# Essays in High and Low Frequency Price Discovery

by

Saad Ali Khan

A thesis submitted to the
Department of Economics
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen's University
Kingston, Ontario, Canada
April 2020

# Abstract

In this thesis, I examine high and low frequency price discovery for US equity markets. In Chapter 3, I propose a methodology to disentangle high frequency price jumps into a permanent and transitory component. Using a variance decomposition and realized estimates I show that while jumps are extremely rare events, jump contribution to intraday price discovery is large. In Chapter 4, co-authored with Ryan Riordan, we examine the market dynamics of price jumps. We find that jumps have a predictable component which is captured by the degree of fragmentation in liquidity in minutes leading up to price jumps. Applying the methodology of Chapter 3, we show that fragmentation predicts noisier jumps. Lastly, in Chapter 5, co-authored with Evan Dudley, Luke Phelps and Ryan Riordan, we use mutual fund fire sales to decompose low frequency quarterly stock prices into an efficient and noise component. We show that more than a quarter of variance in low frequency stock prices is attributable to noise. Overall, this thesis finds that noise is an important component of both high and low frequency stock price variation.

# Acknowledgments

I express my sincere appreciation to my supervisors Prof. Ryan Riordan and Prof. Morten Nielsen.

Prof. Ryan Riordan has been instrumental in my training and development as an academic researcher. His guidance and encouragement helped me withstand the most challenging of times. Prof. Ryan Riordan has been a mentor, instructor and friend. I cannot imagine how my research would have been without him.

Prof. Morten Nielsen's support has been invaluable since the very earliest stages of my degree. Under his guidance I developed technical skills without which this thesis would not have been possible. I am grateful to him for always having time for me, even when overseas.

I would also like to thank Prof. James MacKinnon, Prof. Evan Dudley and Prof. Allan Gregory for their support and feedback.

I am grateful to my colleagues Jenny Watt, Line Drapeau and Ludovic Auger for their support and encouragement, as well as, all the participants of the Econ 999 Graduate Seminar Series and the Quantitative Working Group.

Finally, but most importantly, I would like to thank my family. My parents Akbar Ali Khan and Samima Akbar for their unwavering faith in me and my sisters Asra Akbar and Saima Athar for their kindness.

# Co-Authorship

Chapter 4 of this thesis is co-authored with Ryan Riordan of the Smith School of Business at Queen's University.

Chapter 5 of this thesis is co-authored with Evan Dudley, Luke Phelps and Ryan Riordan of the Smith School of Business at Queen's University.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

The two central functions of financial markets is to provide liquidity and facilitate price discovery. While liquidity refers to matching buyers and sellers, price discovery is the process through which asset prices incorporate new information. Since the flow of information is continuous, price discovery is a dynamic process which involves repeated interaction of traders until all participants agree on the fundamental value of the asset.

Intrinsically, price discovery is a latent process. This is due to the fact that information sets and trading motives of individual traders are unobserved. In particular, while the trades of informed traders push prices towards their efficient value, uninformed traders have liquidity needs which are unrelated to asset fundamentals resulting with transitory deviations (i.e. noise) of prices from their efficient value. This interaction is further complicated by liquidity suppliers' who set prices both to manage inventory risk arising from uninformed traders and adverse selection risk arising from informed traders. While this thesis does not identify the unique source of information leading to a particular price move, we identify the mechanisms of price discovery in a signal extraction model. In particular, this thesis shows that high frequency price discovery is largely dominated by discrete price moves (i.e. jumps). Further, this thesis shows that jumps and their realized signal-to-noise ratios contain a predictable component which is captured degree of dispersion in liquidity (i.e. liquidity fragmentation) in minutes prior to jumps. Lastly, we propose a methodology which uses the distress selling of mutual funds (i.e. fire sales) to disentangle the efficient and noise components of prices at low frequencies. More broadly, we study the market dynamic of high and low frequency price discovery.

## 1.1 Contribution

Financial markets have evolved to unprecedented levels over the past two decades. Rapid technological advances have resulted with traditional sources of information such as analyst reports and print news largely replaced by high frequency machine readable news, social media posts, high frequency changes to variables governing the state of electronic limit order books, as well as other sources of information which can be quickly processed by machines and used to make trading decisions at low latencies. These innovation mean that intraday prices evolve as a continuous process interspersed with discrete price moves (i.e. jumps) in a manner foreseen by the seminal work of Press [1967] and Clark [1973]. While continuous price moves are associated with the slow trickle down of information (e.g. Kyle [1985a]; Glosten and Milgrom [1985]), discrete price moves are associated with high frequency information release of the aforementioned type. Understanding the dynamics of jumps is important as they can affect the efficient functioning of markets by impairing portfolio management, risk management and option-pricing (e.g. Bollerslev and Todorov [2011]; Bollerslev, Li, and Todorov. [2016]; Bégin, Dorion, and Gauthier [2019]). Despite their importance, we know little about the importance of jumps for price discovery and even less about the market dynamics of jump price discovery. This thesis finds that despite being rare events, jumps are extremely importance to the price formation process. We show that jumps have a permanent component entangled with transitory (noise) component. Additionally, the relative size of the two component has a predictable component captured by liquidity fragmentation in minutes prior to the jump.

Disentangling the permanent and transitory jump components is important since these two components have distinct implication for investors. For instance, jumps with large transitory component are largely hedgeable simply by holding the portfolio through mean reversion. In contrast, a large permanent jump can result in significant informational losses particularly for liquidity suppliers who trade against jumps. Further, jumps with large transitory components have the potential to stabilize markets (e.g. Flash Crash of 2010).

A second result of technological advances is the degree of fragmentation in equity markets, with quoting and trading regularly occurring on a multitude of exchanges, trading venues, and broker-dealer platforms (O'Hara and Ye [2011]).[1] The fragmented nature of equity markets can lead to increased fragility, a focus of this thesis, and complexity as market participants attempt to coordinate liquidity supply and demand in real-time (Menkveld and Yueshen [2018]). Often the liquidity demand exceeds the liquidity supply and this mismatch can lead to jumps with large transitory component as liquidity demanding orders consume all the

[1] See SEC Public Statement - https://www.sec.gov/news/statement/us-equity-market-structure.html

available liquidity in a market, pushing prices far away from the efficient value. This thesis shows that the dergree of liquidity fragmentation in minutes leading up to price jumps can predict both jumps and direction of price jumps, thereby capturing the information of liquidity suppliers.[2]

Our research also disentangles low frequency prices into their respective noise and efficient components. Price discovery at lower frequencies has important implication for corporate and policy decisions. Extending this insight we decompose information into a part that is knowable at the time of the investment (public information), and a part that is revealed in the future (private information). Using these information components we show that managers do not have private information about changes in efficient value, implying that managers do not possess information not already in prices.

## 1.2 Research Questions

The goal of this thesis is to study the market dynamics of price discovery. One particularity focus this thesis is intraday price jumps and their contribution to intraday price discovery.

The first research question is as follows:

RQI: *How much do jumps contribute to intraday price discovery and intraday noise?*

RQI is addressed in Chapter 3 by proposing a signal extraction methodology which disentangles both variance contributions and realized time paths of intraday jumps and continuous price moves. I also examine the implication of permanent

[2] We defined liquidity supplier as any trader that post a limit order.

and transitory jump components for information cost and contrarian strategy profits of liquidity suppliers.

The second research question is as follows:

RQII: *Can liquidity fragmentation predict jumps, direction of jumps and jump realized signal-to-noise ratio?*

If liquidity suppliers' are better informed than the average investor, then we should expect that liquidity fragmentation should predict jumps, and direction of jumps. In addition, if fragmented liquidity make markets fragile during times leading up to jumps, then higher levels of fragmentation should predict jumps with smaller realized jump signal-to-noise ratio. RQII is addressed in Chapter 4

The third and final question is as follows:

RQIII: *How informative are low frequency stock prices?*

In Chapter 5, we propose a methodology which uses distressed selling of mutual funds (i.e. Fire Sales) to disentangle low frequency quarterly price into efficient component and noise component. Using these two components we show that that corporate managers do not have private information about future changes in efficient value nor can they identify noise in prices.

## 1.3 Organization of Thesis

Chapter 3 presents a general overview of literature related to price jumps and liquidity fragmentation. In Chapters 3 and 4 we examine high frequency price discovery for US equity markets. In Chapter 3 I propose a methodology to disentangle the permanent and transitory component of jumps and continuous price moves. In Chapter 4 we examine the market dynamic of price jumps. In particular, we examine the importantance of intermarket linkages in liquidity for price jumps and jump component of price discovery. In Chapter 5 use mutual fund fires sales to examine low frequency price discovery.

# Chapter 2

# Literature Review

This chapter provides a general overview of the literature on price jumps and liquidity fragmentation. A more focused literature review and how it relates to the specific chapters is provided again in Chapters 3 and 4.

## 2.1 Price Jumps

Since the seminal work of Press [1967] and Clark [1973], it is widely accepted that stock prices are characterized by extreme price changes over relatively short periods of time. These aforementioned low probability and high impact price moves are commonly known as price jumps. One of the main innovation in the econometrics of price jumps is the availability of high frequency price data which greatly facilitates estimation and detection of jumps which relies on the infill asymptotic properties of volatility estimators. Moreover, the inclusion of jumps in standard stochastic volatility model considerably improve model fit for equity prices (e.g. Andersen, Benzoni, and Lund [2002], Andersen, T., and Ørregaard Nielsen [2010])

Previous work has suggested that jumps might be explained by the arrival of new valuation relevant information (Andersen, Bollerslev, Diebold, and Vega [2007]; Jiang and Yao [2013a]; Jeon and Zhao [2019]) and political news (Lobo [1999]), while others have put forward the idea that limited arbitrage is the main cause of large price movements (Mitchell and Pulvino [2012]). Despite the lack of consensus on the source of price jumps, their importance is widely recognized across a wide range of areas in finance, most notably for option pricing and equity risk premium models.

One of the underlying assumption of the celebrated Black and Scholes [1973] model is that markets operate continuously which precludes the presence of jumps. Merton [1976] shows that the continuous price solution of the Black and Scholes model can be viewed as an asymptotic limit to the case where the price process contains discrete jumps. Assuming a log-normal distributed jump size, Merton derives the option price in the presence of jumps as a weighted average of the Black and Scholes solution. More generally, the inclusion of jumps in both price and stochastic volatility processes can significantly improve model performance compared to continuous time models. For example, the inclusion of jumps in option pricing models explain the U-shape of implied volatilities across different strike prices (Bates [1996]; Kou [2002]), further including jumps in both prices and volatility increase implied volatility of in-the-money options (Eraker, Johannes, and Polson [2003]).

Price jumps also have deep implication for the predictability of equity risk premium. In particular jumps can explain several asset pricing anomalies. For instance, Bollerslev et al. [2016] show that the inclusion of systematic jump risk in the capital asset pricing model (CAPM) significantly improves the performance of the CAPM in explaining cross-sectional return variation. Jumps can also explain the pricing of idiosyncratic risk (Bégin et al. [2019]), as well as size and value premiums (Jiang and Yao [2013b]). In addition to price jumps, recent work also examines the effects of jumps in variance risk premium (VRP) as it pertains to return predictability. For example, Bollerslev, Todorov, and Xu [2015] non-parametrically decompose VRP into a continuous and jump component to show that return predictability of the market portfolio exclusively arise from the jump component of VRP. Rombouts, Stentoft, and Violante [2019] also find no evidence to suggest that the continuous component of VRP is priced.

## 2.2 Market Fragmentation

Modern equity markets are highly fragmented with quoting and trading occurring on multiple exchanges, electronic communicating networks (ECNs), dark pools and broker-dealer platforms. For instance, in the United States, stocks listed on NYSE or NASDAQ can trade on BATS, Direct Edge, NYSE Arca and several other trading venues. When a security trades on several venues its price discovery is fragmented across multiple venues, in addition the information available to the demand side of liquidity is also fragmented across multiple limit order books. This has lead to concerns for policy makers such as excessive price dispersion, search cost and adverse selection costs. On the other hand, proponents of fragmentation argue that it encourages competition amongst venues for orderflow which result in lower transaction cost and improved market quality.

There exist a rich literature examining the effects of market fragmentation on market quality. Early theoretical models supported market consolidation. For instance Pagano [1989] argues that liquidity externality leads to all traders gravitating towards a single exchange hence making fragmented markets unstable.

Chowdhry and Nanda [1991] also support consolidation by arguing that adverse selection cost for liquidity suppliers are higher in fragmented markets since informed trader can hide their orders across venues. More recent literature has examined the beneficial effects of fragmentation. Parlour and Seppi [2003] develop a model where a hybrid market competes with a pure limit order book. Foucault and Menkveld [2008] extend this framework to model competition between two pure limit order markets, the London Stock Exchange and Euronext. The authors show that consolidated limit order book is deeper when both markets coexist. O'Hara and Ye [2011] also find that fragmentation in US equity markets lowered transaction cost with faster execution speeds. Bennett and Wei [2006]; Degryse and van Kervel [2015] also find evidence of lower transaction in fragmented markets.

Recent papers which favor consolidation argue that fragmentation can result in a search cost for market participants and therefore adversely effect market quality. Yin [2005] shows that search costs associated with searching for quotes across multiple venues widen spreads and thereby adversely effect liquidity. Madhavan [2012] argues that weaker intermarket linkages resulting from fragmented liquidity exacerbated the Flash Crash of 2010. Upson and Van Ness [2017] find that quote fragmentation increase quoted and effective spreads.

# Chapter 3

# A Tale of Two Jumps

This chapter proposes a signal extraction model of price discovery to examine intraday price discovery in a state space framework. The variance contributions of two types of price moves are separately considered: (i) continuous returns and (ii) discrete jumps. My results suggest that even though jumps are extremely rare events which happen with probability of 0.38 percent, they contribute 12.87 percent to price discovery and 36 percent to transitory mispricing. Using estimates from the state space model I decompose intraday realized profits from the provision of liquidity into informational losses from permanent jumps and reversal profits from transitory jumps. The results suggest that profits from price reversals and quoted spreads are not sufficient to compensate for large informational losses, thereby liquidity suppliers face net losses during jumps. Profits from the the provision of intraday liquidity arise exclusively from continuous price moves.

## 3.1 Introduction and Literature Review

Using a novel time series representation of intraday price series as distinct continuous and jump parts, I propose a state space model to explicitly extract the ex-ante and realized informational content of jumps. In essence, I disentangle information (i.e. permanent component) and transitory mispricing (i.e. noise component) from observed jump returns. I further use realized estimates of permanent and transitory jump components to decompose profits from the provision of intraday liquidity into informational losses from permanent component and reversal profits (i.e. contrarian strategy profits) from transitory component. The results suggest that profits from price reversals and quoted spreads are insufficient to compensate for large informational losses resulting from the permanent component of jumps. However, since jumps are rare the provision of intraday liquidity remains profitable on average.

Price discovery is the process through which prices incorporate new fundamental information and thereby transition to an informationally efficient equilibrium (Grossman and Stiglitz [1980]). This chapter finds that not all jumps are equal in their contribution to price discovery, the aforementioned finding is important because the dynamics of price discovery for jumps have strong implications for investors. For instance, a large transitory jump can destabilized the market (e.g. Flash Crash, 2010) resulting in loss of investor confidence in the price system, while the large mean reverting component in transitory jumps can also result in large profits from contrarian liquidity supplying strategies. In contrast, a large permanent jump can result in significant informational losses for liquidity suppliers and thereby significantly impair market liquidity. Permanent jumps also differ from transitory jump as the former entails inventory risk for liquidity suppliers while the latter is largely hedgeable simply by holding the inventory position through mean reversion. More generally, the instantaneous nature of jumps make portfolio allocation suboptimal at the instance following the jump, but only to the degree of the permanent component of jumps.

My main contribution is to methodology. I propose a time series representation of the observed price series which allows for a convenient representation of jumps as a signal extraction model in a Non-Gaussian state space framework. Using estimation methods proposed by Shephard and Pitt [1997], and in a series of articles by Durbin and Koopman [1994, 1997, 2000] I estimate variance contributions of jumps to permanent and transitory intraday return variation. My analysis is frequentist and allows convenient estimates of realized paths of permanent and transitory jumps returns. These realized estimators are efficient in the sense of conditioning on future and past information in price series. I use the t-distribution for permanent and transitory jump components which allow for a sufficiently large probability of tail events to match my sample distribution of jump returns. To the best of my knowledge, this is the first article to explicitly model price jumps in a signal extraction framework, where each jumps is uniquely disentangled into its permanent and transitory component.

I find that while jumps are rare events which occur with probability of 0.38 percent, they contribute 12.87 percent to permanent variance of intraday price returns. However, these large contributions to price discovery are also accompanied by a 36 percent contribution to total transitory variance. In addition, the median jump realized signal-to-noise ratio is 6 times smaller as compared to the median continuous signal-to-noise ratio. These results suggest that relative to continuous returns, jumps are large but often imprecise signals of fundamental information. Using realized estimates of permanent and transitory jumps and their continuous counterparts, I examine the provision of liquidity during jump and continuous returns. I find that jumps contribute 6.74 percent to intraday informational losses, which relative to their probability is 17.76 times larger. Examining intraday profits net of informational losses, the results suggest that liquidity suppliers, on average face a small loss during jumps compared to an average profit of 25,422.41 dollars from continuous price moves across each stock and day. Therefore, profits from the provision of intraday liquidity result exclusively during continuous price moves. I also use the proposed methodology to examine price discovery during the 2008 financial crisis. My results suggest that price discovery during the 2008 financial crisis was dominated by a larger proportion of low latency news. This translated into a larger contribution of jumps to permanent variance by 17.62 percent as compared to the relatively tranquil period of 2010.

My work is related to studies which examine price discovery in a state space framework. For instance, Menkveld, Koopman, and Lucas [2007] examine price discovery for Dutch stocks across multiple markets and, Brogaard, Hendershott, and Riordan [2014] examine the contribution of high frequency traders to price discovery. I also contribute to recent yet growing literature which examine the market dynamics of jumps. For instance, Christensen, Oomen, and Podolskij [2014] disentangle jumps from volatility bursts at tick frequency, and Brogaard at al. (2018) examine the liquidity supplying role of high frequency traders during jumps.

The outline of the chapter is as follows. Section 3.2 introduces price discovery as a signal extraction model and details the state space model used to compute variance contributions and realized estimators of permanent and transitory return components. Section 3.3 presents my data and explains my sample selection. Section 3.4 presents results and examines characteristics of estimated permanent and transitory jump components from the state space model. In section 3.5, I present two applications to my methodology. First, I examine informational losses to the provision of liquidity during jumps. Second, I examine price discovery during the 2008 financial crisis. Section 3.6 concludes.

## 3.2 Methodology

### 3.2.1 State Space Model (SSM)

Introducing price discovery as a signal extraction model and assuming markets are rational, it follows that new information about a stock's fundamental value has a permanent impact on the stock price. This permanent component is precisely the component of observed return which contributes to price discovery. Denoting $\{\eta_{i,k}\}_{k=1}^{T}$ as the contributing series to price discovery for stock i, the permanent price (i.e. efficient price) can be modelled as a random walk adapted to the market's aggregate informational set.

$$m_{i,t} = m_{0,i} + \sum_{k=1}^{t} \eta_{i,k} \;\; ; \;\; t \in [0 \mathrel{..} T] \tag{1}$$

Where $m_{0,i}$ is the initial permanent price. Since the permanent price is a martingale, it follows that the contributing series $\{\eta_{i,k}\}_{k=1}^{T}$ is a martingale difference series (MDS). Denoting $s_{i,t}$ as stock i's transitory price, the observed logarithmic price is the sum of permanent and transitory parts.

$$p_{i,t} = m_{i,t} + s_{i,t} \tag{2}$$

Since $p_{i,t}$ is logarithmic price, observed returns are the sum of permanent and transitory returns.

$$\Delta p_{i,t} = \eta_{i,t} + \Delta s_{i,t} \tag{3}$$

Using a SSM framework I can decompose the variance of $\Delta p_{i,t}$ into a part related to new fundamental information $\eta_{i,t}$ and a part related to transitory pricing errors $\Delta s_{i,t}$. Transitory mispricing in returns result from market overreaction to information and/or liquidity considerations. As will be discussed in section 3.5.1, $\Delta s_{i,t}$ result in profits for liquidity suppliers. Since markets eventually corrects all price moves unrelated to the stock's efficient value, $\Delta s_{i,t}$ is the mean reverting part of observed return. In addition to variance decomposition a SSM can also compute realized estimators of permanent and transitory returns by conditioning on all past and future information in observed price. These realized estimators denoted as $\hat{\eta}_{i,t|p}$ and $\Delta\hat{s}_{i,t|p}$ are efficient in the sense of using all past and future information in the observed price series. In the SSM framework realized estimators are commonly known as smoothed innovations. However, in the present framework I want to emphasize that $\hat{\eta}_{i,t|p}$ and $\Delta\hat{s}_{i,t|p}$ are ex-post estimators of the unobserved sample time paths of permanent and transitory returns and therefore I will refer to these estimators as realized estimators.

### 3.2.2 Return decomposition

Intraday price series is characterized by two types of price increments, (i) continuous returns and (ii) discrete jumps. It is important to distinguish between these two increments for two reasons. First, from a price discovery perspective, these two types of returns differ in the characteristics of information each incorporate during price formation. While continuous returns are associated with the slow trickle down of information in continuous trading, jumps incorporate low latency and high informational content news. Second, from a modeling perspective, continuous returns satisfy properties associated with the Normal distribution, whereas jumps being extreme events have heavier tails than the Normal distribution. Denoting $\mathcal{J}_{i,y}$ as the set of jump times for stock i in year y, and denoting its complement $\mathcal{J}_{i,y}^c$, the set of continuous times, it follows that observed returns have the following decomposition into jump and continuous returns.

$$\Delta p_{i,t} = \begin{cases} C_{i,t} + J_{i,t} & \text{if} \quad t \in \mathcal{J}_{i,y} \\ C_{i,t} & \text{if} \quad t \in \mathcal{J}_{i,y}^c \end{cases} \tag{4}$$

Where $C_{i,t}$ is the continuous return for stock i at time t and $J_{i,t}$ is the jump return. To avoid convulsion of jump and continuous distributions I make the following assumption.

**Assumption 1.**

$$\Delta p_{i,t} \approx J_{i,t} \quad \textit{if} \quad t \in \mathcal{J}^{i,y}$$

Assumption 1 states that when there is a jump, the jump return is sufficiently larger than the continuous return, such that the jump return can be approximated by the observed return alone. In my sample, the size of observed returns during a jumps is 9 times larger, on average than the continuous counterpart. Therefore assumption 1 is innocuous for the sample.

**Assumption 2.**

$$(a) \quad J_{i,t|t-1} \sim i.i.d(0, \sigma^2_{i,y})$$

$$(b) \quad C_{i,t|t-1} \sim \; i.i.d \; \mathcal{N}(0, \sigma^2_{i,d})$$

*Where y denotes year and d denotes day.*

Assumption 2(a) states that time t jump return, conditional on time $t-1$ information set, is identically and independently distributed with constant variance for a given stock-year. Assumption 2(b) states that time t continuous return, conditional on time $t-1$ information set, is identically and independently distributed following a Normal distribution with constant variance for a given stock-day. Assumption 2-b allows continuous component of volatility to vary across days but is non-stochastic.

### 3.2.3 Jump State Space Model

To examine the contribution of jumps to price discovery I construct an observed jump price process which contains all relevant information related to permanent and transitory jump returns. Using the method proposed by Lee and Mykland [2008] and outlined in 3.2.9, I estimate a set of jump times $\mathcal{J}_{i,y}$, for each stock i in year y. Under Assumption 1 of the previous section, jump return $J_{i,t}$ can be approximated by the observed return alone when $t \in \mathcal{J}_{i,y}$. Under assumption 2-a, conditional on time $t-1$, time $t$ jump return is identically and independently distributed. In particular, jump return is conditionally independent of past and future continuous return. It follows that all relevant information required for a signal extraction model can be recursively summarized by a jump price series which, at each time t, is the aggregate sum of past jump returns as follows.

$$p^J_{i,t} = p^J_{i,0} + \sum_{\{k \in \mathcal{J}_{i,y} \cap [0\ \ t]\}} J_{i,k} \tag{5}$$

Where $p^J_{i,0}$ is the initial jump price. Since my initialization will be diffuse, estimators of permanent and transitory jump returns will not depend on the precise value of the unknown initial jump price $p^J_{i,0}$. Using (5) I can formulate jump part of the SSM consisting of a permanent jump price (i.e. efficient jump price) as random walk process, and a transitory component which is a stationary and therefore mean reverting process. The increments to the random walk are permanent jump returns which incorporate new information about the security's efficient value in observed price. Whereas, the transitory jump return $\Delta s_t$ is the mean reverting mispricing component. The jump part of the SSM is as follows:

$$p^J_{i,t} = m^J_{i,t} + s^J_{i,t} \tag{6a}$$

$$m^J_{i,t} = m^J_{i,t-1} + \eta^J_{i,t} \tag{6b}$$

Where $\eta^J_{i,t} \sim i.i.d(0, \sigma^{\eta,J}_{i,y})$, $s^J_{i,t} \sim i.i.d(0, \sigma^{s,J}_{i,y})$ and $t \in \mathcal{J}^{i,y}$. Expressions (6a) and (6b) define a parametric model for variance decomposition of jump returns into permanent and transitory parts. Where the permanent variance $(\sigma^{\eta,J}_{i,y})^2$ is the variance contribution of new information and the transitory variance $(\sigma^{s,J}_{i,y})^2$ is the variance contribution of transitory mispricing incorporated in the observed price by jumps. Using $(\sigma^J_{i,y})^2$ to denote the variance of $J_{i,t}$ for $t \in \mathcal{J}_{i,y}$, the signal-to-noise variance decomposition is as follows:

$$(\sigma^J_{i,y})^2 = (\sigma^{\eta,J}_{i,y})^2 + (\sigma^{s,J}_{i,y})^2 \tag{7}$$

Corresponding to the variance contributions in equation (7), the SSM can also decompose jump returns into a MDS, $\{\eta^J_{i,t}\}_{t\in\mathcal{J}^{i,y}}$ which is the informative component of jump returns and a transitory mispricing component of jump returns, $\{\Delta s^J_{i,t}\}_{t\in\mathcal{J}^{i,y}}$. The jump return decomposition is as follows.

$$J_{i,t} = \eta^J_{i,t} + \Delta s^J_{i,t} \tag{8}$$

In order to validate that the stated model is well specified, I test for a permanent component (i.e. unit root) in the jump price series $p^j$ across the 347 stock-year in my sample. My sampling frequency for intraday returns is 1 minute. I choose two unit root tests, the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test and the Phillips Perron (PP) test. The null hypothesis of the KPSS test assumes that a unit root is absent in the series, whereas the null hypothesis of the PP test assumes a unit root is present. Results from these two unit root tests are presented in panel A of table 1. Starting with the KPSS test, the average LM statistic is 28.55 and the null hypothesis of absence of a unit root is rejected in all stock-years at the 5 percent level. For the PP test, the average PP test statistic is -0.23 and the null hypothesis of presence of a unit root in 94 percent of stock-years cannot be rejected. For the remaining 6 percent of stock-years the null hypothesis is rejected after allowing for a constant mean in the PP test. These results provide conclusive evidence that $\{p^J\}_{t \in \mathcal{J}^{i,y}}$ contains a permanent component across all stock-years in my sample and therefore the jump state space model in expressions (6a) and (6b) is well specified.

### 3.2.4 Continuous Part of State Space Model

Correspondingly to the jump observed price series, I construct the continuous price series as follows:

$$p_{i,t} = p^c_{i,0} + \sum_{\{k \in \mathcal{J}^c_{i,y} \cap [0\ \ t]\}} C_{i,k} \tag{9}$$

where $p^c_{i,0}$ is the initial continuous price. Using equation 8, I can formulate a continuous part of the SSM with Normally distributed returns as follows:

$$p^c_{i,t} = m^c_{i,t} + s^c_{i,t} \tag{10a}$$

$$m^c_{i,t} = m^c_{i,t-1} + \eta^c_{i,t} \tag{10b}$$

$$s^c_{i,t} = \phi s^c_{i,t-1} + \epsilon^c_{i,t} \tag{10c}$$

where $\eta^c_{i,t} \sim \mathcal{N}(0, \sigma^{\eta,c}_{i,d})$, $\epsilon^c_{i,t} \sim \mathcal{N}(0, \sigma^{\epsilon,c}_{i,d})$, d denotes day, and $t \in \mathcal{J}^c_{i,y}$.

Following (Hendershott and Menkveld [2014]) and to allow for price pressure arising from persistent order splitting strategies I model transitory price as an autoregressive process. The half life of transitory mispricing in continuous price is $-\dfrac{ln(2)}{ln(|\phi|)}$.

Corresponding to my jump price process, I also test that the continuous SSM is well specified by testing for a unit root in the observed continuous price $p_{i,t}$. Since the continuous SSM is estimated independently for each stock-day, the unit root tests are also independently implemented for each of the 64,889 stock-days in my sample. Panel B of table 3.1 presents results for the KPSS and PP test for the continuous price series $p^c$. The average LM statistic from the KPSS test is 32.76 and the null hypothesis of the absence of unit root in all stock-days in my sample can be rejected. For the PP test the null hypothesis of presence of a unit in all, but 1.94 percent of stock-days, cannot be rejected. These results present strong to suggest that the continuous price $p^c$ contains a permanent component and therefore the continuous SSM is well specified.

**Table 3.1: Unit Root Tests**

This table presents results from the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test and the Phillips Perron (PP) test for the jump price series $p^j$ (Panel A) and continuous price series $p^c$ (Panel B). The null hypothesis of the KPSS test assumes that a unit root is absent in the series, whereas the null hypothesis of the PP test assumes a unit root is present.

KPSS Test: $p_t = m_t + s_t$ ; $m_t = m_{t-1} + \eta_t$ ; where $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$ and $s_t$ is a stationary process.
Null hypothesis of KPSS test is $\sigma_\eta^2 = 0$
PP Test $\Delta p_t = \alpha p_{t-1} + u_t$
Null hypothesis of PP test is $\alpha = 0$

Panel A: Jump Price Series

| | | KPSS Test | | PP Test | |
|---|---|---|---|---|---|
| Year | N | LM Statistic | % Rejected | PP Statistic | % Rejected |
| All | 347 | 28.55 | 100 | -0.23 | 6.34 |
| 2008 | 174 | 28.24 | 100 | -0.04 | 3.44 |
| 2010 | 173 | 28.86 | 100 | -0.41 | 9.24 |

Panel B: Continuous Price Series

| | | KPSS Test | | PP Test | |
|---|---|---|---|---|---|
| Year | N | LM Statistic | % Rejected | PP Statistic | % Rejected |
| All | 64,889 | 32.76 | 100 | 0.02 | 1.94 |
| 2008 | 32,948 | 32.72 | 100 | -0.03 | 2.47 |
| 2010 | 31,941 | 32.79 | 100 | 0.08 | 1.40 |

### 3.2.5 Variance Decomposition of Jumps

I begin with outlining the estimation procedure for jump SSM given in expressions (6a) and (6b). The first step is the choice of a distribution for permanent jump returns denoted by $\eta^J_{i,t}$ and transitory price $s^J_{i,t}$ respectively. Since observed jump returns $\Delta p^J_{i,t}$ are on average 9 times larger than continuous returns $\Delta p^c_{i,t}$, it is important that the choice of distributions have sufficiently large mass in the tails (i.e. large kurtosis). Examining the kurtosis of $\Delta p^J_{i,t}$ in panel A of table 3.1, it can be observed that the average kurtosis across the two years in the sample is 8.46. Since $\Delta p^J_{i,t}$ is the sum of $\eta^J_{i,t}$ and $\Delta s^J_{i,t}$, the distributions of $\eta^J_{i,t}$ and $s^J_{i,t}$ should have larger kurtosis than that of $\Delta p_t$. This is because the sum of two distribution has smaller kurtosis than the individual distributions. At the same time, the estimation procedure leads to poor convergence when the kurtosis is unbounded or the distribution is bi-modal. Given the above considerations, I choose a two parameter t-distribution with substantially heavier tails than the normal distribution. Since the focus of this chapter is on variance contributions I fix the degrees of freedom of the t-distribution to 4.25 and estimate variances of the t-distributed $\eta^J_{i,t}$ and $s^J_{i,t}$ from the SSM [1] . The choice of 4.25 degrees of freedom for the two parameter t-distribution leads to an excess kurtosis over the normal distribution of 24 and therefore substantially larger probability in the tails than the Normal distribution. Given this choice of distribution, the jump-state model in (6a) and (6b) takes the following form.

$$p^J_{i,t} = m^J_{i,t} + s^J_{i,t} \tag{11a}$$

$$m^J_{i,t} = m^J_{i,t-1} + \eta^J_{i,t} \tag{11b}$$

Where $\eta^J_{i,t} \sim t_{4.25}(0, \sigma^{\eta,J}_{i,y})$, $s^J_{i,t} \sim t_{4.25}(0, \sigma^{s,J}_{i,y})$ and $t \in \mathcal{J}^{i,y}$.

Since densities for both the permanent and transitory component are Non-Normal, the standard Kalman Filter no longer yield efficient estimators of realized permanent and transitory jump components. Therefore, I use simulation based

[1]The estimates of permanent and transitory components are similar for degrees of freedom ranging from 4.25 to 5.25 in increments of 0.25

methods proposed by Shephard and Pitt (1997), and in a series of papers by Durbin and Koopman [1994, 1997, 2000]. These methods involve drawing samples from an importance density to numerically maximize the likelihood function. The procedure is as follows.

Denote the vector of unknown parameters as $\Psi^J = \{\sigma_y^{\eta,J}, \sigma_y^{s,J}\}$, the stacked vectors of jump price, permanent jump returns and transitory jump returns as $p^J = (p_1^J, ..., p_T^J)$, $\eta^J = (\eta_1^J, ..., \eta_T^J)$ and $s^J = (s_1^J, ..., s_T^J)$ respectively. In addition, denote $f(p^J|\Psi^J)$ as the jump price density. Where for notational convenience I have dropped the stock index i. Then the likelihood function $\mathcal{L}_f$, takes the following form:

$$\begin{aligned}
\mathcal{L}_f(p^J|\Psi^J) &= \int f(p^J, \eta^J|\Psi^J) d\eta^J \\
&= \int \frac{f(p^J, \eta^J|\Psi^J)}{g(\eta^J|p^J, \Psi^J)} g(\eta^J|p^J, \Psi^J) d\eta^J \\
&= \mathcal{L}_g(p^J|\Psi^J) \mathbb{E}_g[\omega(\eta^J, s^J|\Psi^J)]
\end{aligned} \tag{12}$$

where $\mathcal{L}_g(p^J|\Psi^J)$ is the likelihood function containing the two importance densities, $\mathbb{E}_g$ denotes expectation under density $g$, and $\omega$ are importance sampling weights.

$$\begin{aligned}
\omega(\eta^J, s^J|\Psi^J) &= \frac{f(\eta^J, s^J|\Psi^J)}{g(\eta^J, s^J|\Psi^J)} \\
&= \frac{\prod_t f(\eta_t^J|\Psi^J) f(s_t^J|\Psi^J)}{\prod_t g(\eta_t^J|\Psi^J) g(s_t^J|\Psi^J)}
\end{aligned} \tag{13}$$

For a given $\Psi^J$ I estimate the value of $\mathcal{L}_f$ by drawing N samples of permanent jumps $\{\eta_{(k)}^J\}_{k=1}^N$ from the importance density $g(\eta^J|p^J, \Psi^J)$ and of transitory jumps $\{s_{(k)}^J\}_{k=1}^N$ from the importance density $g(s^J|p^J, \Psi^J)$ . I then proceed to compute the sample counterpart of expression (12).

$$\widehat{\mathcal{L}}_f(p^J|\Psi^J) = \mathcal{L}_g(p^J|\Psi^J)\frac{1}{N}\sum_{k=1}^{N}\omega(\eta^J_{(k)}, s^J_{(k)}) \tag{14}$$

I estimate $\hat{\Psi}^J$, the maximized likelihood estimator of $\Psi^J$ as the value which maximizes expression (14). Since both permanent and transitory jump returns are Non-Normal I choose a large sample size of 500 draws from the importance densities $g(\eta^J|p^J, \Psi^J)$ and $g(s^J|p^J, \Psi^J)$, for each computation of expression (14). In addition, I use diffuse initialization and therefore $\hat{\Psi^J}$ and realized permanent and transitory jump returns do not depend on the initial permanent jump price $m_0^J$. In order to increase the efficiency of the estimators I use 4 antithetic variables which are balanced for scale and location. Details about these antithetic variables are in Durbin and Koopmans (2012). Combined with antithetic variables the complete sample size N is 2,000 for each computation of $\widehat{\mathcal{L}}_f(p^J|\Psi^J)$ in (14).

As in Jungbacker and Koopmans (2007) I use the normal densities as the importance densities to the t-distributions $t_{4.25}(0, \sigma^{\eta,J}_{i,y})$ and $t_{4.25}(0, \sigma^{s,J}_{i,y})$. The parameters of the importance densities are estimated by maximizing the logarithmic conditional density $f(\eta^J, s^J|p^J, \Psi^J)$ with respect to $\eta^J$ and $s^J$. The resulting value $\hat{\eta}^J$ and $\hat{s}^J$ are the mode and therefore the most probably value of $\eta^J$ and $s^J$ under the Non-Gaussian densities $f(\eta^J|p^J, \Psi^J)$ and $f(s^J|p^J, \Psi^J)$. The procedure is as follows.

$$\begin{aligned} max_{\eta^J,s^J}\{logf(\eta^J, s^J|p^J, \Psi^J)\} = max_{\eta^J,s^J}\{&logf(p^J|\eta^J, s^J, \Psi^J)- \\ &logf(\eta^J, s^J|\Psi^J) - logf(p^J|\Psi^J)\} \end{aligned} \tag{15}$$

Using Newton-Rhapson method to maximize expression (15) results with Normal densities $g(\eta^J|p^J, \Psi^J)$ and $g(s^J|p^J, \Psi^J)$ with the same mode as $f(\eta^J|p^J, \Psi^J)$

and $f(s^J|p^J, \Psi^J)$ respectively. Following which I then proceed to use $g(\eta^J|p^J, \Psi^J)$ and $g(s^J|p^J, \Psi^J)$ as the importance densities to draw sample of $\eta^J$ and $s^J$. To put simply, the importance densities $g(\eta^J|p^J, \Psi^J)$ and $g(s^J|p^J, \Psi^J)$ are normally distribution with the same mode as the Non-Normal jump densities $f(\eta^J|p^J, \Psi^J)$ and $f(s^J|p^J, \Psi^J)$. Each single draw from the importance densities $g(\eta^J|p^J, \Psi^J)$ and $g(s^J|p^J, \Psi^J)$ requires a single run of the Kalman Filter and Simulation smoother of Durbin and Koopmans (2002).

### 3.2.6 Realized Permanent and Transitory Jump Returns

In addition to estimating variance contribution of permanent and transitory jump returns, I also compute realized estimators of time paths for jump return components $\hat{\eta}_t^J$ and $\Delta\hat{s}_t^J$ . These realized estimators are expectations of permanent and transitory jump returns conditional on both future and past time paths of observed variables in the SSM. In the state space modeling literature these estimators are known as smoothed innovations. These estimator are efficient in the sense of using all available information in past and future time paths. Using the importance densities $g(\eta^J|p^J, \Psi^J)$ and $g(s^J|p^J, \Psi^J)$ from section 3.2.5 , I compute realized estimator of permanent jump returns as follows:

$$\begin{aligned}
\eta^J_{t|p^J} &= \int \eta_t f(\eta_t^J|p^J, \Psi^J) d\eta_t^J \\
&= \int \eta_t \frac{f(\eta_t^J|p^J, \Psi^J)}{g(\eta_t^J|p^J, \Psi^J)} g(\eta_t^J|p^J, \Psi^J) d\eta_t^J \\
&= \frac{\mathbb{E}_g[\eta_t^J \omega(\eta_t^J, s_t^J|\Psi^J)]}{\mathbb{E}_g[\omega(\eta_t^J, s_t^J|\Psi^J)]}
\end{aligned} \tag{16}$$

where $\omega(\eta_t^J, s_t^J|\Psi^J)$ are importance weights defined in expression (13). The sample counterpart of (14) is as follows:

$$\hat{\eta}^J_{t|p^J} = \frac{\sum_{k=1}^{N} \eta^J_{t,(k)} \omega_k}{\sum_{k=1}^{N} \omega_k} \tag{17}$$

similarly, realized estimator of transitory jump price is as follows:

$$\hat{s}^J_{t|p^J} = \frac{\sum_{k=1}^{N} s^J_{t,(k)} \omega_k}{\sum_{k=1}^{N} \omega_k} \tag{18}$$

Using (15), the realized estimator for transitory jump return is $\Delta\hat{s}^J_{t|p^J}$. I compute realized permanent and transitory jump returns by simulating 500 draws from the importance density $g(\eta^J|p^J, \Psi^J)$, along with 4 antithetic variables which lead to a sample size N of 2,000. As is the case when numerically computing the likelihood function in section 3.2.5, each draw of $\eta_t$ from the importance density requires a of single run of the Kalman Filter and Simulation smoother of Durbin and Koopman [2002].

### 3.2.7 Continuous State Space Model Estimation

Since the continuous SSM in expressions (10a)-(10c) has normally distributed permanent returns $\eta^c_{i,t}$ and innovations to the transitory price $\epsilon^c_{i,t}$, standard Kalman Filtering yields efficient estimators of continuous return components $\hat{\eta}^c_{t|p^c}$ and $\Delta\hat{s^c}_{t|p^c}$. I use Expectation Maximization (EM) algorithm of Dempster, Laird, and Rubin [1977] to compute starting values. These starting values are then used as initial values to maximize the likelihood function using BFGS algorithm. Maximizing the likelihood function which is computed using the Kalman Filter yield estimates for the continuous part of my model for each stock-day in the sample.

### 3.2.8 Permanent Price Estimator

The realized estimator for permanent (i.e efficient price) is recursively defined for each year by combining realized returns from jumps and the continuous component for each stock-day in the sample. Denoting T as the last time period for a given day, the realized permanent price has the following recursive form.

$$\hat{m}_{i,t|T} = \begin{cases} \hat{m}_{i,t-1|T} + \hat{\eta}^J_{i,t|p^J} & \text{if} \quad t \in \mathcal{J}_{i,y} \\ \hat{m}_{i,t-1|T} + \hat{\eta}^c_{i,t|p^c} & \text{if} \quad t \in \mathcal{J}^c_{i,y} \end{cases} \tag{19}$$

The realized estimator $\hat{m}_{i,t|T}$ is the conditional expectation of permanent price conditional on all continuous information in a given day, and all jump information for the given year. If the first observation of a given day is not a jump, I estimate the initial permanent price by $\hat{m}^c_{i,0|T}$, which is the continuous estimator of initial permanent price, whereas if the first observation is a jump, I use $\hat{m}^c_{i,0|T} + \hat{\eta}^J_{i,-1|p^J}$ as the initial permanent price. Since both the jump and continuous SSM are initialized using diffuse initializations, the realized permanent price and all other estimator do not depend on the true realization $m_{0,i}$.

### 3.2.9 Jump Detection

A variety of tests have been proposed to detect the presence of jumps at intraday frequencies (e.g. Barndorff-Nielsen and Shephard [2006]; Jiang and Oomen [2008]; Aït-Sahalia and Jacod [2009]; Andersen et al. [2010]). The earliest contribution to this literature was an estimator of integrated volatility, realized bipower variation (RBV), proposed in Barndorff-Nielsen and Shephard (2004). Unlike realized variation (RV), RBV is robust to jumps in returns and this makes it a valid measure of sample return variation arising from the continuous part of return process. More specifically defining the return process as a jump diffusion.

$$dp(t) = \sigma_t dW_t + Y_t dq_t \tag{20}$$

Where $W_t$ is Brownian motion, $\sigma_t$ is strictly positive instantaneous volatility, $Y_t$ is random jump size and $q_t$ is Poisson-jump process. RBV is constructed by multiplying adjacent returns from discretely sampled data. As shown in Lee and Mykland (2008) a slight modification of RBV yields an estimator of instantaneous volatility over a rolling window of size K as follows.

$$\hat{\sigma}_t^2 = \frac{1}{K-2} \sum_{i=t-K+2}^{t-1} |\Delta p_t||\Delta p_{t-1}| \tag{21}$$

The exact time of a jump along with the jump size can easily be estimated using a non-parametric test statistic proposed by Lee and Mykland [2008]. The intuition behind their test statistic is as follows: at any given time, the observed return can be large if there is a jump, or because instantaneous volatility had a large realization, which suggest that returns standardized by a jump robust estimator of instantaneous volatility are large precisely when there is a jump. Following this institution the authors propose the following test statistic.

$$\mathrm{L}_t = \frac{\Delta p_t}{\hat{\sigma}_t^2} \tag{22}$$

Where $\hat{\sigma}_t^2$ is the jump robust estimator of instantaneous volatility from (20).

Under certain regularity conditions outlined in Lee and Mykland [2008], the authors show that under the null of absence of a jump, $L_t$ converges asymptotically to a known distribution. Using the test statistic $L_t$ I detect detect jumps as follows. Across each stock-year, I compute $L_t$ for all intraday returns sampled at the 1 minute frequency and using a rolling window K=300. This choice of rolling window size is consistent with the authors' suggestion for returns sampled at the 1 minute frequency. Using a critical value of 1 percent I test for a jump, for each stock i and time t return $\Delta p_{i,t}$, for the two sample periods of 2008 and 2010 independently. Therefore, for each stock i and year y, I estimate a set of jump times, $\mathcal{J}_{i,y}$ with associated jump returns, $J_{i,t}$. If a return $\Delta p_{i,t}$ is a not a jump then it is a continuous return.

## 3.3 Sample and Data

I start my sample selection with the Compustat index constituents data and select all stocks listed on the S&P 500 index for least one month in each of the 9 years between 2008 to 2016. This ensures that the sample stocks are sufficiently liquid to trade at intraday frequencies. Following which I remove all stocks that (i) underwent a stock split, (2) paid stock dividends, (3) de-listed or the issuing company underwent a merger or (4) closing price falls below 5 dollars in Center for Research in Security Price (CRSP) daily stock file, between 2008 to 2016. Following these four filters I am left 229 stock, of which 215 uniquely matched with Trade & Quotation (TAQ) Data master files. This concludes my stock sample selection with 215 stocks. I use two temporal samples for the SSM estimation namely (i) 1st January 2008 to 31 December 2008, the year of the financial crisis, and (ii) the non-crisis period of 1st January 2010 to 31st December 2010. The latter sample is sufficiently close to the financial crisis such that the stock issuing companies are not likely to undergo structural changes, whereas the height of the 2008 financial crisis had resolved by January of 2010.

In order to avoid estimation issues related to small sample sizes in the jump part of the SSM, I truncate the bottom half of stocks by number of jumps detected for each of the sample years . In order to avoid a biased sample towards high jump stocks I also truncated the top 10 percent of the stocks by number of jumps for each year in my sample. This preserves the median number of jumps across stocks in a given year. The final sample consist of 174 stocks in my first sample year, 2008 and 173 stocks in my second sample year, 2010.

I sample midpoint price and compute returns at the 1 minute frequency for each day in 2008 and 2010 during regular trading hours hours, 9:30 am to 4:00 pm. I remove the first and last minute of the trading day to avoid the opening and closing batch auction returns. I also remove withdrawn quotes using the methods and code of Holden and Jacobsen [2014]. All market statistics such as spread, orderflow imbalance and traded volume are computed using aggregates or means within the 1 minute interval over which the corresponding return is calculated. In addition to market data I also use high frequency and machine readable news data from Raven Pack News Analytics Database. This database consists of news events timestamped to the millisecond, and used in real time trading by high frequency and other algorithmic traders. The database consist of news feed from sources such as Barron's MarketWatch and Dow Jones Newswire, and cover approximately 90 percent of all global investment news. From RavenPack News Database I filter news that are relevant to the stocks in the sample, by using a unique company level identifier and a relevance score. Relevance score is computed by RavenPack using proprietary text analysis algorithms. Relevance score takes value between 0 and 100 with high scores indicating high relevance. To filter news by relevance, I remove all news with relevance scores smaller than 90. I further filter news based on novelty score computed by RavenPack at the news level. I remove all news with novelty score of less than 100 which is the highest level of novelty. The final news dataset consist of high relevance and highly novel news for the sample stocks during the years 2008 and 2010.

**Table 3.2: Summary Statistics - Sample Stocks**
This table reports descriptive statistics for stocks in my sample by year (Panel A), and by size group (Panel B) as measured by dollar market capitalization (MCAP). Daily vol. is average daily traded volume in dollars, and Price is the average intraday dollar price sampled at the 1 minute frequency. Spread is average quoted spread in percentage basis points, computed from intraday midpoints at the millisecond frequency. N is the number of stocks in the given group.

Panel A: By Year

| Year | N | Number of Jumps | | | MCAP($\times 10^9$) | Price | Daily Vol.($\times 10^6$) | Spread |
|---|---|---|---|---|---|---|---|---|
| | | Min | Mean | Max | | | | |
| All | 347 | 175 | 336.83 | 472 | 36.90 | 46.37 | 265.20 | 5.18 |
| 2008 | 174 | 233 | 333.32 | 472 | 39.20 | 44.72 | 286.28 | 6.34 |
| 2010 | 173 | 175 | 340.36 | 459 | 49.49 | 48.54 | 241.02 | 3.85 |

Panel B: By Size Group

| Size | N | Number of Jumps | | | MCAP($\times 10^9$) | Price | Daily Vol.($\times 10^6$) | Spread |
|---|---|---|---|---|---|---|---|---|
| | | Min | Mean | Max | | | | |
| Small | 98 | 232 | 352.49 | 468 | 9.73 | 42.26 | 102.39 | 6.30 |
| Large | 99 | 199 | 323.73 | 472 | 65.53 | 50.54 | 430.51 | 4.10 |

Table 3.2 presents descriptive statistics for the sample stocks. The minimum number of jumps for a given stock-year in my sample is 175 and the mean is

367. The sample stocks have average market capitalization (MCAP) of 37 billion dollars. The average daily traded volume across stocks in the sample is 265 million dollars. In addition, the average midpoint price is 46.37 dollars. I further examine stocks based on two size groups. Small stocks have average MCAP of 9.73 billion dollars whereas large stocks have MCAP of 65.53 billion dollars. Large stocks in the sample are more liquid with almost four times the daily traded volume of small stocks, and 2.20 basis points smaller quoted spread.

Each 1 minute interval corresponds to a 1 minute realized return, which is either a jump return or a continuous return. Panel A of table 3 presents descriptive statistics for jump time intervals and panel C for continuous intervals. From table 3, the mean jump return size is 67.24 basis points as compared to 7.68 basis points for continuous return size. Therefore on average, jump returns are almost 9 times larger than continuous returns. Jumps are also accompanied by a surge of trading activity. The average volume traded over 1 minute jump intervals is 1.94 million as compared to 0.68 million for continuous intervals. Since jumps are also times of market stress, average quoted spread are also 4 times larger than the corresponding continuous time intervals. As can be expected, 2008 had larger size jumps with an average size of 83.70 basis points as compared to 50.58 basis points for 2010. However, there are slightly fewer jumps detected in 2008 with a probability of 0.36 percent as compared to 0.40 percent in 2010. This result can be explained by taking into account that volatility was larger in 2008, which suggest that a portion of large returns were volatility spikes.

**Table 3.3: Summary Statistics - Jump and Continuous Times**
This table reports descriptive statistics for intraday jumps (Panel A), intraday jumps from the stocks which converged in state space model (SSM) (Panel B) and for and continuous (Panel C) 1 minute time intervals during the trading day. Prob.(%) is the percentage probability for jumps (Panel A) or the continuous return (Panel B). Return size is the average of the absolute value of 1 minute returns and reported in percentage basis points. Traded vol. is average dollar traded volume in the given 1-minute intraday time interval. Spread is average quoted spread in percentage basis points, computed from intraday midpoints at the millisecond frequency. For jumps I compute average kurtosis by year, and by day for the corresponding continuous returns, this is consistent with the Jumps and Continuous SSM estimation frequency.

| Panel A: Jump Returns | | | | | | | |
|---|---|---|---|---|---|---|---|
| Year | Prob.(%) | Return Size | | | Kurtosis | Traded Vol.($\times 10^6$) | Spread |
| | | Min | Mean | Max | | | |
| All | 0.38 | 4.83 | 67.24 | 6,233 | 8.46 | 1.94 | 20.15 |
| 2008 | 0.36 | 9.59 | 83.70 | 6,233 | 6.44 | 1.94 | 29.83 |
| 2010 | 0.40 | 4.83 | 50.58 | 5,629 | 10.49 | 1.93 | 10.34 |
| **Panel B: Jump Returns - SSM Sample** | | | | | | | |
| Year | Prob.(%) | Return Size | | | Kurtosis | Traded Vol.($\times 10^6$) | Spread |
| | | Min | Mean | Max | | | |
| All | 0.38 | 5.54 | 66.65 | 6,233 | 7.96 | 1.88 | 17.51 |
| 2008 | 0.36 | 10.40 | 84.53 | 6,233 | 6.53 | 1.77 | 26.20 |
| 2010 | 0.36 | 5.54 | 50.20 | 5,629 | 9.31 | 1.98 | 9.50 |
| **Panel C: Continuous Returns** | | | | | | | |
| Year | Prob.(%) | Return Size | | | Kurtosis | Traded Vol. | Spread |
| | | Min | Mean | Max | | | |
| All | 99.62 | 0 | 7.68 | 4,991 | 2.48 | 0.68 | 5.12 |
| 2008 | 99.64 | 0 | 9.81 | 1,199 | 2.54 | 0.74 | 6.25 |
| 2010 | 99.60 | 0 | 5.24 | 4,991 | 2.41 | 0.62 | 3.83 |

In addition to being large in size and trading activity, jumps are also rare events which occur with a probability of 0.38% across the two sample years. The average kurtosis for jump returns is 8.46 as compared to 2.48 for continuous returns. This suggest that jumps are Non-Gaussian with significant probability mass on tails of the distribution. In section 3.2.5 I outlined the choice of the t-distribution with 4.25 degrees of freedom for permanent and transitory jump returns. This choice allows for sufficiently large kurtosis in the permanent and transitory jump returns when estimating variance contributions from the SSM. The jump kurtosis in 2008 is 6.44 as compared to 10.49 in 2010. This result is caused by more mass on the left tail of the jump distribution in 2008 as compared to 2010. To conclude this section, it can be observed that jumps are low probability, high impact events characterized by large returns, traded volume and quoted spread.

## 3.4 Results

### 3.4.1 Permanent and Transitory Jumps

I proceed to examine permanent, transitory and median jumps using estimates from the SSM. As outlined in section 3.2.6, in addition to estimates of variance contributions, the proposed SSM also provides realized estimates of permanent and transitory jump returns. These realized measures are expectations of permanent and transitory jump components, conditional on all past and future information contained in observed jump price. To examine this decomposition of individual returns into permanent and transitory parts, I construct a Realized signal-to-noise (Realized SN) ratio for individual jump returns as follows.

$$\text{Realized SN}_{i,y,t} = \frac{|\hat{\eta}^J_{i,y,t|p^J}|}{|\Delta\hat{s}^J_{i,y,t|p^J}|} \text{ ; i denotes stock, y denotes year and t is jump time.} \quad (23)$$

From expressions (8) and (23), it can be observed that corresponding to each jump return $J_{i,y,t}$ is a Realized SN which estimates the size of the permanent component of jump return relative to the size of the transitory component. If the Realized SN is large, then the jump has a large estimated permanent component relative to the estimated transitory component, whereas if this estimate is small than the jump has a small estimated permanent component relative to the estimated transitory component. Using Realized SN I define representative permanent and transitory jumps for each stock-year in 3 steps, as follows. (1) For each stock-year I order jumps based on their Realized SN. (2) I compute the 25th, 50th and 75th percentile values of realized SN for each stock-year. (3) If a jump for a given stock-year has a realized SN smaller than 25th percentile, I assign it the label of transitory jump, whereas, if it has a Realized SN higher than 75th percentile, I label the jumps as permanent. For comparison purposes, I label jumps with Realized SN exactly at the 50th percentile as median jumps. Defining jumps based on percentiles for each stock-year has the advantage that each of the permanent, transitory and median jump categories have representations from all stock-years.

Panel A of table 3.4 presents descriptive statistics for each of the three jump categories. Transitory jumps tend to be largest with mean size of 84.46 basis points, whereas permanent jumps are smaller with mean size of 67.20 basis points. Median jumps are the smallest of the three, with mean jump size of 59.10 basis points. As expected, permanent jumps have large permanent size and transitory jumps have large transitory size. The mean (median) permanent jump Realized SN is 32.90 (12.11) as compared to the mean (median) transitory Realized SN of 0.42 (0.28). Median jumps have twice as large permanent size as compared to transitory size, as measured by their mean realized SN of 2.19. Therefore, jumps generally have large relative permanent component and with approximately a quarter of the mass on each tail of SN distribution.

**Table 3.4: Permanent Jumps, Continuous Jumps & Earning News Jumps**
This table reports average size of realized permanent (Perm. Component) and transitory (Tran. Component) jump components across (1) permanent jumps, (2) transitory jumps and (3) median jumps. Size of permanent and transitory jump return is computed as the absolute value of the estimated return component, reported in percentage basis points. Jump components are expectation of permanent and transitory jump returns, conditional on all past and future information in jump price series. These components are realized estimators computed using importance sampling as outlined in section 3.2.6. N is the number of jumps in each group. Standard errors are given in parenthesis. I use importance sampling to compute standard errors of jump components and delta method to compute standard errors of realized SN. I define permanent (transitory) jumps, as jump with realized signal to noise ratio (Realized SN) larger (smaller) than the 75th (25th) percentile value for a given stock-year. Median jumps have realized SN equal to the median value for a given stock year.

Permanent Jumps and Transitory Jumps

| Jump Type | N | Jump Size | Permanent Size | Transitory Size | Realized S-N Jumps | |
|---|---|---|---|---|---|---|
| | | | | | Mean | Median |
| Permanent | 23,229 | 67.20 | 61.23 | 5.97 | 32.90 | 12.11 |
| | | (1.86) | (0.11) | (0.09) | (16.93) | |
| Transitory | 24,111 | 84.46 | 14.58 | 74.65 | 0.42 | 0.28 |
| | | (0.57) | (0.12) | (0.09) | (0.01) | |
| Median | 995 | 59.10 | 42.49 | 22.75 | 2.19 | 1.99 |
| | | (1.48) | (0.46) | (0.40) | (0.07) | |

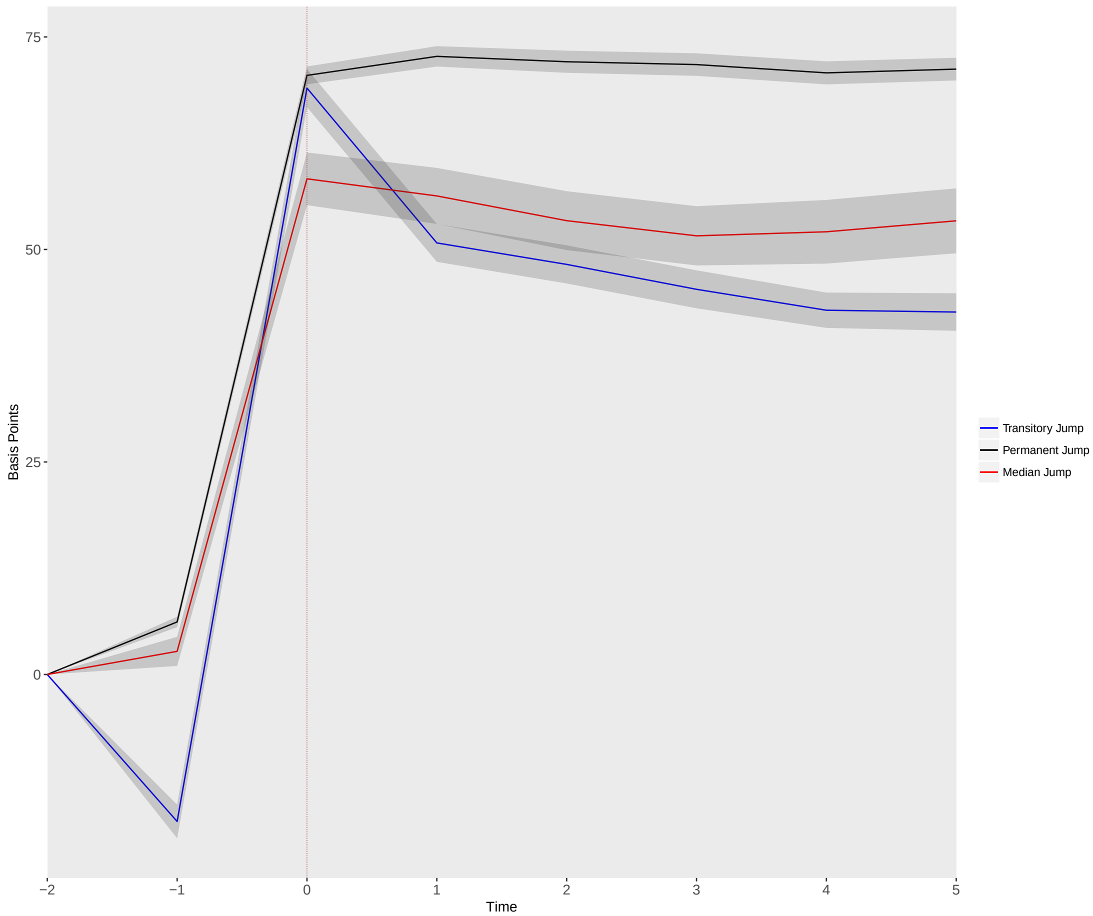



**Figure 3.1:** Average Cumulative Returns - Permanent Jumps versus Transitory Jumps

Figure 1. - Plot of average cumulative returns in event time for jumps identified as permanent, transitory and median jumps by the state space model (SSM)

Figure 3.1 presents plots of average cumulative returns for the three types of jumps in event time. I use a 7 minute event window, with 2 minutes prior and 5 minutes post. Jump time is labeled as $t = 0$. Cumulative returns are normalized to zero at $t = -2$, and negative cumulative returns are flipped to positive. Starting with transitory jumps, two features strike attention. First, transitory jumps tend to have large mean reversion in the 1 minute immediately following the jump. This initial mean reversion at event time $t = 1$ is 18.21 basis points and this is followed by mean reversions of 2.52 and 2.93 basis points at event times $t = 2$ and $t = 3$ respective. From thereon, cumulative returns are almost flat during event times $t = 4$ and $t = 5$. When compared to the the initial midpoint at event time $t = -2$ relative to the jump midpoint at event $t = 0$ a price move of 68.99 basis points can be observed, of which 26.34 basis points mean revert by event time $t = 3$ or equivalently a mean reversion of 38%.

Uninformed orders have low immediacy demand as compared to informed orders. Therefore, uninformed traders are more likely to time their trade when liquidity conditions are favorable. Since transitory jumps are likely to result from large uninformed orders (i.e. price) it can be expected that trading conditions, relative to the direction of the orderflow, to be favorable immediately preceding transitory jumps. This is precisely what can be observed at $t = -1$. Cumulative returns are negative immediately preceding transitory jumps (i.e kink at $t = -1$) which implies that the prevailing midpoint at $t = -1$ is favorable for trades immediately preceding transitory jumps, which are in the identical direction to the transitory jumps; i.e. sell (buy) orders preceding positive (negative) transitory jumps.

Next, I examine permanent jumps. In sharp contrast to transitory jumps, returns initially move in the direction of the permanent jump by 6.18 basis points, followed by a jump return of 64.29 basis points at $t = 0$, and a further move in jump direction of 1.69 basis points at $t = 1$. from thereon cumulative returns are flat. The absence of mean reversion for permanent jumps validates their permanent nature. Lastly, median jumps are similar to permanent jumps except for smaller overall size, and a small mean reversion of 2.92 basis points. To summarize the findings of this section, while the median or typical jump is permanent, approximately 25% of jumps are highly transitory with mean reversion of 38% and correspondingly 25% of jumps are highly permanent with prices continuing to move in the jump direction until 2 minute post jump.

### 3.4.2 Variance Contributions

Following the procedures outlined in section 3.2.5, I estimate ex-ante variance contribution from the SSM. The jump SSM convergence rate is 79.2% of stock-years and continuous SSM convergence rate is 78.3% of days in the sample. Table 3.5 reports estimated variances and realized sizes of jump components. The average permanent jump return variance is 3 times larger than the transitory jump return variance, while the average size of permanent jump is only 8.35 basis points larger. This is consistent with my finding in section 3.2.3 indicating that the average jumps is permanent with a significant proportion of highly permanent and highly transitory jumps. Next, I examine price informativeness for jumps and continuous returns. Recall from the discussion on variance decomposition in section 3.2.5 that observed jump and continuous returns have the following unconditional (i.e. ex-ante) variance decomposition.

$$\text{Var}[J_{i,y,t}] = (\sigma^{\eta,J}_{i,y})^2 + (\sigma^{s,J}_{i,y})^2 \quad (24)$$

$$\text{Var}[C_{i,d,t}] = (\sigma^{\eta,c}_{i,d})^2 + (\sigma^{\epsilon,c}_{i,d})^2 \quad (25)$$

Where $J_{i,y,t}$ and $C_{i,d,t}$ are jump and continuous returns respectively. Using expressions (24) and (25) I define an ex-ante measure of price informativeness,

**Table 3.5: State Space Model (SSM) - Permanent & Transitory Components**
The table reports average size of realized permanent (Perm. Component) and transitory (Tran. Component) jump components across individual jumps. Where size is computed as the absolute value of estimated return component, reported in percentage basis points. Realized jump components are expectation of permanent and transitory jump returns, conditional on all past and future information in jump price series. These components are estimated using importance sampling as outlined in section 3.2.5. N is the number of jumps in each year group. Perm. Var is the average permanent jump return variance, across stock-years. Trans. Var is the average permanent jump return variance, across stock-years. Standard errors are given in parenthesis. e use importance sampling to compute standard errors of jump components

| Year | N | Jump Size | | Perm. Var($\times 10^6$) | Trans. Var($\times 10^6$) |
|---|---|---|---|---|---|
| | | Perm. Component | Tran. Component | | |
| All | 93,030 | 40.82 | 32.48 | 27.16 | 8.12 |
| | | (0.05) | (0.04) | (0.29) | (0.12) |
| 2008 | 44,629 | 54.58 | 39.64 | 44.83 | 13.72 |
| | | (0.10) | (0.08) | (0.57) | (0.24) |
| 2010 | 48,401 | 28.14 | 25.88 | 10.36 | 2.80 |
| | | (0.05) | (0.04) | (0.13) | (0.06) |

signal-to-noise (SN) ratio, as follows.

$$\text{SN Jumps}_{i,y} = \frac{(\sigma_{i,y}^{\eta,J})^2}{(\sigma_{i,y}^{s,J})^2} \ \ ; \text{ i denotes stock and y denotes year} \tag{26}$$

$$\text{SN Cont}_{i,d} = \frac{(\sigma_{i,d}^{\eta,c})^2}{(\sigma_{i,d}^{\epsilon,c})^2} \ \ ; \text{ i denotes stock and d denotes day} \tag{27}$$

The signal to noise ratio is an ex-ante measure of the average fundamental information in observed returns relative to the average mispricing. A large value

**Table 3.6: State Space Model (SSM) - Variance Decomposition**

This table reports results from the state space model (SSM) proposed in section 3.2.3 of this chapter. I compute signal to noise ratio jumps (SN Jumps) for each stock i and year y as the ratio of the permanent and transitory jump return variance estimated from the jump SSM, $\frac{(\hat{\sigma}^{\eta,J}_{i,y})^2}{(\hat{\sigma}^{s,J}_{i,y})^2}$ ; where i denotes stock and y denotes year. The reported SN Jumps is the average across stocks-years. The continuous signal to noise ratio continuous (SN Cont.) is the ratio of the permanent and transitory continuous return variance estimated from the continuous SSM, $\frac{(\hat{\sigma}^{\eta,c}_{i,d})^2}{(\hat{\sigma}^{\epsilon,c}_{i,d})^2}$ ; where i denotes stock and d denotes day. The reported SN Jumps is the average across stocks-days. %Trans is the average amount of transitory mispricing in intraday observed price across days. %Trans-J is the average contribution of jumps to permanent return variance across days, $\frac{(\sigma^{s,J}_{i,y})^2 N^J_d}{\left[(\sigma^{s,J}_{i,y})^2 N^J_d + (\sigma^{\epsilon,c}_{i,y})^2 N^c_d)\right]}$, reported in percent. %Perm-J is the average contribution of jumps to permanent (i.e. efficient) return variance across days, $\frac{(\sigma^{\eta,J}_{i,y})^2 N^J_d}{\left[(\sigma^{\eta,J}_{i,y})^2 N^J_d + (\sigma^{\eta,c}_{i,y})^2 N^c_d)\right]}$, reported in percent. Where $N^J_d$ and $N^c_d$ are the number of jumps and continuous times respectively in a given day d. HL is the half life of transitory mispricing in the continuous price process reported in minutes. Standard errors are reported in parenthesis. I use delta method to compute all standard errors.

| Year | N | SN Jumps | | SN Cont. | | %Trans | %Trans-J | %Perm-J | HL |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Median | Mean($\times 10^5$) | Median | | | | |
| All | 50,844 | 3.85 | 3.51 | 10.04 | 20.72 | 18.69 | 36.05 | 12.87 | 2.85 |
| | | (0.01) | | (0.47) | | (0.05) | (0.23) | (0.30) | (0.02) |
| 2008 | 26,214 | 3.45 | 3.31 | 14.47 | 20.21 | 19.65 | 36.58 | 13.88 | 2.95 |
| | | 0.01 | | (0.86) | (0.07) | (0.33) | (0.30) | (0.40) | (0.02) |
| 2010 | 24,630 | 4.27 | 3.81 | 5.05 | 21.12 | 17.67 | 35.48 | 11.80 | 2.75 |
| | | (0.14) | | (0.29) | (0.06) | (0.33) | (0.34) | (0.44) | (0.02) |

of SN corresponds to more informative observed price increments (i.e. returns), whereas small values corresponds to large mispricing.

Table 3.6 presents SNs for jump and continuous returns. The average SN for jumps is 3.85 and the corresponding continuous SN is 10.04x10$^5$, across both years in the sample. Therefore, the mean continuous SN is several magnitudes higher than jump SN. In addition the median jump SN is 3.85 as compared to the 6 times larger median continuous SN of 20.72. These results suggest that continuous returns though small, have a much higher proportion of fundamental information relative to the amount of mispricing. This implies that in a relative sense, continuous returns are more precise signals, which are released to the market in small increments. This gives markets sufficient time and liquidity to process the informational accurately. Next I compute the transitory variance contribution of jump returns to intraday transitory variance, as well as the permanent variance contribution of jumps to intraday permanent variance. In order to compute intraday jump variance I treat jumps as a compound Poisson process, with permanent and transitory variances estimated from jump SSM and jump intensity computed for each day in the sample using the number of observed jumps. Denoting $N_d^J$ and $N_d^c$ as the number of jumps and continuous times respectively, the jump contribution to total intraday transitory variance is $\frac{(\sigma_{i,y}^{s,J})^2 N_d^J}{\left[(\sigma_{i,y}^{s,J})^2 N_d^J + (\sigma_{i,d}^{\epsilon,c})^2 N_d^c)\right]}$. The corresponding jump contribution to total intraday permanent variance is $\frac{(\sigma_{i,y}^{\eta,J})^2 N_d^J}{\left[(\sigma_{i,y}^{\eta,J})^2 N_d^J + (\sigma_{i,d}^{\eta,c})^2 N_d^c)\right]}$. Table 3.6, presents the average intraday relative transitory contribution of jumps across all stock days in the sample. While the probability of jumps in the sample is 0.38 percent, they contribute 12.87 percent to total intraday permanent return variance or equivalently, the share of jumps in permanent return variance is 34 times jump probability. However, this large amount of fundamental variance is also accompanied by 36.05 percent contribution to total intraday transitory variance. Therefore, the share of jumps in the average return mispricing is 95 times jump probability. These results suggest that jumps have a very large contribution in both price discovery and noise, however, jump contribution to noise is the larger. Lastly, substantial persistence can be observed in transitory continuous price. The estimated half life of transitory continuous innovations is 2.9 minutes.

## 3.5 Applications

In this section I use estimates from the state space model to (i) decompose profits from supplying liquidity into components arising from jumps and continuous returns and (ii) examine characteristics of intraday information during the 2008 financial crisis as compared to 2010.

### 3.5.1 Realized Liquidity Supplier Profits: Jumps versus Continuous Returns

As shown in Section 3.4.1, jump can have an extremely large realized permanent or realized transitory component. Do these large components entail large risk and compensating profits for liquidity suppliers? I address this question in the current section.

The permanent jump component is associated with changes in efficient value of the firm. If some traders have superior information than these informed traders will trade in the direction of the permanent component during times leading up to the jump. Liquidity suppliers, who are in the middle of all trades, make a loss when trading against informed traders. This informational loss is the adverse selection cost of supplying liquidity (Kyle [1985a];Glosten and Milgrom [1985];Easley and O'hara [1987]). In contrast, the transitory jump component is market overreaction to the jump and mean reverts at a future time and consequently, the transitory jump component is part of future observed price change which is negatively serially correlated with the current jump return. This negative serial correlation in returns are profits from the inherently contrarian nature of supplying liquidity as shown in Lehmann [1990], Lo and MacKinlay [1990] as well as Nagel [2012]. Reversal profits arise because liquidity suppliers can reverse their position from the time of the jump at the the more favorable mean reverted future price. The aforementioned argument is not unique to jumps. Corresponding to each continuous price returns the state space model estimates a continuous realized permanent and realized transitory component which can be used to compute realized informational losses and reversal profits for continuous component of returns. Consequently, using the jump and continuous parts of the state space model, I can decompose realized informational losses and reversal profits for liquidity suppliers into distinct components arising during rare jump periods and common continuous periods. This allows me to examine the profitability and risk associated with liquidity supply during jumps, as compared to the frequent continuous price moves. Since my analysis is across all intradays and not restricted to days when there are jumps, I can examine the average contribution to realized profits and losses of the two types of price moves across intradays.

The decomposition of profits into jump and continuous intraday period is similar to the spectral decomposition of profits in Hau (2001), but in the jump-continuous dimension instead of the frequency dimension of Hau [2001]. Since I use mid-quote prices in the state space model a full characterization of profits requires adding a third component, profits from spreads. I define $\Delta p_t$ as the mid-quote return and $L_{i,t}$ as consolidated limit orders for stock i which are executed within the time interval $[t-1, t)$. It follows that profits from liquidity supply on intraday d and for stock i, $\text{Profits}_{i,t}$ is as follows

$$\begin{aligned}
\text{Profits}_{i,d} &= \sum_{t\in d} \text{L}_{i,t}\Delta p_{i,t} + \sum_{t\in d} \frac{\text{QS}_{i,t}}{2}\text{Volume}_{i,t} \\
&= \sum_{t\in d} \text{L}_{i,t}(\eta_{i,t} + \Delta s_{i,t}) + \sum_{t\in d} \frac{\text{QS}_{i,t}}{2}\text{Volume}_{i,t} \\
&= \sum_{t\in d} \text{L}_{i,t}\eta_{i,t} + \sum_{t\in d} \text{L}_{i,t}\Delta s_{i,t} + \sum_{t\in d} \frac{\text{QS}_{i,t}}{2}\text{Volume}_{i,t}
\end{aligned} \tag{28}$$

Where $\frac{\text{QS}_{i,t}}{2}$ is the average quoted half spread and $\text{Volume}_{i,t}$ defined as, $p_{i,t-1} \times$ traded $\text{shares}_{i,t}$, is the traded volume during the time interval $[t-1, t)$. The term $\text{L}_{i,t}\eta_{i,t}$ is the change in liquidity suppliers' inventory resulting from trading against the permanent component of price returns during the time interval $[t-1, t)$, this term is the informational loss for liquidity supplying trades. In contrast, the term $\text{L}_{i,t}\Delta s_{i,t}$ is the profit from reversals. I can further characterize profits and losses from expression (28) into terms arising from jump and continuous component of returns. Defining $J_{i,d}$ and $J^c_{i,d}$ as the set of jump and continuous times within day d, the jump and continuous components of profits are as follows.

$$\begin{aligned}
\text{Profits}_{i,d} = &\sum_{t\in J^c_{i,d}} \text{L}_{i,t}\eta^c_{i,t} + \sum_{t\in J^c_{i,d}} \text{L}_{i,t}\Delta s^c_{i,t} + \sum_{t\in J^c_{i,t}} \frac{\text{QS}_{i,t}}{2}\text{Volume}_{i,t} + \sum_{t\in J_{i,d}} \text{L}_{i,t}\eta^J_{i,t} \\
&+ \sum_{t\in J_{i,d}} \text{L}_{i,t}\Delta s^J_{i,t} + \sum_{t\in J_{i,d}} \frac{\text{Spread}_{i,t}}{2}\text{Volume}_{i,t}
\end{aligned} \tag{29}$$

Using realized estimators from the jump and continuous parts of the SSMs I can estimate realized intraday profit for stock i as follows.

$$\widehat{\text{Profits}}_{i,d} = \underbrace{\sum_{t \in J^c_{i,d}} \text{L}_{i,t}\hat{\eta}^c_{i,t|p^c}}_{\text{Info. Loss - Cont.}} + \underbrace{\sum_{t \in J^c_{i,d}} \text{L}_{i,t}\Delta\hat{s}^c_{i,t|p^c}}_{\text{Reversal Profit - Cont.}} + \underbrace{\sum_{t \in J^c_{i,d}} \frac{\text{QS}_{i,t}}{2}\text{Volume}_{i,t}}_{\text{Spread Profit - Cont.}} + \underbrace{\sum_{t \in J_{i,d}} \text{L}_{i,t}\hat{\eta}^c_{i,t|p^c}}_{\text{Info. Loss - Jumps}} + \underbrace{\sum_{t \in J_{i,d}} \text{L}_{i,t}\Delta\hat{s}^c_{i,t|p^c}}_{\text{Reversal Profit - Jumps}} + \underbrace{\sum_{t \in J_{i,d}} \frac{\text{QS}_{i,t}}{2}\text{Volume}_{i,t}}_{\text{Spread Profit - Jumps}} \tag{30}$$

The terms $\text{L}_{i,t}\hat{\eta}^J_{i,t|p^J}$ and $\text{L}_{i,t}\hat{\eta}^c_{i,t|p^c}$ are the realized informational losses from supplying liquidity during jump and continuous returns respectively, during a given intraday d. $\text{L}_{i,t}\Delta\hat{s}^J_{i,t|p^J}$ and $\text{L}_{i,t}\Delta\hat{s}^c_{i,t|p^c}$ are realized reversal profits around jump and continuous returns. Expression (29) is a complete characterization of intraday profits from liquidity supplying trades into jump and continuous time periods, (30) estimates realized intraday profits by replacing parameters in (29) with realized estimates. Thereby, decomposing realized informational losses, reversal profits and profits from spreads into components associated with jump and continuous returns. Part of these profits are compensation for holding risky inventory (Stoll [1978]). In continuous markets, inventory risk arises because buy and sell orders from liquidity demander do not arrive at the the same time. Therefore, liquidity suppliers have to hold temporary order imbalances in the form of inventory positions. In the time between holding an initial order imbalance and clearing their position with an arriving liquidity seeker, the permanent price may move against the liquidity supplier's position and therefore liquidity suppliers face an ex-ante risk of making a loss even when not trading with an informed trader. From the SSM, I can estimate intraday inventory risk arising from jump and continuous components of returns, as estimates of intraday conditional permanent return standard deviation, $\hat{\sigma}_{i,y}^{\eta,J}\sqrt{(}N_d^J)$ and $\hat{\sigma}_{i,y}^{\eta,c}\sqrt{(}N_d^c)$. Where, as in section 3.4.2, $N_d^J$ and $N_d^c$ are number of jump and continuous time intervals in a given day. Using these estimates of inventory risk I can compute profits adjusted for holding inventory risk as follows.

$$\text{Risk Adj. Profits}_d^J = \frac{\text{Profits}_d^J}{\hat{\sigma}_{i,y}^{\eta,J}\sqrt{(}N_d^J) + \hat{\sigma}_{i,y}^{\eta,c}\sqrt{(}N_d^c)} \tag{31}$$

$$\text{Risk Adj. Profits}_d^c = \frac{\text{Profits}_c^J}{\hat{\sigma}_{i,y}^{\eta,J}\sqrt{(}N_d^J) + \hat{\sigma}_{i,y}^{\eta,c}\sqrt{(}N_d^c)} \tag{32}$$

Expression (31) and (32) are intraday dollar profits per basis points of inventory risk. These expressions can quantify and capture temporal variation in profits adjusted for inventory risk. For the case when there are no jumps in a given day, Risk Adj. Profits$_d^J$, is zero and all risk adjusted compensation arises from the continuous component Risk Adj. Profits$_d^c$. I use Lee and Ready (1991) tick test to sign direction of market orders and therefore direction of executed limit orders.

Table 3.7 presents results of profit decomposition from the supply of liquidity. From panel A, the average realized profit net of informational losses, across stock-days, is 25,369.71 dollars and the risk adjusted realized profit is 183.14 dollars per basis points of inventory risk. Realized profits from reversals accounts for 8,051 dollars or 31.74 percent of total net profits. Whereas, profits from bid-ask spread are much larger with mean of 48,277.69 dollars. I now compare profit decomposition during jumps and continuous returns. From panel B, the jump contribution to informational loss is 2,089.43 dollars which accounts for 6.75 percent of realized intraday informational loss. To illustrate the magnitude of this loss relative to the rarity of jumps, note that probability of a jump in the sample is 0.38 percent. Therefore, jump contribution to realized information loss is 17.76 times larger relative to the probability of jumps. Similarly, examining the risk adjusted jump contribution to realized informational loss, jump contribution to realized risk adjusted informational loss is 18.87 times larger relative to the rarity of jumps. These results suggest that despite being rare events jumps entail a large realized informational loss for liquidity suppliers. Consequently, adverse selection risk resulting from the jump part of the tails of return distribution is important in relative magnitude for liquidity suppliers. As shown in table 3.5, jumps tend to have an average transitory return size of 32.48 basis points and therefore a large mean reverting component. The estimate of realized jump reversal profit in is 925.55 dollars (panel B of table 3.7) which is 11.50 percent of intraday reversal profits. This jump contribution to reversal profits is 30.26 times larger relative to the probability of jumps and suggest that relative to their rarity jumps are substantially important for reversal profits. These findings complement my results from section 3.4.1, where I show that jumps have large contribution to both the permanent and transitory components of intraday returns, the aforementioned results suggest that large permanent and transitory jump components have strong implications for the provision of liquidity in the form of informational losses and reversal profits, respectively.

I now examine total realized profits net of informational losses during jumps versus continuous returns. The estimated intraday net loss from jumps is 52.70 dollars as compared to a profit of 25,422.41 dollars during continuous returns.

Comparing risk adjusted profits, jumps entail a loss of 1 dollar as compared to a profit of 184.14 dollars per basis point of inventory risk. Since these estimates are across all days in the sample, this suggest that intraday profits result exclusively from supplying liquidity during continuous returns. Consequently, the larger spreads and reversals profits during jump returns are not sufficient to compensate for the substantially larger relative realized informational loss from jumps. In essence, my results suggest that intraday liquidity supply consist of frequent profits from continuous returns and occasional losses from jumps. Given the rarity of jumps, continuous time periods make intraday liquidity provision profitable.

**Table 3.7: Profit Decomposition - Liquidity Supply**
The table reports profit decomposition from supplying liquidity, averaged across stock-days and reported in dollars. The three components of profits are as follow. (1) Informational losses (Info) arises from supplying liquidity against the permanent component of return, (2) Profits from price reversals (Rev), Rev profits occur from the transitory component of return which mean reverts allowing liquidity suppliers to reverse their position at the new favorable price and (3) Profits from quoted spreads (Spread). The column labeled 'Total', reports profits net of informational losses. Risk Adjusted column reports profit decomposition in units of total permanent variance (Panel A), jump permanent variance (Panel B) and continuous permanent variance (Panel C) in units of dollars per basis points variance.

Panel A: All Intervals

| | Profits | | | | Risk Adjusted Profits | | | |
|---|---|---|---|---|---|---|---|---|
| Year | Info | Rev | Spread | Total | Info | Rev | Spread | Total |
| All | -30,959.59 | 8,051.60 | 48,277.69 | 25,369.71 | -168.37 | 58.75 | 292.76 | 183.14 |
| | (194.87) | (77.98) | (212.19) | (179.78) | (0.95) | (0.60) | (1.34) | (1.42) |
| 2008 | -34,420.75 | 9,184.63 | 57,443.80 | 32,207.68 | -132.84 | 48.72 | 260.66 | 176.54 |
| | (301.22) | (123.56) | (312.53) | (288.00) | (1.01) | (0.77) | (1.52) | (1.86) |
| 2010 | -27,366.64 | 6,875.43 | 38,762.57 | 18,271.36 | -205.25 | 69.15 | 326.06 | 189.96 |
| | (242.51) | (93.16) | (272.44) | (201.77) | (1.53) | (0.93) | (2.20) | (2.16) |

| Panel B: Jump Intervals | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Profits | | | | Risk Adjusted Profits | | | |
| Year | Info | Rev | Spread | Total | Info | Rev | Spr | Total |
| All | -2,089.43 | 925.55 | 1,111.18 | -52.70 | -12.07 | 4.90 | 6.17 | -1.00 |
| | (34.60) | (20.91) | (12.31) | (31.68) | (0.19) | (0.11) | (0.06) | (0.17) |
| 2008 | -2,289.68 | 1,149.02 | 1,404.61 | 263.95 | -9.92 | 4.70 | 6.24 | 1.02 |
| | (15.04) | (7.63) | (9.06) | (51.61) | (0.26) | (0.16) | (0.09) | (0.24) |
| 2010 | -1,881.56 | 693.57 | 806.58 | -381.41 | -14.31 | 5.10 | 6.09 | -3.12 |
| | (4.62) | (22.27) | (13.20) | (35.88) | (0.27) | (0.14) | (0.08) | (0.24) |

| Panel C: Continuous Intervals | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Profits | | | | Risk Adjusted | | | |
| Year | Info | Rev | Spread | Total | Info | Rev | Spr | Total |
| All | -28,870.16 | 7,126.05 | 47,166.51 | 25,422.41 | -156.30 | 53.85 | 286.59 | 184.14 |
| | (174.43) | (67.00) | (205.90) | (169.09) | (0.85) | (0.24) | (1.31) | (1.39) |
| 2008 | -32,131.07 | 8,035.61 | 56,039.19 | 31,943.73 | -122.92 | 44.02 | 254.42 | 175.52 |
| | (268.43) | (104.55) | (302.29) | (267.52) | (0.95) | (0.69) | (1.49) | (1.78) |
| 2010 | -25,485.08 | 6,181.86 | 37,955.99 | 18,652.77 | -190.94 | 64.05 | 319.97 | 193.08 |
| | (218.66) | (82.44) | (266.15) | (194.87) | (1.38) | (0.88) | (2.16) | (2.16) |

Before concluding this section I compare realized jump profits during the financial crisis to the relatively tranquil period of 2010. During 2008 realized jump profits are 263.95 dollar as compared to a realized loss -381.41 dollars during 2010, which suggest that liquidity provision during jumps was more profitable in 2008 despite the substantially larger realized informational loss. The identical conclusion can be drawn when comparing risk adjusted profits in panel B of table 3.7. The larger net profits during the crisis follow from 66 percent larger realized jump reversal profits and 74.14 percent larger profits from spreads in 2008 as compared to 2010. This result suggest that the substantially larger realized jump informational cost during the criss was offset by larger profits. To summarize the findings of this section, the results suggest that liquidity suppliers face realized losses during jumps and realized profits during continuous returns. However, since jumps are rare as compared to the frequent continuous returns, intraday provision of liquidity is profitable on average. More generally, the adverse selection cost of jumps resulting from the size of the permanent component is sufficiently large to offset profits from quoted spreads and price reversals.

### 3.5.2 Price Discovery during the 2008 Financial Crisis

In this section I use the state space model (SSM) to address the following question. Did characteristics of intraday information change during the 2008 financial crisis and if so, did this impact price discovery? Price discovery is important for market stability in times of crisis, and can effect monetary policy (Mishkin [2009]), corporate decisions (Fishman and Hagerty [1989]) and the demand for accounting research (Lee [2001]). The 2008 financial crisis is also unique as the only major recession of its magnitude marked by the prevalence of high frequency traders, enhancing the importance of intraday price discovery (Brogaard et al. [2014]). The 2008 financial crisis was also characterized by microstructure events such as the 2008 ban on short sale, which adversely effected price discovery (Boehmer, Jones, and Zhang [2013]; Brogaard, Hendershott, and Riordan [2017]). Despite the importance of intraday price discovery for the 2008 financial crisis, to the best of my knowledge, there is no research which examine differences in price discovery resulting from differences in the characteristics of information revealed to the market, between the 2008 financial crisis and a non-crisis period. By means of a variance decomposition the proposed SSM can accomplish this by breaking down the outcome of price discovery into a jump component associated with high information content and low latency news and a continuous component associated with small and/or incrementally revealed information. Comparing the variance decomposition results during 2008 financial crisis with those of 2010 sample period can highlight any shift in price discovery along the dimensions of latency and size of information content. I further validate my results using real time news data from RavenPack News Analytics. In essence, this application section highlights the differences in price discovery emanating from differences in the characteristics of information between the financial crisis period of 2008 and the non-crisis period of 2010.

I first examine differences in revealed information characteristics between 2008 and 2010 by building a news dataset consisting low latency and high information content news events. As outline in section 3.3, my news dataset is from RavenPack News Analytics which consist of news events that are timestamped to the millisecond and used in real time trading by high frequency traders. Given its low latency, RavenPack news is discretely revealed news. Not all discrete news has high information content, therefore, in addition to relevance and novelty filters outlined in section 3.3, I wish to subset news based on high information content. To do this I use RavenPack's composite sentiment score (CSS) which is based on textual analysis of emotionally charged words and phrases. In addition to textual

**Table 3.8: High Frequency News**
The table reports news counts from RavenPack News Analytics Database. Each news is timestamped to the millisecond and used in real time trading by high frequency traders. I filter news on high novelty and high relevance scores for the sample stocks. In addition, I also filter news so that each news events has high sentiment based on sentiment score assigned by RavenPack Analytics.

| | 2008 | | | 2010 | | |
|---|---|---|---|---|---|---|
| News Group | Negative | Positive | Group Total | Negative | Positive | Group Total |
| Acquisitions/Mergers | 4 | 0 | 4 | 2 | 1 | 3 |
| Analyst Ratings | 400 | 2 | 402 | 361 | 8 | 369 |
| Assets | 4 | 5 | 9 | 3 | 2 | 5 |
| Credit | 19 | 1 | 20 | 17 | 2 | 19 |
| Credit Ratings | 214 | 5 | 219 | 268 | 9 | 277 |
| Dividends | 3 | 3 | 6 | 0 | 6 | 6 |
| Earnings | 471 | 388 | 859 | 245 | 76 | 321 |
| Equity Actions | 57 | 3 | 60 | 54 | 2 | 56 |
| Investor Relations | 0 | 0 | 0 | 0 | 1 | 1 |
| Labour Issues | 24 | 3 | 27 | 20 | 4 | 24 |
| Legal | 15 | 0 | 15 | 6 | 1 | 7 |
| Marketing | 2 | 0 | 2 | 1 | 1 | 2 |
| Partnerships | 3 | 2 | 5 | 0 | 1 | 1 |
| Price Target | 0 | 0 | 0 | 3 | 0 | 3 |
| Products/Services | 37 | 23 | 60 | 18 | 23 | 41 |
| Regulatory | 5 | 0 | 5 | 3 | 1 | 4 |
| Total | 1,258 | 435 | 1,693 | 1,001 | 138 | 1,139 |

analysis, stories are further rated by experts as having effect on share price. CSS can take values between 0 and 100, values larger (smaller) than 50 indicate a positive (negative) sentiment, whereas a value of 50 indicates neutral sentiment. To restrict the news sample to news which represent high information content, I restrict the sample to news events with CSS score of larger than 60 (positive news) or with a score smaller than 40 (negative news).

Table 3.8 presents news counts for the years 2008 and 2010 based on news group and news sentiment. The total number of high frequency and high sentiment news is 1,693 in 2008 as compared to 1,139 in 2010. Therefore, it can be observed that 48 percent more high information content and discretely revealed news events in 2008 as compared to 2010. Looking across news groups, it can be observed that there were 859 news events in 2008 labeled as 'earnings' news, as compared to 321 news events in 2010, or equivalently there were 2.7 times more earning news in 2008 as compared to 2010. Earnings news is most likely to contain high information content about a security's fundamental value and therefore, the aforementioned result suggest that the financial crisis period had a higher proportion of low latency and high information content news.

Did a higher latency and informational content of news during the 2008 financial crisis change the outcome of price discovery? In order to address the aforementioned question, I use estimates of variance contribution from the SSM. From table 3.6, jumps contribute 13.88 percent to permanent price (i.e. efficient price) over a given day in 2008 as compared to 11.80 percent in 2010. This translates to a 17.62 percent larger jump contribution to price discovery in 2008 as compared to 2010. Therefore, the variance decomposition results from the SSM provide evidence to suggest that the shift in information towards high information content and discrete news did translate into more discrete price discovery in 2008 as compared to the 2010 benchmark. Examining continuous price discovery during the financial crisis, it can be observed that continuous signal-to-noise ratios (Cont. SN) in 2008 were 3 times larger as compared to 2010. This suggest that while jumps contributed proportionally more to overall price discovery in 2008 as compared to 2010, continuous price increments were much more informed in 2008 as compared to 2010.

## 3.6 Conclusion

Using a time series representation of intraday price series as distinct jump and continuous components, I examined intraday price discovery in a state space framework. I find that while jumps are extremely rare events which happen with a probability of 0.38 percent, they contribute significantly more to both price discovery and transitory mispricing. The realized estimates of permanent and transitory jump components suggest that not all jumps are equal in their contribution to price discovery. While the median or typical jump is permanent, roughly half of all jumps are highly transitory or highly permanent. Since the dynamics of price discovery through jumps are important for investors, the results suggest that allowing for distinct permanent and transitory jumps in risk management and asset pricing models can be an important avenue of future research. In particular, the proposed framework in this chapter can be used to allow for distinct jump CAPM betas arising from co-movement in permanent and transitory jumps with the corresponding market portfolio jump components. In addition, extending the work of Easley, Hvidkjaer, and O'hara [2002] and O'Hara [2003]) the proposed framework can examine the jump risk of price discovery in the cross section of assets along with asset pricing implications.

Using a profit decomposition for the provision of liquidity in the jump-continuous domain, I found that jump entail a large informational risk for liquidity suppliers. Overall, liquidity suppliers face realized net losses during jumps which are compensated by supplying liquidity during continuous price movements. Since jumps are rare, the provision of intraday liquidity is, on average, profitable.

# Chapter 4

# Intraday Liquidity Fragmentation & Price Jumps

This chapter examines the relationship between liquidity fragmentation and price jumps. Unexpected changes in intraday liquidity fragmentation predict jumps and jump direction. A shock to ask (bid) side liquidity fragmentation increases the probability of positive (negative) jumps by 36%. Decomposing jumps into information and noise components we show that fragmented jumps are noisier. Our work suggests that liquidity suppliers predict jumps and actively manage their exposure to large order imbalances accompanying jumps by fragmenting liquidity. This makes jumps predictable as liquidity suppliers' information is reflected in liquidity fragmentation, minutes before the arrival of a jump.

## 4.1 Introduction and Literature Review

Before and during the Flash Crash, liquidity supply was fragmented across more than 50 exchanges and execution venues.[1] There is little consensus on the sources

[1]See Thomson Reuters Business News - https://www.reuters.com/article/idINIndia-49218020100610

and causes of extreme price movements (jumps) in modern financial markets. Understanding the dynamics of jumps is important as they can affect the efficient functioning of markets by impairing portfolio management, risk management and option-pricing (e.g. Bollerslev and Todorov [2011]; Bollerslev et al. [2016]; Bégin et al. [2019]). Despite their importance, we know little about the source of the these price movements and we know even less about how to predict their occurrence. This chapter does not take a stand on the source of jumps but identifies a predictable component of jumps related to liquidity suppliers' information. Further, we apply a signal extraction methodology proposed in Chapter 3 to identify components of price movements related to information and noise, and show that the ratio of information to noise is partially predictable using a proxy for liquidity suppliers' information. Consistent with Jeon and Zhao [2019]; we find that jumps in large and liquid S&P 100 stocks are mostly permanent and therefore related to the arrival of new valuation relevant information. Nevertheless, the average jump return has a 7.44% transitory component.[2] This means that on average markets overreact to positive and negative information arrivals for even the highly liquid S&P 100 stocks.

Modern equity markets are fragmented, with quoting and trading regularly occurring on a multitude of exchanges, trading venues, and broker-dealer platforms (O'Hara and Ye [2011]).[3] The fragmented nature of equity markets can lead to increased fragility, a focus of This chapter, and complexity as market participants attempt to coordinate liquidity supply and demand in real-time (Menkveld and Yueshen [2018]). Often the liquidity demand exceeds the liquidity supply and this

[2]Or equivalently the signal-to-noise ratio (SN ratio) is 12.44

[3]See SEC Public Statement - https://www.sec.gov/news/statement/us-equity-market-structure.html

mismatch can lead to large transitory jump components as liquidity demanding orders consume all the available liquidity in a market, pushing prices far away from the fundamental value.

In This chapter we study the relationship between liquidity fragmentation and jumps. In general, fragmentation has a large persistent component. For instance, unexpected changes in fragmentation persist for at least 30 minutes, suggesting that fragmentation is predictable. We use unanticipated changes in fragmentation (innovation) to explain the arrival and direction of jumps 1-minute into the future. In an instrumental variable (IV) probit model where we control for overall market conditions and account for omitted variables that are jointly determined with fragmentation, we find that a one standard deviation increase in liquidity fragmentation on the ask (bid) side increases the probability of positive (negative) jumps by 36% points. Therefore, liquidity fragmentation can predict future jumps and their direction. Our model can correctly predict 28% of all within sample jumps and their direction.[4] Our paper is the first to show that liquidity fragmentation is informative for jumps and jump direction.

Liquidity suppliers are traditionally closest to the trading mechanism. They view trading in real-time and accumulate inventory (shares) in the opposite direction of the aggregate market.[5] As liquidity suppliers accumulate inventory and otherwise observe trading they may also acquire information about future demand and supply imbalances or about future information arrivals (Harris and

[4]We compute a measure known as recall in the maching learning literature. For binary jump event, *recall* is the proportion of all within sample jumps which are correctly predicted. Where we define a jump as correctly predicted, if the time t-1 ask (bid) side fragmentation innovation predicts a time t jump probability of greater than 50% for a corresponding positive (negative) jump.

[5]We defined liquidity supplier as any trader that post a limit order.

Panchapagesan [2005]). They will condition their liquidity supply and manage risk based on this information. One way for liquidity suppliers to manage their risk is to reduce the likelihood of executing against a large order. They do this by reducing the total depth offered and also by reducing the maximum order size on all markets. The latter leads to fragmented liquidity supply. In order to capture the dynamics of information flow for liquidity suppliers our research uses ex-ante offered liquidity to measure market fragmentation. This approach directly approximates pre-trade information of liquidity suppliers allowing us to examine the relationship between fragmentation and price jumps arising from the information of liquidity suppliers.

To further highlight the importance of fragmentation in managing inventory risk, note that liquidity suppliers have to manage their exposure to both fast and slow traders (Van Kervel [2015]; and Foucault and Menkveld [2008]). Fast traders use technology to monitor and simultaneously route orders to multiple trading venues known as smart order routing (SOR). These traders can access liquidity across multiple venues as if they were a consolidated liquidity pool. While the exact fraction of SOR traders is not known it is estimated to be less than one third of all traders.[6] Unlike fast traders, slow (non-SOR) traders can only access liquidity in multiple venues sequentially. When a substantial fraction of traders are slow, fragmented liquidity can lead to a mismatch between the supply and demand for liquidity during times of extreme price movements. This is because liquidity suppliers have the opportunity to cancel quotes in competing trading venues before slow traders can access them. As an example, take the case of

[6]For European markets Foucault and Menkveld [2008] estimate that 27% of traders use SOR. In a more recent study Van Kervel [2015] estimates the proportions to be 20%.

one liquidity supplier and two trading venues with the liquidity supplier willing to quote 1,000 shares in total. In times leading up to a jump, the liquidity supplier may anticipate a surge of immediacy demand and fragment liquidity equally across both venues by offering to trade 500 shares on each market. When a slow trader sweeps the the first venue by consuming all 500 offered shares, the liquidity supplier may strategically cancel her quote on the second venue before the slow trader can access liquidity at the second venue.[7] As a result, prices are likely to overshoot and therefore become more noisy relative to their information content. This example highlights one reason why fragmented liquidity can lead to a mismatch between the demand and supply of liquidity and how the time-series of fragmentation may provide insight into the information sets of liquidity suppliers.

Since jumps are characterised by large and quick price movement, they are accompanied by a surge in market activity.[8] In particular, jumps are accompanied by a sudden increase in directional demand for immediacy as investors manage their portfolio and risk.[9] This surge in directional trading poses risk for liquidity suppliers as they accumulate large net inventory position in volatile times.[10] Liquidity suppliers will therefore manage their exposure to large expected directional orderflow by fragmenting liquidity across numerous venues. Therefore, liquidity fragmentation should be a reasonable proxy for liquidity suppliers' information.

[7] We find that in times leading up to jumps market orders which consume all available liquidity are disproportional larger than the number quote updates.

[8] By their definition jumps are discrete price moves and therefore of higher frequency then continuous returns.

[9] The average volume based orderflow in the direction of returns is 0.67 million during jumps; as compared to 0.16 million during continuous price moves; see table 4.5

[10] Jumps are often followed by large volatility spikes e.g. Todorov and Tauchen [2012]

Further, if liquidity suppliers are better informed than the general public, it follows that unanticipated shocks to liquidity fragmentation predict jumps. The degree to which they fragment liquidity and the corresponding direction of fragmentation will be based on their information. Natural investors and the trading public only view transactions and cannot reasonably approximate net-buying or selling by liquidity suppliers.

Our work contributes to growing literature examining the effects of market fragmentation on market quality. O'Hara and Ye [2011] find that fragmentation in US equity markets lowered transaction cost and short term return volatility thereby improving market quality. Using the entrance of Euronext in the European equity market as an exogenous fragmentation shock, Foucault and Menkveld [2008] show that ex-post liquidity improved for a set of Dutch stocks. Madhavan [2012] finds that fragmented stocks were disproportionately affected during the flash crash. Our work contributes to this literature by examining the effect of high frequency innovation to fragmentation on market quality during jumps. To the best of our knowledge ours is the first paper to show that innovation to liquidity fragmentation are informative for jumps and their direction. In addition, we show that jumps which follow fragmented liquidity are more noisier than the average jump. Our research complements the work of Van Kervel [2015] who shows that liquidity suppliers manage adverse selection risk by cancelling orders on other exchanges after the arrival of a trade. Our research adds to Van Kervel [2015] by showing that US equity markets are most fragmented precisely when liquidity suppliers anticipate a large directional demand; i.e. high frequency unanticipated changes in liquidity fragmentation are strategic. In a recent paper Brogaard, Hendershott, and Riordan [2019] show that limit orders predict future price movements and transmission of information across markets. We complement this study by showing that the dispersion of liquidity in limit orders predict large price movements.

More generally, we show that market linkages are important, and are particularly important during extreme price movements. Our work also suggest that market fragmentation could lead to noisier prices then a consolidated market. Future work could explore the welfare consequences of exogenous variation in fragmentation.

The rest of the paper is organized as follows. In section 4.2 we present our data and sample selection. Section 4.3 outlines fragmentation and liquidity measures of This chapter. Section 4.4 presents jump detection and decomposition methodologies. In section 4.5, we present a preliminary analysis of the link between fragmentation innovation and jumps and introduce our tests examining the effect of fragmentation innovation for jumps. In sections 4.6, 4.7 and 4.8 we perform our tests and discuss the implications. Section 4.9 presents our conclusion.

## 4.2 Data and Market Shares in Liquidity

### 4.2.1 Data

Our sample time period begins on January 1$^{st}$, 2010 and ends on December 31$^{st}$, 2017. The dataset consist of stocks listed on the S&P 100 index in a given sample year along with the S&P 500 ETF (ticker: SPY) which proxies for the market portfolio. Since stock listing on an index is not necessarily continuous, the set of sample stocks is dynamic across years. On average, 7 percent of the stocks in our sample change each year. To avoid spurious jumps associated with stock splits, we remove a given stock during the sample year in which the split has taken place. This leaves us with a dataset consisting of 128 unique stocks and 768 stock-years. Using the signal extraction methodology proposed in Chapter 3 we estimate permanent and transitory jump components for each stock-year in our dataset with an average convergence rate of 84 percent across stock-years. This concludes our sample selection with 123 unique stocks spread across 606 stock-years.

We define intraday returns as the set of 1 minute logarithmic price moves between 9:32 and 15:59. We remove the first and last minute of the trading day to avoid the opening and closing batch auction. An additional minute at the open is removed to ensure that the first price move of the day corresponds to at least 2 minutes of within-day of prior. Table 4.1 reports descriptive statistics for our sample stocks. Of the 123 stocks in our sample, 96 have primary listing on the NYSE and the remaining 27 have primary listing at NASDAQ. The sample consist of large and liquid stocks, with an average market capitalization (MCAP) of 77.44 billion USD and average daily traded volumes of 475.51 million USD.

Our primary data consist of (i)consolidated quotes, (ii) consolidated NBBO quotes, and (iii) consolidated trades files from NYSE's Trade and Quotation (TAQ) dataset. We use the methodology proposed in Holden and Jacobsen [2014] to account for withdrawn quotes. Time stamped to the millisecond, the TAQ dataset uniquely identifies all quotes and trades originating from any NMS participant.[11] We further augment our dataset by matching TAQ data to Center for Research in Security Prices (CRSP) dataset to obtain daily price, dividend adjusted returns and primary listing information.

[11] All major national and regional exchanges are national market system (NMS) participants.

**Table 4.1: Descriptive Statistics: Sample Stocks**
This table reports descriptive statistics for our sample stocks. MCAP is the average market capitalization computed as price times shares outstanding and reported in billion USD. Price is the daily closing price. Volume is daily traded volume reported in millions. Quoted spread is the daily closing ask price minus bid price relative to the close midpoint. Return size (Ret. Size) is the dividend adjusted size of daily return.

| Stock Descriptive Statistics | | | | | | |
|---|---|---|---|---|---|---|
| Sample Stocks: S&P 100 | | | | | | |
| Number of Stocks: 123 | | | | | | |
| NYSE Listed: 96 | | | | | | |
| NASDAQ Listed: 27 | | | | | | |
| Sample Years: 2010-2017 | | | | | | |
| | | | | Percentiles | | |
| Name | Units | Mean | Std. Dev | $25^{th}$ | $50^{th}$ | $75^{th}$ |
| MCAP (*Size*) | $ B. | 77.44 | 67.81 | 33.44 | 57.61 | 95.12 |
| Price | $ | 80.40 | 139.18 | 34.43 | 55.07 | 79.68 |
| Quoted Spread | % Bps | 3.60 | 2.30 | 2.24 | 2.98 | 4.05 |
| Ret. Size | % Bps | 128.18 | 52.42 | 94.62 | 116.58 | 148.14 |
| Shares Outstanding | # M. | 1,533.04 | 1,781.77 | 492.29 | 983.28 | 1,705.67 |
| Volume | $M. | 475.51 | 381.39 | 240.41 | 349.28 | 575.70 |

### 4.2.2 Market Shares in Liquidity

We define market share in intraday liquidity as the proportion of ask or bid depth offered at NBBO prices during a 1-minute time interval.[12] Table 4.2 reports liquidity shares of the NMS trading platforms for the S&P 100 sample stocks. NASDAQ OMX has the largest share with a 35.58 percent share in intraday liquidity. NYSE falls closely with 28.89 percent share and NYSE Arca follows with a share of 19.48 percent. The two major exchanges NYSE and NASDAQ have a total share of approximated 65 percent leaving the remaining 35 percent split across the remaining NMS trading venues. Considering that all of the sample

[12]National best bid and offer (NBBO) quote price are the lowest (highest) ask (bid) price across all NMS participants for a given security and a given time.

stocks have primary listing on NYSE or NASDAQ, these market shares represent a significant degree of fragmentation, in particular when compared to years prior to 2007 when NYSE alone was executing 79% of volume in its listed stocks.[13][14]

**Table 4.2: Intraday Liquidity - National Market System (NMS)**
This table reports the percentage share in intraday liquidity for all national market system (NMS) platforms. We define intraday liquidity as the number of shares offered by the platform at the national best bid or offer quote price (NBBO).

| TAQ Flag | Platform | Market Share (%) |
|---|---|---|
| A | NYSE MKT LLC / AMEX | 0.01 |
| B | NASDAQ OMX BX, Inc. | 0.10 |
| C | National Stock Exchange Inc. (NSX) | 0.46 |
| D | FINRA Alternative Display (FINRA ADF) | 0.05 |
| I | International Securities Exchange, LLC (ISE) | 2.70 |
| J | Direct Edge A Stock Exchange, Inc. | 0.54 |
| K | Direct Edge X Stock Exchange, Inc. | 5.56 |
| M | Chicago Stock Exchange, Inc. (CHX) | 0.02 |
| N | New York Stock Exchange | 28.89 |
| P | NYSE Arca, Inc. | 19.48 |
| T/Q | NASDAQ Stock Exchange | 35.58 |
| W | CBOE Stock Exchange | 0.02 |
| X | NASDAQ OMX PSX Stock Exchange | 2.56 |
| Y | BATS BYX Exchange, Inc. | 0.07 |
| Z | BATS BZX Exchange, Inc. | 3.94 |

## 4.3 Liquidity and Fragmentation Measures

In this section we outline our measures of intraday liquidity and liquidity fragmentation.

[13] Regulation national market system (NMS) was implemented in the year 2007.
[14] See SEC Litreature Review Regulation NMS - www.sec.gov/marketstructure/research/fragmentation-lit-review-100713.pdf

### 4.3.1 Intraday Liquidity

*Market Depth*

Trading against a limit order that is strictly worse in price than the NBBO quote is prohibited by the Securities and Exchange Commission (SEC), therefore of particular importance for US markets is the liquidity available at NBBO quotes.[15] Amongst all limit orders NBBO quotes have the largest contribution to price discovery (Brogaard et al. [2019]). If there is insufficient liquidity available at NBBO quotes, then traders have to search for liquidity across time and venues. Hence, motivated by the importance of NBBO quotes, we define share based ask (bid) depth as the aggregate number of shares available on the ask (bid) side of the NBBO within a given intraday time interval. Market depth is then defined as the average of ask and bid side depth. Denote $i = 1 \ldots I$ for the set of stocks, $p = 1 \ldots P$ for the set of exchanges and $Q_{i,m,p}$ as the consolidated quote for stock i, at exchange p for a given millisecond m contained in the intraday interval $[t-1,\ t)$; it follows:

$$\begin{aligned} \text{Best Depth Ask}_{i,m} &= \sum_{p=1}^{P} \big(S(\mathrm{Q}_{i,p,m}^{Ask})\big)\ \mathcal{I}_{Q_{i,p,m}^{\text{NBBO, Ask}}} \\ \text{Best Depth Bid}_{i,m} &= \sum_{p=1}^{P} \big(S(\mathrm{Q}_{i,k,p}^{Bid})\big)\ \mathcal{I}_{Q_{i,k,m}^{\text{NBBO, Bid}}} \end{aligned} \tag{4.1}$$

aggregating across each millisecond m in interval $[t-1,\ t)$

[15] See SEC statement on trade-throughs - https://www.sec.gov/rules/proposed/34-49325.htm#III

$$\begin{aligned} \text{Depth Ask}_{i,t} &= \sum_{m \in [t-1,\ t)} \text{Best depth Ask}_{i,m} \\ \text{Depth Bid}_{i,t} &= \sum_{m \in [t-1,\ t)} \text{Best depth Bid}_{i,m} \end{aligned} \tag{4.2}$$

where $\mathcal{I}_{Q_{i,p,m}^{\text{NBBO, Ask}}}$ $\mathcal{I}_{Q_{i,p,m}^{\text{NBBO, Bid}}}$ are indicator function which equal 1 if consolidated quote at venue p is the national best ask (bid) quote and zero otherwise. As in Holden and Jacobsen [2014] we a also construct time weighted volume based market depth, $\text{depth}_{i,t}^{V}$. The aforementioned measure captures both the effect of share size and offered price and serves as a control variable in our following models.

*Liquidity Supply Demand Mismatch (SDM)*

The demand and supply of liquidity is a dynamic matching problem. Liquidity suppliers' provide trading options to the demand side by posting limit orders. As these limit orders are consumed by the demand side in the form of market orders, new limit orders are continuously posted to replenish consumed liquidity. However, if market orders are consuming liquidity at a faster rate than it is replenished by quote updates then there is higher demand for immediacy as compared to liquidity supply. In particular, there exists a supply and demand mismatch (SDM). In order to measure the degree of liquidity SDM we compute a variable which measures the number of trades of size larger than the prevailing liquidity offered at the NBBO relative to the number of quote updates to the NBBO. Denote $S_{i,l,t}$ as the size of l-th trade in intraday interval $[t-1,\ t)$, we define the number of large sized trades as follows:

$$\text{Trades}_{i,t}^{NBBO} = \sum_{l=1}^{L} \mathcal{I}_{[S_{i,l,t} \geq Dept_{i,k,t}^{NBBO}]} \tag{4.3}$$

where $\mathcal{I}_{[S_{i,k,t} \geq Dept_{i,k,t}^{NBBO}]}$ is an indicator variable which equals 1 if the l-th buy (sell) trade has size larger than the offered ask (bid) depth. $\text{Trades}_{i,t}^{NBBO}$ is a count variable of trades which consume all the liquidity available at NBBO quotes. Correspondingly we compute NBBO quote updates as follows:

$$\text{Quotes}_{i,t}^{NBBO} = \sum_{k=1}^{K} \mathcal{I}_{Q_{i,k,t}^{\text{NBBO}}} \tag{4.4}$$

$\text{Quotes}_{i,t}^{NBBO}$ is the liquidity supply side counterpart of $\text{Trades}_{i,t}^{NBBO}$ which counts the number of consolidated quote updates to NBBO. The ratio of these two variables is our SDM measure:

$$\text{SDM}_{i,t} = \frac{\text{Trades}_{i,t}^{NBBO}}{\text{Quotes}_{i,t}^{NBBO}} \tag{4.5}$$

the intuition behind the SDM variable is as follows: when SDM is large for a given intraday time interval there are more liquidity consuming trades relative to liquidity replenishing orders. Therefore, temporal variation in SDM captures the changes in the degree of mismatch between the supply and demand for liquidity.

*Effective Spread*

Effective spread (ES) is the difference between the trade price and the prevailing mid quote. Effective spreads can be interpreted as the total cost of trading which include liquidity supplier profits and the information component of the trade. Denoting $l = 1 \ldots L$ as the total number of trades, ES is computed as follows:

$$\mathrm{ES}_{i,t} = \frac{1}{L}\sum_{l=1}^{L} D_{i,l}\frac{p_{i,l} - M_{i,l}}{M_{i,l}} \quad (4.6)$$

where $D_l$ is an indicator for trade direction, which equals 1 (-1) for a buy (sell) trade. Intraday effective spread is the average cost of trading in a given interval, where the average is computed over all trades in a given intraday interval.

*Price Impact of Trade*

The effective spread can be decomposed into two components; (i) liquidity supplier profits and (ii) price impact (PI). The latter represents the information cost of trade. We define intraday price impact as follows:

$$\mathrm{PI}_{i,t} = \frac{1}{L}\sum_{l=1}^{L} D_{i,l}\frac{M_{i,l,m} - M_{i,l,m+60000}}{M_{i,l,m}} \quad (4.7)$$

where $M_{l,m+60000}$ is the prevailing mid quote 1 minute (or 60,000 milliseconds) after the l-th trade. The price impact can be interpreted as the information content of a trade. If a trade is highly informative than the midpoint will have a large change in the direction of the trade which will persist 1 minute into the future. Since the effective spread is the sum of the price impact and the realized spread, by including the former two in a linear model, we also implicitly also account for temporal variation in realized spreads.

Panel A of table 4.3 reports descriptive statistics for intraday liquidity variables. As expected the S&P 500 ETF is far more liquid than the individual sample stocks. For the sample stocks, the average depth is 0.64 million shares, which is approximately one third of the depth available for the S&P 500 ETF of 2.33 million shares. On average, a given trade has an effective spread and price impact of

2 basis points. Correspondingly, the S&P 500 ETF trades has an effective of also 1.6 basis points with a price impact of 0.51 basis points. On average, there are 272 quote updates for the sample stocks. These updates are newly added quotes at a given exchange which offer either price and/or size improvement relative to the prevailing NBBO quotes at the time of submission. The average SDM is 0.09 which suggests that on average the number of NBBO quote revisions outnumber trades. This further highlights that our sample consists of liquid stocks which trade at high frequencies.

### 4.3.2 Fragmentation

We define intraday liquidity fragmentation as the degree to which visible liquidity is dispersed across NMS trading venues. To proxy for the level of consolidated liquidity we construct a Herfindahl-Hirschman (HHI) index based on the aggregate depth available in a given intraday time interval. Formally, denote $p = 1 \dots P$ as the number of trading platforms; the bid and ask side market share (MS) of platform p in time interval t is as follows:

$$\begin{aligned} \mathrm{MS}^{Ask}_{i,p,t} &= \frac{\text{Best depth Ask}_{i,p,t}}{\sum_{i=1}^{P} \text{Best depth Ask}_{i,p,t}} \\ \mathrm{MS}^{Bid}_{i,p,t} &= \frac{\text{Best depth Bid}_{i,p,t}}{\sum_{i=1}^{P} \text{Best depth Bid}_{i,p,t}}. \end{aligned} \tag{4.8}$$

$\mathrm{MS}_{i,p,t}$ is a proportion between zero and one which measures the the share of p-th exchange in offered depth, in a given interval t. Using intraday market shares we construct HHI for each 1 minute time interval t as follows:

$$\begin{aligned}\text{HHI}_{i,t}^{Ask} &= \sum_{p=1}^{P} (\text{MS}_{i,p,t}^{Ask})^2 \\ \text{HHI}_{i,t}^{Bid} &= \sum_{p=1}^{P} (\text{MS}_{i,p,t}^{Bid})^2\end{aligned} \tag{4.9}$$

HHI is bounded between one and $\frac{1}{P}$. Where one corresponds to completely consolidated markets and $\frac{1}{P}$ corresponds to liquidity equally fragmented across the P platforms. From HHI we construct a measure of fragmentation, $F_t$ such that when liquidity is consolidated on a single platform $F_t$ takes on a value of zero.

$$\begin{aligned}F_{i,t}^{Ask} &= 1 - \text{HHI}_{i,t}^{Ask} \\ F_{i,t}^{Bid} &= 1 - \text{HHI}_{i,t}^{Bid}\end{aligned} \tag{4.10}$$

In order to model the relationship between jumps and liquidity fragmentation, we first note that if liquidity fragments because of liquidity suppliers' information, then it is the unanticipated component of fragmentation; $\big($i.e. $F_{i,t} - \mathbb{E}_{t-1}[F_{i,t}]\big)$ which affects time t price. This is because the expected component, $\mathbb{E}_{t-1}[F_t]$ is already incorporated in market participants' time t-1 information set, and therefore so is its effect on price. We model the unanticipated component of liquidity fragmentation in two steps. First, for each stock-year combination, we demean liquidity fragmentation. Demeaning fragmentation removes both stock fixed and year fixed effects. Second, for each stock-year in our sample we estimate an autoregressive model for both the bid and ask side of demeaned fragmentation as follows:

**Table 4.3: Descriptive Statistics: Intraday Liquidity and Stock-Year Controls**
This table reports descriptive statistics for intraday liquidity (panel A), intraday liquidity fragmentation (panel B) and stock controls (panel C). Dept is the average of ask and bid side shares offered; reported in millions. Average dept ($Depth^V$) is the time weighted average of ask side and bid side offered volume. Effective spread ($ES$) is the signed difference between price and prevailing midpoint, measured relative to the prevailing midpoint at the time of trade. Price impact ($PI$) is the signed changed in midpoint following a trade, relative to prevailing midpoint at time of trade. Fragmentation innovation ($\tilde{F}^D$) is the unanticipated component of liquidity fragmentation estimated using an autoregressive model of order p, where p is determined by AIC criterion. NBBO quotes are the number of quote revision to the national best bid and offer price. Orderflow is the number of trades in the direction of price move. Orderflow$^{\mathrm{V}}$ is the traded volume in the direction of price move reported in millions. Supply and demand mismatch (SDM) is computed as the number of trades larger than NBBO offered liquidity relative to number NBBO quote updates. Return size (Ret. Size) is the size of return computed as the absolute value of returns. Trades are the count of executed trades. Volume is the dollar traded volume reported in millions. The SPY superscript denotes S&P 500 ETF variables which proxies for systematic liquidity.

Panel A: Intraday Liquidity

| | | | | Percentiles | | |
|---|---|---|---|---|---|---|
| Name | Units | Mean | Std. Dev | $25^{\mathrm{th}}$ | $50^{\mathrm{th}}$ | $75^{\mathrm{th}}$ |
| Depth ($Depth$) | # M./1min | 0.64 | 1.61 | 0.04 | 0.12 | 0.44 |
| Depth SPY ($Depth^{SPY}$) | # M./1min | 2.33 | 2.81 | 0.66 | 1.50 | 3.01 |
| Avg Depth ($Depth^V$) | $ M./1min | 0.07 | 0.08 | 0.03 | 0.04 | 0.08 |
| Avg Depth SPY ($Depth^{V,SPY}$) | $ M./1min | 1.74 | 0.99 | 1.12 | 1.49 | 2.14 |
| Eff. Spread ($ES$) | % Bps./1min | 2.23 | 1.46 | 1.32 | 1.85 | 2.67 |
| Eff. Spread SPY ($ES^{SPY}$) | % Bps./1min | 1.58 | 6.49 | 0.47 | 0.60 | 0.82 |
| NBBO Quotes | #/1min | 271.61 | 290.21 | 90.57 | 175.71 | 336.20 |
| Orderflow | #/1min | 16.33 | 36.89 | −1.43 | 9.06 | 26.19 |
| Orderflow$^{\mathrm{V}}$ | $ M./1min | 0.17 | 0.51 | −0.03 | 0.08 | 0.26 |
| Prc. Impact ($PI$) | % Bps./1min | 1.94 | 3.76 | 0.00 | 1.26 | 3.23 |
| Prc. Impact SPY ($PI^{SPY}$) | % Bps./1min | 0.51 | 1.43 | −0.16 | 0.35 | 1.04 |
| SD Mistmatch($SDM$) | Ratio×100/1min | 8.88 | 10.88 | 1.48 | 4.93 | 12.12 |
| Trades | #/1min | 108.71 | 124.21 | 34.39 | 67.66 | 131.16 |
| Volume | $ M./1min | 1.13 | 1.57 | 0.29 | 0.61 | 1.27 |

Panel B: Intraday Liquidity Fragmentation

| | | | | Percentiles | | |
|---|---|---|---|---|---|---|
| Name | Units | Mean | Std. Dev | $25^{\mathrm{th}}$ | $50^{\mathrm{th}}$ | $75^{\mathrm{th}}$ |
| Frag. ($F$) | Prop./1min | 0.57 | 0.18 | 0.49 | 0.62 | 0.69 |

Table Continued

| Name | Units | Mean | Std. Dev | Percentiles 25th | 50th | 75th |
|---|---|---|---|---|---|---|
| Frag. Inno ($\tilde{F}^D$ | Prop./1min | 0.00 | 0.12 | −0.05 | 0.02 | 0.08 |
| IV Frag. ($F^{IV}$) | Prop./1min | 0.55 | 0.06 | 0.51 | 0.56 | 0.60 |
| IV Frag. Inno ($\tilde{F}^{IV,D}$) | Prop./1min | 0.00 | 0.03 | −0.02 | 0.00 | 0.02 |

Panel C: Stock Controls

| Name | Units | Mean | Std. Dev | Percentiles 25th | 50th | 75th |
|---|---|---|---|---|---|---|
| Lograthmic Size (log($MCAP$)) | Log/Stock-year | 17.95 | 0.81 | 17.40 | 17.86 | 18.56 |
| Quoted Spread ($QS$) | % Bps./Stock-year | 2.81 | 1.77 | 1.53 | 2.23 | 3.29 |
| Volatility ($Vol.$) | $ /Stock-year | 0.02 | 0.01 | 0.01 | 0.01 | 0.02 |

$$\begin{aligned} F_{i,t}^{Ask} &= \alpha_1 F_{i,t-1}^{Ask} + \alpha_2 F_{i,t-2}^{Ask} + \cdots + \alpha_p F_{i,t-p}^{Ask} + \tilde{F}_{i,t}^{Ask} \\ F_{i,t}^{Bid} &= \beta_1 F_{i,t-1}^{Bid} + \beta_2 F_{i,t-2}^{Bid} + \cdots + \beta_p F_{i,t-p}^{Bid} + \tilde{F}_{i,t}^{Bid} \end{aligned} \tag{4.11}$$

where the optimal lag length p is determined using the BIC criterion. The innovation component of intraday fragmentation $\tilde{F}_{i,t}^{Ask}$ and $\tilde{F}_{i,t}^{Bid}$ are the estimated residual components from the autoregressive model. Next, we examine the degree of persistence; i.e. predictable component of fragmentation. Figure 4.3.2 plots the average autocorrelation function for up to 30 minutes post innovation. From the plots we note that the first order autocorrelation is 0.38. Which suggests that 38 percent of the innovation to fragmentation persists 1 minute into the future. More interestingly, the autocorrelation 30 minutes following the innovation is 0.18. This highlights that approximately 18 percent of the original innovation persists

30 minutes into the future. These statistics reveal a high degree of persistence and therefore predictability in fragmentation. Table 4.4 reports results from estimates (4.11) using conditional least squares. Consistent with the aforementioned acf

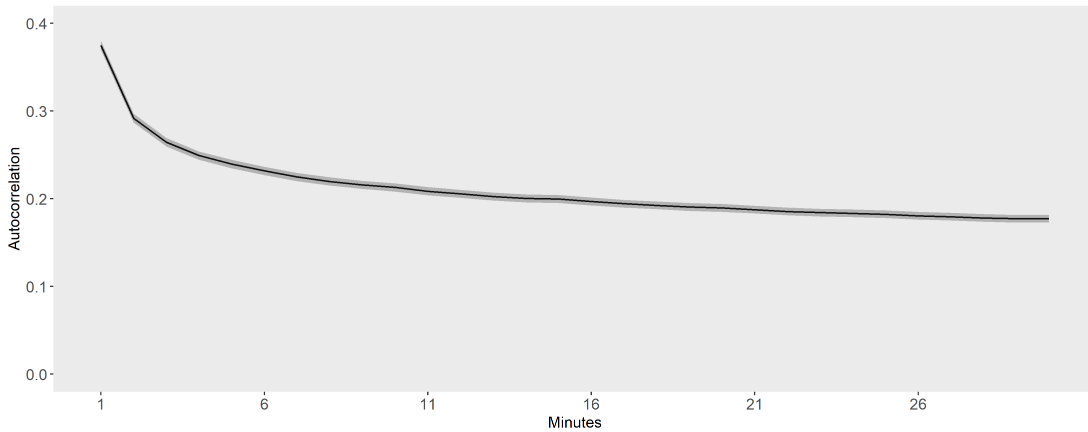
**Table 4.4: Fragmentation - Autoregressive Model**
This table reports estimates from the autoregressive (AR) model given in (4.11). Each stock-year is independently estimated. Order selected is the average AR order p, determined by minimizing the Bayesian information criterion (BIC) criterion. We also report the average of the first five estimated coefficient across stock-year models. White Noise (WN) test is the Ljung-Box for estimated residuals performed on the first 24 lags.

Panel A: Ask Size

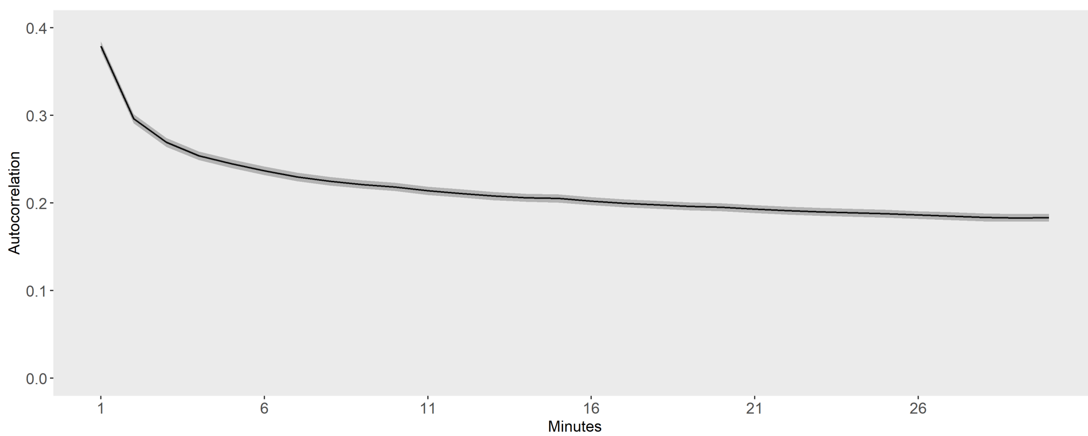
| Variable | Units | Mean | T-statistic | P25 | Median | P75 |
|---|---|---|---|---|---|---|
| Order Selected | # lags | 36.69 | 139.04 | 32.00 | 36.00 | 41.00 |
| Order 1 | Correlation | 0.23 | 131.87 | 0.20 | 0.24 | 0.26 |
| Order 2 | Correlation | 0.08 | 176.75 | 0.08 | 0.08 | 0.09 |
| Order 3 | Correlation | 0.06 | 175.57 | 0.05 | 0.06 | 0.06 |
| Order 4 | Correlation | 0.04 | 180.76 | 0.04 | 0.04 | 0.05 |
| Order 5 | Correlation | 0.04 | 178.43 | 0.03 | 0.04 | 0.04 |
| WN Test | P-value | 0.96 | 131.72 | 0.99 | 1.00 | 1.00 |

Panel B: Bid Size

| Variable | Units | Mean | T-statistic | P25 | Median | P75 |
|---|---|---|---|---|---|---|
| Order Selected | # lags | 37.49 | 135.49 | 32.00 | 37.00 | 41.00 |
| Order 1 | Correlation | 0.23 | 129.67 | 0.20 | 0.24 | 0.26 |
| Order 2 | Correlation | 0.08 | 173.91 | 0.07 | 0.08 | 0.09 |
| Order 3 | Correlation | 0.06 | 174.62 | 0.05 | 0.06 | 0.06 |
| Order 4 | Correlation | 0.04 | 177.04 | 0.04 | 0.04 | 0.05 |
| Order 5 | Correlation | 0.04 | 183.77 | 0.03 | 0.04 | 0.04 |
| WN Test | P-value | 0.96 | 134.10 | 1.00 | 1.00 | 1.00 |




**Panel A: Ask Side Autocorrelation Function**




**Panel B: Bid Side Autocorrelation Function**

**Figure 4.1: Autocorrelation Function - Liquidity Fragmentation**
This figure plots the average autocorrelation function for liquidity fragmentation. Panel A is a plot of ask side autocorrelation function and panel B corresponds to the bid side. Averages are computed across stock-year model. The shaded region shows the confidence intervals for the mean autocorrelation for each lag.

plots, the BIC criterion selects a lag length of 37 minutes for the bid and ask side models. The average p-value for Ljung-Box test for up to 12 lags is 0.96 suggesting that the autoregressive model adequately captures the predictable component of each side of fragmentation.

Our objective is to capture the degree to which the size and direction of fragmentation predict jumps and jump direction, we therefore define directional liquidity fragmentation as follows:

$$\tilde{F}_t^D = \left\{ \begin{array}{ll} \tilde{F}_t^{Ask}, & \Delta p_{t+1} > 0 \\ 1/2\left(\tilde{F}_t^{Ask} + \tilde{F}_t^{Bid}\right), & \Delta p_{t+1} = 0 \\ \tilde{F}_t^{Bid}, & \Delta p_{t+1} < 0 \end{array} \right\} \quad (4.12)$$

Since a strategic liquidity supplier will fragment liquidity in the same direction as the anticipated price move; i.e. ask (bid) side for an anticipated positive (negative) price move, $\tilde{F}_t^D$ relates both the size and direction of fragmentation to the size and direction of future price moves.

Panel B of table 4.3 reports descriptive statistics for liquidity fragmentation measures. The average fragmentation is 0.57, which suggests our sample stocks are moderately fragmented.[16][17] The innovation component of fragmentation has an average standard deviation that is two-thirds as large as fragmentation which suggest that roughly two-thirds of the variation in fragmentation is from the unanticipated component.

### 4.3.3 Stock Level Controls

To control for stock level heterogeneity, we construct four variables using the previous year averages for the given stock. To control for the average level of liquidity for a given stock-year we compute daily spreads from CRSP data as follows:

[16]Fragmentation is bounded between one and $\frac{1}{15}$ with one representing complete consolidation on a single venue.

[17]There are 15 unique trading venues in our data set

$$\text{QS}_{i,y-1} = \frac{1}{D} \sum_{d=1}^{L} \frac{Ask_{i,d} - Bid_{i,d}}{M_{i,d}} \tag{4.13}$$

where d=1...D denotes trading days in year y-1, and $Ask_{i,d}$ and $Bid_{i,d}$ are the end of day ask and bid price. We further compute the previous year's market capitalization, and daily return volatility as follows:

$$\text{MCAP}_{i,y-1} = \frac{1}{D} \sum_{d=1}^{D} p_{i,d} \times \text{shrout}_{i,d} \tag{4.14}$$

$$\text{Vol}_{i,y-1} = \frac{1}{D-1} \sum_{d=1}^{D-1} r_{i,d}^2 \tag{4.15}$$

where $r_{i,d}$ is the daily closing return and shrout is the number of shares outstanding. The control variables account for cross-sectional variation in liquidity, size and return. Our final control variable is an indicator variable *listing*, and it equals 1 if the primary listing of the stock is NASDAQ and 0 if the it is NYSE. This variable controls for the cross-sectional effect of primary listing on all our dependent variables. Panel C of 4.3 reports descriptive statistics on control variables.

## 4.4 Jumps

In this section we outline our jump detection and jump decomposition methodologies.

### 4.4.1 Detecting Jumps

The exact time of a jump along with the jump size can be estimated using a non-parametric test statistic proposed by Lee and Mykland [2008] and outlined

**Table 4.5: Jumps versus Continuous Price moves**
This table reports jump counts (panel A), average estimates from the jump decomposition methodology proposed in Chapter 3 (panel B), liquidity statistics during jumps (panel C) and liquidity statistics during continuous price moves (panel D). We define a jump as systematic if it occurs within a 1 minute interval of a jump in the S&P 500 ETF. Size of permanent return ($|\hat{\eta}|$) is the absolute value of the estimated permanent jump component and size of transitory return ($|\hat{\epsilon}|$) is the estimated size of transitory jump component reported in percentage basis points. Signal-to-noise ratio (SG) is the ratio of estimated permanent to transitory jump size. We define directional fragmentation innovation $\tilde{F}^{D}_{t-1}$, as $\tilde{F}^{Ask}_{t-1}$ ($\tilde{F}^{Bid}_{t-1}$) if time t price move is positive (negative). NBBO quotes are the number of quote revision to the national best bid and offer price. Orderflow is the number of trades in the direction of price move. Orderflow$^{\text{V}}$ is the traded volume in the direction of price move reported in millions. Supply and demand mismatch (SDM) is computed as the number of trades with share size larger than NBBO offered liquidity relative to the number NBBO quote updates. Return size (Ret. Size) is the size of return computed as the absolute value of returns. Trades are the number of trades and volume is the dollar traded volume reported in millions. All variables are winsorized at the %1 level with the exception of listing and return size.

Panel A: Jump Counts

| Name | Count | % Prop. |
|---|---|---|
| Total | 168, 115 | |
| Systematic | 24, 699 | 17.22% |
| Idiosyncratic | 143, 41 | 83.88% |

Panel B: Jump Estimates

| Name | Units | Mean | Std. Dev | Percentiles | | |
|---|---|---|---|---|---|---|
| | | | | $25^{\text{th}}$ | $50^{\text{th}}$ | $75^{\text{th}}$ |
| Size Perm. Ret.($|\hat{\eta}|$) | % Bps./1min | 21.81 | 20.66 | 10.67 | 17.90 | 27.78 |
| Size Trans. Ret.($|\hat{\epsilon}|$) | % Bps./1min | 12.23 | 37.25 | 4.09 | 8.70 | 15.48 |
| Signal to Noise(SG) | Ratio | 12.44 | 44.45 | 0.83 | 2.26 | 5.91 |

Panel C: Jump Liquidity Statistics

| Name | Units | Mean | Std. Dev | Percentiles | | |
|---|---|---|---|---|---|---|
| | | | | $25^{\text{th}}$ | $50^{\text{th}}$ | $75^{\text{th}}$ |
| Frag. Inno ($\tilde{F}^{D}$) | Prop./1min | 0.06 | 0.12 | 0.01 | 0.08 | 0.15 |
| NBBO Quotes | #/1min | 454.70 | 420.71 | 186.22 | 346.72 | 700.79 |

Table 4.5 Continued

| Name | Units | Mean | Std. Dev | Percentiles 25th | 50th | 75th |
|---|---|---|---|---|---|---|
| Orderflow | #/1min | 54.76 | 59.76 | 14.34 | 38.99 | 93.76 |
| Orderflow$^V$ | \$ M./1min | 0.67 | 0.97 | 0.12 | 0.41 | 1.02 |
| SD Mistmatch($SDM$) | #/1minx100 | 19.10 | 13.69 | 8.72 | 16.06 | 26.06 |
| Ret. Size | % Bps./1min | 34.04 | 41.42 | 19.22 | 25.75 | 36.23 |
| Trades | #/1min | 252.97 | 225.25 | 82.17 | 185.61 | 452.59 |
| Volume | \$ M./1min | 3.06 | 3.08 | 0.88 | 1.84 | 4.04 |

Panel D: Continuous Liquidity Statistics

| Name | Units | Mean | Std. Dev | Percentiles 25th | 50th | 75th |
|---|---|---|---|---|---|---|
| Frag. Inno ($\tilde{F}^D$) | Prop./1min | 0.00 | 0.12 | −0.05 | 0.02 | 0.08 |
| NBBO Quotes | #/1min | 270.94 | 289.40 | 90.15 | 174.85 | 334.44 |
| Orderflow | #/1min | 16.19 | 36.71 | −1.48 | 9.01 | 26.15 |
| Orderflow$^V$ | \$ M./1min | 0.16 | 0.50 | −0.03 | 0.08 | 0.26 |
| Ret. Size (Cont. Ret Size) | % Bps./1min | 3.85 | 7.85 | 0.82 | 2.56 | 5.18 |
| SD Mistmatch($SDM$) | Ratio./1min × 100 | 8.84 | 10.85 | 1.48 | 4.92 | 12.10 |
| Trades | #/1min | 108.18 | 123.37 | 34.36 | 67.44 | 131.21 |
| Volume | \$ M./1min | 1.12 | 1.55 | 0.29 | 0.61 | 1.27 |

in section 3.2.9 of this thesis. Using the test statistic $L_t$ we detect detect jumps as follows. Across each stock-year, we compute $L_t$ for all intraday returns sampled at the 1 minute frequency and using a rolling window K=300. This choice of rolling window size is consistent with the authors' suggestion for returns sampled at the 1 minute frequency. Using a critical value of 1 percent we test for a jumps across for in returns $\Delta p_{i,t}$, for each stock-year combination independently. For a given stock, we define a jump as systematic when it is within 1 minute of a S&P 500 ETF jump.[18]

[18] The results are not sensitive to alternate windows up to 5 minute

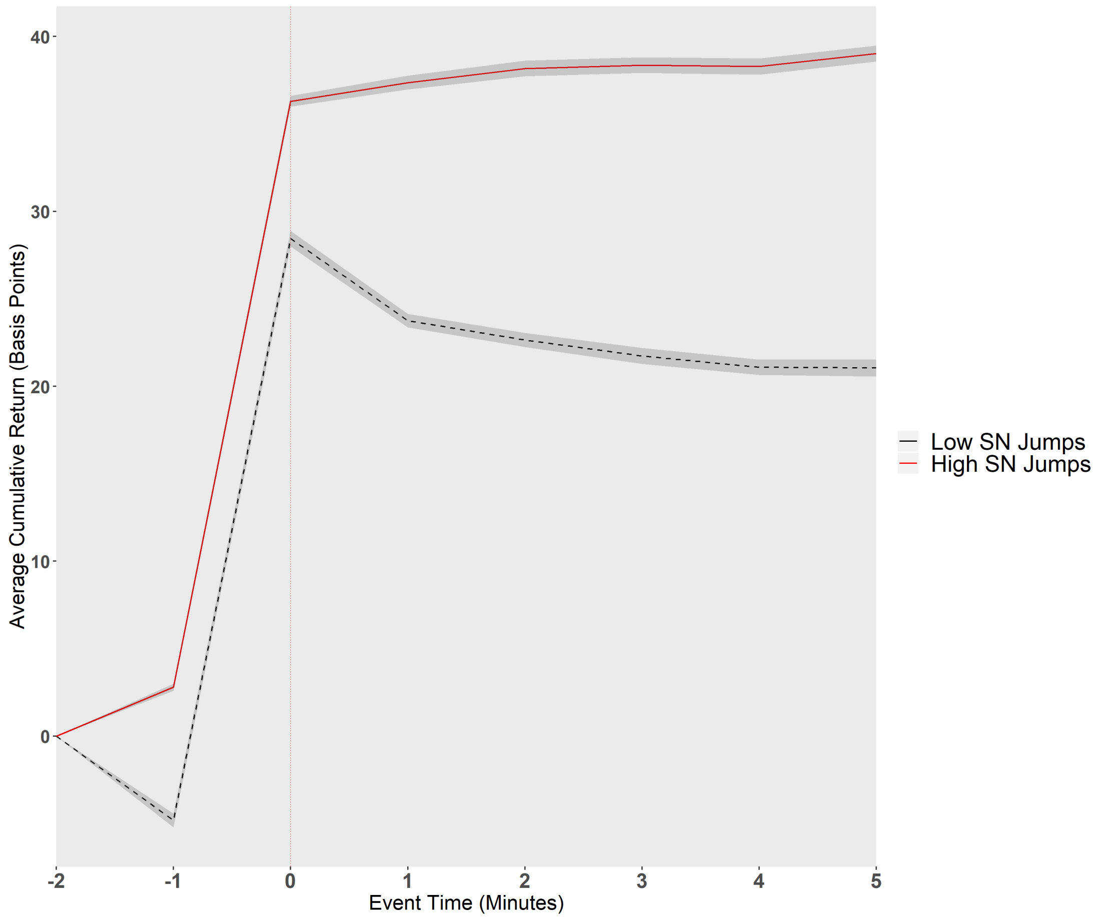



**Figure 4.2: Cumulative Returns - High SN versus Low SN Jumps**
This figure plots cumulative returns in event time for high (solid line) and low (dashed line) signal to noise (SN) jumps. For each jump the realized SN is estimated using the methodology proposed outlined in section 4.2. The horizontal axis shows event time in minutes and the vertical axis shows cumulative returns in minutes. Time 0 denotes time of jump. Cumulative returns are normalized to zero two minute prior to jumps. The shaded region shows the confidence intervals for the mean cumulative returns.

Panel A of table 4.5 reports jump counts for the sample stocks. The total number of jumps detected is 168,115 which represents 0.30% of our sample, amongst these 17.22% are systematic jumps and 83.88% are idiosyncratic. Comparing jump return size (panel C) with the corresponding continuous counterpart (panel D), the average jump size is 34.04 basis points as compared to 3.85 basis point continuous return size. Jumps are therefore roughly nine times larger than continuous price moves. Combined with the low frequency of occurrence, the aforementioned statistics provide convincing evidence that jumps are rare tail events. In addition, jumps are far more likely to be stock specific tail events.

### 4.4.2 Jump Signal-to-Noise (SN) Ratio

Chapter 3 has proposed a methodology to disentangle permanent and transitory jump components, where the permanent component is related to information and the transitory component is the component related to noise (e.g. market overreaction related to liquidity considerations). More specifically the permanent jump component contributes to a price process that is a martingale and therefore efficient, whereas the transitory jump component is a stationary mean reverting process. The identification of the informative (uninformative) part of price movement based on a random walk (stationary) dates back to the early work of Hasbrouck (Hasbrouck [1991a], Hasbrouck [1991b]) and more recently Hendershott and Menkveld [2014] and Chapter 5 of this thesis. The observed and efficient process are modeled as follows:

$$\begin{aligned} p_{i,t^J} &= m_{i,t^J} + s_{i,t^J} \\ m_{i,t^J} &= m_{i,t^J} + \eta_{i,t^J} \end{aligned} \tag{4.16}$$

where $t^J$ is random jump times, $m$ is the efficient jump price process and $s$ is a stationary transitory price process. Using a heavy tailed distribution and importance sampling methods Chapter 3 proposes a methodology to disentangle the equivalent jump return process.

$$\Delta p_{i,t^J} = \eta_{i,t^J} + \Delta s_{i,t^J} \tag{4.17}$$

The individual permanent and transitory jump component in equation (4.17) can be identified from (4.16) which has a state space form. An efficient estimator of the permanent and transitory price process can be obtained of the following form:

$$\begin{aligned} \hat{\eta}_{i,t^J} &= \mathbb{E}[\eta_{i,t^J} | p_1 \dots p_T] \\ \Delta\hat{s}_{i,t^J} &= \mathbb{E}[\hat{s}_{i,t^J} | p_1 \dots p_T] \end{aligned} \tag{4.18}$$

$\hat{\eta}_{i,t^J}$ and $\Delta\hat{s}_{i,t^J}$ are estimators of the conditional expectation form, where the conditioning is on the entire observed prices series, future and past. Since $\hat{\eta}_{i,t^J}$ and $\Delta\hat{s}_{i,t^J}$ are individual estimators for each individual jumps, they are estimators of the realized sample time path of jump time series. We will refer to these estimators as the realized permanent and transitory jump components.

The size of the realized permanent jump component $|\hat{\eta}_{i,t^J}|$ estimates how much new fundamental information is incorporated by the jump at time $t^J$, correspondingly the size of the transitory component $|\Delta\hat{s}_{i,t^J}|$ estimates the mean reverting and illiquid associated jump component at time $t^J$. By this logic we can construct a measure of relative jump informativeness, realized signal-to-noise ratio, $SN$ as follows:

$$SN_{i,t^J} = \frac{|\hat{\eta}_{i,t^J}|}{|\Delta\hat{s}_{i,t^J}|} \tag{4.19}$$

$SN_{i,t^J}$ ratio is a size based relative measure of price informativeness. Jumps which have a large permanent component relative to the transitory component are more informative relative to their liquidity component or conversely have a smaller liquidity (noise) associated component relative to the size of the new information incorporated by the jump. Panel B of table 4.5 reports descriptive statistics for the realized signal-to-noise ratio for intraday jumps in our sample stock. The mean SN is 12.44 while the median is 2.34.[19] This suggests that the SN ratio is skewed left; high information content jumps tend to be far more informative than the median signal-to-noise ratio jumps. To see this explicitly we classify jumps into two categories: (1) high SN jumps with estimated SN of larger than the 75th percentile value and (2) low SN as jumps with estimated SN of smaller than the 25th value; within a given stock. Figure 4.2 presents average cumulative returns in event time by jump type. High SN jumps tend to be larger with a mean size of 36.30 basis points as compared to 28.44 basis points for low SN jumps. As predicted by the estimated model high SN jumps are highly informative with prices continuing to move in the jump direction for up to 5 minutes which adds a further 8 percent to jump impact. In sharp contrast, low SN jumps revert by 27% within 5 minutes of the jump.

## 4.5 Jumps and Fragmentation

Having outlined our jump detection and decomposition methodology, we now focus on the link between jumps and liquidity fragmentation. We first compare intraday jump times to times characterized by extreme unanticipated fragmentation, where we define extreme fragmentation as innovations that are larger than 1.64 standard deviation from the average. Panel A of figure 4.3 plots frequency

[19] Or equivalently the mean (median) of percentage estimated noise is 7.44% (29.77%)

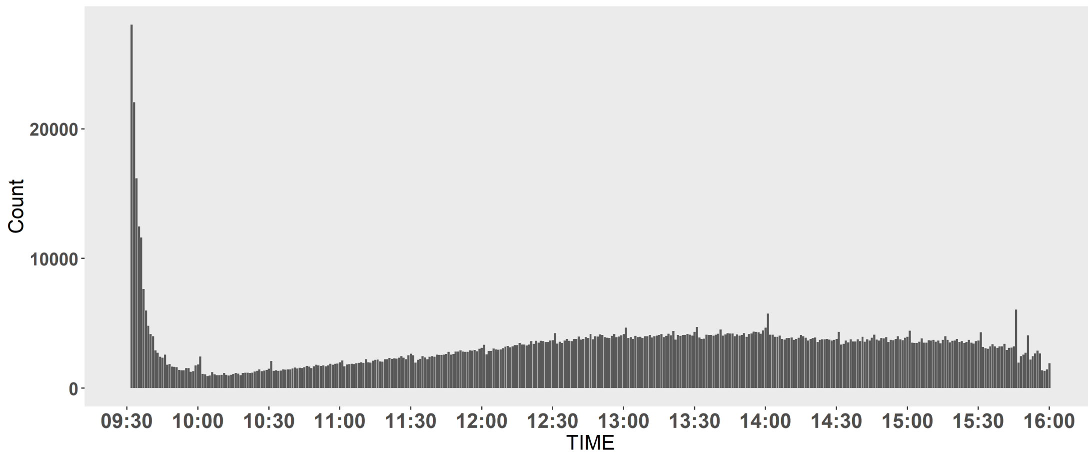
counts of extreme fragmentation innovation in 1 minute intraday time bins.[20][21] Several noteworthy features can be observed from the figure. First, the total count of intraday times when markets are characterized by extreme unanticipated fragmentation is 1,270,932 or 2.35% of our sample as compared to jumps which consist of 0.30% of our sample. Second, large unanticipated fragmentation occur most often within the first thirty minutes of market open. This is consistent with a significant quantity of news generated overnight and before market open. Lastly, comparing panel A with panel B it is evident that both jumps and extreme fragmentation innovation tend to coincide with the first 30 minutes of market open. Hence, intraday times when markets are most fragmented are also times when jumps are likely.

Jumps are mainly attributed to the arrival of new information in the form of news announcements or unanticipated liquidity shocks. In either case jumps trigger large directional immediacy demand as traders' manage the adverse price risk by buying (selling) large quantities of shares during positive (negative) jumps.[22] Table 4.5 reports that the average 1 minute traded volume and number of trades are approximately three times as large as their corresponding continuous counterpart. More importantly, the average volume orderflow imbalance when measured in the direction of price move is 0.67 million during jumps as compared to 0.16 million during continuous price moves. This surge in directional liquidity demand implies that liquidity suppliers risk holding large quantities of unbalanced inventory positions. To manage their

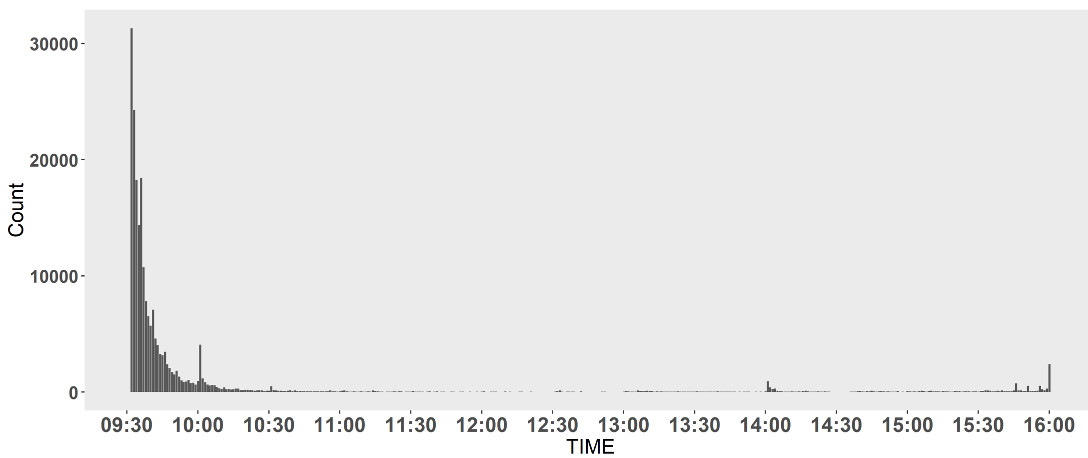
[20]Figure 4.3 is robust to alternate definitions using values between one and two standard deviation.

[21]The mean of fragmentation innovation is zero.

[22]As in Grossman and Miller [1988] immediacy is defined as traders' willingness to trade immediately rather than wait.




**Panel A: Extreme Unanticipated Fragmentation**




**Panel B: Jumps**

**Figure 4.3: Extreme Fragmentation and Price Jumps - Frequency Counts**
This figure plots frequency counts of extreme liquidity fragmentation shock (panel A) and jumps (panel B). We define the latter as times when unanticipated liquidity fragmentation is larger than 1.64 standard deviation. The horizontal axis corresponds to intraday time.

inventory exposure liquidity suppliers may fragment quotes leading to a mismatch between the supply and demand side of liquidity during times leading up to jumps. Figure 4.4 plots SDM in intraday event time, where time 0 corresponds to the time of the jump.[23] The dashed line corresponds to the mean of SDM and the event window is defined as 10 minute prior and post jump. Interestingly, SDM begins to increase 10 minute before jumps and reaches its largest value 0.02 at the time of the jump. Thereon, SDM gradually falls over the following 10 minutes.

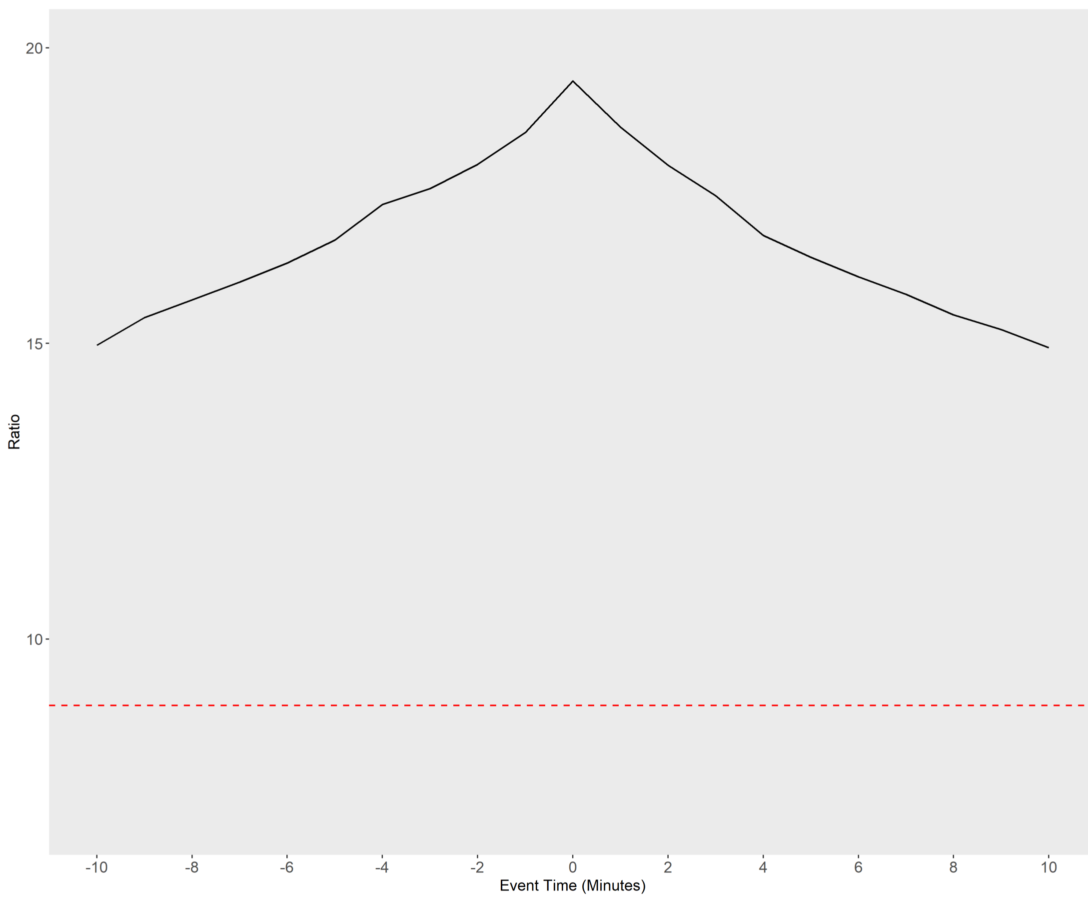



**Figure 4.4: Liquidity Supply and Demand Mismatch and Jumps**
This figure plots liquidity supply and demand mismatch (SDM) in jump event time. We define liquidity supply and demand mismatch as the number of trader larger than NBBO offered liquidity relative to the number NBBO quote updates The horizontal axis corresponds to intraday event time in minutes. Time 0 denotes time of jumps. The dashed line corresponds to average fragmentation shock. The shaded region shows the confidence intervals for the means across jump types.

[23]In section 2.1 we define SDM as the count of trades which consume all liquidity available at NBBO price relative to number of NBBO quote revisions

Figure 4.4 highlights an important feature of jumps: in times leading up to jumps there is a gradual build up of immediacy demand which leads to mismatch between the supply and demand side of liquidity. In particular, as measured by SDM, large sized trades tend to outnumber quote revisions and this trend gradually increases in jump time. In appendix A we show that liquidity fragmentation is associated

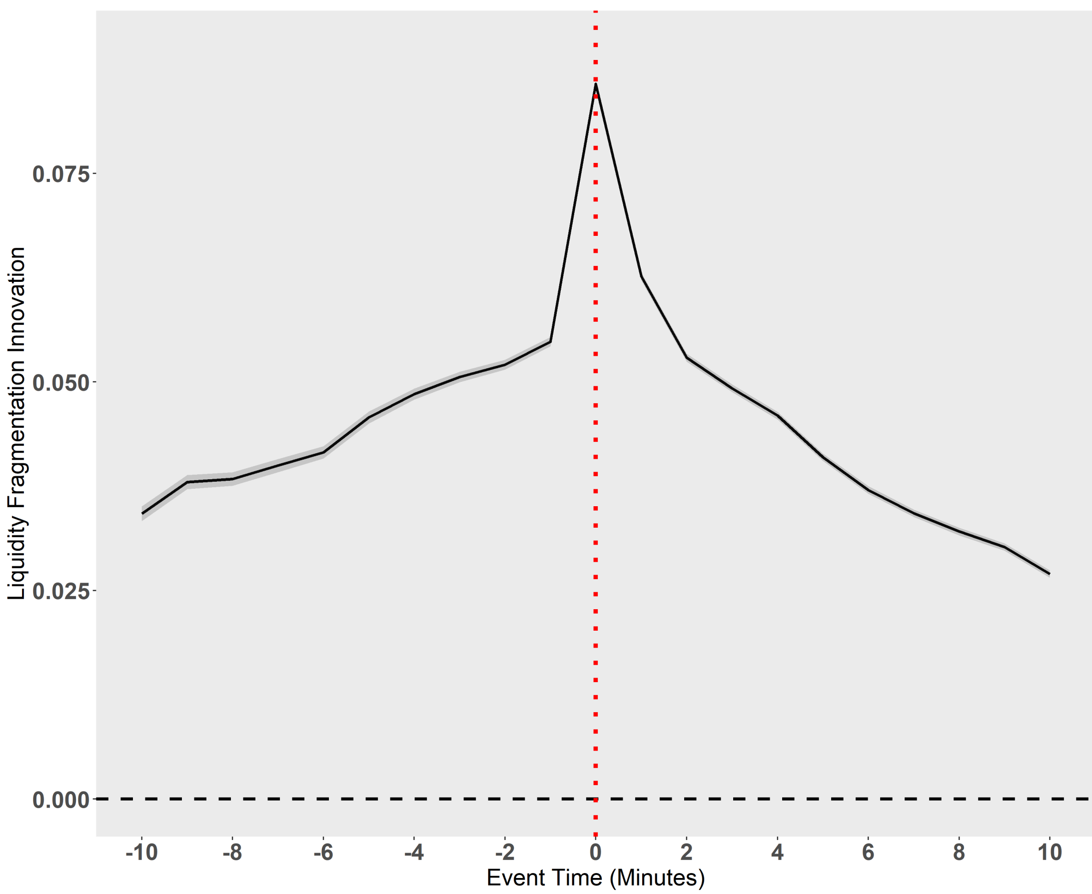



**Figure 4.5: Unanticipated Fragmentation and Jumps**
This figure plots liquidity fragmentation shocks during jumps. Unanticipated Fragmentation (innovation) is estimated using an autoregressive model and equation (4.12). The horizontal axis corresponds to intraday event time in minutes. Time 0 denotes time of jumps. The dashed line corresponds to zero mean of fragmentation shocks. The shaded region shows the confidence intervals for the means across jump types.

with larger future SDM levels which highlights that fragmented liquidity leads to a mismatch between the supply and demand side of liquidity.

If fragmentation is linked to jumps through the liquidity suppliers' information; then a preliminary analysis should show a gradual increase in fragmentation before jumps as was the case for SDM. To examine the timing of unanticipated fragmentation relative to jumps, figure 5 plots fragmentation innovation during each 1-minute time interval in event time, where time 0 corresponds to time of jumps. Between event time -10 and -1 liquidity experiences fragmentation innovation in increasing magnitude. The innovation component at time -1 is 0.06 followed by its maximum value of 0.09 at the time of jump. This suggests that the unanticipated component of fragmentation is informative for future price jumps. Where the information content is increasing gradually in jump event time.

The preliminary analysis of this subsection suggest two characteristics of the association between fragmentation and jumps; (i) periods when jumps are most probable are also periods when large unanticipated fragmentation is also most likely and (ii) while fragmentation and SDM are largest during jumps, they tend to increase gradually during times leading up to jumps. The aforementioned results provide preliminary evidence to suggest that fragmented liquidity is informative for jumps and also lead to a mismatch of liquidity supply and demand.

We formally model the relationship between liquidity fragmentation and price jumps as a series of test:

Test I: *Does unanticipated fragmentation predict future price jumps?*

Test II: *Conditional on observing a jump, does unanticipated fragmentation result with noisier jumps?*

Test III: *Is fragmented liquidity during times leading up to jumps a strategic response of liquidity suppliers?*

If liquidity suppliers are better informed than the general trading public, it follows that unanticipated fragmentation should be positively related to future jump events. This is precisely the question in test 1. Test II examines price informativeness during jumps. If fragmented liquidity leads to a mismatch between the supply and demand side of liquidity then jump price moves will, on average, overshoot relative to their information as the demand side of liquidity will trade at worse prices, making jumps during times of fragmented liquidity noisier. Test III examines if liquidity fragmentation is at least in part related to liquidity suppliers' strategic inventory management. Positive (negative) jumps are accompanied by large buy (sell) direction imbalances in orderflow. If liquidity suppliers anticipate these imbalances they will fragment the ask (bid) side of liquidity to manage their exposure to positive (negative) incoming jumps. Thereon, any measure which incorporates the direction of liquidity fragmentation should contain more information for future price jumps as compared to an average measure. The three tests combined address a more general question as follows. *How important are high frequency market linkages during times leading up to extreme price movements?* In the following sections we perform the three tests respectively.

## 4.6 Can Fragmentation Predict Jumps and Jump Directions?

If liquidity suppliers are better informed than the general trading public, we should expect that unanticipated changes in fragmentation are informative for price jumps. In this section we perform Test 1 of the paper in an instrument variable (IV) probit model. In particular, by using an instrument we estimate the direct predictive effect of fragmentation for jumps arising from the information of liquidity suppliers.

### 4.6.1 Unobserved Endogeneity

Time t-1 fragmentation innovation and time t price jumps are likely related through other decision variables of liquidity suppliers not observed in TAQ data. To see this note that fragmentation depends on liquidity suppliers' information, as do other decision variables of liquidity suppliers such as quote cancellations and dept offered at each level of the limit order book.[24] These variables are largely unobserved in TAQ data and in addition their relationship to fragmentation is of unknown form. This aforementioned omitted variable bias implies that the information content of fragmentation for jumps is likely to be over or underestimated depending on the sign of the correlation of omitted variables with fragmentation and with jumps.[25] In particular, there exist three channels through which liquidity suppliers' information can predict jumps: (i) directly through fragmentation (ii) indirectly through the correlation of fragmentation and omitted variables (iii)

[24] Every variable on the limit order book is at least partially correlated with liquidity suppliers' information.

[25] If the average correlation between (i) omitted variables and jumps and (ii) omitted variables and fragmentation is of identical (opposite) sign then the predictive power of fragmentation for jumps is over (under) estimated when ignoring omitted variable bias.

directly through omitted variables. Our objective is to estimate (i) and while (iii) does not lead to bias in our estimation; (ii) will lead to biased estimates for (i). Put explicitly, the predictive power of fragmentation for price jumps and their direction may arise partially for reasons other then the *direct* effect of liquidity suppliers' information on fragmentation i.e. omitted variables which predict jumps and are also correlated with fragmentation. To address this aforementioned endogeneity we follow an approach similar to Degryse and van Kervel [2015]. In this approach we instrument a given stock's time t fragmentation, with time t systematic component of fragmentation. The instrument is computed by taking the average fragmentation for all stocks except stock i as follows:

$$\tilde{F}_{i,t}^{IV,D} = \frac{1}{I-1} \sum_{j \neq i} \tilde{F}_{j,t}^{D} \tag{4.20}$$

The instrument $\tilde{F}_{i,t}^{IV}$ removes all stock specific endogeneity issues arising from the link between fragmentation and omitted variables. This follows since it is highly unlikely that stock specific (idiosyncratic) variation in omitted variables would be correlated with market liquidity; i.e. liquidity for all other stocks excluding stock i. Therefore, by instrumenting with systematic fragmentation we largely isolate the predictive power of fragmentation arising from the direct effect of liquidity suppliers' information on fragmentation during times leading up to jumps. While it still remains the case that systematic variation in omitted variables are likely correlated with the instrument, this would only lead to omitted variable bias for the case of systematic jumps. However, only 17% of the jumps in our sample are systematic jumps and further our results are robust to the exclusion of systematic jumps.[26] In addition, we control for several dimensions of systematic liquidity using measures computed from the S&P 500 ETF. Therefore, our IV approach to account for the correlation between fragmentation and omitted liquidity variables is a conditional IV model.[27].

Table 4.6 presents our first stage OLS regressions, where in column 2 we include all control variable for systematic and firm specific liquidity.[28] The correlation between fragmentation and its instrument is positive and statistically significant. Despite the significantly smaller standard deviation of the instrument, the results imply that a one standard deviation change in the instrument is associated with a 30 percent change in fragmentation innovation of the same direction. This result is consistent with Coughenour and Saad [2004] who show that NYSE specialist provide liquidity for multiple stocks. Commonality in specialist is likely to create commonality in fragmentation at the stock level, as inventory risk is pooled across multiple stocks and therefore interlinked. The first stage regression also present several results consistent with the claim that liquidity fragmentation is a search cost for the demand side of liquidity. Table 4.6 reports a positive correlation between fragmentation and the two transaction cost measures: effective spreads and price impacts. This suggest that liquidity is likely to fragmented when transaction cost is large. In addition fragmentation is negatively correlated with the size of offered depth, and therefore the size of offered liquidity. Lastly, the

[26]We define a jump as systematic if it occurs within a 1 minute interval of a jump in the S&P 500 ETF. See table 4.5.

[27]The Independence assumption of the instrument is replaced by independence conditional on a set of observable controls; see Deuchert and Huber [2017] for a discussion on conditional IV models

[28]We discuss our control variable in the following section.

number cumulative shares traded at NYSE and NASDAQ is positively correlated with fragmentation. Including this variable in our forthcoming models controls for demand side innovation which may fragment liquidity prior to jumps.

### 4.6.2 Methodology

Having discussed our instrumental variable approach we now present our IV Probit model to parametrically model the predictive relationship between fragmentation and jumps arising directly from the information of liquidity suppliers. In particular, this section examines if fragmentation innovation at time period t-1 have an impact on time t jump probability. First, we define a binary response jump variable as follows:

$$J_{i,t} = \left\{ \begin{array}{ll} 1, & t \in \mathcal{J}_{i,t} \\ 0, & t \notin \mathcal{J}_{i,t} \end{array} \right\} \tag{4.21}$$

where $\mathcal{J}_{i,t}$ is the set of detected jump times for stock i. Corresponding to the binary response model there exist a latent jump state $J^*_{i,t}$ defined as follows:

$$J_{i,t} = \mathbb{I}[J^*_{i,t} \geq 0] \tag{4.22}$$

$\mathbb{I}[J^*_{i,t} \geq 0]$ is an indicator function which equals 1 if $J^*_{i,t} \geq 0$. $J^*_{i,t}$ is an information threshold variable governing the dynamics of jumps. We model $J^*_{i,t}$ as a linear function of an unpredictable random component $\nu_{1,t}$, and a set of time t-1 observables as follows:

$$J^*_{i,t} = \zeta + \delta_{11}\tilde{F}^D_{i,t-1} + \delta_{12}C_{i,t-1} + \nu_{1,i,t}. \tag{4.23}$$

**Table 4.6: First Stage (IV Model) - Fragmentation Innovation**
This table reports OLS results of the first stage IV regression. The dependent variable is directional fragmentation innovation $\tilde{F}^{D}_{i,t}$, defined as $\tilde{F}^{Ask}_{t}$ ($\tilde{F}^{Bid}_{t}$) if time t+1 price move is positive (negative). $\tilde{F}^{IV,D}_{i,t}$ is the instrument for $\tilde{F}^{D}_{i,t}$; defined as the average fragmentation innovation for all stocks except stock i. Effective spread ($ES_t$) is the signed difference between price and prevailing midpoint, measured relative to the prevailing midpoint at the time of trade. Price impact ($PI_t$) is the signed changed in midpoint following a trade, relative to prevailing midpoint at time of trade. Average dept ($Depth^{V}_{t}$) is the time weighted average of ask side and bid side offered volume. $Trades$ is cumulative traded volume at NASDAQ and NYSE in units of ten thousand shares. SPY superscript denotes S&P 500 ETF variables which proxy for systematic liquidity. The data is pooled and includes stock-year level controls as follows: average market capitalization (MCAP), average quoted spread computed from the previous years data; and *listing* defined as a dummy variable which equals one if the stock's primary listing is on NASDAQ and zero if NYSE. SPY return size ($|r|^{SPY}_{t}$) is the absolute value of time t-1 SPY return. All variables are statistically significant at the % 1 level unless otherwise stated.

| Dependent Variable= Fragmentation Innovation ($\tilde{F}^{D}_{i,t}$) | | | | |
|---|---|---|---|---|
| | Reg 1. | | Reg 2. | |
| Variable | Coeff. | T-stat | Coeff | T-stat |
| Intercept | 0.00 | 46.35 | −0.16 | −255.46 |
| IV Frag. Inno. ($\tilde{F}^{IV,D}_{t}$) | 0.76 | 1,387.19 | 0.52 | 884.92 |
| Eff. Spread ($ES_t$) | | | 0.28 | 181.75 |
| Prc. Impact ($PI_t$) | | | 0.17 | 328.81 |
| Depth$^{V}$ ($Depth_t$) | | | −0.33 | −1,095.40 |
| Trades ($Trades_t$) | | | 0.02 | 919.63 |
| SPY Eff. Spread ($ES^{SPY}_{t}$) | | | −0.01 | −19.52 |
| SPY Prc. Impact ($PI^{SPY}_{t}$) | | | 0.04 | 31.46 |
| SPY Depth ($Depth^{SPY}_{t}$) | | | 0.00 | 100.51 |
| SPY Ret. Size ($|r|^{SPY}_{t}$) | | | −0.13 | −150.08 |
| MCAP ($Size$) | | | 0.01 | 244.59 |
| Quoted Spreads ($QS$) | | | 1.01 | 385.76 |
| Volatility ($Vol$) | | | −1.39 | −324.24 |
| Listing | | | −0.00 | 39.40 |
| F-Value | | 1,924,300 | | 144,819 |
| Year Fixed Effect | Yes | | Yes | |
| # R-sq | 0.03 | | 0.06 | |
| # Obs | 54,185,923 | | 54,185,923 | |

Assuming $\nu_{1,i,t}$ is a normally distributed random variables, equation (4.23) defines a probit model for the jump binary response variable $J_{i,t}$. Where $\tilde{F}^{D}_{i,t-1}$ denotes time t-1 innovation component of liquidity fragmentation defined in (4.12) and $C_{i,t-1}$ denotes matrix of control variables. The variable $\nu_{1,i,t}$ corresponds to news or liquidity shocks which are purely unanticipated by the market at time t-1; e.g. the unanticipated component of an earnings announcement. The non-linearity of the probit model makes estimation of stock fixed effects computationally intensive. Therefore, we use an alternative approach to control for stock level heterogeneity using variables which account for stock characteristic. These set include the logarithmic of average market capitalization (size), average daily quoted spread and average daily volatility computed from the previous year's CRSP data. We also include a dummy to capture listing affect. The *listing* dummy variable equals one if the stock's primary listing is NASDAQ and zero if it is NYSE. We control for time fixed effects by including year fixed dummy variables in $C_{i,t-1}$.

Our main parameter of interest is $\delta_{11}$. A positive coefficient on $\delta_{11}$ means that past unanticipated changes in liquidity fragmentation increase the probability of future jumps. However, as discussed in the previous section, a stock's level of liquidity fragmentation is correlated with omitted variables, therefore in order to explicitly examine the predictive effect of fragmentation which directly arises from the information of liquidity suppliers, we use an IV probit model. The endogenous equation is assumed to be a linear model as follows:

$$\tilde{F}^{D}_{i,t-1} = \mu_2 + \delta_{21}\tilde{F}^{IV,D}_{i,t-1} + \delta_{22}C_{i,t-1} + \nu_{2,i,t-1} \tag{4.24}$$

Omitted variable bias is explicitly modelled by assuming that $\nu_{2,i,t-1}$ and $\nu_{1,i,t}$

are correlated with a correlation coefficient denoted by $\rho_{\nu_1,\nu_2}$. We estimate $\rho_{\nu_1,\nu_2}$ using maximum likelihood. Using (4.23) and (4.24) along with the two marginal densities $f(J_{i,t}|\tilde{F}^D_{i,t-1}, C_{i,t-1})$ and $f(\tilde{F}^D_{i,t-1}|\tilde{F}^{IV,D}_{i,t-1}, C_{i,t-1})$ we can explicitly derive the log likelihood function as follows:

$$\begin{aligned}
\mathcal{L}^{IV} &= \sum_{i,t} \log\Big( f(J_{i,t}|\tilde{F}^D_{i,t-1}, C_{i,t-1}) f(\tilde{F}^D_{i,t-1}|\tilde{F}^{IV,D}_{i,t-1}, C_{i,t-1}) \Big) \\
&= \sum_{i,t} \Big( J_{i,t} \log\big(\Phi(\omega_{i,t})\big) + [1 - J_{i,t}] \log\big(1 - \Phi(\omega_{i,t})\big) - \\
&\quad {}^1\!/\!_2 \log(\sigma^2_{\mu_2}) - {}^1\!/\!_{2\sigma^2_{\mu_2}} (\tilde{F}^D_{i,t-1} - \mu_2 - \delta_{21}\tilde{F}^{IV,D}_{i,t-1} - \delta_{22}C_{i,t-1})^2 \Big)
\end{aligned} \tag{4.25}$$

Where:

$$\omega_{i,t} = \frac{[\mu_1 + \delta_{11}\tilde{F}^D_{i,t-1} + \delta_{12}C_{i,t-1} + ({}^{\rho_{\nu_1,\nu_2}}\!/\!_{\sigma_{\mu_2}})(\tilde{F}^D_{i,t-1} - \mu_2 - \delta_{21}\tilde{F}^{IV,D}_{i,t-1} - \delta_{22}C_{i,t-1})]}{(1 - \rho^2_{\mu_1,\mu_2})^{\frac{1}{2}}} \tag{4.26}$$

with $\text{Var}(\nu_{1,i,t}) = \sigma^2_{\nu_1}$; $\text{Var}(\nu_{2,i,t-1}) = \sigma^2_{\nu_2}$; $\text{Corr}(\nu_{1,i,t}, \nu_{2,i,t-1}) = \rho_{\nu_1,\nu_2}$ and $\Phi(.)$ denotes the cdf of normal distribution. We jointly maximize all parameters in the log-likelihood function (4.25) using the BFGS algorithm over our data set consisting of of 54 million data points.

Also included in our set of control variables $C_{i,t}$ are intraday stock liquidity variables which capture the size of liquidity on several different dimensions. In particular, we include the following intraday variables: (i) effective spreads ($ES_{i,t-1}$) which proxy for the overall trading cost, (ii) price impact ($PI_{i,t-1}$) of trade which proxy for the information content of trades and (iii) the time weighed average depth which directly measures the total size of liquidity available at the market best prices ($Depth^V_{t-1}$). In order to control for systematic liquidity we compute both liquidity and price informativeness measures from the S&P 500 ETF. This includes all the above liquidity variables computed using the S&P 500 ETF rather than the stock data. In addition to systematic liquidity, we also control for systematic information shocks using the size of S&P 500 ETF returns ($|r_{t-1}|^{SPY}$). Lastly, liquidity can fragment at time t-1 from demand side shock which consumes a large proportion of available liquidity at the two largest exchange; i.e. NYSE and NASDAQ. Therefore, we control for the cumulative number of shares traded ($Trades_{i,t-1}$) at NYSE and NASDAQ.[29]

### 4.6.3 Results

Columns 1 and 2 of table 4.7 report MLE estimates and t-statistics from maximizing the log-likelihood function given in (4.25). Our main estimate of interest is the coefficient on $\tilde{F}^D_{t-1}$ which corresponds to the effect of fragmentation on the on the binary response variable $J_{i,t}$ through its effect on $J^*_{i,t}$. The estimate is positive which implies that ask (bid) side fragmentation predict higher positive (negative) jump probabilities. Correspondingly, the magnitude of the marginal effect is positive. In order to examine the importance of fragmentation in predicting jumps we compute a measure known as *recall* in the machine learning literature. To compute recall we define two counting variables as follows:

[29] We are agnostic about the source of information leading liquidity suppliers' to fragment liquidity during time periods prior to t-1. Therefore, if demand side shocks fragment liquidity prior to time t-1 then the source of the information for liquidity suppliers' is potentially the demand side. Important to our analysis is controlling for demand side shocks at time t-1.

$$\begin{aligned} \mathcal{J} &:= \{\text{Number of jumps in the sample}\} \\ \mathcal{J}^{\alpha} &:= \{\text{Number of jumps in the sample with } \widehat{P}(J_{i,t}=1) \geq \alpha\} \end{aligned} \tag{4.27}$$

$\mathcal{J}^{\alpha}$ counts the number of in sample jumps which are correctly predicted at a decision rule of $\alpha$. It follows that recall is defined as follows:

$$\text{Recall}(\alpha) = \frac{\mathcal{J}^{\alpha}}{\mathcal{J}} \times 100 \tag{4.28}$$

Recall is therefore the percentage of correctly predicted jumps using a decision rule $\alpha$. While jumps represent 0.3 percent of our sample, table (4.7) reports a recall of 28 percent at an $\alpha$ of 0.5. Put explicitly, while occur with a frequency of 0.3 percent the IV Probit model correctly identifies 28 percent of jumps at a decision rule of 0.5. We next examine how important is fragmentation in predicting jumps? In order to answer the aforementioned question we drop fragmentation from our model. The resulting recall is 0 percent as reported in table (4.7). This suggests that fragmentation is the most important information variable in predicting jumps. Columns 3-4 report results for the simple probit model which severly underestimated the effect of fragmentation for jumps. In particular the estimated correlation between fragmentation innovation and unobserved omitted variables $\rho_{\nu_1,\nu_2}$, is both large and negative implying that omitted variables have correlations with fragmentation and with jumps which have opposite signs. For example, controlling for NBBO depth, depth at levels outside the NBBO is likely to be negatively correlated with NBBO liquidity fragmentation and positively correlated with jumps as liquidity suppliers place limit orders further down the

**Table 4.7: Probit Model - Fragmentation Shocks and Jump Direction**
This table presents estimation results from the IV Probit model (columns 2-4) and Probit model (columns 5-7). The dependent variables is jump binary response which equals one if there is a jump at time t and zero otherwise. We define directional fragmentation innovation $\tilde{F}^{D}_{t-1}$, as $\tilde{F}^{Ask}_{t-1}$ ($\tilde{F}^{Bid}_{t-1}$) if time t price move is positive (negative). Effective spread ($ES_{t-1}$) is the difference between price and prevailing midpoint, measure relative to the prevailing midpoint at the time of trade. Price impact ($PI$) is the signed changed in midpoint following a trade, relative to prevailing midpoint at time of trade. $Trades$ is cumulative traded volume at NASDAQ and NYSE in units of ten thousand shares. SPY superscript denotes S&P 500 ETF variables which proxy for systematic liquidity. The data is pooled with included stock-year level controls: average market capitalization (MCAP), average quoted spread computed from the previous years data; and *listing* defined as a dummy variable which equals one if the stock's primary listing is on NASDAQ and zero if NYSE. SPY return size ($|r|^{SPY}_{t-1}$) is the absolute value of time t-1 SPY return. Columns 3 and 6 report the marginal effect at the mean $\frac{dP\bar{X}}{dX_k}$ defined as the change in jump probability from a small change in the the given dependent variable from it's mean; with all other dependent variable held fixed at their respective means. We define a correctly predicted response if the predicted probability of a jump in the sample is greater than or equal to 0.5 (less than 0.5) and the observed response is a jump (not a jump). % Correctly Predicted is the percentage of correctly predicted response. % Recall is the percent of correctly predicted jumps relative to the number of jumps. We use full information likelihood for estimating both probit models. The reported t-statistics are computed using inverse of the estimated hessian matrix. All variables are winsorized at the %1 level with the exception of listing and return size. All variables are statistically significant at the % 1 unless otherwise stated. All variables are statistically significant at the %1 level unless otherwise stated

| Dependent Variable = Binary Jump Variable ($J_{i,t}$) | | | | | | |
|---|---|---|---|---|---|---|
| | IV Probit Model | | | Simple Probit Model | | |
| Variable | Coeff. | T-stat | $\frac{dP(\bar{X})}{dX_k}$ | Coeff | T-stat | $\frac{dP\bar{X}}{dX_k}$ |
| Intercept | 0.15 | 14.31 | | −3.97 | −136.80 | |
| Frag. Inno. ($\tilde{F}^{D}_{t-1}$) | 7.26 | 2,865.10 | 1.82 | 0.63 | 88.23 | 0.01 |
| Eff. Spread ($\tilde{F}^{D}_{t-1}$) | 1.17 | 57.95 | 0.29 | 16.59 | 463.36 | 0.15 |
| Prc. Impact ($PI_{t-1}$) | −0.96 | −158.44 | −0.24 | 1.13 | 74.13 | 0.01 |
| Depth$^V$ ($Depth_{t-1}$) | 1.87 | 307.12 | 0.47 | −3.29 | −177.91 | −0.03 |
| Trades ($Trades_{t-1}$) | −0.10 | −398.44 | −0.03 | 0.15 | 266.87 | 266.87 |
| SPY Eff. Spread ($ES^{SPY}_{t-1}$) | −0.10 | −16.31 | −0.03 | −0.77 | 37.39 | −0.01 |
| SPY Prc. Impact ($PI^{SPY}_{t-1}$) | −0.59 | −30.52 | −0.15 | −0.88 | −16.34 | −0.00 |
| SPY Depth ($Depth^{SPY}_{t-1}$) | −0.01 | −22.72 | −0.00 | 0.03 | 28.97 | 0.00 |
| SPY Ret. Size ($|r|^{SPY}_{t-1}$) | 0.64 | 54.33 | 0.16 | 1.97 | 63.38 | 0.02 |
| MCAP ($Size$) | −0.05 | −97.63 | −0.01 | 0.07 | 45.93 | 0.00 |

Table Continued

| | | | | | | |
|---|---|---|---|---|---|---|
| Quoted Spreads ($QS$) | −7.60 | −189.78 | −1.90 | −2.63 | −23.00 | −0.02 |
| Volatility ($Vol$) | 5.58 | 76.83 | 1.40 | −19.74 | −95.25 | −0.18 |
| Listing | −0.01 | −3.02 | 0.00 | 0.05 | 22.31 | 0.00 |
| Endogenity Test for $\tilde{F}_{i,t}$ ($\rho_{\nu_1,\nu_2} = 0$) | $-0.944, 340.00$ | | | | | |
| Year Fixed Effect | Yes | | | Yes | | |
| % Recall(Decision Rule: $\widehat{P}(J_{i,t} = 1) \geq 50\%$) | 28.14 | | | 0.43 | | |
| % Recall (Without Frag. Inno) | 0.00 | | | 0.00 | | |
| # Obs | 54,185,923 | | | 54,185,923 | | |

limit order book to manage inventory risk from jumps.

Next we examine the magnitude of predictive ability of fragmentation for jumps by computing predicted jump probabilities using the maximized parameter estimates of our probit model. Table 4.8 reports predicted jump probabilities for both the IV probit and simple probit models. In each row of table 4.8, we compute predicted jump probabilities for values of fragmentation innovation $\tilde{F}_t^D$ which vary in increments of half standard deviation from its mean, with all other explanatory variables held fixed at their mean value. Since the mean of fragmentation innovation is zero by design, we refer to positive (negative) standard deviation increments as fragmentation (consolidation) innovation to liquidity. We compute separate predictions for the two values of the *listing* dummy variable which correspond to predicted probabilities for the NYSE (columns 1 and 3) and NASDAQ (columns 3 and 4) listed stocks. For the IV Probit model, the predicted jump probability for NYSE listed stocks is 16.73% when fragmentation is at its mean.[30] The predicted probability of a positive (negative) jump more than doubles to 34.70% for a half standard deviation fragmentation innovation on the ask

[30]The predicted probability at the mean value is larger than the unconditional jump frequency as a result of the negative estimated correlation $\rho_{\nu_1,\nu_2}$ from the IV probit model in (4.23) and (4.24). As discussed earlier this implies that omitted variables have correlations of opposite magnitudes with fragmentation and jumps. An example is provided in the discussion earlier.

(bid) side of liquidity. A full standard deviation innovation almost quadruples

**Table 4.8: Effect of Fragmentation Innovation on Jump Probaility and Direction**

This table presents predicted probabilities from the IV Probit model (columns 2-4) and probit model (column 3-4). We define directional fragmentation innovation $\tilde{F}^{D}_{t-1}$, as $\tilde{F}^{Ask}_{t-1}$ ($\tilde{F}^{Bid}_{t-1}$) if time t price move is positive (negative). We report predicted probabilities for fragmentation innovation ($\tilde{F}^{D}_{t-1}$) in units of standard deviation from the mean; with all other dependent variables held fixed at their mean value. We individually report predicted values for NASDAQ and NYSE stocks based on the *listing* dummy.

| | Predict Jump Probability $\widehat{Prob}$ (%) | | | |
|---|---|---|---|---|
| | IV Probit Model | | Simple Probit Model | |
| Std Dev. from Mean ($\tilde{F}^{D}_{t-1}$) | NYSE | NASDAQ | NYSE | NASDAQ |
| −1.50 | 0.57 | 0.40 | 0.20 | 0.19 |
| −1.25 | 1.16 | 0.89 | 0.21 | 0.21 |
| −1.00 | 2.23 | 1.84 | 0.23 | 0.23 |
| −0.75 | 4.03 | 3.54 | 0.24 | 0.24 |
| −0.50 | 6.86 | 6.38 | 0.26 | 0.26 |
| −0.25 | 11.02 | 10.73 | 0.28 | 0.28 |
| 0.00 | 16.73 | 16.89 | 0.30 | 0.30 |
| 0.25 | 24.07 | 24.95 | 0.32 | 0.33 |
| 0.50 | 32.88 | 34.70 | 0.35 | 0.35 |
| 0.75 | 42.77 | 45.58 | 0.37 | 0.38 |
| 1.00 | 53.13 | 56.81 | 0.39 | 0.41 |
| 1.25 | 63.28 | 67.52 | 0.42 | 0.44 |
| 1.50 | 72.58 | 76.94 | 0.45 | 0.47 |

the predicted probability to 53.13%. Conversely, a half standard deviation consolidation innovation to ask (bid) side of liquidity reduces the predicted positive (negative) jump probability by more than half with to a value of 6.38%. A full one standard deviation consolidation innovation further reduces this probability to 1.84%. These results provide strong evidence to suggest that (i) time t-1 fragmentation strongly predict jumps and jump direction and (ii) omitted variable bias severely underestimates the predictive ability of fragmentation. Since we largely isolated the direct effect of liquidity supplier's information for jumps through the fragmentation channel, these results suggest that liquidity suppliers' anticipate jumps and their information is strongly captured by fragmentation. More formally, the information content of fragmentation strongly predict jumps. Further, the predicted probabilities for NASDAQ stocks are similar to NYSE stocks.

## 4.7 Does Fragmentation Predict Noisier Jumps?

In appendix A we show that fragmentation also predicts a mismatch between supply and demand for liquidity. Any price impact arising from this mismatch must mean revert as it is unrelated to the efficient price. Further, the eventual arrival of natural counter-parties allow liquidity suppliers to restore their original inventory position (e.g. Ho and Stoll [1981]; Grossman and Miller [1988]; Hendershott and Menkveld [2014]) and therefore restore the balance between the supply and demand for liquidity. Therefore, if fragmentation innovation lead to a mismatch between the supply and demand side of liquidity in times preceding jumps, we should expect fragmentation innovation to predict jumps with smaller signal-to-noise (SN) ratio. In this section we examine Test II: *Conditional on observing a jump, does unanticipated fragmentation result with noisier jumps?*

### 4.7.1 Two Stage Least Square Regression - Signal-to-Noise Ratios

In section 4.4.2 we outline our methodology to decompose each jump in our sample into two components: (i) permanent component, which represent changes in efficient price incorporated by the jump and (ii) noise component. Using these two components we compute an estimator of the information content of jumps relative to their noise component; i.e. the realized SN ratio given in equation

(4.19). Following the unobserved endogeneity discussion in section 4.6.1, we use a two stage least square model (TSLS) to examine predictive power of time t-1 fragmentation innovation on time t SN ratio. The second stage of the TSLS is as follows:

$$\begin{aligned} \tilde{F}^{D}_{i,t-1} &= \mu + \alpha_{11}\tilde{F}^{IV,D}_{i,t-1} + \alpha_{12}C_{i,t-1} + \epsilon_{1,i,t-1} \\ SN_{i,t} &= \zeta + \alpha_{21}\widehat{\tilde{F}^{D}}_{i,t-1} + \alpha_{22}C_{i,t-1} + \epsilon_{1,i,t-1} \end{aligned} \tag{4.29}$$

where the model is estimated conditional on observing a jump at time t, $\widehat{\tilde{F}}^{D}_{i,t-1}$ is the fitted value from the first stage regression which instruments fragmentation innovation by its systematic component. The matrix of control variables $C_i$ is identical to section 4.6.2. Our main parameter of interest is $\alpha_{21}$ which measures the effect of time t-1 fragmentation on time t jump SN ratio. A negative and economically significant value of $\alpha_{21}$ corresponds to the finding that jumps which proceed fragmented liquidity tend to be noisier as measured by their SN ratio.

### 4.7.2 Results

Table 4.9 presents second stage estimates from the 2SLS model (columns 1 and 2) and from the simple OLS model (columns 2 and 3). All intraday variables are normalized by their stock level standard deviations. This normalization allows us to interpret the coefficients as the effect of a one standard deviation change in the given variable on jump SN ratio. Amongst the intraday variables, fragmentation has the largest magnitude effect for jump SN ratios. In particular, a one standard deviation fragmentation innovation is associated with a SN ratio which is smaller by -7.05. Relative to the unconditional mean of SN ratio,

**Table 4.9: Effect of Fragmentation Shocks on Jump Signal to Noise Ratio** This table presents estimation results from the two stage least squares model (columns 3-4) and simple OLS model (columns 2-3). The dependent variable is the signal-to-noise ratio computed using the methodology proposed in 3. We define directional fragmentation innovation $\tilde{F}^{D}_{t-1}$, as $\tilde{F}^{Ask}_{t-1}$ ($\tilde{F}^{Bid}_{t-1}$) if time t price move is positive (negative). Effective spread ($ES_{t-1}$) is the signed difference between price and prevailing midpoint, measured relative to the prevailing midpoint at the time of trade. Price impact ($PI$) is the signed changed in midpoint following a trade, relative to prevailing midpoint at time of trade. Average dept ($Depth^V$) is the time weighted average of ask side and bid side offered volume. $Trades$ is cumulative traded volume at NASDAQ and NYSE in units of ten thousand shares. SPY superscript denotes S&P 500 ETF variables which proxy for systematic liquidity. The data is pooled and includes stock-year level controls as follows: average market capitalization (MCAP), average quoted spread computed from the previous years data; and listing defined as a dummy variable which equals one if the stock's primary listing is on NASDAQ and zero is NYSE. SPY return size ($|r|^{SPY}_{t-1}$) is the absolute value of time t-1 SPY return. The explanatory variables $\tilde{F}$, $ES$, $PI$, $Dept^V$, $ES^{SPY}$, $PI^{SPY}$ and $Dept^{V,SPY}$ are intraday variables computed at the 1 minute frequency. Size is computed as the logarithmic of average daily market capitalization for the previous year. Average quoted spread is computed as the difference between the close of day ask minus bid price, relative to the midpoint and averaged across the previous year. *Listing* dummy variable is defined as a dummy variable which equals one if the stock's primary listing is on NASDAQ and zero if NYSE. Volatility is computed as the daily return volatility of the previous year. Size, Quoted Spreads, Volatility and Listing are stock level controls that vary across stock-years. All intraday variables are normalized by their stock level standard deviation to facilitate economic interpretation. Each regression includes year dummies to control for year fixed effects. All variables are winsorized at the %1 level with the exception of listing and returns. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.001$

| Dependent Variable = Jump Signal to Noise Ratio ($SN_t$) | | | | |
|---|---|---|---|---|
| | 2SLS (Second Stage) | | Simple Regression | |
| Variable | Coeff. | T-stat | Coeff. | T-stat |
| Intercept | $-16.05^{***}$ | $-5.22$ | $-17.64^{***}$ | $-5.79$ |
| Frag. Inno ($\tilde{F}^{D}_{t-1}$) | $-7.05^{***}$ | $-11.33$ | $-0.47^{***}$ | $-4.35$ |
| Eff. Spread ($ES_{t-1}$) | $2.28^{***}$ | 22.32 | $2.08^{***}$ | 20.93 |
| Prc. Impact ($PI_{t-1}$) | $-0.04$ | $-0.36$ | $0.06^{***}$ | 0.56 |
| Depth ($Depth^{V}_{t-1}$) | $-1.88^{***}$ | $-8.43$ | 0.23 | 2.21 |
| Trades ($Trades_{t-1}$) | $1.20^{***}$ | 7.64 | 0.03 | 0.29 |
| SPY Eff. Spread ($ES^{SPY}_{t-1}$) | $-0.12$ | $-1.14$ | $-0.01$ | $-0.16$ |
| SPY Prc. Impact ($PI^{SPY}_{t-1}$) | $-0.02$ | $-0.21$ | $-0.02$ | $-0.23$ |
| SPY Depth ($Depth^{V,SPY}_{t-1}$) | $-0.47^{***}$ | $-4.23$ | $-0.69^{***}$ | $-6.43$ |
| SPY Ret. Size ($\|r\|^{SPY}_{t-1}$) | $0.43^{***}$ | 4.04 | $0.50^{***}$ | 4.78 |
| MCAP ($Size$) | $1.28^{***}$ | 8.84 | $1.29^{***}$ | 8.22 |

Table Continued

| | | | | |
|---|---|---|---|---|
| Quoted Spreads ($QS$) | 283.58*** | 21.57 | 224.01*** | 18.96 |
| Volatility ($Vol$) | −377.94*** | −15.18 | −295.89*** | −12.60 |
| Listing ($NASDAQ$) | −2.83*** | −10.41 | −2.75*** | −10.21 |
| Year Fixed Effect | Yes | | Yes | |
| R-square | 0.02 | | 0.02 | |
| Obs | 168,192 | | 168,192 | |

the aforementioned result corresponds to a 57% smaller SN ratio. This result provides strong evidence to suggest that when jumps follow fragmented liquidity, they tend to be less informed and more noisy, or put explicitly, tend to have smaller informational component as compared to their noise component.

## 4.8 Is Fragmented Liquidity Strategic?

In this section we perform test III: *Is fragmented liquidity during times leading up to jumps strategic?*

Since jumps are accompanied by large immediacy demand in the direction of jumps, liquidity suppliers risk accumulating large inventory positions during jumps. As shown in table 4.5, orderflow imbalance during jumps is both large and in the same direction as the jump; i.e. excess buy (sell) imbalance during positive (negative) jump price moves is large. If liquidity suppliers are fragmenting liquidity using their information to actively manage inventory risk then we should expect that any fragmentation measure which takes the direction of fragmentation into account contains more information for jumps, as compared to a measure which ignores the direction of liquidity fragmentation. As an example take the case where there is one stock and one liquidity supplier at time t-1. Further, a scheduled earning announcement is anticipated at time t. If the liquidity

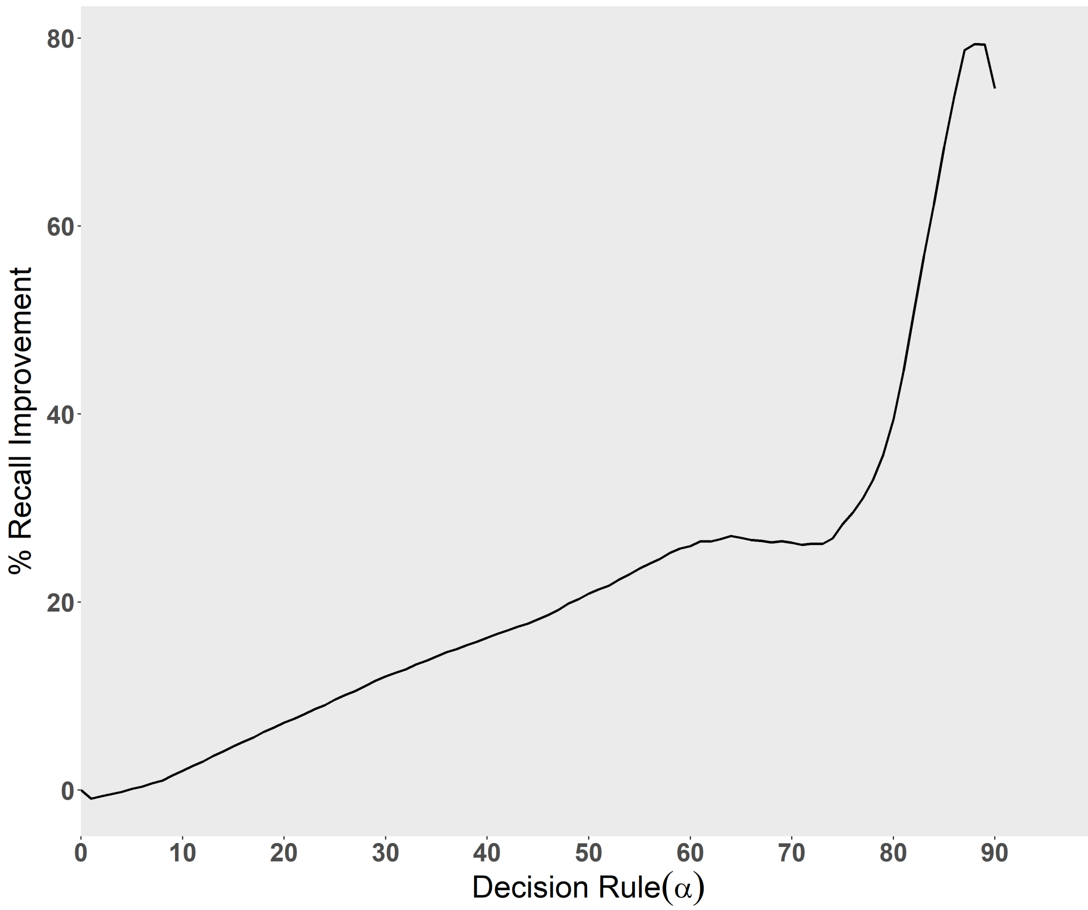



**Figure 4.6: Recall Improvement - Directional Fragmentation versus Average Fragmentation**
This figure plots recall improvement when using directional fragmentation versus average fragmentation as given in (4.31). The vertical axis shows percentage improvement and the vertical axis values of decision rule between 0 and 90%).

supplier's information set predicts a positive earnings surprise then she anticipates a large ask side immediacy demand at time t. Consequently, she will fragment ask side liquidity at time t-1 to manage inventory risk from incoming time t ask side orderflow imbalance. From the preceding example we note that if liquidity suppliers are both strategic and correct in anticipating immediacy demand, on average, then it must follow that the direction of fragmentation should predict the direction of jumps i.e. ask (bid) side fragmentation predict positive (negative) jumps. As a result a measure which takes into account both the size and direction of liquidity fragments should contain more information than a measure which ignores the direction of liquidity fragmentation. One such measure which ignores the direction of liquidity fragmentation is the average level of fragmentation defined as follows:

$$\tilde{F}_{i,t} = \frac{\tilde{F}_{i,t}^{Ask} + \tilde{F}_{i,t}^{Bid}}{2} \tag{4.30}$$

Comparing the directional fragmentation measure $\tilde{F}_{i,t}^{D}$ defined in (4.12) with $\tilde{F}_t$ it follows

that if liquidity fragmentation is strategic than we should expect the directional measure of fragmentation $\tilde{F}_{i,t}^{D}$ to contain more information than the average fragmentation measure $\tilde{F}_{i,t}$. In order to examine the aforementioned statement we estimate the IV probit model using $\tilde{F}_{i,t}$ and compare its recall measure to the model of section 4.6 which uses $\tilde{F}_{i,t}^{D}$. In particular we compute the relative improvement in recall as follows:

$$\text{Recall Improvement} = \frac{\text{Recall}^{D}(\alpha) - \text{Recall}(\alpha)}{\text{Recall}(\alpha)} \times 100 \tag{4.31}$$

Where $\alpha$ is the decision rule as defined in (4.27), $\text{Recall}^{D}(\alpha)$ $\big(\text{Recall}(\alpha)\big)$ is recall computed using directional (average) fragmentation. For a given alpha, a positive value of recall improvement implies that that $\tilde{F}_{i,t-1}^{D}$ is more informative for future jumps as compared to $\tilde{F}_{i,t-1}$. Or equivalent the direction in which liquidity is fragmenting contains more information for jumps in the next minute. Figure 4.6 plots recall improvement for values of alpha between 0% and 90%. We note that recall improvement is positive for all decision rules $\alpha$ between 5% and 90% which implies that $\tilde{F}_{i,t}^{D}$ has more information on average than $\tilde{F}_{i,t}$ in predicting price jumps. In addition, recall improvement is increasing with larger values for $\alpha$ implying that the information content of $\tilde{F}_{i,t}^{D}$ relative to $\tilde{F}_{i,t}$ is increasing with more stringent decision rule. These results provide strong evidence in support of liquidity suppliers strategically fragment liquidity in the same direction as their information is predicting future price jumps.

## 4.9 Conclusion

We examine the relationship between intraday liquidity fragmentation and price jumps for S&P 100 stocks. We find that unanticipated changes to fragmentation predict future price jumps. This suggests that liquidity suppliers are informed about large future price movements. In addition, we find that fragmentation, during times leading up to price jumps, is related to liquidity suppliers actively managing inventory risk based on their information about future liquidity demand and information arrivals.

More generally, we show that market linkages are important during times leading up to jumps. In particular, jumps are more informative and less noisier in consolidated markets as compared to fragmented markets. Future research could examine the welfare consequences of exogenous variations in fragmentation to guide policy.

## 4.10 Appendix: Fragmentation & Supply and Demand Mismatch

In this appendix section we show that liquidity fragmentation predicts mismatch in demand for immediacy and the supply of liquidity. We measure this aforementioned mismatch by constructing a variable, SDM defined in (4.5). Our modelling approach is the two stage least squares (TSLS) model similar to section 4.6.1. Following the discussion in section 4.5.1 we instrument fragmentation innovation by its systematic component and regress time t SDM on time t-1 fitted values from the first stage as follows:

$$\begin{aligned} \tilde{F}_{i,t-1} &= \mu + \gamma_{11}\tilde{F}^{IV}_{i,t-1} + \gamma_{12}C_{i,t-1} + \zeta_{1,i,t-1} \\ SDM_{i,t} &= \zeta + \gamma_{21}\widehat{\tilde{F}}_{i,t-1} + \gamma_{22}C_{i,t-1} + \zeta_{1,i,t-1} \end{aligned} \tag{4.32}$$

where the set of variables $C_{i,t}$ is matrix of control variables. The parameter $\gamma_{21}$ measures the predictive power of fragmentation innovation on SDM. Table 4.10 reports least square results of model (4.32). We normalize $SDM_{i,t}$ and all intraday variables by their respective stock level standard deviations for ease of economics interpretation. The results presented in table (4.10) show that a one standard deviation increase in time t-1 fragmentation increase time t SDM by 0.65 standard deviation. Based on these results we conclude that the degree of liquidity fragmentation predicts future increase in the mismatch between the supply and demand side of liquidity. These results are consistent with section 4.7 where we show that jumps which follow fragemted liquidity tend to be noisier on average.

**Table 4.10: Effect of Fragmentation Innovation on Liquidity Demand and Supply Mismatch**

This table presents estimation results from the two stage least squares model given in equations 4.32 (columns 2-3) and simple OLS model (columns 4-5). The dependent variables is the liquidity demand and supply mismatch normalized by its stock level standard deviation, defined as the number of trades larger than NBBO offered liquidity relative to number NBBO quote updates. We define directional fragmentation innovation $\tilde{F}^{D}_{t-1}$, as $\tilde{F}^{Ask}_{t-1}$ ($\tilde{F}^{Bid}_{t-1}$) if time t price move is positive (negative). Effective spread ($ES$) is the signed difference between trade price and prevailing midpoint, computed relative to the prevailing midpoint. Price impact ($PI$) is the signed change in midpoint 1 minute following a trade, computed relative to the prevailing midpoint. Volume dept ($dept^{V}$) is the time weighted average of ask side and bid side offered volume at NBBO quote.$Trades$ is cumulative traded volume at NASDAQ and NYSE in units of ten thousand shares. The superscript SPY denotes variables computed for the S&P 500 ETF which proxies for the market portfolio. SPY return size $|r^{SPY}|$ is the absolute value of time t-1 SPY return. The explanatory variables $\tilde{F}$, $ES$, $PI$, $Dept^{V}$, $ES^{SPY}$, $PI^{SPY}$ and $Dept^{V,SPY}$ are intraday variables computed at the 1 minute frequency. Size is computed as the logarithmic of average daily market capitalization for the previous year. Average quoted spread is computed as the difference between the close of day ask minus bid price, relative to the midpoint and averaged across the previous year. *Listing* variable is defined as a dummy variable which equals one if the stock's primary listing is on NASDAQ and zero if NYSE. Volatility is computed as the daily return volatility of the previous year. Size, Quoted Spreads, Volatility and Listing are stock level controls that vary across stock-years. All intraday variables are normalized by their stock level standard deviation to facilitate economic interpretation. Each regression includes year dummies to control for year fixed effects. All variables are winsorized at the %1 level with the exception of listing and return size. All variables are statistically significant at the %1 level unless otherwise stated.

| Dependent Variable = Liquidity Demand and Supply Mismatch ($SDM_t \times 100$) | | | | |
|---|---|---|---|---|
| | 2SLS (Second Stage) | | Simple Regression | |
| Variable | Coeff | T-stat | Coeff | T-stat |
| Intercept | 2.58 | 550.79 | 2.41 | 601.23 |
| Frag. Inno ($\tilde{F}^{D}_{t-1}$) | 0.65 | 476.70 | 0.01 | 99.56 |
| Eff. Spread ($ES_{t-1}$) | 0.04 | 257.09 | 0.06 | 430.10 |
| Prc. Impact ($PI_{t-1}$) | −0.02 | −168.71 | −0.00 | −8.27 |
| Depth ($Depth_{t-1}$) | 0.02 | 70.31 | −0.09 | −678.42 |
| Trades ($Trades_{t-1}$) | 0.02 | 61.82 | 0.10 | 669.26 |
| SPY Eff. Spread ($ES^{SPY}_{t-1}$) | −0.00 | −28.12 | −0.01 | −50.13 |
| SPY Prc. Impact ($PI^{SPY}_{t-1}$) | −0.01 | −72.09 | −0.01 | −51.95 |
| SPY Depth ($Depth^{SPY}_{t-1}$) | −0.01 | −62.95 | −0.00[1] | −2.12 |
| SPY Ret. Size ($\|r\|^{SPY}_{t-1}$) | 0.01 | 61.82 | 0.01 | 72.66 |

Table Continued

| | | | | |
|---|---|---|---|---|
| MCAP (*Size*) | −0.05 | −192.22 | −0.04 | −193.71 |
| Quoted Spreads (*QS*) | −29.20 | −1,362.00 | −26.17 | −1,498.10 |
| Volatility (*Vol*) | 26.35 | 698.30 | 23.07 | 726.82 |
| Listing | 0.19 | 418.37 | 0.17 | 451.71 |
| Year Fixed Effect | Yes | | Yes | |
| R-square | 0.07 | | 0.08 | |
| Obs | 53,311,147 | | 53,311,147 | |

[1]Estimate is not statistically significant at the %1 level.

# Chapter 5

# Discovering Efficient Value

Using machine learning methods we identify the efficient value and noise components of quarterly stock prices. We show that 28% of stock price variation is attributable to noise, and that 40% of noise is attributable to mutual fund trading. We find spikes in noise around the dot-com bubble, the 2008 financial crisis, and the European sovereign-debt crisis. Noise is higher for small firms and firms with high R&D expenditures. In an application of our methodology, we show that corporate managers do not have private information about future changes in efficient value nor can they identify noise in prices.

## 5.1 Introduction and Literature Review

Disentangling the efficient value and noise components of equity prices has a long history. Fisher Black famously noted in his 1985 presidential address that in the "...basic model of financial markets, noise is contrasted with information." Markets aggregate information from investors about efficient value in prices. While theoretically appealing, equity prices often deviate from efficient value when uninformed investors trade (Shleifer and Summers [1990]), when markets become illiquid, or when limits to arbitrage are binding (Brunnermeier and Pedersen [2009]). Identifying informed and uninformed investors and their order-flow is difficult as informed investors often hide their trading by mimicking the trading of uninformed investors (Kyle [1985a] and Collin-Dufresne and Fos [2015]).

More informative prices are more useful if they aggregate information not already possessed by managers and investors, thereby improving resource allocation and investment decisions. For instance, firm managers condition investment decisions on prices (Chen, Goldstein, and Jiang [2007]). Would-be acquirers are more likely to retreat from takeover efforts after negative stock price movements (Luo [2005]). Additionally, regulators often use market responses to understand the impacts of regulation on the economy. Disentangling permanent changes in efficient value from temporary deviations of prices away from efficient value is an important step towards understanding the relationship between market-aggregated information and real decision-making.

We contribute to this literature by developing a machine learning state space methodology (SSM) that uses mutual fund flows to identify permanent changes in unobserved efficient value and temporary deviations of the price from efficient value (i.e. noise). We use a simple model of stock prices and mutual fund trading during fire sale quarters to isolate the two components. Mutual fund fire sales have been used to indirectly identify transitory deviations from efficient value (Coval and Stafford [2007]) but clean identification is difficult. Our approach addresses the critique that previous measures of noise using mutual fund fire sales are correlated with returns for mechanical reasons (Wardlaw [2019]).

We decompose the previously-used mutual fund fire sales variable into a liquidity component that satisfies redemption requests and a discretionary component correlated with information. The rationale for this decomposition is that mutual fund managers do not have to sell each stock in proportion to current holdings. This grants managers discretion to sell a greater proportion of holdings in firms with negative expected future returns. This measure of discretionary buying or selling is based on the idea that, when a mutual fund manager must sell a certain proportion of her fund's holdings to fulfill investor redemption requests, deviations in sales from this proportion reveal the mutual fund manager's information.[1]

Using mutual funds' liquidity-driven and discretionary trading (as opposed to all trading) to identify noise in stock prices assumes that mutual fund discretionary trading is equivalent to informed trading. In order to justify this assumption, we show that as long as mutual fund managers observe similar signals to other investors and mutual fund managers correctly interpret the information content of prices and redemption requests by their investors (even with a lag), mutual fund trading is correlated with changes in efficient value.

We establish the empirical validity of our decomposition of mutual fund trading by first showing that actual net imbalances based on all trading by mutual funds correlate positively with current returns and negatively with future returns. This suggests that at least some of the correlation between actual net imbalances and returns is temporary and subsequently reversed. We also show that discretionary net share imbalances are positively correlated with returns. Importantly,

[1] This breakdown is similar to the predicted selling model in Huang, Ringgenberg, and Zhang [2019].

the discretionary measure is not associated with later reversals in stock prices, suggesting mutual funds' discretionary sales are correlated with permanent changes in the efficient value of firms rather than temporary deviations in the price from efficient value.

To confirm that our methodology is capturing information, we separate discretionary net imbalances into stock-quarters in which managers over-sell and under-sell holdings of those stocks, relative to their proportion of the fund's portfolio. Our data allow us to estimate how much a manager should actually sell in order to fulfill investor redemption requests. Managers that sell more of a particular stock than required to fulfill redemption requests may do so because they possess information about the future prospects of the firm.

Indeed, we find that mutual funds over-sell stocks in their portfolio that exhibit permanent decreases in efficient value. Stocks that managers under-sell exhibit positive contemporaneous changes in efficient value. Over-selling is a particularly salient event because while a mutual fund manager is forced to liquidate some of her holdings to satisfy redemption requests, she can also choose to satisfy some of the requests with cash holdings. This leads actual sales to be consistently lower than predicted hypothetical sales. By selling a security more than required to satisfy redemption requests and by opting not to use cash on-hand to meet these requests, thereby incurring liquidity costs to sell, the manager is expressing a strong opinion against the firm being over-sold. Our results show clearly that this behavior is associated with large negative permanent changes in efficient value.

By decomposing mutual fund fire sales into two parts and using these to estimate a state-space model, we merge the strands of literature that suggest fire sales: (1) contain information; and (2) cause prices to become noisier. Consistent with the prior literature, we find that observed mutual fund sales by funds experiencing significant investor redemptions are correlated with transitory price movements. The SSM suggests that noise variance is roughly 28% of total price variance and that 39% of that variance can be explained by actual mutual fund trading in a given stock-quarter. Put differently, roughly 11% of total variance is noise variance that can be attributed to mutual fund fire sales. We find spikes in noise during the dot-com bubble, the 2008-09 financial crisis, and the European sovereign debt crisis.

We also find an informational component in fire-sale-induced mutual fund trading. Mutual fund managers' deviations from predicted selling (i.e., discretionary trading) is correlated with unobserved changes in the efficient value of firms. The SSM parameter Kappa ($\kappa$) measures this correlation between discretionary trading by mutual funds and permanent price changes. Kappas are positive and significant, suggesting that when mutual fund managers exercise discretion in their trading they do so for informed reasons. Moreover, their information varies with observable firm characteristics and the quality of information about the firm, consistent with a noisy rational expectations equilibrium in which investors update their beliefs based on both their private information and observed prices (Hellwig [1980], Admati [1985], Banerjee [2011]). Correlations between mutual fund managers' discretionary trading and changes in efficient value are greater for larger firms, more profitable firms, and firms that invest more in R&D relative to assets.

That mutual fund managers possess private information is discussed in a long literature. For instance, Bollen and Busse [2001] show that mutual fund managers possess the ability to time market movements, whereas Frazzini and Lamont [2008] suggest that mutual fund investors and managers constitute "dumb money." In a related paper, Huang et al. [2019] show that mutual fund fire sales are correlated with permanent price changes in equity prices. Another literature suggests that transitory price movements (noise) due to mutual fund fire sales affect corporate policies. Phillips and Zhdanov [2013] demonstrate the relationship between mutual fund fire sales and R&D expenditures. Derrien, Kecskés, and Thesmar [2013] show how corporate payout policy varies with mutual fund fire sales. Chen et al. [2007] relate the firm's own stock prices to corporate investment.

Our paper is related to Brogaard, Nguyen, Putniņš, and Wu [2020], Bai, Philippon, and Savov [2016] and Dávila and Parlatore [2018]. Brogaard et al. [2020] define noise as the residual term in a return regression, where the residual is possibly correlated with unobserved sources of information. In contrast, we explicitly define noise as the portion of stock price changes correlated with the liquidity component of mutual-fund trading. Our results also differ from Bai et al. [2016], who find that price informativeness (noise) has increased (decreased) over time. However, our findings complement Dávila and Parlatore [2018], who find substantial cross-sectional variation in price informativeness.

Our methodology can be applied to numerous first order questions in finance research. For instance, a significant challenge lies in making inferences about the channel through which stock prices are associated with corporate investment (Dessaint, Foucault, Frésard, and Matray [2019]). Based on their study we present an application of our measures of efficient value and noise and estimate versions of corporate investment-price sensitivity relationships using both the firm's own noise and peer noise. We show that firm investment is sensitive to changes in the efficient price and noise, with a greater sensitivity to changes in efficient value than noise. Extending this insight we decompose information into a part that is knowable at the time of the investment (public information), and a part that is revealed in the future (private information). These results indicate that managers do not have private information about changes in efficient value, implying that managers do not possess information not already in prices.

Our methodology can be applied to better understand the information sets and decision-making of corporate insiders, activists, or arbitrageurs. It is flexible and can be modified to include the trading of other institutional investors (e.g., closed-end funds, exchange-traded funds, firm executives and directors) or retail traders. For instance, we estimate the model for all mutual trading (rather than only those undergoing fire sales) and show that our results also hold in this setting. We estimate our model at the quarterly frequency but the model could also be estimated at the intraday, daily, weekly or monthly frequencies depending on data availability and the research question. Finally, we assume an error correlation structure suggested by a number of theoretical models (Wang [1994] and Hendershott and Menkveld [2014]), however the model can be estimated with more general error correlation assumptions.

## 5.2 Background and model of stock prices

Investors use models to update their beliefs about the efficient value of a firm. Most investors will use some form of Bayesian updating, in that they will have an opinion (prior) about the efficient value of a firm. The prior could be based on current and past prices (or price movements), and/or current and past trading. At some point, new information will arrive that updates the prior information of the investors. The investor will then generate a posterior estimate of the efficient value depending on how much weight they wish to assign to the new information. This representation maps directly into the Kalman filter that we use to estimate efficient value. For instance, the estimate of efficient value generated by a Kalman filter would be similar to the estimate generated by a Bayesian investor where the updating function from the prior to posterior estimate is the Kalman gain.

Our model takes an additional step. Since the Bayesian investor is not endowed with any specific informational advantage, they may not possess useful information about efficient value. We extend the logic to a Bayesian investor who knows all future prices and trading. This type of investor would be closer to a Walrasian who knows all past, current, and future prices, volumes and the relationships between these variables. The Walrasian is comparable to our Kalman smoother who uses these same variables to generate estimates of the efficient value for every firm-quarter in our sample.

Our measure of noise is simply the difference between the estimate of efficient value generated by the Walrasian-type investor (Kalman smoother) and the observed price. This intuitive decomposition identifies information that enters into price through the trading process and noise that is generated by trading.

### 5.2.1 Identifying informed and uninformed trading

In order to motivate the empirical analysis, we consider asset prices in the context of a noisy rational expectations equilibrium (REE). In REE models, investors learn about the value of the firm by observing stock prices and a (noisy) signal of firm value (Grossman and Stiglitz [1980]). In the case of different private signals across many investors, stock prices play an information aggregation role (Hellwig [1980]).

A key feature in these models is that there is “noise” in the price system, either in the form of supply shocks (Hellwig [1980]), traders who trade randomly (Kyle [1985b]), or forecast errors by uninformed investors (Wang [1994]). Noise is an essential feature of the market in order to incentivize private (costly) information production (Grossman and Stiglitz [1980]). The existence of noise in the price system implies that prices respond to both informed and uninformed trading. Uninformed trading will lead prices to deviate from efficient value and subsequent price reversals as investors learn more about the value of the firm. Informed trading will lead prices toward efficient value and will not lead to subsequent price reversals.

Consistent with a noisy REE, Coval and Stafford [2007] find strong price declines following forced assets sales in mutual funds. These forced sales lead to reversals in prices, with the average reversal lasting several calendar quarters. The length and the relative size of the average reversal suggests that these forced sales are not caused purely by noise (either in the form of supply shocks or random trading), implying that fire sales are also correlated with information. Mutual fund managers may therefore be exercising discretion and acting on information (public or private) in the selection of which stocks to sell, and the proportion that they sell, in response to large mutual fund redemption requests.

In order to distinguish between noisy and informed trading, we decompose mutual fund sales into two components based on observed sales of firm i's stock across mutual funds j that hold stock in firm i at time t-1:

$$Act_{i,t} = \sum_{j} (Shares_{i,j,t} - Shares_{i,j,t-1})/Volume_{i,t}. \tag{5.1}$$

Actual sales in quarter t equals the change in shares held from quarter t-1 to t by all funds j that: (1) hold the stock at the end of quarter t-1; and (2) experience flows of assets under management of less than or equal to -5%. We divide the summation by share volume of firm i in quarter t. Note that the scaling variable is measured in shares traded in quarter t and does not rely on price. Constructing the measure of sales (and discretionary sales below) in this way, rather than using dollar changes in holdings and dollar volume as a scaling variable, addresses the Wardlaw [2019] critique that instruments based on mutual fund fire sales are invalid because they mechanically incorporate returns. Conditioning sales on the subset of fund-quarters where the fund experience flows less than or equal to -5% of assets under management identifies fire-sale quarters (FS quarters).

We define discretionary imbalances as the difference between actual sales and what the fund would have sold in response to redemption requests had it sold shares in all of its stocks according to their proportion of AUM. That is,

$$Disc_{i,t} = \sum_{j} (\Delta Shares_{i,j,t} - Flow_{j,t} \times Shares_{i,j,t-1})/Volume_{i,t} \tag{5.2}$$

where $Flow$ is the percentage change in AUM net of returns earned over the quarter, and the summation is across mutual funds experiencing flows less than -5% in quarter t.

We exploit this feature of mutual fund fire sales (i.e., coexistence of noise and informed trading) to decompose stock prices into efficient value and noise via a state space model representation of such prices. The next section provides a brief description of this model.

### 5.2.2 Model of stock prices

Our model of stock prices is based on a dealer-inventory model described in Glosten and Milgrom [1985] and Hendershott and Menkveld [2014]. The latter characterize the dealer's problem as a stochastic optimal linear regulator problem. This intermediary holds inventory of stock and supplies liquidity to investors. The intermediary can either be long or short inventory, though she prefers a net zero position. As she is risk averse, she will bid up prices in response to buy orders (thereby increasing her inventory), and mark down prices in response to sell orders (thereby decreasing her inventory). All market orders are assumed to pass through a dealer (i.e. there is no limit-order book).

Solving the dealer's problem yields the following structural model of stock prices:

$$\begin{aligned} p_t &= m_t + s_t \\ m_t &= m_{t-1} + \kappa(\Delta I_t - E_{t-1}[\Delta I_t]) + \eta_t \\ s_t &= \phi s_{t-1} + \beta \Delta I_t + \epsilon_t \end{aligned} \tag{5.3}$$

where $p_t$ equals the log of price and $\Delta I_t$ equals the change in dealer's inventory resulting from buy and sell orders.

The model of stock prices has two components. Efficient value is represented by $m_t$. This component is adjusted each period for trade and non-trade related information. Here, $\Delta I_t - E_{t-1}[\Delta I_t]$ represents the information conveyed by the unexpected component of inventory changes due to trade, and $\eta_t$ represents non-trade news that arrives between time t-1 and t. The variances of $\eta_t$ and $\epsilon_t$ can be interpreted as the impact of information and market frictions on security prices.

The second component ($s_t$) represents noise, or a transitory deviation of the price from efficient value. This process is mean-reverting and the coefficient $\beta$ reflect the impact of market frictions on price. For example, initiators of buy orders pay more than efficient value to compensate the dealer for the resulting inventory imbalance.

The main feature of the model is that order arrivals convey information and cause prices to deviate from true values.[2] This feature can be seen from the fact that the transitory component $s_t$ is correlated with the trading-related component of information ($\Delta I_t - E_{t-1}[\Delta I_t]$) because $\Delta I_t$ is common to both. In summary, changes attributed to information have a permanent effect on prices while changes attributed to pricing errors have a temporary effect on prices.

### Kalman filter representation

Stock prices in equation (5.3) have a Kalman filter representation, which we use to estimate the parameters. The filtering equation is linear in the prediction errors. Based on this representation, investors update their conditional expectations after observing prices.[3] Let $m_{t|t} := E(m_t|v_t)$, where $v_t$ is the prediction error form observing prices. A similar definition for $s_{t|t}$ obtains. Then the stock price model implies,

$$\begin{pmatrix} m_{t|t} \\ s_{t|t} \end{pmatrix} = \begin{pmatrix} E(m_t|v_{t-1}) \\ E(s_t|v_{t-1}) \end{pmatrix} + \mathbf{K}_t v_t \tag{5.4}$$

[2] George and Hwang [2001] also discuss how to estimate a similar representation of stock prices.

[3] Wang [1994] proposes a similar formulation.

where $\mathbf{K}_t$ is a $2 \times 1$ matrix with positive components and $v_t$ is the prediction error or surprise in prices, $p_t - E(m_t|v_{t-1})$. See Internet Appendix for the derivation.

The first term on the right-hand side gives the expectation of efficient value and noise based on previous information. The second term gives the update in expectations based on new information from surprises in price. The first component of $\mathbf{K}_t$ is the Kalman gain, or the weight given to the price signal about efficient value $m_t$ by investors. Thus, estimates of noise and efficient value are based on investors' updating of the conditional distributions of efficient value and noise after observing prices, where the weight put on prices is given by $\mathbf{K}_{(1)t}$.

Price informativeness in stock prices can be recovered from the recursion estimates in (5.4). Price informativeness, $\tau_{\hat{m},t}$, is defined as the precision of the signal about efficient value contained in price $p_t$ after observing the time-t prediction error $v_t$. Thus

$$\begin{aligned} \tau_{\hat{m},t} &:= Var(m_t|v_t)^{-1} \\ &= [Var(m_t|v_{t-1})(1 - \mathbf{K}_{(1)t}]^{-1} \end{aligned} \tag{5.5}$$

Price informativeness increases with the weight investors put on price surprises (the Kalman gain), and it decreases with the variance of efficient value conditional on time t-1 information. In the limiting case where the Kalman gain is close to one, investors put all the weight on the price as a signal of innovation in efficient value, disregarding all prior information.[4]

[4] Dávila and Parlatore [2018] link price informativeness to the Kalman gain.

### Assumptions required for estimation

The stochastic inventory control model implies that changes in inventory affect the transitory component of stock prices, but only unexpected changes (or innovations) in inventory affect the permanent component of price. Intuitively, only the unexpected portion of information conveyed by inventory changes should matter for efficient value.

However, because we only observe mutual fund trading, we must make assumptions about the information available to different types of investors. More specifically, we assume that different sets of investors' (e.g., mutual funds' and hedge funds') trading is correlated through access to common sources of information. Therefore, trading by informed and less-informed investors differs only in quantity. This assumption implies the independence between the error terms $\epsilon$ and $\eta$, which reduces the number of parameters required to estimate the state space model.

This assumption can be relaxed by allowing observed mutual fund trading and unobserved trading to be only partially correlated through a common source of information. For example, if more-informed investors have access to an orthogonal signal that is unobserved by less-informed investors. We show in the Appendix that this type of information structure generates a correlation between the error terms that is proportional to the variance of information observed by the more informed investor. In terms of estimating the model, allowing this more general information structure across investors adds an additional parameter to the SSM.

## 5.3 Data and descriptive results

Estimation of the state space model uses both stock-level returns and quarterly mutual fund holdings information. We obtain necessary data from CRSP, Compustat, CRSP Mutual Funds and Thomson Reuters to run the state space model and the investment regressions. We use Thomson Reuters to determine the number of shares of firms held by mutual funds, while CRSP Mutual Funds provides information on the performance of these funds. The intersection of these two datasets allows us to calculate FLOW, as well as both hypothetical and actual sales of stocks across funds for each firm-quarter. The intersection of CRSP, Compustat and the mutual fund-based data from 1990-2016 leaves us with 589,184 observations. We remove firms which do not have at least eight years (32 quarters) of non-missing values for our test variables. Next, since we require variation in our mutual fund trading variables to identify our state space estimators, we remove firms that do not have at least three non-zero observations for both hypothetical sales and actual sales.

Finally, we remove firms with average total assets below $5M, as well as financial and utility firms (2-digit SIC codes 49, 60-69, 90+). The sample with which we proceed to state space model estimation spans 273,023 observations (4,417 unique stocks). Estimating the recursion in equation (5.4) eliminates some firms in cases where the maximum likelihood estimator does not converge. Our estimation model converges for 72% of firms in fire sale models and 70% in the all mutual fund trading model, resulting in a combined sample of 2,581 firms.

### 5.3.1 Summary statistics

Summary statistics for the sample are provided in Table 5.1. As shown, firms in the final sample are large, measured either in terms of market capitalization (MCAP) or total assets (Assets), with the average firm having $3.4 billion in assets measured in current dollars. Firms in the sample have average daily bid-ask spreads of 1.55% and daily trading volume of 1.21 million shares (median trading volume of 120,000 shares). Firms are also widely held by mutual funds, with the average firm being held by 62 funds (median is 29).

Table 5.3 presents mutual-fund level statistics for the subset of firm-calendar quarters characterized by fire sales (FS). Following Chen et al. [2007], fire-sale quarters are defined as quarters in which at least one mutual fund holding a given stock in our sample experiences of flows less than or equal to -5% of AUM. As is shown, FS-quarters comprise 25% of the sample, and conditional on a FS-quarter, flows are substantial, amounting to 11.5%of AUM on average (median is 9.5%). Mean net imbalance ($Act$) in response to fire sales is -0.32% of trading volume, and the median is zero, suggesting that funds do not sell stocks proportionally to AUM in response to flows.

### 5.3.2 Discretionary Imbalance

If funds incorporate information in their decision to sell stocks during FS-quarters, then discretionary net imbalances (buying - selling) should be different from zero. As shown in Table 5.3, average discretionary net imbalances ($Disc$) are 0.68% of share volume, indicating that mutual funds' share sales in response to redemption requests are correlated with information about the firm's value. We explore this

**Table 5.1: Summary Statistics - Sample Stocks**
This table reports descriptive statistics for the 2,581 stocks used in our main analysis. The sample time period is 1990-2016. For each variable, the table reports the mean, 25th percentile, median, 75th percentile and standard deviation. Stock-Quarters is the number of quarters a sample stock exists in the CRSP and Compustat universe with non-missing statistics required in our main analysis. Assets is end of quarter book value of total assets reported in COMPUSTAT Fundq file. MCAP is end of quarter price times shares outstanding from CRSP MSF file and averaged across all sample stocks and sample time periods. Returns (Ret.) are computed as the logarithmic difference of end of quarter price. Standard deviation of returns are computed for each stock and quarter using end of day price from CRSP DSF file. Spreads are computed as the average of end of day ask price minus bid price relative to the prevailing bid-ask midpoint in the CRSP DSF file. Volume is total quarterly dollar traded volume for a given stock. Shrout is end of quarter shares outstanding for a given stock, reported in millions of shares. Funds held is the number of mutual funds holding at least one sample stock. FS Funds held is the number of funds experiencing a fire sale and holding at least one sample stock. Investment (Inv.) is capital expenditure (Capxq) over of given quarter relative to property planet and equipment in the previous quarter reported in COMPUSTAT Fundq file. Cash-to-assets (CF/Assets) is defined as end of quarter operating cash flow relative to end of quarter assets reported in the COMPUSTAT Fundq file. Debt-to-assets (Debt/Assets) is total end of quarter debt relative to end of quarter assets reported in the COMPUSTAT Fundq file. All dollar denoted variables are deflated using the quarterly seasonally adjusted GDP deflater from FRED database. Price, Volume and MCAP are adjusted for stock splits and distributions. In addition, all variables are winsorized at the 1% level in each tail.

| # Stocks | Variable | Units | Mean | P25 | Median | P75 | Std |
|---|---|---|---|---|---|---|---|
| 2,581 | | | | | | | |
| | Stock-Quarters | # quarters - across stocks | 62.40 | 42.00 | 58.00 | 82.00 | 22.62 |
| | MCAP | $B. quarterly - across stocks | 2.31 | 0.10 | 0.37 | 1.36 | 6.52 |
| | Assets | $B. quarterly - across stocks | 3.42 | 0.12 | 0.42 | 1.72 | 10.26 |
| | Price | $ quarterly - across stocks | 27.15 | 7.31 | 15.48 | 30.03 | 44.85 |
| | Ret | % quarterly - across stocks | −0.32 | −13.47 | 0.85 | 14.13 | 26.85 |
| | Std. Ret. | $ quarterly - across stocks | 0.03 | 0.02 | 0.03 | 0.04 | 0.02 |
| | Spreads | % daily - across stocks | 1.55 | 0.12 | 0.65 | 2.03 | 2.26 |
| | Volume | $M. quarterly - across stocks | 1.21 | 0.02 | 0.12 | 0.70 | 3.29 |
| | Shrout | M. Shr. quarterly - across stocks | 82.87 | 10.11 | 26.56 | 64.13 | 182.06 |
| | Funds held | # Funds - across quarters | 61.69 | 8.00 | 29.00 | 88.00 | 78.89 |
| | FS Funds held | # FS Funds - across quarters | 14.19 | 1.00 | 4.00 | 16.00 | 24.45 |
| | Inv. (Capxq/PPENTq) | % quarterly - across stocks | 7.44 | 2.62 | 4.90 | 8.96 | 8.44 |
| | CF/Assets | % quarterly - across stocks | 1.05 | 0.72 | 2.06 | 3.31 | 5.07 |
| | Debt/Assets | % quarterly - across stocks | 23.50 | 5.15 | 20.69 | 35.39 | 20.87 |

idea in more detail by estimating residual discretionary net imbalances:

$$Disc_{i,t} = \xi_{i,t} + \gamma_i + \alpha_1 \text{Size}_{i,t-1} + \alpha_2 \text{Spreads}_{i,t-1} + \alpha_2 \text{Shrout}_{i,t-1} + \alpha_3 \text{Volatility}_{i,t-1} + \widetilde{Disc}_{i,t} \quad (5.6)$$

**Table 5.2: Panel Regression - FS Discretionary Imbalance**
This table presents results of OLS panel regressions of time t discretionary net purchases on time t-1 logarithmic assets (size), logarithmic shares outstanding (Shrout) and quoted spread. Each panel regression contains stock fixed and year fixed effects. All explanatory variables are normalized by their standard deviation to facilitate economic interpretation. All variables are winsorized at the %1 level in each tail. Reported T-statistics are computed using Heteroscedasticity-Corrected Covariance Matrix (HCCME 1). * $p < 0.01$; ** $p < 0.05$; *** $p < 0.001$

| Dependent Variable: FS Discretionary Imbalance $_{i,t}$ | | | | | |
|---|---|---|---|---|---|
| Explanatory Variables: | | | | | |
| Log Size $_{i,t-1}$ | $-0.35^{***}$ | | | | $-0.33^{***}$ |
| | $(-11.14)$ | | | | $(-8.98)$ |
| Spread $_{i,t-1}$ | | $0.10^{***}$ | | | $0.15^{***}$ |
| | | $(5.24)$ | | | $(7.73)$ |
| Volatility $_{i,t-1}$ | | | $-0.21^{***}$ | | $-0.26^{***}$ |
| | | | $(-19.13)$ | | $(-22.97)$ |
| Log Shares Outstanding $_{i,t-1}$ | | | | $-0.25^{***}$ | $-0.10^{***}$ |
| | | | | $(-9.65)$ | $(-3.36)$ |
| | | | | | $(7.73)$ |
| $R^2$ | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 |

The residual component removes the predictable component from discretionary net imbalances ($Disc$) related to liquidity and volatility. For example, mutual funds might sell the most liquid stocks in order to minimize on trading costs.

Estimates of this regression are reported in Table 5.2. As expected, larger, more liquid stocks in a mutual fund's portfolio are more likely to be sold in fire sale quarters. By construction, the residual discretionary imbalances are zero on average, but, as shown in the last row of Table 2, residual discretionary sales exhibit substantial variation around the mean with an inter-quartile range of 1.33% of share volume.

We next measure the price impact of discretionary and actual selling as a validation check of our measure of discretionary selling. If discretionary sales are correlated with information, then price changes in FS-quarters with a large discretionary component of observed sales should be permanent. That is, we should not observe price reversals following the FS-quarter. In contrast, if a large component of actual sales in response to flows is uncorrelated with information, then these sales should be more likely to be followed by significant price reversals. We test both ideas in Table 5.5, which reports estimates of the following price-impact regressions:

$$ret_{it} = \alpha + \beta_1 Act_{it} + \beta_2 Act_{it-1} + \varepsilon \tag{5.7}$$

$$ret_{it} = \alpha + \beta_1 \widetilde{Disc}_{i,t} + \beta_2 \widetilde{Disc}_{i,t-1} + \varepsilon \tag{5.8}$$

As shown in Panel A, *Act* is positively and significantly associated with current quarter returns but lagged values of *Act* are negatively and significantly associated with current returns. This result indicates that part of the current impact of mutual fund trading on prices is reversed during the next quarter. Panel B reports equivalent estimates for price impacts of discretionary residual net imbalances

(Disc. residual, see equation 5.6). Consistent with the idea that the discretionary

**Table 5.3: Summary Statistics - Fire Sales Funds**
This table reports descriptive statistics for Fire Sales funds in our sample. The sample time period is 1990-2016. For each variable, the table reports the mean, 25th percentile, median, 75th percentile and standard deviation. In a given quarter we identify fire sale funds as those funds with market performance-adjusted net asset outflow greater than 5 percent (i.e., $Flow_{j,q} < -0.05$). Fire Sales fund quarter frequency (FS fund Qtr. freq.) is the number of fire sale fund-quarters divided by the total number of fund-quarters. Hypothetical Net imbalance is a stock-quarter variable equal to fire sales funds' outflow multiplied by their holdings of the stock at the beginning of the quarter, scaled by share trading volume. The variable Act is a stock-quarter variable equal to fire sales funds' net imbalance of the stock in the quarter, scaled by share trading volume. Discretionary net imabalance (Disc) measures the discretionary component of mutual fund buying-selling. Disc. residual is the residual component from a panel regression of Disc. on size, volatility, spread and shares outstanding. All variables are winsorized at the 1% level in each tail. Further details of the variables in this table are given in Appendix A.

| # Funds | Variable | Units | Mean | P25 | Median | P75 | Std |
|---|---|---|---|---|---|---|---|
| | FS fund Qtr. freq. | % quarterly | 25.12 | 10.28 | 19.17 | 33.35 | 19.63 |
| | Flow | % quarterly | −11.50 | −13.86 | −9.21 | −6.73 | 6.58 |
| | Act | % quarterly | −0.32 | −0.50 | 0.00 | 0.06 | 2.68 |
| | Disc | % quarterly | 0.68 | −0.05 | 0.00 | 0.75 | 3.14 |
| | Disc. residual | % quarterly | −0.00 | −0.95 | 0.00 | 0.38 | 2.96 |

component is correlated with information, the contemporaneous effect on prices of this component of mutual fund net imbalances is significantly positive, but the impact of lagged values of this variable is small and statistically insignificant.

## 5.4 Estimating efficient value

Having established that discretionary trading is correlated with information, we next solve the signal extraction problem described by equation (5.4) for the sample of stocks described in the preceding section. Estimation is done stock-by-stock by obtaining initial estimates of the parameters assuming there is no auto-correlation

**Table 5.4: Panel Regression - Discretionary Imbalance / Net Imbalance**
This table presents results of OLS panel regressions of time t discretionary purchases on time t-1 logarithmic assets (size), logarithmic shares outstanding (Shrout) and quoted spread. In Panel A the dependent variable is discretionary fire sales imbalance, and in Panel B net discretionary imbalance. Each panel regression contains stock fixed and year fixed effects. All explanatory variables are normalized by their standard deviation to facilitate economic interpretation. All variables are winsorized at the 1% level in each tail. Reported t-statistics are computed using a Heteroscedasticity-Corrected Covariance Matrix (HCCME 1) and clustered by stock and year. * $p < 0.01$; ** $p < 0.05$; *** $p < 0.001$

| | | | | | |
|---|---|---|---|---|---|
| Panel A: FS Discretionary Imbalance $_{t-1}$ | | | | | |
| Log Size $_{i,t-1}$ | $-0.35^{***}$ | | | | $-0.33^{***}$ |
| | $(-11.14)$ | | | | $(-8.98)$ |
| Spread $_{i,t-1}$ | | $0.10^{***}$ | | | $0.15^{***}$ |
| | | $(5.24)$ | | | $(7.73)$ |
| Volatility $_{i,t-1}$ | | | $-0.21^{***}$ | | $-0.26^{***}$ |
| | | | $(-19.13)$ | | $(-22.97)$ |
| Log Shrout $_{i,t-1}$ | | | | $-0.25^{***}$ | $-0.10^{***}$ |
| | | | | $(-9.65)$ | $(-3.36)$ |
| | | | | | $(7.73)$ |
| $R^2$ | 0.10 | 0.10 | 0.10 | 0.10 | 0.10 |
| #Stocks | 4,228 | 4,228 | 4,228 | 4,228 | 4,228 |
| #Obs | 214,173 | 214,173 | 214,173 | 214,173 | 214,173 |
| Panel B: Net Discretionary Imbalance $_{t-1}$ | | | | | |
| Log Size $_{i,t-1}$ | $-0.02$ | | | | $-0.40^{***}$ |
| | $(0.34)$ | | | | $(-6.08)$ |
| Spread $_{i,t-1}$ | | $-0.48^{***}$ | | | $-0.33^{***}$ |
| | | $(-14.64)$ | | | $(-9.71)$ |
| Volatility $_{i,t-1}$ | | | $-0.52^{***}$ | | $-0.45^{***}$ |
| | | | $(-22.57)$ | | $(-19.00)$ |

Table Continued

| | | | | | |
|---|---|---|---|---|---|
| Log Shrout $_{i,t-1}$ | | | | 0.08 | 0.22*** |
| | | | | (1.60) | (3.79) |
| | | | | | (7.73) |
| $R^2$ | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |
| #Stocks | 4,417 | 4,417 | 4,417 | 4,417 | 4,417 |
| #Obs | 267,736 | 267,736 | 267,736 | 267,736 | 267,736 |

/noindent in noise using repeated iterations (250) of the Kalman recursion. This produces starting values for the parameters which we combine with a grid search for the auto-correlation coefficient parameter. We then estimate the full set of parameters with maximum likelihood.[5]

### 5.4.1 State space model estimates

Estimates are reported in Table 5.6, where mean and median values and confidence intervals of the parameters over all stocks are reported. Reported estimates are based on log prices in order to facilitate the interpretation (differences in log prices equal returns). The first row reports average and median values of $\kappa$, which measures the relationship between efficient value with discretionary trading. On average kappas are significantly positive and lie in a 95% confidence interval bounded below by 0.83 and above by 1.53. This shows that, a significant portion

[5]Internet Appendix provides more detail on this procedure.

**Table 5.5: FS Price Impacts**
This table reports OLS panel regression estimates of equation(5.7), price impact regressions for our sample of 2,581 stocks. The dependent variable is the stock return. Panel A reports results for actual Net Imbalances (buying-selling) and panel B for the Discretionary Residual Net Imbalances. All explanatory variables are normalized by their standard deviation to facilitate economic interpretation and are winsorized at the 1% level in each tail. Reported t-statistics are computed using Heteroscedasticity-Corrected Covariance Matrix (HCCME 1). * $p < 0.01$; ** $p < 0.05$; *** $p < 0.001$.

Dependent Variable: Qtr. Returns

Panel A: Panel Regression - Actual Net Imbalance

| # Obs | Variable | Estimate | $R^2$ |
|---|---|---|---|
| 158,480 | Act. Net Imbalances$_{i,t}$ | 1.62*** | 0.02 |
| | | (23.05) | |
| | Act. Net Imbalances$_{i,t-1}$ | −0.24*** | |
| | | (−3.40) | |

Panel B: Panel Regression - Discretionary Residual Net Imbalance

| # Obs | Variable | Estimate | $R^2$ |
|---|---|---|---|
| 158,480 | Disc. Res. Net Imbalances$_{i,t}$ | 1.80*** | 0.02 |
| | | (26.69) | |
| | Disc. Res. Net Imbalances$_{i,t-1}$ | −0.03 | |
| | | (−0.47) | |

**Table 5.6: Main Results: FS State Space Model (SSM)**
This table presents estimation results from the Fire Sales state space model given in equation 5.3. The total number of converged stocks were 2,581 which represent 72% of our sample stocks. All reported statistics are cross-sectional means across stock estimates. t-statistics for the means are reported in parenthesis. $\kappa$ is the mean of estimated correlations between stock permanent price innovations and discretionary component of sales. $\beta$ is the mean of estimated correlation between stock transitory price innovation and actual sales and $\phi$ is the mean of estimated autoregressive component in transitory price. % Noise is the size weighted mean of transitory price contribution to total price variation; % Disc. contribution to information is size weighted mean of discretionary sale contribution to permanent price variation; % Act. contribution to noise is size weighted mean of actual sales contribution to transitory price variation. 95% Confidence intervals for the means are presented in column 4 and percent of estimated stock $\kappa$ and $\beta$ which are positive and statistically significant at the 10% level are reported in column 5.

| Variable | Mean | Median | 95% CI | |
|---|---|---|---|---|
| Disc. Res. Comp $_t$ ($\kappa$) | 1.18 | 1.50 | (0.83 | 1.53) |
| | (6.61) | | | |
| Act. Sales $_t$ ($\beta$) | 1.62 | 0.61 | (1.29 | 1.95) |
| | (9.67) | | | |
| AR(1) ($\phi$) | 0.30 | 0.48 | (0.33 | 0.28) |
| | (23.34) | | | |
| % Noise | 28.01 | 16.29 | (26.92 | 29.09) |
| | (50.59) | | | |
| % Disc. contribution to information | 21.68 | 7.39 | (20.46 | 22.89) |
| | (35.05) | | | |
| % Act. contribution to noise | 39.26 | 26.64 | (37.90 | 40.62) |
| | (56.67) | | | |

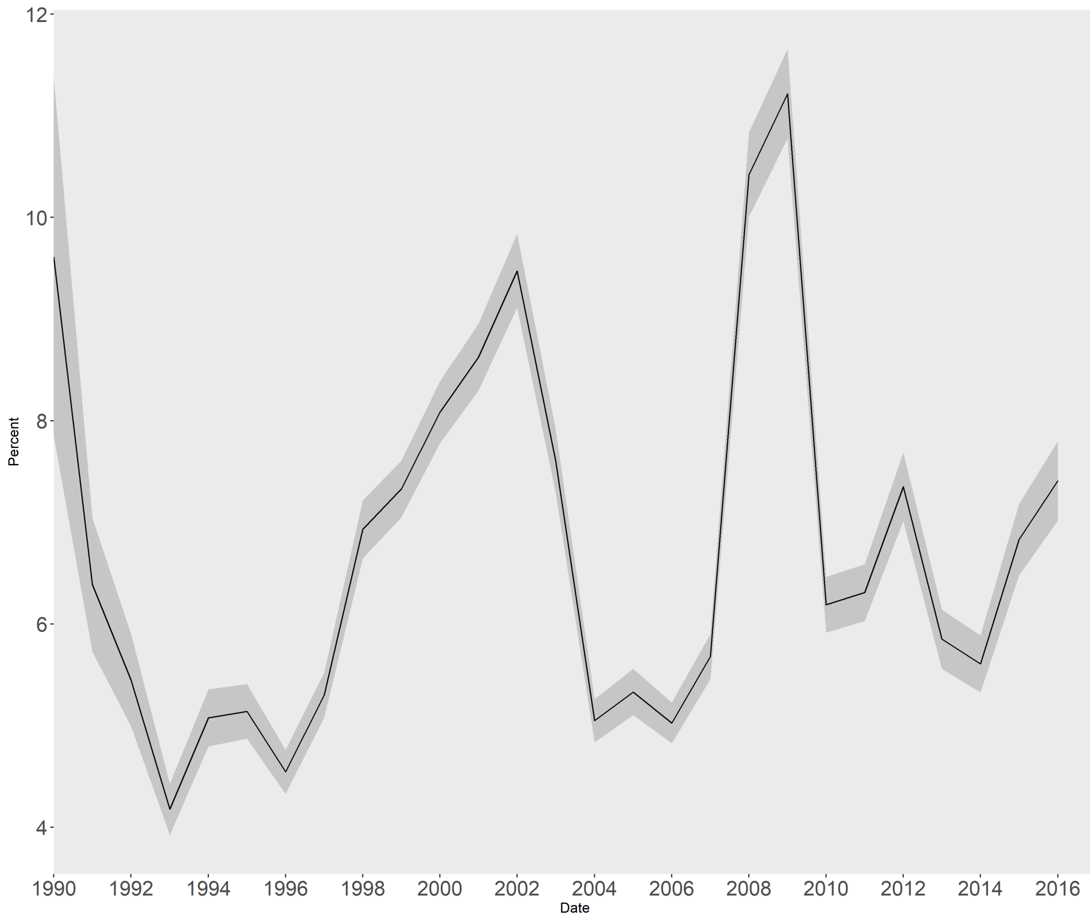



**Figure 5.1: Percentage Transitory Price - FS SSM**
This figure plots transitory price as a percentage of observed price, winsorized at the 1% level in each tail.

of mutual fund trading during FS quarters is correlated with information. The second row reports average and median values of $\beta$ that measures the association between actual sales and the noise term in the price equation ($s_t$). As shown, noise in stock prices is significantly associated with mutual fund imbalances during FS-quarters, as initially observed by Coval and Stafford [2007]. The variance of $s_t$ is a measure of transitory volatility, which we report in the row titled % Noise. We scale this variable by total variance in prices. Based on the average stock, approximately 28% of stock price variation is noise-related. The median value of noise is 16.3%, suggesting that measures of noise are right-skewed. The median percentage of noise in stock prices during our sample is remarkably close to the 10% lower bound obtained by Eckbo and Liu [1993] during the post-war period using different methods.

We next examine the proportion of information-related variance explained by the discretionary component of mutual fund sales.

$$\frac{\kappa^2 Var(\widetilde{Disc_t})}{\sigma_\eta^2 + \kappa^2 Var(\widetilde{Disc_t})} \tag{5.9}$$

On average, 21.7% of the variance in efficient value ($m_t$) is related to the discretionary component of mutual fund sales during FS quarters. We also compute a similar variance ratio for the relationship between actual imbalances and noise and show that almost 40% of noise variance during FS-quarters is attributable to mutual fund trading (% Act. contribution to noise). This result suggests that more than 10% of stock price variance, in fire sale quarters, is related to mutual fund trading. We conclude that mutual fund imbalances during FS-quarters are an important determinant of stock price movements that are associated with both information and noise induced by extreme flows.

We also examine changes in the proportion of noise in stock prices over time. Figure 5.1 plots the ratio of the absolute value of noise to (log) price across stocks in each year of the sample. As shown, the transitory component becomes more important during periods of higher uncertainty about stock market valuations. For example, the average noise-to-price ratio $\left(\frac{s_t}{p_t}\right)$ increased from less than 5% in 1996 to over 9% in 2001 before declining to 5% at the end of 2006. This ratio reached its peak of 11% during the 2007-2008 financial crisis before reverting back to an average of 7% during the post-crisis years.

Figure 5.2 plots average price and smoothed estimates of the permanent component of returns averaged across firms in each quarter of our sample. We report smoothed estimates of $m_t$ in order to incorporate past, current and future information in our estimate of efficient value.[6] As one might expect, the

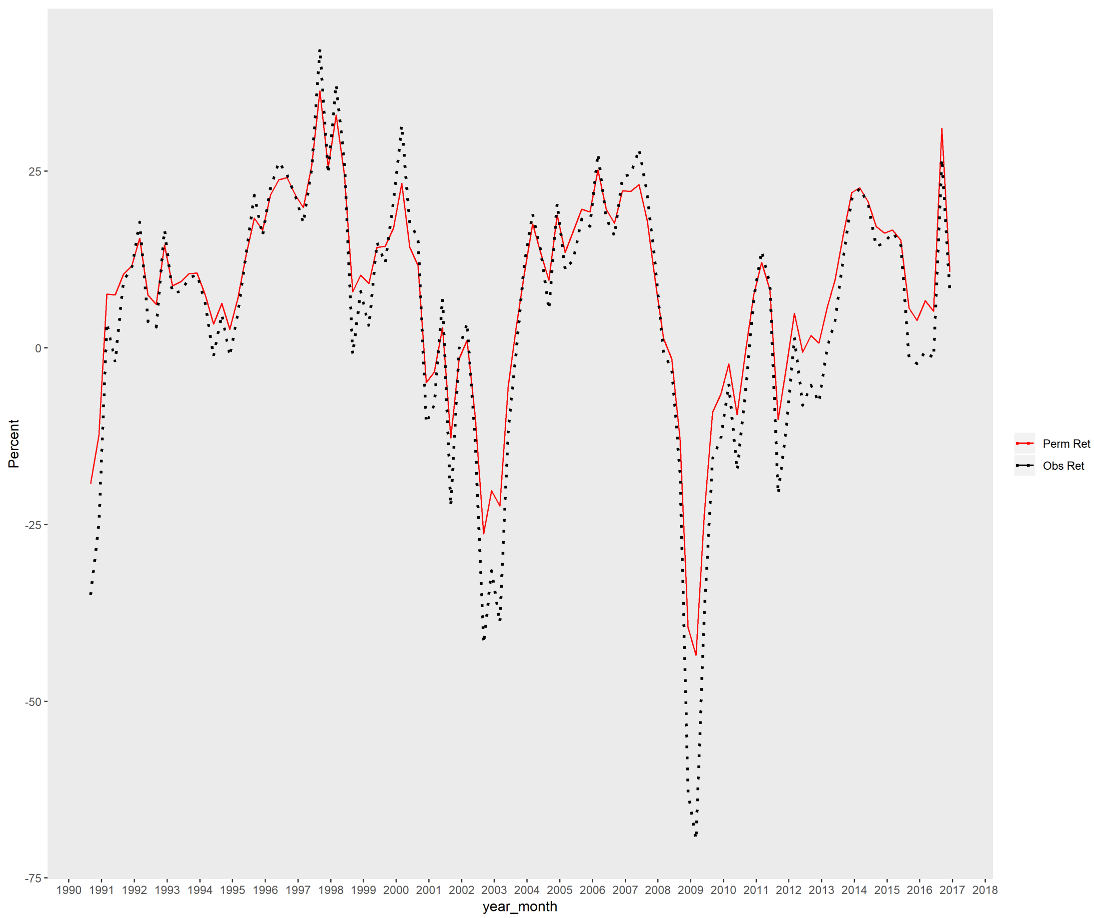



**Figure 5.2: Return decomposition - FS SSM**
This figure plots cumulative permanent returns versus observed returns.

[6]Smoothed estimates are obtained by taking the expectation of $m_t$ conditional on past, current and future information. See Durbin and Koopman [2012] for further details.

### 5.4.2 Validation test: Over- and under-selling during fire sale quarters

time series of the permanent component of returns is smoother than the price-based return series. We argue above that selling of stock by more than its proportional weight in mutual funds' portfolios sends a strong signal about efficient value. Accordingly, in this section, we explore the implications of over-selling by conducting an event study around FS-quarters characterized by either over- or under-selling.

If we have solved the signal extraction problem correctly, we expect both declines in price and efficient value during quarters in which there is over-selling. Likewise we expect to observe increases in efficient value during event quarters characterized by under-selling relative to what one would expect if mutual fund managers sold stocks proportionately to what they owned in the previous quarter. In both cases, the rationale is that over- and under-selling are associated with information about the permanent component of returns $m_t$.

We test this prediction by defining over-selling event quarters as quarters in which the following condition holds.

$$\sum_j (\Delta Shares_{i,j,t} - Flow_{j,t} \times Shares_{i,j,t-1}) < 0 \tag{5.10}$$

where the summation is over mutual funds $j$ such that $Flow_{j,t} < -5\%$. An analogous definition applies for under-selling quarters. We employ the smoothed estimates of $m_t$ and $s_t$, which condition on all information, thus incorporating

**Table 5.7: Event Study - FS Model: Overselling versus Underselling**
This table reports mean and t-statistics for observed returns, permanent price returns and noise price returns around Overselling (Panel A) and Underselling (Panel B) events. Event time t, denotes time of respective buying or selling. Overselling (Underselling) is defined as change in fire sale net holdings corresponding to positive (negative) discretionary residual component. Permanent returns and transitory returns are computed from the fire sales state space model 5.3. * $p < 0.01$; ** $p < 0.05$; *** $p < 0.001$

Panel A: Overselling

| Event Time | Obs. Ret.($\Delta p_t$) | | Perm. Ret($\Delta m_t$) | | Noise. Ret ($\Delta s_t$) | |
|---|---|---|---|---|---|---|
| | Mean | T-Stat | Mean | T-Stat | Mean | T-Stat |
| t-8 | 0.61*** | 5.32 | 0.57 | 6.80 | 0.04 | 0.49 |
| t-7 | 1.17*** | 10.66 | 0.88 | 10.95 | 0.28*** | 4.11 |
| t-6 | 1.13*** | 10.33 | 0.77 | 9.52 | 0.36*** | 5.33 |
| t-5 | −0.12 | −1.11 | 0.00 | 0.01 | −0.12* | −1.81 |
| t-4 | −0.04 | −0.39 | 0.03 | 0.37 | −0.07 | −1.06 |
| t-3 | 0.41*** | 3.85 | 0.21*** | 2.63 | 0.20*** | 3.06 |
| t-2 | −0.73*** | −6.89 | −0.38*** | −4.83 | −0.35*** | −5.23 |
| t-1 | −0.43*** | −4.10 | −0.25*** | −3.23 | −0.18*** | −2.71 |
| t | −1.63*** | −15.72 | −1.06*** | −13.85 | −0.57*** | −8.72 |
| t+1 | 0.11 | 1.11 | −0.08 | −1.11 | 0.19*** | 3.11 |
| t+2 | −0.68*** | −6.39 | −0.44*** | −5.58 | −0.24*** | −3.64 |
| t+3 | 0.30*** | 2.85 | 0.02 | 0.30 | 0.28*** | 4.24 |
| t+4 | 0.13 | 1.27 | −0.06 | −0.75 | 0.19*** | 2.96 |
| t+5 | −0.52*** | −4.82 | −0.38*** | −4.78 | −0.14*** | −2.05 |
| t+6 | −0.39*** | −3.59 | −0.40*** | −4.99 | 0.01 | 0.20 |
| t+7 | −0.24*** | −2.23 | −0.14** | −1.80 | −0.10 | −1.47 |
| t+8 | −0.45*** | −4.11 | −0.35*** | −4.28 | −0.10 | −1.51 |

Panel B: Underselling

| Time | Obs. Ret.($\Delta p_t$) | | Perm. Ret($\Delta m_t$) | | Noise. Ret ($\Delta s_t$) | |
|---|---|---|---|---|---|---|
| | Mean | T-Stat | Mean | T-Stat | Mean | T-Stat |
| t-8 | 0.96*** | 7.31 | 0.60*** | 6.13 | 0.36*** | 4.29 |
| t-7 | −0.71*** | −5.23 | −0.25*** | −2.44 | −0.46*** | −5.39 |
| t-6 | −0.37*** | −2.87 | 0.02 | 0.20 | −0.39*** | −4.63 |
| t-5 | 1.34*** | 10.25 | 0.95*** | 9.70 | 0.39*** | 4.62 |

Table Continued

| | | | | | | |
|---|---|---|---|---|---|---|
| t-4 | 0.57*** | 4.37 | 0.42*** | 4.30 | 0.15* | 1.77 |
| t-3 | 0.05 | 0.41 | 0.12 | 1.25 | −0.07 | −0.82 |
| t-2 | 0.69*** | 5.29 | 0.24*** | 2.52 | 0.44*** | 5.35 |
| t-1 | −1.80*** | −13.96 | −1.11*** | −11.52 | −0.69*** | −8.34 |
| t | 2.63*** | 21.01 | 1.44*** | 14.93 | 1.18*** | 14.86 |
| t+1 | −1.27*** | −9.62 | −0.59*** | −5.92 | −0.68*** | −8.26 |
| t+2 | −0.50*** | −3.96 | −0.46*** | −4.83 | −0.04 | −0.45 |
| t+3 | −1.31*** | −10.10 | −0.74*** | −7.64 | −0.56*** | −6.97 |
| t+4 | −1.25*** | −9.53 | −0.84*** | −8.45 | −0.41*** | −5.10 |
| t+5 | −0.84*** | −6.42 | −0.64*** | −6.44 | −0.21*** | −2.53 |
| t+6 | 0.32*** | 2.44 | 0.04 | 0.37 | 0.29*** | 3.51 |
| t+7 | −1.52*** | −11.45 | −1.12*** | −11.23 | −0.40*** | −4.81 |
| t+8 | 0.15 | 1.13 | −0.18*** | −1.80 | 0.33*** | 4.07 |

past, current and future information. As such, smoothed estimates of efficient value benefit from peering into the future.

Table 5.7 reports mean values of the log change in price ($\Delta p_t$), efficient value ($\Delta m_t$) and noise ($\Delta s_t$). The first column reports quarterly returns, and the second and third columns report the percentage changes in the efficient and transitory (noise) components of price. As expected, both price and the transitory component of price decline significantly during event quarters in which there is over-selling. More importantly, the permanent component $m_t$ declines by more than 1% during over-selling quarters. Consistent with the price impact regressions reported earlier, there is no subsequent reversal in either price or efficient value following the event quarter. Panel B reports results for under-selling quarters, with the direction of changes in price, efficient value and noise being positive in the event quarter. There is a slight reversal in price following under-selling quarters. Figures 5.3 and 5.4 provide a visual

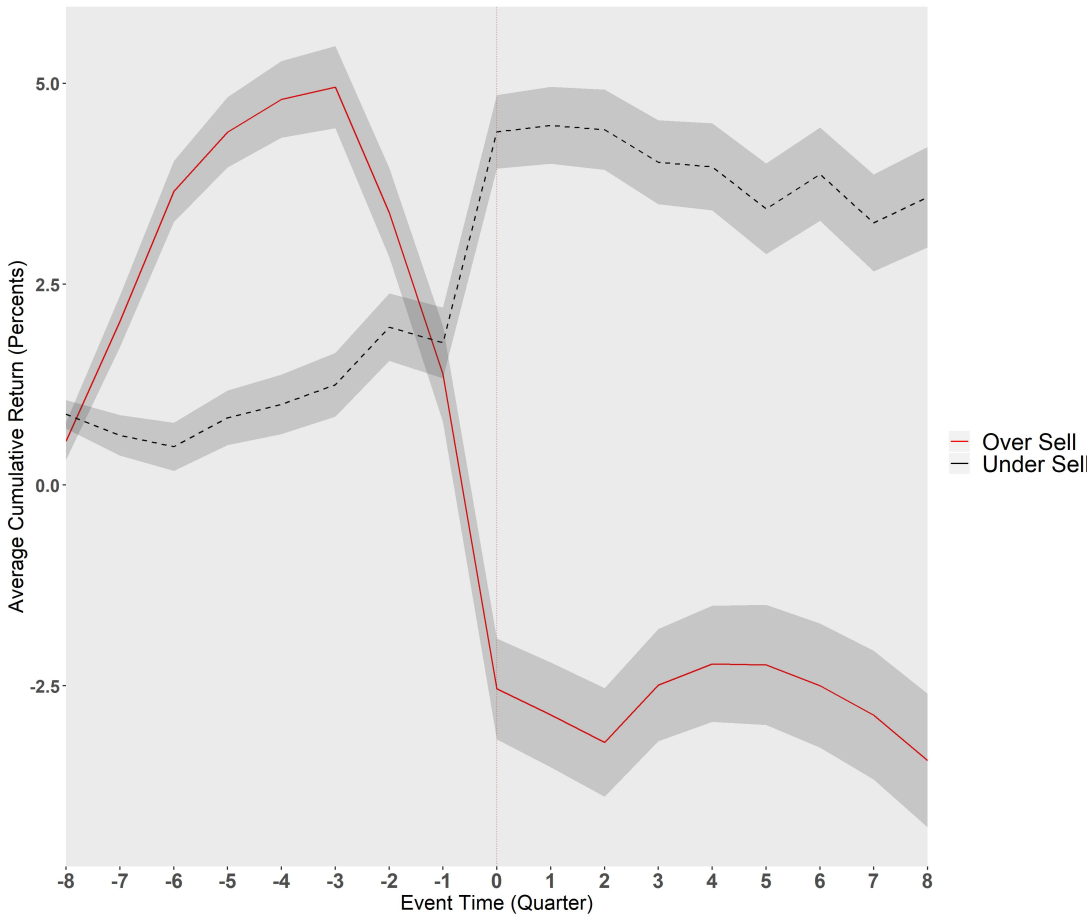



**Figure 5.3: Under-selling versus Over-selling and Observed Returns - FS SSM** This figure plots observed abnormal returns for over-selling versus under-selling in the fire sale sample. We compute abnormal returns as observed returns minus the CRSP value weighted return index. Over-selling (under-selling) of a stock occurs when the sum of funds' discretionary residual trading in a quarter is negative (positive).

representation of returns, efficient value and noise during and around over and under selling quarters. Figure 5.3 shows the evolution of price around over-selling and under-selling quarters. Reversals in price are of limited magnitude over quarters t+1, t+2, etc. Figure 5.4 shows the impact of information on the (smoothed) permanent component of prices following under and over selling quarters.

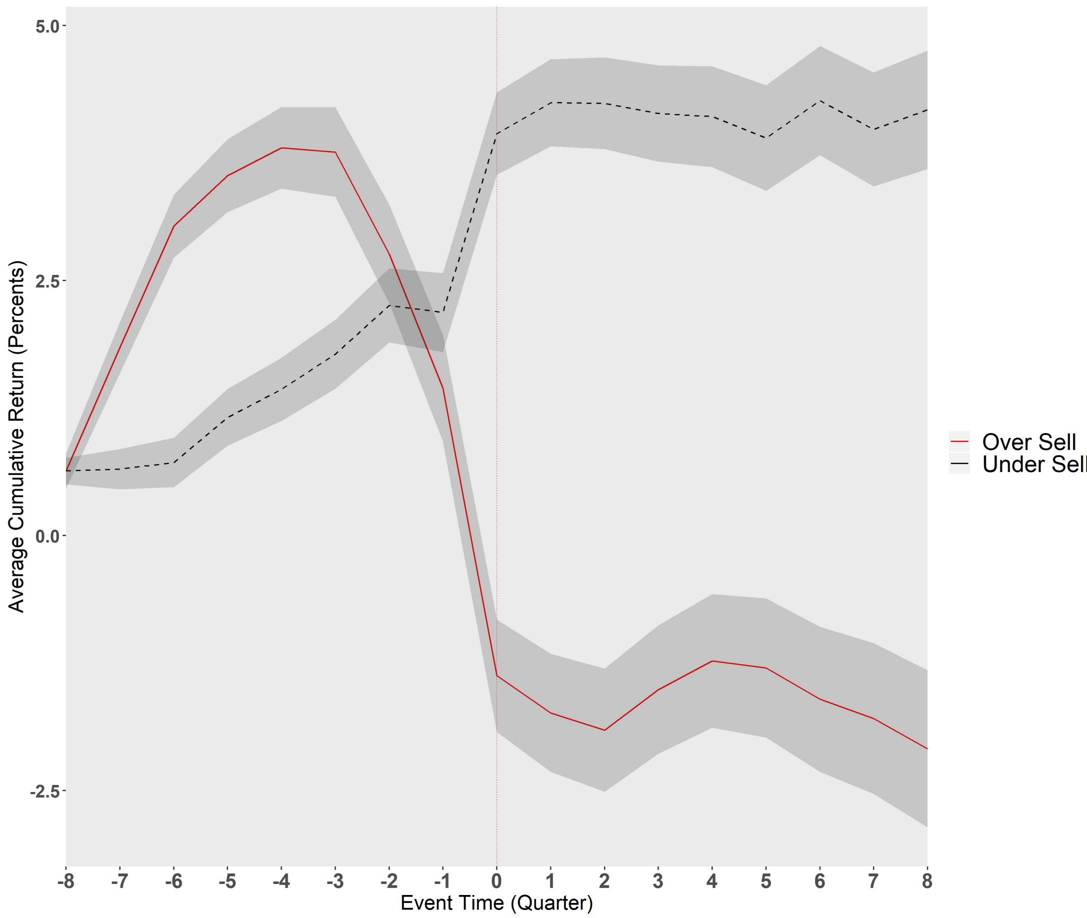



**Figure 5.4: Underselling versus Overselling Permanent Returns - FS SSM**
This figure plots permanent returns from the FS SSM in event time. Over-selling (under-selling) of a stock occurs when the sum of funds' discretionary residual trading in a quarter is negative (positive).

There is no reversal in the permanent component of price following under-selling and over-selling quarters.

### 5.4.3 State space model estimates based on all mutual fund-quarters

Consistent with the literature on fire sales-induced price pressure, we have used FS quarters to construct the two measures of mutual fund trading $Disc_t$ and $Act_t$ by setting the values of these two variables to zero in non FS-quarters. In this section we consider whether our estimates change when all mutual fund trading quarters (i.e., including fund-quarters with $Flow_{j,t} > -5\%$) are used to estimate the permanent and temporary components of returns. Broadening the set of mutual fund calendar quarters that are included in the sample is appealing because the additional information contained in mutual fund trading during

**Table 5.8: Main Results: Net State Space Model (SSM)**
This table presents estimation results from the state space model given in equation 5.3. The total number of converged stocks were 2,581 which represent 72% of our sample stocks. All reported statistics are cross-sectional means across stock estimates. t-statistics for the means are reported in parenthesis. $\kappa$ is the mean of estimated correlations between stock permanent price innovations and discretionary component of sales. $\beta$ is the mean of estimated correlation between stock transitory price innovation and actual sales and $\phi$ is the mean of estimated autoregressive component in transitory price. % Noise is the size weighted mean of transitory price contribution to total price variation; % Disc contribution is size weighted mean of discretionary sale contribution to permanent price variation; % Act contribution is size weighted mean of actual net purchases contribution to transitory price variation. 95% Confidence intervals for the means are presented in column 4 and percent of estimated stock $\kappa$ and $\beta$ which are positive and statistically significant at the 10% level are reported in column 5.

| Variable | Mean | Median | 95% CI | |
|---|---|---|---|---|
| Disc residual$_{\mathrm{t}}$ ($\kappa$) | 1.41 | 1.50 | (1.17 | 1.64) |
| | (11.62) | | | |
| Act$_{\mathrm{t}}$ ($\beta$) | 1.01 | 0.41 | (0.78 | 1.24) |
| | (8.72) | | | |
| AR(1) ($\phi$) | 0.30 | 0.50 | (0.32 | 0.28) |
| | (29.57) | | | |
| % Noise | 27.59 | 13.92 | (26.49 | 28.68) |
| | (33.47) | | | |
| % Disc. contribution to information | 16.41 | 6.08 | (15.45 | 17.37) |
| | (35.06) | | | |
| % Act. contribution to noise | 33.61 | 19.72 | (32.26 | 34.96) |
| | (48.75) | | | |

non FS-quarters can increase the precision of our estimates of the permanent component of prices. However, doing so may come at the cost of loss of power to detect trading-induced price pressures. The principal reason is that positive flows do not commit mutual funds to trading to satisfy redemption requests within one business day, as is the case with redemption requests in open-ended funds. This factor allows funds to break down their purchases of a stock over time in order to minimize the impact of their trades on price (see Simutin [2014] who argues these funds have higher risk-adjusted returns). On the other hand, the impact on prices of mutual-fund trading in response to flows may be muted in funds that are more likely to internalize the negative externality of fire sales. As argued by Chernenko and Sunderam [2020], these funds actively manage their cash holdings in order to minimize the impact of flows on prices. Adding non-fire sale mutual fund quarters and trading can incorporate additional information due to the trading by these funds.

We suspect that trading during non FS-quarters is more likely to be associated with information than during FS quarters. It follows that the variance ratio of noise to price (% noise) should fall when using all mutual fund trading quarters, as should the variance proportion of actual trades to total noise (% Act contribution).

We also expect the variance of discretionary trading to be lower in relation to the variance of the permanent component of returns (% Disc contribution). While this conjecture may seem counterintuitive as FS quarters generate more temporary price pressure than non FS quarters, the ability of mutual fund managers to reduce price impact by breaking up trading during non-FS quarters means that uninformed investors will have more difficulty distinguishing discretionary trading from non-discretionary trading to meet investor flows during non FS quarters. Put differently, discretionary trading $Disc_t$ has lower power to detect information-related trading during non FS quarters, which decreases the ratio of the variance of the discretionary trading ($Disc$) to the permanent component variance.

Table 5.8 reports SSM estimates using all mutual fund trading quarters for each stock in our sample. As shown, the noise proportion in price variance equals 27.6%, which is not significantly lower than during FS-quarter based estimates reported in Table 5.6. However, the proportion of discretionary trading contribution to permanent component variance is significantly lower. Moreover, the proportion of variance of actual trading to noise variance is also lower, as one might expect if non-FS quarters contain information-related trading. The net model also suggests prices are somewhat less noisy than in the FS model and that mutual fund trading contributes to less overall noise (11% versus only 9%).

### 5.4.4 Application: Investment-noise sensitivity

We next consider an application of our price decomposition in explaining corporate investment. A firm's stock price may inform managerial investment decisions if managers and investors have different sets of information, as stock prices may reflect information about investment opportunities that the manager would not have otherwise known (i.e., the active informant hypothesis). If this is the case and managers use prices to make real decisions, noisy prices have the potential to distort corporate investment policy when managers incorrectly attribute noise-based changes in price to information.[7]

Evidence on this question is mixed, with some studies showing a significant

[7] Morck, Shleifer, and Vishny [1990] describe several different hypotheses relating stock prices to corporate investment.

relation between stock prices and investment (Chen et al. [2007]), and others showing no relation (Bakke and Whited [2010]). Answering this question is also challenging because investment and stock prices may be related through information that is known by investors and the firm's managers (i.e. the passive informant hypothesis).

We estimate investment-price sensitivity by decomposing log prices into their permanent and transitory components through the SSM. Finding a significant relation between the transitory component (i.e., noise) and investment would be inconsistent with the passive informant hypothesis because the noise term ($s_t$) is, by construction, impermanent and unrelated to information (see equation 5.3).

Because the permanent component ($m_t$) is non-stationary, we normalize it by subtracting the log of the book value of equity per share ($CEQ$). This variable is equivalent the log of the efficient (or efficient value) equity price-to-book ratio. This normalization also has the benefit of scaling log prices, which may differ by large magnitudes across stocks (e.g., Berkshire Hathaway's A-class shares are priced above $100,000, much higher than the average price of other stocks). Since we cannot perform an equivalent scaling for noise, we also standardize all variables by subtracting the time-series mean and dividing by the within-firm standard deviation.[8] Both efficient value and noise estimates are based on the Kalman smoother, meaning that we incorporate past, present and future information into estimates of $m_t$ and $s_t$.

In order to measure the sensitivity of investment to noise and efficient value, we estimate the following regression:

[8] We standardize using firm-specific mean and standard deviation instead of the pooled mean and standard deviation, since cross-sectional differences in non-standardized noise are related to difference in the dollar magnitudes of stock prices across firms.

**Table 5.9: Investment Sensitivity**
This table reports OLS estimates of investment sensitivity panel regression specified in equation 5.11. Our sample selection is outlined in section 5.3 and the sample time period is 1990-2016. The dependent variable is investment of firm i defined as the quarterly capital expenditures relative to lag property plant and equipment; in units of percent. In column 1 we use the Fire Sales State Space Model (FS SSM) price estimates and in column 2 we use the Net Imbalance State Space Model (Net SSM) price estimates. SSM efficient firm value to replacement cost (SSM Eff.) is defined as the SSM efficient price estimate ($\hat{m}_t$) net of natural logarithmic of the per share value of common equity. We include firm's debt to asset value as control for firm's value of debt. SSM Noise is the SSM estimate of transitory price ($\hat{s}_t$). In column 3 we replicate table 2 of Dessaint et al. (2019) using our sample stocks and sample time period. MFHS is unadjusted for returns as in Dessaint et al. (2019). Tobin Q residuals is a proxy for efficient firm value to replacement cost and is estimated as residuals from panel regression of Tobin's Q on MFHS. CF/Assets is cash flow divided by assets. Size is the natural logarithmic of assets. Peer variable are computed as equally weighted peer averages. Peers are identified using similarity scores from Hoberg and Phillips (2015). All regressions include firm fixed and time fixed effects at the quarter level. All variables are winsorized at the 1% level in each tail and deflated to 2012 dollars. All explanatory variables are normalized by their stock standard deviation to facilitate economic interpretation and are winsorized at the 1% level in each tail. Reported t-statistics are computed using Heteroscedasticity-Corrected Covariance Matrix (HCCME 1) with clustering at the firm and time levels. $^{*}$ $p < 0.01$; $^{**}$ $p < 0.05$; $^{***}$ $p < 0.001$.

Dependent Variable: Investment ($\text{CAPX}_t/\text{PPE}_{t-1}$)

Panel A: Own Sensitivity

| | **FS SSM** | **Net SSM** | **Dessaint et al. (2019)** |
|---|---|---|---|
| SSM Noise ($\hat{s}_{i,t-1}$) | $0.33^{***}$ | $0.38^{***}$ | |
| | (11.82) | (13.23) | |
| SSM Eff. $\left(\hat{m}_{i,t-1} - \log\left(\frac{CEQ_{i,t-1}}{\text{shares}_{i,t-1}}\right)\right)$ | $0.81^{***}$ | $0.78^{***}$ | |
| | (20.07) | (19.46) | |
| $\text{MFHS}_{\text{i,t-1}}$ | | | $0.20^{***}$ |
| | | | (6.22) |
| Tobin Q residuals ($\tilde{Q}_{i,t-1}$) | | | $1.22^{***}$ |
| | | | (27.34) |
| $\text{CF/Assets}_{\text{i,t-1}}$ | $0.28^{***}$ | $0.28^{***}$ | $0.28^{***}$ |
| | (8.05) | (8.19) | (8.15) |
| $\text{Size}_{\text{i,t-1}}$ | 0.07 | 0.06 | 0.00 |
| | (1.41) | (1.30) | (0.07) |

Table Continued

| | | | |
|---|---|---|---|
| Debt/Assets$_{\text{i,t-1}}$ | $-0.68^{***}$ | $-0.68^{***}$ | |
| | $(-18.25)$ | $(-18.25)$ | |
| # Stocks | 2,341 | 2,341 | 2,341 |
| # Obs | 107,204 | 107,204 | 107,470 |
| $R^2$ | 0.30 | 0.28 | 0.28 |

Panel B: Own and Peer Sensitivity

| | **FS SSM** | **Net SSM** | **Dessaint et al. (2019)** |
|---|---|---|---|
| SSM Noise ($\hat{s}_{i,t-1}$) | $0.33^{***}$ | $0.39^{***}$ | |
| | (9.83) | (11.55) | |
| SSM Eff. $\left(\hat{m}_{i,t-1} - \log\left(\frac{CEQ_{i,t-1}}{\text{shares}_{i,t-1}}\right)\right)$ | $0.75^{***}$ | $0.74^{***}$ | |
| | (15.30) | (14.59) | |
| MFHS$_{\text{i,t-1}}$ | | | $0.10^{**}$ |
| | | | (2.39) |
| Tobin Q residuals ($\tilde{Q}_{i,t-1}$) | | | $1.19^{***}$ |
| | | | (21.33) |
| CF/Assets$_{\text{i,t-1}}$ | $0.28^{***}$ | $0.30^{***}$ | $0.28^{***}$ |
| | (6.75) | (7.52) | (7.04) |
| Size$_{\text{i,t-1}}$ | $0.13^{**}$ | 0.02 | 0.07 |
| | (2.43) | (0.45) | (1.42) |
| Debt/Assets$_{\text{i,t-1}}$ | $-0.66^{**}$ | $-0.60^{***}$ | |
| | $(-15.20)$ | $(-13.22)$ | |
| Peer SSM Noise ($\hat{s}_{-i,t-1}$) | $0.12^{***}$ | $0.13^{***}$ | |
| | (3.25) | (3.44) | |
| Peer SSM Eff. $\left(\hat{m}_{-i,t-1} - \log\left(\frac{CEQ_{-i,t-1}}{\text{shares}_{-i,t-1}}\right)\right)$ | $0.13^{***}$ | $0.10^{**}$ | |
| | (3.22) | (2.09) | |
| Peer MFHS$_{\text{-i,t-1}}$ | | | $0.08^{*}$ |
| | | | (1.78) |
| Peer Tobin Q residuals ($\tilde{Q}_{-i,t-1}$) | | | 0.06 |
| | | | (1.46) |
| Peer CF/Assets$_{\text{-i,t-1}}$ | 0.02 | 0.03 | $0.06^{*}$ |
| | (0.60) | (0.90) | (1.77) |
| Peer Size$_{\text{-i,t-1}}$ | $-0.01$ | $-0.00$ | 0.02 |
| | $(-0.39)$ | $(-0.02)$ | (0.53) |

Table 8 Continued

| | | | |
|---|---|---|---|
| Peer Inv. ($\frac{CAPX_{-i,t-1}}{PPE_{-i,t-2}}$) | $0.27^{***}$ | $0.28^{***}$ | $0.28^{***}$ |
| | (7.60) | (7.85) | (7.68) |
| Peer Debt/Assets$_{\text{-i,t-1}}$ | $-0.09^{**}$ | $-0.09^{**}$ | |
| | $(-2.30)$ | $(-2.18)$ | |
| # Stocks | 2,063 | 2,063 | 2,063 |
| # Obs | 74,873 | 74,873 | 74,873 |
| $R^2$ | 0.30 | 0.29 | 0.29 |

$$\begin{aligned}\frac{CAPX_{i,t}}{PPENT_{i,t-1}} &= \alpha_i + \beta_1 \hat{s}_{i,t-1} + \beta_2 \left( \hat{m}_{i,t-1} - \log \left( \frac{CEQ_{i,t-1}}{\text{shares}_{i,t-1}} \right) \right) \\ &+ \beta_3 \frac{CF}{Assets_{i,t-1}} + \beta_4 Size_{i,t-1} + \beta_5 \frac{Debt}{Assets_{i,t-1}} + \delta_t + \epsilon_{i,t}\end{aligned} \tag{5.11}$$

where $\alpha_i$ is a firm fixed effect and $\delta_t$ is a fiscal quarter fixed effect. The dependent variable, investment, is measured as the ratio of capital expenditures to (lagged) net property, plant and equipment. Control variables include the ratio of cash-flow to assets ($CF/Assets$), the log of total assets ($Size$), and financial leverage measured as the ratio of total book value of debt to total assets. As described above, we measure all independent variables in units of within-firm standard deviation to facilitate economic interpretation.

Table 5.9 reports results of this test. As shown in Panel A, investment is positively and significantly related to both components of stock prices ($m_t$ and $s_t$). The estimates suggest that a one standard deviation increase in the firm's noise increases quarterly investment (as a fraction of $PP\&E$) by 0.33% (1.31% annually). In comparison, a one standard deviation increase in the log ratio of efficient price to book value $\left(\hat{m}_{i,t-1} - \log \frac{CEQ_{i,t-1}}{\text{shares}_{i,t-1}}\right)$ increases quarterly investment by 0.81% (median quarterly investment scaled by $PP\&E$ is 4.9%). We obtain slightly greater investment-noise sensitivities when including all mutual fund-quarters beyond fire sale quarters (the net sales model described in Section 5.4.3). As shown in column (2), a one standard deviation in noise increases quarterly investment by 0.38% of PP&E.

For purposes of comparison with previous literature, we report a similar regression in which noise is proxied by mutual fund hypothetical sales (MFHS), normalized so that it is in units of standard deviation. This variable, used in Dessaint et al. [2019] and Lou and Wang [2018] to proxy for noise in stock prices, is constructed in similar fashion to $Act_t$ except that shares are multiplied by time t-1 share prices and dollar volume over quarter t is used in the denominator. MFHS is defined as follows.

$$MFHS_{i,t} = \sum_j (FLOW_{j,t} \times Shares_{i,j,t-1}) \times Price_{i,t-1}/(Price_{i,t} \times Volume_{i,t}). \tag{5.12}$$

where the summation is over mutual funds j experiencing quarterly flows less than -5%. As shown, investment is also sensitive to this measure of non-efficient price pressure, but the economic significance is smaller with one standard deviation increasing investment by 0.20% of PP&E.

The sensitivity of investment to noise reported in Panel A is consistent with two interpretations. Stock prices may aggregate information that would otherwise not be available to the manager (the active informant hypothesis), with managers learning from stock prices and (incorrectly) inferring information from the transitory component of stock price returns. Alternatively, investment may be sensitive to price because access to equity financing required for investment is dependent on favorable stock prices (the financing hypothesis).[9]

One solution to disentangling these two effects is to measure the sensitivity of investment to noise in peer prices. Including peer stock prices controlling for the firm's own stock price allows a clean test of the active informant hypothesis as financing cost effects are captured by the correlation between investment and the firm's own stock price. Foucault and Fresard [2014] and Dessaint et al. [2019] use this strategy and estimate the sensitivity of investment to peer stock prices and peer noise, respectively. These studies find a significant relation between investment, peer stock prices and peer noise. We adopt a similar testing framework except that we measure peer noise as the average of the transitory component $s_t$ across the firm's peers. We identify the set of company peers using the similarity measure developed in Hoberg and Phillips [2016] based on a textual analysis of company 10-K forms. The similarity measure compares firms' product description and we group firms based on this measure. The peer noise and peer efficient value consist of an equally weighted average of each measure across firms in each industry grouping.[10]

Panel B reports estimation results with peer effects. As shown, firm investment is sensitive to noise in peer stock prices. As one might expect, the economic magnitude is smaller than the effect of the firm's own noise, with a one standard deviation increase in peer noise increasing investment by 0.12%, compared with

[9] For evidence on the financing hypothesis, see Baker, Stein, and Wurgler [2003].

[10] We obtain similar results using weighted averages with weights based on product similarity scores.

0.33% for the firm's own noise in stock price. We obtain similar results using all mutual fund-quarter sales instead of only sales in FS-quarters to identify noise (column 2).

Interestingly, firm managers are equally sensitive to changes in the permanent and transitory components of peer prices. The coefficients on peer noise and peer efficient price are both approximately 0.12-0.13 and statistically indistinguishable. This finding suggests that managers cannot distinguish between noise and information in peer prices, yet use these prices (along with their firm's prices) to inform their investment decisions. The last column replicates the Dessaint et al. [2019] regression with peer effects using $PeerMFHS$, equal to the average of $MFHS$ over peer firms, as a measure of noise in stock prices. Consistent with their results, this measure of noise in peer stock prices is (weakly significantly) related to investment.

### 5.4.5 Public or Private Information

Another solution to disentangling whether or not managers possess private information about efficient value is to decompose information into a component knowable at time t (public information) and information revealed between time t and time $\tau > t$ (private information). To do this we exploit the Kalman filter that provides an estimate of efficient value ($m_t$) and the difference between this estimate and the estimate generated by the Kalman smoother that includes information up until time $\tau$.

We estimate a similar investment-sensitivity model as before, except that we split efficient value and noise into public (filtered) and private (unanticipated)

**Table 5.10: Investment Sensitivity - Filtered versus Smooth estimates**
This table reports OLS estimates of investment sensitivity panel regression specified in equation 5.11. Our sample selection is outlined in section 5.3 and the sample time period is 1990-2016. The dependent variable is investment of firm i defined as the quarterly capital expenditures relative to lag property plant and equipment; in units of percent. In column 1 we use the Fire Sales State Space Model (FS SSM) price estimates and in column 2 we use the Net Imbalance State Space Model (Net SSM) price estimates. Filtered estimates are conditional permanent and noise price estimates based on information until time t. Smooth estimates are conditional permanent and noise price estimates based on all available; i.e. past, present, and future. Filtered firm value to replacement cost (Filtered SSM Eff.) is defined as the SSM filtered efficient price estimate ($\hat{m}_t^F$) net of natural logarithmic of the per share value of common equity. Smooth firm value to replacement cost (SSM Eff.) is defined as the SSM smooth efficient price estimate ($\hat{m}_t$) net of natural logarithmic of the per share value of common equity. Unanticipated SSM Eff. (SSM Eff - Filtered SSM Eff) is the difference between firm value to replacement cost computed using filtered permanent price estimate and smooth price estimate. SSM filtered noise is the SSM estimate of filtered transitory price ($\hat{s}_t^F$). SSM Noise is the SSM estimate of smooth transitory price ($\hat{s}_t$). Unanticipated SSM Noise ($\hat{s}_t - \hat{s}_t^F$) is the difference between the filtered and smooth noise price estimate. CF/Assets is cash flow divided by assets. Size is the natural logarithmic of assets. We include firm's debt to asset value as control for firm's value of debt. All regressions include firm fixed and time fixed effects at the quarter level. All variables are winsorized at the 1% level in each tail and deflated to 2012 dollars. All explanatory variables are normalized by their stock standard deviation to facilitate economic interpretation and are winsorized at the 1% level in each tail. Reported t-statistics are computed using Heteroscedasticity-Corrected Covariance Matrix (HCCME 1) with clustering at the firm and time levels. $^*$ $p < 0.01$; $^{**}$ $p < 0.05$; $^{***}$ $p < 0.001$.

Dependent Variable: Investment ($\text{CAPX}_t/\text{PPE}_{t-1}$)

| | **FS SSM** | **Net SSM** |
|---|---|---|
| Filtered SSM Noise ($\hat{s}_{i,t-1}^F$) | 0.30$^{***}$ | 0.36$^{***}$ |
| | (10.53) | (12.46) |
| Unanticipated SSM Noise ($\hat{s}_{i,t-1} - \hat{s}_{i,t-1}^F$) | 0.15$^{***}$ | 0.14$^{***}$ |
| | (5.85) | (5.16) |
| Filtered SSM Eff. $\left(\hat{m}_{i,t-1}^F - \log\left(\frac{CEQ_{i,t-1}}{\text{shares}_{i,t-1}}\right)\right)$ | 0.82$^{***}$ | 0.78$^{***}$ |
| | (19.95) | (19.07) |
| Unanticipated SSM Eff. (SSM Eff - Filtered SSM Eff.) | 0.02 | 0.01 |
| | (0.99) | (0.56) |
| $\text{CF/Assets}_{\text{i,t-1}}$ | 0.26$^{***}$ | 0.27$^{***}$ |
| | (7.70) | |

Table Continued

| | | |
|---|---|---|
| (7.83) | | |
| $Size_{i,t\text{-}1}$ | 0.05 | 0.04 |
| | (0.97) | (0.88) |
| $Debt/Assets_{i,t\text{-}1}$ | $-0.67^{***}$ | $-0.67^{***}$ |
| | $(-18.16)$ | $(-18.14)$ |
| # Stocks | 2,341 | 2,341 |
| # Obs | 107,204 | 107,204 |
| $R^2$ | 0.28 | 0.28 |

components. A positive coefficient on the unanticipated component of noise suggests that managers make investment decisions using a component of price that is revealed as noise in the future. A positive coefficient on the unanticipated component of efficient value would suggest that managers make investment decisions using private information about efficient value. We report the coefficients in Table 5.10.

We find evidence that managers' investment decisions are correlated with unanticipated noise, suggesting that they do not possess private information. The coefficient on unanticipated noise is 0.15 (0.14) and statistically significant for the fire sales model (net model). Managers may assume that they are making decisions based on information or they may realize that they are making decisions based on noise that are advantageous. For instance, they may be exploiting temporary increases in the transitory component of prices to acquire inexpensive funding and invest in projects that nay otherwise not be funded.

Our results also suggest that managers do not possess or act upon information about future efficient value, not already included in prices at time $t$. The coefficients on unanticipated information are close to zero and statistically insignificant.

**Table 5.11: SSM Cross Sectional Regression** This table reports OLS results from a cross-sectional regression of state space model (SSM) estimates on firm fundamentals. In regression (1) and (2) the dependent variable is the estimated SSM coefficient $\kappa$, on discretionary sales (panel A) and discretionary net purchases (panel B). In regression (3) and (4) the dependent variables is percentage estimated noise from the fire sales model (panel A) and net imbalance model (panel B). Log Size is the natural logarithmic of average assets, PPENT/Assets is the average of the ratio of property plant and equipment relative to assets (asset tangibility), cash flow volatility to average cash flow (CF Vol/Avg. CF) is the standard deviation of cash flow to average cash flow, CF to Assets is the average of the ratio of cash flow to assets, Investment is the ratio of capital expenditures to lagged property plant and equipment, R&D to assets (RD Exp. to Assets) is the average of the ratio of R&D expenditures to assets and sales growth is the average of quarterly percent change in net sales. Averages correspond to time-series averages taken firm-by-firm. All variables are winsorized at the 1% level in each tail and deflated to 2012 dollar values. All ratio are expressed in units of percent. * $p < 0.01$; ** $p < 0.05$; *** $p < 0.001$

Panel A : Fire Sales

| Dependent Variable: | Disc. ($\kappa$) | | % Noise | |
|---|---|---|---|---|
| | Reg. 1 | Reg. 2 | Reg. 3 | Reg. 4 |
| Intercept | $-2.29^{***}$ | $-2.35^{***}$ | $40.67^{***}$ | $37.50^{***}$ |
| | $(-4.00)$ | $(-2.91)$ | $(18.53)$ | $(11.90)$ |
| Log Size | $0.60^{***}$ | $0.54^{***}$ | $-1.12^{***}$ | $-1.35^{***}$ |
| | $(6.73)$ | $(5.34)$ | $(-3.30)$ | $(-3.44)$ |
| Sales Growth | | 0.05 | | $0.48^{**}$ |
| | | (0.88) | | (2.27) |
| PPENT/Assets | | $-3.13^{***}$ | | $-2.07$ |
| | | $(-3.71)$ | | $(-0.63)$ |
| CF/Assets | | $0.28^{***}$ | | $1.10^{***}$ |
| | | (5.01) | | (5.10) |
| CF Vol/Avg. CF | | 0.05 | | 0.03 |
| | | (1.17) | | (0.21) |
| Investment | | $0.08^{**}$ | | $0.24^{*}$ |
| | | (2.14) | | (1.72) |
| RD Exp. to Assets | | $0.17^{**}$ | | $0.97^{***}$ |
| | | (1.98) | | (2.91) |

Table Continued

| | | | | |
|---|---|---|---|---|
| Missing RD Exp Indicator | | 0.10 | | −2.00 |
| | | (0.24) | | (−1.28) |
| $R^2$ | 0.02 | 0.04 | 0.00 | 0.02 |
| #Stocks | 2,581.00 | 2,405.00 | 2,581.00 | 2,495.00 |

Panel B : Net Imbalance

| Dependent Variable: | Disc. ($\kappa$) | | % Noise | |
|---|---|---|---|---|
| | Reg. 1 | Reg. 2 | Reg. 3 | Reg. 4 |
| Intercept | $-3.18^{***}$ | $-2.22^{**}$ | $42.12^{***}$ | $37.42^{***}$ |
| | (−4.49) | (−2.24) | (19.38) | (11.98) |
| Log Size | $0.70^{***}$ | $0.55^{***}$ | $-1.20^{***}$ | $-1.15^{***}$ |
| | (6.43) | (4.45) | (−3.56) | (−2.97) |
| Sales Growth | | 0.02 | | $0.61^{***}$ |
| | | (0.38) | | (2.96) |
| PPENT/Assets | | $-3.62^{***}$ | | −1.72 |
| | | (−3.49) | | (−0.53) |
| CF/Assets | | $0.43^{***}$ | | $0.82^{***}$ |
| | | (6.22) | | (3.79) |
| CF Vol/Avg. CF | | 0.04 | | 0.13 |
| | | (0.77) | | (0.85) |
| Investment | | 0.04 | | 0.20 |
| | | (0.82) | | (1.46) |
| RD Exp. to Assets | | $0.26^{**}$ | | $0.99^{***}$ |
| | | (2.48) | | (3.00) |
| Missing RD Exp Indicator | | $-0.65^{**}$ | | −1.32 |
| | | (−1.30) | | (−0.85) |
| $R^2$ | 0.02 | 0.04 | 0.00 | 0.02 |
| #Stocks | 2,581.00 | 2,495.00 | 2,581.00 | 2,495.00 |

The overall evidence suggests that firm-managers do not possess private information about efficient value or noise.

### 5.4.6 Cross-sectional tests

Stocks issued by firms can have characteristics that make them more likely to be associated with information-related trading than trading deriving from liquidity demand. For example, the market liquidity of larger firms makes these firms' stock more likely to be selected for sale by managers who have (negative) information about efficient value. Large firms are less likely to suffer from a fire sale discount during a FS-quarter. Stock prices of larger firms may also be more informationally efficient in the sense that the variance of $m_t$ conditional on price and information is low (Brunnermeier, Markus, and Brunnermeier [2001]). Begenau, Farboodi, and Veldkamp [2018] provide supporting evidence for this assertion. They argue that because large firms disclose more information for the market to process, these firms have higher price-informativeness and lower costs of capital than smaller firms.

Another feature determining the association between information and trading is the discretionary component of total firm investment, measured using research and development expenses (R&D) and investment through capital expenditures (Investment). Changes in R&D in firms with a large proportion of R&D to total investment are likely to be associated with information about efficient value (Phillips and Zhdanov [2013]). For similar reasons, we also expect these features i.e., size, CAPX and R&D) to be associated with the proportion of stock price variance that is attributable to noise.

Table 5.11 reports cross-sectional regressions of $\kappa$ and the variance proportion of noise on firm size (in columns 2 and 4), and firm size, investment, R&D

and other measures of firm fundamentals (in columns 3 and 5). Higher correlations between $\kappa$ and firm fundamentals indicates that stock trading in firms with these features is more likely to be associated with information than with liquidity demand. The coefficient on firm size is positive and significant across all four specifications suggesting that it is robust to including additional cross-sectional variables. As shown, with the exception of profitability (CF/Assets), the association between $\kappa$ and firm fundamentals are in the direction one might expect. Firms with more tangible assets (PPENT/Assets) have lower $\kappa$'s, while larger firms and firms with higher R&D relative to assets have higher $\kappa$'s. Firms with greater cash flows from operations typically have lower information asymmetry between insiders and outsiders, which is predictive of a negative correlation between $\kappa$ and operating cash flows. We find the opposite result, suggesting that trading in high-CF firms is more likely to be information-driven than liquidity-demand driven.

The last column reports associations between firm efficients and the proportion of price variance attributable to noise. Consistent with intuition, the variance proportion declines with size and increases with R&D and investment. Firms with greater average sales growth (a non-price based measure of investment opportunities) have a greater proportion of noise variance, as do firms with greater operating cash flows. This last result, while unexpected, is consistent with the positive relation between cash flows and information-related trading observed in Reg. (2).

## 5.5 Conclusion

We propose a parsimonious model that decomposes stock prices into a transitory (noise) and permanent component (efficient value). The model solves a signal extraction problem in which uninformed investors update their conditional expectations about efficient value based on observed prices and mutual fund trading. We use the model to measure the information content of mutual fund trading during fire sale quarters, defined as stock-quarters in which one or more mutual funds holding the stock experiences outflows of 5% of AUM or more. Our price decomposition indicates that 28% of the variance in stock prices is attributable to noise.

We apply our decomposition to measure the sensitivity of corporate investment to noise and efficient value in the firm's own and peer stock prices. We find that firms respond significantly to noise in their own stock prices, as well as their peers' stock prices, but to a lesser extent. Our results show a stronger reaction to noise in own stock prices than documented in other studies, consistent with theories of learning from stock prices. We also apply our methodology to a cross-asset learning setting and confirm previous results. In a novel test we align manager decision making with public information in prices at decision time. We show that they do not possess private information about future changes in efficient value. In fact, we show that their investment decisions are positively correlated with future noise.

The price decomposition we propose can be applied to other corporate policies such as executive compensation. A fruitful avenue for future research is to apply the price decomposition to the terms of executive compensation contracts..

Compensation for noise may have implications for the design of certain aspects of executive compensation contracts, such as “clawback” clauses in these contracts designed to limit excessive pay following stock-price reversals. Our methodology can also be applied in an asset pricing context. For instance, using our estimates of the noise component of prices, researchers can directly test the relationship between noise, or the component of noise related to trading (illiquidity), and expected returns.

## Appendix A: Identifying assumptions for estimation of structural model of stock prices

### 5.6 Appendix - Inventory Control

The state space model representation is based on Ho and Stoll [1981] and Hendershott and Menkveld [2014] who characterize the intermediary's problem as a stochastic optimal linear regulator problem. The intermediary holds inventory and supplies liquidity to investors. We assume that the intermediary can either be long or short inventory and prefers a net zero position. As he is risk averse, he will bid up prices in response to sell orders, which increase his inventory, and mark down prices in response to buy orders, which decrease his inventory. We assume a dealer market in which all market orders pass through a dealer (i.e. there is no limit-order book).

The solution to the stochastic control problem yields the following structural model of prices.

$$\begin{aligned} p_t &= m_t + s_t \\ m_t &= m_{t-1} + \kappa(\Delta I_t - E_{t-1}[\Delta I_t]) + \eta_t \\ s_t &= \phi s_{t-1} + \beta \Delta I_t + \epsilon_t \end{aligned} \tag{5.13}$$

where $\Delta I_t$ equals the dealer's inventory imbalance, i.e, buy orders minus sell orders. The error terms $\eta_t$ and $\epsilon_t$ are assumed to be independent, but we relax this assumption below. Here, $\Delta I_t - E_{t-1}[\Delta I_t]$ represents the information conveyed by the unexpected component of trade and $\eta_t$ represents non-trade news that arrives between time t and t-1. The unexpected component conveys information while also increasing or deceasing the dealer's inventory. The variances of $\eta_t$ and $\epsilon_t$ can be interpreted as the impact of information and market frictions on security prices. Changes attributed to information have a permanent effect on prices while changes attributed to pricing errors have a temporary effect on prices.

One of the main features of such a model is that changes in inventory convey information and cause prices to deviate from fundamental value. This can be seen from the fact that the transitory component $s_t$ is correlated with the trading-related component of information $(\Delta I_t - E_{t-1}[\Delta I_t])$ because $\Delta I_t$ is common to both. Order arrivals convey information and cause prices to deviate from true values (George and Hwang 2001).

### 5.6.1 Assumptions required for estimation

The following discusses how to estimate such a model when only a portion of changes in dealer inventory $I_t$ are observed. Based on the assumption that all trades pass through a dealer (no limit-order book assumption), the dealer inventory follows the following transition equation:

$$I_t = I_{t-1} - q_{s,t} + q_{b,t} \tag{5.14}$$

where $q_{s,t}$ and $q_{b,t}$ are investor buy and sell orders respectively. In the empirical implementation, we only observe a portion of trades that change the dealer inventory. Suppose that buy and sell orders are comprised of two components, one observed and indexed by one, and a correlated but unobserved component, indexed by two. Then the change in inventory is

$$\Delta I_t = q_{b,t} - q_{s,t} = q_{1,b,t} - q_{s,1,t} + q_{2,b,t} - q_{s,2,t} \tag{5.15}$$

So that inventory changes have two components, one observed and one unobserved.

$$\Delta I_t = x_{1,t} + x_{2,t} \tag{5.16}$$

The state space model cannot be estimated without making some assumptions about the correlation between the observed component, $x_{1,t}$, and the unobserved component $x_{2,t}$ of inventory changes. Suppose that the two components are linearly related through access to a common signal about fundamental value.

$$x_{2,t} = \gamma x_{1,t} + u_t \tag{5.17}$$

where the error term $u_t$ is orthogonal to $x_{1,t}$ and mean zero conditional on time t-1 information. The unconditional variance of this error term is $\sigma_u^2$. The error term represents additional information about fundamental value observed by investor 2 but not by investor 1. Denoting $(\Delta X_t - E_{t-1}[\Delta X_t])$ as $\Delta \tilde{X}_t$,

$$\tilde{x}_{2,t} = \gamma \tilde{x}_{1t} + u_t \tag{5.18}$$

$$\begin{aligned} \Delta \tilde{I}_t &= \tilde{x}_{1t} + \tilde{x}_{2t} \\ &= (1+\gamma)\tilde{x}_{1,t} + u_t \end{aligned} \tag{5.19}$$

$$\begin{aligned} \Delta I_t &= x_{1t} + x_{2t} \\ &= (1+\gamma)x_{1,t} + u_t \end{aligned} \tag{5.20}$$

Rewriting the state space model in terms of $x_{1,t}$ and $x_{2,t}$ we obtain

$$\begin{aligned} p_t &= m_t + s_t \\ m_t &= m_{t-1} + \kappa(\tilde{x}_{1t} + \tilde{x}_{2t}) + \eta_t \\ s_t &= \phi s_{t-1} + \beta(x_{1,t} + x_{2,t}) + \epsilon_t \end{aligned} \tag{5.21}$$

which is equivalent to,

$$\begin{aligned} p_t &= m_t + s_t \\ m_t &= m_{t-1} + \kappa^{'}(\tilde{x}_{1t}) + \eta_t^{'} \\ s_t &= \phi s_{t-1} + \beta^{'}\Delta(x_{1,t}) + \epsilon_t^{'} \end{aligned} \tag{5.22}$$

where $\kappa^{'} = (1+\gamma)\kappa$, $\beta^{'} = (1+\gamma)\beta$ and,

$$\begin{aligned} \eta_t^{'} &= \eta_t + \kappa u_t \\ \epsilon_t^{'} &= \epsilon_t + \beta u_t \end{aligned} \tag{5.23}$$

Therefore having a correlated unobserved transaction component $x_{2,t}$ implies that the error terms $\eta^{'}$ and $\epsilon^{'}$ in the state space mode are correlated with correlation coefficient $\kappa\beta\sigma_u^2$. The correlated unobserved component of trading adds an extra parameter to the state space model that is proportional to the variance of the information observed by investor 2 but not investor 1.[11]

[11]We thank Ioanid Rosu for suggesting this formulation.

### 5.6.2 Special case: investors have the same information set but differ in scale

A more restrictive specification would be to assume that investor one and investor two have the same information set, but trade in different quantities. This implies that trades of the two investors are proportional.

Suppose instead that investor two observes the same signal as investor one, but can trade in either greater or smaller scale, for example with the use of leverage. Then,

$$x_{2,t} = \gamma x_{1,t} \tag{5.24}$$

and

$$\begin{aligned} p_t &= m_t + s_t \\ m_t &= m_{t-1} + \kappa^{'}(\tilde{x}_{1t}) + \eta_t \\ s_t &= \phi s_{t-1} + \beta^{'}\Delta(x_{1,t}) + \epsilon_t \end{aligned} \tag{5.25}$$

where $\kappa^{'} = (1+\gamma)\kappa$, $\beta^{'} = (1+\gamma)\beta$. In this special case we can assume that the non-trade related innnovations $\eta_t$ and $\epsilon_t$ in the permanent and transitory components are independent. In the empirical implementation, $x_{1,t}$ is measured with mutual fund net sales ($Act_t$) and ($I_t - E_{t-1}[I_t]$) is measured with the discretionary component of mutual funds sales ($Disc_t$).

**Internet Appendix: Derivation of state space model**

## 5.7 Appendix - SSM Derivation

### 5.7.1 Notation

**Notation:**

- j: denotes fund.
- i: denotes stock.
- t: denotes quarter.
- $Hyp_{i,t}$: $\sum_j Hyp_{i,j,t}$.
- $Act_{i,t}$: $\sum_j Act_{i,j,t}$.
- $Disc_{i,t}$: $Act_{i,t}$ - $Hyp_{i,t}$.

**Theorem:** Through out the recursions, we will extensively use the following theorem. If $X \sim \mathcal{N}(\mu_X, \sigma_X)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y)$; and cov(X,Y)=$\sigma^2_{X,Y}$ then if follows:

$$u_{X|Y} = \mu_X + \frac{\sigma^2_{X,Y}}{\sigma^2_Y}(Y - \mu_Y) \tag{5.26}$$

$$\sigma^2_{X|Y} = \sigma^2_X - \frac{\sigma^2_{X,Y}}{\sigma^2_Y} \tag{5.27}$$

### 5.7.2 Model Setup

## SSM

For each Permno i=1,....I:

**Observed Logarithmic Price**

$$p_{i,t} = m_{i,t} + s_{i,t} \tag{5.28}$$

**Unobserved Efficient Price**

$$\begin{aligned} m_t &= m_{i,t} + \eta^*_{i,t} \\ \eta^*_{i,t} &= \kappa \widetilde{Disc}_{i,t} + \eta_{i,t} \end{aligned} \tag{5.29}$$

**Unobserved Transitory Price**

$$\begin{aligned} s_{i,t} &= \phi s_{i,t-1} + \epsilon^*_{i,t} \\ \epsilon^*_{i,t} &= \beta Act_{i,t} + \epsilon_{i,t} \end{aligned} \tag{5.30}$$

Where $\eta_{i,t} \sim \mathcal{N}(0, \sigma_{i,\eta})$ and $\epsilon_{i,t} \sim \mathcal{N}(0, \sigma_{i,\epsilon})$

**Discretionary Net Purchases - Panel Regression**

$$D_{i,t} = \xi_{i,t} + \gamma_i + \alpha_1 \text{Size}_{i,t-1} + \alpha_2 \text{Spreads}_{i,t-1} + \alpha_2 \text{Shrout}_{i,t-1} + \alpha_3 \text{Volatility}_{i,t-1} + \widetilde{Disc}_{i,t} \tag{5.31}$$

**SSM Prediction Errors, $\nu_{i,t}$**

$$\nu_{i,t} = p_{i,t} - m_{i,t|t-1} - \kappa \widetilde{Disc}_{i,t} - \phi s_{i,t|t-1} - \beta Act_{i,t} \tag{5.32}$$

**SSM Updating Steps (Updating Recursions)**

Starting from diffuse initial values $m_{i,1|0}, P^{m,\nu}_{i,1|0}$ and $s_{i,1|0}, P^{s,\nu}_{i,1|0}$

$$\begin{aligned} m_{t|t} &= \mathbb{E}[m_{i,t}|p_{i,t}] \\ &= \mathbb{E}[m_{i,t}|p_{i,t-1}, \nu_{i,t}] \\ &= m_{i,t|t-1} + P^{m,\nu}_{i,t|t-1} F^{-1}_{i,t} \nu_{i,t} \end{aligned} \tag{5.33}$$

$$\begin{aligned} s_{i,t|t} &= \mathbb{E}[s_{i,t}|p_{i,t}] \\ &= \mathbb{E}[s_{i,t}|p_{i,t-1}, \nu_{i,t}] \\ &= s_{i,t|t-1} + P^{s,\nu}_{i,t|t-1} F^{-1}_{i,t} \nu_{i,t} \end{aligned} \tag{5.34}$$

Since $m_{i,t}$ is a random walk:

$$\begin{aligned} m_{i,t+1|t} &= \mathbb{E}[m_{i,t}|p_{i,t}] \\ &= m_{t|t} \end{aligned} \tag{5.35}$$

Since $s_{i,t}$ is follows an autoregressive process:

$$\begin{aligned} s_{i,t+1|t} &= \mathbb{E}[s_{i,t}|p_{i,t}] \\ &= \phi s_{t|t} \end{aligned} \tag{5.36}$$

$$P^{m,\nu}_{i,t|t} = P^{m,\nu}_{i,t|t-1} - P^{m,\nu}_{i,t|t-1} F^{-1}_{i,t} P^{m,\nu}_{i,t|t-1} \tag{5.37}$$

$$P^{s,\nu}_{i,t|t} = P^{s,\nu}_{i,t|t-1} - P^{s,\nu}_{i,t|t-1} F^{-1}_{i,t} P^{s,\nu}_{i,t|t} \tag{5.38}$$

$$P^{m,\nu}_{i,t+1|t} = P^{m,\nu}_{i,t|t} + \sigma^2_{i,\eta} \tag{5.39}$$

$$P^{s,\nu}_{i,t+1|t} = P^{s,\nu}_{i,t|t} + \sigma^2_{i,\epsilon} \tag{5.40}$$

$$F_{i,t} = P^{s,\nu}_{i,t|t-1} + P^{s,\nu}_{i,t|t-1} + \sigma_{i,\eta} + \sigma^2_{i,\epsilon} \tag{5.41}$$

Where the recursions $P^{m,\nu}_{i,t|t-1} = \text{cov}(m_{t|t-1}, \nu_{i,t})$, $P^{s,\nu}_{i,t|t-1} = \text{cov}(s_{t|t-1}, \nu_{i,t})$ and $F_{i,t} = \text{var}(\nu_{i,t})$.

**SSM Smoothing Steps (Smoothing Recursions)**

$$\begin{aligned}
\hat{m}_{i,t} &= \mathbb{E}(m_{i,t}|p_{i,T}) \\
&= \mathbb{E}(m_{i,t}|p_{i,t-1}, \nu_{t:T}) \\
&= m_{t|t-1} + \sum_{j=t}^{T} \text{Cov}(m_{t|t-1}, \nu_{i,t}) F^{-1}_{i,t} \nu_{i,t} \\
&= m_{t|t-1} + P^{m,\nu}_{t|t-1} F^{-1}_{i,t} \nu_{i,t} + P^{m,\nu}_{t|t-1} L^m_{i,t} F^{-1}_{i,t+1} \nu_{i,t+1} + \cdots + P^{m,\nu}_{t|t-1} L^m_{i,t} ... L^m_{i,T-1} F^{-1}_{i,t} \nu_{i,t}
\end{aligned} \tag{5.42}$$

$$
\begin{aligned}
\hat{s}_{i,t} &= \mathbb{E}(s_{i,t}|p_{i,T}) \\
&= \mathbb{E}(s_{i,t}|p_{i,t-1}, \nu_{t:T}) \\
&= s_{t|t-1} + \sum_{j=t}^{T} \text{Cov}(s_{t|t-1}, \nu_{i,t}) F_{i,t}^{-1} \nu_{i,t} \\
&= s_{t|t-1} + P_{t|t-1}^{s,\nu} F_{i,t}^{-1} \nu_{i,t} + P_{t|t-1}^{s,\nu} L_{i,t}^{s} F_{i,t+1}^{-1} \nu_{i,t+1} + \cdots + P_{t|t-1}^{s,\nu} L_{i,t}^{s} ... L_{i,T-1}^{s} F_{i,t}^{-1} \nu_{i,t}
\end{aligned}
\tag{5.43}
$$

Where $L_{i,t}^{m} = 1 - P_{t|t-1}^{m,\nu} F_{i,t}^{-1}$ and $L_{i,t}^{s} = \phi(1 - P_{t|t-1}^{s,\nu} F_{i,t}^{-1})$

$$
\begin{aligned}
\hat{V}_{i,t}^{m} &= \text{Var}(m_{i,t}|p_{i,T}) \\
&= \text{Var}(m_{i,t}|p_{i,t-1}, \nu_{t:T}) \\
&= P_{i,t|t-1}^{m,\nu} - P_{i,t|t-1}^{m,\nu} F_{i,t}^{-1} P_{i,t|t-1}^{m,\nu} - P_{i,t|t-1}^{m,\nu} L_{i,t}^{m} F_{i,t}^{-1} L_{i,t}^{m} P_{i,t|t-1}^{m,\nu} - \cdots - P_{i,t|t-1}^{m,\nu} L_{i,t}^{m} \ldots \\
&\quad L_{i,T-1}^{m} F_{i,t}^{-1} L_{i,T-1}^{m} \ldots L_{i,t}^{m} P_{i,t|t-1}^{m,\nu}
\end{aligned}
\tag{5.44}
$$

$$
\begin{aligned}
\hat{V}_{i,t}^{s} &= \text{Var}(s_{i,t}|p_{i,T}) \\
&= \text{Var}(s_{i,t}|p_{i,t-1}, \nu_{t:T}) \\
&= P_{i,t|t-1}^{s,\nu} - P_{i,t|t-1}^{s,\nu} F_{i,t|t-1}^{-1} P_{i,t}^{s,\nu} - P_{i,t|t-1}^{s,\nu} L_{i,t}^{s} F_{i,t}^{-1} L_{i,t}^{s} P_{i,t|t-1}^{s,\nu} - \cdots - P_{i,t|t-1}^{s,\nu} L_{i,t}^{s} \ldots \\
&\quad L_{i,T-1}^{s} F_{i,t}^{-1} L_{i,T-1}^{s} \ldots L_{i,t}^{s} P_{i,t|t-1}^{s,\nu}
\end{aligned}
\tag{5.45}
$$

$$\begin{aligned}\hat{V}^s_{i,t,t+1} &= \text{Cov}(s_{i,t}, s_{i,t-1}|p_{i,T}) \\ &= \text{Cov}(s_{i,t}, s_{i,t-1}|p_{i,t-1}, \nu_{t:T}) \\ &= P^s_{t|t-1}L^s_{i,t}\Big(1 - (F^{-1}_{i,t+1} + L^s_{i,t+1}F^{-1}_{t+2}L^s_{i,t+1} + \cdots + L^s_{i,t+1}\ldots L_{i,T-1}F^{-1}_{i,t}L^s_{i,T-1}\ldots L_{i,t+1}) \\ &\quad P^{s,\nu}_{i,t|t-1}\Big)\end{aligned} \tag{5.46}$$

$$\begin{aligned}\hat{V}^m_{i,t,t+1} &= \text{Cov}(m_{i,t}, m_{i,t-1}|p_{i,T}) \\ &= \text{Cov}(m_{i,t}, s_{i,t-1}|p_{i,t-1}, \nu_{t:T}) \\ &= P^m_{t|t-1}L^m_{i,t}\Big(1 - (F^{-1}_{i,t+1} + L^m_{i,t+1}F^{-1}_{t+2}L^m_{i,t+1} + \cdots + L^m_{i,t+1}\ldots L_{i,T-1}F^{-1}_{i,t}L^m_{i,T-1}\ldots L_{i,t+1}) \\ &\quad P^{m,\nu}_{i,t|t-1}\Big)\end{aligned} \tag{5.47}$$

Where $L^m_{i,t} = 1 - P^{m,\nu}_{t|t-1}F^{-1}_{i,t}$ and $L^s_{i,t} = \phi(1 - P^{s,\nu}_{t|t-1}F^{-1}_{i,t})$

### 5.7.3 Likelihood

Recall from Section 2:

$$\nu_{i,t} = p_{i,t} - m_{i,t|t-1} - \kappa\widetilde{Disc}_{i,t} - \phi s_{i,t|t-1} - \beta Act_{i,t} \tag{5.48}$$

It follows from the SSM model: $\nu_{i,t} \sim \mathcal{N}(0, F_{i,t})$.

Therefore, the log-likelihood function is as follows:

$$\mathcal{L}_i \propto -\frac{1}{2}\sum_{t=1}^{T}\left(\nu_{i,t}F^{-1}_{i,t}\nu_{i,t}\right) \tag{5.49}$$

Note that $\nu_{i,t}$ is a function of $\kappa_i$,$\beta_i$ and $\phi_i$; and $F_{i,t}$ is an implicit function of $\sigma_{i,\eta}$ and $\sigma_{i,\epsilon}$. It follows that the maximization problem is as follows:

$$\max_{\kappa_i,\beta_i,\sigma_{i,\eta},\sigma_{i,\epsilon}} \left\{ \frac{1}{2} \sum_{t=1}^{T} \left( \nu_{i,t} F_{i,t}^{-1} \nu_{i,t} \right) \right\} \tag{5.50}$$

### 5.7.4 Maximization using EM Algorithm

We use the expectation–maximization (EM) algorithm is used to obtain starting values of $\kappa_i, \beta_i, \sigma_{i,\eta}$ and $\sigma_{i,\epsilon}$ . We initially set $\phi = 0$ so that the model is linear and the EM algorithm can be conveniently applied. From the likelihood function the first order conditions (FOC) are as follows:

FOC $\kappa_i$:

$$\hat{\kappa}_i = \frac{\sum_{t=1}^{T-1} \Delta m_{i,t+1|t} \widetilde{Disc}_{i,t+1}}{\sum_{t=1}^{T-1} \widetilde{Disc}_{i,t+1}^{2}} \tag{5.51}$$

FOC $\beta_i$:

$$\hat{\beta}_i = \frac{\sum_{t=0}^{T-1} s_{i,t+1|t} Act_{i,t+1}}{\sum_{t=1}^{T} Act_{i,t+1}^2} \tag{5.52}$$

FOC $\sigma_{\eta,i}^2$:

$$\hat{\sigma_{i,\eta}}^2 = \frac{1}{T-1} \sum_{t=1}^{T-1} \left( m_{t+1|t} - m_{t|t-1} - \kappa \widetilde{Disc}_{i,t+1} \right)^2 \tag{5.53}$$

FOC $\sigma_{\epsilon,i}^2$:

$$\hat{\sigma_{i,\epsilon}}^2 = \frac{1}{T-1} \sum_{t=1}^{T} \left( s_{t+1|t} - \beta Act_{i,t+1} \right)^2 \tag{5.54}$$

The expectation and maximization steps can be combined by taking expectations in (23)-(26) and using smoothing recursions (17)-(20).

$$\hat{\kappa}_i^{EM} = \frac{\sum_{t=1}^{T-1} \Delta\hat{m}_{i,t+1}\widetilde{Disc}_{i,t}}{\sum_{t=1}^{T-1} \widetilde{Disc}_{i,t}^{2}} \tag{5.55}$$

$$\hat{\beta}_i^{EM} = \frac{\sum_{t=0}^{T-1} \hat{s}_{i,t+1} Act_{i,t}}{\sum_{t=1}^{T-1} Act_{i,t+1}^2} \tag{5.56}$$

$$\hat{\sigma}_{i,\eta}^{EM} = \frac{1}{T-1}\sum_{t=1}^{T-1}\left((\hat{m}_{i,t+1} - \hat{m}_{i,t} - \hat{\kappa}_i^{EM}\widetilde{Disc}_{i,t+1})^2 + \hat{V}_{i,t+1}^m + \hat{V}_{i,t}^m - 2\hat{V}_{i,t+1,t}^m\right) \tag{5.57}$$

$$\hat{\sigma}_{i,\epsilon}^{EM} = \frac{1}{T-1}\sum_{t=1}^{T-1}\left((\hat{s}_{i,t} - \hat{\beta}_i^{EM} Act_{i,t+1})^2 + \hat{V_{i,t}}^s\right) \tag{5.58}$$

From equations (27)-(30) it follows that the k-th EM algorithm recursion is as follows:

$$\begin{aligned}
\hat{\kappa}_i^{EM,(k)} &= \frac{\sum_{t=1}^{T-1} \Delta\hat{m}_{i,t}^{(k-1)}\widetilde{Disc}_{i,t}}{\sum_{t=1}^{i,t} \widetilde{Disc}_{i,t}^{2}} \\
\hat{\beta}_i^{EM,(k)} &= \frac{\sum_{t=1}^{T-1} \hat{s}_{i,t}^{(k-1)} Act_{i,t}}{\sum_{t=1}^{T} Act_{i,t}^2} \\
\hat{\sigma}_{i,\eta}^{(k)} &= \frac{1}{T-1}\sum_{t=1}^{T}\left((\hat{m}_{i,t+1}^{(k-1)} - \hat{m}_{i,t}^{(k-1)} - \hat{\kappa}_i^{EM,(k-1)}\widetilde{Disc}_{i,t})^2 + \hat{V}_{i,t+1}^{m,(k-1)} + \hat{V}_{i,t}^{m,(k-1)} - 2\hat{V}_{i,t+1,t}^{m,(k-1)}\right) \\
\hat{\sigma}_{i,\epsilon}^{(k)} &= \frac{1}{T-1}\sum_{t=1}^{T-1}\left((\hat{s}_{i,t+1}^{(k-1)} - \hat{\beta}_i^{EM,(k-1)} Act_{i,t})^2 + \hat{V}_{i,t+1}^{s,(k-1)}\right)
\end{aligned} \tag{5.59}$$

Using 250 iteration of the EM algorithm we obtain EM estimates $\hat{\kappa}_i^{EM}, \hat{\beta}_i^{EM}, \hat{\sigma}_{i\eta}^{EM}$ and $\hat{\sigma}_{i\eta}^{EM}$.

### 5.7.5 Grid Search

Starting with initial values $\hat{\kappa}_i^{EM}, \hat{\beta}_i^{EM}, \hat{\sigma}_{i\eta}^{EM}, \hat{\sigma}_{i\eta}^{EM}$ we use a grid search on the likelihood function $\mathcal{L}_i$ over the interval $\phi_i \in (-0.9, 0.9)$ in increments of 0.05. Using the aforementioned procedure we obstain an initial value for the autoregressive paramter $\phi$.

$$\phi_i^{Grid} := \max_{\phi \in (-0.9, 0.9)} \mathcal{L}(\hat{\kappa}_i^{EM}, \hat{\beta}_i^{EM}, \hat{\sigma}_{i\eta}^{EM}, \hat{\sigma}_{i\eta}^{EM}) \tag{5.60}$$

### 5.7.6 Constrained Maximization

Starting with initial values $\hat{\kappa}_i^{EM}, \hat{\beta}_i^{EM}, \hat{\sigma}_{i\eta}^{EM}, \hat{\sigma}_{i\eta}^{EM}$ and $\phi_i^{Grid}$ we do an unconstrained maximization of $\mathcal{L}_i$ using the Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm with the restriction $\phi \in (-0.9, 0.9)$.

### 5.7.7 Unconstrained Maximization

Starting with initial values from the constrained maximization we do an unconstrained maximization of $\mathcal{L}_i$ using the BFGS algorithm.

### 5.7.8 Diffuse Initialization

We initialize the efficient price process $m$, assuming an uninformative prior as follows:

$$m_{i,0} \sim \mathcal{N}(0, \delta) \tag{5.61}$$

where we let $\delta \to \infty$; i.e. we assume a diffuse prior for $m_{i,t}$. Following Durbin and Koopman [2012] the diffuse initialization equations are given by the following equations:

$$P_{i,2|1} = TP^{\infty}_{i,1}L^{(1)\top}_{i,1} + TP^{*}_{i,1}L^{0\top}_{i,1} + Q \tag{5.62}$$

$$\begin{aligned}\begin{pmatrix} m_{i,2|1} \\ s_{i,2|1} \end{pmatrix} &= K^0_1 \begin{pmatrix} p_{i,1} & \mathbb{E}(s_t) \end{pmatrix} \\ &= K^0_1 \begin{pmatrix} p_{i,1} & 0 \end{pmatrix}\end{aligned} \tag{5.63}$$

where $Q$ is the diagonal matrix of variances $\begin{pmatrix} \sigma^2_\eta & 0 \\ 0 & \sigma^2_\epsilon \end{pmatrix}$, T is the transition matrix $\begin{pmatrix} 1 & 0 \\ 0 & \phi \end{pmatrix}$ and $P^{\infty}_{i,1}, P^{*}_{i,1}, L^{0}_{i,1}, K^0_1$ can be solved using second order Taylor approximation around $F^{-1}_{i,t}$ and the Kalman gain matrix with $\delta \to \infty$. Substituting solutions to $P^{\infty}_{i,1}, P^{*}_{i,1}, L^{0}_{i,1}, K^0_1$ in 5.63 we obtain the diffuse initialization as follows:

$$\begin{aligned}\begin{pmatrix} m_{i,2|1} \\ s_{i,2|1} \end{pmatrix} &= \begin{pmatrix} p_{i,1} \\ 0 \end{pmatrix} \\ P_{i,2|1} &= \begin{pmatrix} \sigma^2_0 & -\phi\sigma^2_0 \\ -\phi\sigma^2_0 & (\phi^2+1)\sigma^2_0 \end{pmatrix}\end{aligned} \tag{5.64}$$

where $\sigma^2_0$ is the unconditional variance of transitory price $\frac{\sigma^2_\epsilon}{1-\phi^2}$.

# Bibliography

S James Press. A compound events model for security prices. *Journal of business*, pages 317–335, 1967.

Peter K Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica: journal of the Econometric Society*, pages 135–155, 1973.

Albert S Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335, 1985a.

Lawrence R Glosten and Paul R Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1):71–100, 1985.

Tim Bollerslev and Viktor Todorov. Tails, fears and risk premia. *Journal of Finance*, 66(6):2165–2211, 2011.

Tim Bollerslev, Sophia Zhengzi Li, and Viktor Todorov. Roughing up beta: Continuous versus discontinuous betas and the cross section of expected stock returns. *Journal of Financial Economics*, 120(3):464–490, 2016.

J.F. Bégin, C. Dorion, and G. Gauthier. Idiosyncratic jump risk matters: Evidence from equity returns and options. *Forthcoming - Review of Financial Studies*, 2019.

Maureen O'Hara and Mao Ye. Is market fragmentation harming market quality? *Journal of Financial Economics*, 100(3):459–474, 2011.

Albert J Menkveld and Bart Zhou Yueshen. The flash crash: A cautionary tale about highly fragmented markets. *Management Science*, 2018.

Torben G Andersen, Luca Benzoni, and Jesper Lund. An empirical investigation of continuous-time equity return models. *The Journal of Finance*, 57(3):1239–1284, 2002.

Bollerslev Andersen, T. G., P. T., Frederiksen, and M. Ørregaard Nielsen. Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns. *Journal of Applied Econometrics*, 25(2):233–261, 2010.

Torben G Andersen, Tim Bollerslev, Francis X Diebold, and Clara Vega. Real-time price discovery in global stock, bond and foreign exchange markets. *Journal of International Economics*, 73(2):251–277, 2007.

George J. Jiang and Tong Yao. Stock price jumps and cross-sectional return predictability. *Journal of Financial and Quantitative Analysis*, 48(5):1519–1544, 2013a.

Thomas H. McCurdy Jeon, Yoontae and Xiaofei Zhao. News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies. *Available at SSRN 3318517*, 2019.

Bento J. Lobo. Jump risk in the us stock market: Evidence using political information. *Review of Financial Economics*, 8(2):149–163, 1999.

Mark Mitchell and Todd Pulvino. Arbitrage crashes and the speed of capital. *Journal of Financial Economics*, 104(3):469–490, 2012.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.

Robert C Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1):125–144, 1976.

David S Bates. Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *The Review of Financial Studies*, 9(1):69–107, 1996.

Steven G Kou. A jump-diffusion model for option pricing. *Management science*, 48(8):1086–1101, 2002.

Bjørn Eraker, Michael Johannes, and Nicholas Polson. The impact of jumps in volatility and returns. *The Journal of Finance*, 58(3):1269–1300, 2003.

George J Jiang and Tong Yao. Stock price jumps and cross-sectional return predictability. *Journal of Financial and Quantitative Analysis*, 48(5):1519–1544, 2013b.

Tim Bollerslev, Viktor Todorov, and Lai Xu. Tail risk premia and return predictability. *Journal of Financial Economics*, 118(1):113–134, 2015.

Jeroen VK Rombouts, Lars Stentoft, and Francesco Violante. Dynamics of variance risk premia: A new model for disentangling the price of risk. *Journal of Econometrics*, 2019.

Marco Pagano. Trading volume and asset liquidity. *The Quarterly Journal of Economics*, 104(2):255–274, 1989.

Bhagwan Chowdhry and Vikram Nanda. Multimarket trading and market liquidity. *The Review of Financial Studies*, 4(3):483–511, 1991.

Christine A Parlour and Duane J Seppi. Liquidity-based competition for order flow. *The Review of Financial Studies*, 16(2):301–343, 2003.

Thierry Foucault and Albert J. Menkveld. Competition for order flow and smart order routing systems. *The Journal of Finance*, 63(1):119–158, 2008.

Paul Bennett and Li Wei. Market structure, fragmentation, and market quality. *Journal of Financial Markets*, 9(1):49–78, 2006.

Frank De Jong Degryse, Hans and Vincent van Kervel. The impact of dark trading and visible fragmentation on market quality. *Review of Finance*, 19(4): 1587–1622, 2015.

Xiangkang Yin. A comparison of centralized and fragmented markets with costly search. *The Journal of Finance*, 60(3):1567–1590, 2005.

Ananth Madhavan. Exchange-traded funds, market structure, and the flash crash. *Financial Analysts Journal*, 68(4):20–35, 2012.

James Upson and Robert A Van Ness. Multiple markets, algorithmic trading, and market liquidity. *Journal of Financial Markets*, 32:49–68, 2017.

Sanford J Grossman and Joseph E Stiglitz. On the impossibility of informationally efficient markets. *The American economic review*, 70(3):393–408, 1980.

Neil Shephard and Michael K Pitt. Likelihood analysis of non-gaussian measurement time series. *Biometrika*, 84(3):653–667, 1997.

J Durbin and SJ Koopman. Filtering, smoothing and estimation for time series when the observations come from exponential family distributions. *London School of Economics Statistics Research Report*, 1994.

James Durbin and Siem Jan Koopman. Monte carlo maximum likelihood estimation for non-gaussian state space models. *Biometrika*, 84(3):669–684, 1997.

James Durbin and Siem Jan Koopman. Time series analysis of non-gaussian observations based on state space models from both classical and bayesian perspectives. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):3–56, 2000.

Albert J Menkveld, Siem Jan Koopman, and André Lucas. Modeling around-the-clock price discovery for cross-listed stocks using state space methods. *Journal of Business & Economic Statistics*, 25(2):213–225, 2007.

Jonathan Brogaard, Terrence Hendershott, and Ryan Riordan. High-frequency trading and price discovery. *The Review of Financial Studies*, 27(8):2267–2306, 2014.

Kim Christensen, Roel CA Oomen, and Mark Podolskij. Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics*, 114(3):576–599, 2014.

Suzanne S Lee and Per A Mykland. Jumps in financial markets: A new non-parametric test and jump dynamics. *The Review of Financial Studies*, 21(6): 2535–2563, 2008.

Terrence Hendershott and Albert J Menkveld. Price pressures. *Journal of Financial Economics*, 114(3):405–423, 2014.

James Durbin and Siem Jan Koopman. A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–616, 2002.

Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

Ole E Barndorff-Nielsen and Neil Shephard. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of financial Econometrics*, 4(1):1–30, 2006.

George J Jiang and Roel CA Oomen. Testing for jumps when asset prices are observed with noise–a “swap variance” approach. *Journal of Econometrics*, 144 (2):352–370, 2008.

Yacine Aït-Sahalia and Jean Jacod. Testing for jumps in a discretely observed process. *The Annals of Statistics*, 37(1):184–222, 2009.

Craig W Holden and Stacey Jacobsen. Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *The Journal of Finance*, 69(4):1747–1785, 2014.

David Easley and Maureen O’hara. Price, trade size, and information in securities markets. *Journal of Financial economics*, 19(1):69–90, 1987.

Bruce N Lehmann. Fads, martingales, and market efficiency. *The Quarterly Journal of Economics*, 105(1):1–28, 1990.

Andrew W Lo and A Craig MacKinlay. When are contrarian profits due to stock market overreaction? *The review of financial studies*, 3(2):175–205, 1990.

Stefan Nagel. Evaporating liquidity. *The Review of Financial Studies*, 25(7): 2005–2039, 2012.

Harald Hau. Location matters: An examination of trading profits. *The Journal of Finance*, 56(5):1959–1983, 2001.

Hans R Stoll. The supply of dealer services in securities markets. *The Journal of Finance*, 33(4):1133–1151, 1978.

Frederic S Mishkin. Is monetary policy effective during financial crises? *American Economic Review*, 99(2):573–77, 2009.

Michael J Fishman and Kathleen M Hagerty. Disclosure decisions by firms and the competition for price efficiency. *The Journal of Finance*, 44(3):633–646, 1989.

Charles MC Lee. Market efficiency and accounting research: a discussion of 'capital market research in accounting'by sp kothari. *Journal of Accounting and Economics*, 31(1-3):233–253, 2001.

Ekkehart Boehmer, Charles M Jones, and Xiaoyan Zhang. Shackling short sellers: The 2008 shorting ban. *The Review of Financial Studies*, 26(6):1363–1400, 2013.

Jonathan Brogaard, Terrence Hendershott, and Ryan Riordan. High frequency trading and the 2008 short-sale ban. *Journal of Financial Economics*, 124(1): 22–42, 2017.

David Easley, Soeren Hvidkjaer, and Maureen O'hara. Is information risk a determinant of asset returns? *The journal of finance*, 57(5):2185–2221, 2002.

Maureen O'Hara. Presidential address: Liquidity and price discovery. *The Journal of Finance*, 58(4):1335–1354, 2003.

Lawrence E. Harris and Venkatesh Panchapagesan. The information content of the limit order book: evidence from nyse specialist trading decisions. *Journal of Financial Markets*, 8(1):25–67, 2005.

Vincent Van Kervel. Competition for order flow with fast and slow traders. *The Review of Financial Studies*, 28(7):2094–2127, 2015.

Viktor Todorov and George Tauchen. Inverse realized laplace transforms for nonparametric volatility density estimation in jump-diffusions. *Journal of the American Statistical Association*, 107(489):622–635, 2012.

J. Brogaard, T. Hendershott, and Riordan. Price discovery without trading: Evidence from limit orders. *Forthcoming - Journal of Finance*, 2019.

Joel Hasbrouck. Measuring the information content of stock trades. *Journal of Finance*, 46(1):179–207, 1991a.

Joel Hasbrouck. The summary informativeness of stock trades: An econometric analysis. *Review of Financial Studies*, 4(3):571–595, 1991b.

Sanford J. Grossman and Merton H. Miller. Liquidity and market structure. *Journal of Finance*, 43(3):617–633, 1988.

Eva Deuchert and Martin Huber. A cautionary tale about control variables in iv estimation. *Oxford Bulletin of Economics and Statistics*, 79(3):411–425, 2017.

Jay F. Coughenour and Mohsen M. Saad. Common market makers and commonality in liquidity. *Journal of Financial Economics*, 73(1):37–69, 2004.

Thomas Ho and Hans R. Stoll. Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, 9(1):47–73, 1981.

Andrei Shleifer and Lawrence H Summers. The noise trader approach to finance. *Journal of Economic perspectives*, 4(2):19–33, 1990.

Markus K Brunnermeier and Lasse Heje Pedersen. Market liquidity and funding liquidity. *The review of financial studies*, 22(6):2201–2238, 2009.

Pierre Collin-Dufresne and Vyacheslav Fos. Do prices reveal the presence of informed trading? *The Journal of Finance*, 70(4):1555–1582, 2015.

Qi Chen, Itay Goldstein, and Wei Jiang. Price informativeness and investment sensitivity to stock price. *Review of Financial Studies*, 20(3):619–650, 2007.

Yuanzhi Luo. Do insiders learn from outsiders? evidence from mergers and acquisitions. *Journal of Finance*, 60(4):1951–1982, 2005.

Joshua Coval and Erik Stafford. Asset fire sales (and purchases) in equity markets. *Journal of Financial Economics*, 86(2):479–512, 2007.

Malcolm Wardlaw. Measuring mutual fund flow pressure as shock to stock returns. *Available at SSRN 3248750*, 2019.

Sheng Huang, Matthew C. Ringgenberg, and Zhe Zhang. The information in asset fire sales. *Unpublished working paper*, 2019.

Martin F. Hellwig. On the aggregation of information in competitive markets. *Journal of Economic Theory*, 22:477–498, 1980.

Anat R Admati. A noisy rational expectations equilibrium for multi-asset securities markets. *Econometrica: Journal of the Econometric Society*, pages 629–657, 1985.

Snehal Banerjee. Learning from prices and the dispersion in beliefs. *The Review of Financial Studies*, 24(9):3025–3068, 2011.

Nicolas PB Bollen and Jeffrey A Busse. On the timing ability of mutual fund managers. *The Journal of Finance*, 56(3):1075–1094, 2001.

Andrea Frazzini and Owen A Lamont. Dumb money: Mutual fund flows and the cross-section of stock returns. *Journal of financial economics*, 88(2):299–322, 2008.

Gordon M Phillips and Alexei Zhdanov. R&d and the incentives from merger and acquisition activity. *The Review of Financial Studies*, 26(1):34–78, 2013.

François Derrien, Ambrus Kecskés, and David Thesmar. Investor horizons and corporate policies. *Journal of Financial and Quantitative Analysis*, 48(6):1755–1780, 2013.

Jonathan Brogaard, Thanh Huong Nguyen, Tālis J Putniņš, and Eliza Wu. What moves stock prices? the role of news, noise, and information. In *AFA Annual Meeting*, 2020.

Jennie Bai, Thomas Philippon, and Alexi Savov. Have financial markets become more informative? *Journal of Financial Economics*, 122(3):625–654, 2016.

Eduardo Dávila and Cecilia Parlatore. Identifying price informativeness. Technical report, National Bureau of Economic Research, 2018.

Olivier Dessaint, Thierry Foucault, Laurent Frésard, and Adrien Matray. Noisy stock prices and corporate investment. *The Review of Financial Studies*, 32(7): 2625–2672, 2019.

Jiang Wang. A model of competitive stock trading and volume. *Journal of Political Economy*, 102(1):127–168, 1994.

Albert S Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335, 1985b.

Thomas J George and Chuan-Yang Hwang. Information flow and pricing errors: A unified approach to estimation and testing. *The Review of Financial Studies*, 14(4):979–1020, 2001.

B Espen Eckbo and Jian Liu. Temporary components of stock prices: new univariate results. *Journal of Financial and Quantitative Analysis*, 28(2):161–176, 1993.

James Durbin and Siem Jan Koopman. *Time series analysis by state space methods*. Oxford university press, 2012.

Mikhail Simutin. Cash holdings and mutual fund performance. *Review of Finance*, 118:1425–1464, 2014.

Sergey Chernenko and Adi Sunderam. Do fire sales create externalities? *Journal of Financial Economics*, 135(3):602–628, 2020.

R. Morck, A. Shleifer, and R. Vishny. The stock market and investment: Is the market a side show? *Brookings Papers on Economic Activity*, 2:157–215, 1990.

Tor-Erik Bakke and Toni M Whited. Which firms follow the market? an analysis of corporate investment decisions. *The Review of Financial Studies*, 23(5): 1941–1980, 2010.

Xiaoxia Lou and Albert Y. Wang. Flow-induced trading pressure and corporate investment. *Journal of Financial and Quantitative Analysis*, 53(1):171–201, 2018.

Malcolm Baker, Jeremy C Stein, and Jeffrey Wurgler. When does the market matter? stock prices and the investment of equity-dependent firms. *The Quarterly Journal of Economics*, 118(3):969–1005, 2003.

Thierry Foucault and Laurent Fresard. Learning from peers' stock prices and corporate investment. *Journal of Financial Economics*, 111(3):554–577, 2014.

Gerard Hoberg and Gordon Phillips. Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5):1423–1465, 2016.

Markus K Brunnermeier, K Markus, and Markus Konrad Brunnermeier. *Asset pricing under asymmetric information: Bubbles, crashes, technical analysis, and herding*. Oxford University Press on Demand, 2001.

Juliane Begenau, Maryam Farboodi, and Laura Veldkamp. Big data in finance and the growth of large firms. *Journal of Monetary Economics*, 97:71–87, 2018.