be free of biases, since misreporting and failure to comply with the SEC requirements imposes significant reputational and legal risks and may result in criminal penalties.

Form D filings are publicly available. I download and process all the electronic form D filings from the SEC's Electronic Data Gathering, Analysis, and Retrieval system (EDGAR).[2] I start in January 2010, when all hedge funds were required to submit forms electronically. Thus, the downloaded sample covers period from January 2010 to December 2016.

Each fund in the EDGAR system is uniquely identified by its *Central Index Key* or *CIK number*. Thus, by knowing the name of the fund or its CIK number, one gets access to information about its fundraising. For example, a search for Citadel Global Equities fund will produce ten form D filings over the period from July 2009 to September 2016. From the filings, we learn that the fund was originated with Citadel Advisors in July 2009. The fund raised $100 millions from one investor at the origination date. Then, it raised $153 millions from seven investors by August 2010 and $446 millions from fifty-nine investors by September 2016.

In imposing strict standards on the marketing of hedge funds, the SEC requires funds to disclose in their form D filings information about any entity which is directly or indirectly compensated for advertising and offering a fund to investors. This information allows one to differentiate between the funds sold to investors by brokers and the funds offered to investors directly.[3] The disclosed information consists of brokers' biographical information, their Central Registration Depository ("CRD") number within the Financial Industry Regulatory Authority ("FINRA") system and the list of states in which they advertise offerings. For example, I classify Citadel Global Equities Fund as a directly-sold fund, since it does not employ any intermediary in the fundraising process, while Renaissance Institutional Equities Fund is an example of a broker-sold fund, since it is sold to clients by Renaissance Institutional

[2]The EDGAR depository is accessible at https://www.sec.gov/edgar/searchedgar/webusers.htm.

[3]Hedge funds report information about intermediary brokers that are involved in fundraising process in Item 12 of form D filings, *Sales Compensation*

Management LLC.

Table 1.2 displays the largest open directly-sold and broker-sold funds in 2015. For example, Medallion fund of Renaissance Technologies raised $6.5 billions by 2015, while D.E. Shaw Oculus International fund of D.E. Shaw & Co that raised $13 billions with the help of broker.

I classify intermediary brokers into *in-house brokers* and *outside brokers* based on the similarity of the names of the fund and the broker. For example, Fortress Convex Asia fund LP uses the capital introduction services of Fortress Capital Formation LLC. In this case, I classify Fortress Capital Formation LLC as an in-house broker. ING Clarion Market Neutral LP is sold by Citigroup Global Markets and Merrill Lynch, Pierce, Fenner and Smith Inc. In this case, I classify both brokers as outside brokers. Funds are classified as being sold by in-house brokers when they employ only in-house brokers. If a fund is sold by outside brokers, I refer to such fund as outside broker-sold fund. Thus, Fortress Convex Asia fund LP is classified as an in-house broker-sold fund and ING Clarion Market Neutral LP is classified as an outside broker-sold fund.

Table 1.3 displays ten broker firms in the capital introduction business which market the largest number of hedge funds. The list of the top brokers in this business comprises top investment banks such as Goldman Sachs, Morgan Stanley, and J.P. Morgan. For example, over the considered period, Goldman Sachs intermediates as many as 377 hedge funds. The average (median) amount of capital raised by funds that are intermediated by Goldman Sachs is $350 millions ($98 millions). The average (median) number of investors in funds that are intermediated by Goldman Sachs is 149 (30) investors. According to anecdotal evidence, big broker firms often offer their wealthy clients opportunities to invest in hedge funds through online platforms without having to go to the funds themselves.

Figure 1.3 shows the fundraising dynamics over the period from January 2010 to December 2015 comparing hedge fund industry with other alternative investments. I analyze four main alternative investment business types: hedge funds, private equity, venture capital and other investment funds, which includes fund of funds, commodity trading advisors(“CTAs”) and commodity trading operators (“CTOs”). Figure 1.3 is split into four panels. Panels A, B, C and D display hedge funds, other investment funds, private equity funds, and venture capital funds, respectively. Focusing on the difference between the fundraising channels, the figure visualizes the amount of capital that was raised by directly-sold and broker-sold funds over the considered period.

To estimate the amount of capital inflows, I use reported information on the *Total Amount Sold* that the fund reports in form D filings. I consider two cases: capital inflows at the fund’s inception and capital inflows during the life of the fund. In the first case, the amount of capital raised at inception is directly reported in the Total Amount Sold variable. In the second case, it may be estimated as an increment of the Total Amount Sold variable between two consecutive fund’s filings. I outline the methodology on capital inflows estimation in Appendix.

The hedge fund industry enjoyed capital inflows which steadily grew from 2010 to 2015, spiking above the average level in 2014 and recovering to the previous trend of inflows at $300 billions per year. The spike in capital inflows in 2014 coincides with the lifting of the SEC’s advertisement ban, which was implemented in September 2013, following the JOBS Act directive.

1.1.2 Morningstar database and risk-adjusted returns

I use the Morningstar CISDM hedge fund database available from Wharton Research Data Service (“WRDS”). The database contains fund-level information on live and liquidated hedge funds. It keeps the most recent snapshot of fund’s administrative information, such as name, address, inception date, compensation structure, minimum investment size, and liquidity restrictions. It also records the fundâĂŹs after-fee performance and assets under management at a monthly frequency.

I use Morningstar data to estimate the performance and skill of the hedge fund. Hedge funds usually employ various risky trading strategies. Thus, to make a sensible comparison of hedge funds, I control for their exposure to systematic risk factors and calculate their alphas. I estimate the tradable alpha regressing the annualized monthly excess return, R^e_{it}, on seven tradable risk factors, as suggested by Fung and Hsieh (2004):

$$\begin{aligned}R^e_{it} = \alpha_i + \beta_{Mkt} \cdot SNPMRF_t + \beta_{SmB} \cdot SMB_t + \beta_{T10y} \cdot BD10RET_t + \\ \beta_{Cr.Spr.} \cdot BAAMTSY_t + \beta_{pBD} \cdot PTFSBD_t + \beta_{pFX} \cdot PTFSFX_t + \\ \beta_{pCOM} \cdot PTFSCOM_t + \tilde{\epsilon}_{it}.\end{aligned} \tag{1.1}$$

To account for market exposure, I use annualized returns on the S&P500 index, $SNPMRF_t$. Adjusting for exposure to the size factor, I use an annualized return spread between the Russell 2000 and the S&P500 index, SMB_t, obtaining a time series for the Russell 2000 and the S&P500 indexes from Thomson Reuters Datastream.

To control for yield curve exposure, I follow the literature and use the annualized excess returns of the U.S. 10-year Treasury constant maturity bond, $BD10RET_t$. A tradable yield curve level factor that is used in this paper is Bank of America Merrill Lynch's U.S. 10-year Treasury constant maturity bond returns, which I download from Thomson Reuters Datastream. As a robustness check I used 10-year discount factors from the Federal Reserve Bank of St.Louis' Treasury yield curve estimates.[4] The correlation between the two time series is 0.96.

Accounting for credit spread exposure, I use an annualized return spread between Moody's Baa-rated corporate bond, $BAAMTSY_t$, and the U.S. 10-year Treasury constant maturity bond. To proxy Moody's Baa-rated corporate bond, I use the tradable Barclays Long Baa U.S. Corporate index, which can be downloaded from Thomson Reuters Datastream.

Finally, adjusting for the dynamic nature of the hedge funds' strategies, I follow Fung and Hsieh (2004) and use a trend-following bond factor, $PTFSBD_t$, a trend-following currency factor, $PTFSFX_t$, and a trend-following commodity fac-

[4]FED's yield curve can be downloaded from Federal Reserve Economic Data (FRED): http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html.

tor, $PTFSCOM_t$, which are constructed from look-back options and can be downloaded from David A. Hsieh's Data Library.[5]

For every fund i in month t, I estimate its annualized monthly alpha, $\hat{\alpha}_{it}$, with a two-year rolling-window regression (1.1). The final sample consists of 29,051 fund-month observations.

Although, investors care about after-fee returns on their hedge fund investment, skills of funds are reflected in pre-fee returns. Hedge fund databases usually take the perspective of investors and report fund performance and net asset values ("NAV") after accounting for fees. To reconstruct pre-fee returns, I apply the modification of methodology that was used in Brooks, Clare and Motson (2007), Hodder, Jackwerth and Kolokolova (2012), and Kolokolova (2010)

I make several assumptions that reflect the general practice on the calculation of hedge funds' fees. [1] Pro-rata management fees are paid at the end of the month on pre-fee net asset value at the end of the month. [2] Incentive fees are accrued on a monthly basis, but are only paid at the end of the calendar year; reported after-fee net asset value and performance account for accrued incentive fees. [3] Hedge funds use the high-watermark provision and incentive fees are paid in case pre-fee net asset value adjusted for management fees are above the current high water mark. [4] The high-water mark is reset to a pre-fee net asset value if it exceeds the current high water mark; otherwise the high-water-mark stays as in the previous month. [5] Management and incentive fees remain constant over time.[6] [6] The equalisation credit/contingent redemption scheme is used to calculate net asset value to ensure that the fund managers are compensated correctly for positive performance, while investors, who might invest in funds at different time are treated fairly and equally.[7]

For each fund I estimate the pre-fee net asset value, $NAV^*(t)$, and the pre-fee return, $R^*(t)$, using available data on after-fee net asset value, $NAV(t)$, after-fee return, $R(t)$, management fee (in percentage terms), f_M, and incentive fee (in

[5]David A. Hsieh's Data Library is accessible at https://faculty.fuqua.duke.edu/ dah7/HFRFData.htm.

[6]In reality hedge funds may update their compensation structure as documented by Deuskar et al. (2011), Agarwal and Ray (2012) and Schwarz (2007).

[7]âĂŸEqualisation Credit/Contingent RedemptionâĂŹ accounting procedure is described and discussed in McDonnell (2003).

percentage terms), f_I.

The hedge fund database reports after-fee net asset value, which is calculated as a pre-fee net asset value adjusted for management fees (in dollars), $F_M(t)$, and accrued incentive fees (in dollars), $F_I(t)$:

$$NAV(t) = NAV^*(t) - F_M(t) - F_I(t). \tag{1.2}$$

Dollar management fees are calculated based on the net assets of the fund at the end of the month, as per assumption [1]:

$$F_M(t) = NAV^*(t) \cdot f_M/12. \tag{1.3}$$

Incentive fees accrue if the net asset value after management fees and net capital flows are above the high water mark, following assumptions [2], [3], and [4]:

$$F_I(t) = \max(0; NAV^*(t) - F_M(t) - \text{Netflows}(t) - \text{HWM}(t)) \cdot f_I. \tag{1.4}$$

Solving the system of equations (1.2), (1.3), and (1.4), I express the pre-fee net asset value, dollar management fees, and the dollar incentive fees

$$\begin{cases} NAV^*(t) = NAV(t) + F_M(t) + F_I(t) & (1.5) \\ F_M(t) = [NAV(t) + F_I(t)] \cdot \dfrac{f_M/12}{1 - f_M/12} & (1.6) \\ F_I(t) = [NAV(t) - \text{Netflows}(t) - \text{HWM}(t)] \cdot \dfrac{f_I}{1 - f_I} \cdot & \\ \qquad \mathbb{I}\left[NAV(t) - \text{Netflows}(t) > \text{HWM}(t)\right] & (1.7) \end{cases}$$

Dollar incentive fees (1.7) are accumulated only if the assets of the fund are above the high water mark, $NAV(t) - \text{Netflows}(t) > \text{HWM}(t)$; otherwise, the fund does not get any incentive fees.

Finally, I estimate the pre-fee return, $R^*(t)$, as a growth rate between the pre-fee assets under management at the beginning of the month and the pre-fee assets under management at the end of the month, adjusted for dollar netflows during the period:

$$1 + R^*(t) = \frac{NAV^*(t) - \text{Netflows}(t)}{NAV^*(t-1) - F_M(t-1)}. \tag{1.8}$$

At the beginning of the investment period, assets under management are equal to pre-fee net assets at the end of the previous period adjusted for management fees. Also, the pre-fee net asset value has to be adjusted for netflows, which I estimate as in the literature on fund flows (Sirri and Tufano (1998), Agarwal, Daniel and Naik (2004)).

$$\text{Netflows}(t) = NAV(t) - NAV(t-1) \cdot (1 + R(t)). \tag{1.9}$$

Finally, Substituting (1.2) and (1.9) into (1.8), I estimate the pre-fee return $R^*(t)$.

1.1.3 Matching form D filings and Morningstar database

I match the form D filings with Morningstar database by the name of the fund using a fuzzy matching method.

First, I estimate the pairwise generalization of Levenshtein (1966) edit distance, a measure of dissimilarity, between the funds in Form D and Morningstar databases. I eliminate the pairs that have a dissimilarity score above 200. Second, I eliminate pairs of matched form D and Morningstar funds that report inception dates of more than six months apart from each other. Finally, I manually verify the results of the matching procedure.

The matched sample consists of 1,728 individual funds that in total submitted 7,824 form D filings. It represents 15% of Reg D funds and 8% of funds that are listed in the Morningstar database. Among the matched funds 92% of funds are identified as hedge funds and 8% of funds are identified as other investment funds. A low match rate is explained by the fact that the universe of Reg D funds consists only of funds that are open for investment and have at least one US investor. Additionally, not all form D funds may choose to be listed in Morningstar database.

Jorion and Schwarz (2015) are able to match in total 3,816 form D funds with 14,581 form D filings, using the Hedge Fund Research (HFR) and Lipper TASS databases. The match rate between the form D funds and Morningstar funds is consistent with the match rates of form D funds with hedge funds that report to TASS (1,896 funds).

In the matched sample there are 1,103 of directly-sold funds and 625 of broker-sold funds.

Focusing on the heterogeneity of brokers, I further differentiate the broker-sold funds into funds that are offered to investors through in-house brokers and funds that are sold to by outside brokers. In the matched sample of broker-sold funds I identify in total 537 funds that are sold by outside brokers, 56 fund that are sold by in-house brokers and 32 funds that are sold through both.

The matched database inherits all the biases that are usually associated with Morningstar database.

First, the information that hedge funds report to Morningstar database is not verifiable. Fund managers usually list their funds in hedge fund databases to market their funds and attract potential investors. Agarwal, Mullally and Naik (2015) and Getmansky, Lee and Lo (2015) provide a comprehensive review of the limitations and potential biases in hedge fund data.

Often funds backfill returns prior to the date when they starts reporting to the data vendor. Thus, a fund manager has an incentive to list his hedge fund in a database after a period of good performance. As discussed in Edwards and Park (1996), this potentially leads to misleadingly good track records and may result in upward bias in expected returns due to this instant history or backfill bias.

Joenväärä, Kosowski and Tolonen (2014) estimate a backfill bias of around twenty months by analyzing snapshots of databases that have been taken on different dates. Following the literature practice, I exclude the first twenty-four months of returns observations since the inception of the funds to mitigate this bias.

Second, there is also survivorship bias. Funds have an incentive to stop reporting their performance after a period of bad performance. Therefore, underperforming funds may be under-represented, again biasing upwards the estimates of expected returns. To mitigate this bias, I consider both live and defunct funds moved to hedge fund graveyard files.

Third, Morningstar hedge fund data, unfortunately, contains significant numbers of missing assets under management. Following Joenväärä, Kosowski and Tolonen (2014), I fill in any missing observations with the most recent observations of the past.

Table 1.4 presents summary statistics on annual capital inflows, the number of investors, and the number of new investors across funds that are directly sold to investors and funds that are offered to investors through brokers from form D filings. Panel A presents the summary statistics for the whole sample of form D funds. Panel B presents summary statistics for the matched sample in order to examine any potential biases introduced by the matching procedure.

Annual capital inflows into hedge funds do not differ significantly across distribution channels. On average, directly-sold funds and broker-sold funds raise $49 millions per year. The median amount of capital raised by directly-sold funds is $3 millions and $5 millions for broker-sold funds. There are on average 12 investors in directly-sold funds and 33 investors in broker-sold funds. The average size of investment in a broker-sold fund is 2.75 less than that of a directly-sold fund.

I do not find significant differences between the matched sample and the total form D sample of funds, comparing a sample that consists of matched funds and sample of all form D funds on their observable characteristics.

1.2 Empirical evidence

This section provides an empirical description of the fundraising process of the hedge funds, focusing on the differences between "direct" and "brokered" distributions.

1.2.1 After-fee performance across distribution channels

To compare the performance of funds between fundraising channels, I construct two portfolios of funds. The first one consists of directly-sold funds, representing the anti-intermediation view. The second one comprises hedge funds that are offered to investors through brokers, representing the pro-intermediation view. The portfolios of funds are rebalanced monthly, so that newly originated funds are included and liquidated funds are excluded appropriately. Assuming an initial investment of $100, I track the portfolios of the funds' after-fee performance from January 2010 to December 2015.

Figure 1.4 plots the after-fee performance dynamics for the portfolios of funds. Panel A shows the performance of the portfolio of funds where the constituent funds are equally-weighted. Panel B displays the performance of portfolios of funds where the constituent funds are value-weighted. Portfolio of directly sold funds outperforms portfolio of broker sold funds over considered five year period. For the equally-weighted scheme, the portfolio of directly-sold funds increases from $100 to $130, with an annualized return of 5.38% per year over five years, while the portfolio of broker-sold funds rises from $100 to $125, with an annualized return of 4.56% per year. The difference is more pronounced when the value-weighted scheme is considered. Portfolio of directly sold funds increases from $100 to $136 with annualized return of 6.34% per year, while portfolio of broker sold funds increases from $100 to $126 with annualized return of 4.73% per year. The results also hold when I consider the full sample of hedge fund returns without adjusting for backfill bias. I present the results in Figure 1.7 in the Appendix.

Investors, however, should care about risk-adjusted returns. I estimate two-year rolling alpha of the portfolios of funds, adjusting performance for systemic risk exposure using equation (1.1). Figures 1.5 presents the time-series dynamics of the after-fee alphas of the portfolio of directly-sold funds and the portfolio of broker-sold funds. The figure is split into two sub-figures, which correspond to the equally-weighted scheme in Panel A and the value-weighted scheme in Panel B. The after-fee alpha of directly-sold hedge funds is persistently higher than the after-fee alpha of the broker-sold hedge funds regardless of portfolio-weighting scheme. For the equally-weighted scheme, the after-fee alpha of the directly-sold hedge funds is equal on average to 4.42% per year versus 3.37% per year for the broker-sold hedge funds. For the value-weighted scheme, the average alpha of the portfolio of directly-sold funds is equal to 4.43% as opposed to 3.55% for the portfolio of broker-sold funds.

I implement another robustness check and perform panel data analysis. For each hedge fund i in month t, I estimate its annualized monthly alpha, $\hat{\alpha}_{it}$, with a two year rolling-window regression (1.1). Then I estimate the difference between the alphas of the directly-sold funds and the broker-sold funds with a panel regression

$$\hat{\alpha}_{it} = \beta_0 + \beta_B \cdot B_{it} + \beta_X \cdot X_{it-1} + \beta_t + \tilde{\epsilon}_{it}, \tag{1.10}$$

where B_{it} is a dummy variable that is equal to one if fund i is sold through brokers and it is equal to zero if the fund raises capital directly. I use a set of controls, X_{it-1}, which includes the assets under management of hedge fund in a previous month, the age of the fund, and its vintage. I also control for aggregate demand shocks with time fixed effects, β_t. The coefficient of interest that measures the difference in the alphas of directly-sold and broker sold-funds is β_B.

Panel A of Table 1.7 presents the results of the estimation of regression (1.10). I find that the after-fee alpha of the broker-sold funds is, on average, 1.6% per year lower than that of directly-sold funds. The results are economically significant and robust for inclusion of the fund's size, age, vintage year controls and time fixed effects. I also find consistent results (reported in Appendix Table 1.16) for the sample of funds without correction for backfill bias.

I also compare the dollar value added measure of Berk and van Binsbergen (2013) for directly-sold funds and that of broker-sold funds. I find monthly dollar value added to investors, $\hat{S}_{it}$, as a product of the after-fee alpha of the hedge fund and its assets under management in a given month. I perform panel data analysis and report results in Panel A of Table 1.8. I find that investors in the broker-sold fund receive, on average, \$210,000 per month less than investors in directly-sold funds. The results are robust when controlling for the age of the fund, its vintage and time fixed effects.

Exploiting heterogeneity across brokers, I analyze the difference in performance between funds that are sold by in-house brokers and funds that are offered by outside brokers. I perform a formal analysis with the following panel regression:

$$Y_{it} = \beta_0 + \beta_I \cdot B^I_{it} + \beta_O \cdot B^O_{it} + \beta_X \cdot X_{it} + \beta_t + \tilde{\epsilon}_{it}, \tag{1.11}$$

where $Y_{it} = \hat{\alpha}_{it}$ denotes the fund's annualized risk-adjusted performance. B^I_{it} is a dummy variable that is equal to one when the fund is offered to investors by an in-house broker and is equal to zero otherwise. B^O_{it} is a dummy variable that is equal to one when the fund is sold to investors through an outside broker and is equal to zero otherwise.

Table 1.9 displays the results of the estimation of regression (1.11). I find that the result of the under-performance of broker-sold funds is mostly driven by funds that are sold through in-house brokers. The average after-fee alpha of funds that are sold through in-house brokers is 2% lower than that of directly-sold funds, while average after-fee alphas of funds that are offered through outside brokers is 1.4% lower than that of directly sold funds. Performing a formal F-test and comparing the difference between in-house broker-sold and outside broker-sold funds, I find that the alpha of funds that are sold by in-house brokers is statistically different from the alpha of funds that are sold by outside brokers. The results are robust when the fund's size, vintage, and year-month controls. Furthermore, I perform additional robustness checks by estimating the regression (1.11) on the sample that does not correct for backfill bias, which is displayed in Table 1.18 in the Appendix.

The above findings on the underperformance of broker-sold hedge funds relative to directly-sold funds are consistent with the findings in the mutual funds literature. Bergstresser, Chalmers and Tufano (2009) were the first to establish that broker-sold mutual funds, with an average after-fee alpha of -2.28% per year, underperform directly-sold mutual funds, with an average after-fee alpha of -1.07% per year, by 1.21% per year. Del Guercio and Reuter (2014) and Reuter (2015) find similar results when considering different weighting schemes. Authors document the difference in equally-weighted after-fee alphas between the two groups of funds of 1.15% and that of the value-weighted after-fee alphas 0.64% per year. Christoffersen, Evans and Musto (2013) find that a 1% increase in the excess load paid to broker decreases mutual fund after-fee future performance by 0.24% over the next year. In contrast to my results, the authors find that the underperformance is mostly driven by mutual funds that are sold through outside brokers rather than in-house brokers.[8]

1.2.2 Pre-fee performance across distribution channels

Addressing the question of whether brokers help to identify skillfull hedge funds, I analyze the pre-fee risk-adjusted performance of funds across distribution channels. I estimate the two-year rolling pre-fee alpha of portfolios of funds, adjusting their pre-fee returns for systemic risk exposure using equation (1.1). Figures 1.6 presents the time-series dynamics of the pre-fee alphas of the portfolio of directly-sold funds and the portfolio of broker-sold funds. The figure is split into two sub-figures, which correspond to the equally-weighted scheme in Panel A and the value-weighted scheme in Panel B.

The pre-fee alpha of the portfolio of directly-sold hedge funds is persistently higher than the pre-fee alpha of the portfolio of broker-sold hedge funds regardless of the portfolio-weighting scheme. I find that for the equally-weighted scheme, the alpha of the portfolio of directly-sold hedge funds is equal, on average, to 5.78% versus 4.48% per year for the portfolio of broker-sold funds. For the value-weighted scheme, the average alpha of directly-sold funds is equal to 5.53% versus 4.95% for the broker-sold funds.

I implement another robustness check and compare the skill of the funds across

[8] Christoffersen, Evans and Musto (2013) refer to outside brokers as non-affiliated brokers and in-house brokers as captive brokers.

distribution channels with a panel regression (1.10). Panel B of Table 1.7 presents the estimation results of the panel regression. I find that the funds that are sold to investors through brokers underperform funds that are offered to investors directly by 2% per year before accounting for fees. The results are robust for the inclusion of fund-level controls and time fixed effects. I perform a robustness check, using sample without adjusting for backfill bias and find consistent results reported in Panel B of Table 1.16 in the Appendix.

I also compare the dollar value added by directly-sold hedge funds and broker-sold hedge funds. I find the monthly dollar value added of the hedge fund as a product of the pre-fee alpha of the hedge fund and its assets under management in a given month. The dollar value added measure estimates the amount of money that the hedge fund extracts from the financial markets. I perform a panel data analysis and report the results in Panel B of Table 1.8. I estimate that the value added by a broker-sold fund is, on average, $190,000 per month lower than the value added by a directly-sold hedge fund. The result is robust in controlling for the age of the fund, its vintage and the time fixed effects.

Next, analyzing heterogeneity across brokers, I study the difference in skill between funds that are sold by in-house brokers and funds that are offered by outside brokers. Table 1.10 displays the estimation of the regression (1.11). I find that hedge funds that are offered by in-house brokers, on average, have the same pre-fee alpha as hedge funds that are sold through non-affiliated brokers and underperform directly-sold hedge funds by 2% per year. The results are robust for the inclusion of the size of the fund and its vintage year and controlling for time-variant demand shocks. Furthermore, I perform an additional robustness checks by the estimating regression (1.11) on the sample that does not correct for backfill bias and find similar results, which I report in Table 1.19 in the Appendix.

1.2.3 Fees across distribution channels

Next, I assess whether intermediaries help investors to find funds that charge lower fees. To answer this question, I use information about management fees and incentive fees that hedge funds report in Morningstar database. Since only the most recent contract characteristics are kept in the database, I perform a formal comparison using the following cross-sectional regression:

$$Y_i = \beta_0 + \beta_B \cdot B_i + \lambda_t + \tilde{\epsilon}_i, \tag{1.12}$$

where B_i is a dummy variable that is equal to one when fund is broker-sold and is equal to zero otherwise. The regression includes a control for the fund's vintage year, λ_t.

Table 1.11 compares the fees of hedge funds across the distribution channels. Columns (1) and (2) estimates the difference in the management fees of broker-sold and directly-sold hedge funds. On average, hedge funds charge their investors 1.4% management fees, but I do not find any significant difference between funds with different distribution channels. I also do not find any significant difference between the management fees that funds sold through in-house brokers and funds offered through outside brokers charge their investors. These results are not surprising since hedge funds uses management fees to cover their operational expenses.

Next, I estimate the difference in incentive fees that directly-sold funds and broker-sold funds charge their investors and present the results in columns (3) and (4). I find that directly-sold funds, on average, charge a incentive fee of 18.35%, which is 1.4% higher than the incentive fee of broker-sold funds. Analyzing the heterogeneity of broker-sold funds, I establish that funds that are sold by outside brokers charge incentive fees that are, on average, 1.5% lower than fees that directly-sold funds charge, while funds that are sold by in-house brokers charge the same incentive fees as directly sold funds. Performing an F-test, I find that the incentive fee that funds sold by in-house brokers charge are significantly different from the incentive fees that funds sold by outside brokers charge.

The above results differ from the findings of the mutual fund literature. Bergstresser, Chalmers and Tufano (2009) establish that the non-distributional expenses of mutual funds that are sold through intermediaries are 23 basis points higher than those of mutual funds that are sold to investors directly, concluding that brokers do not help investors to identify mutual funds with lower non-distribution fees.

1.2.4 Clientele across distribution channels

I complete the empirical analysis by analyzing whether investors of broker-sold hedge funds differ from investors of directly-sold hedge funds. Since hedge funds are very secretive and do not disclose information about their investors, I use a minimum investment size and an average investment size as empirical proxies of the size of the hedge fund's marginal investor and average investor. To estimate the difference in the hedge funds' clientele across the distribution channels, I estimate a regression (1.12).

Columns (1) and (2) of Table 1.12 estimate the difference in the minimum investment size of broker-sold and directly-sold hedge funds. The minimum investment size of directly-sold funds is, on average, \$1 million, which is \$0.27 millions more than that of directly-sold funds. Further, analyzing the heterogeneity of brokers, I find that the minimum investment size of funds sold through in-house brokers does not differ from that of directly-sold funds, while the minimum investment size of funds sold through outside brokers is \$0.21 millions lower than that of directly-sold funds. Performing an F-test, I find that the minimum investment size of in-house broker-sold funds is statistically different from the minimum investment size of outside broker-sold funds.

Columns (3) and (4) of Table 1.12 estimate the difference in the average investment size of broker-sold and directly-sold hedge funds. Comparing the average investment size, I find that broker-sold funds have a \$12 millions lower average investment size than directly-sold funds.

These findings suggest that funds may target a different clientele.

1.3 Theoretical motivation

I presents a simple model of fundraising in the hedge fund industry. I then reconcile empirical findings with the model implications and estimate the compensation that brokers receive for capital introduction services.

1.3.1 Model of fundraising

Suppose there are three types of agents: hedge funds, investors, and brokers, who intermediate between hedge funds and investors. There are two risk-neutral funds that differ in their portfolio management skills: a good fund and a bad fund. Let θ denote a type of fund, where $\theta \in \{\mathrm{G}, \mathrm{B}\}$ corresponds to the good fund and the bad fund, respectively. The good and the bad funds deliver positive pre-fee risk-adjusted returns, α_G and α_B, respectively, with $\alpha_G > \alpha_B > 0$. I assume that alphas are known to the funds themselves, but unobservable to investors and the broker.

The fund does not have capital and has to raise it from investors. It can either directly raise capital from investors or use capital introduction services offered by the broker. For its portfolio management services, the fund charges performance-based fees, which are calculated as the fraction of generated profits. The fund chooses a fee and capital raising channel to maximize the total dollar fees that it collects from its investors.

There is also a continuum of risk-neutral investors. Each investor is endowed with a unit of capital, which he may either invested in one of the hedge funds or in an outside option (return of the outside option is normalized to zero). All investors qualify for the status of accredited investor and may invest in hedge funds. To capture heterogeneity among clientele, I assume that investors differ in their search and due diligence costs. There are professional investors with low search and due diligence costs and mainstream accredited investors who have high search and due diligence costs. I assume that the search and due diligence costs of investors, c, are uniformly distributed at interval from 0 to $\bar{C}$, $c \sim U[0; \bar{C}]$.

The investor has the following options. He may search for a fund himself and invest on his own after paying due diligence costs. Or, he may hire the intermediary broker and invest his money into a fund recommended by the broker. In the latter case, the broker performs due diligence and certifies the quality of the fund.

Due diligence is important since the hedge fund industry is opaque and there are fraudulent funds that investors should be aware of. Analyzing form ADV disclosures of registered hedge funds, Brown et al. (2008) find that approximately 16% of hedge funds have committed a felony or have financial-related charges or convictions. As pointed out by Garleanu and Pedersen (2016), hedge fund prospective investors usually undertake extensive analysis by studying the track record and evaluating the investment process and the risk management of funds. Fraudulent, negative alpha funds exist on the off-equilibrium path. Therefore, investors who do not perform due-diligence may loose money investing in these funds.

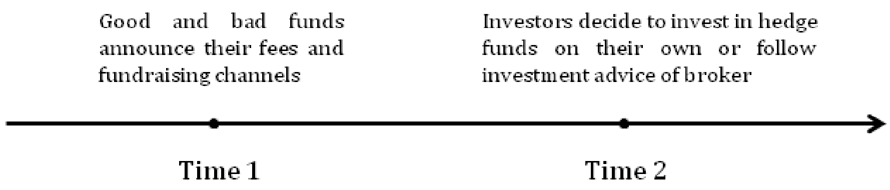
The broker performs due diligence and a certification of the fund at cost, $c_I > 0$. For the capital introduction service, the broker charges the fund some fraction of the fund's fees. The broker and the fund bargain with each other and split the collected dollar fees. I assume that the bargaining power of the broker is an exogenous parameter, $G \in (0; 1)$. Although I do not solve for an optimal contract for the broker, the performance-related compensation ensures that the broker acts in the interest of investors and allows for avoiding a moral hazard problem between the broker and the investors.

The fundraising game has a simple sequential structure, which is illustrated in Figure 1.1. At time 1, the good fund and the bad fund simultaneously announce fees that they charge for portfolio management services and their choices of capital raising channels. At time 2, the investors decide whether to invest into the hedge fund industry on their own or hire an intermediary broker.

Strategies.

Let f_θ be a fee that a type-θ fund charges its investors. Let X_θ be the fund's choice of capital raising channel. If the type-θ fund is sold to investors directly then $X_\theta = 0$. If the type-θ fund is sold to investors by the broker, then $X_\theta = 1$. The strategy of type-θ is a vector, $s_\theta = (f_\theta, X_\theta)$, such that $s_\theta \in \mathbb{R}^+ \times \{0, 1\}$. The good fund and the

Figure 1.1: Time line of the fundraising game



bad fund have strategies s_G and s_B, respectively.

The investor decides either to perform a costly due diligence of the hedge fund industry at cost c and invest into one of the funds on his own or to approach the intermediary broker and follow his investment advice. In both cases, the investor pays a portfolio management fee, f_θ, upon investing into the type-θ hedge fund. The decision of the investor depends on his search and due diligence costs c and the strategies of the funds s_G and s_B.

Payoffs of players.

Let's denote the profit of type-θ hedge fund $\Pi_\theta\Big(s_\theta; s_{-\theta}; \mathbb{C}(s_\theta, s_{-\theta})\Big)$. It depends on the strategy of the type-θ fund s_θ, the strategies of the other fund $s_{-\theta}$, and a proportion of investors, who decide to invest in the fund, denoted as $\mathbb{C}(s_\theta, s_{-\theta}) \subset [0; \bar{C}]$. Given strategy $s_\theta = (f_\theta, X_\theta)$, the profit of the type-θ fund is determined as

$$\Pi_\theta\Big(s_\theta; s_{-\theta}; \mathbb{C}(s_\theta, s_{-\theta})\Big) = \Pi_\theta\Big((f_\theta, X_\theta); s_{-\theta}; \mathbb{C}(s_\theta, s_{-\theta})\Big) = \tag{1.13}$$

$$\begin{cases} f_\theta \cdot \displaystyle\int_{\mathbb{C}(s_\theta, s_{-\theta})} dc, & \text{if } X_\theta = 0 \quad (1.13\text{a}) \\ (1-G) \cdot f_\theta \cdot \displaystyle\int_{\mathbb{C}(s_\theta, s_{-\theta})} dc, & \text{if } X_\theta = 1. \quad (1.13\text{b}) \end{cases}$$

If the type-θ fund decides to be sold to investors directly ($X_\theta = 0$), then its profits are equal to the total dollar fees raised from the investors, as in (1.13a). If the type-θ fund decides to be sold to investors through the broker ($X_\theta = 1$), then the fund and the broker split the total dollar fees and the fund gets a fraction $1 - G$, which is determined by its bargaining power, as in (1.13b).

Let's denote $U_{\theta c}$ the utility of the investor with due diligence cost c, who allocates his endowment into the type-θ fund. It is equal to

$$U_{\theta c} = \alpha_\theta - f_\theta - c \cdot \mathbb{I}\{X_\theta = 0\}. \tag{1.14}$$

If the investor invests on his own, then his utility equals to the after-fee return of the type-θ fund adjusted for due-diligence costs. If the investor follows financial advice, then his utility equals to the after-fee return on the type-θ fund.

Let's denote the profit that the broker gets $\Pi_I\Big(s_\theta; s_{-\theta}; \mathbb{C}(s_\theta, s_{-\theta})\Big)$. It is equal to the compensation that the broker gets for the capital introduction service adjusted for due diligence cost c_I. The profit of the broker may be expressed in terms of the profit that the fund receives as follows:

$$\Pi_I\Big(s_\theta; s_{-\theta}; \mathbb{C}(s_\theta, s_{-\theta})\Big) = \Big(\frac{G}{1-G} \cdot \Pi_\theta\Big(s_\theta; s_{-\theta}; \mathbb{C}(s_\theta, s_{-\theta})\Big) - c_I\Big) \cdot \mathbb{I}\{X_\theta = 1\}. \tag{1.15}$$

The broker makes a profit when the fund is broker-sold ($X_\theta = 1$) and he gets no profit when the fund is directly-sold to investors ($X_\theta = 0$).

Definition of "cut-off" equilibrium.

I define the Nash equilibrium of the fundraising game as follows:

(i) The good fund chooses strategy s_G to maximize its profits

$$\Pi_G\Big(s_G; s_B; \mathbb{C}(s_G, s_B)\Big) \geq \Pi_G\Big(s'_G; s_B; \mathbb{C}(s'_G, s_B)\Big) \text{ for any}$$
$$s'_G \in \mathbb{R}^+ \times \{0,1\}/\{s'_G \neq s_G\}.$$

(ii) The bad fund chooses strategy s_B to maximize its profits

$$\Pi_B\Big(s_B; s_G; \mathbb{C}(s_B, s_G)\Big) \geq \Pi_B\Big(s'_B; s_G; \mathbb{C}(s'_B, s_G)\Big) \text{ for any}$$
$$s'_B \in \mathbb{R}^+ \times \{0,1\}/\{s'_B \neq s_B\}.$$

(iii) There is a cut-off marginal investor with due diligence cost $\hat{c}(s_\theta, s_{-\theta})$ who is indifferent about investing on his own or using the advice of a broker (or investing in an outside option). Investors with costs that are lower than the cost of the marginal investor, i.e. $\mathbb{C}(s_G, s_B) = \Big[0; \min\{\hat{c}(s_G, s_B), \bar{C}\}\Big]$ will invest on their own. Investors with costs that are greater than the cost of the marginal investor, i.e $\mathbb{C}(s_B, s_G) = \Big(\min\{\hat{c}(s_B, s_G), \bar{C}\}; \bar{C}\Big]$ will approach the broker for investment advice.

(iv) The profit of the broker covers his due diligence cost, c_i.

Note that I restrict a space of the investor's strategies to "cut-off" strategy, which is determined by the marginal investor with a search and due diligence cost, $\hat{c}(s_\theta, s_{-\theta})$. Since the investors base of the fund $\mathbb{C}(s_\theta, s_{-\theta})$ may be fully described by a threshold search and due-diligence cost $\hat{c}(s_\theta, s_{-\theta})$ of the marginal investor, it allows me to simplify the notation for the profit of the type-θ fund in the following way, $\Pi_\theta\Big(s_\theta; s_{-\theta}; \hat{c}(s_\theta, s_{-\theta})\Big)$.

PROPOSITION. There exists a separating pure strategies "cut-off" equilibrium in the fundraising game. A good fund is directly-sold to investors and charges fee $f_G^* = \frac{\alpha_G}{2}$. *A bad fund raises capital through a broker and charges fees* $f_B^* = \alpha_B$.

$$s_G^* = \Big(\frac{\alpha_G}{2}, 0\Big), \tag{1.16}$$

$$s_B^* = \Big(\alpha_B, 1\Big). \tag{1.17}$$

A marginal investor with due diligence cost $\hat{c}^*$ *gets zero utility and is indifferent between investing into the hedge fund industry on his own or using the investment advice of a broker:*

$$\hat{c}^* = \frac{\alpha_G}{2}, \tag{1.18}$$

$$U_{Gc^*} = U_{Bc^*} = 0. \tag{1.19}$$

Investors with costs $c < \hat{c}^*$ *invest by themselves and those with* $c > \hat{c}^*$ *follow the recommendation of broker.*

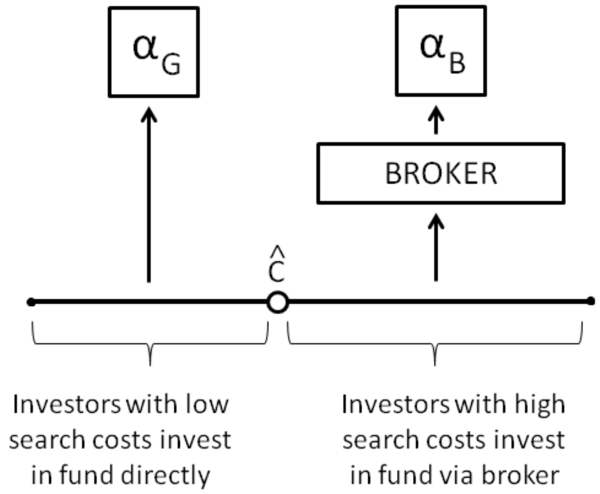
The necessary conditions for the existence of separating equilibrium are as follows:

$$\max\left\{1-\frac{\alpha_G}{4\cdot\bar{C}};\frac{c_I}{\alpha_B\cdot(\bar{C}-\frac{\alpha_G}{2})}\right\}\leqslant G<1 \tag{1.20}$$

$$\alpha_B<\hat{c}^*=\frac{\alpha_G}{2}<\bar{C}. \tag{1.21}$$

This separating "cut-off" equilibrium of the fundraising game is illustrated in Figure 1.2.

Figure 1.2: Separating equilibria of the fundraising game



Solution.

I verify the existence of the separating "cut-off" equilibrium by confirming the optimality of strategies of the players' strategies.

Good fund.

The good fund chooses optimally its fee and capital raising channel to maximize its profits (1.13). Since the capital raising choice of the fund is binary, the profit maximization over a two-dimensional vector-strategy $s_G = (f_G, X_G)$ simplifies to two one-dimensional maximization problems. The first optimization corresponds to the choice by the good fund of engaging in direct capital raising. The second optimization corresponds to a choice by the good fund of raising capital through the broker.

First, let's calculate the profits that the good fund gets if it chooses to be directly-sold ($X_\theta = 0$). Its investor base includes either all the investors with due diligence costs that are smaller than threshold $\hat{c}$ or the entire population of investors, $\mathbb{C}(s_G, s_B) = \Big[0; min\{\hat{c}(s_G, s_B), \bar{C}\}\Big]$. The good fund chooses fee f_G to maximize its profits subject to the feasibility condition on fees and the participation constraint of the marginal investor.

$$\Pi_G\Big((f_G, 0); s_B; \hat{c}(s_G, s_B)\Big) = \max_{f_G} \quad f_G \cdot \int_0^{min\Big\{\hat{c}\big((f_G,0);s_B\big),\bar{C}\Big\}} dc \tag{1.22}$$

subject to

$$0 \leqslant f_G \leqslant \alpha_G \tag{1.22a}$$

$$\alpha_G - f_G - \hat{c}((f_G, 0); s_B) = 0. \tag{1.22b}$$

The fee feasibility constraint (1.22a) states that the fund can not charge a fee f_G that is bigger than the return α_G that it generates. The participation constraint (1.22b) says that the marginal investor has to be indifferent about receiving utility $\alpha_G - f_G - \hat{c}$ upon investment into the fund and the utility of zero upon investment in an outside option.

Solving the maximization (1.22), I am interested in the interior case. There is also a less interesting corner case when even the highest cost investor decides to invest into the hedge fund on his own ($\hat{c} > \bar{C}$). In this case, all investors, after performing their due-diligence, invest in the good fund only. I consider a more realistic case when $\hat{c} < \bar{C}$. Then the optimization problem (1.22) is equivalent to the following quadratic optimization:

$$\max_{f_G} \quad f_G \cdot (\alpha_G - f_G) \tag{1.23}$$

subject to

$$0 \leqslant f_G \leqslant \alpha_G. \tag{1.23a}$$

The hedge fund's choice of fee affects its profits directly through fee f_G and indirectly through the size of its investors base $\alpha_G - f_G$. The good fund exercises its monopoly power and sets a fee optimally at, $f_G = \frac{\alpha_G}{2}$. Thus, the strategy of the good fund that chooses to be sold to investors directly is $s_G = (\frac{\alpha_G}{2}, 0)$ and its profits are:

$$\Pi_G\Big((\frac{\alpha_G}{2}, 0); s_B; \hat{c}(s_G, s_B)\Big) = \frac{\alpha_G^2}{4}. \tag{1.24}$$

The threshold search and due diligence costs are equal to

$$\hat{c} = \frac{\alpha_G}{2}. \tag{1.25}$$

To ensure the interior case occurs, which makes it suboptimal for high-cost investors to invest on their own, the following condition has to be satisfied:

$$\hat{c} < \bar{C}. \tag{1.26}$$

Substituting (1.25) into (1.26), I get the second condition in (1.21).

Second, let's calculate the profits that the good fund gets if it chooses to be sold through broker ($X_G = 1$). In this case, both funds are offered to investors through a broker. However, the broker will only market the good fund, since in this case, he will receive higher compensation. Thus, all investors will be channelled to the good fund and $\mathbb{C}(s_G, s_B) = [0; \bar{C}]$. The good fund that is sold through the broker will choose fee f_G to maximize its profits subject to the feasibility condition on the fee and the participation constraint of the broker.

$$\Pi_G\Big((f_G, 1); s_B; \hat{c}(s_G, s_B)\Big) = \max_{f_G} \quad (1-G) \cdot f_G \cdot \int_0^{\bar{C}} dc \tag{1.27}$$

subject to

$$0 \leqslant f_G \leqslant \alpha_G \tag{1.27a}$$

$$G \cdot f_G \cdot \int_0^{\bar{C}} dc \geqslant c_I. \tag{1.27b}$$

The fee feasibility constraint (1.27a) is similar to (1.22a). The broker helps to attract all investors to the good fund and gets a fraction G of the total dollar fees. The participation constraint of the broker (1.27b) ensures that the compensation that he receives is enough to cover his due diligence cost c_I.

Since the good fund gets all the investors regardless of the fees that it charges, it optimally sets a fee to extract all profits, leaving investors indifferent about investing into the fund or investing into the outside option. Thus, the good fund that chooses to be sold to investors through the broker sets fee $f_G = \alpha_G$. Its optimal strategy is $s_G = (\alpha_G, 1)$ and its profits are equal to the $(1-G)$ fraction of the generated surplus $\alpha_G \cdot \bar{C}$.

$$\Pi_G\Big((\alpha_G, 1); s_B; \hat{c}(s_G, s_B)\Big) = (1-G) \cdot \alpha_G \cdot \bar{C}. \tag{1.28}$$

The profits of the broker equals the fraction G of the generated surplus after accounting for the due diligence costs of the broker.

$$\Pi_I\Big((\alpha_G, 1); s_B; \hat{c}(s_G, s_B)\Big) = G \cdot \alpha_G \cdot \bar{C} - c_I. \tag{1.29}$$

Finally, the good fund optimally chooses the capital-raising channel by comparing profits (1.24) that it gets if it is directly-sold to investors with the profits (1.28) that it gets if it is sold to investors through a broker. For the good fund to become directly-sold, the following incentive compatibility condition must be met:

$$\Pi_G\Big((\frac{\alpha_G}{2},0);s_B;\hat{c}(s_G,s_B)\Big) > \Pi_G\Big((\alpha_G,1);s_B;\hat{c}(s_G,s_B)\Big). \tag{1.30}$$

Substituting (1.24) and (1.28) into condition (1.30) gives the first constraint on the bargaining power (1.20) of the broker:

$$G \geq 1 - \frac{\alpha_G}{4 \cdot \bar{C}}. \tag{1.31}$$

Bad fund.

The bad fund optimally chooses a fee and capital raising channel which maximizes its profits (1.13). Similar to the analysis for the good fund, I consider two separate cases, which correspond to the choice of fundraising of the bad fund.

First, let's calculate the profits that the bad fund gets if it chooses to be sold to investors through broker ($X_B = 1$). Investors with search and due diligence costs $c > \hat{c}$ approach the broker and invest their capital in the fund that the broker recommends. Its investor base is $\mathbb{C}(s_B, s_G) = (\hat{c}(s_B, s_G); \bar{C}]$ for the interior case when $\hat{c} < \bar{C}$. The bad fund chooses fee f_B to maximize its profit subject to the feasibility condition on the fee and the participation constraint of the broker.

$$\Pi_B\Big((f_B,1);s_G;\hat{c}(s_B,s_G)\Big) = \max_{f_B} \quad (1-G)\cdot f_B \cdot \int_{\hat{c}(s_B,s_G)}^{\bar{C}} dc \tag{1.32}$$

subject to

$$0 \leqslant f_B \leqslant \alpha_B \tag{1.32a}$$

$$G \cdot f_B \cdot \int_{\hat{c}(s_B,s_G)}^{\bar{C}} dc \geqslant c_I. \tag{1.32b}$$

The fee feasibility constraint (1.32a) states that the fund cannot charge a fee f_B bigger than the return α_B that it generates. The broker brings investors $\mathbb{C}(s_B, s_G) = (\hat{c}(s_B, s_G); \bar{C}]$ to the bad fund and receives a fraction G of the total dollar fees that the fund charges. The participation constraint of the broker (1.32b) ensures that the compensation that he receives is enough to cover his due diligence cost c_I.

The choice of fees of the bad fund has only a direct effect on its profit, since its investors' base comes from the broker. Thus, it maximizes its profits by extracting all profits through fees and making its investors indifferent about investing into the fund or investing in an outside option. Thus, the bad fund that chooses to be sold to investors through the broker sets the fee $f_B = \alpha_B$. Its strategy is $s_B = (\alpha_B, 1)$ and its profits are equal to the $(1-G)$ fraction of the generated surplus $\alpha_B \cdot [\bar{C} - \frac{\alpha}{2}]$

$$\Pi_B\Big(s_G; (\alpha_B, 1); \hat{c}(s_B, s_G)\Big) = (1-G) \cdot \alpha_B \cdot [\bar{C} - \frac{\alpha_G}{2}]. \tag{1.33}$$

The profits that the broker gets is a fraction G of the generated surplus after accounting for the due diligence costs of the broker.

$$\Pi_I\Big(s_G; (\alpha_B, 1); \hat{c}(s_B, s_G)\Big) = G \cdot \alpha_B \cdot [\bar{C} - \frac{\alpha_G}{2}] - c_I > 0. \tag{1.34}$$

Condition (1.34) yields the second constraint (1.20) on the bargaining power of the broker.

$$G \geqslant \frac{c_I}{\alpha_B \cdot (\bar{C} - \frac{\alpha_G}{2})} \tag{1.35}$$

Second, consider the case when the bad fund chooses to be directly sold ($X_B = 0$) and its strategy is described as $s_B = (f_B, 0)$. When the bad fund decides to be directly sold, we have to insure that it will not attract any investors regardless of the fee that it sets. To attract more investors, the bad fund may set zero fees $f_B = 0$. In this case, its strategy is $s_B = (0, 0)$.

I need to ensure that the marginal investor $\hat{c}$ still prefers to invest into the good fund that is sold directly rather than into the bad fund that is sold directly and charges no fees. The marginal investor invests into the good directly-sold fund if

$$\alpha_B - f_B - \hat{c} < \alpha_G - f_G - \hat{c}. \tag{1.36}$$

Since $f_B = 0$ and $f_G = \frac{\alpha_G}{2}$, I get

$$\alpha_B < \frac{\alpha_G}{2}. \tag{1.37}$$

The combination of conditions (1.26), (1.31), (1.35), and (1.37) determine the necessary conditions for the existence of a pure strategy separating the "cut-off" equilibrium in Proposition 1.

Discussion of equilibrium. I consider several cases in relation to the parameters of the model to illustrate equilibrium. When the bargaining power of the broker is high $G \to 1$, the broker extracts all generated surplus. In this case, condition (1.20) is always satisfied and the good fund never wants to use the capital introduction services of the broker.

In the case of competition among the brokers, the broker should make enough profit to cover his due diligence cost c_I. If the fund hires a competitive broker, then the profit of the fund equals the generated surplus adjusted by the due diligence cost of the broker.

$$\frac{\alpha_G^2}{4} > \alpha_G \cdot \bar{C} - c_I. \tag{1.38}$$

If the due diligence cost is high, then the good fund and the bad fund separate:

$$c_I > \alpha_G \cdot [\bar{C} - \frac{\alpha_G}{4}]. \tag{1.39}$$

If the due diligence cost is low and condition (1.39) is violated, then only the good fund survives.

1.3.2 Model implications

Next, I discuss the implications of the the fundraising model and reconcile the model predictions with the empirical results from Section II.

First, the model has implications for the after-fee return that investors receive on their hedge fund investments, $\alpha_\theta - f_\theta$. The equilibrium strategy of the good fund (1.16) implies that the after-fee returns of investors in the directly-sold fund are determined by the reservation value for the marginal investor and are equal to $\frac{\alpha_G}{2}$. The equilibrium strategy of the bad fund (1.17) implies that broker-sold fund extracts all generated surplus through fees, making its investors indifferent about investing in the fund and the outside option. Therefore, the after-fee return of the broker-sold fund investor is equal to 0. Thus, the after-fee returns of directly-sold funds are higher than the after-fee returns of broker-sold funds $\frac{\alpha_G}{2} > 0$.

The empirical patterns that are documented in Figure 1.4, Figure 1.5 and Panel A of Table 1.7 support the prediction about the after-fee performance of directly-sold and broker-sold funds.

Second, the model also makes predictions about the pre-fee return of directly-sold and broker-sold funds. The equilibrium strategies of the good fund (1.16) and that of the bad fund (1.17) imply that the good fund raises capital directly, while the bad fund raises funds through the broker. Together with condition (1.37), it implies that broker-sold funds underperform directly-sold funds, even before accounting for portfolio management fees $\alpha_G > \alpha_B$.

The empirical findings of Figure 1.6 and Panel B of Table 1.7 support the model prediction about the pre-fee performance of directly-sold and broker-sold funds.

Third, the model makes a prediction about portfolio management fees that funds charge. The equilibrium strategy of the good fund (1.16) implies that the directly-sold fund charges fee, $f_G = \frac{\alpha_G}{2}$. The equilibrium strategy of the bad fund (1.17) states that the broker-sold fund charges fee, $f_B = \alpha_B$. Condition (1.37) from Proposition 1 implies that the fees that directly-sold funds charge their investors are higher than the fees that broker-sold funds charge their investors $f_G = \frac{\alpha_G}{2} > \alpha_B = f_B$.

Table 1.11 presents the results of testing the above prediction. I find that directly-sold funds charge higher incentive fees than broker-sold funds. I do not find, however, any significant difference between the management fees of directly-sold and broker-sold funds.

Fourth, the model makes predictions about clientele of the funds. In equilibrium, investors with costs smaller than the costs of the marginal investor invest in the directly-sold fund $\mathbb{C}(s_G, s_B) = [0; \hat{c}]$, while investors with costs higher than cost of the marginal investor invest in the broker-sold fund $\mathbb{C}(s_B, s_G) = [\hat{c}; \bar{C}]$. Thus, the model predicts that the marginal and average investor of the directly-sold fund has lower costs than the marginal and average investor of the broker-sold fund. If the sizes of the investors is negatively correlated with their due diligence and search costs, then the model implies that the marginal investor of the directly-sold fund with cost $\hat{c}$ is bigger than the marginal investor of the broker-sold fund with cost $\bar{C}$. Also, the average investor of the directly-sold fund with cost $\frac{\hat{c}}{2}$ is bigger than average investor of the broker-sold fund with cost $\frac{\hat{c}+\bar{C}}{2}$.

Using minimum investment size as an empirical proxy of the size of the marginal investor and the average investment size as a proxy of the size of average investor, I test the model predictions of the clientele of hedge funds. Table 1.12 displays the tests of the above prediction.

1.3.3 Compensation for the broker

I estimate the economic magnitude of compensation that broker receives for capital introduction services. In the fundraising model, the broker and the fund split the dollar profits. Compensation for the broker is proportional to the total dollar fees that hedge fund collects from its investors, with the proportionality constant being equal to the bargaining power of the broker, as in (1.15).

I use information about the fund's assets under management, performance, and compensation structure to estimate the total dollar fees. Using methodology for the reconstruction of the pre-fee returns that is described in detail in the section Data, I estimate the dollar management fees using equations (1.3) and dollar incentive fees using equation (1.4). I find the total dollar fees collected as a sum of the annual dollar management fees and the dollar incentive fees. I consider the bargaining power of the broker to be in the range of 5% to 95%. The lower bound corresponds to the low bargaining power and the upper bound to the high bargaining power. Knowing the total annual dollar fees and the bargaining power of the broker, I estimate the fees that the broker gets for a capital introduction service using equation (1.15).

For every broker-sold fund in the matched sample, I estimate the annual compensation that broker receives. I report the average annual compensation in Table 1.13. Depending on the bargaining power, the estimates of the annual compensation of the broker vary from $241,000 to $4.58 million. For a bargaining power of 1\3, which corresponds to the equilateral division of surplus among the fund, its investors, and the broker, I estimate the average compensation that the broker receives to be $1.45 million per year.

1.4 Conclusion

This paper analyzes empirically and theoretically the fundraising process in the hedge fund industry. I analyze form D filings that hedge funds report to the SEC with regard to their fundraising process. Information that the funds report in their filings allows me to differentiate between the funds that raise capital directly from investors and those that use the capital introduction services of intermediary brokers. I find that funds that are sold to investors through intermediaries underperform funds that are offered to investors directly on a risk-adjusted basis, both before and after accounting for fees. I also find that hedge funds that are sold to investors directly on average have a larger average investorâĂŹs size, a larger minimum investment size and charge higher incentive fees compared to funds offered to investors by brokers. These findings provide empirical description of the equilibrium.

I also present a stylized model that has a simple intuition and reconciles the above empirical findings. In equilibrium, sophisticated investors who are better at due diligence will sort themselves into better funds, which avoid having to internalize the high cost of hiring a broker, while bad funds hire a broker, which mitigates capital-raising inefficiency, but requires compensation for capital introduction services. Brokers' bargaining power and the relative outperformance of the good fund over the bad fund are essential for the existence of separating equilibria. The calibrate model implies that average broker compensation is $1.5 million per year, which is consistent with the empirically estimate, value-added difference between the broker-sold funds and the directly-sold funds.

1.5 Tables and figures

Table 1.1: Outline of form D filings

Item	Description
1. Issuer's identity	Name and type of entity that initiates fundraising.
2. Principal place of business and contact information	Administrative information about the fundraising entity.
3. Related persons	Information about all executive officers, directors, and promoters associated with the fundraisning offer.
4. Industry group	Information on the entity's industry group that most accurately reflects the use of capital raised. *Banking and financial services* includes *pooled investment funds*, which comprises *hedge funds*, *private equity funds*, *venture capital funds*, and *other investment funds*.
5. Issuer size	Information of revenue range or aggregate net asset value of fundraising entity. Hedge funds and other investment funds may decline to response to this question.
6. Federal exemptions and exclusions claimed	Provision(s) that are claimed to exempt the capital raising from formal offering registration.
7. Type of filing	Information on whether the entity is filing a new notice or an amendment to a notice.
8. Duration of offering	Information on duration of fundraising offering.
9. Type(s) of securities offered	Information on the type of security offered, which includes equity, debt, options, and pooled investment fund interests.
10. Business combination transactions	Information on whether the fundraising offering is made in connection with business combination transactions, such as merger or acquisition.
11. Minimum investment size	Minimum dollar amount of investment that will be accepted from any outside investor.
12. Sales compensation	Information about each person that has been or will be paid directly or indirectly any commission in connection with fundraising.
13. Offering and sales amounts	Dollar amount of capital raised up to date.
14. Investors	Total number of investors who already have invested in the offering and number of non-accredited investors.
15. Sales commissions and finders' fees expenses	Information on estimate of sales commissions and finders' fee expenses.
16. Use of proceeds	Estimation of commissions that are paid to related persons.

Table 1.1 describes information about their fundraising process that hedge funds disclose in form D filings. Column Item outlines main categories of the form D. Column Description provides key information that fundraising entity reports in item.

Table 1.2: Largest funds by distribution channel

Fund	Fund family	Capital raised
PANEL A: Directly sold funds		
VERDE ALPHA FUND LTD	Verde Asset Management	20,221
GLOBAL ASCENT LTD	Global Ascent	16,524
OZ OVERSEAS FUND II LTD	OZ Management	15,290
CANYON VALUE REALIZATION FUND LTD	Canyon Capital Advisors	14,745
ADAGE CAPITAL PARTNERS LP	Adage Capital Management	14,049
CONVEXITY CAPITAL OFFSHORE LP	Convexity Capital GP	11,155
ABERDEEN FIXED INCOME FUNDS POOLED TRUST	Aberdeen Asset Management	10,783
DYMON ASIA MACRO FUND	Dymon Asia Capital	10,733
TUDOR BVI GLOBAL FUND LTD	Tudor Investment Corp	10,587
LONE CASCADE LP	Lone Pine Capital	10,347
ANCHORAGE CAPITAL PARTNERS OFFSHORE LTD	Anchorage Capital Group	10,063
GLENVIEW CAPITAL PARTNERS CAYMAN LTD	Glenview Capital Management	9,495
KING STREET CAPITAL LP	King Street Capital	9,473
BROOKSIDE CAPITAL PARTNERS FUND LP	Brookside Capital Management	8,905
BAUPOST VALUE PARTNERS LP IV	The Baupost Group	8,603
PANEL B: Broker sold funds		
D.E. SHAW COMPOSITE INTERNATIONAL FUND	D.E. Shaw & Co	18,235
RENAISSANCE INSTITUTIONAL EQUITIES FUND LLC	Renaissance Technologies LLC	16,192
MESIROW ABSOLUTE RETURN FUND LTD	Mesirow Advanced Strategies Inc	15,096
D.E. SHAW OCULUS INTERNATIONAL FUND	D.E. Shaw & Co	13,390
RENAISSANCE INSTITUTIONAL DIVERSIFIED ALPHA	Renaissance Technologies LLC	10,232
GRAHAM GLOBAL INVESTMENT FUND II SPC LTD	Graham Capital Management	10,199
GRAHAM GLOBAL INVESTMENT FUND I SPC LTD	Graham Capital Management	9,227
BREVAN HOWARD FUND LTD	Brevan Howard Capital Management LP	8,412
MESIROW ABSOLUTE RETURN FUND (INSTITUTIONAL)	Mesirow Advanced Strategies Inc	8,196
D.E. SHAW COMPOSITE FUND LLC	D.E. Shaw & Co	7,779
DRAWBRIDGE SPECIAL OPPORTUNITIES FUND LP	Fortress Investment Group LLC	7,056
MILLENNIUM USA LP	Millennium Management LLC	6,868
PERMAL FIXED INCOME HOLDINGS NV	Permal Asset Management Inc	6,847
WEATHERLOW FUND I LP	Evanston Capital Management LLC	6,804
PAULSON ADVANTAGE PLUS LP	Paulson & Co	6,419

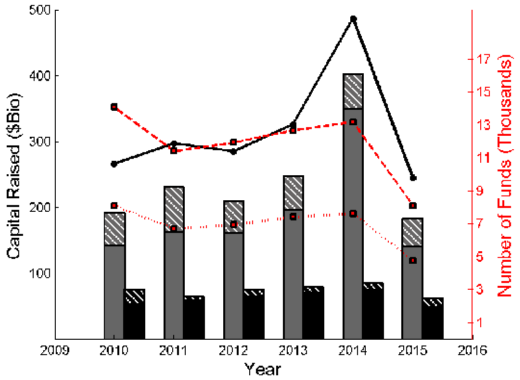
Table 1.2 presents fifteen directly sold hedge funds (Panel A) and broker sold hedge funds (Panel B) that were open for investment and raised maximum amount of capital by 2015. Table reports fund's name, name of management company and total amount of capital raised (in millions of dollars).

Table 1.3: Top players in fundraising industry

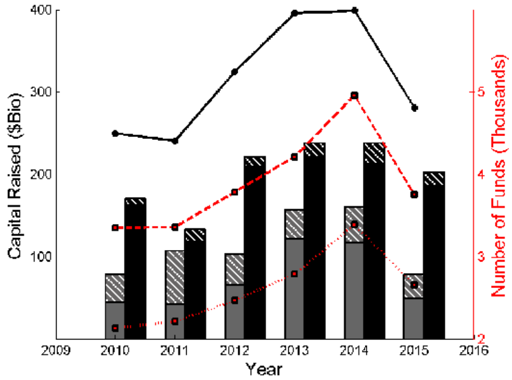
	Name	# Funds	Capital raised	# Investors
1.	GOLDMAN SACHS & CO	377	350	149
			[98]	[30]
2.	WELLS FARGO ADVISORS, LLC	364	176	271
			[25]	[16]
3.	MORGAN STANLEY & CO	359	428	436
			[77]	[99]
4.	J.P. MORGAN SECURITIES LLC	295	765	248
			[256]	[69]
5.	MERRILL LYNCH	275	319	469
			[118]	[158]
6.	CITIGROUP GLOBAL MARKETS INC	242	403	453
			[87]	[81]
7.	CREDIT SUISSE SECURITIES LLC	210	367	433
			[97]	[57]
8.	UBS FINANCIAL SERVICES INC	191	443	347
			[193]	[128]
9.	DEUTSCHE BANK SECURITIES INC	170	385	76
			[23]	[6]
10.	BARCLAYS CAPITAL INC.	114	395	144
			[156]	[75]

Table 1.3 provides information on the top broker firms that intermediate fundraising process. Top broker firms are defined as those companies that intermediate the largest number of funds. Table reports broker's name, average [median] amount of capital raised by funds that are intermediated by the same broker firm (in millions of dollars) and average [median] number of investors in funds with the same broker. Statistics are calculated using sample of Form D filings from January 2009 to December 2015 for hedge funds and other investment companies. For each broker statistics are calculated on sample of funds that are intermediated by this broker, using information that is available in the latest available form D filings where the broker is reported.

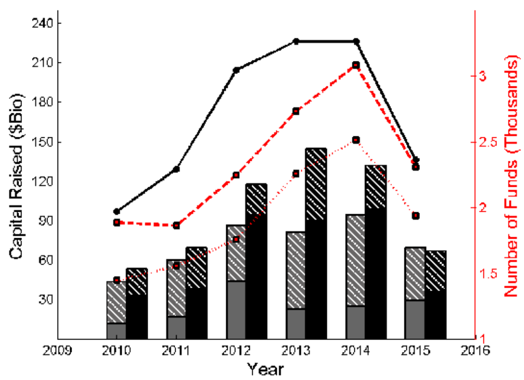
Figure 1.3: Fundraising in alternative investment industry



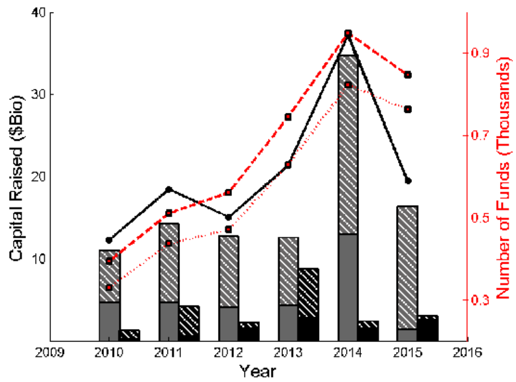
Panel A: Hedge funds



Panel B: Other investment funds



Panel C: Private equity funds



Panel D: Venture capital funds

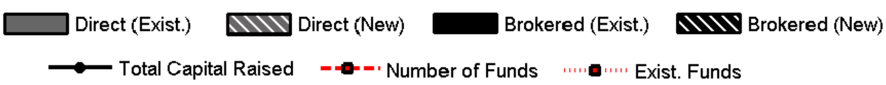


Figure 1.3 displays fundraising dynamics in alternative investment industry from 2010 to 2015, using information that funds report in form D filings. Panel A, B, C and D displays evolution of hedge funds, other investment funds, private equity and venture capital industries, respectively. Bars indicate amount of capital (in billions of dollars, left y-axis) that funds have raised from investors during a given year. Grey solid bars indicate capital that was raised by existing directly-sold funds. Grey hatched bars display capital that was raised by newly opened directly-sold funds. Black solid bars indicate capital inflows into existing broker-sold funds. Black hatched bars show capital that was raised by newly opened broker-sold funds. Black solid line (right y-axis) indicates total amount of capital raised in a given year. Red dashed line displays total number of funds that raise capital from investors in a given year (in thousands, right y-axis). Appendix describes methodology that is used to estimate capital inflows. Red dotted line indicates total number existing funds (in thousands, right y-axis).

Table 1.4: Summary statistics

	Direct	Brokered	Diff.	P-value
PANEL A: form D filings				
Average Inflows	47.80	48.50	0.70	(0.92)
Median Inflows	2.66	5.00	2.34	
Average [Inflows >0]	66.80	63.30	-3.50	(0.74)
Median [Inflows >0]	9.63	12.00	2.37	
Average # Investors	48	142	94***	(0.00)
Median # Investors	15	42	27	
Average # New Investors	12	33	21***	(0.00)
Median # New Investors	5	7	2	
# Filings	31,031	9,283		
# Funds	9,650	1,925		
PANEL B: form D filings and Morningstar				
Average Inflows	45.50	47.31	1.81	(0.71)
Median Inflows	3.43	4.23	0.80	
Average [Inflows >0]	60.30	59.91	-0.39	(0.95)
Median [Inflows >0]	9.04	8.50	-0.54	
Average # Investors	75	118	43***	(0.00)
Median # Investors	42	74	32	
Average # New Investors	14	27	13***	(0.00)
Median # New investors	6	7	1	
# Filings	2,872	1,129		
# Funds	1,103	625		

Table 1.4 describes information that funds report in form D filings for directly sold funds and broker sold funds over the period from January 2009 to December 2015. Panel A focuses on the sample of all hedge funds that file forms D. Panel B presents results for the sample of funds that file forms D and list their funds at Morningstar database. Table presents information about the average and median annual capital inflows(in millions of dollars), average and median annual positive capital inflows (in millions of dollars), average and median number of investors and average positive minimum investment size (in thousands of dollars). Methodology that is used to estimate annual capital inflows is outlined in Appendix. Column Diff. reports difference between the values for directly sold and broker sold funds. Column P-value reports p-value (in parenthesis) of T-test for means across directly sold and broker sold funds groups. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively.

Figure 1.4: Performance of directly sold and broker sold hedge funds

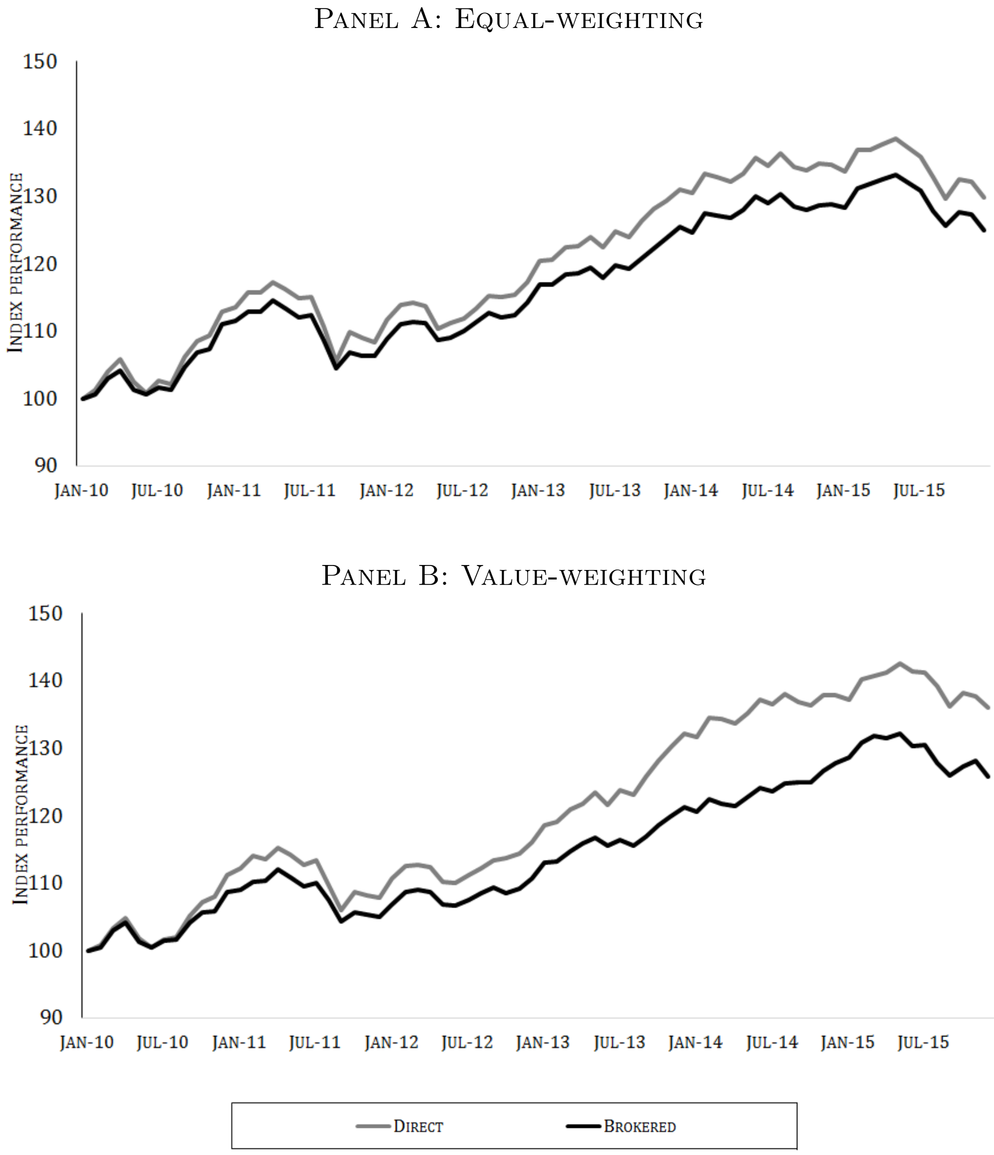


Figure 1.4 displays after-fee performance of fund of directly sold hedge funds (grey solid line) relative to performance of fund of broker sold hedge funds (black solid line) over the period from January 2010 to December 2015, assuming initial investment of $100. The sample of funds consists of funds that are listed in Morningstar database and file form D filings. Panel A displays after-fee performance of funds of funds where constituent hedge funds are equally-weighted. Panel B displays after-fee performance of funds of funds where constituent hedge funds are value-weighted. Returns of funds are adjusted for backfill bias.

Table 1.5: After-fee systematic risk exposure of hedge funds

	$\bar{R}$	$\hat{\alpha}$	$\hat{\beta}_{Mkt}$	$\hat{\beta}_{SmB}$	$\hat{\beta}_{T10y}$	$\hat{\beta}_{Cr.Spr.}$	$\hat{\beta}_{pBD}$	$\hat{\beta}_{pFX}$	$\hat{\beta}_{pCOM}$	R^2
	PANEL A: Equal-Weighting									
Direct	4.79%**	4.42%**	0.12*	0.38***	0.10	0.25***	-0.02*	0.01	-0.01	68%
	(0.02)	(0.02)	(0.06)	(0.04)	(0.07)	(0.09)	(0.01)	(0.01)	(0.01)	
Brokered	3.97%*	3.37%*	0.12**	0.32***	0.07	0.18**	-0.07*	0.01	-0.01*	68%
	(0.02)	(0.02)	(0.05)	(0.03)	(0.06)	(0.07)	(0.01)	(0.01)	(0.01)	
	PANEL B: Value-Weighting									
Direct	5.39%	4.433%**	0.13***	0.31***	0.07	0.16**	-0.02*	0.01	-0.01	66%
	(0.02)	(0.02)	(0.05)	(0.03)	(0.06)	(0.07)	(0.01)	(0.01)	(0.01)	
Brokered	4.16%	3.552%**	0.12***	0.25***	0.05	0.15**	-0.01*	0.01	-0.01	62%
	(0.02)	(0.01)	(0.04)	(0.03)	(0.05)	(0.06)	(0.01)	(0.01)	(0.01)	

Table 1.5 presents estimation of Fung and Hsieh (2004) seven-factor model for fund of directly sold (row "Direct") and broker sold funds (row "Brokered"). Panel A displays results for funds of funds where constituent funds are equally-weighted. Panel B reports results for funds of funds where constituent funds are value-weighted. The sample of funds is restricted to funds that are listed in Morningstar database and file form D filings. The seven-factor model (1.1) is estimated using after-fee monthly returns between January 2010 and December 2015, where the first 24-months of fund's performance are excluded to adjust for backfill bias. Table displays estimated annualized excess after-fee return of fund of fund, $\bar{R}$, estimated annualized alpha, $\hat{\alpha}$, estimated exposures to market factor, $\hat{\beta}_{Mkt}$, estimated exposure to size spread factor, $\hat{\beta}_{SmB}$, estimated exposure to yield curve level factor, $\hat{\beta}_{T10y}$, estimated exposure to credit spread factor, $\hat{\beta}_{Cr.Spr.}$, and estimated exposures to bond, commodity and forex trend-following factors, $\hat{\beta}_{pBD}$, $\hat{\beta}_{pFX}$ and $\hat{\beta}_{pCOM}$, as well as the adjusted R^2. Figures in parentheses are the Newey and West (1987) heteroscedasticity and autocorrelation consistent standard errors. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively.

Table 1.6: Pre-fee systematic risk exposure of hedge funds

	$\bar{R}$	$\hat{\alpha}$	$\hat{\beta}_{Mkt}$	$\hat{\beta}_{SmB}$	$\hat{\beta}_{T10y}$	$\hat{\beta}_{Cr.Spr.}$	$\hat{\beta}_{pBD}$	$\hat{\beta}_{pFX}$	$\hat{\beta}_{pCOM}$	R^2
	PANEL A: Equally-weighted									
Direct	6.17%***	5.78%***	0.12*	0.39***	0.11	0.25***	-0.02*	0.01	-0.01	69%
	(0.02)	(0.02)	(0.06)	(0.04)	(0.07)	(0.09)	(0.01)	(0.01)	(0.01)	
Brokered	5.12%***	4.48%**	0.17**	0.33***	0.07	0.18**	-0.01*	0.01	-0.01*	69%
	(0.02)	(0.02)	(0.05)	(0.03)	(0.06)	(0.07)	(0.01)	(0.01)	(0.01)	
	PANEL B: Value-weighted									
Direct	6.62%***	5.53%***	0.14***	0.32***	0.07	0.16**	-0.02*	0.01	-0.01	65%
	(0.02)	(0.02)	(0.05)	(0.03)	(0.06)	(0.07)	(0.01)	(0.01)	(0.01)	
Brokered	5.50%***	4.95%***	0.11***	0.26***	0.05	0.15**	-0.01	0.01	-0.01	61%
	(0.02)	(0.01)	(0.04)	(0.03)	(0.05)	(0.06)	(0.01)	(0.01)	(0.01)	

Table 1.6 presents estimation of Fung and Hsieh (2004) seven-factor model for fund of directly sold (row "Direct") and broker sold funds (row "Brokered"). Panel A displays results for funds of funds where constituent funds are equally-weighted. Panel B reports results for funds of funds where constituent funds are value-weighted. The sample of funds is restricted to funds that are listed in Morningstar database and file form D filings. The seven-factor model (1.1) is estimated using pre-fee monthly returns between January 2010 and December 2015, where the first 24-months of fund's performance are excluded to adjust for backfill bias. Table displays estimated annualized excess pre-fee return of fund of fund, $\bar{R}$, estimated annualized alpha, $\hat{\alpha}$, estimated exposures to market factor, $\hat{\beta}_{Mkt}$, estimated exposure to size spread factor, $\hat{\beta}_{SmB}$, estimated exposure to yield curve level factor, $\hat{\beta}_{T10y}$, estimated exposure to credit spread factor, $\hat{\beta}_{Cr.Spr.}$, and estimated exposures to bond, commodity and forex trend-following factors, $\hat{\beta}_{pBD}$, $\hat{\beta}_{pFX}$ and $\hat{\beta}_{pCOM}$, as well as the adjusted R^2. Figures in parentheses are the Newey and West (1987) heteroscedasticity and autocorrelation consistent standard errors. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively.

Figure 1.5: After-fee alphas of directly sold and broker sold hedge funds

PANEL A: EQUAL-WEIGHTING

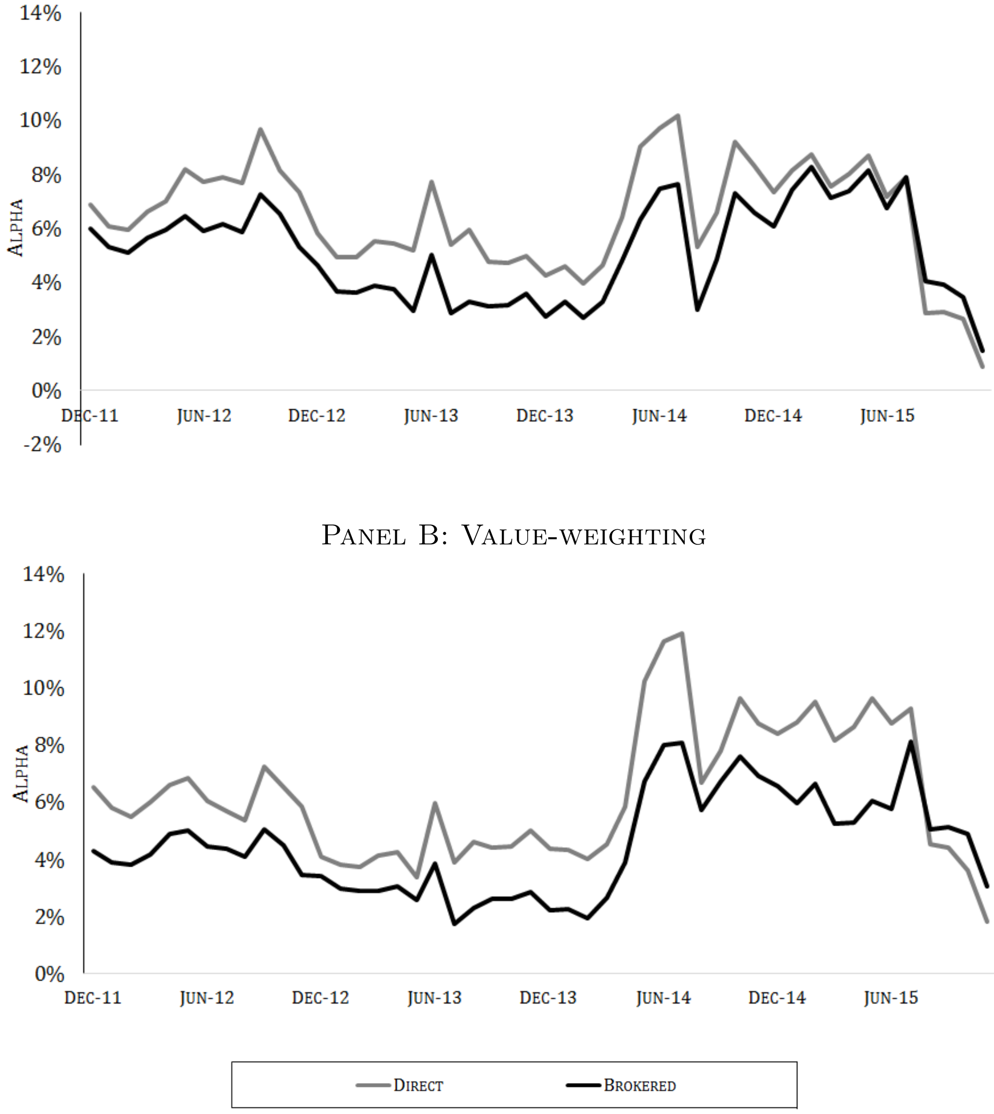
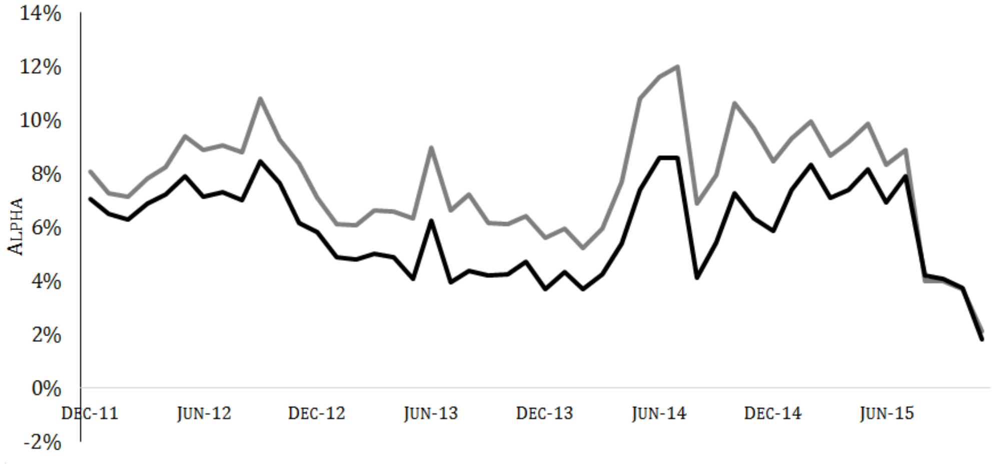


FIGURE 1.5 DISPLAYS A TIME VARYING RISK-ADJUSTED PERFORMANCE (ALPHA) FOR THE EQUALLY-WEIGHTED AND VALUE-WEIGHTED FUNDS OF HEDGE FUNDS THAT ARE DISPLAYED IN PANEL A AND PANEL B, ACCORDINGLY. ALPHAS OF FUNDS OF FUNDS ARE ESTIMATED WITH THE ROLLING-WINDOW FUNG AND HSIEH (2004) SEVEN-FACTOR MODEL (1.1). THE ROLLING-WINDOW REGRESSIONS (WITH 24 MONTHS WINDOW) ARE ESTIMATED FOR EACH PORTFOLIO USING MONTHLY AFTER-FEE RETURNS BETWEEN JANUARY 2010 AND DECEMBER 2015 (ADJUSTED FOR BACKFILL BIAS). ROLLING AFTER-FEE ALPHA OF FUND OF DIRECTLY SOLD FUNDS IS DISPLAYED WITH GREY SOLID LINE AND THAT OF FUND OF BROKER SOLD FUNDS IS DISPLAYED WITH BLACK SOLID LINE.

Figure 1.6: Pre-fee alphas of directly sold and broker sold hedge funds

Panel A: Equal-Weighting



Panel B: Value-Weighting

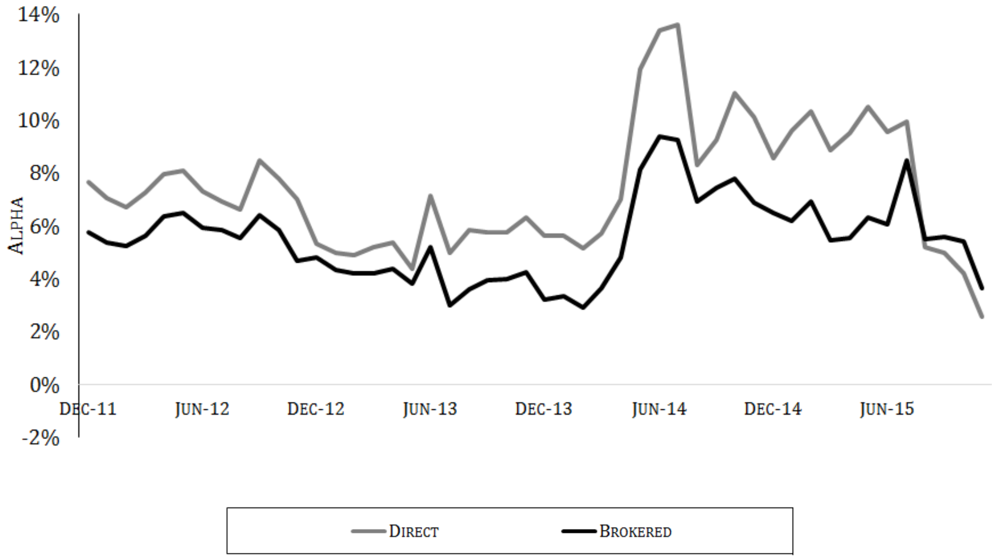


Figure 1.6 displays a time varying risk-adjusted performance (alpha) for the equally-weighted and value-weighted funds of hedge funds that are displayed in Panel A and Panel B, accordingly. Alphas of funds of funds are estimated with the rolling-window Fung and Hsieh (2004) seven-factor model (1.1). The rolling-window regressions (with 24 months window) are estimated for each fund of funds using monthly pre-fee returns between January 2010 and December 2015 (adjusted for backfill bias). Rolling pre-fee alpha of fund of directly sold funds is displayed with grey solid line and that of fund of broker sold funds is displayed with black solid line.

Table 1.7: Alphas of directly and broker sold hedge funds

	Alpha		
	(1)	(2)	(3)
PANEL A: After-fee			
B_{it}	-0.013***	-0.016***	-0.016***
	(0.002)	(0.002)	(0.002)
$\ln(Asset_{it-1})$	—	0.007***	0.007***
	—	(0.001)	(0.001)
Age_{it}	—	-0.0001	-0.0005**
	—	(0.0002)	(0.0002)
Vintage	No	Yes	Yes
Time FE	No	No	Yes
# Obs.	29,051	29,051	29,051
R^2	0.02%	4%	7%
PANEL B: Pre-fee			
B_{it}	-0.016***	-0.021***	-0.021***
	(0.002)	(0.001)	(0.001)
$\ln(Asset_{it-1})$	—	0.008***	0.008***
	—	(0.001)	(0.001)
Age_{it}	—	-0.0001	0.0007***
	—	(0.0002)	(0.0002)
Vintage	No	Yes	Yes
Time FE	No	No	Yes
# Obs.	28,493	28,493	28,493
R^2	0.3%	4%	7%

Table 1.7 presents estimates of difference in risk-adjusted performance between directly sold and broker sold hedge funds with panel regression $\hat{\alpha}_{it} = \beta_0 + \beta_B \cdot B_{it} + \beta_s \cdot X_{it-1} + \beta_t + \tilde{\epsilon}_{it}$. Fund level controls X_{it-1} include logarithm of assets under management in the previous period, age, and vintage year and time fixed effects β_t. Panel A displays results for after-fee alphas of hedge funds. Panel B displays results for pre-fee alphas of hedge funds. The sample covers hedge funds that are listed in Morningstar database and file form D filings over period from January 2010 to December 2015. Figures in parentheses are the Newey and West (1987) heteroscedasticity and autocorrelation consistent standard errors clustered by month. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively.

Table 1.8: Value added by directly and broker sold hedge funds

	Dollar value added		
	(1)	(2)	(3)
PANEL A: After-fee			
B_{it}	-0.214***	-0.209***	-0.211***
	(0.051)	(0.048)	(0.048)
Age_{it}	—	-0.0004	-0.017**
	—	(0.003)	(0.004)
Vintage	No	Yes	Yes
Time FE	No	No	Yes
# Obs.	29,051	29,051	29,051
R^2	1%	4%	5%
PANEL B: Pre-fee			
B_{it}	-0.198***	-0.182***	-0.189***
	(0.058)	(0.056)	(0.056)
Age_{it}	—	-0.001	0.014***
	—	(0.004)	(0.004)
Vintage	No	Yes	Yes
Time FE	No	No	Yes
# Obs.	28,493	28,493	28,493
R^2	0.06%	3%	4%

Table 1.8 presents estimates of difference in dollar value added (in millions of dollars) by directly sold and broker sold hedge funds with panel regression $\hat{S}_{it} = \beta_0 + \beta_B \cdot B_{it} + \beta_s \cdot X_{it} + \beta_t + \tilde{\epsilon}_{it}$. Fund level controls X_{it} include fund's age, vintage year and time fixed effects β_t. Panel A displays results for after-fee dollar value added by hedge funds. Panel B displays results for pre-fee dollar value added of hedge funds. The sample covers hedge funds that are listed in Morningstar database and file form D filings over period from January 2010 to December 2015 with an adjustment for backfill bias. Figures in parentheses are the Newey and West (1987) heteroscedasticity and autocorrelation consistent standard errors clustered by month. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively.

Table 1.9: Heterogeneity of brokers

	After-fee alpha		
	(1)	(2)	(3)
B^I_{it}	-0.020***	-0.020***	- 0.021***
	(0.003)	(0.003)	(0.003)
B^O_{it}	-0.014***	-0.014***	-0.014***
	(0.002)	(0.002)	(0.002)
$\ln(Asset_{it-1})$	0.006***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)
Age_{it}	0.0000	-0.0001	-0.0008
	(0.0000)	(0.0002)	(0.0005)
Vintage	No	Yes	Yes
Time FE	No	No	Yes
# Obs.	28,854	28,854	28,854
R^2	1%	3%	4%
Ho: In-house = Outside			
F-test	3.73*	4.36**	4.74**
p-value	0.06	0.04	0.03

Table 1.9 estimates difference in after-fee risk adjusted performance between directly sold hedge funds and funds that are sold through in-house broker or outside broker with panel regression: $\hat{\alpha}_{it} = \beta_0 + \beta_{in} \cdot B^I_{it} + \beta_{out} \cdot B^O_{it} + \beta_x \cdot X_{it} + \beta_t + \tilde{\epsilon}_{it}$. B^I_{it} is a dummy variable that is equal to one when the fund is sold through in-house broker and is equal to zero otherwise. B^O_{it} is a dummy variable that is equal to one when the fund is sold through outside broker and is equal to zero otherwise. Regression includes fund level controls, X_{it}, such as fund's age, vintage year and time fixed effects, β_t. The sample of funds is restricted to funds that are listed in Morningstar database and file form D filings over period from January 2010 to December 2015, using backfill corrected sample of hedge fund returns observations. Figures in parentheses are the Newey and West (1987) heteroscedasticity and autocorrelation consistent standard errors clustered by month. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively. Table presents results of F-test for hypothesis that alphas of funds that are sold through in-house brokers is equal to alphas of funds that are sold through outside brokers.

Table 1.10: Heterogeneity of brokers

	Pre-fee alpha		
	(1)	(2)	(3)
B^I_{it}	-0.020***	-0.018***	-0.020***
	(0.003)	(0.003)	(0.003)
B^O_{it}	-0.019***	-0.020***	-0.020***
	(0.002)	(0.002)	(0.002)
$\ln(Asset_{it-1})$	0.008***	0.008***	0.008***
	(0.001)	(0.001)	(0.001)
Age_{it}	0.0000	-0.0001	-0.0006
	(0.0000)	(0.0002)	(0.0004)
Vintage	No	Yes	Yes
Time FE	No	No	Yes
# Obs.	28,304	28,304	28,304
R^2	1%	4%	5%
Ho: In-house = Outside			
F-test	0.02	-0.26	0.11
p-value	0.89	0.61	0.74

Table 1.10 estimates difference in pre-fee risk-adjusted performance between directly sold hedge funds and funds that are sold through in-house broker or outside broker with panel regression: $\hat{\alpha}_{it} = \beta_0 + \beta_{in} \cdot B^I_{it} + \beta_{out} \cdot B^O_{it} + \beta_x \cdot X_{it} + \beta_t + \tilde{\epsilon}_{it}$. B^I_{it} is a dummy variable that is equal to one when the fund is sold through in-house broker and is equal to zero otherwise. B^O_{it} is a dummy variable that is equal to one when the fund is sold through outside broker and is equal to zero otherwise. Regression includes fund level controls, X_{it}, such as fund's age, vintage year and time fixed effects, β_t. The sample of funds is restricted to funds that are listed in Morningstar database and file form D filings over period from January 2010 to December 2015, using backfill corrected sample of hedge fund returns observations. Figures in parentheses are the Newey and West (1987) heteroscedasticity and autocorrelation consistent standard errors clustered by month. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively. Table presents results of F-test for hypothesis that alphas of funds that are sold through in-house brokers is equal to alphas of funds that are sold through outside brokers.

Table 1.11: Fees of directly sold and broker sold funds

	Management fee		Incentive fee	
	(1)	(2)	(3)	(4)
B_i	0.000	—	-0.014***	—
	(0.000)	—	(0.004)	—
B_{it}^I	—	-0.000	—	0.006
	—	(0.000)	—	(0.006)
B_{it}^O	—	0.000	—	-0.015***
	—	(0.000)	—	0.004
Vintage	Yes	Yes	Yes	Yes
R^2	5%	5%	4%	5%
#Obs.	1,376	1,370	1,289	1,283
Ho: In-house = Outside				
F-test	—	0.95	—	5.95**
p-value	—	0.33	—	0.01

Table 1.11 presents estimation of cross-sectional regressions (1.12) and (1.11), comparing fee structure of directly sold and broker sold hedge funds. Columns (1) and (2) present results for management fees. Columns (3) and (4) present results for incentive fees. Figures in parentheses are the Newey and West (1987) heteroscedasticity and autocorrelation consistent standard errors. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively.

Table 1.12: Clientele of directly sold and broker sold funds

	Min. investment size		Aver. investment size	
	(1)	(2)	(3)	(4)
B_i	-0.272***	—	-12.033***	—
	(0.086)	—	(3.608)	—
B_{it}^I	—	-0.472***	—	-15.566***
	—	(0.217)	—	(4.623)
B_{it}^O	—	-0.282**	—	-5.716*
	—	(0.091)	—	(3.293)
Vintage	Yes	Yes	Yes	Yes
R^2	3%	3%	3%	3%
#Obs.	1,365	1,338	1,577	1,570
Ho: In-house = Outside				
F-test	—	0.69	—	4.76**
p-value	—	0.40	—	0.03

Table 1.12 presents estimation of cross-sectional regressions (1.12) and (1.11), comparing clientele of directly sold and broker sold hedge funds. Columns (1) and (2) present results for minimum investment size (in millions of $). Columns (3) and (4) present results for average investment size (in millions of $). Figures in parentheses are the Newey and West (1987) heteroscedasticity and autocorrelation consistent standard errors. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively.

Table 1.13: Average broker fee: bargaining power

Bargaining Power	5%	10%	20%	30%	50%	60%	70%	80%	90%	95%
Dollar Fee	$0.241	$0.482	$0.964	$1.446	$2.410	$2.893	$3.375	$3.857	$4.339	$4.580

Table 1.13. This table presents estimates of average annual fee (in millions $) that fund pays to broker, who intermediates fund's capital raising process. Fee is estimated for a given broker's bargaining power. The sample of funds is restricted to funds that are listed in Morningstar database and may be classified as broker-sold funds according to information in form D filings. Annual dollar broker fees are estimated under considered fee specification, using the methodology that is described in Appendix. For a given bargaining power table displays average annual dollar fee across broker-sold funds.

1.6 Capital inflows estimation

To estimate capital inflows into industry, I use the following methodology. Among various information that fund reports in its form D filings, is up-to-date information on total amount of capital raised from investors, which is reported in the field Total Amount Sold.[9] To estimate the amount of capital raised by the fund, we should consider two cases: capital inflows at fund's inception and capital inflows during the life of the fund. In the first case, amount of capital raised at inception is directly reported in Total Amount Sold variable. In the second case, it may be estimated as an increment of Total Amount Sold variable between two consecutive fund's filings. For example, Citadel Global Equities Fund[10], that was opened in July, 2009, reports capital inflow of $100 millions in its first filing. The fund reports $ 153 millions as total amount sold to investors in its next filing in August, 2010. Thus, total capital inflows into the fund between July, 2009 and August, 2010 build up to $53 millions. As funds sometimes file amendment to their form D filings more than once a year, I estimate an amount of capital raised, using information from the latest filing in a given year.

Due to self-reporting nature of form D filings, there are some funds in the sample that mistakenly report their yearly inflows instead of up-to-date total amount of money raised from investors, which is required by Regulation D. I identify those funds when inflow that are estimated using the introduced methodology are negative.[11] Funds that misreport information about total amount of capital raised are excluded from analysis.

Unfortunately, form D filings do not allow to recover an exact timing of capital inflows, but rather estimate capital inflows during the period between the filings. Therefore, additional assumptions are required to determine the year of capital inflows into the fund. As above, I consider two scenarios separately. The first case corresponds to capital raising at fund's inception. In this case, I assume that capital

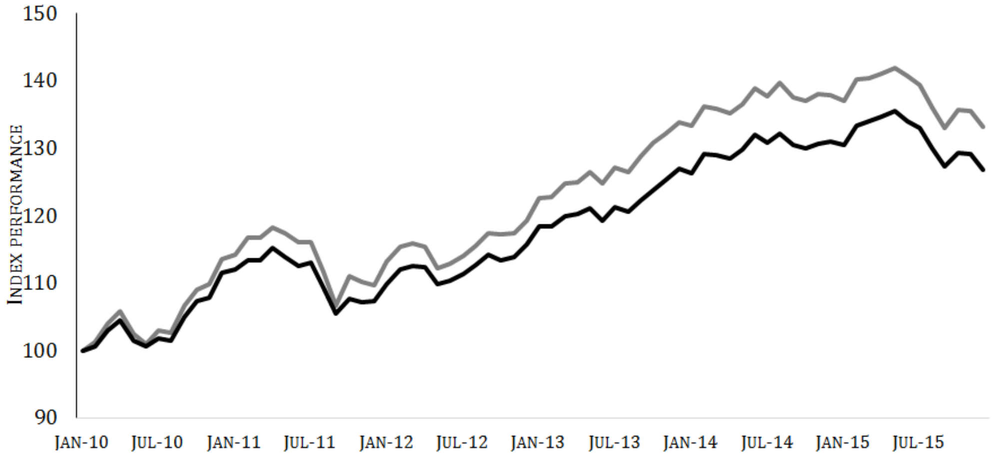
[9]Total Amount Sold is reported in field (b) of form D Item 13 (Offering and Sales Amounts).
[10]Citadel Global Equities Fund LLC is identified by Central Index Key (CIK) 1468448.
[11]By construction capital inflows is non-negative variable.

inflows happened in the year of the first fund's form D filing. The second scenario corresponds to the situation when fund is already in operation, meaning that fund has filed several form D filings. Specifically, the earlier filing of the fund is registered in month, m_1, of year, y_1, while the next consecutive filing occurs in month, m_2, of year, y_2. In this scenario, I assume that capital inflows occurred in year $y_1(y_2)$ if the period between the two filings mostly belongs to year $y_1(y_2)$. Using the example of Citadel fund, I estimate that capital inflows of $100 millions happened in 2009 (corresponds to the first case) and $53 millions were raised in 2010 (corresponds to the second case).

1.7 Robustness checks

Figure 1.7: Performance of hedge fund portfolios: after fee + no bias correction

Panel A: Equally-Weighted Portfolios



Panel B: Value-Weighted Portfolios

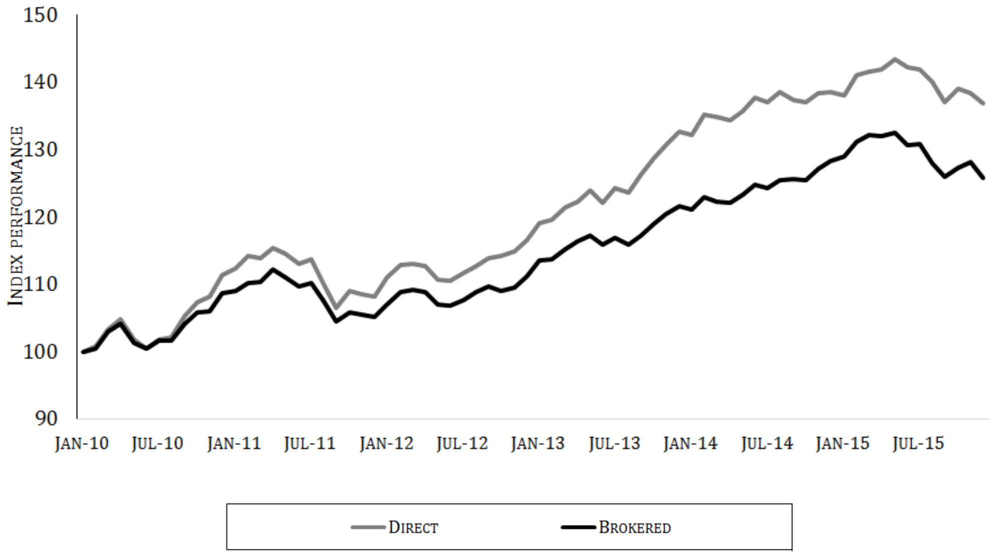


Figure 1.7 displays after-fee performance of fund of directly sold hedge funds (grey solid line) relative to performance of fund of broker sold hedge funds (black solid line) over the period from January 2010 to December 2015, assuming initial investment of $100. The sample of funds consists of funds that are listed in Morningstar database and file form D filings. Panel A displays after-fee performance of funds of funds where constituent hedge funds are equally-weighted. Panel B displays after-fee performance of funds of funds where constituent hedge funds are value-weighted. Returns of funds are adjusted for backfill bias.

Table 1.14: Performance of Hedge Fund Portfolios: After Fee + Bias

	$\bar{R}$	$\hat{\alpha}$	$\hat{\beta}_{Mkt}$	$\hat{\beta}_{SmB}$	$\hat{\beta}_{T10y}$	$\hat{\beta}_{Cr.Spr.}$	$\hat{\beta}_{pBD}$	$\hat{\beta}_{pFX}$	$\hat{\beta}_{pCOM}$	R^2
	PANEL A: Equally-Weighted Portfolio									
Direct	4.793%**	4.421%**	0.12*	0.38***	0.10	0.25***	-0.02*	0.01	-0.01	68%
	(0.02)	(0.02)	(0.06)	(0.04)	(0.07)	(0.09)	(0.01)	(0.01)	(0.01)	
Brokered	3.968%*	3.366%*	0.12**	0.32***	0.07	0.18**	-0.07*	0.01	-0.01*	68%
	(0.02)	(0.02)	(0.05)	(0.03)	(0.06)	(0.07)	(0.01)	(0.01)	(0.01)	
	PANEL B: Value-Weighted Portfolio									
Direct	5.391%	4.433%**	0.13***	0.31***	0.07	0.16**	-0.02*	0.01	-0.01	66%
	(0.02)	(0.02)	(0.05)	(0.03)	(0.06)	(0.07)	(0.01)	(0.01)	(0.01)	
Brokered	4.157%	3.552%**	0.12***	0.25***	0.05	0.15**	-0.01*	0.01	-0.01	62%
	(0.02)	(0.01)	(0.04)	(0.03)	(0.05)	(0.06)	(0.01)	(0.01)	(0.01)	

Table 1.14. Results of Fung and Hsieh (2004) seven-factor models estimation for portfolio of directly sold and broker sold funds are presented in Table 1.14. Panel A displays results for the equally-weighted portfolio of funds, while Panel B reports results for the value-weighted portfolio of funds. Portfolios of directly sold and broker sold funds (that is constructed using a sub-sample of funds that report to Morningstar and file forms D) are reported in row Direct and row Brokered, respectively. The seven-factor model (1.1) is estimated using after-fee monthly returns between January 2010 and December 2015, where the first 24-months of fund's performance are excluded to adjust for back-fill bias. Table displays estimated annualized expected annualized excess return of portfolio,$\bar{R}$, estimated annualized alpha, $\hat{\alpha}$, the estimated exposures to the market, $\hat{\beta}_{Mkt}$, the estimated exposure to size spread factor, $\hat{\beta}_{SmB}$, the estimated exposure to yield curve level factor, $\hat{\beta}_{T10y}$, the estimated exposure to credit spread factor, $\hat{\beta}_{Cr.Spr.}$, and the estimated exposures to bond, commodity and forex trend-following factors, $\hat{\beta}_{pBD}$, $\hat{\beta}_{pFX}$ and $\hat{\beta}_{pCOM}$, as well as the adjusted R^2. Figures in parentheses are the Newey and West (1987) heteroscedasticity and autocorrelation consistent standard errors. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively.

Table 1.15: Performance of hedge fund portfolios: pre fee + bias

	$\bar{R}$	$\hat{\alpha}$	$\hat{\beta}_{Mkt}$	$\hat{\beta}_{SmB}$	$\hat{\beta}_{T10y}$	$\hat{\beta}_{Cr.Spr.}$	$\hat{\beta}_{pBD}$	$\hat{\beta}_{pFX}$	$\hat{\beta}_{pCOM}$	R^2
	PANEL A: EQUALLY-WEIGHTED PORTFOLIO									
Direct	6.167%***	5.781%***	0.12*	0.39***	0.11	0.25***	-0.02*	0.01	-0.01	69%
	(0.02)	(0.02)	(0.06)	(0.04)	(0.07)	(0.09)	(0.01)	(0.01)	(0.01)	
Brokered	5.120%***	4.481%**	0.17**	0.33***	0.07	0.18**	-0.01*	0.01	-0.01*	69%
	(0.02)	(0.02)	(0.05)	(0.03)	(0.06)	(0.07)	(0.01)	(0.01)	(0.01)	
	PANEL B: VALUE-WEIGHTED PORTFOLIO									
Direct	6.620%***	5.532%***	0.14***	0.32***	0.07	0.16**	-0.02*	0.01	-0.01	65%
	(0.02)	(0.02)	(0.05)	(0.03)	(0.06)	(0.07)	(0.01)	(0.01)	(0.01)	
Brokered	5.504%***	4.948%***	0.11***	0.26***	0.05	0.15**	-0.01	0.01	-0.01	61%
	(0.02)	(0.01)	(0.04)	(0.03)	(0.05)	(0.06)	(0.01)	(0.01)	(0.01)	

TABLE 1.15. RESULTS OF FUNG AND HSIEH (2004) SEVEN-FACTOR MODELS ESTIMATION FOR PORTFOLIO OF DIRECTLY SOLD AND BROKER SOLD FUNDS ARE PRESENTED IN TABLE 1.15. PANEL A DISPLAYS RESULTS FOR THE EQUALLY-WEIGHTED PORTFOLIO OF FUNDS, WHILE PANEL B REPORTS RESULTS FOR THE VALUE-WEIGHTED PORTFOLIO OF FUNDS. PORTFOLIOS OF DIRECTLY SOLD AND BROKER SOLD FUNDS (THAT IS CONSTRUCTED USING A SUB-SAMPLE OF FUNDS THAT REPORT TO MORNINGSTAR AND FILE FORMS D) ARE REPORTED IN ROW DIRECT AND ROW BROKERED, RESPECTIVELY. THE SEVEN-FACTOR MODEL (1.1) IS ESTIMATED USING PRE-FEE MONTHLY RETURNS BETWEEN JANUARY 2010 AND DECEMBER 2015. TABLE DISPLAYS ESTIMATED ANNUALIZED EXPECTED ANNUALIZED EXCESS RETURN OF PORTFOLIO, $\bar{R}$, ESTIMATED ANNUALIZED ALPHA, $\hat{\alpha}$, THE ESTIMATED EXPOSURES TO THE MARKET, $\hat{\beta}_{Mkt}$, THE ESTIMATED EXPOSURE TO SIZE SPREAD FACTOR, $\hat{\beta}_{SmB}$, THE ESTIMATED EXPOSURE TO YIELD CURVE LEVEL FACTOR, $\hat{\beta}_{T10y}$, THE ESTIMATED EXPOSURE TO CREDIT SPREAD FACTOR, $\hat{\beta}_{Cr.Spr.}$, AND THE ESTIMATED EXPOSURES TO BOND, COMMODITY AND FOREX TREND-FOLLOWING FACTORS, $\hat{\beta}_{pBD}$, $\hat{\beta}_{pFX}$ AND $\hat{\beta}_{pCOM}$, AS WELL AS THE ADJUSTED R^2. FIGURES IN PARENTHESES ARE THE NEWEY AND WEST (1987) HETEROSCEDASTICITY AND AUTOCORRELATION CONSISTENT STANDARD ERRORS. STATISTICAL SIGNIFICANCE AT THE 1%, 5% AND 10% LEVELS IS DENOTED BY *, **, AND *** RESPECTIVELY.

Table 1.16: Alphas of directly and broker sold hedge funds

	Alpha		
	(1)	(2)	(3)
PANEL A: After-fee			
B_{it}	-0.012***	-0.013***	-0.013***
	(0.002)	(0.002)	(0.002)
$\ln(Asset_{it-1})$	—	0.007***	0.007***
	—	(0.001)	(0.001)
Age_{it}	—	-0.0002	-0.001**
	—	(0.0002)	(0.0004)
Vintage	No	Yes	Yes
Time FE	No	No	Yes
# Obs.	26,572	26,572	26,572
R^2	0.1%	4%	6%
PANEL B: Pre-fee			
B_{it}	-0.015***	-0.018***	-0.019***
	(0.002)	(0.002)	(0.002)
$\ln(Asset_{it-1})$	—	0.009***	0.009***
	—	(0.001)	(0.001)
Age_{it}	—	-0.0002	0.0007
	—	(0.0002)	(0.0004)
Vintage	No	Yes	Yes
Time FE	No	No	Yes
# Obs.	25,712	25,712	25,712
R^2	0.2%	4%	6%

Table 1.16 presents estimates of difference in risk-adjusted performance between directly sold and broker sold hedge funds with panel regression $\hat{\alpha}_{it} = \beta_0 + \beta_B \cdot B_{it} + \beta_s \cdot X_{it-1} + \beta_t + \tilde{\epsilon}_{it}$. Fund level controls X_{it-1} include logarithm of assets under management in the previous period, age, and vintage year and time fixed effects β_t. Panel A displays results for after-fee alphas of hedge funds. Panel B displays results for pre-fee alphas of hedge funds. The sample covers hedge funds that are listed in Morningstar database and file form D filings over period from January 2010 to December 2015. Figures in parentheses are the Newey and West (1987) heteroscedasticity and autocorrelation consistent standard errors clustered by month. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively.

Table 1.17: Value added by directly and broker sold hedge funds

	Dollar value added		
	(1)	(2)	(3)
PANEL A: After-fee			
B_{it}	-0.135***	-0.160***	-0.169***
	(0.060)	(0.054)	(0.055)
Age_{it}	—	0.002	-0.031***
	—	(0.003)	(0.009)
Vintage	No	Yes	Yes
Time FE	No	No	Yes
# Obs.	26,472	26,472	26,472
R^2	0.02%	3%	4%
PANEL B: Pre-fee			
B_{it}	-0.101***	-0.127***	-0.141***
	(0.068)	(0.065)	(0.066)
Age_{it}	—	0.001	-0.026**
	—	(0.004)	(0.009)
Vintage	No	Yes	Yes
Time FE	No	No	Yes
# Obs.	25,712	25,712	25,712
R^2	0.01%	4%	4%

Table 1.17 presents estimates of difference in dollar value added (in millions of dollars) by directly sold and broker sold hedge funds with panel regression $\hat{S}_{it} = \beta_0 + \beta_B \cdot B_{it} + \beta_s \cdot X_{it} + \beta_t + \tilde{\epsilon}_{it}$. Fund level controls X_{it} include fund's age, vintage year and time fixed effects β_t. Panel A displays results for after-fee dollar value added by hedge funds. Panel B displays results for pre-fee dollar value added of hedge funds. The sample covers hedge funds that are listed in Morningstar database and file form D filings over period from January 2010 to December 2015 with an adjustment for backfill bias. Figures in parentheses are the Newey and West (1987) heteroscedasticity and autocorrelation consistent standard errors clustered by month. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively.

Table 1.18: Heterogeneity of brokers

	After-fee alpha		
	(1)	(2)	(3)
B^I_{it}	-0.023***	-0.022***	-0.022***
	(0.003)	(0.002)	(0.002)
B^O_{it}	-0.015***	-0.014***	-0.014***
	(0.002)	(0.002)	(0.002)
$\ln(Asset_{it-1})$	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)
Age_{it}	-0.0001***	-0.0002	-0.0010**
	(0.0000)	(0.0002)	(0.0004)
Vintage	No	Yes	Yes
Time FE	No	No	Yes
# Obs.	32,026	32,026	32,026
R^2	0.7%	3%	4%
Ho: In-house = Outside			
F-test	3.73*	4.36**	4.74**
p-value	0.06	0.04	0.03

Table 1.18 estimates difference in after-fee risk-adjusted performance between directly sold hedge funds and funds that are sold through in-house broker or outside broker with panel regression: $\hat{\alpha}_{it} = \beta_0 + \beta_I \cdot B^I_{it} + \beta_O \cdot B^O_{it} + \beta_x \cdot X_{it} + \beta_t + \tilde{\epsilon}_{it}$. B^I_{it} is a dummy variable that is equal to one when the fund is sold through in-house broker and is equal to zero otherwise. B^O_{it} is a dummy variable that is equal to one when the fund is sold through outside broker and is equal to zero otherwise. Regression includes fund level controls X_{it}, such as fund's age, vintage year and time fixed effects β_t. The sample of funds is restricted to funds that are listed in Morningstar database and file form D filings over period from January 2010 to December 2015, using full sample of hedge fund returns observations. Figures in parentheses are the Newey and West (1987) heteroscedasticity and autocorrelation consistent standard errors clustered by month. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively. Table presents results of F-test for hypothesis that alphas of funds that are sold through in-house brokers is equal to alphas of funds that are sold through outside brokers.

Table 1.19: Heterogeneity of brokers

	Pre-fee alpha		
	(1)	(2)	(3)
B^I_{it}	-0.022***	-0.019***	-0.020***
	(0.003)	(0.002)	(0.002)
B^O_{it}	-0.021***	-0.020***	-0.020***
	(0.001)	(0.001)	(0.001)
$\ln(Asset_{it-1})$	0.008***	0.009***	0.009***
	(0.001)	(0.001)	(0.001)
Age_{it}	-0.0000*	-0.0002	-0.0008
	(0.0000)	(0.0002)	(0.0004)
Vintage	No	Yes	Yes
Time FE	No	No	Yes
# Obs.	30,929	30,929	30,929
R^2	1%	4%	5%
Ho: In-house = Outside			
F-test	0.04	0.18	0.38
p-value	0.83	0.67	0.53

Table 1.19 estimates difference in pre-fee risk-adjusted performance between directly sold hedge funds and funds that are sold through in-house broker or outside broker with panel regression: $\hat{\alpha}_{it} = \beta_0 + \beta_I \cdot B^I_{it} + \beta_O \cdot B^O_{it} + \beta_x \cdot X_{it} + \beta_t + \tilde{\epsilon}_{it}$. B^I_{it} is a dummy variable that is equal to one when the fund is sold through in-house broker and is equal to zero otherwise. B^O_{it} is a dummy variable that is equal to one when the fund is sold through outside broker and is equal to zero otherwise. Regression includes fund level controls X_{it}, such as fund's age, vintage year and time fixed effects β_t. The sample of funds is restricted to funds that are listed in Morningstar database and file form D filings over period from January 2010 to December 2015, using full sample of hedge fund returns observations. Figures in parentheses are the Newey and West (1987) heteroscedasticity and autocorrelation consistent standard errors clustered by month. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively. Table presents results of F-test for hypothesis that alphas of funds that are sold through in-house brokers is equal to alphas of funds that are sold through outside brokers.

Chapter 2

Order Shredding, Invariance, and Stock Returns

2.1 Introduction

There is a long-standing debate on what is a good way to model security price dynamics. It is crucial for our understanding of financial markets. Progress has been made in this important area, but there is still no fully satisfactory answer as to the mechanism of how returns process is generated. We propose a novel structural model for price dynamics within the paradigm of market microstructure invariance, developed recently by Kyle and Obizhaeva (2016) and found to be successful in explaining a number of empirical regularities in the data.

It is known that empirical price processes depart from the Brownian motion, and price changes are not distributed as normal random variables. Several alternative models have been proposed in the literature. Mandelbrot (1963) suggests that price changes may be better described by a stable Pareto distribution with fat tails. Mandelbrot and Taylor (1967) and Clark (1973) propose that price processes seem to be closely related to the Brownian motions that evolves not in calendar time but rather in some business time, which is linked to either arrival of transactions or trading volume, respectively. Jones, Kaul and Lipson (1994), Hasbrouck (1999), Ané and Geman (2000), Andersen et al. (2015) study what business clock best fits the data. In comparison to these approaches, our structural model of returns dynamics comes from explicit modelling of how traders trade in real financial markets.

The backbone of our model is the arrival process of investment ideas, or bets, placed by fundamental traders into the market. This process has been earlier calibrated within the microstructure invariance paradigm by Kyle and Obizhaeva (2016) who suggested that bets arrive according to a stochastic process with an expected arrival rate per day approximately proportional to the 2/3 exponent of trading volume and volatility, and the distribution of bet sizes closely resemble log-normal random variables with log-variance of 2.53. This large log-variance implies frequent arrivals of very large bets. We assume that traders execute bets by splitting them into sequences of transactions according to some bet-shredding algorithm in order to reduce price impact; we model price impact in response to each transaction as suggested by invariance-based market impact model. We also introduce arbitrageurs who implement order anticipation algorithms based on predictive models to detect execution of large bets and trade ahead of them with hope to make some money. Market makers clear the market.

The core idea of market microstructure invariance is that business time runs faster in liquid markets and slower in illiquid markets, whereas a trading game itself that traders play remains invariant. Our structural model ultimately differs across stocks and time periods, because it is based on different arrival processes of bets. We also calibrate bet shredding parameters using the method of simulated moments in order to match the cross-sectional and time-series variation in empirical moments of stock returns.

We update the evidence on cross-sectional and time-series properties of moments of daily U.S. stock returns using the Center for Research in Security Prices (CRSP) database. We find that idiosyncratic excess kurtoses tend to be positive and decrease with trading activity of stocks; the ratio of idiosyncratic kurtosis for the median least active stocks to that of the most active stocks is almost always greater than one, but this difference becomes less pronounced over time. The total kurtosis without any adjustment for market returns is also larger for the less active stocks; these patterns reverse over during market crashes, when kurtosis of liquid stocks becomes bigger relative to kurtosis of illiquid stocks, possibly due to staleness of prices. The idiosyncratic skewness does not exhibit any distinctive cross-sectional patterns and fluctuates over time around a small positive value, often dropping to negative values during market crashes.

Our calibration allows us to discuss the properties of implied bet shredding parameters. Under the assumption that traders target a fixed proportion of overall expected trading volume, we find that traders target a bigger proportion when executing bets in less liquid securities. We also find that bet-shredding has intensified over time, and now traders choose to execute bets over two or three times longer horizons than in 1950s. The prevalence of shredding in modern markets have been also documented empirically in Kyle, Obizhaeva and Tuzun (2016), Angel, Harris and Spatt (2015), and Garvey, Huang and Wu (2017). Bet shredding is also optimal for traders who seek to minimize transaction costs, as shown theoretically by Kyle, Obizhaeva and Wang (2017). Our structural model can be used as a vehicle to gain insight into hard-to-observe parameters of trading.

There are two different approaches to modelling securities returns. The first approach, usually preferred by economists, relies on calibration of structural equilibrium models in order to make sure that models are internally consistent with market clearing and strategic optimizing behavior of traders; the example is a structural framework of Campbell and Kyle (1993) that helps to model permanent and temporary shocks to prices. The second approach, usually preferred by statisticians and econophysicists, relies on agency-based models that simulate actions and interactions of traders to study their effects on the system as a whole, but often assume mechanic—rather than driven by economic incentives—order placement strategies and price formation process; examples include Cont and Bouchaud (2000), Farmer, Patelli and Zovko (2005), Cont, Stoikov and Talreja (2010), and Ladley (2012), among others.

Our model is a combination of these two approaches, taking the best of both of them. On the one hand, we pay careful attention to the modelling of how people trade as in agency-based models. Indeed, our groups of market participants closely resemble the classification of Kirilenko et al. (2017), who provide a micro-level empirical description of the structure of trading in the market of the E-mini S&P 500 futures during the Flash crash on May 6, 2010. On the other hand, each part of our model is guided by the insights of existing theories of financial economics. Bets arrive according to general invariance predictions, which one can derive within a number of equilibrium models such as the dynamics model of Kyle and Obizhaeva (2017*c*) and the one-period model of Kyle, Obizhaeva and Wang (2017). Bet-shredding algorithms are similar to optimal trading strategies suggested by the literature on optimal execution, such as Bertsimas and Lo (1998), Almgren and Chriss (2000), and Obizhaeva and Wang (2013) among others. Arbitrageurs insure that prices follow a martingale and markets are efficient. Market makers insure that markets clear.

This paper is organized as follows. Section I presents empirical analysis of time-series and cross-sectional properties of moments for returns of the U.S. stock market. Section II describes a structural model of returns dynamics based on market microstructure invariance with bet shredding and arbitrage trading. Section III discuss its calibration and properties of implied parameters. Section IV concludes.

2.2 Moments of daily returns: empirical analysis

2.2.1 Data

We examine cross-sectional and time-series properties of moments of daily U.S. stock returns using the Center for Research in Security Prices (CRSP) database. Only common stocks (CRSP share codes of 10 and 11) listed on the New York Stock Exchange (NYSE), the American Stock Exchange (Amex), NASDAQ, and NYSE Arca in the period of January 1926 through December 2016 are included in the sample. ADRs, REITS, and closed-end funds are excluded.

Estimates of higher moments are very sensitive to large price changes, outliers, and errors in the data. We do not windsorize of the data, because we do not want to eliminate most important large but rare observations. Instead, we carefully clean the data by trying to filter out outliers and errors, while keeping large observations caused by execution of large bets, market crashes, or other events.

First, we adjust for stale prices. For each security, the CRSP mixes two time series. For days with transactions, the database reports last transaction prices at the close. For days with no transactions, the database reports averages of bid and ask prices, marking these averages with a negative sign; these observations are often not representative of true prices, at which traders could actually transact during that day. The mixture two price series often leads to large temporary deviations in the composite series. For example, for the six days from May 17, 2010 to May 24, 2010, one finds the following prices in the CRSP for the stock of the firm Ikonics: \$7.1, \$6.52, -\$7.225, -\$12.76, -\$7.07, and \$6.81; the three negative prices mean that there were no transactions at these three days and the average bid-ask prices are reported instead of actual transaction prices. If one would simply change their negative signs into positives sign and calculate time-series of returns, then he will get -8, 11, 77, -45, and -4 percents with large positive price change followed by large negative price changes in the middle of the sample. At the same time, Yahoo Finance reports \$7.1, \$6.52, \$6.52, \$6.52, \$6.52, and \$6.81 for the same days implying returns of -8, 0, 0, 0, and 4 percents. The two time series will have very different estimates of moments, especially for higher moments such as kurtosis. To circumvent this problem, we use only transaction prices when available, accumulate returns from the very last transaction price reported, and assign returns of zero to all days with no transactions.

Second, there remain many large zigzag price changes in the sample. It is usually unclear whether these are actual prices that we need to keep or errors that we need to eliminate. As describe in Fischer (1963), the process of creating the CRSP database required a lot of effort and involved a lot of data cleaning. Some errors though may still exist due to mistakes in original data collected by exchanges, incorrect conversion of the data from paper books into electronic databases, inconsistent adjustment for splits and dividends, confusion with tickers, inaccurate treatment of trades in error accounts that are often cancelled within a few days, and many other reasons. We checked manually whether large zigzag price deviations in the CRSP coincide with price patterns in other datasets or whether they can be attributed to some events. Since unexplained temporary price swings occur especially often in the earlier pre-war part of the sample, we choose to focus on the data from January 1950 to December 2016.

Third, we eliminate daily observations with fewer than 100 shares traded, because transaction prices on these days may also be not representative of true prices. Small trades may be used as vehicles for side payments between traders, soft commissions, or transactions by market makers who are required to maintain some minimal trading activity in illiquid stocks.

Finally, we exclude stocks with more than fifteen no-trade days in a month and daily volatility of less than one percent. We also exclude stocks with the median of prices being less than $5, because estimates of their returns moments are very unstable, as errors are especially critical for these stocks.

We excluded about 45% of observations from the original sample. The final sample includes 1,576,834 observations for 1,089 months and 19,922 stocks. The number of stocks vary significantly throughout the sample. Initially, there were only NYSE stocks. The number of stocks rose steadily from 500 stocks in 1926 to 1,100 by 1962, then jumped to about 2000 in July 1962 and 5000 in November 1982, when the Amex and NASDAQ stocks were included into the sample, respectively. The number of stocks slightly declined after the market crash of October 1987 and increased during the dot-com bubble 1995 though 2000, peaking at 7300 in 1997. Afterwards, the number of stocks dropped, and it is equal to about 4000 at the end of the sample.

2.2.2 Estimation of moments

The estimate moments of log-returns are known to be sensitive to outliers. We next obtain these estimates using robust estimation methods.

We modify the sample estimates of higher moments that usually use the sample estimates of means and standard deviations and that are prone to several biases. First, the sample means introduce forward-looking biases by making returns look less volatile than they are in reality. In our estimation of higher moments, we instead assume that daily stock returns have zero mean.

Second, the sample standard deviation tends to be overestimated during volatile periods, and these biased estimates of volatility in turn make the sample estimates of kurtosis underestimated. We assume means of zero instead sample means and pre-estimate volatility over the previous three-month period, using one of the robust iterative estimation methods; we also consider only three-month periods with more than fifteen non-zero observations of returns and average price above \$5. We first estimate volatility over the entire three-month sample, then exclude observations with absolute values bigger than two sigma, estimate volatility again and repeat this procedure until either the difference in subsequent volatility estimates becomes less than one basis point or the number of excluded outliers exceeds five percent of the original sample. These are conservative measure of volatility robust to outliers. For robustness, we also consider volatility estimated using Inter Quantile Range methods (IQR-α methods), as proposed by Aucremanne(2004) and Kimber(1990), respectively, as well as Median Absolute Deviation methods (MAD-β methods), as proposed by Iglewicz and Hoaglin (1993) and Hampel(1974); all results (not reported) are qualitatively and quantitatively similar to our main reported findings.[1]

For each month and each stock, we then calculate the estimates of skewness and kurtosis using the formulas for sample moments but replacing sample means and sample standard deviations with our robust estimates. We apply this procedure

[1]In IQR-α method, volatility is estimated on reduced sample $[P_{25} - \alpha \cdot [P_{75} - P_{50}], P_{75} + \alpha \cdot [P_{50} - P_{25}]]$, where P_x denote the percentile x, with most outliers excluded ($\alpha = 3$ and $\alpha = 1.5$). In MAD-β method, volatility is estimated on entire sample excluding observations with $|M_i| > \beta$, where $M_i = 0.6745 \cdot (x_i - med(X))/MAD$ and $MAD = med(|x_i - med(X)|)$ ($\beta = 3$ and $\beta = 2$).

for both the sample of returns and the sample of idiosyncratic returns, obtained by subtracting the contemporaneous values of index returns under the assumption that all stocks' betas are equal to one.

2.2.3 Time-series and cross-section of empirical moments

To examine empirically cross-sectional patterns, we split all stocks in ten groups based on daily trading activity, an important characteristic of securities reflecting the speed with which markets operate and levels of liquidity. Trading activity is defined as the product of dollar volume and volatility and represent the total amount of risk transferred per day. For each stock and each month, we calculate trading activity as the product of the average daily dollar volume and volatility over the previous three months. We then sort all stocks each month into ten groups based on trading activity. The breakpoints are chosen to be 30th, 50th, 60th, 70th, 75th, 80th, 85th, 90th, 95th of the NYSE traded stocks. Group 1 consists of least actively traded stocks. Group 10 consists of most actively traded stocks.

Table 2.1 presents a detailed time-series and cross-sectional summary statistics for high moments of idiosyncratic daily returns. The medians of sample moments (volatility, skewness, and kurtosis) are shown for seven decades between 1950 and 2016 and for trading activity groups 1, 3, 5, 8, and 10.

Figure 2.1 shows the monthly time series of the 12-month moving averages of the sample medians of sample kurtosis of idiosyncratic daily stock returns for the same trading activity groups. The estimates are averaged over a twelve month period to smooth out unstable estimates. Figure 2.3 shows similar moving averages of monthly kurtosis estimates for daily stock returns without any adjustment for market movements. We can draw several conclusions from table 2.1 and figure 2.1.

First, idiosyncratic kurtoses tend to decrease with trading activity. The daily kurtoses of the least liquid stocks are stable, ranging between 6.60 and 8.47 across decades and thus implying fat tails. The daily kurtoses of the most liquid stocks slightly increase over time from 2.66 in decade 1950-1960 to 4.23 in 2010-2016;

their values remain being close to 3, suggesting that distributions of their daily idiosyncratic returns closely resemble the log-normal. Figure 2.2 reveals similar patterns. The figure shows that, depicted by the solid horizontal line, the ratio of idiosyncratic kurtoses of stocks in group 1 to kurtoses of stocks in group 10 is bigger than one for each month throughout the sample, except for the month of September 2008 when uncertainty reached its peak during the financial crisis. The difference in kurtoses of least and most active stocks becomes less pronounced over time. Similar patterns are observed for kurtoses of total daily returns in figure 2.4. Ratios of kurtoses of stocks in group 1 to kurtoses of stocks in group 10 drop below one only during a few episodes in 1956, 1962-1963, 1987-1988, and 1993-1994; these breaks might be attributed to Kennedy slide in 1962, market crash in October 1987, and mini-crash in October 1989, respectively.

Second, the monthly time series of the 12-month moving averages of the sample kurtosis medians in figures 2.1 and 2.3 are relatively stable over time, but exhibit several significant spikes in May 1962, October 1987, August 1998, September 2008, and August 2011. Even though the events that triggered large price changes are relatively short lived, these spikes continue for twelve months due to our calculations of moving averages using the twelve lags. These spikes correspond to volatile times mentioned above as well as to the LTCM collapse in 1998. During these periods, the idiosyncratic kurtoses continue to be larger for less liquid stocks, but the patterns for kurtosis sometimes flip, and kurtosis of liquid stocks becomes bigger relative to kurtosis of illiquid stocks, possibly due to staleness of their price.

Figure 2.5 shows the time series of idiosyncratic skewness for the trading activity groups. Idiosyncratic skewness is usually slightly positive, fluctuating between 0.06 to 0.36 across decades and decreasing over time, on average, from 0.26 in decade 1950-1960 to 0.10 in 2010-2016, as shown in table 2.1. During market dislocations, skewness tends to drop. Skewness does not exhibit any distinctive cross-sectional patterns. It remains to be close to zero, thus suggesting that the distribution of returns is close to a log-normal.

Figure 2.6 shows monthly time series of 12-month moving average of median sample volatility of idiosyncratic daily for the five trading activity groups with two pronounced spikes during the dot-com bubble in 2000-2001 and financial crisis of 2008-2009.

In what follows, we will propose a structural model of price dynamics and calibrate it to match the cross-sectional and time-series patterns of higher moments in table 2.1.

2.3 Invariance-implied structural model of price dynamics

In this section, we describe a structural model of stock returns dynamics in financial markets. There are three market participants: traders, intermediaries, and arbitragers. Traders are institutional asset managers and retail investors who arrive to the market with some trading ideas, or bets, and execute these bets by shredding them over time based on bet shredding algorithms. We assume that these bets are generated according to the implications of market microstructure invariance. Intermediaries such as traditional market makers and high-frequency traders clear the market by taking the other side of these transactions. Meanwhile arbitrageurs try to detect large bets of traders in the order flow and profit by trading ahead of them.

2.3.1 Bets of traders

We start by describing trading strategies of institutional asset managers and retail investors. These traders submit bets based on either some investment ideas or their needs to rebalance portfolios. Bets move prices and induce volatility. Small bets lead to small price changes, large bets trigger large price changes. Invariance implies a specific structure of order flow, i.e. the number of bets and distribution of their size for different markets.

Consider a stock i at day t with returns volatility σ_{it}, share volume V_{it}, dollar price P_{it}, and trading activity

$$W_{jt} = \sigma_{jt} \cdot P_{jt} \cdot V_{jt}. \tag{2.1}$$

Let γ_{it} denote the number of bets placed at day t in the market of stock i. Suppose that a sequence of bets executed at day t is Q_{it1}, Q_{it2},..$Q_{it\gamma_{it}}$; each kth bet Q_{itk} is measured in shares, bets are positive for buys and negative for sells, both arriving with equal probabilities of 1/2. Let $\tilde{Q}_{it}$ denote a random variable whose probability distribution represents the signed size of bets and let $\tilde{\gamma}_{it}$ denote a random variable whose probability distribution represents the expected arrival rate.

Kyle and Obizhaeva (2016) calibrate these distributions using the sample of portfolio transitions executed over the period 2001 through 2005 in the U.S. stock market as the main benchmark sample. As the first-order approximation, they find that $|\tilde{Q}_{it}|$ is well described by a log-normal distribution with log-variance $\sigma_Q^2 = 2.53$ and $\tilde{\gamma}_{it}$ is a Poisson variable with the mean $\bar{\gamma}_{it}$; the means of both of these random variables vary across days t and stocks i,

$$\bar{\gamma}_{it} = 85 \cdot \Big[\frac{W_{it}}{(0.02)(40)(10^6)}\Big]^{2/3}. \tag{2.2}$$

$$\ln\Big[\frac{|\tilde{Q}_{it}|}{V_{jt}}\Big] \approx -5.71 - \frac{2}{3} \cdot \ln\Big[\frac{W_{it}}{(0.02)(40)(10^6)}\Big] + \sqrt{2.53} \cdot \tilde{Z}, \qquad \tilde{Z} \sim N(0,1). \tag{2.3}$$

The 2/3 exponents in these formulas are implications of invariance; the constants 85, −5.71, and 2.53 are calibrated from the data. For the benchmark stock with daily volatility $\sigma = 0.02$, volume $V = 10^6$, and price $P = 40$, for example, there are on average 85 bets per day, their median dollar size is $exp(-5.71) \cdot V \cdot P$ or \$132,000, and their average dollar size is $exp(-5.71 + 0.5\sigma_Q^2) \cdot V \cdot P$ or \$470,000. Both the number of bets $\tilde{\gamma}_{it}$ and their size $|\tilde{Q}_{it}|$ increase with dollar volume and returns volatility.

Intermediaries take the other side of these bets by setting market clearing prices.[2]

[2]Under the assumption that the volume multiplier $\zeta = 2$, as consistent with our assumption that intermediaries take the other side of these bets, and the portfolio transition size multiplier $\delta = 1$.

Kyle and Obizhaeva (2016) analyse by how much each bet on average moves prices and calibrate several price impact models. The first model is the linear price impact model. According to its log-linear version, buying or selling Q shares of a stock with a current stock price P moves the price on average by $\Delta P(Q)$ such that

$$\ln\left(1+\frac{\Delta P(Q)}{P}\right)=\frac{\sigma_{it}}{0.02}\Big(\frac{\bar{\kappa}}{10^4}\cdot\left[\frac{W_{it}}{(0.02)(40)(10^6)}\right]^{-1/3}+ 2\cdot\frac{\bar{\lambda}}{10^4}\cdot\left[\frac{W_{it}}{(0.02)(40)(10^6)}\right]^{1/3}\frac{Q}{(0.01)V_{it}}\Big), \tag{2.4}$$

where $\bar{\kappa}=8.21$ and $\bar{\lambda}=2.50$ are calibrated from the data and exponents $-1/3$ and $1/3$ are implications of invariance. The first model is the square root price impact model. According to its log-linear version, buying or selling Q shares of a stock with a current stock price P moves the price on average by $\Delta P(Q)$ such that

$$\ln\left(1+\frac{\Delta P(Q)}{P}\right)=\frac{\sigma_{it}}{0.02}\Big(\frac{\bar{\kappa}}{10^4}\cdot\left[\frac{W_{it}}{(0.02)(40)(10^6)}\right]^{-1/3}+2\cdot\frac{\bar{\lambda}}{10^4}\cdot\big[\frac{Q}{(0.01)V_{it}}\big]^{1/2}\Big), \tag{2.5}$$

where $\bar{\kappa}=2.08$ and $\bar{\lambda}=12.08$ are calibrated from the data and exponents $-1/3$ and $1/2$ are implications of invariance.

Equations (2.2) and (2.3) describe the order-flow process for traders. Equations (2.4) and (2.5) describe how intermediaries update prices in response to each bet. Combining price impact of all bets executed during the day, one can calculate implied daily price changes. The set of these equations thus describe a basic structural model for daily returns, as implied by invariance.

2.3.2 Price changes upon execution of one bet

We next examine moments of price changes induced by one bet. Since buys and sells arrive with equal probabilities, the distribution of signed bet sizes $\tilde{Q}_{it}$ is symmetric, and all of its odd moments are equal to zero. For example, $E\left[\tilde{Q}_{it}\right]=0$ and $E\left[\tilde{Q}_{it}^3\right]=0$.

Since the distribution (2.3) of unsigned bet size $|\tilde{Q}_{it}|=\exp(\mu_Q+\sigma_Q\cdot\tilde{Z})$ is a log-normal with a log-mean of μ_Q and a log-variance of $\sigma_Q^2 = 2.53$, its moments can be calculated as,

$$E\left[|\tilde{Q}_{it}|^p\right] = \int q^p \cdot \frac{1}{q} \cdot \frac{1}{\sqrt{2\pi\sigma_Q^2}} \exp\left(-\frac{(\ln(q) - \mu_Q)^2}{2\sigma_Q^2}\right) dq = e^{p^2\sigma_Q^2/2 + p\mu_Q}. \tag{2.6}$$

This implies the kurtosis of price changes upon execution of a bet. For the linear price impact model, it is equal to kurtosis of a bet size itself,

$$\text{kurt}\left[\Delta P(\tilde{Q}_{it})\right] = \text{kurt}\left[|\tilde{Q}_{it}|\right] = \frac{\mathbb{E}\left[|\tilde{Q}_{it}|^4\right]}{\mathbb{E}\left[|\tilde{Q}_{it}|^2\right]^2} = e^{4\sigma_Q^2} = 22,000. \tag{2.7}$$

For the square root price impact model, it is equal to

$$\text{kurt}\left[\Delta P(\tilde{Q}_{it})\right] = \text{kurt}\left[|\tilde{Q}_{it}|^{1/2}\right] = \frac{\mathbb{E}\left[|\tilde{Q}_{it}|^2\right]}{\mathbb{E}\left[|\tilde{Q}_{it}|\right]^2} = e^{\sigma_Q^2} = 12. \tag{2.8}$$

These values are much larger than kurtosis of 3 for a normal distribution, especially for the linear model.

2.3.3 Price changes upon execution of bet sequences with no bet shredding

We next examine moments of price changes induced by one bet. Since buys and sells arrive with equal probabilities, the distribution of signed bet sizes $\tilde{Q}_{it}$ is symmetric, and all of its odd moments are equal to zero. For example, $E\left[\tilde{Q}_{it}\right] = 0$ and $E\left[\tilde{Q}_{it}^3\right] = 0$.

Since the distribution (2.3) of unsigned bet size $|\tilde{Q}_{it}| = \exp(\mu_Q + \sigma_Q \cdot \tilde{Z})$ is a log-normal with a log-mean of μ_Q and a log-variance of $\sigma_Q^2 = 2.53$, its moments can be calculated as,

$$E\left[|\tilde{Q}_{it}|^p\right] = \int q^p \cdot \frac{1}{q} \cdot \frac{1}{\sqrt{2\pi\sigma_Q^2}} \exp\left(-\frac{(\ln(q) - \mu_Q)^2}{2\sigma_Q^2}\right) dq = e^{p^2\sigma_Q^2/2 + p\mu_Q}. \tag{2.9}$$

This implies the kurtosis of price changes upon execution of a bet. For the linear price impact model, it is equal to kurtosis of a bet size itself,

$$\text{kurt}\left[\Delta P(\tilde{Q}_{it})\right] = \text{kurt}\left[|\tilde{Q}_{it}|\right] = \frac{\mathbb{E}\left[|\tilde{Q}_{it}|^4\right]}{\mathbb{E}\left[|\tilde{Q}_{it}|^2\right]^2} = e^{4\sigma_Q^2} = 22,000. \tag{2.10}$$

For the square root price impact model, it is equal to

$$\text{kurt}\left[\Delta P(\tilde{Q}_{it})\right] = \text{kurt}\left[|\tilde{Q}_{it}|^{1/2}\right] = \frac{\mathbb{E}\left[|\tilde{Q}_{it}|^2\right]}{\mathbb{E}\left[|\tilde{Q}_{it}|\right]^2} = e^{\sigma_Q^2} = 12. \tag{2.11}$$

These values are much larger than kurtosis of 3 for a normal distribution, especially for the linear model.

2.3.4 Price changes upon execution of bet sequences with no bet shredding

Daily price change ΔP is equal to the sum of all price changes in response to execution of independent and identically distributed bets. If there are γ bets executed in day t and stock i, then kurtosis of daily returns is

$$\text{kurt}\left[\Delta P|\tilde{\gamma}_{it} = \gamma\right] = \text{kurt}\left[\sum_{k=1}^{\gamma} \Delta P(Q_{kit})\right] = \frac{\text{kurt}\left[\Delta P(\tilde{Q}_{it})\right]}{\gamma}, \tag{2.12}$$

where $\text{kurt}\left[\Delta P(\tilde{Q}_{it})\right]$ is defined in equations (2.10) and (2.11).

To find unconditional kurtosis, we should integrate out γ in equation (2.12), because the number of bets executed per day is a random variable. If no bet arrives, then we should not update our estimates of kurtosis. The kurtosis of the random sum of random variables with expected Poisson arrival rate $\bar{\gamma}_{it}$ is given by

$$\text{kurt}\left[\Delta P\right] = E_{\gamma}\left(\text{kurt}\left[\Delta P|\tilde{\gamma}_{it} = \gamma\right]\right) = \sum_{j=1}^{+\infty}\left[\frac{\text{kurt}\left[\Delta P(\tilde{Q}_{it})\right]}{j} \cdot \frac{\bar{\gamma}_{it}^j}{j!} \cdot e^{-\bar{\gamma}_{it}}\right]. \tag{2.13}$$

The infinite sum $\sum_{j=1}^{+\infty}\left[\gamma^j/j!\,\right]$ is a converging series, a series $\{1/j\}$ is a bounded from above, monotone sequence, and $\text{kurt}\left[\Delta P(\tilde{Q}_{it})\right]$ is a constant. Applying Abel's convergence test, we find that the infinite sum (2.13) converges, though it does not have a close form solution.

It is possible to derive the lower bound for the unconditional kurtosis using Jensen's inequality. Indeed,

$$\text{kurt}\left[\Delta P\right] = E_{\gamma}\left(\frac{\text{kurt}\left[\Delta P(\tilde{Q}_{it})\right]}{\tilde{\gamma}_{it}}\right) \geq \frac{\text{kurt}\left[\Delta P(\tilde{Q}_{it})\right]}{E_{\gamma}\left(\tilde{\gamma}_{it}\right)} = \frac{\text{kurt}\left[\Delta P(\tilde{Q}_{it})\right]}{\bar{\gamma}_{it}}. \qquad (2.14)$$

The lower bound is equal to the kurtosis of daily returns (2.12) conditional of the assumption that the number of bets $\tilde{\gamma}_{it}$ coincides with the average arrival rate $\bar{\gamma}_{it}$.

Our simulation analysis shows that this lower bound provides a good approximation for the daily kurtosis of most stocks, as implied by a structural model. The differences created by uncertainty in the Poisson arrival rates and non-linearities of log-returns are insignificant. For a stock with median dollar volume and returns variance in each of the ten trading activity groups, we run 1000 Monte-Carlo simulations and calculate the average theoretical kurtosis with its standard errors. The simulations are done based on a structural model of price process with bet arrival rate in equation (2.2), distribution of bet sizes in equation (2.3), and price impact model (2.4). We also calculate the lower bound using equation (2.14). Table 2.2 shows that the lower bound tracks closely the average kurtosis for all groups, except the group of least actively traded stocks. The percentage differences in the series of two estimates are 29%, 3%, 2%, 1%, and 0% for groups 1, 3, 5, 8, and 10, respectively. A large difference for the first group may reflect an upward bias in theoretical estimates of kurtosis. The bet arrival rates for other groups range from 23 to 232, and this effect is less pronounced. As long as the arrival rate of bets is not too low, the lower bound is a reasonable proxy for kurtosis of daily returns. We get the following approximation,

$$\text{kurt}\left[\Delta P\right] \approx \frac{\text{kurt}\left[\Delta P(\tilde{Q}_{it})\right]}{\bar{\gamma}_{it}}. \qquad (2.15)$$

Using equations (2.10) and (2.11), the lower bound for daily kurtosis is equal to

$22,000/\bar{\gamma}_{it}$ and $12/\bar{\gamma}_{it}$ for the linear and square root models, respectively. Kurtosis of price changes per each bet is the same across stocks, but the number of bets per day is larger for more liquid stocks. Therefore daily returns of more liquid stocks have lower kurtosis. The number of bets $\bar{\gamma}_{it}$ per day increases with trading activity at a rate of 2/3 in equation (2.2). Equation (2.15) then implies that kurtosis decreases with the trading activity approximately at the same rate, i.e., with 2/3 power of the trading activity. In table 2.2, for example, the ratio of kurtosis of most inactively traded stock to kurtosis of most actively traded stock is about 77 (= 7214/95); it is similar to the ratio of their trading activities in 2/3 power equal to 59 ($= (3600/8)^{2/3}$).

Similar intuition suggests that kurtosis decrease with tenor of returns for a given security. For example, kurtosis of weekly returns is expected to be lower than kurtosis of daily returns, which in turn is expected to be lower than kurtosis of one-minute returns.

Our basic structural model implies the values of kurtosis that are too high relative to empirical estimates. The average theoretical kurtosis in table 2.2 ranges between 95 and 7,214, whereas empirical estimates in table 2.1 do not exceed 8.47, the level of average kurtoses for stocks in group 1 for decade 1960-1970.

2.3.5 Price changes upon execution of bet sequences with bet shredding

So far we have assumed each bet is executed instantaneously. In reality, traders shred orders and execute them over time in sequences of transactions to reduce transaction costs. Bet shredding smooths out spikes in price dynamics and tends to make returns kurtoses smaller. We next consider several modifications of our basic model that incorporate order shredding and arbitrage trading. These models are more realistic and more flexible in their ability to match empirical estimates.

Traders decide on "target" inventories and bets based on either their private information or inventories shocks. Then, they gradually adjust actual inventories towards their targets. Let S^*_{it} denote the cumulative target order imbalances for stock i at the end of day t, calculated as the signed sum of all *bets* placed into the market place by that time,

$$S^*_{it} = \sum_{m \leq t} Q_{imk}. \tag{2.16}$$

Suppose next that each bet Q_{imk} is shredded into a sequence of transactions $x_{imk}(s)$, where s is a day count in execution package.

Let S_{it} denote cumulative realized order imbalance for stock i at the end of day t, calculated as the signed sum of all *transactions* placed into the market by that time.

$$S_{it} = \sum_{m \leq t, s \leq t} x_{imk}(s). \tag{2.17}$$

The structural model of trading (2.16) and (2.17) is consistent with the equilibrium strategies in a continuous-time model of smooth trading of Kyle, Obizhaeva, Wang (2016). In that model, symmetric, relatively overconfident, oligopolistic informed traders calculate target inventories based on how their own estimates of the long-term dividend growth rate differ from the estimates of other traders. Since the market offers no instantaneous liquidity for block trades, each trader only partially adjusts his inventory in the direction of a target inventory; the rate of adjustment is determined by the deep parameters of the model, it is larger when private information decays faster and when there is more disagreement between traders.

The difference between the time series of S^*_{it} and S_{it} depends on the specifics of bet shredding algorithms. Bet shredding algorithms are not directly observable. We assume that each algorithm is characterized by two main decisions. For each bet, traders first choose execution horizon and then parameters of shredding method. We consider several alternative specifications.

First, traders determine an appropriate execution horizon T_{itk} for each bet Q_{itk}. For example, traders may target a fixed time horizon t, say one day,

$$\text{Method-}T(t): \qquad T_{itk} = t. \tag{2.18}$$

We refer to this algorithm as "Method-$T(t)$"; for example, "Method-$T(1)$" or "Method-$T(5)$" correspond to cases when all trades are executed their bets over one day or one week.

Traders may also target a small fraction η, say equal to 5%, of expected contemporaneous volume $T_{itk} \cdot V_{it}$ or $T_{itk} \cdot \bar{\gamma}_{it} \cdot E\left[|\tilde{Q}_{it}|\right]$,

$$|Q_{itk}| = \eta \cdot T_{itk} \cdot \bar{\gamma}_{it} \cdot E\left[|\tilde{Q}_{it}|\right]. \tag{2.19}$$

This implies the execution horizon that is linearly proportional to bet size,

$$\text{Method-}V(\eta): \qquad T_{kit} = \frac{|Q_{kit}|}{\eta \cdot \bar{\gamma}_{it} \cdot E\left[|\tilde{Q}_{it}|\right]}. \tag{2.20}$$

We refer to this algorithm as "Method-$V(\eta)$"; for example, "Method-$V(0.05)$" or "Method-$V(0.10)$" for execution algorithms targeting 5 percent and 10 percent of daily volume, respectively.

Traders may also target to induce a small fraction η, say equal to 5%, of expected returns variance $T_{kit} \cdot \sigma_{it}^2$ under the assumption that each transaction is expected to move price by $\lambda \cdot |Q_{kit}|$,

$$(\lambda \cdot |Q_{kit}|)^2 = \eta \cdot T_{kit} \cdot \gamma_{it} \cdot \lambda^2 \cdot E\left[|\tilde{Q}_{it}|^2\right]. \tag{2.21}$$

This implies that the execution horizon is proportional to the square of bet size,

$$\text{Method-}\sigma^2(\eta): \qquad T_{kit} = \frac{|Q_{kit}|^2}{\eta \cdot \gamma_{it} \cdot E\left[|\tilde{Q}_{it}|^2\right]}. \tag{2.22}$$

We refer to this algorithm as "Method-$\sigma^2(\eta)$"; for example, "Method-$\sigma^2(0.05)$" or "Method-$\sigma^2(0.10)$" for execution algorithms targeting 5 percent and 10 percent of daily volatility, respectively.

In all cases, larger bets are executed over longer period of time. In the third case (2.20) larger bets are spread over longer periods of time than in the second case (2.22) and returns distribution is expected to exhibit smaller kurtosis. For the square root impact model, targeting a given fraction of returns variance is equivalent to targeting a given fraction of volume, so we do not consider this case separately.

Next, traders have to choose an appropriate shredding method. We consider two bet shredding methods. Each bet Q_{kit} can be shredded at a uniform rate and executed in equally-sized transactions $x_{kit}(s)$,

$$x_{itk}(s) = \frac{|Q_{itk}|}{T_{itk}}, \qquad s = 1, ..T_{itk}. \tag{2.23}$$

Bertsimas and Lo (2001) find that this simple execution is optimal when a risk-neutral trader needs to execute an order.

Alternatively, each bet Q_{itk} can be shredded at a monotonically decreasing rate,

where sinh and cosh are the hyperbolic sine and cosine functions. Each day a trader executes some fraction of the remaining part of the bet, determined by parameter ρ. This parameter is related to the speed of information decay, risk aversion, and riskiness of securities. The larger is parameter ρ, the faster the bet is executed. Almgren and Chriss (2000) finds that execution is optimal when a risk-averse trader executes a bet. Similar solution can be also found in Grinold and Kahn (1999).

We choose to focus on simple execution strategies. In reality, execution strategies are more complicated. Execution algorithms are often price dependent, as discussed in Obizhaeva (2012). Other order shredding algorithms are for example discussed in Gatheral and Schied (2013), Schied and Schoeneborn (2009), and Obizhaeva and Wang (2013). If necessary, sophisticated execution strategies may be built into our structural model as well.

The structural model for bet arrival (2.2) and (2.3) augmented with specific order shredding algorithm represent the structural model describing the order-flow process. Together with price impact model, they allow to construct implied time-series of prices. In what follows, we consider linear price impact rule.

2.3.6 Price dynamics with shredding and arbitrageurs

Bet shredding introduces positive autocorrelation in stock return process and makes future price changes predictable. For example, execution of a large buy bet is expected to inject a positive trend into the price dynamics, while execution of a large sell bet induces a downward price dynamics. Arbitrageurs notice that prices are not martingales and construct order anticipation algorithms to detect execution of orders.

We next describe how to model trading by arbitrageurs. If intermediaries observed target bet imbalances, they would set prices according to their market clearing rule,

$$P^*_{it} = \lambda_{it} \cdot S^*_{it}, \tag{2.24}$$

and price changes would be unpredictable. In reality, intermediaries may at best identify only actual signed order imbalances S_{it} and set prices as,

$$\hat{P}_{it} = \lambda_{it} \cdot S_{it}. \tag{2.25}$$

To the extent that unexecuted order imbalance $S^*_{it} - S_{it}$ are predictable based on past information, these price process is not a martingale.

Arbitrageurs build a model to forecast $S^*_{it} - S_{it}$ and trade $\mathbb{E}_t\{S^*_{it} - S_{it}\}$ at day t. When target order imbalances are higher than actual order imbalances, arbitrageurs buy ahead of other traders. When target inventories are lower than actual inventories, arbitrageurs sell ahead of other traders. Market makers set clearing prices based on the aggregate order flow of both traders and arbitrageurs,

$$P_{it} = \mathbb{E}_t\{P^*_{it}\} = \lambda_{it} \cdot S_{it} + \lambda_{it} \cdot \mathbb{E}_t\{S^*_{it} - S_{it}\} = \lambda_{it} \cdot \mathbb{E}_t\{S^*_{it}\}. \tag{2.26}$$

Trading by arbitrageurs restore martingale properties of stock prices and makes price process $P_{it} = \mathbb{E}\{P^*_{it}\}$ a martingale based on arbitrageurs' filtration. Essentially, the price is set based on the market's forecasts of current target imbalances.

Our structural model is flexible to be consistent with various predictive models of arbitrageurs. We suppose that arbitrageurs know daily volatility and daily volume of an asset. They are also familiar with all invariance formulas and bet shredding algorithms that traders use. Arbitrageurs thus can simulate hypothetical bet arrival process and how bets are shredded into sequences of transactions. Then, they can perform a large estimation on the simulated sample to build a model for forecasting unexecuted order imbalances.

This procedure can be summarized as follows,

1. Simulate N paths of bet histories for an asset with volume V_{it} and volatility σ_{it} based on formulas (2.2) and (2.3);

2. Using the conjectured parameters of bet shredding algorithm, aggregate bets and transactions, calculating histories of target imbalances and actual imbalances, $S^*_{it,n}$ and $S_{it,n}$ for each of simulated paths $n = 1, ..N$;

3. Run a rolling-window predictive regression for unexecuted imbalances with k lags of linear and quadratic terms of realized past imbalances,

$$\mathbb{E}_t\{S^*_{it,n} - S_{it,n}\} = \alpha + \sum_{j=1}^{k} \beta_{1j} \cdot S_{i,t-j,n} + \sum_{j=1}^{k} \beta_{2j} \cdot S^2_{i,t-j,n} + \epsilon_{tn}, \qquad t = 1, ..T, n = 1, ..N, \tag{2.27}$$

to estimate coefficients $\hat{\beta}_{1j}$ and $\hat{\beta}_{2j}$, $j = 1, ..k$. For example, we use $k = 5$ as our benchmark model, i.e. an arbitrageur using information on actual inventories over the previous week.

Equipped with estimated model $\hat{\beta}_{1j}$ and $\hat{\beta}_{2j}$, $j = 1, ..k$, arbitrageurs construct forecasts based on current information about past order imbalances $S_{i,t-j}$, $j = 1, ..k$, as

$$\mathbb{E}_t\{S^*_{it} - S_{it}\} = \alpha + \sum_{j=1}^{k} \hat{\beta}_{1j} \cdot S_{i,t-j} + \hat{\beta}_{2j} \cdot S^2_{i,t-j}. \tag{2.28}$$

This is a model for forecasting an unexecuted order imbalances.

In what follows, we mostly apply bet shredding method that targets a given fraction of daily volume, split all bets into equally-sized transactions, and assume a linear price impact function.

2.3.7 Properties of simulated returns

We first illustrate our structural model using the example of a hypothetical benchmark stock with price P of \$40 per share, daily volume V of one million shares, and daily volatility σ of 2% per day. This benchmark stock would belong to the bottom tercile of S&P 500.

We simulate 1,000 paths of 90-day bet arrival histories for the benchmark stock using formulas (2.2) and (2.3). We then apply several bet shredding algorithms by first selecting the execution horizon depending on the fraction of daily volume targeted and second by shredding each bet into a sequence of equally-sized transactions. The execution of some packages extends beyond the boundary of a 90-day sample. We cut tails of these unfinished packages at the end of each sample path, multiply remaining sequences by 1 or -1 with equal probabilities to model buy and sell orders, and insert them into the beginning of the same sample path. This mimics a typical situations when some of large bets arrived in the past are continuing to get executed at the beginning of selected sample paths.

We then estimate forecasting model (2.27) of arbitrageurs, who seek to predict unexecuted bet imbalances at each point of time using the last five realized bet imbalances and their squares. This estimation is done on the entire simulated sample on a rolling-window basis.

Table 2.3 reports the results for the three bet shredding algorithms with $\eta =$ 1%, $\eta = 5\%$, and $\eta = 10\%$. The lower is fraction η of volume targeted in the execution, the more execution is extended over time, the more past imbalances are autocorrelated with current unexecuted imbalances, and the larger are estimated coefficients. For example, when $\eta = 1\%$, the coefficients are 1.98, 0.97, 1.02, 1.25, and 5.32. When $\eta = 10\%$, the coefficients are only 0.17, 0.16, 0.19, 0.30, and 0.89. Using these estimates, we construct predictive model (2.28) and price paths using equation 2.4.

Figure 2.7 shows the averages, medians, and standard error bounds for returns autocorrelation coefficients at different lags, ranging from one day to forty days for the simulated sample under the assumption that there are no arbitrageurs. The four panels show the results for the cases of $\eta = 1\%$, $\eta = 5\%$, $\eta = 10\%$, and the case with no shredding, i.e. $\eta = \infty$. As expected, when there is no shredding, autocorrelations are equal to zero at all lags. In the other panels, autocorrelations are high at fist lags, decaying with time. The lower is the fraction η of bet shredding algorithm and the longer are execution horizons of large bets, the bigger autocorrelations at first lags are and the slower they decay.

Figure 2.8 show the same statistics but under the assumption that there are arbitrageurs. Most of the autocorrelation coefficients are now close to zero, since arbitrageurs eliminate most of returns predictability. Based on the linear terms and squared terms, their forecasting model works reasonably well, except for reducing autocorrelations at the boundaries of their forecasting window, which is assumed to have the length of five days in our example.

Table 2.4 presents the autocorrelations and their standard errors. As before, in panel A when the model has no arbitrageurs, many of the coefficients are statistically bigger than zeros, especially when η is small. In panel B when we introduce arbitrageurs, most coefficients become insignificant. For example, when $\eta = 1\%$, the first-order autocorrelation is equal to 0.696 with standard errors of 0.092 with no arbitrageurs and 0.033 with standard errors of 0.119 with arbitrageurs.

Figures 2.9 and 2.10 present distributions of the four moments of simulated returns for the cases without and with arbitrageurs, respectively. There are distributions of the four moments in the four columns. Each of the four rows corresponds to different bet-shredding methods with $\eta = 1\%$, $\eta = 5\%$, $\eta = 10\%$ as well as the case with no shredding. Table 2.5 reports the summary statistics for these distributions. On both figures, the means and the skewness are centered around zero, since the base model is symmetric for buy and sell orders. The volatility is much lower than initially assumed daily volatility of $\sigma = 2\%$ when there are no arbitrageurs, especially when η is low. Intuitively, bet-shredding converts returns volatility into the price drift. Trading by arbitrageurs "restores" martingale properties of prices and brings volatility back to the assumed levels. For example, when $\eta = 1\%$, the daily volatility of simulated returns is equal to 0.005 with no arbitrageurs and 0.022 with arbitrageurs.

2.4 Properties of implied shredding parameter

The properties of daily returns depend on the assumptions about parameters of the bet-shredding algorithm. We next use the method of simulated moments and calibrate these parameters to match empirical moments of daily returns.

As before, we assume that traders generate bets according to invariance, design execution to target a given fraction η of expected daily volume, and split bets into equally-sized transactions. Meanwhile, arbitrageurs apply the forecasting model described in section 2.3.6 and market makers clear the market. We generate $N =$ *** paths of daily returns. The bet-shredding parameter η is then estimated by matching the kurtoses of simulated returns $\mathrm{kurt}(\Delta P|\eta, n)$ to the empirical estimates of kurtoses $\mathrm{kurt}(\Delta P|\mathrm{Data})$,

$$\eta^* = \mathrm{argmin}_\eta \Big(\frac{\sum_{n=1}^{N} \mathrm{kurt}(\Delta P|\eta, n)}{N} - \mathrm{kurt}(\Delta P|\mathrm{Data}) \Big). \tag{2.29}$$

The empirical estimates are taken from table 2.1 for different trading activity groups and time periods.

Table 2.6 reports the estimates of implied parameter η for median stocks in the five out of ten trading activity groups and for the seven decades from 1950 to 2017. The table also presents information about trading activity used for simulation of daily returns; its values coincide with statistics reported in table 2.1. There are two patterns.

First, the implied parameter η decreases over time. For the stocks in group 1, parameter η decreased from 8.875 during 1950–1960 to 5.04 during 2010–2017.

For the stocks in group 10, parameter η decreased from 3.225 during 1950–1960 to 1.39 during 2010–2017. This implies that, conditional of bet size, bet shredding increased over time. Similarly, **?** document a significant change in trading patterns in the Trades and Quotes (TAQ) dataset, as the decimalization and use of electronic interfaces has recently led to a significant increase in order shredding; the market for block trades seems almost to have disappeared, and most trading is now dominated by transactions of 100 shares, the minimum lot size. The feature of increased bet shredding implied by our structural model suggests that it has reasonable properties.

Second, the implied parameter η decreases with trading activity W. For example, for the time period 1990 through 2000, η is equal to 8.36, 4.22, 3.46, 2.52, and 1.66 for groups 1, 3, 5, 8, and 10, respectively. For the time period 2010 through 2017, η is equal to 5.04, 2.69, 2.32, 1.80, and 1.39, respectively. This implies that, conditional on bet size, execution of bets is spread over longer periods for more actively traded stocks.

Table 2.7 shows implied execution horizons for the two periods before and after decimalization. Panel A shows results for 1990-2000. Panel B shows results for 2010-2017. We calculate bet sizes using equation (2.3) and then calculate implied execution horizons using equation (2.20) and calibrated bet-shredding parameters $\hat{\eta}$ from table 2.6. For the median stock in group 1, it takes 2.69, 13.22, 64.86, and 318.23 minutes to execute 4-std, 5-std, 6-std, and 7-std bets during 1990–2000, respectively, and 0.79, 3.86, 18.93, and 92.87 minutes for similar bets during 2010–2017. For the median stock in group 10, it takes 0.17, 0.85, 4.16, and 20.39 minutes to execute 4-std, 5-std, 6-std, and 7-std bets during 1990–2000, respectively, and 0.05, 0.26, 1.29, and 6.32 minutes for similar bets during 2010–2017. The speed of execution increased by a factor of 3.

The inspection of estimates in table 2.7 reveals that differences in bet-shredding parameters are similar to differences in trading activity in $-1/3$ power. For example, the ratio in parameters η_i and η_j for stock i and j are related to the ratio of their trading activities W_i and W_j as approximately,

$$\frac{\eta_i}{\eta_j} \approx \Big(\frac{W_i}{W_j}\Big)^{-1/3}. \tag{2.30}$$

We can also extrapolate these estimates to the overall market with daily trading volume of $292 billion (futures and stocks combined) and daily volatility of 2 percent, as noted in Kyle and Obizhaeva (2017*b*). Using parameters for the median stock in group 10 during 2010–2017 as the benchmark, equation (2.30) implies that bet-shredding parameter for the entire U.S. market $\eta_{\text{mkt}} \approx 1.39 \cdot (\frac{W_i}{7{,}141{,}896})^{-1/3} \approx 0.20$, i.e. traders are targeting about 20 percent of expected contemporaneous volume when executing bets in the U.S. market. This is broadly consistent with information in Staffs of the CFTC and SEC (2010b) that the large trader whose trading caused the Flash crash on May 6, 2010, has been targeting 9 percents of contemporaneous volume when executing a bet in the E-mini S&P500 futures market.

2.5 Conclusions

We propose a new structural model of stock returns dynamics, which is inspired by the recently developed ideas of market microstructure invariance. Traders generate investment ideas, or bets, and execute them by shredding large orders over time to minimize transaction costs, arbitrageurs trade to profit on any detectable trends in prices, and market makers clear the market. Bets are assumed to arrive according to the processes calibrated by Kyle and Obizhaeva (2016); parameters of bet-shredding algorithms are chosen to match empirical moments of stock returns.

Our structural model captures realistically the economics of trading. It is the model of stochastic volatility, because arrival of bets and their sizes are stochastic, and large bets lead to bursts in volume, volatility, and intermediation. The model is flexible in terms of modelling trading behavior of arbitrageurs and bet-shredding algorithms, while precise and grounded in theory in terms of using a specific structure of bet flow from traders and intermediaries. It can be calibrated either to fit the data or to infer the implied parameters of trading, for example, such as hard-to-observe bet-shredding parameters.

We focus mostly on the price dynamics, but the framework also generates quantitative predictions about overall trading volume and order flow generated by different groups of traders. As an extension, it is possible to calibrate the model to match cross-sectional and time-series properties of both stock returns and trading volume, or even some empirical findings about trading by different groups of traders, for example, such as defined in Kirilenko et al. (2010).

2.6 Tables and figures

Figure 2.1: Idiosyncratic kurtosis of daily stock returns for 1950 through 2016.

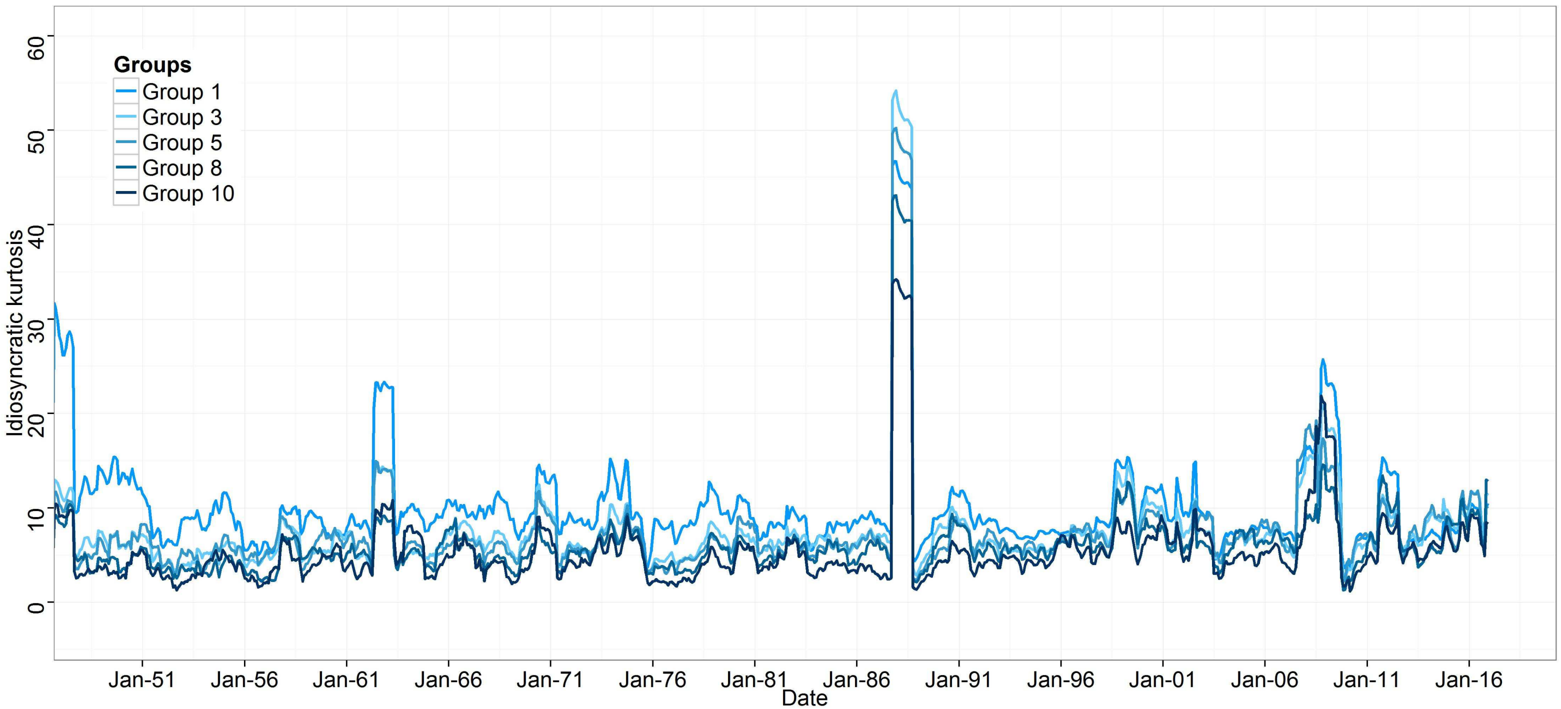


Figure shows five monthly time series of 12-month moving averages of median sample kurtosis for idiosyncratic daily stock returns for each of the five trading activity groups (groups 1, 3, 5, 8, and 10 out of ten groups). Group 1 (10) contains the least (most) actively traded stocks. The period ranges from January 1950 to December 2016.

Figure 2.2: Ratio of idiosyncratic kurtosis (Group 1 to Group 10) for 1950 through 2016.

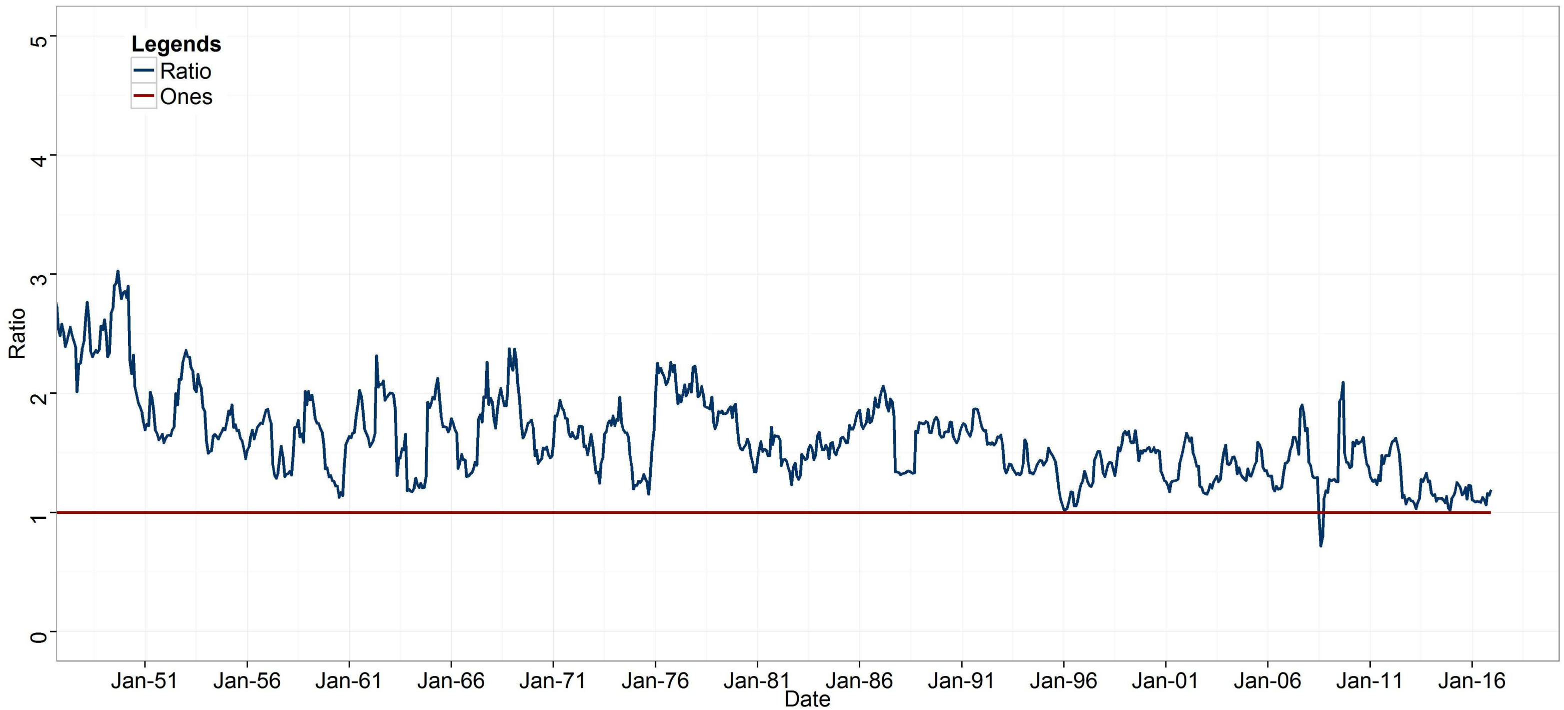


Figure shows the time series of ratio of median sample kurtosis of idiosyncratic daily stock returns for stocks in Group 1 (least actively traded stocks) to the one of Group 10 (most actively traded stocks). The horizontal line marks the value of one. The sample ranges from January 1950 to December 2016.

Figure 2.3: Kurtosis of daily stock returns for 1950 through 2016.

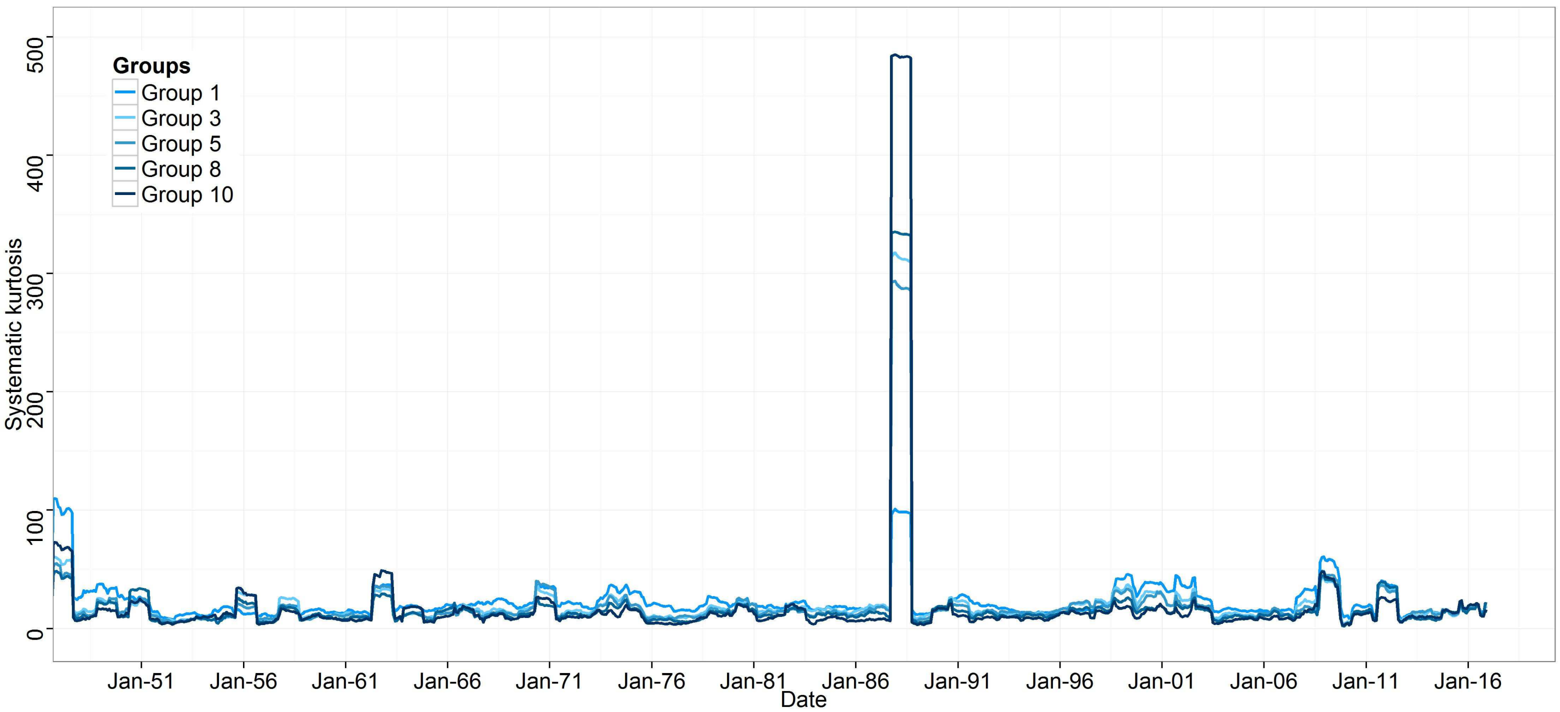


Figure shows five monthly time series of 12-month moving average of median sample kurtosis of daily stock returns for each of the five trading activity groups (groups 1, 3, 5, 8, and 10 out of ten groups). Group 1 (10) contains the least (most) actively traded stocks. The period ranges from January 1950 to December 2016.

Figure 2.4: Ratio of kurtosis (Group 1 to Group 10) for 1950 through 2016.

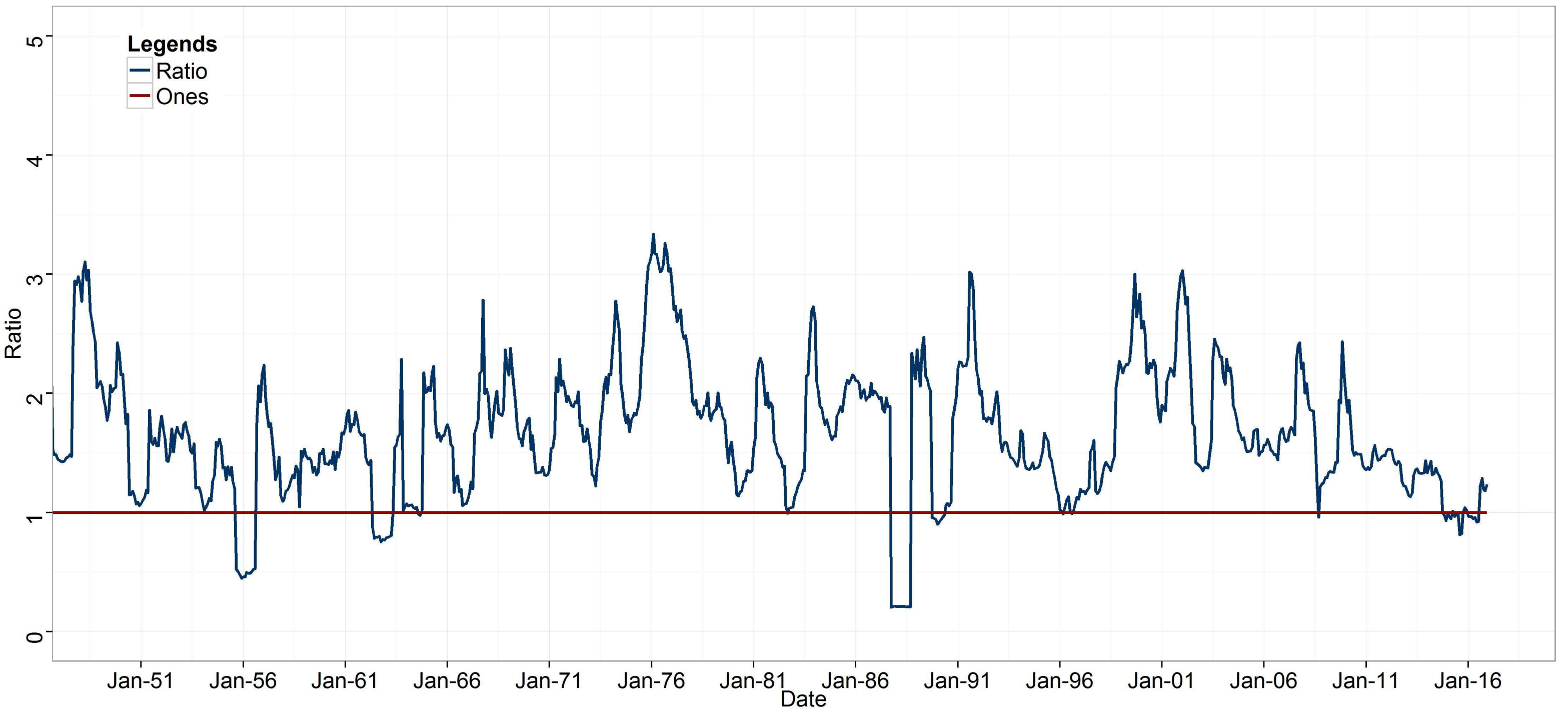


Figure shows the time series of ratio of median sample kurtosis of daily stock returns for stocks in Group 1 (least actively traded stocks) to the one of Group 10 (most actively traded stocks). The horizontal line marks the value of one. The sample ranges from January 1950 to December 2016.

Figure 2.5: Idiosyncratic skewness of daily stock returns for 1950 through 2016.

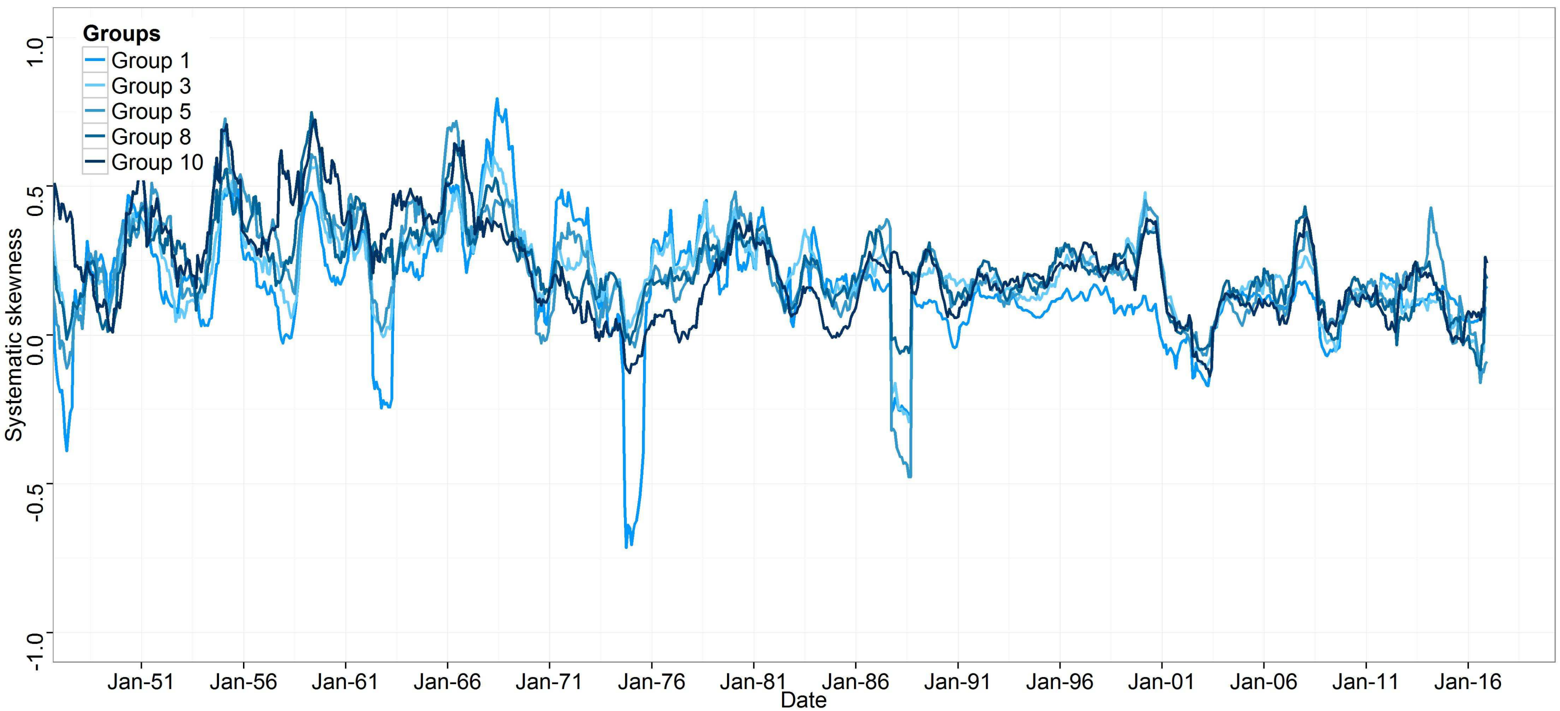


Figure shows five monthly time series of 12-month moving average of median sample skewness of idiosyncratic daily stock returns for each of the five trading activity groups (groups 1, 3, 5, 8, and 10 out of ten groups). Group 1 (10) contains the least (most) actively traded stocks. The period ranges from January 1950 to December 2016.

Figure 2.6: Idiosyncratic volatility of daily stock returns for 1950 through 2016.

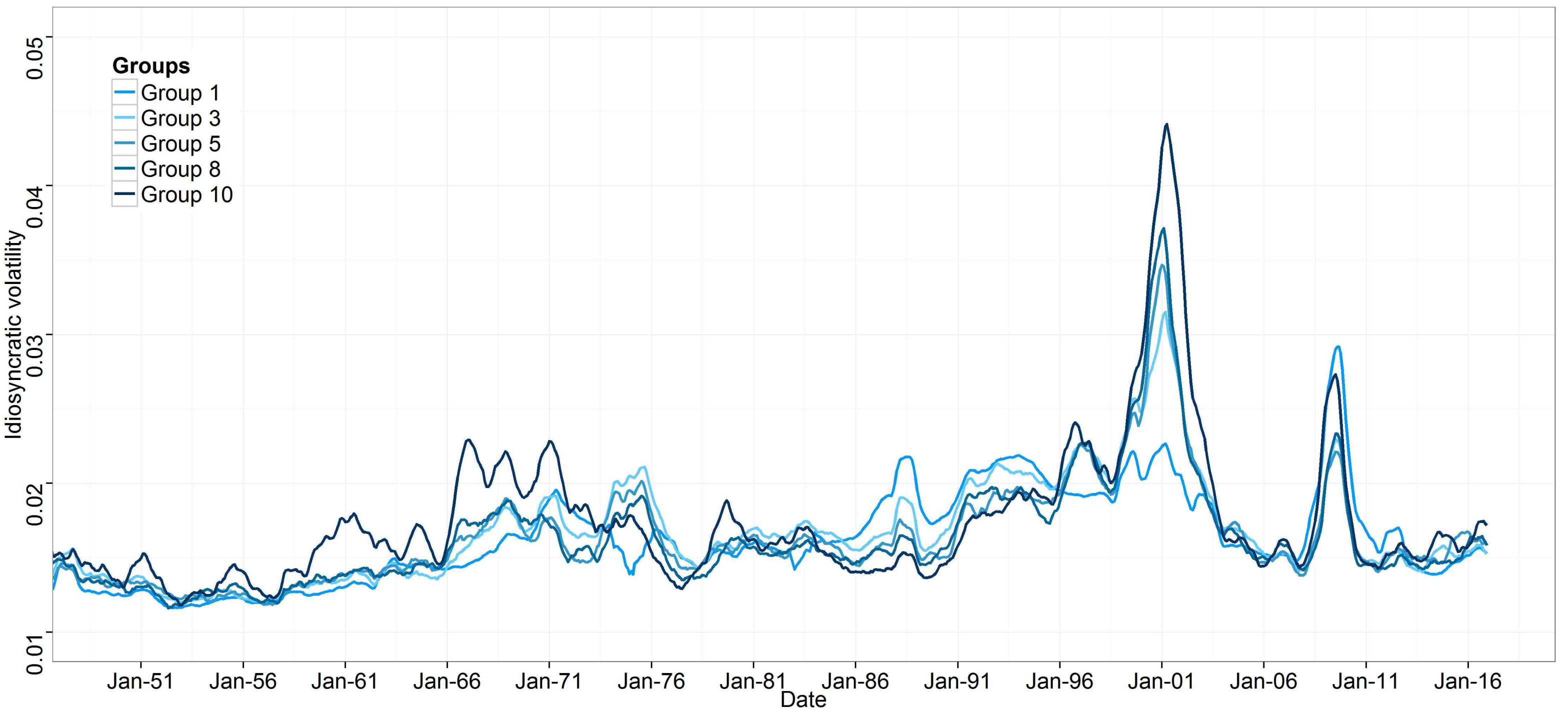


Figure shows five monthly time series of 12-month moving average of median sample volatility of idiosyncratic daily stock returns for each of the five trading activity groups (groups 1, 3, 5, 8, and 10 out of ten groups). Group 1 (10) contains the least (most) actively traded stocks. The period ranges from January 1950 to December 2016.

Table 2.1: Kurtosis and skewness for trading activity groups across decades

	Group 1	Group 3	Group 5	Group 8	Group 10	Total
			Decade 1950-1960			
Activity	101	517	967	1,926	5,925	440
Kurtosis	7.174	4.253	4.387	3.205	2.656	4.825
Skewness	0.199	0.243	0.305	0.317	0.357	0.262
# Stocks	676	766	623	511	328	
			Decade 1960-1970			
Activity	277	1,650	3,375	6,999	26,727	1,197
Kurtosis	8.472	5.419	5.188	4.141	2.972	6.002
Skewness	0.303	0.306	0.303	0.302	0.338	0.310
# Stocks	2,126	1,839	1,378	1,056	557	
			Decade 1970-1980			
Activity	210	2,249	5,103	11,519	39,913	1,284
Kurtosis	6.605	5.511	4.928	4.327	3.289	5.627
Skewness	0.182	0.210	0.178	0.159	0.076	0.179
# Stocks	3,697	1,894	1,435	1,028	583	
			Decade 1980-1990			
Activity	1,176	12,893	29,588	70,470	271,798	4,935
Kurtosis	7.133	5.734	4.939	4.321	2.851	5.938
Skewness	0.144	0.201	0.192	0.205	0.157	0.173
# Stocks	5,642	3,084	1,852	1,236	677	
			Decade 1990-2000			
Activity	2,895	38,104	88,975	245,021	1,232,159	14,012
Kurtosis	8.102	6.450	6.174	5.226	4.220	6.884
Skewness	0.098	0.184	0.202	0.197	0.192	0.143
# Stocks	7,721	5,082	3,265	2,237	1,089	
			Decade 2000-2010			
Activity	8,640	190,720	468,578	1,364,880	6,704,911	62,135
Kurtosis	7.583	5.963	5.491	4.699	4.002	6.542
Skewness	0.059	0.098	0.087	0.106	0.096	0.076
# Stocks	5,894	3,545	2,398	1,736	820	
			Decade 2010-2017			
Activity	18,363	411,722	940,703	2,216,188	7,141,896	83,929
Kurtosis	6.808	6.315	6.085	4.918	4.232	6.396
Skewness	0.107	0.074	0.075	0.077	0.098	0.096
# Stocks	3,364	1,659	1,018	766	441	

Table presents the sample medians of trading activity, idiosyncratic skewness, and idiosyncratic kurtosis as well as the number of stocks for the ten groups of U.S. stocks, based on their trading activity. The sample ranges from January 1950 to December 2016 and split into decades. Group 1 (10) consists of stocks with lowest (highest) trading activity in the previous three months.

Table 2.2: Simulated theoretical kurtosis and low bounds.

	Group 1	Group 3	Group 5	Group 8	Group 10
Trading Activity	8,000	210,000	460,000	1,000,000	3,600,000
Number of Bets	4	35	59	99	232
Avg Kurtosis	7,214	651	381	225	95
Stand. Error	(4.81)	(0.12)	(0.06)	(0.02)	(0.01)
Low Bound	5,576	631	374	259	95
%Δ	29%	3%	2%	1%	0%

Table reports trading activity $\sigma \cdot V \cdot P$, bet arrival rate per day γ, the average daily returns kurtosis and its standard errors of the means from Monte-carlo simulations, low bound for kurtosis, and percentage difference between the average kurtosis and the low bound for the median stock in each of the five trading activity groups (groups 1, 3, 5, 8, and 10 out of ten groups). Group 1 (10) contains the least (most) actively traded stocks.

Table 2.3: Imbalance forecasting model of arbitrageurs.

Shredding	const	$S_{i,t-1}$	$S^2_{i,t-1}$	$S_{i,t-2}$	$S^2_{i,t-2}$	$S_{i,t-3}$	$S^2_{i,t-3}$	$S_{i,t-4}$	$S^2_{i,t-4}$	$S_{i,t-5}$	$S^2_{i,t-5}$	R^2
$\eta = 1\%$	39,719	1.98	0.00	0.97	0.00	1.02	0.00	1.25	0.00	5.32	0.00	12%
$\eta = 5\%$	32,925	0.37	0.00	0.29	0.00	0.34	0.00	0.48	0.00	1.49	0.00	13%
$\eta = 10\%$	12,190	0.17	-0.00	0.16	-0.00	0.19	-0.00	0.30	0.00	0.89	-0.00	13%

Table reports estimates $\hat{\beta}_{1j}$ and $\hat{\beta}_{2j}$, $j = 1, .5$ of arbitrageurs' model for forecasting unexecuted imbalances

$$S^*_{it,n} - S_{it,n} = \alpha + \sum_{j=1}^{5} \beta_{1j} \cdot S_{i,t-j,n} + \sum_{j=1}^{5} \beta_{2j} \cdot S^2_{i,t-j,n} + \epsilon_{tn}, \qquad t = 1, ..T, n = 1, ..N,$$

estimated based on the simulated sample for a benchmark stock with daily volatility 2 percent, price \$40, and daily volume 1 million shares. The simulated sample consist of 1,000 of 90-day paths. The three bet-shredding algorithms are used: "Method-$V(1\%)$", "Method-$V(5\%)$", and "Method-$V(10\%)$".

Figure 2.7: Returns autocorrelations without arbitrager.

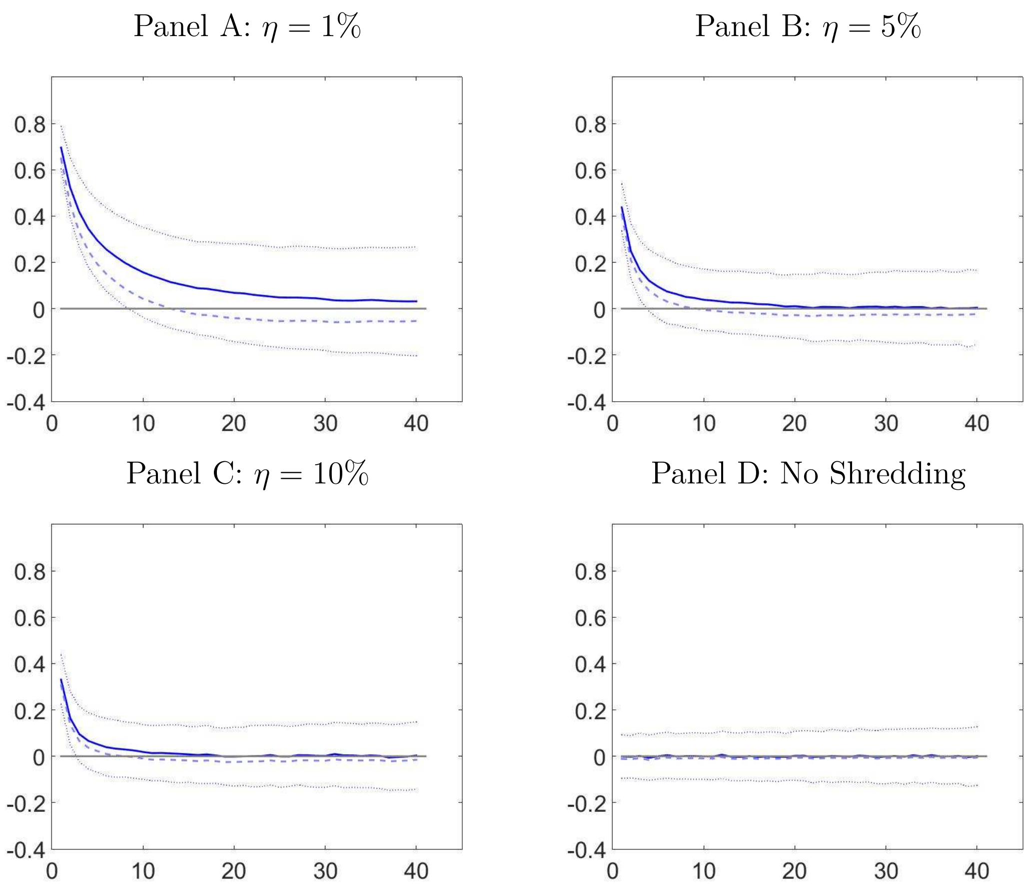


Figure shows autocorrelation coefficients of daily returns at different lags for different models of bet shredding without arbitragers: "Method-$V(1\%)$", "Method-$V(5\%)$", "Method-$V(10\%)$", and no bet shredding. The simulation consists of 1,000 of 90-day paths. There are averages, medians, and standard errors of autocorrelation coefficients in dark solid, dashed, and light solid lines, respectively.

Figure 2.8: Returns autocorrelations with arbitrager.

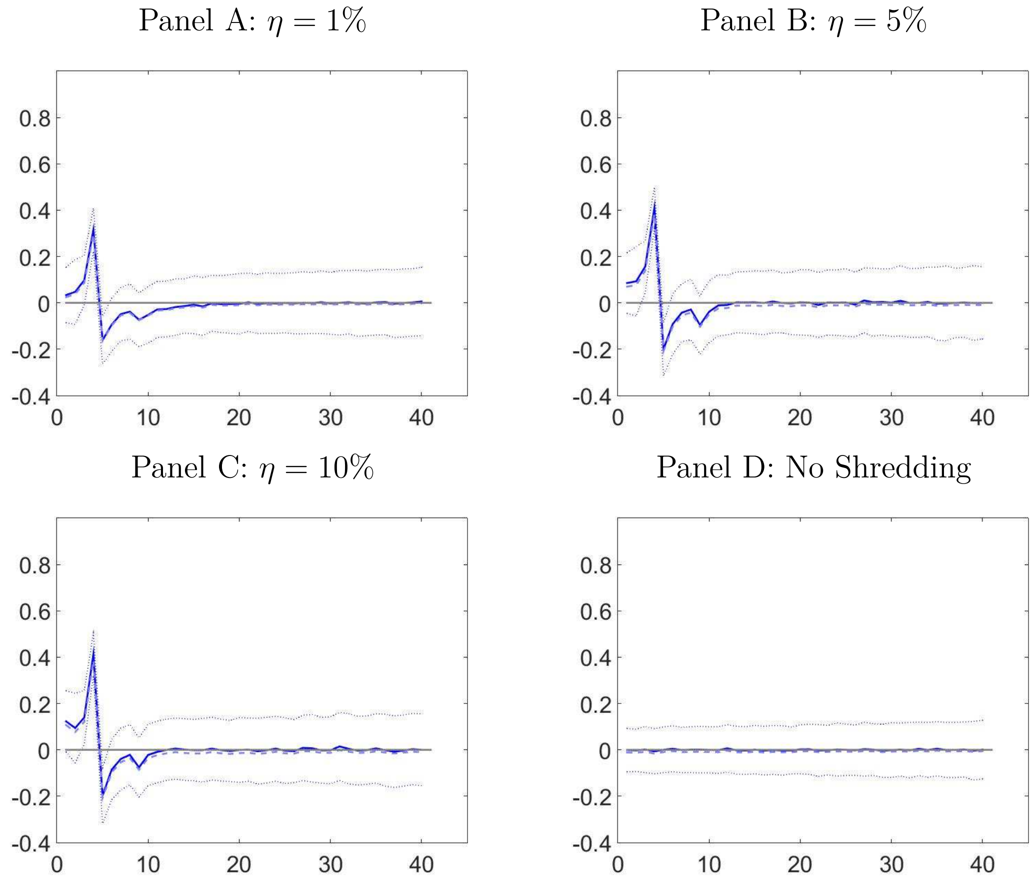


Figure shows average autocorrelation coefficients of daily returns at different lags for different models of bet shredding with arbitragers: "Method-$V(1\%)$", "Method-$V(5\%)$", "Method-$V(10\%)$", and no bet shredding. The simulation consists of 1,000 of 90-day paths. There are averages, medians, and standard errors of autocorrelation coefficients in dark solid, dashed, and light solid lines, respectively.

Table 2.4: Returns autocorrelations.

	Order of autocorrelation					
	lag 1	lag 2	lag 3	lag 5	lag 10	lag 20
Panel A: Model without arbitragers						
$\eta = 1\%$	0.696	0.523	0.417	0.294	0.157	0.068
	(0.092)	(0.128)	(0.152)	(0.172)	(0.194)	(0.211)
$\eta = 5\%$	0.437	0.249	0.167	0.094	0.038	0.009
	(0.101)	(0.119)	(0.127)	(0.137)	(0.133)	(0.138)
$\eta = 10\%$	0.331	0.125	0.096	0.051	0.025	-0.001
	(0.106)	(0.114)	(0.119)	(0.121)	(0.118)	(0.127)
No Shredding	-0.001	-0.002	0.001	0.000	0.000	0.000
	(0.093)	(0.091)	(0.099)	(0.099)	(0.099)	(0.10)
Panel B: Model with arbitragers						
$\eta = 1\%$	0.033	0.047	0.098	-0.16	-0.052	-0.005
	(0.119)	(0.140)	(0.106)	(0.103)	(0.122)	(0.130)
$\eta = 5\%$	0.085	0.093	0.157	0.41	-0.04	0.002
	(0.131)	(0.150)	(0.111)	(0.114)	(0.133)	(0.141)
$\eta = 10\%$	0.123	0.094	0.14	0.413	-0.022	0.000
	(0.132)	(0.150)	(0.115)	(0.120)	(0.132)	(0.140)
No Shredding	-0.001	-0.002	0.001	0.000	0.000	0.000
	(0.093)	(0.091)	(0.099)	(0.099)	(0.099)	(0.10)

Table reports average autocorrelation coefficients of daily returns at different lags for different models of bet shredding: "Method-$V(1\%)$", "Method-$V(5\%)$", "Method-$V(10\%)$", and no bet shredding. Panel A presents results for the model without arbitragers. Panel B presents results for the model with arbitragers. The simulation consists of 1,000 of 90-day paths.

Figure 2.9: Distributions of simulated moments without arbitrageurs

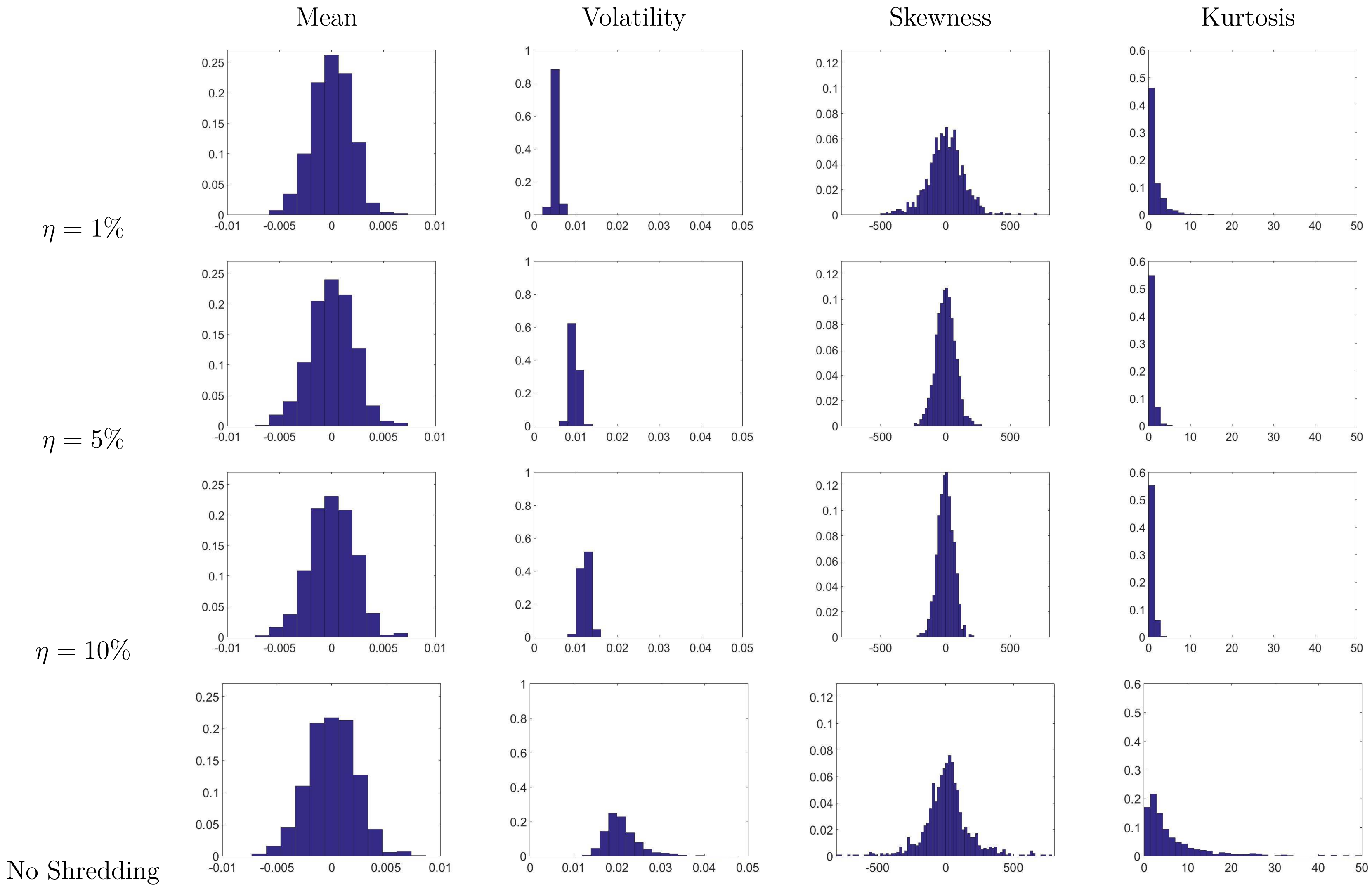


FIGURE SHOWS DISTRIBUTIONS OF SIMULATED MOMENTS OF ORDER FLOW FOR A BENCHMARK STOCK. THERE ARE 1,000 SIMULATIONS OF 90-DAY PATHS OF RETURNS. THE CASE WITH NO BET SHREDDING AND THE THREE BET-SHREDDING ALGORITHMS ARE USED: "METHOD-$V(1\%)$", "METHOD-$V(5\%)$", AND "METHOD-$V(10\%)$".

Figure 2.10: Distributions of simulated moments with arbitrageurs

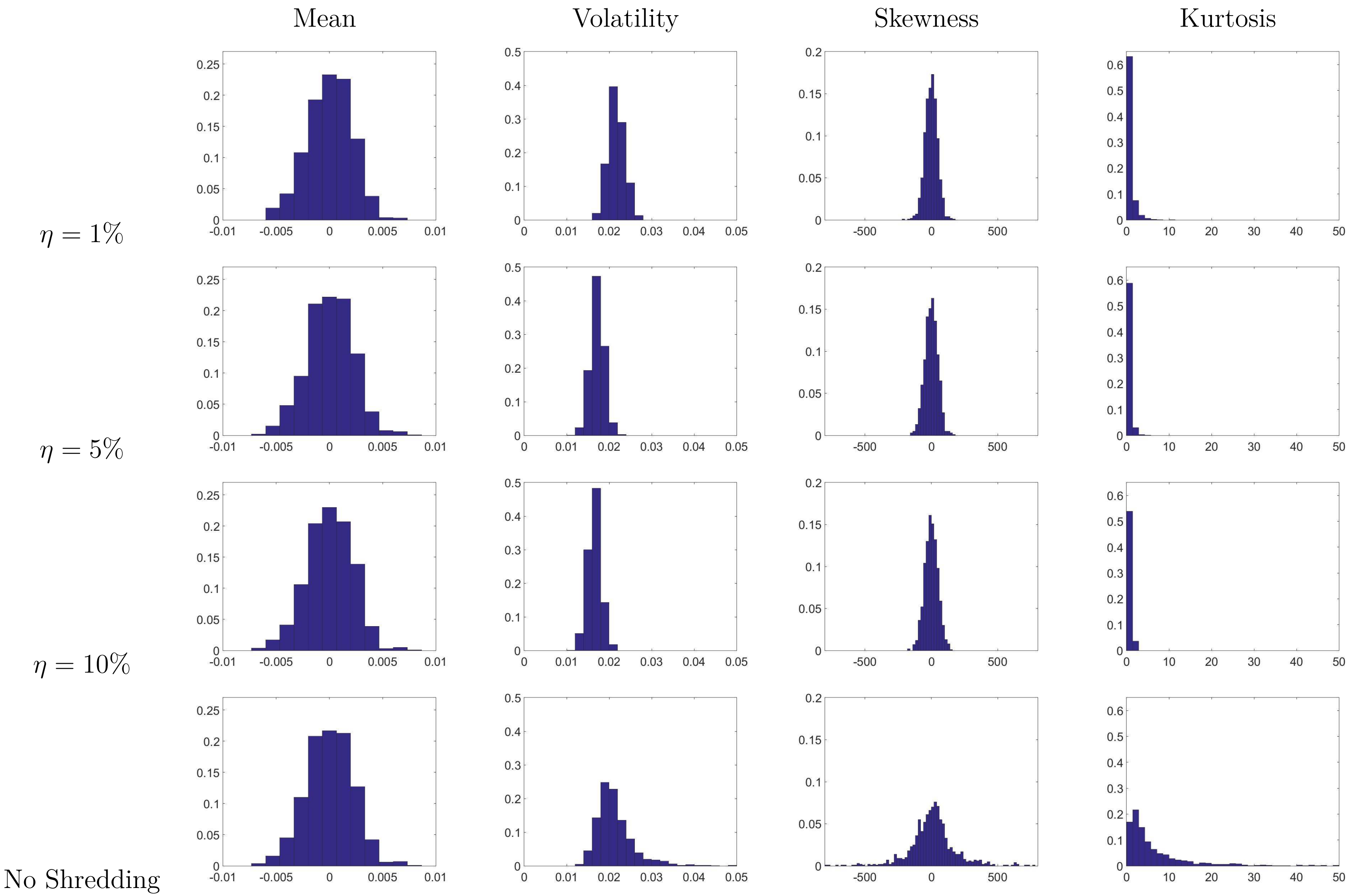


FIGURE SHOWS DISTRIBUTIONS OF SIMULATED MOMENTS OF ORDER FLOW FOR A BENCHMARK STOCK WITH ARBITRAGEURS. THERE ARE *** SIMULATIONS OF 90-DAY PATHS OF RETURNS. THE CASE WITH NO BET SHREDDING AND THE THREE BET-SHREDDING ALGORITHMS ARE USED: “METHOD-V(1%)”, “METHOD-V(5%)”, AND “METHOD-V(10%)”.

Table 2.5: Summary statistics for daily returns.

	$\eta = 1\%$	$\eta = 5\%$	$\eta = 10\%$	No Shredding
PANEL A: MODEL WITHOUT ARBITRAGER				
Mean	0.000	0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)
St.dev	0.005	0.010	0.012	0.021
	(0.001)	(0.001)	(0.001)	(0.005)
Skewness	0.252	0.722	0.451	8.564
	(140)	(74)	(61)	(178)
Kurtosis	0.927	0.338	0.281	7.331
	(1.745)	(0.632)	(0.556)	(10.421)
PANEL B: MODEL WITH ARBITRAGER				
Mean	0.000	0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)
St.dev	0.022	0.017	0.017	0.021
	(0.002)	(0.002)	(0.002)	(0.005)
Skewness	-1.152	-0.497	-0.457	8.564
	(46)	(48)	(49)	(178)
Kurtosis	0.517	0.229	0.171	7.331
	(0.943)	(0.553)	(0.556)	(10.421)

TABLE REPORTS STATISTICS FOR SIMULATED DAILY RETURNS SUCH AS MEAN, STANDARD DEVIATION, SKEWNESS, AND KURTOSIS FOR DIFFERENT MODELS OF BET SHREDDING: "METHOD-$V(1\%)$", "METHOD-$V(5\%)$", "METHOD-$V(10\%)$", AND NO BET SHREDDING. PANEL A PRESENTS RESULTS FOR THE MODEL WITHOUT ARBITRAGERS. PANEL B PRESENTS RESULTS FOR THE MODEL WITH ARBITRAGERS. THE SIMULATION CONSISTS OF *** 90-DAY PATHS.

Table 2.6: Calibrated bet-shredding parameters.

	Group 1	Group 3	Group 5	Group 8	Group 10
			Decade 1950-1960		
W	101	517	967	1,926	5,925
$\hat{\eta}$	8.875	6.288	6.500	3.675	3.225
			Decade 1960-1970		
W	277	1,650	3,375	6,999	26,727
$\hat{\eta}$	9.400	6.888	5.200	4.013	2.588
			Decade 1970-1980		
W	210	2,249	5,103	11,519	39,913
$\hat{\eta}$	9.40	6.89	5.20	4.01	2.59
			Decade 1980-1990		
W	1,176	12,893	29,588	70,470	271,798
$\hat{\eta}$	9.17	4.84	3.71	2.79	1.59
			Decade 1990-2000		
W	2,895	38,104	88,975	245,021	1,232,159
$\hat{\eta}$	8.36	4.22	3.46	2.52	1.66
			Decade 2000-2010		
W	8,640	190,720	468,578	1,364,880	6,704,911
$\hat{\eta}$	6.34	2.94	2.41	1.79	1.37
			Decade 2010-2017		
W	18,363	411,722	940,703	2,216,188	7,141,896
$\hat{\eta}$	5.04	2.69	2.32	1.80	1.39

Table presents calibrated parameter η and trading activity W for the median stocks in the five trading activity groups and for each decade for the period 1950 through 2017.

Table 2.7: Implied execution horizons.

	Group 1	Group 3	Group 5	Group 8	Group 10
			Panel A: Decade 1990-2000		
W	2,895	38,104	88,975	245,021	1,232,159
$\hat{\eta}$	8.36	4.22	3.46	2.52	1.66
std-1	0.02	0.01	0.01	0.00	0.00
std-2	0.11	0.04	0.02	0.02	0.01
std-3	0.55	0.18	0.12	0.08	0.04
std-4	2.69	0.91	0.60	0.39	0.17
std-5	13.22	4.44	2.93	1.92	0.85
std-6	64.86	21.80	14.36	9.42	4.16
std-7	318.23	106.96	70.46	46.22	20.39
			Panel B: Decade 2010-2017		
W	18,363	411,722	940,703	2,216,188	7,141,896
$\hat{\eta}$	5.04	2.69	2.32	1.80	1.39
std-1	0.01	0.00	0.00	0.00	0.00
std-2	0.03	0.01	0.01	0.00	0.00
std-3	0.16	0.04	0.03	0.02	0.01
std-4	0.79	0.19	0.12	0.09	0.05
std-5	3.86	0.91	0.61	0.44	0.26
std-6	18.93	4.46	2.98	2.17	1.29
std-7	92.87	21.88	14.63	10.65	6.32

Table presents implied execution horizons for bets of different sizes for different trading activity groups and time periods. There are calibrated parameter η, trading activity W, and execution horizons (in minutes) for 1 through 7 standard deviation bets.

Chapter 3

Size of Share Repurchases and Market Microstructure

3.1 Introduction

Share repurchases are among the most important corporate decisions. This study concerns what determines the size of repurchase programs and interprets these programs in the context of the market microstructure, as bets on the valuation of companies that managers place in the marketplace. Inspired by market microstructure invariance of Kyle and Obizhaeva (2016), the study documents quantitative empirical relationships between the size of share repurchase programs and trading activity of company stocks, volatility, and the duration of repurchase programs.

According to invariance theory, stock trading can be described as a trading game, in which market participants place bets on asseta. The number of bets and distribution of their sizes differs across assets with different levels of trading volume and volatility in a particular manner. If share repurchases are simply a special type of buy bets, then the insights of invariance theory have to be applicable to these corporate decisions as well. This interpretation of share repurchases allows us to formulate several hypotheses about their sizes.

The first hypothesis of target size is based on the intuition that the size of a repurchase program is simply proportional to the size of bet, typical for the underlying stock market; it predicts that size of repurchase programs as a fraction of expected trading volume is proportional to trading activity to the power of -2/3, where trading activity is defined as the product of dollar volume and return volatility.

The second hypothesis of target imbalance says that the size of repurchase programs is proportional to some percentile of the expected sum of all buy bets that company managers expect to generate over the duration of the repurchase program; it predicts that size of repurchase programs as a fraction of expected trading volume is proportional to trading activity to the power of -1/3 and also depends on the duration of the repurchase program.

The third hypothesis of target cost says that the size of repurchase program is determined by the execution costs of repurchases; it predicts that size of repurchase programs as a fraction of expected trading volume is proportional to trading activity to the power of -1/3 and also depends on the volatility of the underlying security.

I test the hypotheses using the sample of U.S. share repurchase programs over the period from March 1985 to January 2014. I find that trading activity does indeed have high explanatory power for the authorised and realised size of share repurchase programs; the regression r-square is equal to 41 percent for authorised sizes and 26 percent for realised sizes. The estimated coefficient on trading activity is -0.33, which conforms to predictions of target imbalance and target (linear) cost hypotheses. The formal statistical tests, however, reject these hypotheses.

I implement a formal model selection procedure with a Bayesian information criterion. Target imbalances and target (linear) costs hypotheses fit repurchase data best. Furthermore, the target imbalance hypothesis is selected on the open market repurchase programs, which are the most popular type of such programs.

This paper relates to literature on corporate payout policies and share repurchases. Dittmar (2000) argues that companies repurchase shares for various reasons. Firstly, there is market undervaluation theory. Vermaelen (1981), Brav et al. (2005), Buffa and Nicodano (2008) argue that companies initiate share repurchases to signal disagreement with current market valuations of their stocks. Ikenberry, Lakonishok and Vermaelen (1995), Mitchell and Stafford (2000) document positive abnormal returns several years after repurchase announcements. Secondly, there is a free cashflow theory of share repurchases. Jensen (1986), Jensen and Meckling (1976), and Stephens and Weisbach (1998) argue that firms should distribute all available cash to the shareholders through dividends or share repurchase to avoid agency costs arising due to a conflict of interests between management and shareholders. Thirdly, there is optimal corporate structure theory. Modigliani and Miller (1958), Bagwell and Shoven (1988), and Opler and Titman (1996) argue that companies repurchase shares to adjust their capital structures.

These theories identify several factors that influence share repurchase activity, such as past performance of underlying stock, volatility of stock, free cashflows, and market capitalization of the company. I assess the explanatory power of these theories relative to the hypotheses motivated by invariance theory. I find that the R^2 of regression specification that includes only trading activity equals 41 percent. Adding control motivated by theories of share repurchases increases the R^2 by 12 percent. It implies that trading activity of the stock is an important determinant of the size of a share repurchase program.

This paper also contributes to growing literature on market microstructure invariance. Kyle and Obizhaeva (2016) introduce market microstructure invariance principles that explain a substantial fraction of the cross-sectional and time-series variation in bet size and transaction costs across stocks. Kyle and Obizhaeva (2017*c*) derive invariance relationships in the infinite-horizon model of informed trading, noise trading, market making, and endogenous production of information. Kyle and Obizhaeva (2017*a*) establish invariance principles through dimensional analysis arguments. Andersen et al. (2014) document robust empirical patterns of intra-day trading in E-mini futures S&P 500 futures market. Kyle et al. (2011) apply invariance intuition to explain cross-sectional and time-series variation in news arrival rates. Bae et al. (2014) uncover invariant patterns in Korean stock market data. Kyle and Obizhaeva (2017*b*) apply invariance motivated market impact costs to explain market crashes. Kyle, Obizhaeva and Tuzun (2016) apply invariance principles to explain the number of trades and the distribution of trade sizes in the Trades and

Quotes database.

The rest of the paper is organized as follows. Section I provides a brief description of key principles of the market microstructure invariance framework and formulates three hypotheses about distribution of the size of repurchase program. Section II describes the data used in empirical tests. Section III empirically tests the hypotheses, selects the model that best fits the data, and compares alternative determinants of size of share repurchase programs. Section IV concludes the discussion.

3.2 Share repurchases and invariance

This section reviews market microstructure invariance and suggests several ways of how one can think about sizes of share repurchase programs in the context of this paradigm.

3.2.1 Review of market microstructure invariance

Market participants such as institutional and retail investors trade for various reasons. They trade to profit on their information or to meet their hedging needs. Invariance theory implies that the order flow of stocks is determined by risky bets. Bets arrive into market place according to the Poisson process with expected number of bets per day γ. The size of the bet Q is measured in shares. It is positive for buy and negative for sell bets. The order flow of stocks differs according to how many bets arrive and to the distribution of bet size.

Kyle and Obizhaeva (2017c) express bet size $\tilde{Q}$, expected number of bets γ, and execution cost of bet of size Q, $C(Q)$, in terms of observable trading activity of the stock W.

$$W = P \cdot V \cdot \sigma, \tag{3.1}$$

where σ is the volatility of the stock, V is trading volume of the stock, and P is its price level.

The distribution of bet size $\tilde{Q}$ as a fraction of expected daily trading volume V adjusted for trading activity of stock as $\frac{\tilde{Q}}{V} \cdot W^{2/3}$ has invariant distribution. It implies that

$$\frac{\tilde{Q}}{V} \sim W^{-2/3} \cdot \tilde{I}, \tag{3.2}$$

where $\tilde{I}$ is a random variable with an invariant probability distribution. Kyle and Obizhaeva (2016) use the sample of portfolio transition trades in the U.S. stock market over the period from 2001 to 2005 and calibrate the distribution of bet size. They find that $\tilde{I}$ is close to a log-normal distribution with log-variance $\sigma_Q^2 = 2.53$.

The expected number of bets γ is predicted to be proportional to $W^{2/3}$,

$$\gamma \sim W^{2/3}. \tag{3.3}$$

The execution cost of a bet of size Q, $C(Q)$, is predicted to be

$$C(Q) = \sigma \cdot W^{-1/3} \cdot \imath^2 \cdot \bar{C}_B \cdot f\left(\frac{W^{2/3}}{\imath} \cdot \frac{Q}{V}\right), \tag{3.4}$$

where $f(\cdot)$ is the invariant average cost function and $\imath := (E[|\tilde{I}|])^{-1/3}$ and $\bar{C}_B$ are some constants. Invariance is consistent with any functional form of $f(\cdot)$, but most often assume linear or square root cost functions.

For the benchmark stock with daily returns volatility $\sigma^* = 2\%$, trading volume $V^* = 10^6$ shares, price $P^* = \$40$, and trading activity $W^* = 800,000$ there are on average $\gamma^* = 85$ bets per day, the average size of bet is equal to $33,000$ shares, and the execution cost of an average sized bet for a linear cost model is equal to 14 basis points.

3.2.2 Share repurchases hypotheses

Under the assumption that company management actively participates in the trading process of company's stock, I formulate three hypotheses about the size of share repurchase programs.

Hypothesis 1: target size hypothesis. The first hypothesis says that a share repurchase program represents a buy bet that the company executes in the market. Thus, the invariance predictions have to apply to the size of repurchase program. Let X denote the size of repurchase program. Equation (3.2) implies the following relation between the size of the repurchase and trading activity of the underlying stock.

$$\frac{X}{V} \sim W^{-2/3} \cdot \tilde{I}. \tag{3.5}$$

The size of repurchase program as a fraction of expected daily volume adjusted for trading activity of stock has invariant distribution.

This hypothesis can be tested with the following log-linear regression using the panel of share repurchase programs.

$$\ln\left[\frac{X_{it}}{V_{it}}\right] = \alpha_0 + \alpha_W \cdot \ln\left[\frac{W_{it}}{W^*}\right] + \tilde{\epsilon}_{it}, \tag{3.6}$$

where X_{it} denotes the size of share repurchase program i initiated at time t. At time t the stock i has expected daily trading volume V_{it}, expected price level P_{it}, expected daily return volatility σ_{it}, expected trading activity W_{it}, and trading activity of the benchmark stock $W^* = 800,000$. Target size hypothesis predicts that $\alpha_W = -2/3$ in regression model (3.6).

Hypothesis 2: target imbalance hypothesis. The second hypothesis says that the size of repurchase program is related to total size of buy bet-ideas that the company is expected to generate over the duration of a repurchase program.

Let T denote duration of repurchase program. It is most likely that a manager will come up with several bets on the company over the length of a repurchase program. For example, over a three year repurchase program the company generates the following bets. In the first year the company decides to buy back 1 million shares since it believes that its stock is undervalued. In the second year the firm repurchases 400,000 shares to adjust its capital structure. In the third year the company repurchases 600,000 shares to pay back cash to its shareholders. In this case $T = 3$ and the size of repurchase program is equal to $X = 1,000,000 + 400,000 + 600,000 = 2,000,000$ shares.

Target imbalance hypothesis asserts that at the announcement of a repurchase program, the size of program may reflect forward looking estimates of the total size of buy bets that the company expects to generate over the duration of the program.

Let γ_f denote the expected number of bets that management of a company is expected to generate every day. The actual number of bets that a company generates $\tilde{\gamma}_f$ is a random variable that has a Poisson distribution with expected number of bets γ_f. I assume that the expected number of bets that a company generates γ_f is proportional to expected number of bets in the underlying stock market γ.

$$\gamma_f = z_f \cdot \gamma, \tag{3.7}$$

where z_f denotes some proportionality constant.

Suppose that on day t the manager of the company generates $\tilde{\gamma}_{ft} = \gamma_{ft}$ bets $\tilde{Q}_{1t} = Q_{1t}, \tilde{Q}_{2t} = Q_{2t}, ..., \tilde{Q}_{\gamma_{ft}t} = Q_{\gamma_{ft}t}$. Since the number of bets and their sizes are random variables, the bet imbalance at time t $\tilde{\Psi}_t$ - a sum of all bets that company generates at time t - is a random variable as well.

$$\tilde{\Psi}_t = \sum_{i=1}^{\tilde{\gamma}_{ft}} \tilde{Q}_{it}. \tag{3.8}$$

Suppose the size of repurchase program represents some upper percentile estimate of the distribution $\tilde{\Psi}_t$, such as the 95th percentile of total bet imbalance. Therefore, the manager should estimate the standard deviation of bet imbalance $\tilde{\Psi}_t$. I apply the law of total variance to equation (3.8) to estimate variance of daily bet imbalance.

$$\mathbb{V}ar\left[\tilde{\Psi}_t\right] = \mathbb{E}\left(\mathbb{V}ar\left[\sum_{i=1}^{\tilde{\gamma}_{ft}} \tilde{Q}_{it}\right]\right) + \mathbb{V}ar\left(\mathbb{E}\left[\sum_{i=1}^{\tilde{\gamma}_{ft}} \tilde{Q}_{it}\right]\right). \tag{3.9}$$

The second term in equation (3.9) is equal to zero, since the average size of each bet $\tilde{Q}_{it}$ is equal to zero $\mathbb{E}\left[\tilde{Q}_{it}\right] = 0$ and $\mathbb{E}\left[\sum_{i=1}^{\tilde{\gamma}_{ft}} \tilde{Q}_{it}\right] = 0$.

Since the number of bets has a Poisson distribution with expected number of bets γ_f , I apply equation (3.11) and compute the first term of equation (3.9) as

$$\mathbb{E}\left(\mathbb{V}ar\left[\sum_{i=1}^{\tilde{\gamma}_{ft}}\tilde{Q}_{it}\right]\right)=\sum_{k=1}^{\infty}\mathbb{V}ar\left[\sum_{i=1}^{\tilde{\gamma}_{ft}}\tilde{Q}_{it}\middle|\tilde{\gamma}_{ft}=k\right]\cdot\frac{\gamma_f^k}{k!}\cdot e^{-\gamma_f}= \tag{3.10}$$

$$=\mathbb{V}ar\left[\tilde{Q}_{it}\right]\cdot\sum_{k=1}^{\infty}\frac{k\cdot\gamma_f^k}{k!}\cdot e^{-\gamma_f}=\gamma_f\cdot\mathbb{V}ar\left[\tilde{Q}_{it}\right].$$

Since bets are independent, the variance of bet imbalance conditional on the number of bets equals

$$\mathbb{V}ar\left[\sum_{i=1}^{\tilde{\gamma}_{ft}}Q_{it}\middle|\tilde{\gamma}_{ft}=k\right]=k\cdot\mathbb{V}ar[\tilde{Q}_{it}]. \tag{3.11}$$

Since the share repurchase program lasts for T days, I calculate standard deviation of bet imbalance generated by the company over the T days of the program $\tilde{\Psi}(T)$. Substituting (3.7) and (3.10) into (3.9), gives the following standard deviation of bet imbalance.

$$\mathrm{std}\left[\tilde{\Psi}(T)\right]=\sqrt{T\cdot\gamma_f\cdot\mathbb{V}ar\left[\tilde{Q}_{it}\right]}. \tag{3.12}$$

If the size of repurchase program is proportional to the standard deviation of bet imbalance that the company is expected to generate over the T days of share repurchase program, then

$$\bar{X}\sim\sqrt{T\cdot\gamma_f\cdot\mathbb{V}ar\left[\tilde{Q}_{it}\right]}. \tag{3.13}$$

For example, the proportionality constant may be equal to 1.96 if the manager wants to target the 95th percentile.

Finally, I express the authorised size X of share repurchase program as a fraction of daily volume V. I substitute equations (3.2), (3.3), and (3.7) into equation (3.13).

$$\frac{\bar{X}}{V}\sim\sqrt{T\cdot\gamma\cdot\frac{\mathbb{V}ar\left[\tilde{Q}_{it}\right]}{V^2}}\sim\sqrt{T}\cdot W^{-1/3}. \tag{3.14}$$

This hypothesis can be tested with the following log-linear regression using the panel of share repurchase programs

$$\ln\left[\frac{X_{it}}{V_{it}}\right] = \alpha_0 + \alpha_W \cdot \ln\left[\frac{W_{it}}{W^*}\right] + \alpha_\tau \cdot \ln(T_{it}) + \tilde{\epsilon}_{it}. \tag{3.15}$$

Target imbalance hypothesis predicts that $\alpha_W = -1/3$ and $\alpha_\tau = 1/2$ in the regression model (3.15).

Hypothesis 3: target cost hypothesis. The execution cost of share repurchase program is an important factor. The third hypothesis asserts that managers choose the size of repurchase program to target the percentage execution cost of share repurchase program, z_c. For example, the company may want to repurchase 1 million shares and target execution cost of 50 basis points. In this case, $X = 1,000,000$ shares and $z_c = 50$ basis points. I apply invariance percentage cost formula (3.4) to formalise this hypothesis. Inverting equation (3.4), I express the size of repurchase program X as a fraction of expected daily volume V as

$$\frac{X}{V} = \imath \cdot W^{-2/3} \cdot f^{-1}\left(z_c \cdot \sigma^{-1} \cdot W^{1/3} \cdot \frac{1}{\imath^2 \cdot \bar{C}_B}\right). \tag{3.16}$$

Next, I assume that execution cost function $f(x)$ has a power functional form,

$$f(x) = \lambda \cdot x^\beta. \tag{3.17}$$

The literature typically considers linear and square root market impact functions, as in Kyle (1985) and Gabaix et al. (2006), respectively. The linear execution cost function corresponds to the case of $\beta = 1$. The square root execution cost function corresponds to the case of $\beta = 1/2$. The inverse of the power function (3.17) is $f^{-1}(x) = \left(\frac{x}{\lambda}\right)^{\frac{1}{\beta}}$.

Substituting equation (3.17) into equation (3.16), yields the size of repurchase program as a fraction of trading volume,

$$\frac{X}{V} \sim \sigma^{-\frac{1}{\beta}} \cdot W^{\frac{1-2\cdot\beta}{3\cdot\beta}}. \tag{3.18}$$

For the linear execution costs function $\beta = 1$ and expression (3.18) simplifies to

$$\frac{X}{V} \sim \sigma^{-1} \cdot W^{-1/3}. \tag{3.19}$$

For the square root execution costs $\beta = 1/2$ and expression (3.18) simplifies to

$$\frac{X}{V} \sim \sigma^{-2}. \tag{3.20}$$

This hypothesis can be tested with the following log-linear regression using the panel of share repurchase programs

$$\ln\left[\frac{X_{it}}{V_{it}}\right] = \alpha_0 + \alpha_W \cdot \ln\left[\frac{W_{it}}{W^*}\right] + \alpha_\sigma \cdot \ln(\sigma_{it}) + \tilde{\epsilon}_{it}. \tag{3.21}$$

Target cost hypothesis makes the following predictions for the cases of linear and square root execution cost functions. For the case of linear cost function $\alpha_W = -1/3$ and $\alpha_\sigma = -1$ in the regression model (3.21). For the case of square root cost function $\alpha_W = 0$ and $\alpha_\sigma = -2$ in the regression model (3.21).

3.2.3 Nested models

All hypotheses make different predictions about the relationship between size of repurchase program and trading activity of the underlying stock. The target size hypothesis (3.6), target imbalance hypothesis (3.15), and both target cost hypotheses (3.21) may be described by the nested regression model

$$\ln\left[\frac{X_{it}}{V_{it}}\right] = \alpha_0 + \alpha_W \cdot \ln\left[\frac{W_{it}}{W^*}\right] + \alpha_\tau \cdot \ln(T_{it}) + \alpha_\sigma \cdot \ln(\sigma_{it}) + \tilde{\epsilon}_{it}. \tag{3.22}$$

The target size hypothesis predicts that the only factor determining the size of repurchase program is trading activity of the stock, while the duration of a repurchase program and volatility of the stock should not matter. It predicts that $\alpha_W = -2/3$, while $\alpha_\tau = 0$ and $\alpha_\sigma = 0$.

The target imbalance hypothesis predicts that trading activity of the stock and duration of repurchase program determine the size of the repurchase program, while volatility of the stock should not matter. It predicts that $\alpha_W = -1/3$, while $\alpha_\tau = 1/2$ and $\alpha_\sigma = 0$.

The target linear impact cost hypothesis predicts that trading activity of the stock and volatility of the stock determine size of the repurchase program, while duration of the program should not matter. It predicts that $\alpha_W = -1/3$, $\alpha_\tau = 1/2$, while $\alpha_\sigma = 0$.

Target square root impact cost hypothesis predicts that only volatility of the stock should determine the size of the repurchase program, while trading activity of the stock and duration of the repurchase program should not matter. It predicts that $\alpha_W = 0$ and $\alpha_\tau = 0$, while $\alpha_\sigma = -2$.

Table 3.1 summarizes hypotheses predictions for the nested regression (3.22).

Table 3.1: Hypotheses predictions

Hypothesis	α_W	α_τ	α_σ
1. Target size hypothesis	-2/3	0	0
2. Target imbalance hypothesis	-1/3	1/2	0
3. Target cost hypothesis:			
Case 1: Linear cost	-1/3	0	-1
Case 2: Square root cost	0	0	-2

In the next section I test these hypotheses empirically.

3.3 Data

I use data from the Securities Data Company Platinum (SDC) database on the U.S. share repurchase programs covering period from March 1985 to January 2014. It contains information about the date when a repurchase program was authorised by the board of directors, the end date of a repurchase program, the authorised size of share repurchase program, the realised size of the program, methods of repurchase, and reasons for share repurchase.

I merge the repurchase database with information on trading activity of the underlying stock from the Center for Research in Securities Prices (CRSP) database using 8-digit historical CUSIP numbers. Only US ordinary common shares (with share codes 10 or 11) that are listed on NYSE, AMEX, Nasdaq, and NYSE Arca are considered. At the initial merge stage I cannot match 3,477 observations and exclude 2 observations due to obvious typographical errors. Furthermore, I exclude 2,069 observations because because information required for construction of the explanatory variables is absent from the CRSP database. The resulting sample contains 14,369 repurchase programs.

I estimate trading activity of stocks using CRSP data. For each share repurchase program i, $i = 1, ..., 14369$, I estimate expected daily trading volume V_i of corresponding stock as a sample average daily volume in a calendar month prior to the authorisation date of the share repurchase program. I estimate expected volatility of stock returns σ_i as a sample standard deviation of daily log-returns in a calendar month before the initiation of the repurchase program. To account for possible stock splits, I estimate expected dollar volume of the stock $P_i \cdot V_i$ as an average daily dollar trading volume in a calendar month before the share repurchase authorisation. I estimate expected trading activity of the stock W_i as a product of expected dollar volume and expected return volatility $W_i = P_i \cdot V_i \cdot \sigma_i$.

Trading activity of the stock may be affected by the initiation of a repurchase program. For example, it may substantially increase the trading volume because of amplified public attention. As a robustness check I also consider an alternative estimate of expected trading activity and expected volume using data in the calendar month prior to the completion of the share repurchase program. The results that use alternative estimates are qualitatively and quantitatively similar to the base case results. These results are available upon request.

I do analysis on the sample of U.S. repurchase programs and on three sub-periods. The first period is from March 1985 to December 2000. The second period is from January 2001 to December 2007. The third period is from January 2008 to September 2014.

The first period starts shortly after implementation of the safe-harbor Rule 10b-18 by the Security and Exchange Commission in November 1982. Grullon and Michaely (2002) argue that the new legislation stimulated repurchase activity in the United States. The aggregate amount of cash spent on share repurchases tripled a year after the adoption of this rule and the number of companies that initiated share repurchases increased from 14 percent during the 1970s to 30 percent during the 1990s.

The second period corresponds to the period of low dividend taxes introduced by George W. Bush in 2001, also known as "Bush tax cuts". Chetty and Saez (2006) finds that the tax reduction from 35 percent to 15 percent resulted in a sharp increase in the size of repurchase programs. Authors also established no substitution effect between dividends and share repurchases.

The third period corresponds to the post 2008 financial crisis period. Adverse economic conditions affect the payout policy of companies. As illustrated in Figure 3.1 the number of share repurchase programs that were initiated during the financial crisis in the United States dropped from 700 to 500. The dollar size of repurchase programs experienced a sharp decline from $8 billions in 2008 to $3 billions in 2009. Share repurchase activity recovered to the pre-crises levels in 2010.

Table 3.2 reports characteristics of the considered sample of share repurchase programs over the period from 1985 to 2000 as well as for the three sub-periods. Panel A of Table 3.2 reports characteristics such as the authorised size of repurchase programs, the realised size of repurchase programs, and the duration of the programs. Over the considered period the median authorised size of repurchase program is 1.76 million shares. It increased from 1 million shares in the first period to 4 million shares in the third period. The Authorised size of repurchase program as a fraction of trading volume, however, decreases from 31 in the first period to 12 in the third period, because the trading volume of the stocks increases faster than size of repurchase program.

Usually companies do not repurchase all the shares that they are authorised to repurchase. The median realised size of repurchase program is 1 million shares. It increased from 0.69 million shares in the first period to 2 millions shares in the third period.

I also estimate the duration of the repurchase program as the number of days between the authorisation date and end date of the repurchase program. The median duration of the program is 317 days. It decreased from 334 days in the first period to 248 days in the third period, which implies that companies are buying back shares faster.

Panel B of Table 3.2 characterises trading activity of the stocks in the SDC repurchases sample. Median daily dollar volume is \$1.27 million. It increased from \$0.5 million in the first period to \$8.6 million in the third period. Median daily volatility of log returns is stable at 2%. Median trading activity of stocks is $30,000$. Stocks become more actively traded over time, as trading activity increased from $12,000$ in the first period to $195,000$ in the third period.

Companies may use different methods to buy back their stocks. There are six different methods that companies usually use: open market, negotiated, Dutch auction tender offers, accelerated, first price tender offers, and odd lot repurchases. Sometimes companies employ several methods to repurchase shares within one repurchase program.

Table 3.3 provides summary statistics of share repurchase programs across different repurchase methods. The most popular is open market repurchases. According to Stephens and Weisbach (1998) open market share repurchase programs represent approximately 90 percent of the dollar value of all announced share repurchase programs. Open market repurchases usually last around one year, which gives a company flexibility on the timing and quantity of actual shares repurchased. The median authorised size of share repurchases is 1.58 million shares and the number of actual shares repurchased by companies is 1 million shares. Consistent with Stephens and Weisbach (1998), this study finds a completion rate of approximately

64 percent of repurchase programs.

Dutch auction, accelerated and first price repurchases require commitment by the company to repurchase shares. To achieve effective signaling, these repurchase programs are larger and of shorter duration - usually around 2 months - than open market share repurchase programs. The median authorised size is equal to 3.55 million shares for Dutch auction, 6.15 million shares for accelerated repurchases, and 3.49 million shares for first price repurchases. The median total number of shares repurchased is 2.29 million, 6.54 million, and 1.85 million shares for Dutch auction, accelerated, and first price repurchases, respectively.

Another method is negotiated repurchases. Pursuing this method, a company negotiates the deal privately and repurchases stock from one or a few large shareholders. Dittmar (2000) suggests that this method is usually used to prevent a takeover threat. On average negotiated repurchases last for approximately a year. The median authorised size is 2.00 million shares and the median total number of shares repurchased is 1.19 million shares.

Finally, some companies engage in odd-lot share repurchases to eliminate odd-lot shareholders. Vermaelen (2000) argues that companies use this method to reduce shareholder servicing costs. Odd-lot programs usually last about two months. Their median authorised size is 1.6 million shares and the median total number of shares repurchased is 1.00 million shares.

3.4 Results

This section tests the hypotheses concerning the size of repurchase programs and discusses which model conforms best to the data.

3.4.1 Hypothesis testing

Figure 3.2 shows the relationship between the size of a share repurchase program and the trading activity of the underlying stock. Panel A of Figure 3.2 displays the logarithm of authorised sizes of repurchase programs on the x-axis versus the logarithm of trading activity of the stocks on the y-axis. Apart from the four outliers that correspond to small odd-lot repurchase programs, all observations cluster tightly around the line with the slope of -1/3. This implies that trading activity of stock has a high explanatory power for the authorised size of a share repurchase program, which is consistent with implications of invariance theory.

Table 3.5 presents the estimates of regression (3.6) on the whole sample of repurchase programs over the period from 1985 to 2014 and the sub-periods from 1985 to 2000, from 2001 to 2007, and from 2008 to 2014. Panel A of Table 3.5 displays results for the authorised size of repurchase program. On the whole sample of repurchase programs the estimated $\hat{\alpha}_W = -0.33$ with a standard error of 0.005. Stock trading activity explains 40 percent of variation in the authorised size of repurchase programs. The estimated coefficient estimate $\hat{\alpha}_W$ is -0.30, -0.33, and -0.33 with standard errors of 0.01, 0.01, and 0.02 for the considered sub-periods. Trading activity explains 28 percent, 45 percent, and 46 percent of total variation of the authorized size of repurchase program over the considered sub-periods. The estimates are economically close to -1/3. This is consistent with the predictions of target imbalance and target costs hypotheses, but not with the target size hypothesis.

Table 3.5 presents the estimates of regression (3.6), controlling for time and industry fixed effects, on the whole sample of repurchase programs over the period from 1985 to 2014 and the sub-periods from 1985 to 2000, from 2001 to 2007, and from 2008 to 2014. Panel A of Table 3.5 displays results for the authorised size of repurchase programs. On the whole sample of repurchase programs the estimated $\hat{\alpha}_W = -0.34$ with a standard error of 0.007. The estimated coefficient $\hat{\alpha}_W$ is -0.32, -0.33, and -0.36 with standard errors of 0.01, 0.01, and 0.01 respectively for the considered sub-periods. The results are robust, controlling for time and industry variation.

Figure 3.3 reports the estimates of regression (3.6) every year from 1994 to 2014. I exclude 121 share repurchase programs that were initiated before 1994 to avoid the problem of small sample bias. I find that estimate $\hat{\alpha}_W$ is outside 95 percent confidence intervals of -1/3 in only 5 out of 20 years. Year by year estimates of $\hat{\alpha}_W$

are between -0.381 and -0.275.

The trading activity of stock appears to explain well the cross-sectional variation in the authorised size of repurchase programs. The estimated $\hat{\alpha}_W$ is economically close to -1/3. A formal F-test rejects the targeted size hypothesis that predicts that $\alpha_W = -2/3$ on the whole sample of repurchase programs and over the considered sub-periods. However, the above empirical findings are consistent with the target imbalance and target execution costs hypotheses that predict $\alpha_W = -1/3$.

I also analyze relationship between trading activity of stock and the realised size of repurchase programs. Panel B of Figure 3.2 displays the relationship between the logarithm of the total number of shares repurchased on the y-axis versus the logarithm of trading activity of the underlying stocks on the x-axis. Trading activity does not explain well the total number of repurchased shares. Limited explanatory power may be attributed to the endogous termination of share repurchase programs that may be affected by various economic factors.

Panel B of Table 3.5 presents estimates of regression (3.6) for the realised number of shares repurchased. On the whole sample of repurchase programs $\hat{\alpha}_W = -0.31$ with a standard error of 0.007. The trading activity of the underlying stocks explains 26 percent of variation in the realised size of repurchase programs. The estimated $\hat{\alpha}_W$ is -0.31, -0.27, and -0.29 with the standard errors of 0.01, 0.01, and 0.01 for the considered sub-periods. Results, controlling for time and industry fixed effects, are found to be robust.

I estimate the nested regression model (3.22) to test the predictions of the target imbalance hypothesis and two specifications of the target costs hypothesis. Table 3.6 reports regression estimates for the total sample of repurchase programs and the considered sub-periods. The estimate $\hat{\alpha}_W$ is economically close to -1/3. I find that $\hat{\alpha}_\tau = 0.22$ with a standard error of 0.01 and $\hat{\alpha}_\sigma = -0.21$ with standard error of 0.09. Estimates of $\hat{\alpha}_\tau$ and $\hat{\alpha}_\sigma$ outside the 95 percent confidence interval predicted by the target imbalance hypothesis are $\alpha_\tau = 0.5$ and $\alpha_\sigma = 0$. The target imbalance hypothesis is rejected with the joined F test. The estimates $\hat{\alpha}_\tau$ and $\hat{\alpha}_\sigma$ are outside the 95 percent confidence interval predicted by the target linear execution costs hypothesis, at $\alpha_\tau = 0$ and $\alpha_\sigma = -1$. The target linear costs hypothesis is also rejected with the joined F test. The estimate $\hat{\alpha}_W$ is outside the 95 percent confidence interval predicted by the target square root execution costs hypothesis, at $\alpha_W = 0$. The targeted square root costs hypothesis is rejected with the joined F test as well.

Table 3.7 reports estimates of nested regression (3.22) across different repurchase methods: open market, negotiated, Dutch auction, accelerated, first price and odd lot repurchases programs. For the sample of open market share repurchases, I find $\hat{\alpha}_W = -0.328$, $\hat{\alpha}_\tau = 0.403$, and $\hat{\alpha}_\sigma = -0.216$ with clustered by year standard errors of 0.006, 0.021, and 0.027, respectively. The estimates are economically close to predicted $\alpha_W = -1/3$, $\alpha_\tau = 1/2$, and $\alpha_\sigma = 0$. However, a formal joint F test rejects the target imbalance hypothesis with F statistics of 41. The target imbalance and both specifications of the target costs hypothesis are also rejected with the F test. I find no evidence that supports the considered hypotheses for negotiated, Dutch auction, accelerated, first price and odd lot repurchase programs. Although formal statistical tests reject the considered hypotheses, open market repurchase programs conform to the target imbalance hypothesis.

3.4.2 Model selection

Next I implement model selection with Bayesian information criterion to identify which model best conforms with the data. I estimate the Bayesian information criterion on the total sample of repurchase programs and sub-samples that correspond to different repurchase methods. $BIC = -2\ln \hat{L} + k \cdot \ln(n)$, where $\hat{L}$ is the likelihood of the corresponding regression model, k is the number of parameters in the model, and n is the number of observations. For the target size hypothesis, k is equal to 1. For the target imbalance and both target costs hypotheses, k is equal to 2. The model that has the lowest estimate of BIC describes data the best.

Table 3.8 reports the results of model selection. The Bayesian information criteria selects the target imbalance hypothesis on the total sample of share repurchase programs and sub-samples of open market, first price, Dutch auction, and odd lot repurchase programs. The target size hypothesis is selected only for the case of accelerated share repurchases. The Target cost hypothesis (linear case) is selected for negotiated and accelerated share repurchases.

3.4.3 Alternative hypotheses

Previous sections established that trading activity of the stock is an important factor determining the size of a share repurchase program. In this section I consider other factors motivated by alternative theories and assess their explanatory power relative to the considered hypotheses inspired by invariance theory.

The literature on share repurchases identifies several factors that influence share repurchase activity of companies. Firstly, signalling and market undervaluation theories imply that a company repurchases its shares after a period of underperformance when the stock is undervalued. It predicts a negative relation between past performance of the stock and the size of repurchase program. This study uses past log-return of stock over three months R_{it}^{3m} to assess signalling theory.

Secondly, free cashflow theory implies that a company repurchases its shares when it has excess cash to avoid an agency problem. It predicts that cash of the company should be positively related to the size of its share repurchase program. Consistent with Stephens and Weisbach (1998), I use the cash of the company in the quarter before the repurchase announcement C_{it-1} reported in Compustat as a proxy for free cashflows of the company to assess the free cashflow theory.

Thirdly, targeted corporate structure theory predicts that a company repurchases its shares to alter its debt-to-equity ratio. It implies that the size of repurchase program relates negatively to the leverage of the company and the volatility of its stock σ_{it}. I use the volatility of the stock σ_{it} to assess targeted corporate structure theory.

Fourthly, the study examines the economy of scale argument, implying that a company may repurchase more shares when it has a larger market capitalisation M_{it}. I use the market capitalization of the company M_{it} to assess the economy of scale argument.

Finally, I consider the target size, the target imbalance, and both target costs hypotheses. These hypotheses state that trading activity of the underlying stock W_{it}, volatility of the stock σ_{it}, and duration of a repurchase program τ_{it} determine the size of repurchase program. I use these variables to assess the invariance motivated hypotheses.

To assess explanatory power of different theories, I estimate the nested regression model

$$\ln\left[\frac{X_{it}}{V_{it}}\right] = \alpha_0 + \alpha_W \cdot \ln\left[\frac{W_{it}}{W^*}\right] + \alpha_\tau \cdot \ln\left[T_{it}\right] + \alpha_\sigma \cdot \ln\left[\frac{\sigma_{it}}{\sigma^*}\right] + \\ + \alpha_{3m} \cdot R_{it}^{3m} + \alpha_\$ \cdot C_{it-1} + \alpha_M \cdot M_{it} + \tilde{\epsilon}_{it}. \tag{3.23}$$

I assess the explanatory power of the alternative theories relative to the baseline cases of regression (3.23) that corresponds to target imbalance and target costs (linear case) hypotheses. The baseline specification corresponds to regression (3.23) with imposed constraints on regression coefficients $\alpha_W = -1/3$, $\alpha_\tau = 0$, $\alpha_\sigma = 0$, $\alpha_{3m} = 0$, $\alpha_\$ = 0$, and $\alpha_M = 0$. To assess the explanatory power of alternative theories, I compare how much the R^2 increases relative to the R^2 in the base case specification.

Table 3.9 reports estimates of regression (3.23) for specifications that correspond to the considered theories. Column (3) of Table 3.9 assesses signalling and market undervaluation theory. Consistent with the prediction of the theory, I find that the size of repurchase program is negatively related to past performance of the underlying stock. However, after controlling for trading activity W of the stock, the estimated coefficient $\hat{\alpha}_{3m} = -0.2$ is not statistically significantly different from zero. Regression specification that controls for past performance increases R^2 by 3 percent relative to the baseline specification that has $R^2 = 41$ percent.

Column (4) of Table 3.9 assesses free cashflow theory. Consistent with the prediction of the theory, I find that the size of repurchase program is positively related to the cash of the company. However, after controlling for trading activity W of the stock, the estimated coefficient $\hat{\alpha}_{\$} = 0.4$ is not statistically significantly different from zero. The regression specification that controls for the cash of the company increases R^2 by 4 percent relative to the baseline specification that has $R^2 = 41$ percent.

Column (1) of Table 3.9 assesses targeted corporate structure theory. Consistent with the prediction of the theory, I find that the size of repurchase program is negatively related to volatility of the underlying stock. After controlling for trading activity W of the stock, estimated coefficient $\hat{\alpha}_{\sigma} = -0.32$. The regression specification that controls for volatility of the company's stock increases R^2 by 5 percent relative to the baseline specification that has $R^2 = 40$ percent.

Column (5) of Table 3.9 assesses the economy of scale argument. Consistent with this argument, I find that the size of repurchase program is positively related to the market capitalization of the company. After controlling for trading activity W of the stock, the estimated coefficient $\hat{\alpha}_M = 0.02$. The regression specification that controls for cash of the company increases R^2 by 6 percent relative to the baseline specification that has $R^2 = 40$ percent.

The regression specification that includes the five considered control variables has $R^2 = 53$ percent, which is 12 percent higher than the baseline regression specification that controls only for trading activity of the underlying stock. Applying Occam's razor principle, the trading activity of the underlying stock is the key factor that explains variation in the size of share repurchase programs.

3.5 Conclusion

This paper proposes an innovative way to think about company share repurchases. Using the insights of market microstructure invariance of Kyle and Obizhaeva (2016), I interpret companies' share repurchases as buy bets placed by the company management and formulate three hypotheses about the size of repurchase programs: target size, target imbalance, and target cost hypotheses. I find that trading activity of the stock is an important determinant of the size of a share repurchase program. Formal tests establish that target imbalances and target (linear) costs hypotheses fit repurchase data the best. Furthermore, the target imbalance hypothesis is selected on the open market repurchases – the most popular type of repurchase programs.

In future research would be interesting to analyse other corporate decisions, such as secondary share issuances and dividend payouts, from the perspective of market microstructure invariance. For example, share issuances may represent sell bets placed by the company management. Similar to the analysis of this paper, I conjecture that the sizes of secondary issuance programs depend on trading activities of the stocks.

3.6 Tables and figures

Figure 3.1: Historical share repurchase activity.

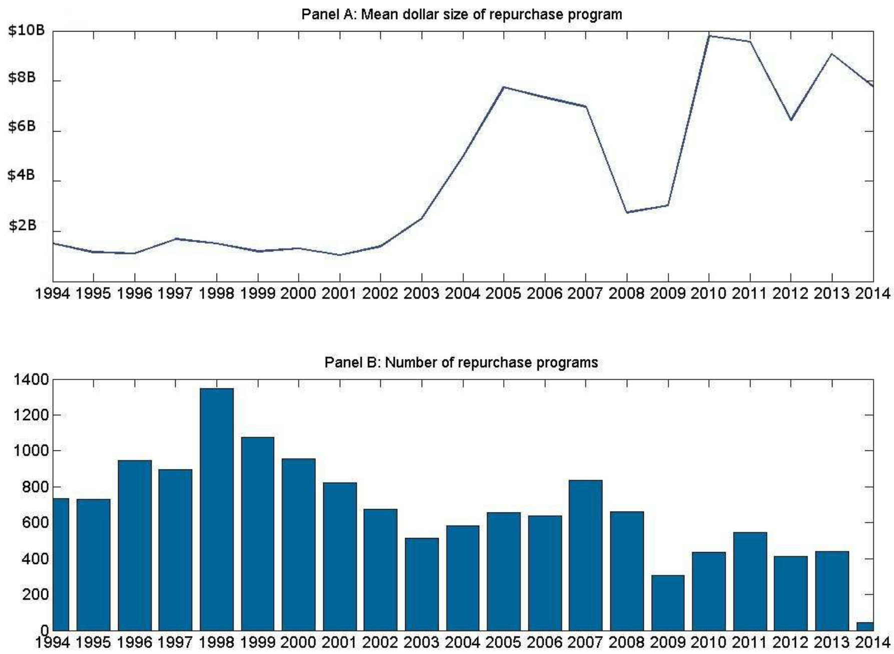


Figure 3.1 displays historical yearly share repurchase activity of US companies over period from 1994 to 2014. Panel A displays average size of repurchase program (in billions of dollars). Panel B displays number of repurchase programs initiated.

Table 3.2: Descriptive statistics.

	ALL	1985-2000	2001-2007	2008-2014
PANEL A: REPURCHASE PROGRAM CHARACTERISTICS				
$\bar{X}$	1.76	1.00	2.20	3.93
$\bar{X}/V$	21.97	30.82	16.53	12.10
X	1.02	0.69	1.50	2.10
X/V	14.36	21.69	11.55	6.66
DURATION	317	334	334	248
PANEL B: STOCK CHARACTERISTICS				
$V \cdot P$	1.27	0.48	3.83	8.59
σ	0.022	0.024	0.019	0.023
W	30.5	11.8	80.7	194.9
#Obs.	14,182	5,751	5,616	2,815

TABLE 3.2. TABLE DESCRIBES THE SDC PLATINUM SAMPLE OF SHARE REPURCHASE PROGRAMS OVER THE PERIOD FROM JANUARY 1985 TO JANUARY 2014 AND THREE SUB-PERIODS FROM 1985 TO 2000, FROM 2001 TO 2007, AND FROM 2008 TO 2014. PANEL A REPORTS CHARACTERISTICS OF SHARE REPURCHASE PROGRAMS, SUCH AS THE MEDIAN AUTHORISED SIZE OF SHARE REPURCHASE PROGRAM $\bar{X}$ (IN MILLIONS OF SHARES AND AS A FRACTION OF DAILY VOLUME), THE MEDIAN REALISED SIZE OF SHARE REPURCHASE PROGRAM X (IN MILLIONS OF SHARES AND AS A FRACTION OF DAILY VOLUME), AND THE MEDIAN DURATION OF REPURCHASE PROGRAM (IN DAYS).PANEL B REPORTS CHARACTERISTICS OF THE REPURCHASED STOCK, SUCH AS THE MEDIAN AVERAGE DAILY DOLLAR VOLUME (IN MILLIONS OF DOLLARS), THE MEDIAN DAILY VOLATILITY, AND THE MEDIAN EXPECTED TRADING ACTIVITY (IN THOUSANDS OF DOLLARS).

Table 3.3: Descriptive statistics by repurchase method.

	$\bar{X}$	X	Duration	#Obs.
Open Market	1.58	1.00	351	12,965
Negotiated	2.00	1.19	327	6,730
Dutch Auction	3.55	2.29	34	342
Accelerated	6.15	6.54	66	162
First Price	3.49	1.85	37	164
Odd Lot	1.59	1.00	41	272

Table 3.3. Table presents characteristics of the SDC Platinum share repurchase programs over the period from January 1985 to January 2014 for different share repurchase methods. The median authorised size of share repurchase program $\bar{X}$ (in millions of shares), the median realised size of share repurchase program $\bar{X}$ (in millions of shares), and the median duration of repurchase program (in days) are reported for different types of share repurchases.

Figure 3.2: Sizes of share repurchase programs versus trading activity of stocks

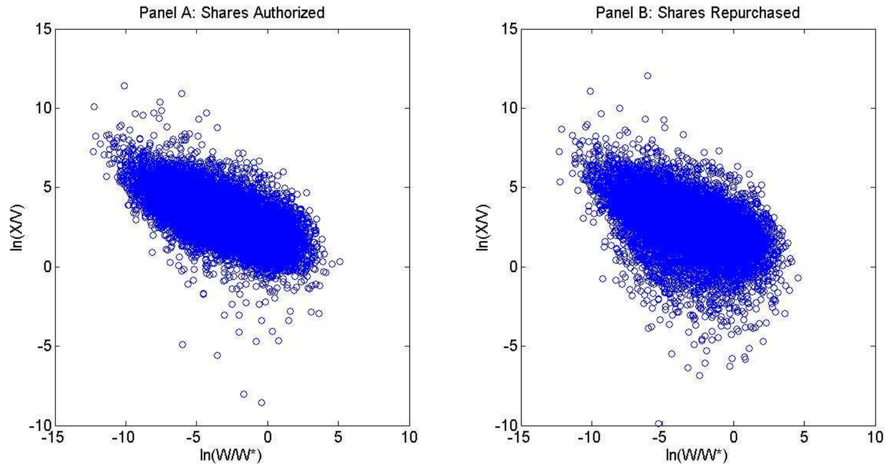


FIGURE 3.2 VISUALIZES RELATIONSHIP BETWEEN THE LOGARITHM OF THE SIZE OF SHARE REPURCHASE PROGRAM AND TRADING ACTIVITY OF STOCKS. PANEL A DISPLAYS RELATION BETWEEN THE LOGARITHM OF AUTHORISED SIZE OF SHARE REPURCHASE PROGRAM $\ln(\bar{X})$ ON THE Y-AXIS AND LOGARITHM OF TRADING ACTIVITY OF THE STOCK $\ln(W)$. PANEL B DISPLAYS RELATION BETWEEN THE LOGARITHM OF THE REALISED SIZE OF SHARE REPURCHASE PROGRAM $\ln(X)$ ON THE Y-AXIS AND LOGARITHM OF TRADING ACTIVITY OF THE STOCK $\ln(W)$.

Table 3.4: Size of repurchase program and trading activity.

	All	1985-2000	2001-2007	2008-2014
PANEL A: Authorised size				
$\hat{\alpha}_0$	2.014***	2.22***	1.95***	1.883***
	(0.037)	(0.044)	(0.017)	(0.05)
$\hat{\alpha}_W$	-0.329***	-0.303***	-0.327***	-0.332***
	(0.005)	(0.008)	(0.007)	(0.017)
R^2	40%	28%	45%	45%
# Obs.	14,200	6,703	4,677	2,820
PANEL B: Realised size				
$\hat{\alpha}_0$	1.595***	1.768***	1.624***	1.314***
	(0.052)	(0.064)	(0.061)	(0.074)
$\hat{\alpha}_W$	-0.311***	-0.306***	-0.273***	-0.289***
	(0.007)	(0.010)	(0.011)	(0.010)
R^2	26%	20%	23%	24%
# Obs.	11,884	6,018	3,949	1,917

Table 3.4 presents estimation of regression: $\ln\left[\frac{X_{it}}{V_{it}}\right] = \alpha_0 + \alpha_W \cdot \ln\left[\frac{W_{it}}{W^*}\right] + \tilde{\epsilon}_{it}$ on total sample of share repurchase programs and sub-samples corresponding to periods 1985-2000, 2001-2007, and 2008-2014. Panel A reports estimation for the authorised size of share repurchase program. Panel B reports estimation for the realised size of share repurchase program. Stock has expected daily volume V_{it}, expected price level P_{it}, expected daily return volatility σ_{it}, expected trading activity, W_{it}. The benchmark stock has expected trading activity W^*. I report estimates $\hat{\alpha}_0$ and $\hat{\alpha}_W$ along with standard errors that are clustered at industry group, regression R^2 and number of observations. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively.

Table 3.5: Size of repurchase program and trading activity.

	All	1985-2000	2001-2007	2008-2014
PANEL A: Authorised size				
$\hat{\alpha}_0$	2.273***	3.033***	0.686***	2.643***
	(0.067)	(0.09)	(0.000)	(0.320)
$\hat{\alpha}_W$	-0.335***	-0.321***	-0.333***	-0.358***
	(0.007)	(0.010)	(0.008)	(0.014)
FE	Yes	Yes	Yes	Yes
R^2	44%	33%	48%	49%
# Obs.	14,200	6,703	3,949	1,917
PANEL B: Realised size				
$\hat{\alpha}_0$	1.478***	4.33***	1.007***	2.037***
	(0.121)	(0.323)	(0.075)	(0.160)
$\hat{\alpha}_W$	-0.293***	-0.302***	-0.269***	-0.306***
	(0.007)	(0.013)	(0.011)	(0.011)
FE	Yes	Yes	Yes	Yes
R^2	30%	25%	28%	29%
# Obs.	11,884	6,018	3,949	1,917

Table 3.5 presents estimation of regression: $\ln\left[\frac{X_{it}}{V_{it}}\right] = \alpha_0 + \alpha_W \cdot \ln\left[\frac{W_{it}}{W^*}\right] + Z_{it} + \tilde{\epsilon}_{it}$ on total SDC share repurchase sample and sub-samples that correspond to periods from 1985 to 2000, from 2001 to 2007 and from 2008 to 2014. Panel A reports estimation for the authorised size of share repurchase program. Panel B reports estimation for the realised size of share repurchase program. Stock has expected daily volume V_{it}, expected price level P_{it}, expected daily return volatility σ_{it}, expected trading activity, W_{it}. The benchmark stock has expected daily volume of 1 million shares, expected price level \$40, expected daily return volatility of 2%, and expected trading activity W^*. Regression includes time and industry fixed effects Z_{it}. I report estimates $\hat{\alpha}_0$ and $\hat{\alpha}_W$ and standard errors that are clustered by industry group, regression R^2 and number of observations. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively.

Figure 3.3: Size of repurchase program and trading activity (yearly).

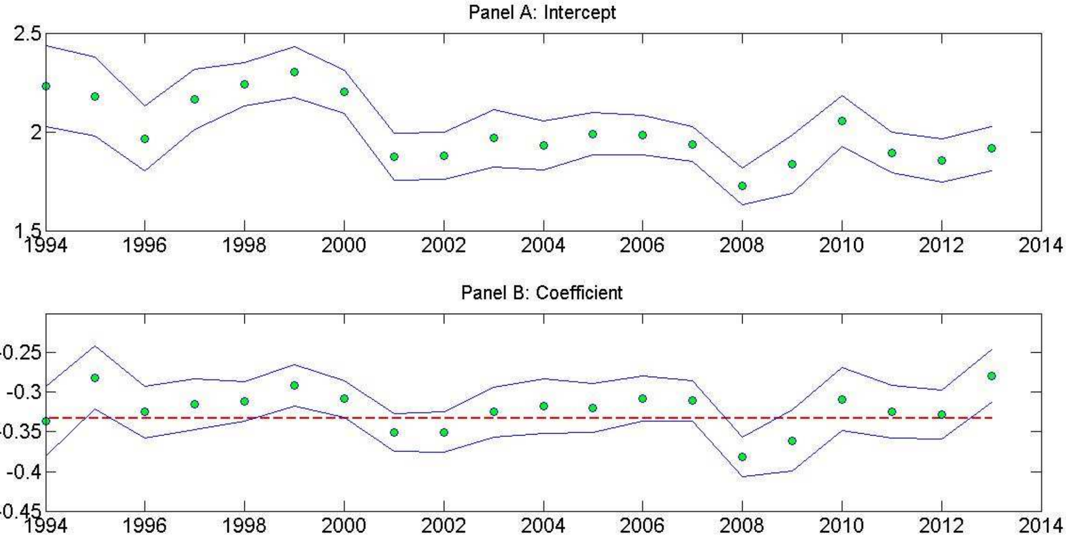


Figure 3.3 presents year by year estimation of regression: $\ln\left[\frac{X_{it}}{V_{it}}\right] = \alpha_0 + \alpha_W \cdot \ln\left[\frac{W_{it}}{W^*}\right] + Z_{it} + \tilde{\epsilon}_{it}$ on total SDC share repurchase sample over the period from 1994 to 2014. Where X_{it} is an authorised size of a repurchase program and V_{it}, P_{it}, σ_{it}, and W_{it} are expected daily trading volume, expected price level, expected daily return volatility, and expected trading activity of the stock, respectively. The benchmark stock has expected daily volume of 1 million shares, expected price level \$40, expected daily return volatility of 2%, and expected trading activity W^*. Panel A displays year by year variation of estimates $\hat{\alpha}_W$ (green circles) with a 95% confidence interval. Red dashed line displays the -1/3 level. Panel B displays year by year estimates $\hat{\alpha}_0$ (green circles) with a 95% confidence interval.

Table 3.6: Nested regression.

	All	1985-2000	2001-2007	2008-2014
PANEL A: Authorised size				
$\hat{\alpha}_0$	0.888***	0.844***	0.971***	1.176***
	(0.082)	(0.172)	(0.084)	(0.092)
$\hat{\alpha}_W$	-0.333***	-0.316***	-0.326***	-0.348***
	(0.013)	(0.015)	(0.012)	(0.012)
$\hat{\alpha}_\tau$	0.217***	0.254***	0.193***	0.147***
	(0.013)	(0.026)	(0.010)	(0.021)
$\hat{\alpha}_\sigma$	-0.210**	-0.287***	-0.186**	-0.144
	(0.094)	(0.094)	(0.105)	(0.087)
R^2	47%	37%	50%	49%
# Obs.	12,351	5,367	5,010	1,974
PANEL B: Realised size				
$\hat{\alpha}_0$	0.328***	0.434***	0.382***	0.579***
	(0.082)	(0.169)	(0.114)	(0.149)
$\hat{\alpha}_W$	-0.318***	-0.313***	-0.283***	-0.303***
	(0.007)	(0.013)	(0.011)	(0.018)
$\hat{\alpha}_\tau$	0.239***	0.259***	0.231***	0.145***
	(0.011)	(0.024)	(0.016)	(0.026)
$\hat{\alpha}_\sigma$	-0.492***	-0.573***	-0.497***	-0.399***
	(0.107)	(0.110)	(0.129)	(0.084)
R^2	33%	29%	33%	28%
# Obs.	11,884	5,187	4,780	1,917

Table 3.6 presents estimation of regression model presents estimation of regression: $\ln\left[\frac{X_{it}}{V_{it}}\right] = \alpha_0 + \alpha_W \cdot \ln\left[\frac{W_{it}}{W^*}\right] + \alpha_\tau \cdot \ln(T_{it}) + \alpha_\sigma \cdot \ln(\sigma_{it}) + Z_{it} + \tilde{\epsilon}_{it}$ on total SDC share repurchase sample and sub-samples that correspond to periods from 1985 to 2000, from 2001 to 2007 and from 2008 to 2014. Panel A presents estimation for authorised size of share repurchase program. Panel B presents results for total number of shares repurchased in the program. Stock has expected daily volume V_{it}, expected price level P_{it}, expected daily return volatility σ_{it}, and expected trading activity, W_{it}. The benchmark stock has expected daily volume of 1 million shares, expected price level $40, expected daily return volatility of 2%, and expected trading activity W^*. Regression includes time and industry fixed effects Z_{it}. I report estimates $\hat{\alpha}_0$, $\hat{\alpha}_W$, $\hat{\alpha}_\tau$ and $\hat{\alpha}_\sigma$, standard errors that are clustered by industry group, regression R^2 and number of observations. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively.

Table 3.7: Nested regression across repurchase types.

	Open market	Negotiated	Dutch auction	Accelerated	First price	Odd Lot
PANEL A: Authorised size						
$\hat{\alpha}_0$	-1.105***	0.347	1.731***	0.807	2.919***	3.991***
	(0.180)	(0.123)	(0.502)	(0.531)	(0.445)	(0.761)
$\hat{\alpha}_W$	-0.328***	-0.358***	-0.419***	-0.396***	-0.412***	-0.374***
	(0.006)	(0.008)	(0.024)	(0.048)	(0.031)	(0.039)
$\hat{\alpha}_\tau$	0.403***	0.177***	0.062	0.010	0.152	-0.096
	(0.021)	(0.018)	(0.058)	(0.026)	(0.094)	(0.118)
$\hat{\alpha}_\sigma$	-0.216***	-0.182***	-0.123	-0.218	0.238	0.420***
	(0.027)	(0.029)	(0.090)	(0.125)	(0.121)	(0.188)
R^2	54%	50%	63%	28%	62%	26%
# Obs.	10,719	5,614	307	113	137	203
PANEL B: Realised size						
$\hat{\alpha}_0$	-1.613***	0.592***	1.939***	1.621***	1.995***	1.754***
	(0.191)	(0.080)	(0.314)	(0.252)	(0.433)	(0.554)
$\hat{\alpha}_W$	-0.309***	-0.342***	-0.428***	-0.279***	-0.418***	-0.437***
	(0.011)	(0.009)	(0.024)	(0.160)	(0.048)	(0.050)
$\hat{\alpha}_\tau$	0.557***	0.179***	0.022	0.029	0.092	-0.126
	(0.037)	(0.014)	(0.080)	(0.058)	(0.109)	(0.151)
$\hat{\alpha}_\sigma$	-0.523***	-0.438***	-0.150	-0.404	0.103	0.134
	(0.077)	(0.101)	(0.112)	(0.236)	(0.225)	(0.218)
R^2	43%	34%	43%	15%	44%	22%
# Obs.	10,719	5,431	306	107	125	197

Table 3.7 presents estimation of regression model presents estimation of regression $\ln\left[\frac{X_{it}}{V_{it}}\right] = \alpha_0 + \alpha_W \cdot \ln\left[\frac{W_{it}}{W^*}\right] + \alpha_\tau \cdot \ln(T_{it}) + \alpha_\sigma \cdot \ln(\sigma_{it}) + Z_{it} + \tilde{\epsilon}_{it}$ for different types of share repurchases, such as open market, negotiated, dutch auction, first price, and odd lot repurchases in SDC Platinum database over the period from 1985 to 2014. Panel A presents estimation for authorised size of share repurchase program. Panel B presents results for total number of shares repurchased in the program. Stock has expected daily volume V_{it}, expected price level P_{it}, expected daily return volatility σ_{it}, and expected trading activity, W_{it}. The benchmark stock has expected daily volume of 1 million shares, expected price level $40, expected daily return volatility of 2%, and expected trading activity W^*. Regression includes time and industry fixed effects Z_{it}. I report estimates $\hat{\alpha}_0$, $\hat{\alpha}_W$, $\hat{\alpha}_\tau$ and $\hat{\alpha}_\sigma$, standard errors that are clustered by industry group, regression R^2, and number of observations. Statistical significance at the 1%, 5% and 10% levels is denoted by *, **, and *** respectively.

Table 3.8: Model selection.

	Target size	Target imbalance	Target cost	
			Linear	Square root
All	44,299	**38,433**	39,588	41,282
Open Market	40,031	**30,819**	35,328	41,282
Negotiated	19,691	17,927	**17,866**	20,918
Dutch Auction	941	**852**	914	1,107
Accelerated	**313**	408	**319**	374
First Price	466	**421**	473	556
Odd Lot	877	**859**	877	942

Table 3.8 reports the results of model selection according to bayesian information criteria for different types of repurchase programs, such as open market, negotiated, dutch auction, accelerated, first price, and odd lot repurchase programs. Considered models are bet hypothesis, targeted imbalance hypothesis, and two versions of targeted cost hypothesis. Bayesian information criteria $BIC = -2\ln\hat{L} + k \cdot \ln(n)$, where $\hat{L}$ is likelihood, k is number of parameters in a model and n is number of observations. Best fitted model is highlighted in bold.

Table 3.9: Alternative hypotheses.

	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: UNRESTRICTED SPECIFICATION						
$\hat{\alpha}_0$	2.200***	5.011***	2.906***	3.248***	2.341***	1.003***
	(0.075)	(0.115)	(0.067)	(0.092)	(0.095)	(0.195)
$\hat{\alpha}_W$	-0.364***	-0.349***	-0.337***	-0.331***	-0.334***	-0.373***
	(0.008)	(0.010)	(0.009)	(0.012)	(0.008)	(0.010)
$\hat{\alpha}_\sigma$	-0.317***	—	—	—	—	-0.19
	(0.095)					(0.115)
$\hat{\alpha}_\tau$	—	0.208***	—	—	—	0.216***
		(0.012)				(0.012)
$\hat{\alpha}_{3m}$	—	—	-0.196	—	—	-0.128
			(0.105)			(0.104)
$\hat{\alpha}_\$$	—	—	—	0.392	—	0.018
				(0.390)		(0.309)
$\hat{\alpha}_M$	—	—	—	—	0.015***	0.011***
					(0.002)	(0.002)
FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	45%	49%	44%	45%	46%	53%
#Obs.	14,200	12,351	13,844	8,561	14,200	7,227
PANEL B: RESTRICTED SPECIFICATION						
FE	No	No	No	No	No	No
R^2	40%	40%	41%	41%	40%	41%

TABLE 3.9 PRESENTS ESTIMATION OF REGRESSION THAT IS MOTIVATED BY ALTERNATIVE HYPOTHESES ON SHARE REPURCHASE $\ln\left[\frac{X_{it}}{V_{it}}\right] = \alpha_0 + \alpha_W \cdot \ln\left[\frac{W_{it}}{W^*}\right] + \alpha_{3m} \cdot R_{it}^{3m} + \alpha_\$ \cdot C_{it-1} + \alpha_M \cdot M_{it} + \alpha_\tau \cdot \ln\left[\tau_{it}\right] + \alpha_\sigma \cdot \ln\left[\frac{\sigma_{it}}{\sigma^*}\right] + \tilde{\epsilon}_{it}$ ON A SAMPLE OF SHARE REPURCHASE PROGRAMS FROM SDC PLATINUM THAT COVERS PERIOD FROM 1985 TO 2014. X_{it} IS AN AUTHORISED SIZE OF THE SHARE REPURCHASE PROGRAM WITH THE STOCK THAT HAS EXPECTED DAILY TRADING VOLUME V_{it}, EXPECTED PRICE LEVEL P_{it}, EXPECTED DAILY RETURN VOLATILITY σ_{it}, AND EXPECTED TRADING ACTIVITY, W_{it}. R_{it}^{3m} IS LOG-RETURN DURING 3 MONTHS PRIOR TO THE REPURCHASE ANNOUNCEMENT. C_{it-1} IS CASH OF THE COMPANY FOR THE QUARTER BEFORE THE REPURCHASE ANNOUNCEMENT FROM COMPUSTAT. M_{it} IS MARKET CAPITALIZATION OF THE COMPANY. T_{it} IS DURATION OF THE REPURCHASE PROGRAM (NUMBER OF DAYS BETWEEN ANNOUNCEMENT DATE AND END OF REPURCHASE PROGRAM DATE). THE BENCHMARK STOCK HAS EXPECTED DAILY VOLUME OF 1 MILLION SHARES, EXPECTED PRICE LEVEL \$40, EXPECTED DAILY RETURN VOLATILITY OF 2%, AND EXPECTED TRADING ACTIVITY W^*. PANEL A REPORTS ESTIMATES OF THE REGRESSION WITH STANDARD ERRORS CLUSTERED BY INDUSTRY, REGRESSION R^2 AND NUMBER OF OBSERVATIONS. PANEL B REPORTS R^2 OF REGRESSION WITH IMPOSED CONSTRAINT OF $\alpha_W = -1/3$ AND ZERO COEFFICIENTS ON CORRESPONDING CONTROL VARIABLES.

Bibliography

Agarwal, Vikas, and Sugata Ray. 2012. “Determinants and Implications of Fee Changes in the Hedge Fund Industry.” *Working Paper.*

Agarwal, Vikas, Kevin Mullally, and Narayan Naik. 2015. “Hedge Funds: A survey of Academic Literature.” *Foundations and Trends in Finance.*

Agarwal, Vikas, Naveen Daniel, and Narayan Naik. 2004. “Flows, Performance and Managerial Incentives in Hedge Funds.” *Working Paper.*

Agarwal, Vikas, Vikram Nanda, and Sugata Ray. 2013. “Institutional Investment and Intermediation in the Hedge Fund Industry.” *Working Paper.*

Aiken, Adam L., Christopher P. Clifford, and Jesse A. Ellis. 2015. “Hedge Funds and Discretionary Liquidity Restrictions.” *Journal of Financial Economics*, 197–218.

Almgren, Robert, and Neil Chriss. 2000. “Optimal Execution of Portfolio Transactions.” *Journal of Risk*, 3(2): 5–39.

Andersen, Torben G., Oleg Bondarenko, Albert S. Kyle, and Anna A. Obizhaeva. 2015. “Intraday Trading Invariance in the E-Mini S&P 500 Futures Market.” Working Paper. Available at http://dx.doi.org/10.2139/ssrn.2693810.

Andersen, Torben, Oleg Bondarenko, Albert S. Kyle, and Anna A. Obizhaeva. 2014. “An Invariance Relationship in Intraday Trading Patterns of E-mini SnP 500 Futures.” Working Paper, Northwestern University, University of Illinois, University of Maryland, New Economic School.

Ané, Thierry, and Hťelyette Geman. 2000. "Order Flow, Transaction Clock, and Normality of Asset Returns." *The Journal of Finance*, 55(5): 2259Ű–2284.

Angel, James J, Lawrence E Harris, and Chester S Spatt. 2015. "Equity Trading in the 21st Century: An Update." *The Quarterly Journal of Finance*, 5(01): 1550002.

Bae, Kyoung hun, Albert S. Kyle, Eun Jung Lee, and Anna A. Obizhaeva. 2014. "Invariance Relationship in the Number of Buy-Sell Switching Points in the South Korean Stock Market." Working Paper, University of Maryland, Hanyang University, New Economic School.

Bagwell, Laurie, and John Shoven. 1988. "Share Repurchases and Acquisitions: An Analysis of Which Firms Participate." *National Bureau of Economic Research*, Corporate Takeovers: Causes and Consequences: 191–220.

Baquero, Guillermo, and Marno Verbeek. 2009. "A Portrait of Hedge Fund Investors: Flows, Performance and Smart Money." *Working paper.*

Baquero, Guillermo, and Marno Verbeek. 2015. "Hedge Fund Flows and Performance Streaks: How Investors Weigh Information." *ESMT Working Paper*, , (15-01).

Bergstresser, Daniel, John M. R. Chalmers, and Peter Tufano. 2009. "Assessing the Costs and Benefits of Brokers in the Mutual Fund Industry." *The Review of Financial Studies*, 22(10): 4129–4156.

Berk, Jonathan B., and Jules van Binsbergen. 2013. "Measuring Skill in the Mutual Fund Industry." *Working Paper.*

Berk, Jonathan B., and Richard C. Green. 2004. "Mutual Fund Flows and Performance in Rational Markets." *Journal of Political Economy*, 12(6).

Bertsimas, Dimitris, and Andrew Lo. 1998. "Optimal Control of Execution Costs." *Journal of Financial Markets*, 1(1): 1–50.

Booth, James R., and Richard L. Smith. 1986. "Capital raising, underwriting and the certification hypothesis." *Journal of Financial Economics*, 15: 261–281.

Brav, Alon, John R. Graham, Campbell R. Harvey, and Roni Michaely. 2005. "Payout Policy in the 21 Century." *Journal of Financial Economics*.

Brooks, Chris, Andrew Clare, and Nick Motson. 2007. "The Gross Truth About Hedge Fund Performance and Risk: The Impact of Incentive Fees." *Working Paper*.

Brown, Stephen, William Goetzmann, Bing Liang, and Christopher Schwarz. 2008. "Mandatory Disclosure and Operational Risk: Evidence from Hedge Fund Registration." *The Journal of Finance*, 63(6): 2785–2815.

Buffa, Andrea M., and Giovanna Nicodano. 2008. "Should Insider Trading be Prohibited When Share Repurchases are Allowed?" *Review of Finance*, 12(4): 735Ű765.

Campbell, John Y., and Albert S. Kyle. 1993. "Smart Money, Noise Trading, and Stock Price Behavior." *The Review of Economic Studies*, 60: 1–Ű34.

Chetty, Raj, and Emmanuel Saez. 2006. "The Effects of the 2003 Dividend Tax Cut on Corporate Behavior: Interpreting the Evidence." *The American economic review*, 124Ű–129.

Chevalier, Judith, and Glenn Ellison. 1997. "Risk Taking by Hedge Funds as a Response to Incentives." *The Journal of Political Economy*, 105(6): 1167–1200.

Christoffersen, Susan E.K., Richard Evans, and David K. Musto. 2013. "What Do Consumers' Fund Flows Maximize? Evidence from Their Brokers' Incentives." *The Journal of Finance*, 68: 201Ű235.

Clark, Peter. 1973. "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices." *Econometrica*, 41: 135Ű–155.

Cont, Rama, and Jean-Philipe Bouchaud. 2000. “Herd Behavior and Aggregate Fluctuations in Financial Markets.” *Macroeconomic Dynamics*, 4: 170–196.

Cont, Rama, Sasha Stoikov, and Rishi Talreja. 2010. “A Stochastic Model for Order Book Dynamics,.” *Operations Research*, 58(3): 549–563.

Del Guercio, Diane, and Jonathan Reuter. 2014. “Mutual Fund Performance and the Incentive to Generate Alpha.” *Journal of Finance*, 1673–1704.

Deuskar, Prachi Z., Jay Wang, Youchang Wu, and Quoc H. Nguyen. 2011. “The Dynamics of Hedge Fund Fees.” *Working Paper.*

Dittmar, Amy. 2000. “Why Do Firms Repurchase Stocks?” *The Journal of Business*, 73(3): 331Ű–355.

Duffie, Darrell. 2010. “Presidential Address: Asset Price Dynamics with Slow-Moving Capital.” *The Journal of Finance*, 65(4).

Edwards, Franklin R., and James M. Park. 1996. “Do Managed Futures Make Good Investments?” *The Journal of Futures Markets*, 475–517.

Farmer, Doyne J., Paolo Patelli, and Ilija I. Zovko. 2005. “The Predictive Power of Zero Intelligence in Financial Markets.” *Proceedings of the National Academy of Sciences of the USA*, 102(11): 2254–2259.

Fischer, Lawrence. 1963. “Use of electronic computers in the quality control of financial data.” *Presentation at the American Statistical Association.*

Fung, William, and David A. Hsieh. 2004. “Hedge Fund Benchmarks: A Risk Based Approach.”

Fung, William, David A. Hsieh, Narayan Y. Naik, and Tarun Ramadorai. 2008. “Hedge Funds: Performance, Risk, and Capital Formation.” *The Journal of Finance*, LXIII(4): 1777–1803.

Gabaix, Xavier, Parameswaran Gopikrishnan, Vasiliki Plerou, and Eugene H. Stanley. 2006. "Institutional Investors and Stock Market Volatility." *The Quarterly Journal of Economics*, 121: 461Ű–504.

Garella, Paolo. 1989. "Adverse Selection and the Middleman." *Economica*, 56: 395–400.

Garleanu, Nicolae, and Lasse Heje Pedersen. 2016. "Efficiently Inefficient Markets for Assets and Asset Managers." Working Paper.

Garvey, Ryan, Tao Huang, and Fei Wu. 2017. "Why Do Traders Split Orders?" *Financial Review*, 52(2): 233–258.

Gatheral, Jim, and Alex Schied. 2013. "Dynamical models for market impact and algorithms for optimal order execution." *Handbook on Systemic Risk (eds.: J.-P. Fouque and J. Langsam)*, 579–602.

Getmansky, Mila. 2002. "The Life Cycle of Hedge Funds: Fund Flows, Size, Competition, and Performance." *Quarterly Journal of Finance*, 2(1).

Getmansky, Mila, Bing Liang, Chris Schwarz, and Russ Wermers. 2015. "Share Restrictions and Investor Flows in the Hedge Fund Industry."

Getmansky, Mila, Peter A. Lee, and Andrew W. Lo. 2015. "Hedge Funds: A Dynamic Industry in Transition." *Working Paper.*

Goetzmann, William, Jonathan Ingersoll, and Stephen Ross. 2003. "High-Water Marks and Hedge Fund Management Contracts." *Journal of Finance*, 58.

Grullon, Gustavo, and Roni Michaely. 2002. "Dividends, share repurchases, and the substitution hypothesis." *The Journal of Finance*, 57: 1649–1684.

Hasbrouck, Joel. 1999. "Trading Fast and Slow: Security Market Events in Real Time." Working Paper, New York University.

Hodder, James E., Jens Carsten Jackwerth, and Olga Kolokolova. 2012. "Recovering Delisting Returns of Hedge Funds." *Working Paper Series.*

Horst, Jenke ter, and Galla Salganik-Shoshan. 2014. "Style Chasing by Hedge Fund Investors." *Journal of Banking and Finance*, 29: 29–42.

Ikenberry, David, Josef Lakonishok, and Theo Vermaelen. 1995. "Market Underreaction to Open Market Share Repurchases." *Journal of Financial Economics*, 39(4): 181–208.

Jensen, Michael. 1986. "Agency Cost Of Free Cash Flow, Corporate Finance, and Takeovers." *American Economic Review*, 76.

Jensen, Michael C., and William H. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics*, 3.

Joenväärä, Juha, Robert Kosowski, and Pekka Tolonen. 2013. "The Effect of Investment Constraints on Hedge Fund Investor Returns."

Joenväärä, Juha, Robert Kosowski, and Pekka Tolonen. 2014. "Hedge Fund Performance: What Do We Know?" *Working Paper.*

Jones, Charles M., Gautam Kaul, and Marc L. Lipson. 1994. "Transactions, Volume, and Volatility." *The Review of Financial Studies*, 7(4): 631–651.

Jorion, Philippe, and Christopher Schwarz. 2015. "Who are the smart investors in the room? Evidence from U.S. Hedge Funds Soliciation." *Working paper.*

Judge, Kathryn. 2014. "Intermediary Influence." *University of Chicago Law Review.*

Kirilenko, Andrei, Albert S. Kyle, Tugkan Tuzun, and Mehrdat Samadi. 2010. "The Flash Crash: High-Frequency Trading in an Electronic Market." Working Paper, University of Maryland.

Kirilenko, Andrei, Albert S. Kyle, Tugkan Tuzun, and Mehrdat Samadi. 2017. "The Flash Crash: High-Frequency Trading in an Electronic Market." *Journal of Finance*, 72(3): 967–998.

Kolokolova, Olga. 2010. "Strategic Behavior within Families of Hedge Funds." *Working Paper.*

Kyle, Albert S. 1985. "Continuous Auctions and Insider Trading." *Econometrica*, 53(6): 1315–1335.

Kyle, Albert S., and Anna A. Obizhaeva. 2017*a*. "Dimensional Analysis, Leverage Neutrality, and Market Microstructure Invariants." Working Paper, University of Maryland and New Economic School.

Kyle, Albert S., and Anna A. Obizhaeva. 2017*b*. "Large Bets and Stock Market Crashes." Available at SSRN: https://ssrn.com/abstract=2023776 or http://dx.doi.org/10.2139/ssrn.2023776.

Kyle, Albert S., and Anna A. Obizhaeva. 2017*c*. "Market Microstructure Invariance: A Dynamic Equilibrium Model." Working Paper, University of Maryland and New Economic School.

Kyle, Albert S., and Anna Obizhaeva. 2016. "Market Microstructure Invariance: Empirical Hypotheses." *Econometrica*, 84(4): 1345Ű1404.

Kyle, Albert S., Anna A. Obizhaeva, and Tugkan Tuzun. 2016. "Microstructure Invariance in U.S. Stock Market Trades." Working Paper, University of Maryland.

Kyle, Albert S., Anna A. Obizhaeva, and Yajun Wang. 2017. "Smooth Trading with Overconfidence and Market Power."

Kyle, Albert S., Anna A. Obizhaeva, Nitish Sinha, and Tugkan Tuzun. 2011. "News Articles and Equity Trading." Working Paper, University of Maryland.

Ladley, Dan. 2012. "Zero Intelligence in Economics and Finance." *Agent-Based Computational Economics*, 27(2): 273–286.

Levenshtein, Vladimir I. 1966. "Binary codes capable of correcting deletions, insertions, and reversals." *Soviet Physics Doklady*, 10(8): 707–710.

Lu, Yan, David Musto, and Sugata Ray. 2013. "Alternative marketing for alternative investments." *The Journal of Finance*.

Mandelbrot, Benoît. 1963. "The Variation of Certain Speculative Prices." *The Journal of Business*, 36(4): 394–419.

Mandelbrot, Benoît, and Howard M. Taylor. 1967. "On the Distribution of Stock Price Differences." *Operations Research*, 15(6): 1057Ű–1062.

McDonnell, Tony. 2003. "Performance Fee Equalisation." *AIMA Journal*.

Mitchell, Mark L., and Erik Stafford. 2000. "Managerial Decisions and Long-Term Stock Price Performance." *Journal of Business*, 73: 287–329.

Modigliani, Franco, and Merton H. Miller. 1958. "The Cost of Capital, Corporation Finance and the Theory of Investment." *American Economic Review*, 48: 261–297.

Nanda, Vikram, M.P. Narayanan, and Vincent A. Warther. 2000. "Liquidity, investment ability, and mutual fund structure." *Journal of Financial Economics*, 57: 417–443.

Newey, Whitney K., and Kenneth D. West. 1987. "A Simple, Positive Semi-Ddefinite, Hetoroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55: 703–708.

Obizhaeva, Anna A. 2012. "Liquidity Estimates and Selection Bias." Working Paper. Available at https://ssrn.com/abstract=1178722 or http://dx.doi.org/10.2139/ssrn.1178722.

Obizhaeva, Anna A., and Jiang Wang. 2013. "Optimal Trading Strategy and Supply/Demand Dynamics." *Journal of Financial Markets*, 16(1): 1–32.

Opler, Tim, and Sheridan Titman. 1996. "The Debt-Equity Choice: Ananalysis of Issuing Firms." *Working Paper.*

Pastor, Lubos, and Robert Stambaugh. 2012. "On the Size of the Active Management Industry." *Journal of Political Economy*, 120(4): 740–781.

Reuter, Johnathan. 2015. "Revisiting the Performance of Broker-Sold Mutual Funds." *Working Paper.*

Rubinstein, Ariel, and Asher Wolinsky. 1987. "Middlemen." *The Quarterly Journal of Economics*, 102: 581–593.

Schied, Alex, and Torsten Schoeneborn. 2009. "Risk aversion and the dynamics of optimal trading strategies in illiquid markets." *Finance and Stochastics*, 13(2): 181–204.

Schwarz, Christopher. 2007. "Hedge Fund Fees." *Working Paper.*

Sirri, Erik R., and Peter Tufano. 1998. "Costly Search and Mutual Fund Flows." *The Journal of Finance*, LIII(5).

Spulber, Daniel F. 2001. "Market Microstructure: Intermediaries and the Theory of the Firm." *Working Paper.*

Staffs of the CFTC and SEC. 2010b. *Findings Regarding the Market Events of May 6, 2010.* Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues. September 30, 2010. "Final Report".

Stephens, Clifford, and Michael Weisbach. 1998. "Actual Share Reacquisitions and Open-Market Repurchase Programs." *The Journal of Finance*, 53.

Stoughton, Neal M., Youchang Wu, and Josef Zechner. 2011. "Intermediated Investment Management." *The Journal of Finance*, LXVI(3): 947–980.

Vayanos, Dimitri. 2004. "Flight to Quality, Flight to Liquidity, and the Pricing of Risk."

Vayanos, Dimitri, and Paul Woolley. 2013. "An Institutional Theory of Momentum and Reversal." *The Review of Financial Studies*, 26(5): 1087Ű1145.

Vermaelen, Theo. 1981. "Common stock repurchases and market signaling." *Journal of Financial Economics*, 9: 139–183.

Vermaelen, Theo. 2000. "Share Repurchases." *Foundations and Trends in Finance.*