

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Linking Historical Market Crashes: A Market Microstructure Model and Statistical Evidence

Permalink

<https://escholarship.org/uc/item/3dr744vc>

Author

Mao, Yuan

Publication Date

2015

Peer reviewed|Thesis/dissertation

**Linking Historical Market Crashes:
A Market Microstructure Model and Statistical Evidence**

by

Yuan Mao

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Engineering - Industrial Engineering and Operations Research

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Xin Guo, Chair
Professor Terrence Hendershott
Professor Ilan Adler

Summer 2015

**Linking Historical Market Crashes:
A Market Microstructure Model and Statistical Evidence**

Copyright 2015
by
Yuan Mao

Abstract

Linking Historical Market Crashes:
A Market Microstructure Model and Statistical Evidence

by

Yuan Mao

Doctor of Philosophy in Engineering - Industrial Engineering and Operations Research

University of California, Berkeley

Professor Xin Guo, Chair

Studies of stock market crashes are as sparse as the occurrence of crashes. The mainstream theoretical models on stock market crashes are rooted in rational expectations equilibrium models and classical market microstructure models. Compared to theoretical works, there are even fewer works done on the empirical side. This is because most of the theoretical models do not provide straightforward tests against empirical data. Secondly, the relatively small sample size (rare occurrence) of stock market crashes is always an obstacle for empirical testing.

In this dissertation, we build a strategic trading model to link two major US stock market crashes: the 1987 crash and the 2010 Flash Crash. We then provide cross-sectional empirical evidence to verify our model hypothesis and evaluate price impact due to the information asymmetry effect and the limited risk-bearing capacity effect. We use statistical learning methods to compare our model based predictors with other predictors for the maximum cross-sectional price drawdown of SP500 stocks during the 2010 Flash Crash and check the robustness of our findings.

Contents

Contents	i
List of Figures	iii
List of Tables	v
1 Introduction	1
2 Literature	4
2.1 Rational Expectations Equilibrium (REE) Models	4
2.1.1 The Walrasian Model	4
2.1.2 Extension	6
2.1.3 Noisy (Supply) REE	7
2.2 Classic Market Microstructure Models	9
2.2.1 Adverse Selection	9
2.2.2 Inventory and Order Handling Cost	11
2.2.3 A Synthetic Model	12
2.3 Crashes	13
2.3.1 Crashes in Competitive REE Models	13
2.3.2 Crashes in Sequential Trade Models	14
2.3.3 Physical Models of Crashes	15
2.3.4 Recent Research	15
3 Strategic Trading Models and Empirical Test	17
3.1 A Single Asset Model and Empirical Test	17
3.1.1 Model	17
3.1.2 Empirical Test of the Model	18
3.1.3 Empirical Implementation and Data	19
3.1.4 Estimation of the Market-Wide Liquidity Shock	20
3.1.5 Estimation of the Limited Risk-Bearing Capacity Induced Price Impact	22
3.2 A Multi-Asset Model and Empirical Discussion	24
3.2.1 Model	24

3.2.2	Empirical Discussion	26
3.3	Modified γ^* under imperfect hedging	30
3.4	Proofs for Section 3.1	31
3.5	Proofs for Section 3.2	32
3.6	Summary	36
4	Statistical Data Analysis	37
4.1	Association Analysis	37
4.1.1	Data	37
4.1.2	Exploratory Data Analysis	38
4.1.3	Association Analysis Results	44
4.2	Statistical Model Selection	47
4.2.1	Model Selection	47
4.2.2	Data	50
4.2.3	Model Selection Results	53
4.2.4	Comparison with Betas of Risk Factors	55
4.2.5	Comparison with Corporate Default Predictors	57
4.2.6	Discussion	59
4.3	Other statistical methods for 39 predictors	62
4.4	Summary	70
5	Conclusion and Future Work	72
	Bibliography	75
A		80

List of Figures

1.1	The Oct. 1987 Crash, Source: US Federal Reserve.	1
1.2	The 2010 Flash Crash, Source: Staffs of the CFTC and SEC (2010b).	2
1.3	Comparison of the Oct. 1987 crash and the 2010 Flash Crash.	3
3.1	The Oct. 1987 Crash.	21
3.2	The 2010 Flash Crash.	22
4.1	Barplot of frequency of three categorical parameters for the 2010 Flash Crash.	38
4.2	Boxplots of PRC_CHG, ESPRD, and VOLUME_RATIO, grouped by GSECTOR for the 2010 Flash Crash. Sector definition: 10-Energy; 15-Materials; 20-Industrials; 25-Consumer discretionary; 30-Consumer staples; 35-Healthcare; 40-Financials; 45-Information technology; 50-Telecommunication services; 55-Utilities.	39
4.3	Boxplots of PRC_CHG, ESPRD, and VOLUME_RATIO, grouped by PRIMEXCH for the 2010 Flash Crash.	39
4.4	Boxplots of PRC_CHG, ESPRD, and VOLUME_RATIO, grouped by SPCSRC for the 2010 Flash Crash (NA means stocks without a SPCSRC).	40
4.5	Correlation matrix of numerical parameters in the 2010 Flash Crash.	41
4.6	Barplot of frequency of GSECTOR and SPCSRC for the 1987 crash.	42
4.7	Boxplots of PRC_CHG, ESPRD, and VOLUME_RATIO, grouped by GSECTOR for the 1987 crash. Sector definition: 10-Energy; 15-Materials; 20-Industrials; 25-Consumer discretionary; 30-Consumer staples; 35-Healthcare; 40-Financials; 45-Information technology; 50-Telecommunication services; 55-Utilities.	42
4.8	Boxplots of PRC_CHG, ESPRD, and VOLUME_RATIO, grouped by SPCSRC for the 1987 crash (NA means stocks without a SPCSRC).	43
4.9	Correlation matrix of numerical parameters in the 1987 crash.	44
4.10	Best subset selection for 8 predictors.	53
4.11	The Lasso for 8 predictors. Top left: Standardized coefficients of 8 predictors against $\log(\lambda)$; Top right: Standardized coefficients of 8 predictors against R^2 explained; Bottom left: best tuning parameter selection with minimum square root cross-validation MSE; Bottom right: best tuning parameter selection with one-standard-error rule (MSE with one standard error band at each point).	54
4.12	Best subset selection for 15 predictors.	55

4.13	The Lasso for 15 predictors. Top left: Standardized coefficients of 15 predictors against $\log(\lambda)$; Top right: Standardized coefficients of 15 predictors against R^2 explained; Bottom left: best tuning parameter selection with minimum square root cross-validation MSE; Bottom right: best tuning parameter selection with one-standard-error rule (MSE with one standard error band at each point). . . .	56
4.14	Forward stepwise selection for 39 predictors.	58
4.15	The Lasso for 39 predictors. Top left: Standardized coefficients of 39 predictors against $\log(\lambda)$; Top right: Standardized coefficients of 39 predictors against R^2 explained; Bottom left: best tuning parameter selection with minimum square root cross-validation MSE; Bottom right: best tuning parameter selection with one-standard-error rule (MSE with one standard error band at each point). . . .	59
4.16	Ridge regression for 39 predictors. Left: cross-validation MSE and best tuning parameter; Right: Standardized coefficients of 39 predictors against $\log(\lambda)$, blue dotted lines are best tuning parameter selection with minimum CV-MSE and one-standard-error rule respectively.	63
4.17	Principal component regression for 39 predictors. Left: square root cross-validation MSE and best number of components is 14; Right: Standardized coefficients of 39 predictors against number of components.	65
4.18	Partial least squares for 39 predictors. Left: square root cross-validation MSE and best number of components is 2; Right: Standardized coefficients of 39 predictors against number of components.	66
4.19	Recursive feature elimination for SVR: use cross-validation to choose the optimal number of features. The optimal number of features selected is 9.	69
4.20	Recursive feature elimination for SVR: use cross-validation to choose the optimal number of features. The hyperparameters are chosen by a grid search. The optimal number of features selected is 14.	69

List of Tables

3.1	Data handling details: ¹ We did not count Oct. 20 in the crash window as the trading activity during Oct. 20 was very unusual. ² We define the lowest price during the crash to be the transaction price below which the security traded for at least certain cumulative time period during the crash window. ³ For 2010 Flash Crash, we filter out stocks with maximum drawdown of more than 15%, as that can be categorized as broken trades. For 1987 crash, we filter out stocks whose ISSM prices do not match CRSP price and also stocks with less than 30 trades on Oct. 19.	20
3.2	Estimates of α : market drawdown (percentage) and θ : market-wide liquidity shock (\$ Billion). <i>Note</i> : *p<0.1; **p<0.05; ***p<0.01	20
3.3	Estimates of α : percentage and θ_1 : market-wide liquidity shock (\$ Billion). <i>Note</i> : *p<0.1; **p<0.05; ***p<0.01	23
3.4	Estimates of α : percentage and θ_1 : market-wide liquidity shock (\$ Billion). <i>Note</i> : *p<0.1; **p<0.05; ***p<0.01	24
4.1	Summary Statistics	52
4.2	Selection Sequence for 8 candidate predictors.	55
4.3	Selection Sequence for 15 candidate predictors. The sequence is not necessarily nested for best subset selection, here β_{smd} was in the best five-predictor set, but was dropped in the best six-predictor set.	57
4.4	Selection Sequence for 39 candidate predictors.	58
4.5	Model selection results for 345 stocks in regression. The standard errors of regression coefficients are estimated by bootstrapping 1000 samples and are shown in the parentheses.	60
4.6	A comparison of model selection result with different data sets. Both Set 1-2 and Set 1-3 have 457 SP500 stocks, Set 1-4 has 345 stocks. We show the result of the Lasso method that applies one-standard-error rule. The results of the best model of selected by the Lasso method are shown in quote.	61
4.7	Model selection results for 457 stocks in regression. The standard errors of regression coefficients are estimated by bootstrapping 1000 samples and are shown in the parentheses.	62

4.8	A comparison of model prediction errors with different statistical methods for Set 1-4 predictors. We show the results of the Lasso method and ridge regression method that apply one-standard-error rule. The results of the best model of selected by the Lasso method and ridge regression method are shown in quote. The SVR result is from the RFE-SVR algorithm with hyperparameters set by cross-validation and applies a linear kernel.	70
A.1	Association analysis: Parameters and data source, 2010 Flash Crash	81
A.2	Association analysis: Parameters and data source, 1987 crash	82
A.3	Statistical model selection: Corporate default parameters abbreviation and data Source	83

Acknowledgments

First I want to thank my advisor, Professor Xin Guo for all her contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating. She has taught me, both consciously and un-consciously, how to be a good academic researcher. I am also thankful for the encouragement and comfort she provided during tough times in my Ph.D. pursuit. Through her, I learned there is so much more to be appreciated in life besides academic success.

I also would like to thank Professor Terrence Hendershott. As my co-advisor, Professor Hendershott has been a constant source of great ideas and inspiration. He has shown tremendous enthusiasm in my research. He is always there asking me tough questions, providing me with constructive feedback, and helping me overcome obstacles one after another. I could not have completed this dissertation without his incredible help and support.

I am grateful to have Professor Ilan Adler in my dissertation and qualifying exam committee for his insightful comments. I thank the other two members of my qualifying exam committee, Professor Shmuel Oren and Professor Craig Evans, for their interest, time and helpful guidance. I am indebted to Professor Adam Szeidl for his help on identifying research topics and preparation of my qualifying exam. I would also like to thank Professor Zuo-Jun Shen for his support and encouragement.

I gratefully acknowledge the funding sources that made my Ph.D. work possible. I was funded by IEOR departmental fellowship for my first year and received subsequent funding both from the department and the Graduate Division at Berkeley. I also thank NASDAQ OMX educational foundation for providing me an honorable one-year Ph.D. dissertation fellowship. For this dissertation, I also would like to thank Wharton Research Data Service for providing excellent data source, and Haas School of Business, Berkeley for granting me access to WRDS.

My time at Berkeley was made enjoyable in large part due to the many friends and groups that became a part of my life. I am grateful for time spent with roommates, classmates and colleagues. I am especially thankful to members of RADAR lab and fellow IEOR students for stimulating research discussions and sharing of academic/professional experience. I also thank staffs in the IEOR department for all their help that facilitates my study and research. I cannot overstate my appreciation of support and caring from brothers and sisters of the HOS fellowship and the CFC Berkeley church. My life at Berkeley was also enriched by weekly pickup soccer games.

Lastly, I would like to thank my family for all their love and encouragement. For my parents who raised me with the best they have and supported me in all my pursuits. For the presence of my lovely and brave son Jacob, that made the final stage of my Ph.D. even more colorful. And most of all for my loving, supportive, and encouraging wife Lu, without whom I would not have accomplished this. Thank you.

Chapter 1

Introduction

What is a crash? Obviously, a stock market crash is a significant drop in asset prices. Not so obviously, a crash often occurs even without major news events. During the 200 years of US stock market history, there were a few famous historical market crashes: one of them being the Oct. 1987 crash, which occurred during Oct. 14-Oct. 19/20. Within a week, the market went down about 30%. A more recent one is the May 6th 2010 Flash Crash, which occurred intraday during the afternoon. Within roughly 15 minutes, the market drawdown was about 5%.

Figure 1.1 is a graph of the Oct. 1987 Crash from the US Federal Reserve. It labels important events around the crash. Oct. 19 is the infamous *Black Monday* and we can see the Chicago Board Options Exchange (CBOE) and the Chicago Mercantile Exchange (CME) trading suspensions on Oct. 20.

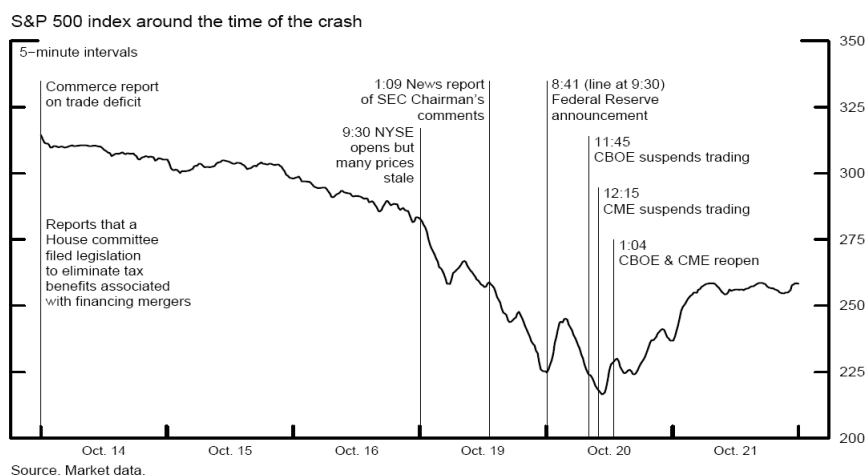


Figure 1.1: The Oct. 1987 Crash, Source: US Federal Reserve.

“On Wednesday morning, October 14, 1987, the U.S. equity market began the most severe one-week decline in its history. The Dow stood at over 2,500 on Wednesday morning. By noon on Tuesday of the next week, it was just above 1,700, a decline of almost one third. Worse still, at the same time on Tuesday, the S&P 500 futures contract would imply a Dow level near 1,400...” (U.S. Presidential Task Force on Market Mechanisms 1988)

After the October 1987 stock market crash, the “Brady Report” (U.S. Presidential Task Force on Market Mechanisms 1988) documented quantities of stock index futures contracts and baskets of stocks sold by portfolio insurers during the crash. According to the report, from Oct. 14 to Oct. 19, the net dollar volume of SP500 futures sold by portfolio insurers and arbitragers is about \$3 billion.

Figure 1.2 is a graph of 2010 Flash Crash. We can see the Dow Jones Index, S&P500 Index and the E-mini S&P500 Index all went down significantly between 2:32 pm-2:45 pm ET. At 2:45:28 pm ET, trading on the E-mini was paused for five seconds when the CME Stop Logic Functionality was triggered in order to prevent a cascade of further price declines.

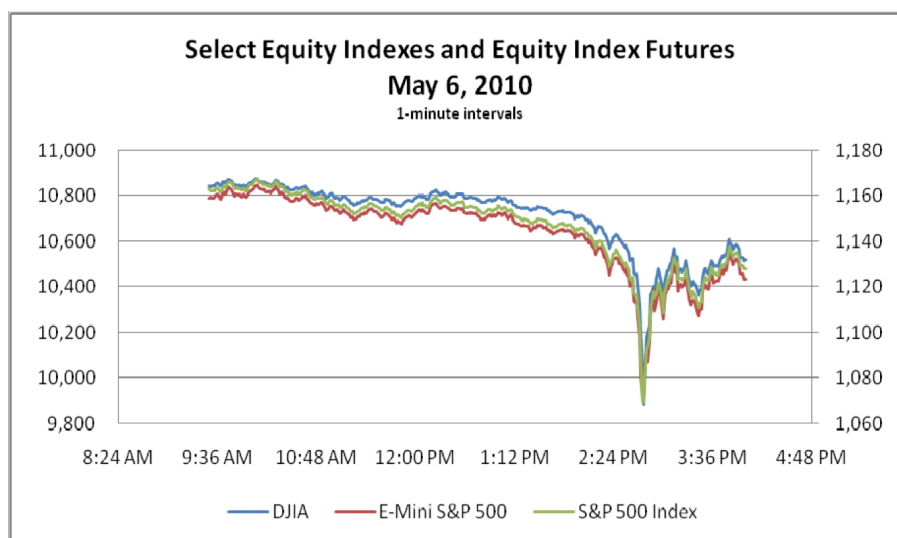


Figure 1.2: The 2010 Flash Crash, Source: Staffs of the CFTC and SEC (2010b).

“On May 6, 2010, the prices of many U.S.-based equity products experienced an extraordinarily rapid decline and recovery. That afternoon, major equity indices in both the futures and securities markets, each already down over 4% from their prior-day close, suddenly plummeted a further 5-6% in a matter of minutes before rebounding almost as quickly...” (Staffs of the CFTC and SEC 2010a)

After the flash crash of May 6, 2010, Staffs of the CFTC and SEC (2010a); Staffs of the CFTC and SEC (2010b) cited as a trigger large sales of futures contracts by one entity, identified in the press as Waddell & Reed: “At 2:32 p.m., against this backdrop of unusually high volatility and thinning liquidity, a large fundamental trader (a mutual fund complex)

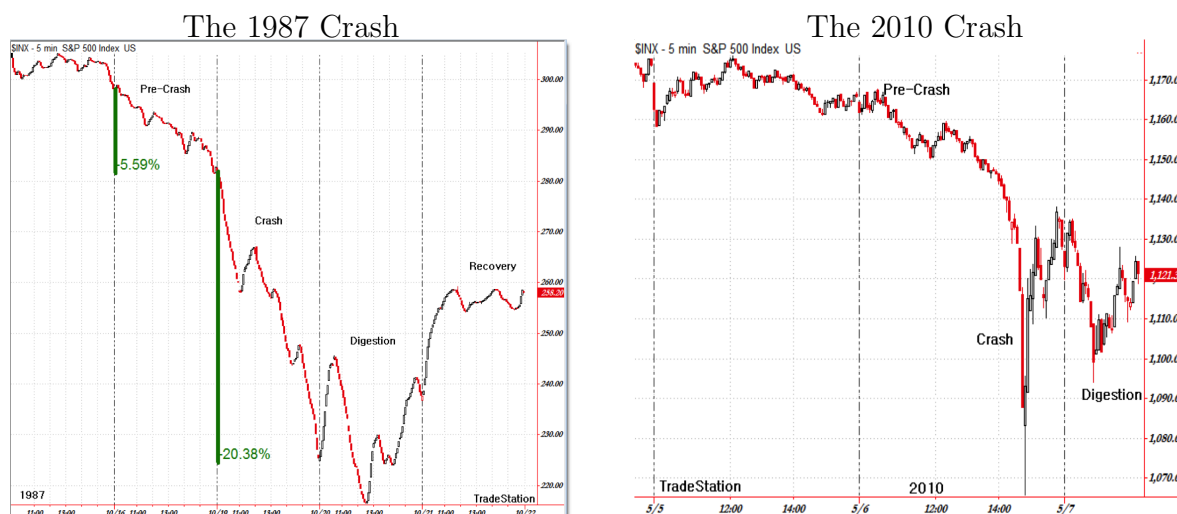


Figure 1.3: Comparison of the Oct. 1987 crash and the 2010 Flash Crash.

initiated a sell program to sell a total of 75,000 E-mini contracts (valued at approximately \$4.1 billion) as a hedge to an existing equity position...Between 2:32 p.m. and 2:45 p.m., as prices of the E-mini rapidly declined, the sell algorithm sold about 35,000 E-mini contracts (valued at approximately \$1.9 billion) of the 75,000 intended...The sell algorithm continued to execute the sell program until about 2:51 p.m. as the prices were rapidly rising in both the E-mini and SPY...” From the above statement we can estimate that the large seller sold around \$1.9-4.1 billion dollar volume of futures between 2:32 pm-2:51 pm ET, and the market declined about 5-6%.

Unlike another famous crash, the 1929 crash, which signaled the beginning of the 10-year Great Depression, during the Oct. 1987 crash and the May 6, 2010 Flash Crash, the shocks the market experienced were relatively temporary, as the price recovered in a short period of time. Leland (2011) mentions the similarities between the events of the Flash Crash and October 19, 1987 are remarkable in nature, if not in speed. Figure 1.3 demonstrates a common pattern in the two crashes. Both of the crashes can be divided into several sessions: pre-crash, crash, digestion and recovery. This motivates us to investigate and explain the links behind the two crashes.

The remaining of this dissertation is organized as follows: Chapter 2 surveys the related traditional market microstructure and crash literature, as well as some recent research findings; in Chapter 3 we build a single asset model and a multi-asset model in a strategic trading setting, and discuss the empirical implication and tests based on the models; in Chapter 4 we use several statistical analysis methods to compare two model-based predictors from Chapter 3, with other candidate predictors for the cross-sectional price drawdown of SP500 stock from other literature; Chapter 5 concludes and discusses future research opportunities.

Chapter 2

Literature

2.1 Rational Expectations Equilibrium (REE) Models

Competitive REE models are the foundation of market microstructure theories. The role of information and learning is extensively studied. The important results in this section are developed by multiple papers, among them: Grossman (1976); Grossman and Stiglitz (1980); Hellwig (1980); Admati (1985). Vives (2008) provides a good review of the existing literature.

2.1.1 The Walrasian Model

The building blocks of the modern market microstructure models are so called Rational Expectation Models. These are static competitive models. The models are formulated as follows:

Given two periods $t = 0, 1$, one risky asset pays a normal distributed dividend d in period 1, $d \sim N(\bar{d}, \sigma^2)$, and it is traded at $t = 0$ at a price p . The supply of the risky asset is random $S > 0$. The riskless rate between periods 0 and 1 is r . There are N agents in the market, with CARA (constant absolute risk aversion) utility and the risk aversion coefficient is α . Among them, N_I are informed and N_U are uninformed, where $N_I + N_U = N$. The informed observe a signal of the risky asset $s = d + \epsilon$, ϵ is independent of d and $\epsilon \sim N(0, \sigma_\epsilon^2)$. The uninformed observe no signal.

This structure of uncertainty enables us to study how the uninformed learn from prices. Consider an uninformed agent with wealth W_0 , if at $t = 0$ she buys x shares, her period 1 wealth is:

$$W = (W_0 - xp)(1 + r) + xd \tag{2.1}$$

and her expected utility is:

$$- E \exp(-\alpha W) \tag{2.2}$$

This form of utility (CARA) has the good property that if $y \sim N(\mu, \sigma^2)$, then

$$E \exp(y) = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$$

The expected utility is therefore:

$$- \exp\left[-\alpha\left((W_0 - xp)(1 + r) + x\bar{d} - \frac{\alpha}{2}\sigma^2x^2\right)\right] \quad (2.3)$$

Thus, the uninformed can be viewed as maximizing:

$$\max_x (W_0 - xp)(1 + r) + x\bar{d} - \frac{\alpha}{2}\sigma^2x^2 \quad (2.4)$$

This is a concave function and by solving F.O.C, we have the uninformed demand:

$$x_U(p) = \frac{\bar{d} - (1 + r)p}{\alpha\sigma^2} \quad (2.5)$$

Here the price of the risky asset enters through budget constraint and the initial wealth is not in this demand.

The informed face a conditional probability, suppose (x, y) are jointly normal, then conditional on x , y is normal and

$$E(y|x) = E(y) + (x - E(x))\frac{\text{cov}(x, y)}{\sigma_x^2} \quad (2.6)$$

$$\sigma_{y|x}^2 = \sigma_y^2 - \frac{\text{cov}(x, y)^2}{\sigma_x^2} \quad (2.7)$$

So in this model, we have

$$E(d|s) = \bar{d} + (s - \bar{d})\frac{\sigma^2}{\sigma^2 + \sigma_\epsilon^2} \quad (2.8)$$

$$\sigma_{d|s}^2 = \frac{\sigma^2\sigma_\epsilon^2}{\sigma^2 + \sigma_\epsilon^2} \quad (2.9)$$

Hence, for the informed, the expected utility is:

$$- \exp\left[-\alpha\left((W_0 - xp)(1 + r) + xE(d|s) - \frac{\alpha}{2}\sigma_{d|s}^2x^2\right)\right] \quad (2.10)$$

and therefore

$$x_I(p) = \frac{E(d|s) - (1 + r)p}{\alpha\sigma_{d|s}^2} \quad (2.11)$$

By equating demand with supply, $N_I x_I(p) + N_U x_U(p) = S$, we solve for price:

$$p = \frac{\bar{d}}{1 + r} + N_I k \left(1 - \frac{\sigma_{d|s}^2}{\sigma^2}\right) (s - \bar{d}) - S \alpha k \sigma_{d|s}^2 \quad (2.12)$$

$$k = \frac{1}{(1 + r)(N_I + N_U \frac{\sigma_{d|s}^2}{\sigma^2})} \quad (2.13)$$

In this model, the price fully reveals the signal of the informed (price is linear in signal). For the uninformed, they could use price to infer informed signal.

2.1.2 Extension

The next model is an extension that let people actually learn from prices. There are N agents, the i th agent has expected utility U^i , W_0^i and signal s^i . A REE is a price function $p(s^1, \dots, s^N)$, and a vector of demands, $(x^1(p), \dots, x^N(p))$ such that

$$x^i(p) = \arg \max_x E\{U^i[(W_0^i - xp)(1 + r) + xd] | s_i, p\} \quad (2.14)$$

and

$$\sum_{i=1}^N x^i(p) = S \quad (2.15)$$

In this model, agents condition on their private information and the price, not just the private information as in the previous case. Price plays a dual role: it affects budget constraint and also improves inferences about others' signals (realizing that assets are not private values but rather common values).

It turns out a REE is given by:

$$p = \frac{\bar{d} + \beta_s(s - \bar{d}) - \frac{S\alpha\sigma_{d|s}^2}{N}}{(1 + r)} \quad (2.16)$$

$$x_I(p) = \frac{\bar{d} + \beta_s(s - \bar{d}) - (1 + r)p}{\alpha\sigma_{d|s}^2} \quad (2.17)$$

$$x_U(p) = \frac{S}{N} \quad (2.18)$$

where $\beta_s = \frac{\text{cov}(s, d)}{\text{var}(s)} = \frac{\sigma^2}{\sigma^2 + \sigma_\epsilon^2}$.

Here is a brief proof, assume Equation 2.16 holds. Since the uninformed can learn from price, the optimal uninformed demand is:

$$x_U(p) = \frac{E(d|p) - (1 + r)p}{\alpha\sigma_{d|p}^2} \quad (2.19)$$

$$= \frac{\bar{d} + \beta_p(p - \bar{p}) - (1 + r)p}{\alpha\sigma_{d|p}^2} \quad (2.20)$$

$$= \frac{\bar{d} + (1 + r)(p - \bar{p}) - (1 + r)p}{\alpha\sigma_{d|s}^2} \quad (2.21)$$

$$= \frac{S}{N} \quad (2.22)$$

as

$$\begin{aligned}\beta_p &= \frac{\text{cov}(p, d)}{\text{var}(p)} \\ &= \frac{1+r}{\beta_s} \frac{\text{cov}(s, d)}{\text{var}(s)} \\ &= 1+r\end{aligned}\tag{2.23}$$

$$\bar{p} = \frac{\bar{d}}{1+r} - \frac{S\alpha\sigma_{d|s}^2}{(1+r)N}\tag{2.24}$$

$$\begin{aligned}\sigma_{d|p}^2 &= \text{var}(d) - \beta_p^2 \text{var}(p) \\ &= \text{var}(d) - (1+r)^2 \left(\frac{\beta_s}{(1+r)}\right)^2 \text{var}(s) \\ &= \text{var}(d) - \beta_s^2 \text{var}(s) \\ &= \sigma_{d|s}^2\end{aligned}\tag{2.25}$$

The optimal informed demand is the same as previous model, as signal tells informed everything they need to know.

$$x_I(p) = \frac{\text{E}(d|s) - (1+r)p}{\alpha\sigma_{d|s}^2}\tag{2.26}$$

$$= \frac{\bar{d} + \beta_s(s - \bar{d}) - (1+r)p}{\alpha\sigma_{d|s}^2}\tag{2.27}$$

Finally, plug the assumed price Equation (2.16) in Equation (2.26) we have $x_I(p) = \frac{S}{N}$ and therefore the market clearing condition Equation (2.15) holds.

Note this is one possible REE but might not be the unique one. In this REE, price function fully reveals the information of the informed. The price is the same as it would be in a Walrasian economy where all agents are informed. Supply allocations are the same across agents. This is related to the Grossman-Stiglitz paradox: if agents have to pay for information, why bother if all information gets into price anyway?

2.1.3 Noisy (Supply) REE

Now we give the supply of the risky asset a random noise: $S + u$ and $u \sim N(0, \sigma_u^2)$. This can be interpreted as the supply from noise traders. Now, the uninformed will make lower profits than the informed traders. This provides incentive to acquire information.

A noisy REE is a price function $p(s^1, \dots, s^N, u)$ and a vector of demands $(x^1(p), \dots, x^N(p))$ such that :

$$x^i(p) = \arg \max_x \text{E}\{U^i[(W_0^i - xp)(1+r) + xd] | s_i, p\}\tag{2.28}$$

and

$$\sum_{i=1}^N x^i(p) = S + u \quad (2.29)$$

Assume $N_I > 0$, there exists a noisy supply REE in which the price is given by $p = A + B[(s - \bar{d}) - Cu]$, for three constants A, B, C . Again, by using conditional distributions and the market clearing condition, we can show that:

$$A = \frac{\bar{d}}{1+r} - \frac{S\alpha\sigma_{d|s}^2}{(1+r)(N_I + N_U\frac{\sigma_{d|s}^2}{\sigma_{d|p}^2})} \quad (2.30)$$

$$B = \frac{N_I\beta_s}{(1+r)N_I + (1+r - \beta_p)N_U\frac{\sigma_{d|s}^2}{\sigma_{d|p}^2}} \quad (2.31)$$

$$C = \frac{\alpha\sigma_{d|s}^2}{N_I\beta_s} = \frac{\alpha\sigma_\epsilon^2}{N_I} \quad (2.32)$$

The optimal uninformed demand is:

$$x_U(p) = \frac{E(d|p) - (1+r)p}{\alpha\sigma_{d|p}^2} \quad (2.33)$$

$$= \frac{\bar{d} + \beta_p(p - A) - (1+r)p}{\alpha\sigma_{d|p}^2} \quad (2.34)$$

The optimal informed demand is the same as previous model:

$$x_I(p) = \frac{E(d|s) - (1+r)p}{\alpha\sigma_{d|s}^2} \quad (2.35)$$

$$= \frac{\bar{d} + \beta_s(s - \bar{d}) - (1+r)p}{\alpha\sigma_{d|s}^2} \quad (2.36)$$

Price informativeness is $\tau = \frac{1}{\sigma_{d|p}^2} = \frac{\sigma^2 + \sigma_\epsilon^2 + C^2\sigma_u^2}{\sigma^2(\sigma_\epsilon^2 + C^2\sigma_u^2)}$. It decreases in the supply noise σ_u^2 and in C .

The price-sensitivity of the demand of the uninformed is:

$$\frac{dx_U(p)}{dp} = \frac{\beta_p - (1+r)}{\alpha\sigma_{d|p}^2} \quad (2.37)$$

So, decrease in price increases demand (keeping information constant) and decrease in price provides negative news which decreases demand. The second effect is weaker because price is not fully informative.

2.2 Classic Market Microstructure Models

Market microstructure studies how, in the short term, the transaction price converges to (or deviates from) the long-term equilibrium values. Short-term deviations between transaction prices and long-term fundamental values arise because of frictions reflecting order-handling costs, as well as asymmetric information or strategic behavior (Biais, Glosten, and Spatt 2005).

2.2.1 Adverse Selection

Kyle (1985) proposes a sequential equilibrium model. Here we review the essence: there are three agents in the market, a risk neutral (monopolistic) informed trader, who knows the distribution of the payoff of the risky asset $\tilde{v} \sim N(p_0, \Sigma_0)$; a noisy trader, who submits random demand/supply $\tilde{u} \sim N(0, \sigma_u^2)$; and risk neutral (competitive) market makers, who set the price according to the order flow, but earn zero profits (indicating that the market is efficient). That is, let \tilde{x} be the order flow from the informed trader, then market makers set price $\tilde{p} = E[\tilde{v}|\tilde{x} + \tilde{u}] = P(\tilde{x} + \tilde{u})$.

Now, the informed trader wants to solve the maximization problem

$$\max_x E[(v - P(x + \tilde{u}))x|\tilde{v} = v] \quad (2.38)$$

subject to

$$P(x + \tilde{u}) = E[\tilde{v}|x + \tilde{u}] \quad (2.39)$$

It turns out that there is an equilibrium in the linear form:

$$\tilde{p} = P(\tilde{x} + \tilde{u}) = p_0 + \lambda(\tilde{x} + \tilde{u}) \quad (2.40)$$

$$\tilde{x} = X(\tilde{v}) = -\frac{p_0}{2\lambda} + \frac{\tilde{v}}{2\lambda} \quad (2.41)$$

where

$$\lambda = \frac{\sqrt{\Sigma_0}}{2\sigma_u} \quad (2.42)$$

$\text{var}[\tilde{v}|\tilde{p}] = \frac{1}{2}\Sigma_0$ so half the price information is revealed in price. λ is the amount market makers adjust price if the order flow increase by 1 unit. Alternatively, $\frac{1}{\lambda}$ measures the market depth, i.e. the amount of order flow needed to increase price by \$1. All trades have impact: informed trades have permanent price impact; uninformed trades have temporary price impact. A big advantage of this parsimonious model is that it provides a testable linear price impact scheme.

Glosten and Milgrom (1985) provide another model explaining how information gets impounded into price. Unlike papers in the next section, Glosten and Milgrom (1985) show that bid-ask spread can arise purely from adverse selection. As in Kyle (1985), price is set by risk neutral market makers who earn zero profits, and the model only considers market

orders. By restricting that all types of market participants can trade only one unit of security at a time, their model does not need to assume a specific form of the informed trader's information, and can have few restrictions on the arrival process of traders.

The model is formulated as follows: V is the unit value of a security at liquidation, which satisfies $[V \geq 0, \text{var}(V) < \infty]$, let H_t be the public information available up to time t , J_t be the private information. So the valuations Z_t of the asset for informed and uninformed are:

$$Z_t^I = \rho_t \mathbb{E}[V | H_t, J_t, A, B] \quad (2.43)$$

$$Z_t^U = \rho_t \mathbb{E}[V | H_t, A, B] \quad (2.44)$$

Here ρ_t denotes a trader's preference of future consumption with respect to current consumption. It is a random variable independent of V and any information about V . The informed or uninformed will buy if $Z_t > A$ and sell if $Z_t < B$.

Given the trader's behavior, and the information available to the market maker at time t : S_t , the market maker's expected profit from an arrival of trade at time t is:

$$\mathbb{E}[(A - V)I_{Z_t > A} + (V - B)I_{Z_t < B} | S_t] \quad (2.45)$$

Therefore under the expected zero-profit condition, the equilibrium bid and ask prices are given as follows:

$$A_t = \mathbb{E}[V | Z_t > A_t] \quad (2.46)$$

$$B_t = \mathbb{E}[V | Z_t < B_t] \quad (2.47)$$

Glosten and Milgrom (1985) show that the ask price is greater and the bid price is less than the expectation of V , i.e. $A_t \geq \mathbb{E}_t[V] \geq B_t$. The inequality is strict if adverse selection is possible. Anything that increases adverse selection would increase spread: i.e. when the informed private information becomes better; the ratio of informed to uninformed arrival rates is increased; the elasticity of uninformed supply and demand increases.

They also show in the model that the first difference of transaction price process is serially uncorrelated. Thus the spreads due to monopoly power, transaction costs and risk aversion lead to negative serial correlation, while spreads solely due to adverse selection do not.

Easley and O'Hara (1987) develop a model demonstrating that trade size also can affect trade price, not because of market makers' inventory imbalance, but because block trades are correlated with adverse selection.

In their model, there are two risk neutral market makers that set the price and competition ensures they each have zero expected profit. An information event about the asset V occurs before the trading day with probability α ($0 < \alpha \leq 1$). The event contains a signal that the value of the asset is either H (with probability $1 - \delta$) or L (with probability δ). According to the signal, we have:

$$\bar{V} = \mathbb{E}[V | s = H] \quad (2.48)$$

$$\underline{V} = \mathbb{E}[V | s = L] \quad (2.49)$$

Assume μ is the fraction of trades made by the informed risk neutral traders. The market makers and the uninformed traders do not know whether an information event has occurred, nor do they observe the signal. But they do know the information event will eventually occur and the information structure. So their initial unconditional expectation of the asset value is $V^* = \delta \underline{V} + (1 - \delta) \bar{V}$. They further assume the uninformed traders desire to buy (sell) $B^1(S^1), B^2(S^2)$, with $0 < B^1 < B^2$ and $0 < S^1 < S^2$. The fraction of uninformed traders for each trade quantity is $X_S^i > 0$ and $X_B^i > 0$, $i = 1, 2$. For market maker j , she charges $c^j(q)$ for q units and pays for $d^j(q)$ for q units. The two market makers play a simultaneous move game against each other.

There are two forms of equilibria that can occur: a separating equilibrium if informed traders trade only large quantities; a pooling equilibrium if the informed trade either small or large quantities with positive probability.

This simple model explains why different trade quantities face different prices, where the only friction is adverse selection.

2.2.2 Inventory and Order Handling Cost

While adverse selection contributes to the permanent impact on the transaction price, order handling and inventory cost contributes to the transitory impact on price (Biais, Glosten, and Spatt 2005).

Here we review three papers that describe how inventory and order handling cost affect price.

Roll (1984) assumes in an information efficient market, the fundamental value of a security fluctuates randomly. However, trading costs can induce negative serial dependence in successive observed market price changes.

$$\text{cov}(\Delta p_t, \Delta p_{t-1}) = -\frac{s^2}{4} \quad (2.50)$$

where s is the effective spread.

Ho and Stoll (1981) examine the bid ask dynamics set by a single market maker who maximizes her expected utility of terminal wealth. The demand the market maker faces is modeled by a continuous time Poisson jump process as in Garman (1976). And she also faces the return risk on her stock and on the rest of her portfolio (which is modeled by diffusion processes). Numerical solutions can be derived from dynamic programming and an interesting result is that in an inactive stock, it is possible for the market maker to refuse to make the market when she is required to trade a minimum amount, because the expected profit from trading may not be enough to offset the risk.

Amihud and Mendelson (1980) build a similar model where a single market maker sets the bid ask prices to maximize her expected average profit per unit-time. The arrivals of buy and sell orders from liquidity traders (uninformed) are characterized by two independent Poisson processes, with arrival rates $D(P_a)$ and $S(P_b)$. D and S are stationary price-dependent rate functions representing the market demand and supply, with $D'(\cdot) < 0$, $S'(\cdot) > 0$.

Under the setting of their model, it can be shown that there exists a ‘preferred’ inventory position for the market maker and that when the market maker finds herself in a position different from the ‘preferred’ position, she will quote prices which will tend to bring her back to that position. Also, they show that for linear demand and supply functions, the transaction prices are serially correlated and that this is still consistent with the market efficiency hypothesis.

2.2.3 A Synthetic Model

Biais, Glosten, and Spatt (2005) propose an interesting synthetic model that incorporates both the adverse selection and inventory/order handling cost. In the model, there are risk averse informed traders and risk averse market makers (no uninformed traders for simplicity). The informed trader is endowed with L shares of the risky asset and has observed a signal s on the final value of the risky asset v . The market maker is endowed with I shares of the risky asset. Assume $v = \pi + s + \epsilon$, and $s \sim N(0, \sigma_s^2)$, $\epsilon \sim N(0, \sigma_\epsilon^2)$. Also, assume market makers incur an identical cost cq^2 to trade q shares.

The informed trader wants to submit a market order that maximizes her expected utility:

$$\max_Q -E\exp[-\gamma(P(Q)Q + (L - Q)v)|s] \quad (2.51)$$

subject to

$$P(Q) = \arg \max_P -E\exp[-\kappa((v - P)Q + Iv)|Q] \quad (2.52)$$

It turns out the information revealed by the market order is equivalent to that contained by the summary statistics: $\theta = s - \gamma^2\sigma_\epsilon^2L$. θ reflects the valuation of the strategic informed trader for the asset, which is increasing in her private signal, and decreasing in her inventory. Denote:

$$\delta = \frac{\sigma_s^2}{\sigma_s^2 + (\gamma\sigma_\epsilon^2)^2V(L)} \quad (2.53)$$

δ quantifies the relative weight of the noise and signal in the summary statistic θ . It also measures the magnitude of the adverse-selection problem. For example, $\delta = 0$ corresponds to the case in which there is no private information.

As in Kyle (1985), if $\delta < \frac{1}{2}$, there exists a perfect Bayesian equilibrium that:

$$E[v|Q] = [\delta m + (1 - \delta)\pi] + \delta(2\lambda - \gamma\sigma_\epsilon^2)Q \quad (2.54)$$

$$Q = \frac{(\pi - m) + \theta}{2\lambda - \gamma\sigma_\epsilon^2} \quad (2.55)$$

$$P = m + \lambda Q \quad (2.56)$$

$$m = \pi - \frac{\kappa V(v|\theta)}{1 - \delta} I \quad (2.57)$$

$$\lambda = \frac{c + \kappa V(v|\theta) + \gamma\sigma_\epsilon^2\delta}{1 - 2\delta} \quad (2.58)$$

When $\delta = 0$, i.e., there is no private information, the result is similar to Roll (1984), Ho and Stoll (1981). Symmetrically, in the case where market makers are risk neutral ($\kappa = 0$), and there is no order handling cost ($c = 0$), we can obtain a specification similar to Kyle (1985), where prices are equal to updated expectations of the value of the asset, conditional on the order flow.

2.3 Crashes

After reviewing literature about how information and liquidity cost get impounded into price in a general setting, now we move on to check the price formation process in extreme events (e.g. stock market crashes).

Brunnermeier (2001) reviews a broad set of literature related to *Crashes*, here we adapt his review of two categories of crash models that are related to the previous sections.

2.3.1 Crashes in Competitive REE Models

In a competitive REE model, many traders simultaneously submit orders. They take prices as given and can trade any quantity of shares in each trading round. In this setting, crashes can occur because of temporary liquidity shortage, multiple equilibria due to portfolio insurance trading and sudden information revelation by prices.

Grossman (1988) shows a REE model where poor information about hedging demand leads to a large price decline. The equilibrium price path and the volatility of a risky asset are driven by news announcements about its liquidation value as well as by investors' risk aversion.

There are three periods, $t = 1, 2, 3$. There are public announcements about the value of the stock in period $t = 2, 3$. After the second announcement in $t = 3$, every investor knows the final liquidation value of the stock. In this model, large price movements can occur when the market makers underestimate the extent of sales due to portfolio insurance trading and reduce liquidity provision. Grossman's model also predicts that the price would rebound immediately after the temporary liquidity shortage is overcome.

Genotte and Leland (1990) show that the market crashes because some other market participants incorrectly interpret the price drop as a bad signal about the fundamental value of the stock. Because of adverse selection, traders wrongly attribute the price drop to a low fundamental value rather than to liquidity shortage. They might think that many other traders are selling because they received bad information about the fundamental value of the stock, while actually many sell orders are triggered by portfolio insurance trading.

They develop a static model to describe the market crashes' dynamics. This approach has some validity because the repetition of a static model can often be considered as a sufficient representation of a dynamic setting. The comparative statics in a static model can be viewed as dynamic changes over time.

In their model, there are uninformed traders and two types of informed traders:

- price-informed traders, each of whom receives a private signal $p'_i = p + \epsilon^i$ about the liquidation value $p \sim N(\bar{p}, \Sigma)$;
- supply informed traders, who know better whether the limit order book is due to informed trading or uninformed noise trading.

The aggregate supply in the limit order book is given by the normally distributed random variable $m = \bar{m} + L + S + \pi$. \bar{m} is known to everybody, $S \sim N(0, \Sigma_s)$ is only known to the supply-informed traders, and the liquidity supply $L \sim N(0, \Sigma_L)$ is not known to anybody. There is an exogenous demand from portfolio insurance traders. Their demand $\pi(p_0)$ rises as the price increases and declines as price falls.

It turns out that as long as $\pi(p_0)$ is linear and common knowledge, the equilibrium price $p_0 = f(p - \bar{p} - k_1 L - k_2 S)$ is a linear function with constants k_1 and k_2 . However, if $\pi(p_0)$ is nonlinear and there is a lack of knowledge of the amount of $\pi(p_0)$, the demand curve is like an “inverted S ” that there are multiple equilibria for a certain range of aggregate supply. Thus as aggregate supply shifts, the equilibrium with the high price vanishes and the price discontinuously falls to a lower equilibrium level.

Similar to Gennotte and Leland (1990), Barlevy and Veronesi (2003) build a model with multiple equilibria to explain stock market crashes. The difference is that in their model, multiple equilibria is not from exogenous hedging, but from information asymmetry effect that makes uninformed traders’ demand function to be backwards-bending.

2.3.2 Crashes in Sequential Trade Models

Sequential trade models are more tractable and thus enable us to focus on the dynamic aspects of crashes.

Avery and Zemsky (1998) illustrate a sequential trade model where a fraction μ of traders are informed while $1 - \mu$ are uninformed liquidity traders. Liquidity traders buy, sell or stay inactive with equal probability. They show when each informed trader receives a noisy individual signal about the value of the stock $\nu \in \{0, 1\}$ and the signal is correct with probability $q > \frac{1}{2}$, in other words, when traders have private information on only one dimension of uncertainty (the effect of a shock to the asset value), price adjustments prevent herd behavior. This is because the market maker and the insiders learn at the same rate from past trading rounds.

Then Avery and Zemsky (1998) study an information structure with higher order of uncertainty. Informed traders receive either a perfect signal that no new information has arrived, that is, the value of stock remains $\nu = \frac{1}{2}$, or a noisy signal with reports the correct liquidation value $\nu \in \{0, 1\}$ with probability q . The market maker does not know whether an information event occurred or not. This asymmetry enables insiders to learn more from the price process (trading sequence) than the market maker. Since market maker sets the price, the price adjustment is slower. Consequently, traders might herd in equilibrium. However, herding in this setting increases the market makers’ awareness of information events and does not distort the asset price.

In order to simulate crashes, Avery and Zemsky (1998)'s model considers a more complex information structure, in their words, a third dimension of uncertainty (the quality of traders' information) in addition to the existence and effect of a shock. There are two types of informed traders. One group of them receives their signals with low precision q_L , whereas the other receives them with high precision $q_H = 1$, that is, the signal is perfect. The proportion of informed traders with the perfect signal is either high or low and it is not known to the market maker. This information structure makes it difficult for the market maker to differentiate between a market composed of well-informed traders following their perfect signal from one with poorly informed traders who herd. When the low probability event that poorly informed traders generate a chain of buy orders makes the market maker think that a large fraction of traders is perfectly informed, she increases the price. If the unlikely event occurs in with only poorly informed traders herd, the asset price may exceed its liquidation value ν . The market maker can infer only after many trading rounds that the uninformed traders have herded. In that case, the asset price crashes.

2.3.3 Physical Models of Crashes

Besides financial economists, physicists have done some research in market crashes too.

Feigenbaum and Freund (1996) and Johansen, Sornette, and Ledoit (1999) use discrete scale invariance to predict crashes. They find characteristic log-periodic signatures of growing bubbles in a variety of markets can predict subsequent market crashes.

Bouchaud and Cont (1998) propose a non-linear Langevin equation as a model for stock market fluctuations and crashes. The model leads to a specific shape of the falldown of the price during a crash, which they compare with the October 1987 data.

2.3.4 Recent Research

Huang and Wang (2009) develop an equilibrium model for stock market liquidity and its impact on asset prices when constant market presence is costly. They show costly market presence prevents traders from synchronizing their trades and hence gives rise to endogenous order imbalances and the need for liquidity. Moreover, the endogenous liquidity need is characterized by excessive selling of significant magnitudes and such liquidity-driven selling leads to market crashes in the absence of any aggregate shocks.

Kyle and Obizhaeva (2013b) propose a *Market Microstructure Invariance* that the following relation holds for all stocks at all times:

$$\gamma^{-1/2} \bar{\sigma} \cdot P \tilde{Q} \sim \tilde{I} \quad (2.59)$$

where γ is the bet arrival rate, $\bar{\sigma}$ is the proportional standard deviation of returns results from order flow imbalances, P is the price of the stock, and \tilde{Q} is the probability distribution of the signed size of bets of the stock. \tilde{I} denotes a random variable representing the "invariance" distribution.

With the above invariance law, Kyle and Obizhaeva (2013a) use portfolio transition order data to extrapolate a set of benchmark parameters, and then use those parameters to examine historical crashes.

Kyle and Obizhaeva (2013a) propose a simple formula to predict the expected percentage price impact by input of observed daily dollar trading volume V , daily return volatility σ and the size of an individual large bet X .

$$\frac{\Delta P(X)}{P} = \exp[\bar{\lambda}/10^4 \cdot (\frac{P \cdot V}{40 \cdot 10^6})^{1/3} \cdot (\frac{\sigma}{0.02})^{4/3} \cdot \frac{X}{0.01V}] - 1 \quad (2.60)$$

Their work addresses the issue that volatility and the arrival rate of a bet affect the price formation process. However, Kyle and Obizhaeva (2013a) point out several implementation issues when apply the invariance to real data. Among them, it is likely that the price impact of an order, especially its transitory component, is related to the speed with which the order is executed. The market impact Equation (2.60) assumes that orders are executed at a typical speed in the relevant units of business time. If execution is speeded up relative to typical speed in business time, then Equation (2.60) may underestimate the transitory market impact.

Madhavan (2012) shows that cross-sectional impact of the May 2010 Flash Crash is positively related to the degree of market fragmentation. The paper highlights the role of equity market structure and the changing nature of liquidity provision in exacerbating the impact of an external liquidity shock, without taking a view as to its catalyst.

Easley, Prado, and O'Hara (2011) argue that the flash crash was the result of the new dynamics at play in the current market structure. They highlight the role played by order toxicity in affecting liquidity provision, and they show that a measure of this toxicity, the volume synchronized probability of informed trading (VPIN), captures the increasing toxicity of the order flow in the hours and days prior to collapse. They subsequently develop VPIN metric in Easley, Prado, and O'Hara (2012). However, there is argument about the usefulness of this measure for predicting market turbulence (Easley, Prado, and O'Hara 2014; Andersen and Bondarenko 2014b; Andersen and Bondarenko 2014a; Andersen and Bondarenko 2015).

Nagel (2012) shows returns of short-term reversal strategies in equity markets can be interpreted as a proxy for the returns from liquidity provision, and that the return of liquidity provision can be predicted by VIX. He shows withdrawal of liquidity supply and an associated increase in the expected returns from liquidity provision, as a main driver behind the evaporation of liquidity during times of financial market turmoil.

Menkveld and Yueshen (2015) provide an alternative explanation of the flash crash. They argue that it was the broken of cross-market arbitrage (the link that connects an E-mini seller to SP500 buyers), that made the large fundamental seller overpaid for immediacy.

There are also some ad-hoc attempt to predict the Flash Crash, for example, Barany et al. (2012) and Aldridge (2014) try to use the sequence and duration of trades from high frequency data to detect market crashes.

Chapter 3

Strategic Trading Models and Empirical Test

3.1 A Single Asset Model and Empirical Test

In this section, we build a single asset strategic trading model to link two major US stock market crashes: the 1987 crash and the 2010 Flash Crash. We provide empirical evidence to verify our model hypothesis and evaluate price impact due to the information effect and the limited risk-bearing capacity effect.

3.1.1 Model

Consider an asset market with a single risky asset in zero net supply, a riskless asset in perfect elastic supply at zero interest rate. There are three groups of risk neutral market participants: informed traders, market makers and liquidity traders.

The value of the risky asset at the end of period 1 is:

$$\tilde{v}_1 = v_0 + \tilde{s} + \beta \tilde{s}_m \quad (3.1)$$

which is paid as a terminal dividend. Here \tilde{s} is an idiosyncratic private information component known only to the informed trader; \tilde{s}_m is a market-wide private information component known only to the informed trader; and β is a constant to which the asset is affected by the market-wide private information.¹

The order flow imbalance for the risky asset observed by market makers is:

$$\tilde{x} = \tilde{y} + \tilde{z} + \gamma \tilde{z}_m \quad (3.2)$$

where \tilde{y} is the demand from the informed trader; \tilde{z} is the demand from idiosyncratic liquidity traders; \tilde{z}_m is a market-wide liquidity shock (from market-wide liquidity traders) and γ is a

¹As in Kyle (1985), we use $\tilde{\alpha}$ to denote a random variable α .

constant to which the demand of the risky asset is affected by the market-wide liquidity shock. Both \tilde{z} and \tilde{z}_m are exogenous and we assume $\tilde{s} \sim N(0, \sigma_s^2)$, $\tilde{s}_m \sim N(0, \sigma_{s_m}^2)$, $\tilde{z} \sim N(0, \sigma_z^2)$, $\tilde{z}_m \sim N(0, \sigma_{z_m}^2)$, and they are independent of each other.

In equilibrium, we have the following linear solution (See proof in Section 3.4):

$$\begin{aligned}
 P &= v_0 + \lambda \tilde{x} \\
 \tilde{x} &= \tilde{y} + \tilde{z} + \gamma \tilde{z}_m = \underbrace{\frac{1}{2\lambda} \tilde{s} + \frac{\beta}{2\lambda} \tilde{s}_m}_{\tilde{y}} + \tilde{z} + \gamma \tilde{z}_m \\
 \lambda &= \frac{\sqrt{\sigma_s^2 + \beta^2 \sigma_{s_m}^2}}{2\sqrt{\sigma_z^2 + \gamma^2 \sigma_{z_m}^2}} \tag{3.3}
 \end{aligned}$$

where λ is the price impact factor. If $\sigma_{s_m}^2 = 0$ and $\sigma_{z_m}^2 = 0$, that is, if we assume market-wide effects do not have impact on the individual risky asset, we get the same result as in Kyle (1985).

3.1.2 Empirical Test of the Model

Based on our model, the empirical test consists of two steps.

Step 1: Estimate the ex-ante price impact factor

We run the following OLS regression to get realized λ from historical data (prior to crash) for the SP500 stocks:

$$\Delta P_{it} = \lambda \Delta X_{it} + \epsilon_{it} \tag{3.4}$$

where ΔP_{it} is the t-interval return of the stock i and ΔX_{it} is the t-interval buy-sell dollar volume imbalance (Hasbrouck 2007).

Step 2: Cross-sectional regression to estimate the market-wide liquidity shock during crashes

We run a cross-sectional OLS regression as follows:

$$\Delta P_i = \lambda_i \Delta X_i + \epsilon_i \tag{3.5}$$

$$= \lambda_i \gamma_i \Delta z_m + \left(\frac{1}{2} \Delta s_i + \frac{\beta_i}{2} \Delta s_m + \lambda_i \Delta z_i + \epsilon_i \right) \tag{3.6}$$

$$= \bar{\beta} \Delta s_m + \lambda_i \gamma_i \Delta z_m + u_i \tag{3.7}$$

That is:

$$\Delta P_i = \alpha + \theta \lambda_i \gamma_i + u_i \tag{3.8}$$

As presumably for stock i , $\Delta X_i \approx \gamma_i \Delta z_m$ during crashes, if data fits the model, we should have $\alpha \approx$ market-wide price change and $\theta \approx$ market-wide liquidity shock. Here we assume γ_i

equals the ratio of the stock i 's market capital to the SP500 index capital at the beginning of the crash.²

3.1.3 Empirical Implementation and Data

All the data we use are from Wharton Research Data Services (WRDS). We use intraday trades and quotes data from the Institute for the Study of Security Markets (ISSM) database to estimate the price impact factor λ_i and to calculate the price drawdown ΔP_i for the 1987 crash. For the 2010 crash, we use intraday trades and quotes data from the NYSE Trade and Quote (TAQ) database to estimate the price impact factor λ_i and to calculate the price drawdown ΔP_i during the crash. Daily SP500 index/constituent and Open/Close information from the Center for Research in Security Prices (CRSP) database are used to calculate the γ_i and the price drawdown ΔP_i during the 1987 crash.

In the data preprocessing step, for the 1987 crash, we only include SP500 stocks that were listed on NYSE/AMEX, as for the ISSM data, intraday data quality for NASDAQ symbols is compromised. For the NYSE/AMEX stocks, we also filter out their intraday trades and quotes that were not originated from NYSE/AMEX, as to exclude irregularity of trades/quotes originated from regional exchanges.

We follow the 5 seconds rule of Lee and Ready (1991) and the 1 seconds rule of Henker and Wang (2006) to label the direction of each trade for the 1987 crash and the 2010 crash respectively.

In Step 1 regression, we use a 30-day horizon prior to crash for the 1987 crash and a 5-day horizon prior to crash for the 2010 crash. The choice is based on a balance between estimating the most recent realization of price impact factor (so that we cannot infer on data that is too far away in the history) and retaining enough data to get statistically significant inference.

For the 2010 Flash Crash, we use a similar intraday time window 14:30-15:00 to control the intraday seasonality of price impact factor.

To eliminate the noise that bid-ask spreads can bring to the estimation of price impact factor in Step 1, we use mid-quote price for both crashes.

For both crashes, from Step 1 to Step 2 cross-sectional regression, we filter out stocks that we cannot get a statistically significant (p-value > 0.001) price impact factor.

In Step 2 regression, a crucial issue is to determine the maximum price drawdown during crashes. The first step is to determine the crash window. For the 1987 crash, we do not count in Oct. 20th as the trading activity on that day was very unusual (Leland and Rubinstein 1988; Gammill and Marsh 1988). The second step is to find the minimum price during the crash. Here we need to filter out prices that are due to broken trades. So instead of naively using the minimum price during crashes, we use *stable price minimums* (if there were trades below that price for at least 300 seconds and 30 seconds respectively for the 1987 crash and

²This assumption is discussed in Section 3.3.

the 2010 crash). As the price drawdown is more significant for 1987 crash, it is calculated in percent price change instead of log price difference.

For the 2010 Flash Crash, we filter out stocks with maximum drawdown of more than 15%, as that can be categorized as broken trades. For 1987 crash, we filter out stocks whose ISSM prices do not match CRSP price and also stocks with less than 30 trades on Oct. 19th.

Table 3.1 summarizes the data handling details:

		The Oct. 1987 Crash	The 2010 Flash Crash
Preprocess	Symbols Trade/Quote filter Label buy-sell trades	NYSE/AMEX originated from NYSE/AMEX 5 sec (Lee and Ready 1991)	1 sec (Henker and Wang 2006)
Step 1	regression horizon intraday window regression ΔP λ filter	30 days prior 9:30-16:00 ET tick by tick log-midquote difference p-value ≤ 0.001	5 days prior 14:30-15:00 ET 15 sec log-midquote difference p-value ≤ 0.001
Step 2	Crash window ¹ max drawdown ² market capital ΔP calculation ΔP filter ³	Oct. 14-Oct. 19 trade below for at least 300 sec Close Oct. 13, 1987 percent price change from Oct. 13 close price	14:30-15:00 ET trade below for at least 30 sec Open May 6, 2010 log price change from 14:30 ET price ≥ -0.15

Table 3.1: Data handling details: ¹We did not count Oct. 20 in the crash window as the trading activity during Oct. 20 was very unusual. ²We define the lowest price during the crash to be the transaction price below which the security traded for at least certain cumulative time period during the crash window. ³For 2010 Flash Crash, we filter out stocks with maximum drawdown of more than 15%, as that can be categorized as broken trades. For 1987 crash, we filter out stocks whose ISSM prices do not match CRSP price and also stocks with less than 30 trades on Oct. 19.

3.1.4 Estimation of the Market-Wide Liquidity Shock

Table 3.2 reports the market-wide liquidity shock estimation from the Step 2 cross-sectional regression for the two crashes. It shows the data fits our model. Specifically, the estimated α and θ are quite close to the market drawdown and the market-wide liquidity shock documented in the official reports and are both statistically significant.

		The Oct. 1987 Crash	The 2010 Flash Crash
Empirical Result	α θ	-26.77*** -4.875***	-5.076*** -2.544***
Official Report	market drawdown market-wide liquidity shock	about -30 about -3	about -5 about -1.9 to -4
Number of SP500 Stocks		436	457

Table 3.2: Estimates of α : market drawdown (percentage) and θ : market-wide liquidity shock (\$ Billion). *Note:* *p<0.1; **p<0.05; ***p<0.01

Figure 3.1 and Figure 3.2 are graphs that contain distributions of independent and dependent variables in the regression.

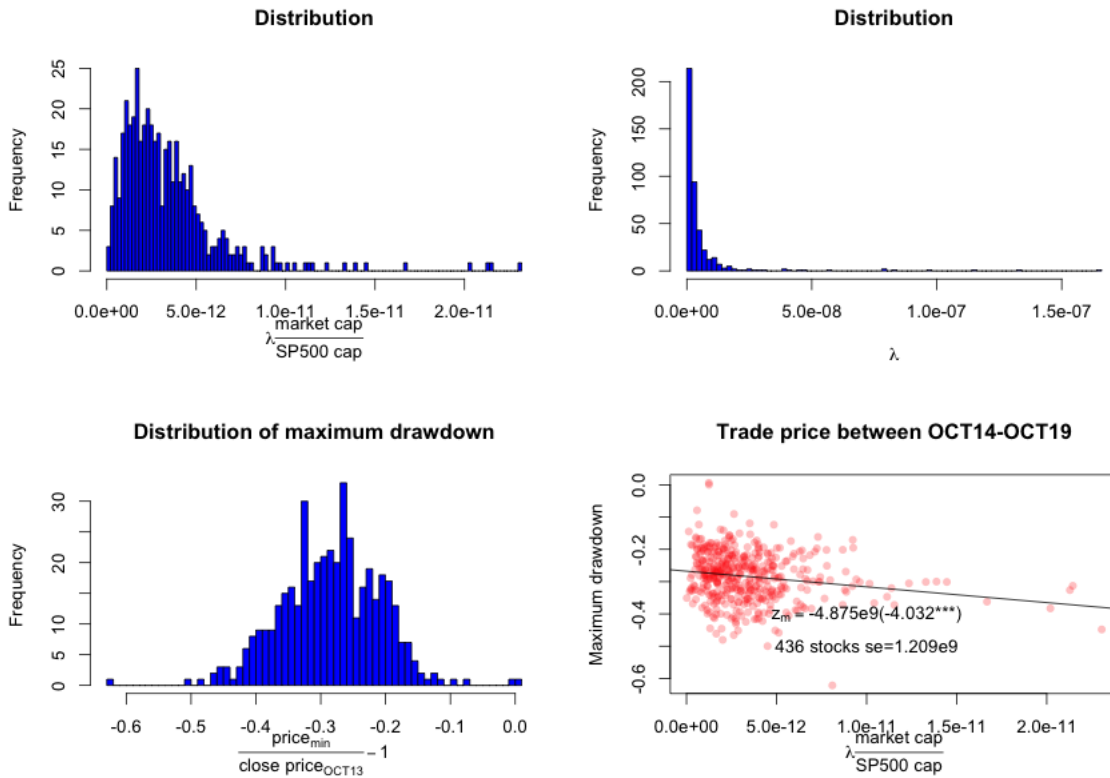


Figure 3.1: The Oct. 1987 Crash.

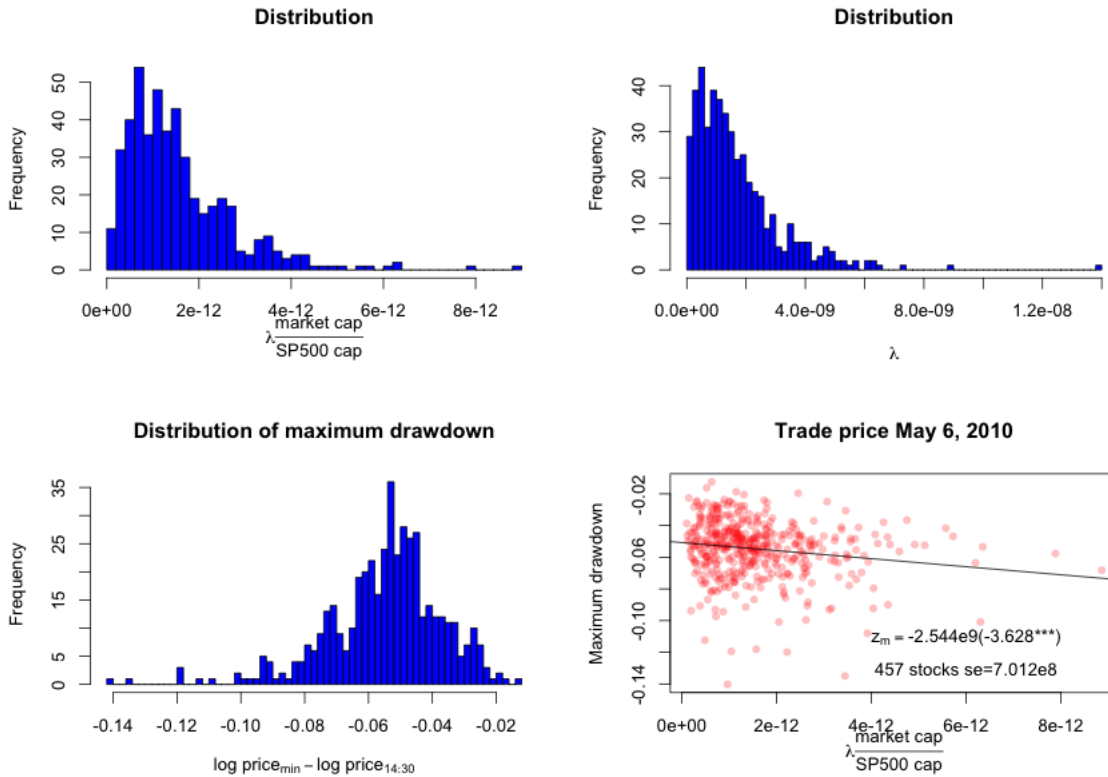


Figure 3.2: The 2010 Flash Crash.

3.1.5 Estimation of the Limited Risk-Bearing Capacity Induced Price Impact

Our model is about information asymmetry induced price impact, another source of price impact is limited risk-bearing capacity (Grossman and Miller 1988; Greenwood 2005). We would like to separate the two kinds of price impact empirically.

In Greenwood (2005), there is no information asymmetry related price impact. The paper states³ that under a publicly informed demand shock, price can be affected by the risk aversion of market makers. The cross-sectional regression of the affected securities' returns can be described as:

$$\Delta P_i = \alpha + \beta(\Sigma \Delta X)_i + \epsilon_i \tag{3.9}$$

where Σ is the covariance matrix of the fundamentals of the affected securities.

We can see from the equation that there is a spill-over risk effect that can impact prices under a multi-asset, risk-aversion setting without any informational effect. While our previ-

³See Proposition 3 in Greenwood (2005)

ous model is built under a single-asset, informational setting without any limited risk-bearing effect, an intuitive way to identify the two effects is to combine the two in a linear way and run a multivariate cross-sectional regression.

The multivariate regression is structured as follows:

$$\Delta P_i = \alpha + \beta(\Sigma\Delta X)_i + \lambda_i\Delta X_i + \epsilon_i \quad (3.10)$$

$$= \alpha + \beta(\Sigma\gamma\Delta z_m)_i + \lambda_i\gamma_i\Delta z_m + \epsilon_i \quad (3.11)$$

$$= \alpha + \theta_1\lambda_i\gamma_i + \theta_2(\Sigma\gamma)_i + \epsilon_i \quad (3.12)$$

For the 2010 Flash Crash, we use 2009/11/01-2010/05/05 CRSP daily stock return without dividend (RETX) data to estimate covariance matrix Σ ; for 1987 crash, we use 1987/01/01/-1987/10/13 CRSP RETX data to estimate covariance matrix Σ .

The results are as follows:

		Oct. 1987 Crash	2010 Flash Crash
Empirical Result	α	-21.90***(t=-17.069)	-3.936***(t=-16.200)
	θ_1	-3.582***(t=-2.921)	-2.393***(t=-3.522)
	θ_2	-620.1***(t=-4.189)	-131.0***(t=-5.609)
Number of SP500 Stocks		436	457

Table 3.3: Estimates of α : percentage and θ_1 : market-wide liquidity shock (\$ Billion). *Note:* *p<0.1; **p<0.05; ***p<0.01

As in Greenwood (2005), we also look into the idiosyncratic and hedging contributions to arbitrage risk respectively. Rewrite:

$$(\Sigma\Delta X)_i = \sigma_i^2\Delta X_i + \sum_{j \neq i} \sigma_{ij}\Delta X_j \quad (3.13)$$

and run a regression as follows:

$$\Delta P_i = \alpha + \theta_1\lambda_i\gamma_i + \theta_2\sigma_i^2\gamma_i + \theta_3 \sum_{j \neq i} \sigma_{ij}\gamma_j + \epsilon_i \quad (3.14)$$

It turns out that the hedge contributions to arbitrage risk is statistically significant, but the idiosyncratic contribution is not. Table 3.4 shows the result.

		Oct. 1987 Crash	2010 Flash Crash
Empirical Result	α	-21.90***(t=-17.069)	-3.936***(t=-16.200)
	θ_1	-3.338**(t=-2.527)	-2.272***(t=-3.157)
	θ_2	-3078(t=-0.623)	-711.2(t=-0.628)
	θ_3	-608.6***(t=-4.057)	-129.2***(t=-5.458)
Number of SP500 Stocks		436	457

Table 3.4: Estimates of α : percentage and θ_1 : market-wide liquidity shock (\$ Billion). *Note:* *p<0.1; **p<0.05; ***p<0.01

3.2 A Multi-Asset Model and Empirical Discussion

In this section, we build a novel multi-asset strategic trading model that captures the informational price impact and the price impact from limited risk-bearing capacity of market makers. This model is an extension of Kyle (1985)'s single asset strategic trading model and can be reduced to special cases in the multi-asset setting of Caballé and Krishnan (1994) and Greenwood (2005).

3.2.1 Model

There are n risky assets in the market, each risky asset is in zero net supply and we assume the risk free rate to be zero.

We assume the *ex post* liquidation value of risky assets is $\tilde{v} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_n)$, which is multivariate normally distributed with mean that we normalize to zero and a nonsingular covariance matrix Σ_v .

The quantity traded by noise traders, denoted by vector $\tilde{u} = (\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_n)$ is assumed to be multivariate normally distributed with mean zero and nonsingular covariance matrix Σ_u .⁴ We also assume \tilde{u} is independent from \tilde{v} .

As in Kyle (1985), at period 0, the exogenous values of \tilde{v} and \tilde{u} are realized and a single informed trader observes \tilde{v} but not \tilde{u} . Being risk neutral, the informed trader solves the following profit optimization problem:

$$\max_y E[(\tilde{v} - \tilde{p})^T y | \tilde{v}] \quad (3.15)$$

At period 1, the risk averse uninformed market maker⁵ observes the aggregate order flow

⁴Covariance matrices are positive semi-definite. Nonsingular covariance matrices are positive definite.

⁵ ρ captures her degree of risk aversion.

$\tilde{y} + \tilde{u}$ and solves the following CARA utility optimization problem:

$$\begin{aligned} & \max_m \mathbb{E}[-\exp[-\rho((\tilde{v} - p)^T m)] | \tilde{y} + \tilde{u}, p] \\ \Rightarrow & \max_m \mathbb{E}[(\tilde{v} - p)^T m | \tilde{y} + \tilde{u}, p] - \frac{\rho}{2} \text{var}[(\tilde{v} - p)^T m | \tilde{y} + \tilde{u}, p] \end{aligned} \quad (3.16)$$

In equilibrium, the market maker sets the price such that we would have the market clearing condition:

$$m + y + u = 0 \quad (3.17)$$

As usual, we conjecture the following linear form for equilibrium price \tilde{p} and informed trader's demand \tilde{y} .

$$\begin{aligned} \tilde{p} &= P(\tilde{y} + \tilde{u}) = A(\tilde{y} + \tilde{u}) \\ \tilde{y} &= B\tilde{v} \end{aligned}$$

It turns out that under this general setting, B can always be represented by $B = \frac{A^{-1}}{2}$ from Equation (3.15) and we can solve for a symmetric linear equilibrium as follows:

Proposition 1. *If a linear equilibrium with symmetric A exists, then A must satisfy the following equation:*

$$A\Sigma_v^{-1}A - \rho A = \frac{\Sigma_u^{-1}}{4} \quad (3.18)$$

Remark: If we do not assume A is symmetric, then we need to solve the following equation:

$$\rho(A + A^T)\Sigma_u(A + A^T) = (A + A^T)\Sigma_u(A + A^T)\Sigma_v^{-1}A - A^T \quad (3.19)$$

Remark: For the second order condition of informed trader's optimization problem to be satisfied, A must be positive definite.

Proposition 2. *A linear equilibrium for which A is symmetric positive definite always exists. A symmetric solution is given by:*

$$A = \frac{1}{2}\Sigma_v^{1/2}(\Sigma_v^{-1/2}\Sigma_u^{-1}\Sigma_v^{-1/2} + \rho^2 I)^{1/2}\Sigma_v^{1/2} + \frac{\rho}{2}\Sigma_v \quad (3.20)$$

Proofs of Proposition 1 and Proposition 2 can be found in Section 3.5

From Proposition 2, we can see in equilibrium, the price impact matrix A is a function of Σ_v , Σ_u and ρ . When market maker is risk neutral (i.e. $\rho = 0$), the solution is reduced to the one obtained in Caballé and Krishnan (1994) in the single insider with perfect private information case: $A = \frac{1}{2}\Sigma_v^{1/2}(\Sigma_v^{-1/2}\Sigma_u^{-1}\Sigma_v^{-1/2})^{1/2}\Sigma_v^{1/2}$. While in the limit case $\Sigma_u^{-1} \rightarrow 0$, the noise trading becomes infinite and agents rationally ignore price information (Admati 1985). Then there are no information asymmetry, and A can be simplified as $A = \rho\Sigma_v$, we get similar result as in Greenwood (2005).

3.2.2 Empirical Discussion

Our model indicates the following price impact equation: $\Delta P = A\Delta X$ where

$$A = \frac{1}{2}\Sigma_v^{1/2}(\Sigma_v^{-1/2}\Sigma_u^{-1}\Sigma_v^{-1/2} + \rho^2 I)^{1/2}\Sigma_v^{1/2} + \frac{\rho}{2}\Sigma_v.$$

Now we make some assumptions to simplify the structure of Σ_v and Σ_u , by decomposing the \tilde{v}_i and the \tilde{u}_i into an idiosyncratic component and a market-wide component as in Section 3.1.

Assume $\tilde{v}_i = \tilde{s}_i + \beta_i\tilde{s}_m$ and $\tilde{u}_i = \tilde{z}_i + \gamma_i\tilde{z}_m$, with $\tilde{s}_i \sim N(0, \sigma_{s_i}^2)$, $\tilde{s}_m \sim N(0, \sigma_{s_m}^2)$, $\tilde{z}_i \sim N(0, \sigma_{z_i}^2)$, $\tilde{z}_m \sim N(0, \sigma_{z_m}^2)$, and they are independent of each other. We have:

$$\Delta X = \frac{1}{2}A^{-1}(\tilde{s} + \beta\tilde{s}_m) + \tilde{z} + \gamma\tilde{z}_m \quad (3.21)$$

$$\Delta P = \frac{1}{2}(\tilde{s} + \beta\tilde{s}_m) + A(\tilde{z} + \gamma\tilde{z}_m) \quad (3.22)$$

$$\begin{aligned} \text{var}(\Delta P) &= \frac{1}{4}\Sigma_v + A\Sigma_u A \\ &= \frac{1}{4}\Sigma_v(I + (I - \rho A^{-1}\Sigma_v)^{-1}) \end{aligned} \quad (3.23)$$

with

$$\Sigma_v = \begin{pmatrix} \sigma_{s_1}^2 + \beta_1^2\sigma_{s_m}^2 & \beta_1\beta_2\sigma_{s_m}^2 & \cdots & \beta_1\beta_n\sigma_{s_m}^2 \\ \beta_2\beta_1\sigma_{s_m}^2 & \sigma_{s_2}^2 + \beta_2^2\sigma_{s_m}^2 & \cdots & \beta_2\beta_n\sigma_{s_m}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_n\beta_1\sigma_{s_m}^2 & \beta_n\beta_2\sigma_{s_m}^2 & \cdots & \sigma_{s_n}^2 + \beta_n^2\sigma_{s_m}^2 \end{pmatrix} \quad (3.24)$$

$$\Sigma_u = \begin{pmatrix} \sigma_{z_1}^2 + \gamma_1^2\sigma_{z_m}^2 & \gamma_1\gamma_2\sigma_{z_m}^2 & \cdots & \gamma_1\gamma_n\sigma_{z_m}^2 \\ \gamma_2\gamma_1\sigma_{z_m}^2 & \sigma_{z_2}^2 + \gamma_2^2\sigma_{z_m}^2 & \cdots & \gamma_2\gamma_n\sigma_{z_m}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_n\gamma_1\sigma_{z_m}^2 & \gamma_n\gamma_2\sigma_{z_m}^2 & \cdots & \sigma_{z_n}^2 + \gamma_n^2\sigma_{z_m}^2 \end{pmatrix} \quad (3.25)$$

Estimation of the market wide liquidity shock

Step 1: Before crash

Given $\Delta P = A\Delta X$ in our model, presumably we have:

$$\begin{aligned} \Delta P_{it} &= (A\Delta X_t)_i \\ &= \sum_{j=1}^n A_{ij}\Delta X_{jt} \end{aligned} \quad (3.26)$$

$$= \lambda_i\Delta X_{it} + \sum_{j \neq i} A_{ij}\Delta X_{jt} \quad (3.27)$$

where $\lambda_i := A_{ii}$ is the idiosyncratic price impact factor.

If $\text{cov}(\Delta X_{it}, \Delta X_{jt}) = 0$ for every $j \neq i$, then we can run a following OLS regression to get an unbiased estimator of λ_i .

$$\Delta P_{it} = \alpha + \lambda_i \Delta X_{it} + \epsilon_{it} \quad (3.28)$$

However if $\text{cov}(\Delta X_{it}, \Delta X_{jt}) \neq 0$ for every $j \neq i$, we need to run a full scale regression to estimate λ_i .

$$\Delta P_{it} = \alpha + \lambda_i \Delta X_{it} + \sum_{j \neq i} A_{ij} \Delta X_{jt} + \epsilon_{it} \quad (3.29)$$

From our specified structure of ΔX , $\text{cov}(\Delta X_i, \Delta X_j) \neq 0$, and therefore ideally we need to run Equation (3.36) to get an unbiased estimator of λ_i .

Step 2: During crash

$$\begin{aligned} \Delta P_i &= (A \Delta X)_i \\ &= \alpha + (A \gamma \Delta z_m)_i + \epsilon_i \\ &= \alpha + (\text{diag}(A) \gamma \Delta z_m)_i + ((A - \text{diag}(A)) \gamma \Delta z_m)_i + \epsilon_i \\ &= \alpha + \lambda_i \gamma_i \Delta z_m + \eta_i \\ &= \alpha + \theta_1 \lambda_i \gamma_i + \eta_i \end{aligned} \quad (3.30)$$

Let $A = M + \frac{\rho}{2} \Sigma_v$, where $M = \frac{1}{2} \Sigma_v^{1/2} (\Sigma_v^{-1/2} \Sigma_u^{-1} \Sigma_v^{-1/2} + \rho^2 I)^{1/2} \Sigma_v^{1/2}$.

Since $A - \text{diag}(A) = M - \text{diag}(M) + \frac{\rho}{2} (\Sigma_v - \text{diag}(\Sigma_v))$, and $\lambda_i = \text{diag}(A)_i = \text{diag}(M + \frac{\rho}{2} \Sigma_v)_i$, by the structure of Σ_v , $E[\eta_i | \lambda_i \gamma_i] \neq 0$, we have endogeneity in the cross-sectional regression. So θ_1 is a biased estimator of Δz_m .

To see $E[\eta_i | \lambda_i \gamma_i] \neq 0$, we have:

$$\begin{aligned} \text{diag}(\Sigma_v) \gamma &= \begin{pmatrix} (\sigma_{s_1}^2 + \beta_1^2 \sigma_{s_m}^2) \gamma_1 \\ (\sigma_{s_2}^2 + \beta_2^2 \sigma_{s_m}^2) \gamma_2 \\ \vdots \\ (\sigma_{s_n}^2 + \beta_n^2 \sigma_{s_m}^2) \gamma_n \end{pmatrix} = \begin{pmatrix} \sigma_{s_1}^2 \gamma_1 \\ \sigma_{s_2}^2 \gamma_2 \\ \vdots \\ \sigma_{s_n}^2 \gamma_n \end{pmatrix} + \phi \quad (3.31) \\ (\Sigma_v - \text{diag}(\Sigma_v)) \gamma \Delta z_m &= \begin{pmatrix} \beta_1 (\beta_2 \gamma_2 + \beta_3 \gamma_3 + \cdots + \beta_n \gamma_n) \\ \beta_2 (\beta_1 \gamma_1 + \beta_3 \gamma_3 + \cdots + \beta_n \gamma_n) \\ \vdots \\ \beta_n (\beta_1 \gamma_1 + \beta_2 \gamma_2 + \cdots + \beta_{n-1} \gamma_{n-1}) \end{pmatrix} \sigma_{s_m}^2 \Delta z_m \\ &= \begin{pmatrix} \beta_1 (K - \beta_1 \gamma_1) \\ \beta_2 (K - \beta_2 \gamma_2) \\ \vdots \\ \beta_n (K - \beta_n \gamma_n) \end{pmatrix} \sigma_{s_m}^2 \Delta z_m = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} K \sigma_{s_m}^2 \Delta z_m - \phi \Delta z_m \quad (3.32) \end{aligned}$$

where

$$K = \sum_{i=1}^n \beta_i \gamma_i \quad (3.33)$$

$$\phi = \begin{pmatrix} \beta_1^2 \gamma_1 \\ \beta_2^2 \gamma_2 \\ \vdots \\ \beta_n^2 \gamma_n \end{pmatrix} \sigma_{sm}^2 \quad (3.34)$$

The error term η has been contaminated by the off-diagonal terms of Σ_v and can be correlated with the regressor.

Add risk aversion effect:

Let $\Sigma_{hedge} = \Sigma_v - \text{diag}(\Sigma_v)$. To get a better estimator of θ_1 as well as checking the role of risk aversion effect predicted by our model, we use the historical return covariance estimation $\text{var}(\Delta P) = \frac{1}{4} \Sigma_v (I + (I - \rho A^{-1} \Sigma_v)^{-1})$ as a proxy of Σ_v and run the following regression:

$$\begin{aligned} \Delta P_i &= (A \Delta X)_i \\ &= \alpha + (A \gamma \Delta z_m)_i + \epsilon_i \\ &= \alpha + (\text{diag}(A) \gamma \Delta z_m)_i + ((A - \text{diag}(A)) \gamma \Delta z_m)_i + \epsilon_i \\ &= \alpha + (\text{diag}(A) \gamma \Delta z_m)_i + ((\Sigma_v - \text{diag}(\Sigma_v)) \frac{\rho}{2} \gamma \Delta z_m)_i + ((M - \text{diag}(M)) \gamma \Delta z_m)_i + \epsilon_i \\ &= \alpha + \lambda_i \gamma_i \Delta z_m + \frac{\rho \Delta z_m}{2} (\Sigma_{hedge} \gamma)_i + \eta_i \\ &= \alpha + \theta_1 \lambda_i \gamma_i + \theta_2 (\Sigma_{hedge} \gamma)_i + \eta_i \end{aligned} \quad (3.35)$$

Strictly speaking, the θ_1 and θ_2 estimated by the above equation are still biased as the error term η is still contaminated.

Unless we can estimate the full price impact matrix A , it is difficult to fully separate the information effect and the risk aversion effect in the regression.

We derive the following corollaries that can potentially simplify the structure of matrix A and facilitate the estimation. Proofs can be found in Section 3.5.

Corollary 1. *If Σ_v and Σ_u commute, then there exists a unique symmetric linear equilibrium, with $A = (\frac{1}{4} \Sigma_u^{-1} \Sigma_v + \frac{\rho^2}{4} \Sigma_v^2)^{\frac{1}{2}} + \frac{\rho}{2} \Sigma_v$.*

Corollary 2. *If Σ_v and Σ_u commute, and their corresponding eigenvalues satisfy: $\rho^2 \lambda_{v,i} \lambda_{u,i} \gg 1$ for every i , then we have $A \approx \rho \Sigma_v + \frac{1}{4\rho} \Sigma_u^{-1}$.*

Corollary 3. *If Σ_v and Σ_u commute, and their corresponding eigenvalues satisfy: $\rho^2 \lambda_{v,i} \lambda_{u,i} \ll 1$ for every i , then we have $A \approx \frac{1}{2} \Sigma_v^{\frac{1}{2}} \Sigma_u^{-\frac{1}{2}} + \frac{\rho^2}{4} \Sigma_v^{\frac{3}{2}} \Sigma_u^{\frac{1}{2}} + \frac{\rho}{2} \Sigma_v$.*

Below, we make a few simplification assumptions so that the above corollaries may hold and see if they can shed some light on achieving our goal of a better estimation.

Assumption 1 (a.k.a. Corollary 1): If $\sigma_{z_i}^2 = \sigma_u^2 \gg \sigma_{z_m}^2$, that is, if the idiosyncratic noise trading is far larger than the market-wide noise trading, then $\Sigma_u \approx \sigma_u^2 I$ and $\Sigma_v \Sigma_u \approx \Sigma_u \Sigma_v$. $A \approx (\frac{1}{4} \Sigma_u^{-1} \Sigma_v + \frac{\rho^2}{4} \Sigma_v^2)^{\frac{1}{2}} + \frac{\rho}{2} \Sigma_v$.

Assumption 1 basically assumes the variance of idiosyncratic liquidity shock is far larger than the market wide liquidity shock. Then the liquidity covariance matrix Σ_u can be approximated by a diagonal matrix. Therefore, Σ_v and Σ_u approximately commute and Corollary 1 holds.

Assumption 2(a) (a.k.a. Corollary 2): In addition to Assumption 1, if $\rho^2 \lambda_{v,i} \lambda_{u,i} \gg 1$ for every i , then $A \approx \rho \Sigma_v + \frac{1}{4\rho} \Sigma_u^{-1} = \underbrace{\rho \Sigma_v}_{\text{hedge}} + \underbrace{\frac{1}{4\rho \sigma_u^2} I}_{\text{hedge:infor}}$.

Under Assumption 2(a), the price impact matrix can be decomposed into a hedge (risk aversion) component and a hedge and information interaction component.

Test of Assumption 2(a):

Based on Assumption 2(a), we can run the following two-step empirical test:

Step 1:

$$\Delta P_{it} = \alpha + \lambda_i \Delta X_{it} + \sum_{j \neq i} A_{ij} \Delta X_{jt} + \epsilon_{it} \quad (3.36)$$

Step 2:

$$\begin{aligned} \Delta P_i &= \alpha + \lambda_i \gamma_i \Delta z_m + \frac{\rho \Delta z_m}{2} (\Sigma_{\text{hedge}} \gamma)_i + \epsilon_i \\ &= \alpha + \theta_1 \lambda_i \gamma_i + \theta_2 (\Sigma_{\text{hedge}} \gamma)_i + \epsilon_i \end{aligned} \quad (3.37)$$

where $\lambda_i \approx \rho \Sigma_v(i, i) + \frac{1}{4\rho \sigma_u^2} = \rho(\sigma_{s_i}^2 + \beta_i^2 \sigma_{s_m}^2) + \frac{1}{4\rho \sigma_u^2}$.

Basically this suggests a similar test as is shown previously (Equation (3.35)), but the endogeneity problem disappears under the assumptions.

Assumption 2(b) (a.k.a. Corollary 3): In addition to Assumption 1, if $\rho^2 \lambda_{v,i} \lambda_{u,i} \ll 1$ for every i , then $A \approx \frac{1}{2} \Sigma_v^{\frac{1}{2}} \Sigma_u^{-\frac{1}{2}} + \frac{\rho^2}{4} \Sigma_v^{\frac{3}{2}} \Sigma_u^{\frac{1}{2}} + \frac{\rho}{2} \Sigma_v = \underbrace{\frac{1}{2 \sigma_u}}_{\text{infor}} \Sigma_v^{\frac{1}{2}} + \underbrace{\frac{\rho^2 \sigma_u}{4}}_{\text{hedge:infor}} \Sigma_v^{\frac{3}{2}} + \underbrace{\frac{\rho}{2}}_{\text{hedge}} \Sigma_v$.

Assumption 2(b) separates the price impact matrix into three components: a hedge, an information and an interaction component.

Test of Assumption 2(b):

Under Assumption 2(b), we can run the following empirical test:

$$\begin{aligned}
\Delta P_i &= (A\Delta X)_i \\
&= \alpha + (A\gamma\Delta z_m)_i + \epsilon_i \\
&= \alpha + \frac{\Delta z_m}{2\sigma_u}(\Sigma_v^{\frac{1}{2}}\gamma)_i + \frac{\rho^2\sigma_u\Delta z_m}{4}(\Sigma_v^{\frac{3}{2}}\gamma)_i + \frac{\rho\Delta z_m}{2}(\Sigma_v\gamma)_i + \epsilon_i \\
&= \alpha + \beta_1(\Sigma_v^{\frac{1}{2}}\gamma)_i + \beta_2(\Sigma_v^{\frac{3}{2}}\gamma)_i + \beta_3(\Sigma_v\gamma)_i + \epsilon_i
\end{aligned} \tag{3.38}$$

However, in practice, as $\Sigma_v^{\frac{1}{2}}\gamma$, $\Sigma_v\gamma$ and $\Sigma_v^{\frac{3}{2}}\gamma$ can be highly correlated, it is difficult to separate them in linear regression.

3.3 Modified γ^* under imperfect hedging

In Section 3.1, we have assumed that γ_i equals the ratio of the stock i 's market capital to the SP500 index capital at the beginning of the crash. This assumes the index futures arbitragers can perfectly hedge their risk and transmit the E-mini 500 sale into SP500 underlying stocks. What if they cannot be perfectly hedged? What will the modified γ_i^* 's be?

We consider a three-period model:

- At period 0, the arbitrageur has a wealth shock of Q dollar amount of E-mini 500 futures.
- At period 1, the arbitrageur sells $Q_i = x_{1i}p_{0i}$ dollar amount of each underlying stock i to hedge the risk.
- At period 2, each stock i pays off a dividend with independent normal distribution $\epsilon_i \sim N(0, \sigma_i^2)$ and the arbitrageur finalizes the hedge.

We assume the arbitrageur has CARA utility of risk aversion constant α .

- At period 0: $W_0 = Q = \sum_i \gamma_i Q = \sum_i x_{0i} p_{0i}$.
- At period 1: $W_1 = \sum_i (x_{0i} - x_{1i}) p_{0i} + (x_{1i} p_{0i} - \frac{1}{2} \lambda_i x_{1i}^2)$.
- At period 2: $W_2 = \sum_i (x_{0i} - x_{1i}) (p_{0i} + \epsilon_i) + (x_{1i} p_{0i} - \frac{1}{2} \lambda_i x_{1i}^2)$.

And the arbitrageur is to solve the following maximization problem:

$$\max_{x_{1i}} -E\exp[-\alpha W_2] \tag{3.39}$$

which is equivalent to:

$$\begin{aligned}
&\max_{x_{1i}} \alpha E[W_2] - \frac{1}{2} \alpha^2 \text{var}[W_2] \\
\iff \max_{x_{1i}} \alpha \left(Q - \frac{1}{2} \sum_i \lambda_i x_{1i}^2 \right) - \frac{1}{2} \alpha^2 \sum_i (x_{0i} - x_{1i})^2 \sigma_i^2
\end{aligned} \tag{3.40}$$

This is a concave function and by taking the derivative with respect to x_{1i} , we can solve for the optimal with the first order condition:

$$-\lambda_i x_{1i} + \alpha \sigma_i^2 (x_{0i} - x_{1i}) = 0, \forall i \quad (3.41)$$

Therefore, we have:

$$x_{1i} = \frac{\alpha \sigma_i^2}{\lambda_i + \alpha \sigma_i^2} x_{0i} \quad (3.42)$$

$$\gamma_i^* = \frac{\alpha \sigma_i^2}{\lambda_i + \alpha \sigma_i^2} \gamma_i < \gamma_i \quad (3.43)$$

We can see that if $\lambda_i = 0, \forall i$, then $\gamma_i^* = \gamma_i$. The existence of the price impact factor λ_i makes the arbitrage hedge the risk with a smaller amount on the underlying stock i .

Now if we assume the returns of SP500 stocks follow a general form of covariance matrix $\Sigma_{n \times n}$ such that they are not necessarily independent. Then the maximization problem becomes:

$$\max_{x_1} \alpha (Q - \frac{1}{2} x_1^T \Lambda x_1) - \frac{1}{2} \alpha^2 (x_0 - x_1)^T \Sigma (x_0 - x_1) \quad (3.44)$$

and we have:

$$x_1 = (\alpha \Sigma + \Lambda)^{-1} \alpha \Sigma x_0 \quad (3.45)$$

and the modified $\gamma^* = x_1^T p_0 / Q$.

3.4 Proofs for Section 3.1

Proof. Conjecture a linear equilibrium. The risk neutral competitive market makers set the price conditioning on the aggregate order flow imbalance:

$$\begin{aligned} P(x) &= \mathbb{E}[\tilde{v}_1 | x] = \mathbb{E}[\tilde{v}_1 | x = \tilde{y} + \tilde{z} + \gamma \tilde{z}_m] \\ &= v_0 + \lambda x \end{aligned} \quad (3.46)$$

where y is the demand from risk neutral informed traders given by:

$$y = \arg \max_y \mathbb{E}[y(\tilde{v}_1 - P(y + \tilde{z} + \gamma \tilde{z}_m)) | \tilde{s} = s, \tilde{s}_m = s_m] \quad (3.47)$$

Plug in the linear form of the price into the informed trader's maximization problem (which becomes quadratic) and solve for the first order condition we have:

$$\begin{aligned} y &= \arg \max_y [y(s + \beta s_m - \lambda y)] \\ &= \frac{1}{2\lambda} s + \frac{\beta}{2\lambda} s_m \end{aligned} \quad (3.48)$$

Plug $y = \frac{1}{2\lambda}s + \frac{\beta}{2\lambda}s_m$ in Equation (3.46) and apply the Projection Theorem, we have:

$$\lambda = \frac{\sqrt{\sigma_s^2 + \beta^2 \sigma_{s_m}^2}}{2\sqrt{\sigma_z^2 + \gamma^2 \sigma_{z_m}^2}} \quad (3.49)$$

□

3.5 Proofs for Section 3.2

Proof of Proposition 1. First, for the informed trader's optimization problem, with the linear conjecture, we have:

$$\max_y \mathbb{E}[(\tilde{v} - \tilde{p})^T y | \tilde{v}] = \max_y \mathbb{E}[(\tilde{v} - A(y + \tilde{u}))^T y | \tilde{v} = v] \quad (3.50)$$

And the first order condition is:

$$v - 2Ay = 0 \quad (3.51)$$

where we have used the symmetry of A.

And therefore we have:

$$\begin{aligned} y &= \frac{A^{-1}}{2} v \\ B &= \frac{A^{-1}}{2} \end{aligned} \quad (3.52)$$

Second, for the market maker's optimization problem, we have m that satisfies the first order condition:

$$\begin{aligned} m &= \frac{1}{\rho} \text{var}[\tilde{v} | \tilde{y} + \tilde{u}]^{-1} (\mathbb{E}[\tilde{v} | \tilde{y} + \tilde{u}] - p) \\ &= \frac{1}{\rho} \text{var}[\tilde{v} | \tilde{y} + \tilde{u}]^{-1} (\mathbb{E}[\tilde{v} | \tilde{y} + \tilde{u}] - A(\tilde{y} + \tilde{u})) \\ &= -(\tilde{y} + \tilde{u}) \end{aligned} \quad (3.53)$$

The last equality is the market clearing condition.

By Projection theorem, we know:

$$\mathbb{E}[\tilde{v} | (\tilde{y} + \tilde{u})] = 2\Sigma_v (A^{-1})^T (A^{-1}\Sigma_v(A^{-1})^T + 4\Sigma_u)^{-1} (\tilde{y} + \tilde{u}) \quad (3.54)$$

$$\text{var}[\tilde{v} | (\tilde{y} + \tilde{u})] = \Sigma_v - \Sigma_v (A^{-1})^T (A^{-1}\Sigma_v(A^{-1})^T + 4\Sigma_u)^{-1} A^{-1}\Sigma_v \quad (3.55)$$

Therefore in equilibrium, from Equation (3.53), A should satisfy:

$$\begin{aligned} &A - 2\Sigma_v (A^{-1})^T (A^{-1}\Sigma_v(A^{-1})^T + 4\Sigma_u)^{-1} \\ &= \rho\Sigma_v - \rho\Sigma_v (A^{-1})^T (A^{-1}\Sigma_v(A^{-1})^T + 4\Sigma_u)^{-1} A^{-1}\Sigma_v \end{aligned} \quad (3.56)$$

Since $(A^{-1})^T = (A^T)^{-1}$, we have:

$$\begin{aligned}\Sigma_v(A^{-1})^T(A^{-1}\Sigma_v(A^{-1})^T + 4\Sigma_u)^{-1} &= (A^T\Sigma_v^{-1})^{-1}(A^{-1}\Sigma_v(A^{-1})^T + 4\Sigma_u)^{-1} \\ &= [(A^{-1}\Sigma_v(A^{-1})^T + 4\Sigma_u) \cdot (A^T\Sigma_v^{-1})]^{-1} \\ &= [A^{-1} + 4\Sigma_u A^T \Sigma_v^{-1}]^{-1}\end{aligned}\quad (3.57)$$

Plug (3.57) into the LHS of (3.56), we have:

$$\begin{aligned}A - 2[A^{-1} + 4\Sigma_u A^T \Sigma_v^{-1}]^{-1} &= A - 2[A^{-1}(I + 4A\Sigma_u A^T \Sigma_v^{-1})]^{-1} \\ &= [I - 2(I + 4A\Sigma_u A^T \Sigma_v^{-1})^{-1}]A\end{aligned}\quad (3.58)$$

Plug (3.57) into the RHS of (3.56), we have:

$$\begin{aligned}\rho\Sigma_v - \rho\Sigma_v(A^{-1})^T(A^{-1}\Sigma_v(A^{-1})^T + 4\Sigma_u)^{-1}A^{-1}\Sigma_v &= \rho\Sigma_v - \rho[A^{-1} + 4\Sigma_u A^T \Sigma_v^{-1}]^{-1}A^{-1}\Sigma_v \\ &= \rho\Sigma_v - \rho[\Sigma_v^{-1}A \cdot (A^{-1} + 4\Sigma_u A^T \Sigma_v^{-1})]^{-1} \\ &= \rho\Sigma_v - \rho[\Sigma_v^{-1} + 4\Sigma_v^{-1}A\Sigma_u A^T \Sigma_v^{-1}]^{-1} \\ &= \rho\Sigma_v - \rho[\Sigma_v^{-1}(I + 4A\Sigma_u A^T \Sigma_v^{-1})]^{-1} \\ &= \rho[I - (I + 4A\Sigma_u A^T \Sigma_v^{-1})^{-1}]\Sigma_v\end{aligned}\quad (3.59)$$

Multiplying both sides by $(I + 4A\Sigma_u A^T \Sigma_v^{-1})$, we get:

$$\begin{aligned}(I + 4A\Sigma_u A^T \Sigma_v^{-1} - 2I)A &= \rho[4A\Sigma_u A^T \Sigma_v^{-1}]\Sigma_v \\ \Leftrightarrow 4A\Sigma_u A^T \Sigma_v^{-1}A - A &= 4\rho A\Sigma_u A^T \\ \Leftrightarrow A^T \Sigma_v^{-1}A - \rho A^T - \frac{1}{4}\Sigma_u^{-1} &= 0 \\ \Leftrightarrow A\Sigma_v^{-1}A - \rho A - \frac{1}{4}\Sigma_u^{-1} &= 0\end{aligned}\quad (3.60)$$

The last equality is again by the symmetry of A .

□

Proof of Proposition 2. Let $G := \frac{1}{4}\Sigma_u^{-1} + \frac{\rho^2}{4}\Sigma_v$, then G is symmetric and positive definite. Let $A = \Sigma_v^{1/2}(\Sigma_v^{-1/2}G\Sigma_v^{-1/2})^{1/2}\Sigma_v^{1/2} + \frac{\rho}{2}\Sigma_v$, and $\Sigma_v^{1/2}$, $\Sigma_v^{-1/2}$ be the unique symmetric positive definite square root of Σ_v and Σ_v^{-1} , then by definition we know A is symmetric positive definite.

Now we have:

$$\begin{aligned}
& A\Sigma_v^{-1}A - \rho A - \frac{1}{4}\Sigma_u^{-1} \\
&= A\Sigma_v^{-1/2}\Sigma_v^{-1/2}A - \rho A - \frac{1}{4}\Sigma_u^{-1} \\
&= (\Sigma_v^{1/2}(\Sigma_v^{-1/2}G\Sigma_v^{-1/2})^{1/2} + \frac{\rho}{2}\Sigma_v^{1/2})((\Sigma_v^{-1/2}G\Sigma_v^{-1/2})^{1/2}\Sigma_v^{1/2} + \frac{\rho}{2}\Sigma_v^{1/2}) - \rho A - \frac{1}{4}\Sigma_u^{-1} \\
&= \Sigma_v^{1/2}(\Sigma_v^{-1/2}G\Sigma_v^{-1/2})\Sigma_v^{1/2} + \rho\Sigma_v^{1/2}(\Sigma_v^{-1/2}G\Sigma_v^{-1/2})\Sigma_v^{1/2} + \frac{\rho^2}{4}\Sigma_v - \rho A - \frac{1}{4}\Sigma_u^{-1} \\
&= G + \rho A - \frac{\rho^2}{4}\Sigma_v - \rho A - \frac{1}{4}\Sigma_u^{-1} \\
&= G + \rho A - \rho A - G = 0
\end{aligned} \tag{3.61}$$

And it is easy to show that:

$$\begin{aligned}
A &= \Sigma_v^{1/2}(\Sigma_v^{-1/2}G\Sigma_v^{-1/2})^{1/2}\Sigma_v^{1/2} + \frac{\rho}{2}\Sigma_v \\
&= \frac{1}{2}\Sigma_v^{1/2}(\Sigma_v^{-1/2}\Sigma_u^{-1}\Sigma_v^{-1/2} + \rho^2 I)^{1/2}\Sigma_v^{1/2} + \frac{\rho}{2}\Sigma_v
\end{aligned} \tag{3.62}$$

□

Remark: Mathematically, this equation belongs to a larger family of algebraic Riccati equations in control theory. General discussions can be found in Lancaster and Rodman (1995). A special case without linear terms is discussed as a symmetric word equation in Armstrong and Hillar (2007).

Proof of Corollary 1. Since Σ_v and Σ_u commute, and they are both symmetric positive definite, we know they can be simultaneously diagonalized. Then it is easy to see that G can commute with $\Sigma_v^{1/2}$ and $\Sigma_v^{-1/2}$ as well.

So A can be expressed as:

$$A = \left(\frac{1}{4}\Sigma_u^{-1}\Sigma_v + \frac{\rho^2}{4}\Sigma_v^2\right)^{\frac{1}{2}} + \frac{\rho}{2}\Sigma_v \tag{3.63}$$

□

Proof of Corollary 2. By spectral theorem, we can rewrite $A = Q\Lambda_A Q^T$, where $\Lambda_A = \frac{1}{2}\left(\frac{\Lambda_v}{\Lambda_u} +$

$\rho^2 \Lambda_v^2)^{\frac{1}{2}} + \frac{\rho}{2} \Lambda_v$. Moreover, if $\rho^2 \lambda_{v,i} \lambda_{u,i} \gg 1$ for every i , then by Taylor expansion, we have:

$$\begin{aligned}
\Lambda_A &= \frac{1}{2} \left(\frac{\Lambda_v}{\Lambda_u} + \rho^2 \Lambda_v^2 \right)^{\frac{1}{2}} + \frac{\rho}{2} \Lambda_v \\
&= \frac{\rho}{2} \Lambda_v \left(1 + \frac{1}{\rho^2 \Lambda_v \Lambda_u} \right)^{\frac{1}{2}} + \frac{\rho}{2} \Lambda_v \\
&\approx \frac{\rho}{2} \Lambda_v \left(1 + \frac{1}{2\rho^2 \Lambda_v \Lambda_u} \right) + \frac{\rho}{2} \Lambda_v \\
&= \rho \Lambda_v + \frac{1}{4\rho \Lambda_u}
\end{aligned} \tag{3.64}$$

So

$$A = Q \Lambda_A Q^T \approx Q \left(\rho \Lambda_v + \frac{1}{4\rho \Lambda_u} \right) Q^T = \rho \Sigma_v + \frac{1}{4\rho} \Sigma_u^{-1} \tag{3.65}$$

□

Proof of Corollary 3. By spectral theorem, we can rewrite $A = Q \Lambda_A Q^T$, where $\Lambda_A = \frac{1}{2} \left(\frac{\Lambda_v}{\Lambda_u} + \rho^2 \Lambda_v^2 \right)^{\frac{1}{2}} + \frac{\rho}{2} \Lambda_v$. Moreover, if $\rho^2 \lambda_{v,i} \lambda_{u,i} \ll 1$ for every i , then by Taylor expansion, we have:

$$\begin{aligned}
\Lambda_A &= \frac{1}{2} \left(\frac{\Lambda_v}{\Lambda_u} + \rho^2 \Lambda_v^2 \right)^{\frac{1}{2}} + \frac{\rho}{2} \Lambda_v \\
&= \frac{1}{2} \left(\frac{\Lambda_v}{\Lambda_u} \right)^{\frac{1}{2}} \left(1 + \rho^2 \Lambda_v \Lambda_u \right)^{\frac{1}{2}} + \frac{\rho}{2} \Lambda_v \\
&\approx \frac{1}{2} \left(\frac{\Lambda_v}{\Lambda_u} \right)^{\frac{1}{2}} \left(1 + \frac{\rho^2 \Lambda_v \Lambda_u}{2} \right) + \frac{\rho}{2} \Lambda_v \\
&= \frac{1}{2} \left(\frac{\Lambda_v}{\Lambda_u} \right)^{\frac{1}{2}} + \frac{\rho^2}{4} \Lambda_v^{\frac{3}{2}} \Lambda_u^{\frac{1}{2}} + \frac{\rho}{2} \Lambda_v
\end{aligned} \tag{3.66}$$

So

$$A = Q \Lambda_A Q^T \approx Q \left(\frac{1}{2} \left(\frac{\Lambda_v}{\Lambda_u} \right)^{\frac{1}{2}} + \frac{\rho^2}{4} \Lambda_v^{\frac{3}{2}} \Lambda_u^{\frac{1}{2}} + \frac{\rho}{2} \Lambda_v \right) Q^T = \frac{1}{2} \Sigma_v^{\frac{1}{2}} \Sigma_u^{-\frac{1}{2}} + \frac{\rho^2}{4} \Sigma_v^{\frac{3}{2}} \Sigma_u^{\frac{1}{2}} + \frac{\rho}{2} \Sigma_v \tag{3.67}$$

□

3.6 Summary

In this chapter, we build a single period, single asset strategic trading model that captures the information asymmetry induced price impact. Our model prediction that a temporary market wide liquidity shock can lead to the cross-sectional price drawdown of SP500 stocks is verified empirically. We also incorporate the limited risk-bearing capacity induced price impact in the empirical test and show that it is another important factor.

In addition, we build a more general single period, multi-asset strategic trading model and solve for a symmetric linear equilibrium. We show that under certain assumptions, the price impact matrix can be decomposed into an information component, a hedge component and their interaction component analytically. However, the empirical test of this model is difficult.

We also build a simple model to calculate the constant to which the demand of an individual risky asset is affected by the market-wide liquidity shock, under imperfect hedging.

Chapter 4

Statistical Data Analysis

In this section, we present some modern machine learning technique applications on market crashes. We use association analysis as an unsupervised learning method to explore the association patterns behind the two crashes. We also use statistical model selection methods to compare previous model based information asymmetry and risk aversion effect predictors with other plausible predictors in finance literature.

4.1 Association Analysis

Association analysis is the task of finding interesting relationships in large data sets. These interesting relationships can take two forms: frequent item sets or association rules. Frequent item sets are a collection of items that frequently occur together. A second way to view interesting relationships is association rules. Association rules suggest that a strong relationship exists between two frequent items.

4.1.1 Data

For the 2010 Flash Crash, we collect the following parameters (Table A.1) from different data sources that represent the as-is state for SP500 stocks during the crash.

We can group the parameters into three categories:

- Numerical parameters collected during the crash:
PRC_CHG, ESPRD, VOLUME_RATIO
- Numerical parameters collected right before the crash:
EPS_RATIO, FEPS_RATIO, INST_HHI, INST_PERC, LAMBDA,
MKTCAP, MEANREC, SELLPCT, VOLATILITY
- Categorical parameters: GSECTOR, PRIMEXCH, SPCSRC

For the 1987 crash, the parameters are the same as those of the 2010 crash, except we do not have MEANREC, SELLPCT, and EPS_RATIO for the 1987 crash. Also in the association analysis we do not include the PRIMEXCH parameter, as more than 90% of sample stocks are listed in NYSE. Table A.2 lists the details about the parameters.

4.1.2 Exploratory Data Analysis

Exploratory data analysis (EDA) is a useful way to visualize patterns that may be otherwise not obvious in traditional summary statistics of parameters.

The 2010 Flash Crash

Figure 4.1 shows the absolute frequency of observations on each of the three categorical parameters respectively. Specifically, we can see that over 80% of the SP500 stocks are primarily listed in NYSE.

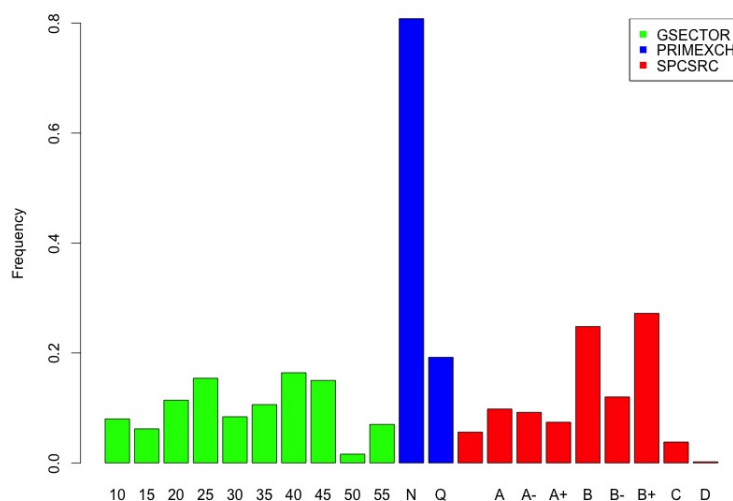


Figure 4.1: Barplot of frequency of three categorical parameters for the 2010 Flash Crash.

We use boxplots to reflect the variations of in-crash parameter distribution across different categorical parameters.¹ Figure 4.2 shows the box-plot of the three in-crash parameters against GSECTOR. We can see the medians of PRC_CHG and ESPRD do not vary much across GSECTOR.

¹We exclude an outlier stock CNP with extreme PRC_CHG and ESPRD

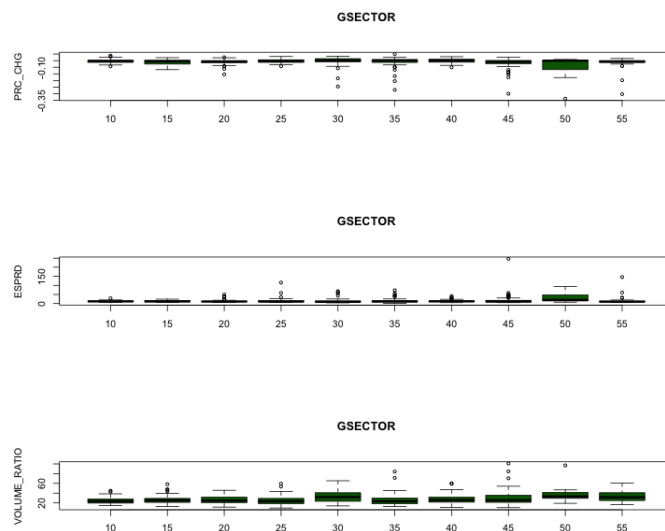


Figure 4.2: Boxplots of PRC_CHG, ESPRD, and VOLUME_RATIO, grouped by GSECTOR for the 2010 Flash Crash. Sector definition: 10-Energy; 15-Materials; 20-Industrials; 25-Consumer discretionary; 30-Consumer staples; 35-Healthcare; 40-Financials; 45-Information technology; 50-Telecommunication services; 55-Utilities.

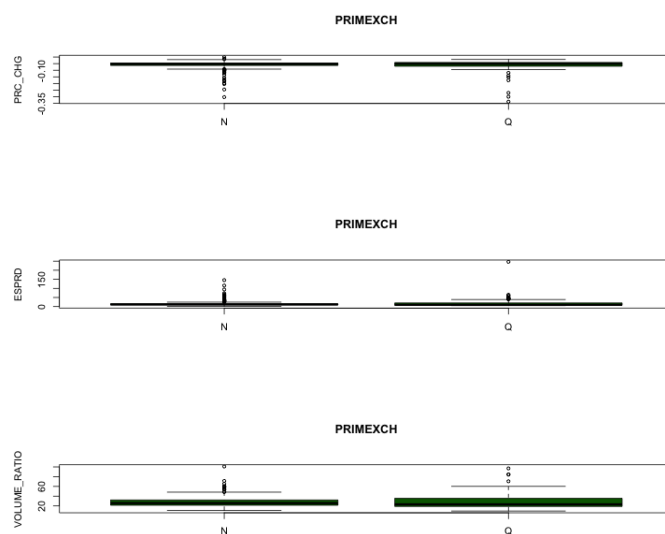


Figure 4.3: Boxplots of PRC_CHG, ESPRD, and VOLUME_RATIO, grouped by PRIMEXCH for the 2010 Flash Crash.

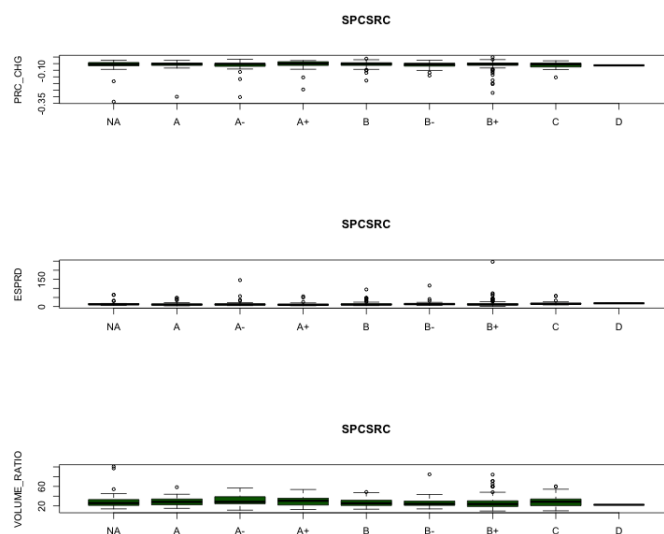


Figure 4.4: Boxplots of PRC_CHG, ESPRD, and VOLUME_RATIO, grouped by SPCSRC for the 2010 Flash Crash (NA means stocks without a SPCSRC).

Figure 4.3 shows the boxplots of three in-crash parameters against PRIMEXCH. The medians are similar, with stocks primarily listed on NASDAQ have larger variation of VOLUME_RATIO. Figure 4.4 shows the boxplots of the three numerical in-crash parameters distribution grouped by SPCSRC.

The correlation matrix of numerical parameters is presented in Figure 4.5. It is not surprising to see that PRC_CHG is highly (negatively) correlated with ESPRD, as the higher the effective spread, the more illiquid the stock during the crash, and that associates with larger price decline. Not surprisingly, we also see the price impact factor LAMBDA is negatively correlated with MKTCAP.

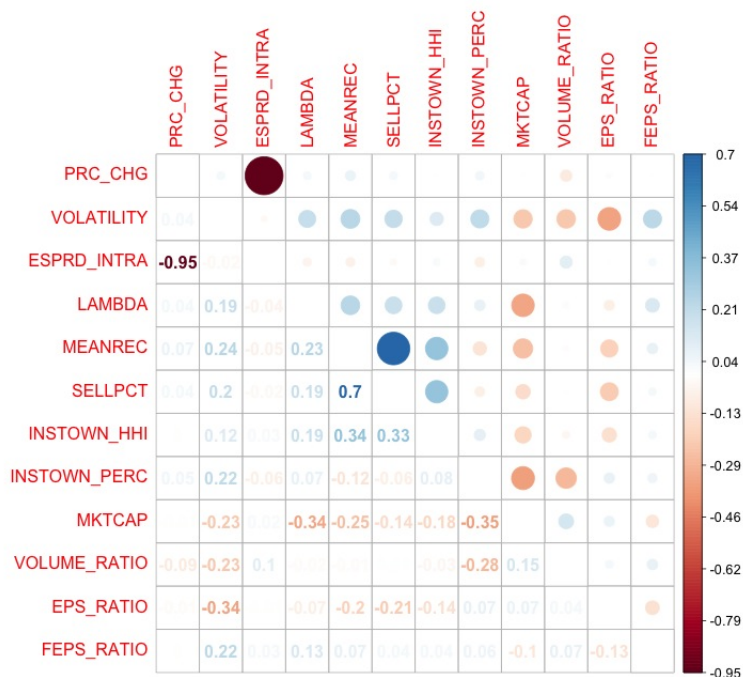


Figure 4.5: Correlation matrix of numerical parameters in the 2010 Flash Crash.

The 1987 Crash

Figure 4.6 shows the absolute frequency of observations on GSECTOR and SPCSRC respectively.

We use boxplots to reflect the variations of in-crash parameter distribution across GSECTOR and SPCSRC. Figure 4.7 shows the boxplots of the three in-crash parameters against GSECTOR. As in the 2010 Flash Crash, the medians of PRC_CHG and ESPRD do not vary much across GSECTOR. Figure 4.8 shows the boxplots of the three numerical in-crash parameters distribution grouped by SPCSRC.

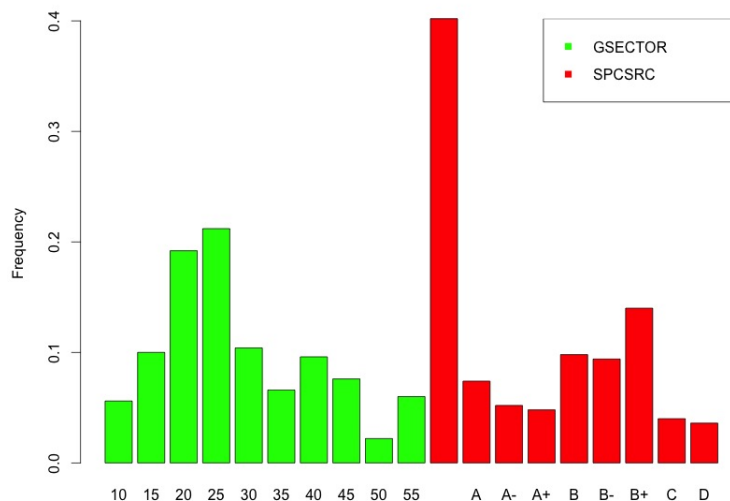


Figure 4.6: Barplot of frequency of GSECTOR and SPCSRC for the 1987 crash.

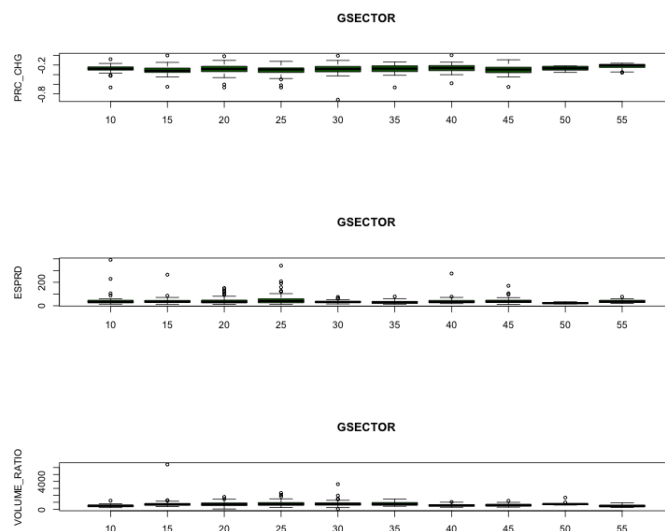


Figure 4.7: Boxplots of PRC_CHG, ESPRD, and VOLUME_RATIO, grouped by GSECTOR for the 1987 crash. Sector definition: 10-Energy; 15-Materials; 20-Industrials; 25-Consumer discretionary; 30-Consumer staples; 35-Healthcare; 40-Financials; 45-Information technology; 50-Telecommunication services; 55-Utilities.

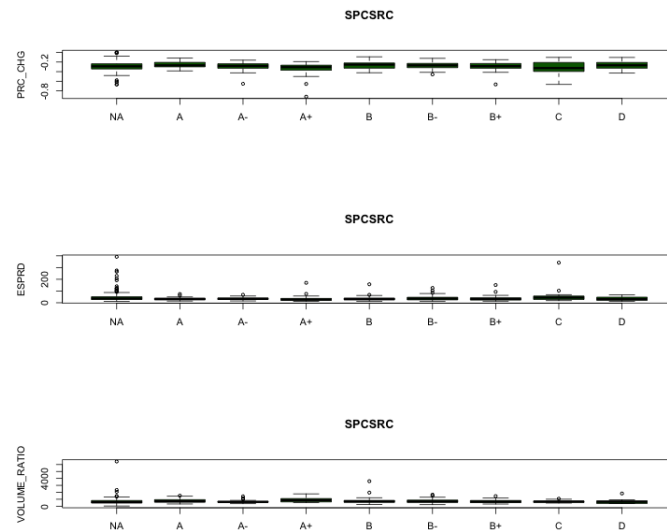


Figure 4.8: Boxplots of PRC_CHG, ESPRD, and VOLUME_RATIO, grouped by SPCSRC for the 1987 crash (NA means stocks without a SPCSRC).

The correlation matrix of numerical parameters is presented in Figure 4.9. PRC_CHG is negatively correlated with ESPRD and VOLATILITY. It is also interesting to see FEPS_RATIO seems to have relatively high correlation with the volatility of the stock.

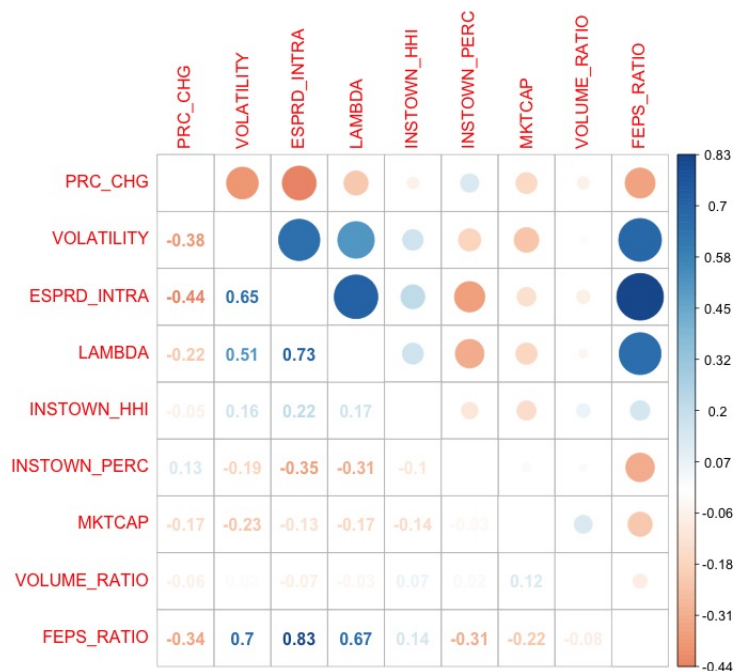


Figure 4.9: Correlation matrix of numerical parameters in the 1987 crash.

4.1.3 Association Analysis Results

For association analysis, we use the Apriori algorithm mentioned in Harrington (2012) and Hastie, Tibshirani, and Friedman (2008).

We use the same data set as in Section 4.1.1, with one additional parameter: $SHOCK = LAMBDA \times MKTCAP$. The purpose is to see if this additional parameter can play a role in our association analysis.

To apply the Apriori algorithm, we convert each numerical parameter into two categories: above and below the median of the original numerical parameter. We keep the original categorical parameters.

After removing observations of missing parameter values, we got a 475×16 matrix of 475 stocks on 16 categorical parameters.

The criteria we use to select frequent itemsets or association rules has three main concepts:

- The *support* of an itemset is defined as the percentage of the data set that contains this itemset.
- The *confidence* is defined for an association rule like $A \Rightarrow B$. The confidence for this rule is defined as $\text{support}(\{A, B\}) / \text{support}(\{A\})$, which can be viewed as an estimate

of $\Pr(B|A)$.

- The *lift* for an association rule $A \Rightarrow B$ is defined as $\text{support}(\{A, B\}) / (\text{support}(\{A\}) \times \text{support}(\{B\}))$, which can be viewed as an estimate of the association measure $\Pr(A \text{ and } B) / \Pr(A)\Pr(B)$.

We are interested in getting a relatively high support frequent itemset, which indicates a frequent pattern in the observation; we are also interested in producing association rules with both relatively high confidence and lift, which can reveal potentially important relations between parameters.

We find 60 frequent itemsets of more than two items with support of at least 30%, to list a few:

- {PRC_CHG below (above) median, ESPRD above (below) median}
- {SELLPCT above median, MEANREC above median}
- {MKTCAP below (above) median, INSTOWN_HHI above (below) median}
- {VOLUME_RATIO below (above) median, VOLATILITY above (below) median}
- {LAMBDA below (above) median, MKTCAP above (below) median}

These frequent itemsets are all quite intuitively interpretable:

SELLPCT is positively correlated with MEANREC by definition; MKTCAP is negatively correlated with INSTOWN_HHI, as large capital stocks are harder to be controlled by a few institutions and thus embrace more fragmentation in their stock ownership. Higher VOLUME_RATIO indicates higher liquidity during the crash, and that is correlated with lower VOLATILITY. Also we expect that large capital stocks have smaller price impact reflected by LAMBDA.

We also list a few notable association rules of interest:

- Association rule 1: Support 10.9%, confidence 86.7%, lift 2.12.
 $\{\text{VOLATILITY below median, LAMBDA below median, INSTOWN_PERC below median, EPS_RATIO above median}\} \Rightarrow \{\text{MKTCAP above median, PRIMEXCH=N}\}$
- Association rule 2: Support 10.9%, confidence 77.9%, lift 2.40.
 $\{\text{PRC_CHG below median, LAMBDA below median, VOLATILITY below median}\} \Rightarrow \{\text{MKTCAP above median, INSTOWN_HHI below median}\}$
- Association rule 3: Support 10.1%, confidence 80%, lift 1.97.
 $\{\text{SHOCK above median, ESPRD above median, EPS_RATIO above median}\} \Rightarrow \{\text{PRC_CHG below median, PRIMEXCH=N}\}$

The first association rule is intuitive. It says that a large market capital stock is associated with low volatility, low price impact, low institutional holding percentage, and high EPS_RATIO.²

The second association rule is a little bit surprising. It basically says, for an SP500 stock with large price decline, small price impact, low historical volatility, most of the time it has large market capital and therefore more fragmented stock ownership structure. Why does a stock with small price impact have larger price decline? Our model in Section 3.1 explains this unusual effect. Namely, it was the interaction term of price impact and market capital that affects the price decline during crashes through the information asymmetry effect.

The third association rule somewhat echoes our point. It is the only association rule we find that has an association result including PRC_CHG. It shows the above median SHOCK is associated with larger price decline.

We apply association analysis to the 1987 crash as well. As we do not have MEANREC, SELLPCT, and EPS_RATIO for the 1987 crash, and we exclude the PRIMEXCH parameter (because more than 90% of our sample stocks are listed in NYSE), the sample data set is a matrix of 436 stocks on 12 categorical parameters.

There are 43 frequent itemsets of more than two items with at least 30% support. As those found in the 2010 crash, they are all intuitively interpretable.

We find the following representative association rules that involve PRC_CHG parameter in the 1987 crash:

- Association rule 1: Support 10.6%, confidence 80.7%, lift 2.91.
 $\{\text{VOLATILITY above median, ESPRD above median, INSTOWN_HHI above median, SHOCK above median}\} \Rightarrow \{\text{PRC_CHG below median, LAMBDA above median}\}$
- Association rule 2: Support 11.9%, confidence 82.5%, lift 2.24.
 $\{\text{PRC_CHG below median, ESPRD below median, INSTOWN_HHI below median}\} \Rightarrow \{\text{MKT CAP above median, LAMBDA below median}\}$

Association rule 1 is a representative of rules that contain PRC_CHG as consequence. We find for support of at least 10% and confidence of at least 75%, all the association rules with PRC_CHG in the consequent has SHOCK and VOLATILITY in the antecedent.

Association rule 2 for the 1987 crash resembles the puzzle of association rule 2 for the 2010 crash. It reveals some stocks with larger price decline, smaller effective spread, and more fragmentation in the stock ownership structure, have larger market capital and smaller price impact.

Both of the above observations are in support of our previous model that SHOCK is a important factor that affected PRC_CHG during the 1987 crash. In addition, association rule 1 suggests VOLATILITY may be another important factor that affects PRC_CHG. This

²We can ignore the PRIMEXCH factor as it is a natural result of the high proportion (80%) of SP500 stocks listed in NYSE.

on the other hand echoes that limited risk bearing capacity is another source of temporary price impact.³

4.2 Statistical Model Selection

In this section, we study the predictors of cross-sectional price drawdown of SP500 stocks during the 2010 Flash Crash by best subset selection and the Lasso model selection methods. We identify a set of three predictors out of eight predictors (by combining the information-asymmetry induced price impact factor, the risk-aversion induced price impact factor and six other predictors used in Madhavan (2012)) that have strong effects in predicting cross-sectional price drawdown. We evaluate the robustness of this result by applying model selection methods on an augmented candidate predictors set of betas with respect to well known risk factors. Finally, additional predictors used in corporate default prediction analysis are also evaluated in the model selection.

4.2.1 Model Selection

In the regression setting, the standard linear model:

$$Y = \beta_0 + \sum_{i=1}^p \beta_i X_i + \epsilon \quad (4.1)$$

is commonly used to describe the relationship between a response Y and a set of variables X_1, X_2, \dots, X_p . However, it is often the case that some or many of the variables used in a multiple regression model are in fact not associated with the response. That makes our model unnecessarily complex. Below, we will apply several variable selection approaches that enable us to exclude irrelevant variables in a multivariate regression model. As a result, we can obtain a model that can be better interpreted.

³Although we only include idiosyncratic volatility in this section's analysis, we discuss further the role of risk in general in Section 3.1.5.

Best subset selection

To perform best subset selection, we fit a separate least squares regression for each possible combination of the p predictors. We then look at all of the resulting models, with the goal of identifying the one that is best.

This is described in the following algorithm James et al. (2013):

Best Subset Selection

1. Let M_0 denote the null model, which contains no predictors. This model simply predicts the sample mean of each observation.
 2. For $k = 1, 2, \dots, p$:
 - (a) Fit all $\binom{p}{k}$ models that contain exactly k predictors.
 - (b) Pick the best among $\binom{p}{k}$ models, and call it M_k . Here the best is defined as having the smallest RSS, or equivalently largest R^2 .
 3. Select a single best model from among M_0, \dots, M_p using cross-validated prediction error, $C_p(\text{AIC})$, (BIC), or adjusted R^2 .
-

Forward stepwise selection

A close relative of best subset selection is forward stepwise selection. As a greedy algorithm, forward stepwise selection is computationally more efficient, and is useful when the number of candidate predictors becomes very large. Forward stepwise selection begins with a model containing no predictors, and then adds predictors to the model, one at a time, until all of the predictors are in the model. In particular, at each step the variable that gives the greatest additional improvement to the fit is added to the model.

This is described in the following algorithm James et al. (2013):

Forward Stepwise Selection

1. Let M_0 denote the null model, which contains no predictors.
 2. For $k = 0, 1, \dots, p - 1$:
 - (a) Fit all $p - k$ models that augment the predictors in M_k with one additional predictor.
 - (b) Pick the best among $p - k$ models, and call it M_{k+1} . Here the best is defined as having the smallest RSS, or equivalently largest R^2 .
 3. Select a single best model from among M_0, \dots, M_p using cross-validated prediction error, $C_p(\text{AIC})$, (BIC), or adjusted R^2 .
-

Model selection criteria

The training set mean squared error (MSE) is generally an underestimate of the test MSE (MSE=RSS/n). This is because training error decreases as more variables are included in a regression model, but the test error may not. We need to adjust the training error to

prevent overfitting. C_p (AIC), (BIC), and adjusted R^2 are useful measures for adjusting the training error against the model size. They can be used to select among a set of models with different number of variables.

For a fitted least squares model containing d predictors, the C_p estimate of test MSE is:

$$C_p = \frac{1}{n}(RSS + 2d\hat{\sigma}^2) \quad (4.2)$$

C_p is also proportional to another famous criterion AIC.⁴

BIC is derived from a Bayesian point of view, but ends up looking similar to C_p (AIC).

$$BIC = \frac{1}{n}(RSS + \log(n)d\hat{\sigma}^2) \quad (4.3)$$

Adjusted R^2 is calculated as

$$Adjusted R^2 = 1 - \frac{RSS/(n-d-1)}{TSS/(n-1)} \quad (4.4)$$

where $TSS = \sum(y_i - \bar{y})^2$ and $RSS = \sum(y_i - \hat{y})^2$.

Both C_p and BIC tend to take on a small value for a model with a low test error. But a large value of adjusted R^2 indicates a model with a small test error.

Cross-validation provides a direct estimate of the test error, and makes fewer assumptions about the true underlying model. As we will see in the Lasso, it can also be used in a wider range of model selection tasks, even in cases where it is hard to pinpoint the model degree of freedom or hard to estimate the error variance σ^2 . In a k -fold cross-validation, first we divide the data randomly into k folds equally. Then we train the model on any $k-1$ folds of the data, and calculate the mean test error on the remaining one fold of the data. Finally, the test MSE is estimated by the average of each fold's test MSE.

The Lasso

The Lasso shrinkage method is proposed by Tibshirani (1996). The Lasso minimizes the loss function subject to the sum of the absolute value of the coefficients being less than a constant. As ridge regression, the Lasso solution can yield a reduction in variance at the expenses of a small increase in bias, and consequently can generate more accurate predictions.

$$\hat{\beta}_\lambda^L = \arg \min \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (4.5)$$

$$= RSS + \lambda \sum_{j=1}^p |\beta_j| \quad (4.6)$$

⁴ $AIC = \frac{1}{n\hat{\sigma}^2}(RSS + 2d\hat{\sigma}^2)$

This is equivalent to minimizing the sum of squares with a constraint of the form $\sum_{j=1}^p |\beta_j| \leq s$. For ridge regression, this constraint becomes $\sum_{j=1}^p \beta_j^2 \leq s$. The Lasso and ridge regression do not focus on subsets but define a continuous shrinking operation. But because of the nature of the Lasso's L1 norm constraint, it tends to produce some zero coefficients and hence can perform variable selection.

Therefore the Lasso enjoys some of the favorable properties of both subset selection and shrinkage method in the sense that it produces interpretable models like subset selection and yet exhibits the stability of shrinkage methods like ridge regression. In general, consider a penalty of the form $(\sum_{j=1}^p \beta_j^q)^{1/q}$, then the Lasso uses $q = 1$ and ridge regression has $q = 2$. Subset selection emerges as $q \rightarrow 0$, and the Lasso uses the smallest value of q (closest to subset selection) that yields a convex problem, which facilitates computation. The optimization problem can be solved via standard convex optimizer, Least angle regression (LAR) algorithm and coordinate descent algorithms (Tibshirani 2011).

Implementing the Lasso requires a method for selecting a value for the tuning parameter λ . Cross-validation provides a simple way to tackle this problem.

4.2.2 Data

The data can be divided into two parts in general: predictors and response, both are at firm level.

There are three set of predictors.

Set 1: Model based predictors:

- SHOCK: $\lambda_i \gamma_i$, information asymmetry induced price impact, scaled by 10^{-9} .
- RISK: $(\Sigma \gamma)_i$, limited risk-bearing capacity induced price impact, scaled by 10^{-2} .

We already have this set of predictors from previous session.

Set 2: Predictors from Madhavan (2012)⁵

- Q_HHI: H_i^q , average of daily quote Herfindahl index.
- V_HHI: H_i^v , average of daily volume Herfindahl index.
- LADV: log of average daily volume in millions of dollars.
- VOLATILITY: average of the daily standard deviation of five-minute return intervals scaled by 10^{-6} in the period 1:30 pm-4:00 pm ET.

⁵Except for INVP, all the predictors are calculated from control period 4/7/2010-5/5/2010. Specifically, Q_HHI, V_HHI and VOLATILITY are calculated in intraday window 1:30 pm-4:00 pm ET. We exclude another predictor ETP used in Madhavan (2012) because it does not apply to our setting.

- INVP: price inverse using the open price on 5/6/2010.
- ISO: intermarket Sweep Order Activity measured by the dollar weighted proportion of volume accounted by Condition Code F orders.

We derive this set of predictors following Madhavan (2012) closely with the TAQ database.

Set 3: Beta Predictors

In the capital asset pricing model (CAPM), it is well known that a risky asset's expected return is driven by risk. In addition to the market risk factor, Fama and French (1993); Carhart (1997) further identify three risk factors (size, value, and momentum) that affect the expected return of a risky asset. In a recent paper, Pelger (2015) identifies four continuous factors that are highly persistent during 2007-2012 based on high-frequency data. The top three of these factors can be replicated by an equally weighted market, an oil and a finance portfolio.⁶

We use CRSP daily excess return data and the Fama-French-Carhart factors daily data from 01/02/2009 to 04/30/2010 from WRDS to get the following betas for each SP500 stocks: β_{mkt} , β_{smb} , β_{hml} , β_{umd} , β_{ewmkt} , β_{oil} , $\beta_{finance}$.

The first four betas are based on Fama-French-Carhart factor data, and the next three betas are based on equal weighted market, oil, and finance portfolios constructed by SP500 stocks according to Pelger (2015).⁷

Set 4: Predictors from US corporate default analysis

We also include a set of 24 predictors that are used in literature to predict US corporate default event, as both default and market crash are related to the stability of the market.

In order to construct default predictors, we combine yearly accounting data from the COMPUSTAT database with monthly equity market data from the CRSP database. Abbreviation and data source descriptions are in Table A.3. As in Wu (2011), the 24 candidate predictors are grouped into eight categories that capture default causal factors at the individual firm level: Size, Capital structure, Growth, Profitability, Debt Coverage, Liquidity, Business Operations, Market Performance. We construct measures of firm's size as: $\log(\text{Sale})$, $\log(\text{TA})$ and ME ; firm's capital structure is characterized by: ME/BD , ME/TA , TL/TA and SD/BD ; Sale Gth , NI Gth and OM CH represent firm's growth traits; RE/TA , EBIT/TA and NI/TA describe firm's profitability; ICR and NI/TL gauge firm's debt coverage; WC/TA , CR , CH/TA , QR are liquidity measures; INVT/SALE , AR/SALE and SALE/TA depict business operations; market performance is characterized by SRT and SIGMA .

Table 4.1 shows the summary statistics for the four sets of predictors.

⁶We do not include the fourth continuous factor (electricity factor) and the jump market factor in this analysis because their effects are weaker and less robust according to Pelger (2015).

⁷See definition of oil and finance stocks with respect to their SIC code in Pelger (2015).

Category	Predictor	SP500 Firms				
		Mean	Median	Std.Dev	5%	95%
Set 1: Model based	SHOCK(e-12)	1.475	1.206	1.192	0.145	3.677
	RISK(e-05)	9.144	8.689	3.569	4.216	15.787
Set 2: Madhavan (2012)	Q_HHI	0.177	0.167	0.047	0.116	0.269
	V_HHI	0.245	0.237	0.029	0.210	0.304
	LADV	4.929	4.836	0.952	3.497	6.558
	VOLATILITY	86.139	75.756	48.469	39.152	161.481
	INVP	0.037	0.026	0.040	0.011	0.091
	ISO	0.281	0.328	0.144	0	0.428
Set 3: Betas	β_{mkt}	1.210	1.072	0.637	0.410	2.526
Fama-French-Carhart	β_{hml}	1.545	1.217	1.127	0.321	3.877
	β_{umd}	-0.775	-0.643	0.558	-1.971	-0.118
	β_{smb}	0.646	0.619	0.485	-0.004	1.545
Pelger (2015)	β_{ewmkt}	1.000	0.884	0.542	0.319	2.105
	$\beta_{finance}$	0.466	0.385	0.311	0.123	1.139
	β_{oil}	0.666	0.588	0.364	0.188	1.366
Set 4: Wu (2011)	log(Sale)	8.917	8.867	1.206	7.057	11.132
Size	log(TA)	9.515	9.352	1.351	7.602	12.019
	ME(e09)	22.146	10.164	37.101	2.819	85.609
Capital Structure	ME/BD	30.783	4.005	166.516	0.605	54.786
	ME/TA	1.275	0.980	1.095	0.116	3.288
	TL/TA	0.605	0.607	0.209	0.257	0.921
	SD/BD	0.134	0.070	0.194	0	0.577
Growth	SALE Gth	-0.042	-0.060	0.538	-0.365	0.181
	NI Gth	-0.272	-0.168	3.340	-2.248	1.045
	OM CH	0.206	0.194	0.155	0.021	0.474
Profitability	RE/TA	0.159	0.234	1.926	-0.174	0.838
	EBIT/TA	0.092	0.080	0.083	-0.015	0.228
	NI/TA	0.047	0.043	0.071	-0.044	0.158
Debt Coverage	ICR	23.290	6.275	72.566	-0.748	83.124
	NI/TL	0.112	0.073	0.208	-0.075	0.427
Liquidity	WC/TA	0.154	0.115	0.159	-0.044	0.472
	CR	1.931	1.583	1.261	0.737	4.153
	CH/TA	0.102	0.079	0.097	0.004	0.287
	QR	1.540	1.211	1.138	0.490	3.564
Business Operations	INVT/SALE	0.151	0.075	0.650	0	0.301
	AR/SALE	0.708	0.145	2.299	0.016	4.464
	SALE/TA	0.790	0.605	0.685	0.068	2.188
Market Performance	SRT	0.526	0.443	0.481	-0.022	1.321
	SIGMA	0.087	0.074	0.054	0.039	0.170

Table 4.1: Summary Statistics

The response variable of our data set is the maximum price drawdown, which is the same as defined in Section 3.13. Also, we will focus on the same filtered set of stocks (457 out of 500) as in our previous linear regression. We use this definition of response and the same set of stocks as before, because: 1. This facilitates us to compare the newly added predictors with the previous two predictors; 2. Madhavan (2012)'s definition of maximum price drawdown does not enable us to get any statistical significance for predictors proposed in his paper, if we run regression only on the SP500 stocks; 3. As Madhavan (2012) is about the 2010 Flash Crash, and the related predictors are derived from TAQ data, which are not available for the 1987 crash, we focus on the 2010 Flash Crash but not the 1987 crash.

4.2.3 Model Selection Results

We start with eight Set 1 and Set 2 candidate predictors.

For best subset selection method, the Adjusted R^2 , C_p , and 10-fold cross-validation all select five predictors as the best model (See Figure 4.10): SHOCK, RISK, Q_HHI, V_HHI and INVP. The BIC tends to impose heavier penalty on model complexity and it selects the three-predictor model: SHOCK, RISK, Q_HHI.

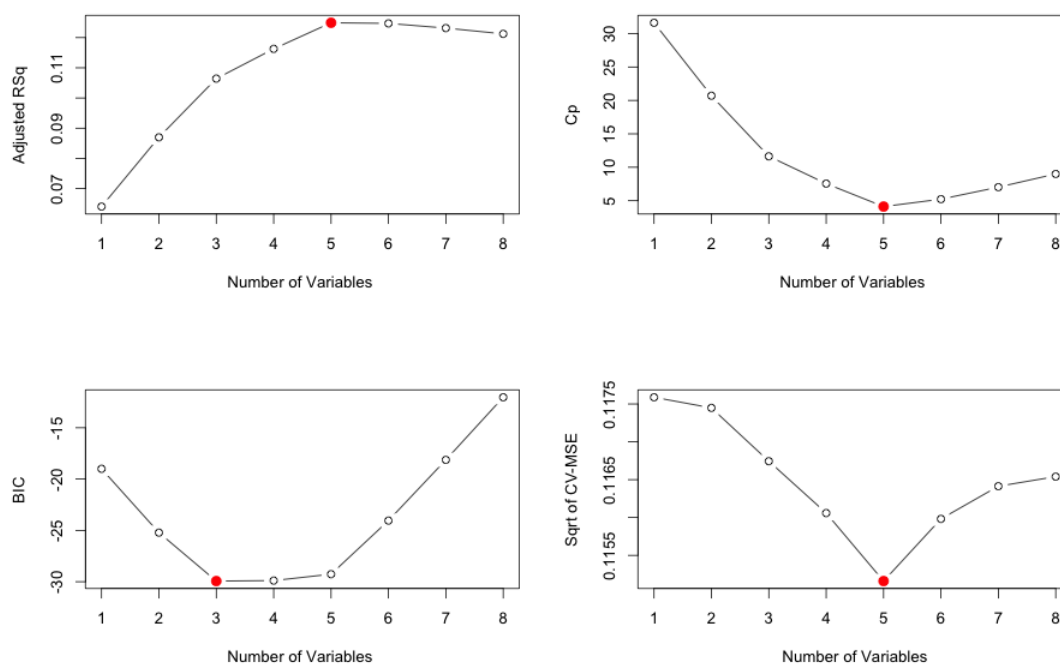


Figure 4.10: Best subset selection for 8 predictors.

For the Lasso method, we use 10-fold cross-validation to choose the best tuning parameter and estimate the test error. It appears that RISK, SHOCK and Q_HHI remain top three predictors and the best model selected via cross-validation in the Lasso has five predictors: SHOCK, RISK, Q_HHI, INVP and ISO.

Figure 4.11 shows the result of the Lasso method.

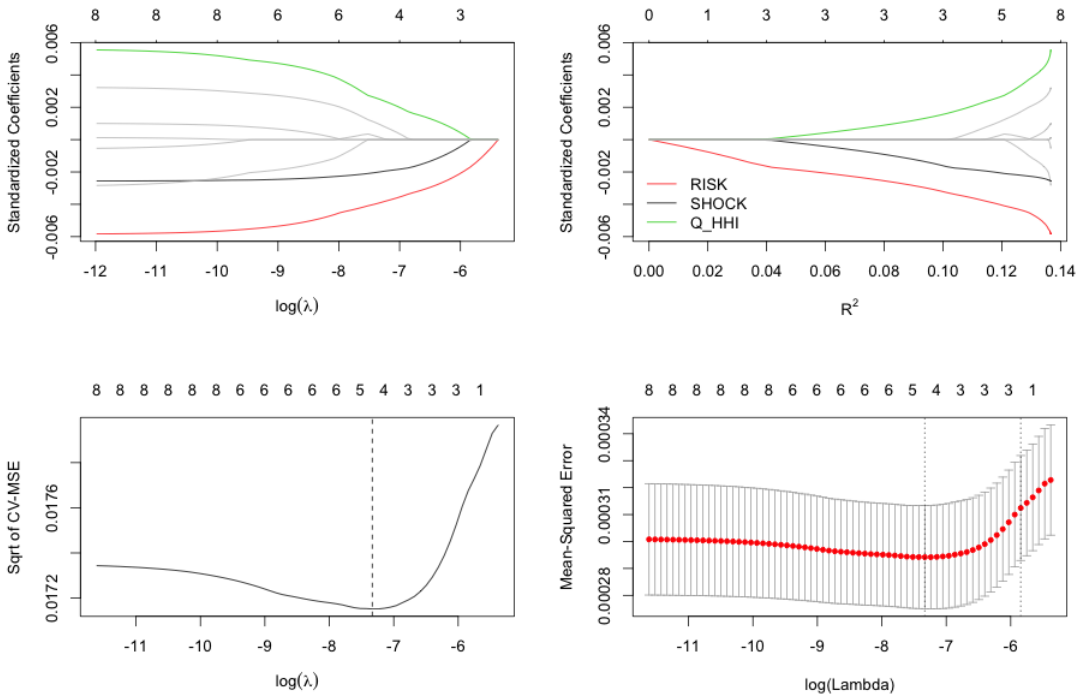


Figure 4.11: The Lasso for 8 predictors. Top left: Standardized coefficients of 8 predictors against $\log(\lambda)$; Top right: Standardized coefficients of 8 predictors against R^2 explained; Bottom left: best tuning parameter selection with minimum square root cross-validation MSE; Bottom right: best tuning parameter selection with one-standard-error rule (MSE with one standard error band at each point).

We can see from Figure 4.11 that the top three predictors: RISK, SHOCK, Q_HHI explain the most of the R^2 . Also if we apply the *one-standard-error rule* that picks the smallest model within one standard error of the MSE of the best model to prevent overfitting for the training data⁸, we get a model with three predictors: RISK, SHOCK and Q_HHI.

Table 4.2 lists the predictor selection procedure of best subset selection and the Lasso method for 8 candidate predictors. RISK and SHOCK are chosen first by both best subset selection method and the Lasso.

⁸ The best λ according to the one-standard-error rule is shown by the right vertical dash line in the bottom right window in Figure 4.11.

Selection Sequence	Best subset selection	The Lasso
1	RISK	RISK
2	SHOCK	SHOCK
3	Q_HHI	Q_HHI
4	INVP	INVP
5	V_HHI	ISO

Table 4.2: Selection Sequence for 8 candidate predictors.

4.2.4 Comparison with Betas of Risk Factors

Putting the Set 3 betas into the previous pool of predictors, we get a total of 15 predictors. We then run best subset selection and the Lasso method with the 15 predictors.

With 10-fold cross-validation, the best model selected by best subset selection method has 8 predictors (See Figure 4.12): SHOCK, RISK, Q_HHI, INVP, β_{hml} , β_{umd} , β_{smb} and β_{oil} .

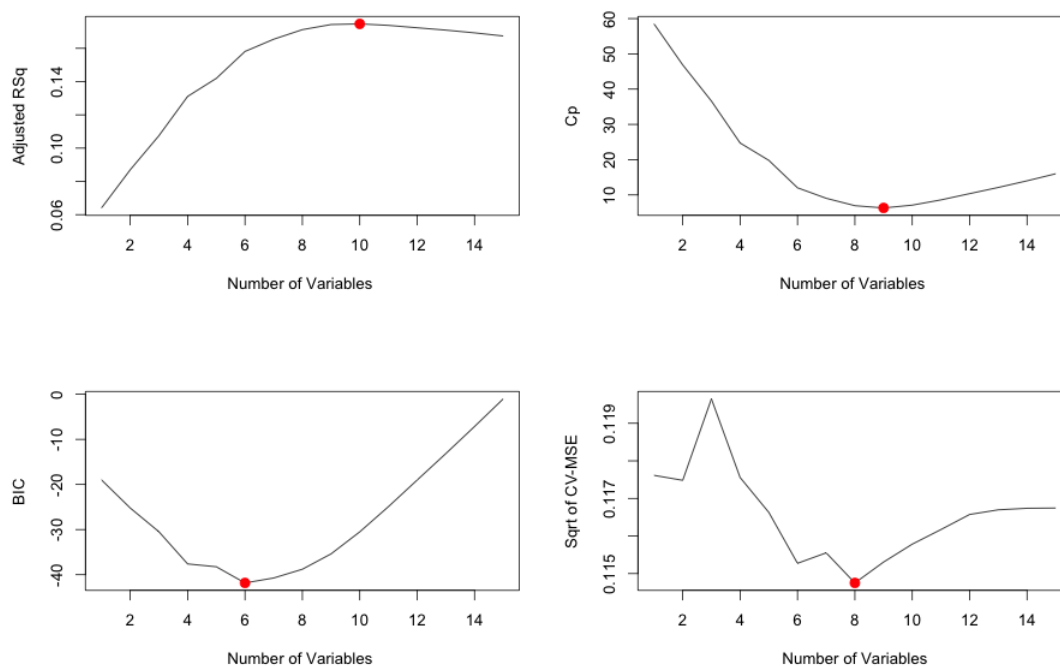


Figure 4.12: Best subset selection for 15 predictors.

For the Lasso method, we use 10-fold cross-validation to choose the best tuning parameter

and estimate the test error. The best model selected via cross-validation has all but two predictors: β_{mkt} and β_{ewmkt} . The best model selected via *one-standard-error rule* for the Lasso has four predictors: SHOCK, RISK, Q_HHI, β_{smb} .

Figure 4.13 shows the result of the Lasso method. We can see that the cross-validation MSEs of the best model selected and the best model selected via *one-standard-error rule* are very close. This suggests that it is sufficient to have the model with a smaller set of predictors.

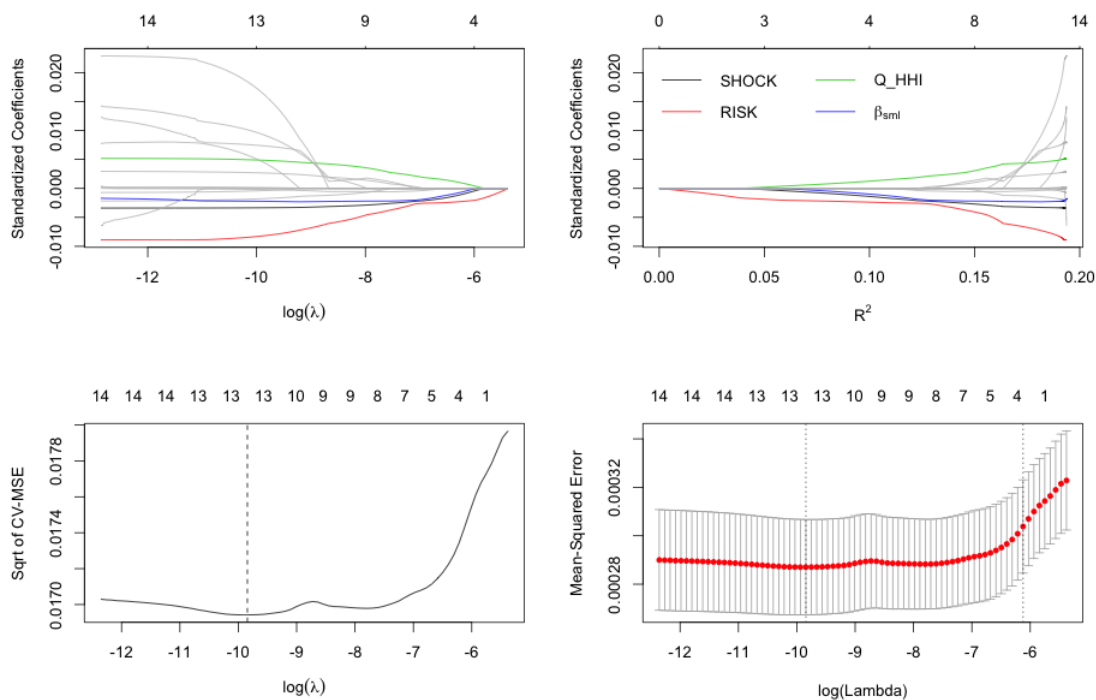


Figure 4.13: The Lasso for 15 predictors. Top left: Standardized coefficients of 15 predictors against $\log(\lambda)$; Top right: Standardized coefficients of 15 predictors against R^2 explained; Bottom left: best tuning parameter selection with minimum square root cross-validation MSE; Bottom right: best tuning parameter selection with one-standard-error rule (MSE with one standard error band at each point).

Table 4.3 lists the predictor selection procedure of best subset selection and the Lasso method for 15 candidate predictors. We can see RISK and SHOCK consistently remain on the top, indicating they are more important predictors compared to betas. Q_HHI remains on the top 4 predictors as well.

Selection Sequence	Best subset selection	The Lasso
1	RISK	RISK
2	SHOCK	SHOCK
3	β_{oil}	Q_HHI
4	Q_HHI	β_{smb}
5	β_{smb}	-
6	$\beta_{hml}, \beta_{umd}, (\beta_{smb})$	-
7	INVP	-
8	β_{smb}	-

Table 4.3: Selection Sequence for 15 candidate predictors. The sequence is not necessarily nested for best subset selection, here β_{smb} was in the best five-predictor set, but was dropped in the best six-predictor set.

4.2.5 Comparison with Corporate Default Predictors

In this section, in addition to Set 2 and Set 3 predictors, we put in Set 4 predictors from the US corporate default analysis to compare their effect on the cross-sectional price drawdown during the stock market crashes, as both events test the vulnerability of a company under unfavorable market environment.

However as corporate default predictors are usually only available for poor grade stocks, we cannot get some corporate default predictors for all the SP500 stocks. After filtering out stocks with missing values for 24 added predictors, we have 345 stocks remaining.⁹ As the total number of predictors increases to 39, here we use forward stepwise selection (which is more computationally efficient) and the Lasso method to perform model selection.

Forward stepwise selection method selects a 18-predictor model by 10-fold cross-validation: SHOCK, RISK, Q_HHI, V_HHI, LADV, INVP, $\log(\text{TA})$, ME/BD, RE/TA, WC/TA, AR/SALE, SRT, SIGMA, ME, β_{mkt} , β_{umd} , $\beta_{finance}$, β_{oil} . The best model selected under 10-fold cross-validation by the Lasso method has 16 predictors. The model with *one-standard-error rule* selected by the Lasso method has 7 predictors: SHOCK, RISK, Q_HHI, $\log(\text{TA})$, WC/TA, SRT, ME.

Table 4.4 compares top 10 predictors selection procedure of forward stepwise selection and the Lasso method out of 39 candidate predictors. Compared to Table 4.3 of 15 predictors selection, we can see SHOCK and RISK are still selected first. Also we notice both forward stepwise selection and the Lasso method select three corporate default predictors: WC/TA (liquidity), $\log(\text{TA})$ (size) and SRT (market performance), which are also among the eight predictors picked by corporate default analysis (Wu 2011). They can be viewed as important

⁹The reduction of number of stocks is mainly due to the missing value of ACT (76 symbols missing), LCT (75 symbols missing) and XINT (38 symbols missing) that construct the default predictors.

indicators of the stability of a firm in hostile market environments, be it economic recession, or market crash.

Selection Sequence	Forward stepwise selection	The Lasso
1	SHOCK	SHOCK
2	RISK	RISK
3	WC/TA(Liquidity)	SRT(Market Performance)
4	log(TA)(Size)	log(TA)(Size)
5	SRT(Market Performance)	WC/TA(Liquidity)
6	Q_HHI	Q_HHI
7	SIGMA(Market Performance)	ME(Size)
8	β_{oil}	-
9	β_{umd}	-
10	$\beta_{finance}$	-

Table 4.4: Selection Sequence for 39 candidate predictors.

The result of forward stepwise selection is shown in Figure 4.14.

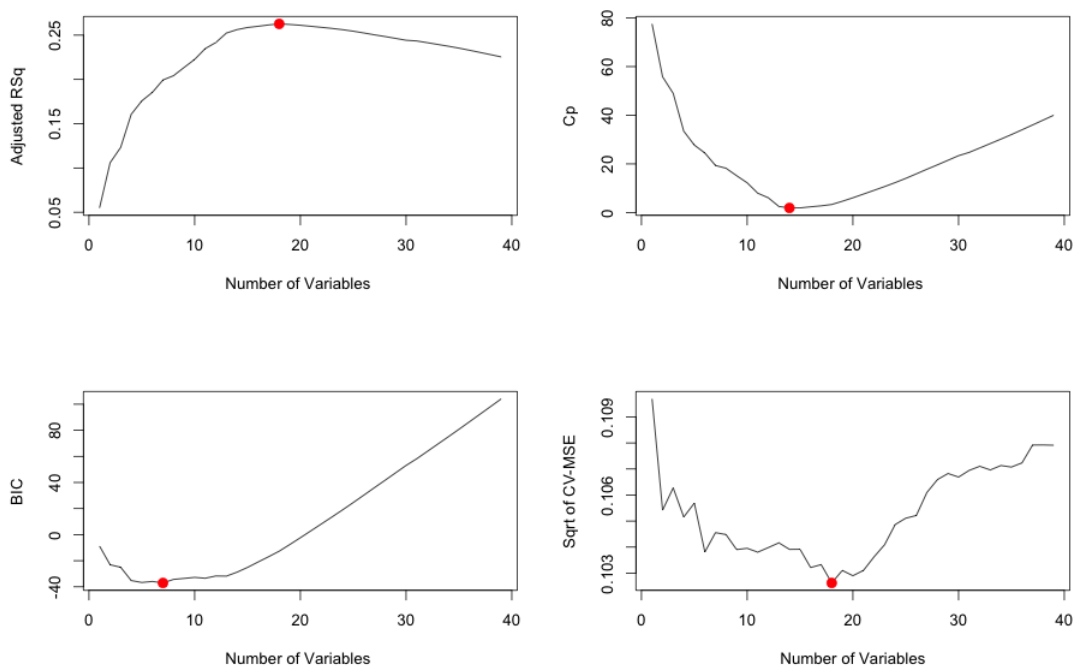


Figure 4.14: Forward stepwise selection for 39 predictors.

The result of the Lasso method is shown in Figure 4.15.

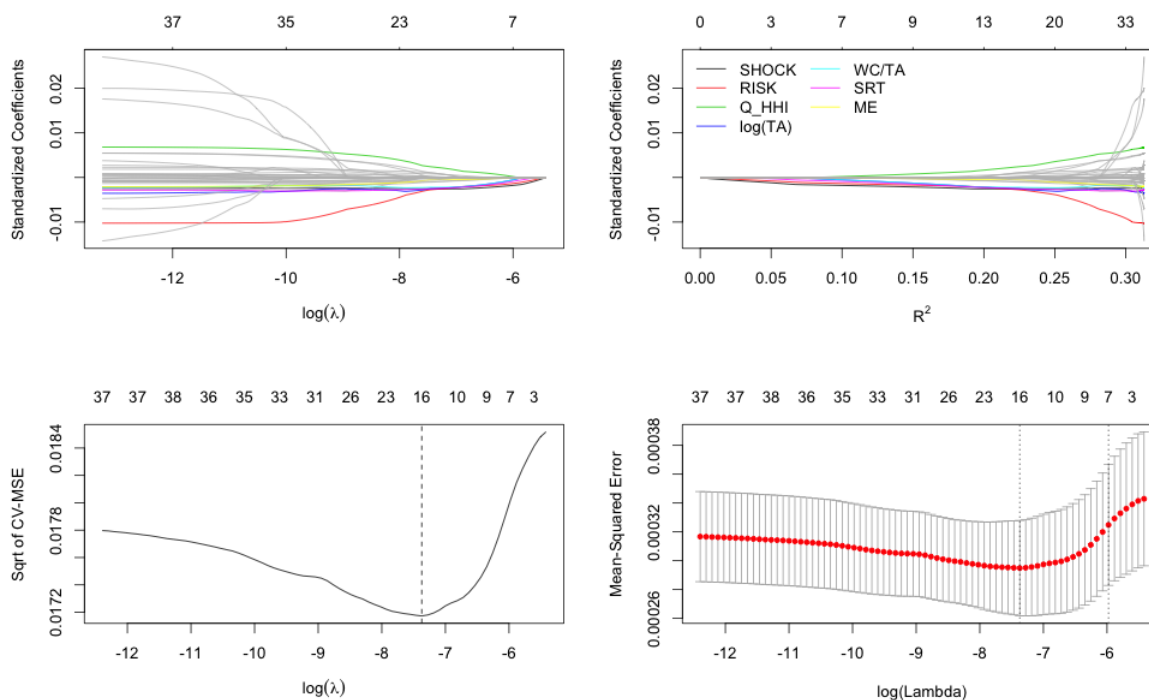


Figure 4.15: The Lasso for 39 predictors. Top left: Standardized coefficients of 39 predictors against $\log(\lambda)$; Top right: Standardized coefficients of 39 predictors against R^2 explained; Bottom left: best tuning parameter selection with minimum square root cross-validation MSE; Bottom right: best tuning parameter selection with one-standard-error rule (MSE with one standard error band at each point).

4.2.6 Discussion

From the above statistical analysis, we can see for the 2010 Flash Crash, the cross-sectional price drawdown of SP500 stocks is not related to their exposure (betas) to certain well-known risk factors, such as size, value or momentum effect. Pelger (2015) uses 15-min intraday interval return to identify major persistent economic factors during 2003-2012, which covers the 2010 Flash Crash. However, the 15-min sampling frequency is too low to find the hidden market microstructure effect, which can be otherwise important in explaining stock performance during crashes or some temporary events. The model selection results of Set 1-3 and Set 1-4 predictors demonstrate this point, where we can see the SHOCK predictor (derived by sampling at 15-second time intervals) is more important compared to betas.

According to the statistical model selection result, we can see RISK and SHOCK remain top predictors of SP500 stock cross-sectional price drawdown when we compare them with three other set of predictors. We find in addition to the two model based predictors, Q_HHI is another important predictor associated with the price drawdown of SP500 stocks. For those stocks that we can construct corporate default predictors, WE/TA, log(TA), SRT are useful in prediction.

We run a linear regression on the six common predictors that are selected by both forward stepwise selection and the Lasso method. The standard errors of regression coefficients are estimated by bootstrapping 1000 samples and are shown in the parentheses. The results are shown in Table 4.5.

	<i>Dependent variable:</i>
	Price drawdown
SHOCK(e09)	−2.923*** (1.020)
RISK(e01)	−7.998*** (2.859)
Q_HHI	0.048** (0.019)
SRT	−0.007*** (0.003)
WC/TA	−0.029*** (0.006)
log(TA)	−0.003*** (0.001)
Constant	−0.012 (0.013)
Observations	345
R ²	0.200
Adjusted R ²	0.185
Residual Std. Error	0.017 (df = 338)
F Statistic	14.050*** (df = 6; 338)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 4.5: Model selection results for 345 stocks in regression. The standard errors of regression coefficients are estimated by bootstrapping 1000 samples and are shown in the parentheses.

We can see that the coefficients of RISK and SHOCK are similar to previous regression in Table 3.3. The coefficient of Q_HHI is positive, meaning the larger the quote fragmentation (the smaller the Q_HHI), the larger the maximum price drawdown. This is consistent with the finding in Madhavan (2012). The signs of the coefficients of the three corporate default predictors, however, are quite counter-intuitive compared to Wu (2011). This may imply that those predictors are not the major ones that drive the cross-sectional price drawdown of SP500 stocks.

Table 4.6 compares the model selection results for different sets of predictors. The Lasso method produces uniformly smaller cross-validation root MSE compared to subset selection

methods. This is because the Lasso not only performs variable selection, it also tends to shrink the coefficients towards zero relative to the least square estimates. This will effectively reduce the variance of the predictions at the expense of a slight increase in bias. As the test MSE is a function of the variance plus the squared bias, a shrinkage method like the Lasso can achieve significantly lower cross-validation MSE.

Secondly, we also notice that the prediction error by including extra corporate default predictors (Set 1-4) increases prediction error. The above observations make us believe that RISK, SHOCK and Q_HHI are most important predictors for SP500 stocks in general. And we could avoid potential overfitting by only include these three predictors.

Model Selection		Best subset	Forward stepwise	The Lasso
Model selected via CV	Set 1-2	5 predictors	-	3 predictors(5)
	Set 1-3	8 predictors	-	4 predictors(13)
	Set 1-4	-	18 predictors	7 predictors(16)
Top 5 predictors	Set 1-2	RISK,SHOCK,Q_HHI, INVP,V_HHI	-	RISK,SHOCK,Q_HHI (INVP, ISO)
	Set 1-3	RISK,SHOCK, β_{oil} , Q_HHI, β_{smb}	-	RISK,SHOCK,Q_HHI, β_{smb} (LADV)
	Set 1-4	-	SHOCK,RISK,WC/TA, log(TA),SRT	SHOCK,RISK,SRT, log(TA),WC/TA
Sqrt of CV-MSE	Set 1-2	0.1152	-	0.0177(0.0172)
	Set 1-3	0.1148	-	0.0174(0.0169)
	Set 1-4	-	0.1026	0.0180(0.0172)

Table 4.6: A comparison of model selection result with different data sets. Both Set 1-2 and Set 1-3 have 457 SP500 stocks, Set 1-4 has 345 stocks. We show the result of the Lasso method that applies one-standard-error rule. The results of the best model of selected by the Lasso method are shown in quote.

Therefore we also run a linear regression on the three predictors only. The standard errors of regression coefficients are estimated by bootstrapping 1000 samples and are shown in the parentheses. The results are shown in Table 4.7. Again, all the estimated coefficients are consistent with findings in Table 3.3 and Madhavan (2012).

	<i>Dependent variable:</i>
	Price drawdown
SHOCK(e09)	−2.312*** (0.711)
RISK(e01)	−11.975*** (25.132)
Q_HHI	0.057*** (0.016)
Constant	−0.050*** (0.004)
Observations	457
R ²	0.112
Adjusted R ²	0.106
Residual Std. Error	0.017 (df = 453)
F Statistic	19.102*** (df = 3; 453)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 4.7: Model selection results for 457 stocks in regression. The standard errors of regression coefficients are estimated by bootstrapping 1000 samples and are shown in the parentheses.

4.3 Other statistical methods for 39 predictors

Although not for variable selection, we use ridge, principal component regression (PCR) and partial least squares (PLS) method to visualize how the relative importance of predictors change through the evolution of their standardized coefficients. Economically, this enables us to see how large one unit change of standardized predictors can affect the best prediction (chosen by smallest cross-validation MSE).

We also try to apply support vector regression (SVR) with feature selection algorithm on this data set. We present the result in this section.

Ridge regression

As mentioned in the previous section, ridge regression adds a L2 penalty to the RSS:

$$\hat{\beta}_\lambda^L = \arg \min \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \quad (4.7)$$

$$= RSS + \lambda \sum_{j=1}^p \beta_j^2 \quad (4.8)$$

This has the following analytical solution:

$$\hat{\beta}^{ridge} = (X^T X + \lambda I)^{-1} X^T y \quad (4.9)$$

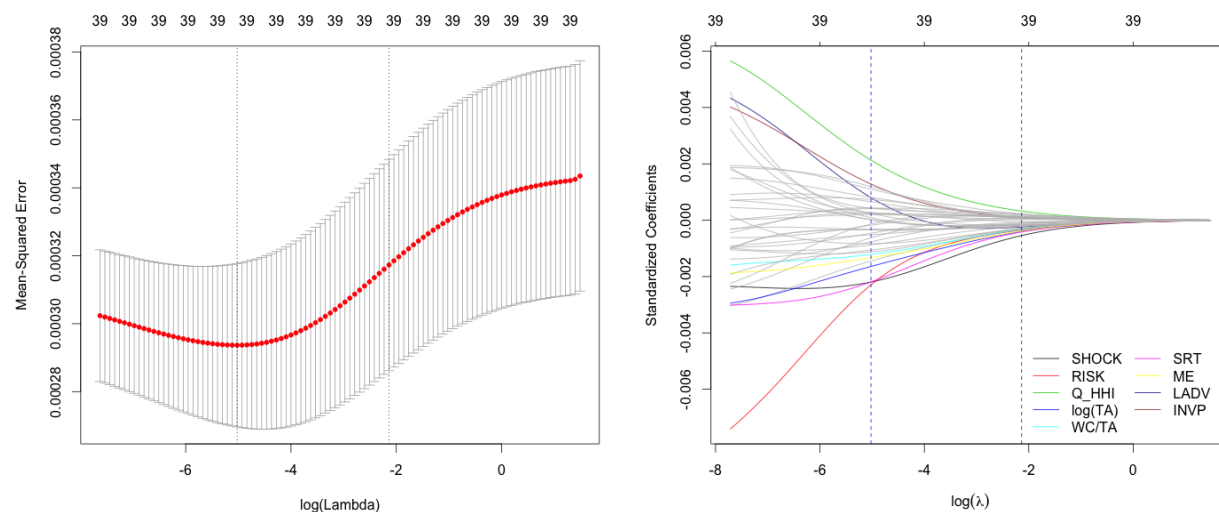


Figure 4.16: Ridge regression for 39 predictors. Left: cross-validation MSE and best tuning parameter; Right: Standardized coefficients of 39 predictors against $\log(\lambda)$, blue dotted lines are best tuning parameter selection with minimum CV-MSE and one-standard-error rule respectively.

Figure 4.16 shows the result of ridge regression. On the right panel, we can see the standardized coefficients continuously shrink with respect to the tuning parameter λ . Unlike the Lasso, ridge regression does not yield zero standardized coefficients. The absolute standardized coefficients of SHOCK, RISK, Q_HHI, WC/TA, log(TA), SRT are relatively larger than other predictors at the best tuning parameter level via 10-fold cross-validation. And at the biggest λ of the *one-standard-error rule*, absolute value of standardized coefficients of SHOCK, RISK, Q_HHI are larger than other predictors.

Principal component regression

The principal components regression (PCR) approach involves constructing the first M principal components, Z_1, \dots, Z_M , and then using these components as the predictors in a linear regression model that is fit using least squares. If a small number of principal components can explain most of the variability in the data, as well as the relationship with the response, in other words, if the directions in which X_1, \dots, X_p show the most variation are the directions that are associated with Y , then PCR will perform well. As a result, we can choose a smaller set of principal features to predict response Y , which is useful in dimension reduction (James et al. 2013).

As in principal component analysis (PCA), we can derive the principal components of

the data matrix ($n \times p$) X via singular vector decomposition (SVD).

$$X = UDV^T \quad (4.10)$$

Here U and V are $n \times p$ and $p \times p$ orthogonal matrices. D is a $p \times p$ diagonal matrix, with diagonal entries $d_1 \geq d_2 \geq \dots \geq d_p \geq 0$ called the singular values of X .

As the sample covariance matrix is given by $S = X^T X/n$, and we have:

$$X^T X = VD^2V^T \quad (4.11)$$

which is the eigen decomposition of $X^T X$ and of S , up to a factor n . The columns of V : V_j are eigenvectors of $X^T X$ and are called the principal components directions of X .

Then we can derive *principal components* Z_j as

$$Z_j = XV_j = U_j d_j \quad (4.12)$$

where the U_j are the columns of U .

By definition, the first principal component Z_1 has the largest sample variance:

$$\text{var}(Z_1) = \frac{d_1^2}{n}$$

In PCR, we regress y on Z_1, Z_2, \dots, Z_M for some $M \leq p$, and use cross-validation to choose the optimal M . If $M = p$, we would just get back the usual least squares estimates, since the columns of $Z = UD$ span the column space of X . For $M < p$, we get a reduced regression.

Analytically, since the Z_m are orthogonal, we have:

$$\hat{\beta}^{pcr}(M) = \sum_{m=1}^M \hat{\theta}_m V_m \quad (4.13)$$

where

$$\hat{\theta}_m = \langle Z_m, Y \rangle / \langle Z_m, Z_m \rangle \quad (4.14)$$

Basically $\hat{\theta}_m$ are the regression coefficients of response Y on the M principal components, and $\hat{\beta}^{pcr}(M)$ are the regression coefficients of response Y on the X_j .

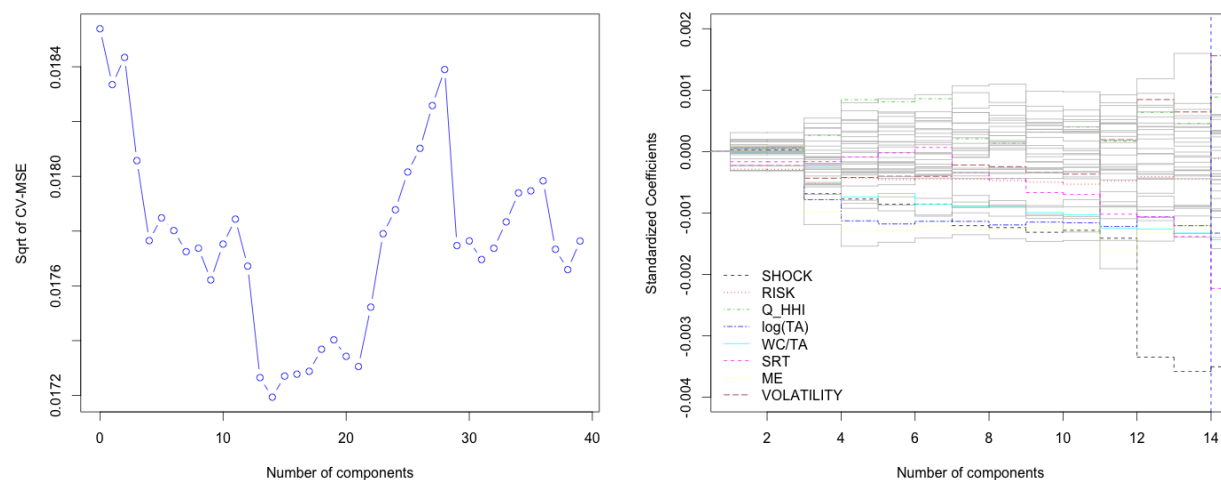


Figure 4.17: Principal component regression for 39 predictors. Left: square root cross-validation MSE and best number of components is 14; Right: Standardized coefficients of 39 predictors against number of components.

Figure 4.17 shows that the best PCR (according to 10-fold cross-validation MSE) for our data set is of 14 principal components. And after we recover the standardized coefficients of original features (the right panel), we can see only SHOCK and SRT have significantly larger absolute standardized coefficients.

Partial least squares

Partial least squares (PLS) is a supervised alternative to PCR. Unlike PCR, PLS identifies a new set of features Z_1, \dots, Z_M in a supervised way, that is, it makes use of the response Y to identify new features that not only approximate the old features well, but also are related to the response.

PLS computes $\hat{\phi}_{j1} = \langle X_j, Y \rangle$ for each j . The first direction $Z_1 = \sum_{j=1}^p \hat{\phi}_{j1} X_j$. For the second direction, PLS orthogonalize X_1, \dots, X_p with respect to Z_1 and computes Z_2 using the orthogonalized X_j . By doing this iteratively, M new features can be constructed.

The following PLS algorithm is described in Hastie, Tibshirani, and Friedman (2008):

Partial Least Squares

1. Standardize X_j to have mean zero and variance one. Set $\hat{Y}^{(0)} = \bar{Y}1$, and $X_j^{(0)} = X_j$, $j = 1, \dots, p$.
 2. For $m = 1, 2, \dots, p$
 - (a) $Z_m = \sum_{j=1}^p \hat{\phi}_{jm} X_j^{(m-1)}$, where $\hat{\phi}_{jm} = \langle X_j^{(m-1)}, Y \rangle$.
 - (b) $\hat{\theta}_m = \langle Z_m, Y \rangle / \langle Z_m, Z_m \rangle$.
 - (c) $\hat{Y}^{(m)} = \hat{Y}^{(m-1)} + \hat{\theta}_m Z_m$.
 - (d) Orthogonalize each $X_j^{(m-1)}$ with respect to Z_m : $X_j^{(m)} = X_j^{(m-1)} - \frac{\langle Z_m, X_j^{(m-1)} \rangle Z_m}{\langle Z_m, Z_m \rangle}$, $j = 1, 2, \dots, p$.
 3. Output the sequence of fitted vectors $\{\hat{Y}^{(m)}\}_1^p$. Since the $\{Z_j\}_1^m$ are linear in the original X_j by construction, so is $\hat{Y}^{(m)} = X\hat{\beta}^{pls}(m)$. These linear coefficients can be recovered from the sequence of PLS transformations.
-

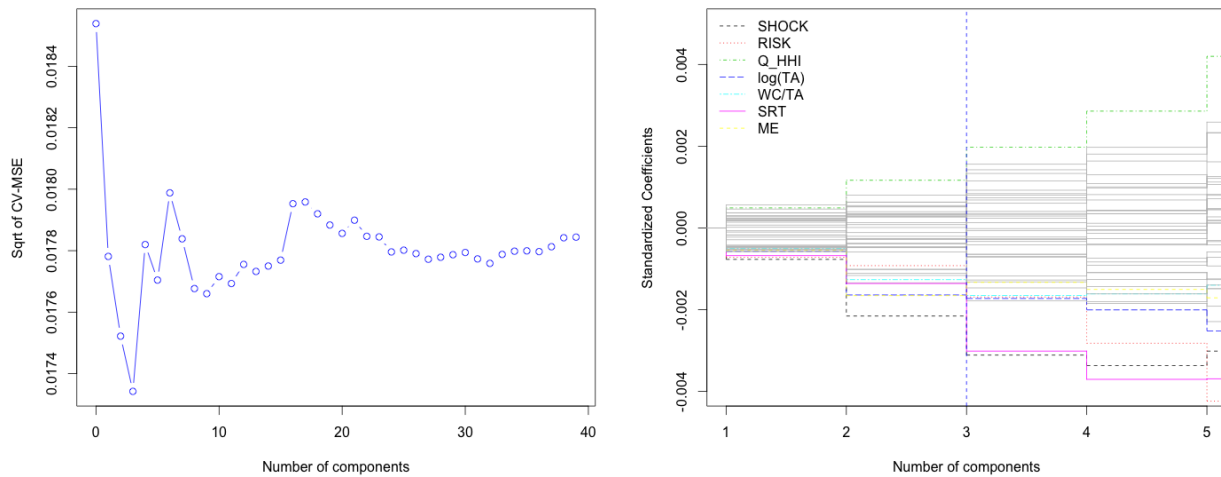


Figure 4.18: Partial least squares for 39 predictors. Left: square root cross-validation MSE and best number of components is 2; Right: Standardized coefficients of 39 predictors against number of components.

As shown in Figure 4.18, PLS gives similar result as ridge regression. The best number of components selected via 10-fold cross-validation is three. We can see the absolute standardized coefficients of SHOCK, RISK, Q_HHI, WC/TA, log(TA), SRT are relatively larger than other predictors at the best tuning parameter level.

Support vector regression

Support vector regression (SVR) is an application of support vector machine (SVM) for regression. Compared to ordinary least squares and its variants like the Lasso and ridge regression, that try to obtain the best fit through minimization of squared error loss function (i.e. $L(y, f(x)) = (y - f(x))^2$) with some regularization terms, SVR uses another form of loss function called ϵ -insensitive loss proposed by Vapnik (1995):

$$L_\epsilon(y, f(x)) = \max\{0, |y - f(x)| - \epsilon\} \quad (4.15)$$

SVR estimates β by

$$\hat{\beta}_{\epsilon, C} = \arg \min \sum_{i=1}^n L_\epsilon(y_i - f(x_i)) + \frac{1}{2C} \|\beta\|^2 \quad (4.16)$$

This is equivalent to the following minimization problem:

$$\min_{\beta, \beta_0, \xi_i, \xi_i^*} \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (4.17)$$

$$\text{subject to } \begin{cases} y_i - f(x_i) \leq \epsilon + \xi_i^*, \\ f(x_i) - y_i \leq \epsilon + \xi_i, \\ \xi_i, \xi_i^* \geq 0. \end{cases}$$

For $f(x) = x^T \beta + \beta_0$, the corresponding Lagrangian is:

$$\begin{aligned} L = & \frac{1}{2} \beta^T \beta + C \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n \alpha_i^* [y_i - \beta^T x_i - \beta_0 + \epsilon + \xi_i^*] \\ & - \sum_{i=1}^n \alpha_i [\beta^T x_i + \beta_0 - y_i + \epsilon + \xi_i] - \sum_{i=1}^n (\gamma_i^* \xi_i^* + \gamma_i \xi_i) \end{aligned} \quad (4.18)$$

where $\gamma_i, \gamma_i^*, \alpha_i, \alpha_i^*$ are Lagrange multipliers.

By differentiating L with respect to β, β_0 , and ξ , we have the resulting equivalent maximization of the following dual objective function:

$$\max_{\alpha_i, \alpha_i^*} -\epsilon \sum_{i=1}^n (\alpha_i^* + \alpha_i) + \sum_{i=1}^n y_i (\alpha_i^* - \alpha_i) - \frac{1}{2} \sum_{i,j=1}^n (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) \langle x_i, x_j \rangle \quad (4.19)$$

$$\text{subject to } \begin{cases} 0 \leq \alpha_i, \alpha_i^* \leq C, \\ \sum_{i=1}^n (\alpha_i^* - \alpha_i) = 0, \\ \alpha_i \alpha_i^* = 0. \end{cases}$$

And the solution will have the form:

$$\hat{\beta} = \sum_{i=1}^n (\hat{\alpha}_i^* - \hat{\alpha}_i) x_i \quad (4.20)$$

$$\hat{f}(x) = \sum_{i=1}^n (\hat{\alpha}_i^* - \hat{\alpha}_i) \langle x, x_i \rangle + \beta_0 \quad (4.21)$$

Due to the nature of these constraints, typically only a subset of the solution values $(\hat{\alpha}_i^* - \hat{\alpha}_i)$ are non-zero, and the associated data values are called the support vectors. As was the case in the SVM classification setting, the solution depends on the input values only through the inner products $\langle x_i, x_j \rangle$. We can generalize the methods to richer spaces by defining an appropriate kernel of inner product.

In practice, we need to determine hyperparameters ϵ and C to estimate $\hat{\beta}$. For ϵ , it depends on the scale of response y ; for C , we can use cross-validation to find optimal setting.

For our data set, in order to compare with previous statistical method we use, we apply linear kernel for SVR. According to Cherkassky and Ma (2004), we preset $\epsilon = \epsilon_0 = 3\sigma \sqrt{\frac{\ln n}{n}}$ and $C = C_0 = \max\{|\bar{y} + 3\sigma_y|, |\bar{y} - 3\sigma_y|\}$, where σ is the estimated noise level, \bar{y} and σ_y are the mean and the standard deviation of response. We estimate σ via K-nearest neighbor regression: $\hat{\sigma}^2 = \frac{n^{1/5}k}{n^{1/5}k-1} \cdot \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$, where the \hat{y}_i are estimated by K-nearest neighbor regression and we use $k = 3$.

In order to perform feature selection, we use the recursive feature elimination (RFE) algorithm which at each step eliminates a feature with the smallest weight out of the SVR fitting of the training data. And we use cross-validation to determine the optimal number of features.

The whole process can be described in the following algorithm:

Recursive feature elimination for SVR

1. Standardize X_j to have mean zero and variance one. Let M_p denote the null model, which contains all p predictors.
 2. For $k = p, p - 1, \dots, 1$:
 - (a) Fit SVR model with all the predictors in M_k .
 - (b) Eliminate the predictor with the smallest estimated weight from M_k , and call it M_{k-1} .
 3. Select a single best model from among M_0, \dots, M_p using cross-validated prediction error.
-

With 10 fold cross-validation, the RFE-SVR algorithm selects a set of 9 predictors: RISK, LADV, log(TA), WC/TA, β_{mkt} , β_{umd} , β_{ewmkt} , $\beta_{finance}$, and β_{oil} .

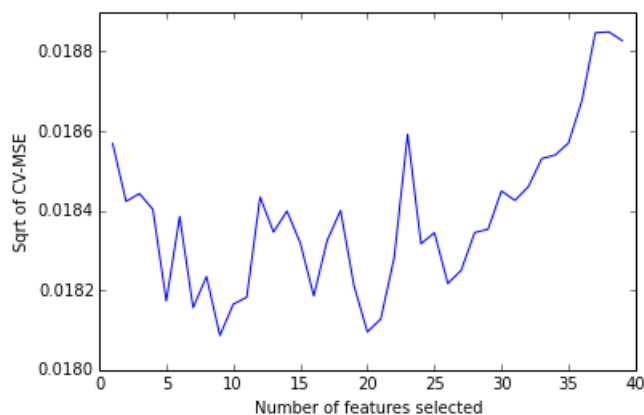


Figure 4.19: Recursive feature elimination for SVR: use cross-validation to choose the optimal number of features. The optimal number of features selected is 9.

Based on the preset ϵ_0 and C_0 , we also use a grid search algorithm to determine optimal hyperparameters via 10-fold cross-validation. The grid we use is: $\epsilon \in \{0, \epsilon_0, 10\epsilon_0\}$ $C \in \{C_0, 10C_0, 100C_0, 1000C_0\}$. We first use grid search to determine the optimal hyperparameters for a certain number of features, then use RFE-SVR algorithm to find optimal number of features with cross-validation. This time the algorithm selects 14 features: RISK, Q_HHI, LADV, INVPRC, $\log(\text{TA})$, NI/TA, WC/TA, CR, QR, β_{mkt} , β_{umd} , β_{ewmkt} , $\beta_{finance}$, β_{oil} .

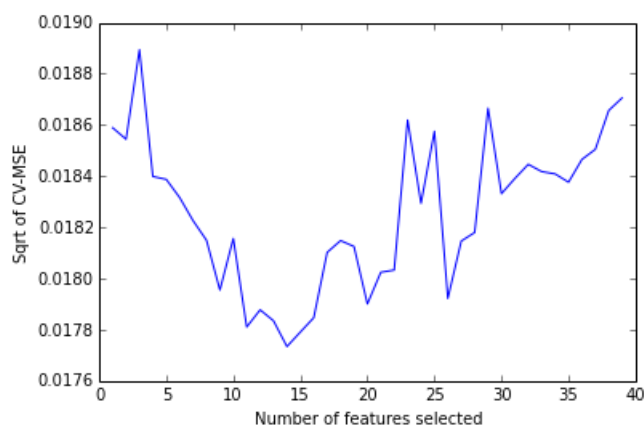


Figure 4.20: Recursive feature elimination for SVR: use cross-validation to choose the optimal number of features. The hyperparameters are chosen by a grid search. The optimal number of features selected is 14.

Examine the individual weight of coefficients more closely, we see SVR tends to put higher

weight on betas but not on *SHOCK*. The cross-validation errors shown in Figure 4.19 and Figure 4.20 are less stable across different number of features selected.

Table 4.8 compares the prediction errors of different statistical methods. In terms of cross-validation error, we can see SVR has slightly higher MSE compared to other methods. This may due to the fact that unlike shrinkage methods (e.g. the Lasso and ridge regression) and dimension reduction methods (e.g. PCR and PLS) are based on least squares, which minimize certain square loss functions, for SVR, it minimizes the new ϵ -insensitive loss function. As we use the cross-validation MSE to measure the prediction accuracy, this metric may favor those methods that are based on least squares.

Model Prediction	The Lasso	Ridge	PCR	PLS	SVR
Sqrt of CV-MSE	0.0180(0.0172)	0.0180(0.0173)	0.0172	0.0173	0.0177

Table 4.8: A comparison of model prediction errors with different statistical methods for Set 1-4 predictors. We show the results of the Lasso method and ridge regression method that apply one-standard-error rule. The results of the best model of selected by the Lasso method and ridge regression method are shown in quote. The SVR result is from the RFE-SVR algorithm with hyperparameters set by cross-validation and applies a linear kernel.

4.4 Summary

In this chapter, we apply association analysis and statistical model selection methods on the stock market crashes data. The association analysis reveals that the interaction term *SHOCK*, which is the product of price impact factor and the market capital, is associated with the cross-sectional price drawdown of SP500 stocks. This echoes the prediction of our single asset model in Section 3.1.

We study the predictors of cross-sectional price drawdown of SP500 stocks during the 2010 Flash Crash by best subset selection, forward stepwise selection and the Lasso model selection methods. We first identify a set of three predictors out of eight predictors (by combining the information asymmetry induced price impact factor, the limited risk bearing capacity induced price impact factor and six other predictors used in Madhavan (2012)) that have strong effects in predicting cross-sectional price drawdown. We then evaluate the robustness of this result by including more candidate predictors such as betas with respect to different economic factors and corporate default predictors. It turns out that the information-asymmetry induced price impact factor and the limited risk bearing capacity induced price impact factor still remain statistically important compared to other predictors.

We also use ridge regression, principal component regression, and partial least squares method to show the evolution of the standardized coefficients of predictors. This enables us to visualize the influence of a unit change of standardized predictor to the best prediction. Finally, we use recursive feature elimination algorithm on support vector regression method, trying to identify a set of predictors that fit the model well with respect to the ϵ -insensitive

loss function. The results show that support vector regression tends to fit the model with more beta predictors and is less stable in terms of cross-validation error.

Chapter 5

Conclusion and Future Work

Stock market crashes are infamous and hard to model. While some crashes are due to economic fundamentals, some are caused by temporary price impact. Due to the limited number of macro stock market crashes, there is even less empirical research than theoretical one on this topic.

In this dissertation, we build a single asset model that explains how a large market-wide liquidity shock can lead to temporary stock market crashes like the 1987 crash and the 2010 Flash Crash. Empirically (Table 3.2), our model predictions are quite close to the real large dollar volume sale and the market price decline. This shows that in a market-wide adverse selection setting, large uninformed trading can have different temporary price impact on different stocks. We include the limited risk-bearing capacity induced price impact from Greenwood (2005)'s model in the cross-sectional regression as to separate causes of temporary price impact. Our regression shows significant effects of both the information asymmetry induced price impact and the limited risk-bearing capacity induced price impact.

We also build a multi-asset model to incorporate the informational cross-trading impact with the Greenwood (2005) model, in order to better decompose the information effect and the limited risk-bearing capacity effect on temporary price impact in our regression. However empirical test of this model is not as easy as it seems to be. Admati (1985) builds a multi-asset noisy rational expectation equilibrium model that accepts a rich set of correlation matrix structure across assets. Bernhardt and Taub (2008) create a strategic analogue of Admati's model, along the lines of Kyle (1985) and Kyle (1989). Based on Admati (1985), Burlacu et al. (2012) transforms unobservable rational expectation equilibrium model parameters (information precision and supply uncertainty) into a single variable that is correlated with expected returns and that can be estimated with recently observed data. Caballé and Krishnan (1994) demonstrate a multi-asset cross-trading impact model as an extension of Kyle (1985)'s single asset model; based on that, Pasquariello and Vega (2013) test information asymmetry induced cross-trading impact factors and find statistical evidence. A possible future research direction would be to decompose the stock returns with a factor fashion along the line of Bernhardt and Taub (2008), and/or aggregate stocks in groups (e.g. by industry) along the line of Pasquariello and Vega (2013), so that we can not only get analytical results

from the model but also conduct feasible empirical tests under a cross-asset trading context. Based on the assumption that information is either long-lived or short-lived, in the future research, we can also extend the multi-asset model with a dynamic version, as in Kyle (1985) and Admati and Pfleiderer (1988).

In Section 3.1, we have assumed that γ_i equals the ratio of the stock i 's market capital to the SP500 index capital at the beginning of the crash. This assumes the index futures arbitragers can perfectly hedge their risk and transmit the sales of E-mini 500 into SP500 underlying stocks. We also calculate the modified γ_i^* 's for the arbitragers under the imperfect hedging scenario. Empirical test of this modification is left for future research.

Using the machine learning method of association analysis, we reveal a weird association that the price drawdown of a SP500 stock is not only associated with its price impact factor, but also its market capital. For example, although a large market cap stock usually has a low price impact factor, it can have larger price drawdown because of the interaction effect of its market cap and the price impact factor. Our model in Chapter 3 explains this unusual effect. We also use best subset selection, forward stepwise selection and the Lasso method to test two model based predictors SHOCK and RISK, which represent the information asymmetry effect and the limited risk-bearing capacity effect respectively, against three sets of other candidate predictors. We show that compared with other predictors, our model based predictors are relatively more important in predicting the SP500 stocks' price drawdown during the 2010 Flash Crash. The quote fragmentation factor (Madhavan 2012) is also identified as a third important factor.

There are a few interesting questions for discussion. First, does our model capture HFT's role during crashes? Our model fits the data well without capturing the role of high frequency trading (especially during the 2010 Flash Crash). As our model gives a mechanism that the 2010 Flash Crash could occur in a similar way as the 1987 crash did (which at that time the HFT activity was insignificant¹), we do not see any unusual effect of HFT that could be the major cause of the 2010 Flash Crash. This is consistent with the finding in Kirilenko et al. (2014).

Does our model apply all the time? In principle, our model is not only restricted to two crashes, rather, it can be generally applied. However it is usually hard to find a relatively accurate documentation of a large market-wide liquidity shock with a certain time scale. In this regard, the Oct. 1987 Crash and the May 2010 Flash Crash are two excellent experiments for our model.

On Apr. 17, 2015, Navinder Singh Sarao, a trader in London was accused by CFTC for manipulating the E-mini SP500 futures contract by "spoofing" tactics— On the day of the crash, May 6, 2010, the trader allegedly entered more than 32,000 orders to sell futures contracts, then canceled the vast majority of them (CFTC 2015). Under the setting of our single period strategic trading models, every order is a market order and is executed in a batched auction way. In practice, price impact is from traders' inference about the size of

¹A close relative of HFT—"program trading" activity during the Oct. 1987 crash was greatly disrupted because of the breakdown of DOT system. U.S. Presidential Task Force on Market Mechanisms (1988)

order flow imbalance through the status of the limit order book and the arrival intensity of market orders. Our model serves as an approximation for this, and aims to decompose the information part and hedging part of the price impact. It would be interesting to build a model that depicts how traders learn from limit and market orders, and the resulting price impact, to capture the “spoofing” effect of limit orders and test it empirically.

Bibliography

- [1] Anat R. Admati. “A Noisy Rational Expectations Equilibrium for Multi-Asset Securities Markets”. In: *Econometrica* 53.3 (1985), pp. 629–658.
- [2] Anat R. Admati and Paul Pfleiderer. “A Theory of Intraday Patterns: Volume and Price Variability”. In: *The Review of Financial Studies* 1.1 (1988), pp. 3–40.
- [3] Irene Aldridge. “High-Frequency Runs and Flash-Crash Predictability”. In: *Journal of Portfolio Management* 40.3 (2014), pp. 113–123.
- [4] Yakov Amihud and Haim Mendelson. “Dealership Market: Market-Making with Inventory”. In: *Journal of Financial Economics* 8 (1980), pp. 31–53.
- [5] Torben Andersen and Oleg Bondarenko. “Assessing Measure of Toxic Order Flow and Early Warning Signals for Market Turbulence”. In: *Review of Finance* 19.1 (2015), pp. 1–54.
- [6] Torben Andersen and Oleg Bondarenko. “Reflecting on the VPIN Dispute”. In: *Journal of Financial Markets* 17 (2014), pp. 53–46.
- [7] Torben Andersen and Oleg Bondarenko. “VPIN and the Flash Crash”. In: *Journal of Financial Markets* 17.1-46 (2014).
- [8] Scott N. Armstrong and Christopher J. Hillar. “Solvability of symmetric word equations in positive definite letters”. In: *Journal London Mathematical Society* 76.3 (2007), pp. 777–796.
- [9] Christopher Avery and Peter Zemsky. “Multidimensional Uncertainty and Herd Behavior in Financial Markets”. In: *The American Economic Review* 88.4 (1998), pp. 724–748.
- [10] E. Barany et al. “Detecting market crashes by analysing long-memory effects using high-frequency data”. In: *Quantitative Finance* 12.4 (2012), pp. 623–634.
- [11] Gadi Barlevy and Pietro Veronesi. “Rational panics and stock market crashes”. In: *Journal of Economic Theory* 110.2 (2003), pp. 234–263.
- [12] Dan Bernhardt and Bart Taub. “Cross-Asset Speculation in Stock Markets”. In: *The Journal of Finance* LXIII.5 (2008).

- [13] Bruno Biais, Larry Glosten, and Chester Spatt. “Market Microstructure: A Survey of Microfoundations, Empirical Results, and Policy Implications”. In: *Journal of Financial Markets* 8.2 (2005), pp. 217–264.
- [14] J.P. Bouchaud and R. Cont. “A Langevin approach to stock market fluctuations and crashes”. In: *The European Physical Journal B* 6 (1998), pp. 543–550.
- [15] Markus K. Brunnermeier. *Asset pricing under asymmetric information: bubbles, crashes, technical analysis, and herding*. Oxford University Press, 2001.
- [16] Radu Burlacu et al. “Risk and the cross section of stock returns”. In: *Journal of Financial Economics* 105 (2012), pp. 511–522.
- [17] Jordi Caballé and Murugappa Krishnan. “Imperfect competition in a multi-security market with risk neutrality”. In: *Econometrica* 62.3 (1994), pp. 695–704.
- [18] Mark M. Carhart. “On Persistence in Mutual Fund Performance”. In: *The Journal of Finance* 52.1 (1997), pp. 57–82.
- [19] CFTC. *Complaint for Injunctive Relief, Civil Monetary Penalties, and other Equitable Relief*. Legal Pleading. 2015.
- [20] Vladimira Cherkassky and Yunqian Ma. “Practical selectoin of SVM parameters and noise estimation for SVM regression”. In: *Neural Networks* 17 (2004), pp. 113–126.
- [21] David Easley and Maureen O’Hara. “Price, Trade Size, and Information in Securities Markets”. In: *Journal of Financial Economics* 19 (1987), pp. 69–90.
- [22] David Easley, Marcos López de Prado, and Maureen O’Hara. “Flow Toxicity and Liquidity in a High-frequency World”. In: *Review of Financial Studies* 25.5 (2012), pp. 1457–1493.
- [23] David Easley, Marcos López de Prado, and Maureen O’Hara. “The Microstructure of the ‘Flash Crash’: Flow Toxicity, Liquidity Crashes, and the Probability of Informed Trading”. In: *Journal of Portfolio Management* 37.2 (2011), pp. 118–128.
- [24] David Easley, Marcos López de Prado, and Maureen O’Hara. “VPIN and the Flash Crash: A rejoinder”. In: *Journal of Financial Markets* 17 (2014), pp. 47–52.
- [25] Eugene F. Fama and Kenneth R. French. “Common risk factors in the returns on stocks and bonds”. In: *Journal of Financial Economics* 33.1 (1993), pp. 3–56.
- [26] James A. Feigenbaum and Peter G.O. Freund. “Discrete Scale Invariance in Stock Markets Before Crashes”. In: *International Journal of Modern Physics B* 10.27 (1996), p. 3737.
- [27] James F. Gammill and Terry A. Marsh. “Trading Activity and Price Behavior in the Stock and Stock Index Futures Markets in October 1987”. In: *Journal of Economic Perspectives* 2.3 (1988), pp. 25–44.
- [28] Mark Garman. “Market microstructure”. In: *Journal of Financial Economics* 3.3 (1976), pp. 257–275.

- [29] Gerard Gennotte and Hayne Leland. “Market Liquidity, Hedging, and Crashes”. In: *The American Economic Review* 80.5 (1990), pp. 999–1021.
- [30] Lawrence R. Glosten and Paul R. Milgrom. “Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders”. In: *Journal of Financial Economics* 14 (1985), pp. 71–100.
- [31] Robin Greenwood. “Short- and long-term demand curves for stocks: theory and evidence on the dynamics of arbitrage”. In: *Journal of Financial Economics* 75 (2005), pp. 607–649.
- [32] Sanford J. Grossman. “An Analysis of the Implications for Stock and Futures Price Volatility of Program Trading and Dynamic Hedging Strategies”. In: *The Journal of Business* 61.3 (1988), pp. 275–298.
- [33] Sanford J. Grossman. “On the Efficiency of Competitive Stock Markets Where Trades Have Diverse Information”. In: *The Journal of Finance* 31.2 (1976), pp. 573–585.
- [34] Sanford J. Grossman and Merton H. Miller. “Liquidity and Market Structure”. In: *The Journal of Finance* 43.3 (1988), pp. 617–633.
- [35] Sanford J. Grossman and Joseph E. Stiglitz. “On the Impossibility of Informationally Efficient Markets”. In: *The American Economic Review* 70.3 (1980).
- [36] Peter Harrington. *Maching Learning in Action*. 1st. Manning Publications, 2012.
- [37] Joel Hasbrouck. *Empirical Market Microstructure*. Oxford University Press, 2007.
- [38] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. 2nd. Springer, 2008.
- [39] Martin F. Hellwig. “On the aggregation of information in competitive markets”. In: *Journal of Economic Theory* 22.3 (1980), pp. 477–498.
- [40] Thomas Henker and Jian-Xin Wang. “On the importance of timing specifications in market microstructure research”. In: *Journal of Financial Markets* 9 (2006), pp. 162–179.
- [41] Thomas Ho and Hans R. Stoll. “Optimal Dealer Pricing under Transactions and Return Uncertainty”. In: *Journal of Financial Economics* 9 (1981), pp. 41–73.
- [42] Jennifer Huang and Jiang Wang. “Liquidity and Market Crashes”. In: *Review of Financial Studies* 22.7 (2009), pp. 2607–2643.
- [43] Gareth James et al. *An Introduction to Statistical Learning with Application in R*. Springer, 2013.
- [44] Anders Johansen, Didier Sornette, and Olivier Ledoit. “Predicting financial crashes using discrete scale invariance”. In: *Journal of Risk* (1999).
- [45] Andrei Kirilenko et al. “The Flash Crash: The Impact of High Frequency Trading on an Electronic Market”. 2014.

- [46] Albert S. Kyle. “Continuous Auctions and Insider Trading”. In: *Econometrica* 53.6 (1985), pp. 1315–1335.
- [47] Albert S. Kyle. “Informed Speculation with Imperfect Competition”. In: *The Review of Economic Studies* 56.3 (1989), pp. 317–355.
- [48] Albert S. Kyle and Anna Obizhaeva. “Large Bets and Stock Market Crashes”. 2013.
- [49] Albert S. Kyle and Anna Obizhaeva. “Market Microstructure Invariance: Theory and Empirical Tests”. 2013.
- [50] Peter Lancaster and Leiba Rodman. *Algebraic Riccati Equations*. Oxford Science Publications, 1995.
- [51] Charles M. Lee and Mark J. Ready. “Interring Trade Direction from Intraday Data”. In: *Journal of Finance* 46.2 (1991), pp. 733–746.
- [52] Hayne Leland. *Leverage, Forced Asset Sales, and Market Stability: Lessons from Past Market Crises and the Flash Crash*. Tech. rep. UK Government Office for Science, Foresight Project, 2011.
- [53] Hayne Leland and Mark Rubinstein. “Comments on the Market Crash: Six Months After”. In: *Journal of Economic Perspectives* 2.3 (1988), pp. 45–50.
- [54] Ananth Madhavan. “Exchange-Traded Funds, Market Structure and the Flash Crash”. In: *Financial Analysts Journal* 68.4 (2012), p. 16.
- [55] Albert J. Menkveld and Bart Zhou Yueshen. “The Flash Crash: A Cautionary Tale about Highly Fragmented Markets”. 2015.
- [56] Stefan Nagel. “Evaporating Liquidity”. In: *The Review of Financial Studies* 25.7 (2012), pp. 2005–2039.
- [57] Paolo Pasquariello and Clara Vega. “Strategic Cross-Trading in the U.S. Stock Market”. In: *Review of Finance* (2013), pp. 1–54.
- [58] Markus Pelger. “Large-dimensional factor modeling based on high-frequency observations”. 2015.
- [59] Richard Roll. “A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market”. In: *Journal of Finance* 39.4 (1984), pp. 1127–1139.
- [60] Staffs of the CFTC and SEC. *Findings Regarding the Market Events of May 6, 2010*. Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues. 2010.
- [61] Staffs of the CFTC and SEC. *Preliminary Findings Regarding the Market Events of May 6, 2010*. Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues. 2010.
- [62] Robert Tibshirani. “Regression shrinkage and selection via lasso”. In: *Journal of Royal Statistical Society Series B*.58(1) (1996).

- [63] Robert Tibshirani. “Regression shrinkage and selection via lasso: a retrospective”. In: *Journal of Royal Statistical Society B.73* (2011), pp. 273–282.
- [64] U.S. Presidential Task Force on Market Mechanisms. *Report of the Presidential Task Force on Market Mechenisms*. The “Brady Report”. Washington: U.S Government Priting Office., 1988.
- [65] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.
- [66] Xavier Vives. *Information and Learning in Markets*. Princeton University Press, 2008.
- [67] Xingwei Wu. “Corporate Default Prediction”. PhD thesis. UC Berkeley, 2011.

Appendix A

Variable Name	Description	Database Input	Measurement Time
ESPRD	Effective spread	TAQ Intraday, in basis point	2010/05/06 14:30-15:00 ET
EPS_RATIO	Earnings per share ratio	Compustat/CRSP merged: ESP/OPENPRC (100%)	2010/05/06
FEPS_RATIO	Forecasted EPS ratio	I/B/E/S: MEANEST/OPENPRC (100%)	2010/04/15 (or 2010/03/18 if not available)
GSECTOR	Stock Sector	Compustat/CRSP merged: GSECTOR	2010/05/06
INST_HHI	Ownership concentration index	Thomson Reuters Institutional (13F) Holding	2010/03/31
INST_PERC	Total Inst. ownership, percent of SHROUT	Thomson Reuters Institutional (13F) Holding	2010/03/31
LAMBDA	Price Impact factor	TAQ Intraday	2010/04/29-2010/05/05
MKTCAP	Market capital	CRSP: SHROUT*OPENPRC	2010/05/06
MEANREC	Mean recommendation	I/B/E/S: MEANREC	2010/04/15 (or 2010/03/18 if not available)
PRC_CHG	Price change	TAQ Intraday	2010/05/06 14:30-15:00 ET
PRIMEXCH	Primary listed exchange	CRSP: PRIMEXCH	2010/05/06
SELLPCT	Sell Percent	I/B/E/S: SELLPCT	2010/04/15 (or 2010/03/18 if not available)
SPCSRC	S&P quality ranking	Compustat/CRSP merged: SPCSRC	2010/05/06
VOLATILITY	Daily stock volatility	CRSP daily stock: RETX	2010/01/01-2010/05/05
VOLUME_RATIO	Volume traded as percent of ADV	TAQ Intraday, CRSP daily: VOLUME/ADV	2010/05/06 14:30-15:00 ET

Table A.1: Association analysis: Parameters and data source, 2010 Flash Crash

Variable Name	Description	Database Input	Measurement Time
ESPRD	Effective spread	ISSM Intraday, in basis point	1987/10/14-1987/10/19
FEPS_RATIO	Forecasted EPS ratio	I/B/E/S: MEANEST/OPENPRC (100%)	1987/09/17
GSECTOR	Stock Sector	Compustat/CRSP merged: GSECTOR	1987/10/13
INST_HHI	Ownership concentration index	Thomson Reuters Institutional (13F) Holding	1987/09/30
INST_PERC	Total Inst. ownership, percent of SHROUT	Thomson Reuters Institutional (13F) Holding	1987/09/30
LAMBDA	Price Impact factor	ISSM Intraday	1987/08/31-1987/10/13
MKTCAP	Market capital	CRSP: SHROUT*OPENPRC	1987/10/13
PRC_CHG	Price change	TAQ Intraday	1987/10/14-1987/10/19
PRIMEXCH	Primary listed exchange	CRSP: PRIMEXCH	1987/10/13
SPCSRC	S&P quality ranking	Compustat/CRSP merged: SPCSRC	1987/10/13
VOLATILITY	Daily stock volatility	CRSP daily stock: RETX	1987/01/01-1987/10/13
VOLUME_RATIO	Volume traded as percent of ADV	ISSM Intraday, CRSP daily: VOLUME/ADV	1987/10/14-1987/10/19

Table A.2: Association analysis: Parameters and data source, 1987 crash

Abbreviation	Description	Database Input
ME	Market Value of Equity	Compustat: PRCCM*CSHO
BD	Book Value of Total Debt	Compustat: DLC+DLTT
TA	Total Asset	Compustat: AT
TL	Total Liability	Compustat: LT
SD	Short Term Debt	Compustat: DLC
SALE	Sales	Compustat: SALE
NI	Net Income	Compustat: IB
RE	Retained Earnings	Compustat: RE
EBIT	Earnings Before Interest and Taxes	Compustat: EBIT
ICR	Interest Coverage Ratio	Compustat: EBIT/XINT
WC	Working Capital	Compustat: ACT-LCT
AR	Account Receivable	Compustat: RECT
INVT	Inventories	Compustat: INVT
CR	Current Ratio	Compustat: ACT/LCT
CH	Cash	Compustat: CH
QR	Quick Ratio	Compustat: (ACT-INVT)/LCT
OM CH	Change in Operating Margin	Compustat: OIBDP/SALE
SRT	Trailing One Year Stock Return	CRSP Monthly Stock
SIGMA	One Year Monthly Stock Volatility	CRSP Monthly Stock

Table A.3: Statistical model selection: Corporate default parameters abbreviation and data Source