Funding Liquidity and Limits to Arbitrage

by

Bassam Aoun

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2012

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Arbitrageurs play an important role in keeping market prices close to their fundamental values by providing market liquidity. Most arbitrageurs however use leverage. When funding conditions worsen they are forced to reduce their positions. The resulting selling pressure depresses market prices, and in certain situations, pushes arbitrage spreads to levels exceeding many standard deviations. This phenomenon drove many century old financial institutions into bankruptcy during the $2007-2009$ financial crisis. In this thesis, we provide empirical evidence and demonstrate analytically the effects of funding liquidity on arbitrage. We further discuss the implications for risk management.

To conduct our empirical studies, we construct a novel Funding Liquidity Stress Index (FLSI) using principal components analysis. Its constituents are measures representing various funding channels. We study the relationship between the FLSI index and three different arbitrage strategies that we reproduce with *real* and *daily* transactional data. We show that the FLSI index has a strong explanatory power for changes in arbitrage spreads, and is an important source of contagion between various arbitrage strategies. In addition, we perform "event studies" surrounding events of changing margin requirements on futures contracts. The "event studies" provide empirical evidence supporting important assumptions and predictions of various theoretical work on market micro-structure.

Next, we explain the mechanism through which funding liquidity affects arbitrage spreads. To do so, we study the liquidity risk premium in a market micro-structure framework where market prices are determined by the supply and demand of securities. We extend the model developed by Brunnermeier and Pedersen [BP09] to multiple periods and generalize their work by considering all market participants to be risk-averse. We further decompose the liquidity risk premium into two components: 1) a *fundamental risk premium* and 2) a *systemic risk premium*. The fundamental risk premium compensates market participants for providing liquidity in a security whose fundamental value is volatile, while the systemic risk premium compensates them for taking positions in a market that is vulnerable to funding liquidity. The first component is therefore related to the nature of the security while the second component is related to the fragility of the market micro-structure (such as leverage of market participants and margin setting mechanisms).

Acknowledgements

Foremost, I would like to thank my supervisors, Professor Donald McLeish and Professor Phelim Boyle, for their continuous support and guidance. I was enriched and influenced by their approach to solving problems and original thinking that will accompany me throughout my career. I would like also to thank my colleagues at the department of Statistics and Actuarial Science for providing a wonderful environment. I thank Mary Lou Dufton and Mary Flatt for assisting me in administrative matters.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Arbitrageurs are active market participants and play an important role in keeping market prices close to their fundamental values by providing market liquidity. Most arbitrageurs however use external financing, mainly in the form of leverage, to conduct arbitrage. [Gea09] and [FG08b] studied the impacts of leverage on the general functioning of financial markets and fluctuations in asset prices. They argue that collateral rates[1] are more important to the stability of financial markets during periods of stress than interest rates[2]. The author, John Geanakoplos, proceeds to say: *"Despite the cries of newspapers to lower the interest rates, the Fed would sometimes do much better to attend to the economy-wide leverage and leave the interest rate alone"*.

One form of leverage is collateralized borrowing through margin accounts. Typically, when an arbitrageur purchases a security S_t, he can borrow L_t by posting the security as a collateral. The difference between S_t and L_t is called dollar margin, $m_t = S_t - L_t$. In other words, the arbitrageur can effectively purchase the security by paying only m_t and posting the security as collateral. Highly leveraged arbitrageurs are however vulnerable to changes in margin requirements. As margin requirements increase, they are forced to reduce their positions to de-lever, leading to a downward pressure on market prices. For example, the increase in margin requirements for silver future contracts by the Chicago Mercantile Exchange (CME) the first week of May 2011 led to over 20% drop in the price. As one news article puts it "*Silver prices also continued to retreat with the July silver*

[1] The *collateral rate* is the ratio of the collateral value to the amount borrowed. Hence, as the *collateral rate* decreases, the leverage increases. A variety of assets can be purchased through collateralized borrowing such as purchasing a stock on margin or using a mortgage loan to buy a house.

[2] Most economists consider the *federal funds target rate*, widely known as the "interest rate", as the variable that has the most impact on the economy.

contract in New York down $3.48 *to US*$35.90 *an ounce. The drop in silver came after the main U.S. metals exchange announced further hikes to margin requirements. The latest hike amounts to an* 84 *per cent increase in margin requirements in two weeks by CME Group Ltd., spread over four separate changes. Silver prices are down over* 20 *per cent this week and analysts say volatility in the sector has spread to other areas since investors have been forced to sell other securities to meet higher margin calls*"[3]. On the same topic, Dave Meger, director of metals trading at Vision Financial Markets, added "*this was panic-type selling. As the market started falling in the midst of thin volumes it drew more sellers to the market, and this created a vacuumed to the downside*"[4]. The anecdotal evidence of spiralling effects between margins and market prices is further supported by theoretical studies such as [BP09] and [BS09]. In this thesis, we provide further empirical evidence on the relationship between funding liquidity and arbitrage. Funding liquidity is the "*ease with which investors can obtain funding*" [BP09]. We show that funding liquidity 1) has a strong explanatory power for changes in arbitrage spreads, 2) is a principal source of contagion between various arbitrage spreads, and 3) is correlated with the probability of being in a period of stress.

Various theoretical studies have attempted to explain the effects of funding liquidity on the deviation of market prices from their fundamental values by studying the micro-structure of capital markets. The literature on market micro-structure evolved gradually towards an accurate depiction of the functioning of capital markets. In 1988, Grossman and Miller [GM88] proposed a market micro-structure model where market prices are determined by the supply and demand of securities. Their model served as a foundation for future work in the literature. The work by De Long *et al.* [DLSSW90] is perhaps the first formal attempt in the literature to assert that arbitrageurs have short term horizons and are therefore deterred from betting aggressively on mispriced assets. [SV97] and [BP09] further argued that the aggregate wealth of arbitrageurs affects their ability to conduct arbitrage and keep markets liquid. Key to their thesis is that arbitrage requires capital which is controlled by an "agency relationship". They both successfully show that markets are vulnerable to *liquidity shocks* when financiers[5], uninformed about the fundamentals, use past performance [SV97] or market price volatility [BP09] to control funding of arbitrage positions. For instance, after periods of high volatility or after losses on existing positions, financiers indiscriminately withdraw capital at the very moment when it is needed the most and when opportunities are the most attractive.

[3] "TSX negative, demand concerns, new silver margin requirements roil commodities" - *The Canadian Press*, Thursday May 5, 2011

[4] "Silver Prices Tumble 12% After CME Hikes Margin" - *The Wall Street Journal*, Monday May 2, 2011

[5] i.e., those who control the capital

The analytical studies in [SV97] and [BP09] were however conducted in a single-period setting, restricted to illustrating "liquidity shocks". "Liquidity shocks" occur when the arbitrageur is forced to de-lever due to binding funding conditions, and market prices drop far below their fundamental values due to the selling pressure. Arbitrageurs however require a *liquidity risk premium*[6] at all time in order to be compensated for the risk of future losses, in particular for those losses that result from future "liquidity shocks". In this thesis, we extend the market micro-structure model of [BP09] to multiple periods and compute the *liquidity risk premium* at various points in time. We further decompose the *liquidity risk premium* into two components: 1) a *fundamental risk premium* and 2) a *systemic risk premium.* The fundamental risk premium compensates market participants for providing liquidity in a security whose fundamental value is volatile, while the systemic risk premium compensates them for taking positions in a market vulnerable to liquidity shocks. The first component is therefore related to the nature of the security while the second component is related to the fragility of the market micro-structure (such as leverage of market participants and margin setting mechanisms). The decomposition of the *liquidity risk premium* into its two components is novel because the *systemic risk premium* only appears in multi-period market micro-structure models with financial constraints.

The basis to those theoretical studies is that the market price is determined by the supply and the demand for the security [GM88]. As a result, any change to either the supply or the demand causes the market price to change. Clearly, a change to the fundamental value of the security is most often accompanied by a change in the supply and demand such that the market price moves in tandem with the fundamental value. However, in certain situations, the supply or the demand changes, not in response to changes to the fundamental value, but due to factors related to the market micro-structure[7] in which the security trades. In those situations, the market price deviates from its fundamental value. The $2007-2009$ financial crisis provided ample evidence of large deviations of market prices from their fundamental values as the market micro-structure came under stress. In most cases, investors were forced to sell some of their holdings causing the supply of securities to overwhelm the demand. For instance, in Figure 1.1 we show a closed-end fund[8] trading

[6]A *risk premium* is the component of the future expected return that compensates the investor for a specific risk. As the *risk premium* increases, the price drops to make room for a larger future expected return.

[7]Such as binding funding constraints restricting the number of shares that can be purchased.

[8]A Closed-End Fund (CEF) is an investment company that raises capital by selling a fixed number of shares through an Initial Public Offering (IPO). The capital is used to purchase an investment portfolio. After the IPO, the shares trade on the stock exchange where the market price is affected by the supply and demand of the security. The Net Asset Value (NAV) per share is equal to the value of the investment portfolio divided by the number of shares outstanding.

at a substantial discount from its net asset value. According to the Wall Street Journal[9], certain hedge funds were forced to liquidate their equity Quant positions (such as closed-end funds arbitrage positions) in order to raise capital after receiving margin calls from their brokers to cover for losses in other strategies (mainly subprime credit). In other cases, investors were forced to buy a large number of shares in a market with insufficient supply. For instance, in Figure 1.2 we show the "*meteoric rise*" of Volkswagen shares as a result of a short squeeze. "*Short sellers desperate to close their positions paid as much as* $1,005$ *Euros a share during the session following Sunday's news that there was less than 6 percent of VW stock still floating in the market*"[10]. Porsche had been discreetly accumulating a large number of VW shares to gain control of the firm. Those hedge funds are currently suing Porsche for more than 2 Billion dollars in damages[11]. As shown in Figure 1.1 and Figure 1.2, *liquidity shocks* are characterized by a sharp deviation of the market price followed by a quick reversal. It is interesting to note that the price deviation does not revert to its "pre-shock" level. In the aftermath, market participants are reminded of the fragility of capital markets and, as a result, price-in a larger *systemic risk premium.* With respect to Figure 1.1, we first note that the discount didn't shrink completely after August 15$^{\text{th}}$ as investors priced-in a *systemic risk premium* (i.e., the risk of future liquidity shocks). Second, the volatility of the fundamental value (i.e., the NAV) didn't change significantly as to justify an abrupt change in the *fundamental risk premium.* The rise in the *liquidity risk premium* is therefore attributed solely to the rise in the *systemic risk premium.*

1.1 Outline

The thesis is composed of five chapters. The introduction was Chapter 1. In Chapter 2 we provide empirical evidence on the effects of funding liquidity on arbitrage spreads. In Chapter 3 we explain the mechanism through which funding liquidity affects arbitrage spreads by conducting analytical studies in a market micro-structure setting. In Chapter 4 we discuss the implications for risk management, and we conclude the thesis in Chapter 5. In what follows, we provide a more detailed break-down.

Chapter 2 is organized into six sections. In Section 2.1, we conduct a literature review. In Section 2.2, we present and motivate the constituents of our funding liquidity index. The constituents are classified into four categories: 1) margin requirements on futures contracts, 2) general collateral repo spreads, 3) LIBOR-based measures, and 4) corporate credit yield

[9] "The Minds Behind the Meltdown" - *Wall Street Journal*, January 23, 2010
[10] "Short sellers make VW the world's priciest firm" - *Reuters*, October 28, 2008
[11] "Hedge Funds Sue Porsche for Billions Lost on VW" - *Reuters*, April 29, 2010

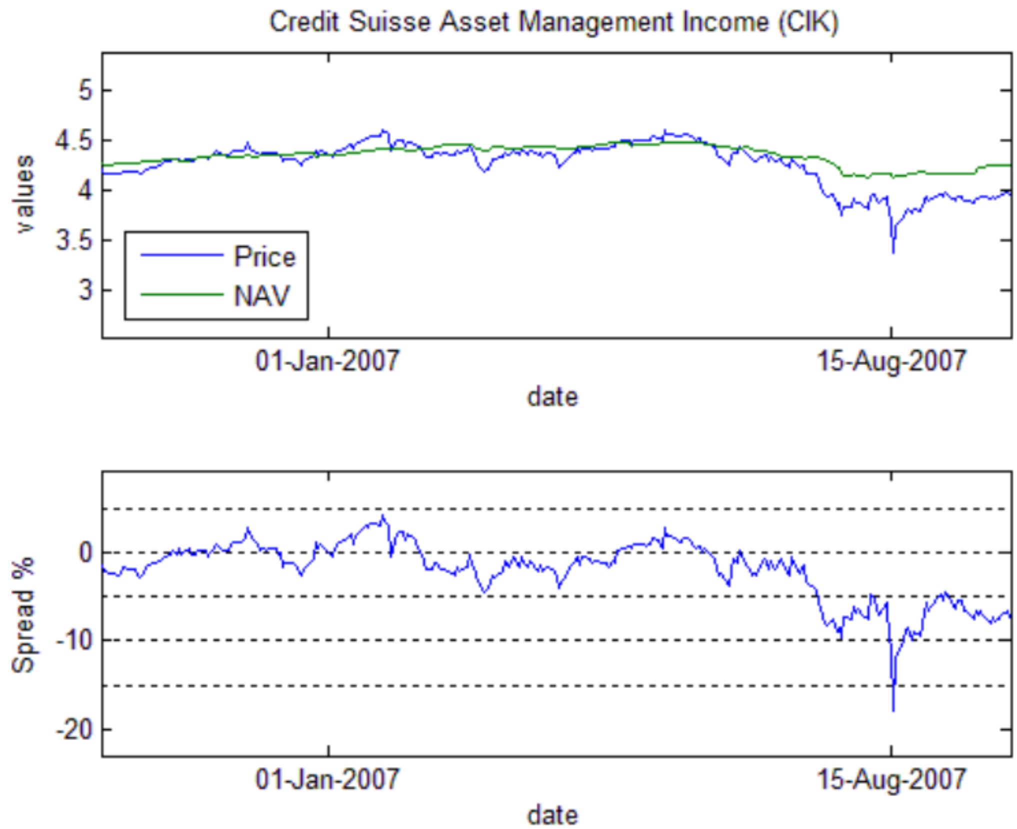


Figure 1.1: Forced Sellers. In the first plot, we show the market price and the net-asset-value (NAV) of a closed-end fund, Credit Suisse Asset Management Income Fund. In the second plot, we show the discount of the market price from its NAV. The discount suddenly widened on August 15^{th} and rebounded shortly after. First, we note that the discount didn't shrink completely after August 15^{th} as investors priced-in a *systemic risk premium* (i.e., the risk of future liquidity shocks). Second, the volatility of the fundamental value (i.e., the NAV) didn't change significantly as to justify an abrupt change in the *fundamental risk premium*. The rise in the *liquidity risk premium* is therefore attributed to a rise in the *systemic risk premium*. The likely cause for the rise in liquidity risk premium is a decline in funding liquidity 1) as arbitrageurs reduced the capital allocated for the CEF arbitrage strategy to cover for losses in other strategies, 2) as financiers made more strict their margin setting mechanism to protect themselves from counter-party risk, 3) as the collateral value of the CEF book declined due an increase in volatility and drop in market prices.

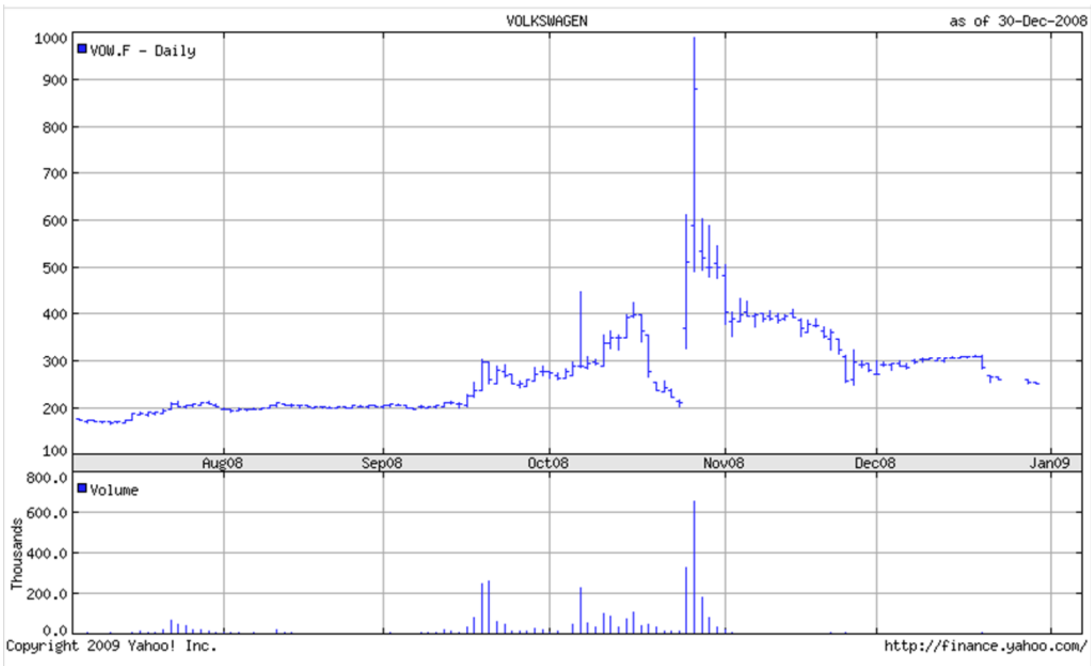


Figure 1.2: Forced Buyers. Volkswagen briefly became the world's biggest company by market value on Tuesday October 29, 2008. Short sellers desperate to close their positions paid as much as 1,005 Euros a share during the session following Sunday's news that there was less than 6 percent of VW stock still floating in the market after a buying spree by Porsche.

spreads. In Section 2.3, we construct the Funding Liquidity Stress Index (FLSI) using principal component analysis. In Section 2.4, we reproduce and study three arbitrage strategies: 1) Closed-End Funds (CEF) arbitrage, 2) Mergers and Acquisitions (M&A) arbitrage, and 3) on-the-run/off-the-run Treasury arbitrage. In Section 2.5, we study the relationship between the FLSI index and the three arbitrage strategies. In Section 2.6, we conduct event-studies surrounding changes in margin requirements.

Chapter 3 is organized into five sections. In Section 3.1, we conduct a literature review and describe the market micro-structure model. In Section 3.2, we construct the multiperiod optimization problem and show that it is an instance of "finite dynamic optimization problems". We also derive a recursive formulation which we use to obtain the numerical results in Section 3.4. In Section 3.3, we derive the analytical solution at time $T-1$ (essentially, the one-step problem) by solving the general multiperiod problem. In Section 3.4, we discuss our numerical approach to solve the problem at times $t < T-1$ and we compute the numerical results. In Section 3.5, we summarize our findings.

Chapter 4 is composed of two sections. In Section 4.1, we fit a Markov regime-switching model and show that the FLSI index has a strong explanatory power for the probability of being in a "bad regime". In Section 4.2, we decompose the variance and covariance of market prices and show that the increase in volatility and correlation during periods of stress is primarily due to changes in the liquidity discounts.

Chapter 2

Empirical Evidence

2.1 Literature Review

2.1.1 Related Work

The main objective of our empirical studies is to motivate and to provide empirical evidence supporting the theoretical studies on market micro-structure models which we extend in the next chapter. In particular, we provide empirical evidence 1) of the effects of funding liquidity on arbitrage spreads, supporting the theoretical work of [SV97], [GV09] and [BP09], and 2) of the commonality across various arbitrage spreads, supporting other aspects of the theoretical work in [BP09].

We also contribute to the large empirical literature that attempts to explain the properties of hedge fund returns. Exposure to classic factors was explored by the early work of [FF93], [Aga04] and [FH04]. More recently, many have explored the effects of funding liquidity on hedge fund returns and the resulting contagion across various hedge fund styles. For instance, [DN10a] shows that margins on futures contracts have significant explanatory power for the returns of hedge funds. In a follow up paper, [DN10b] examines the role of changes in margin requirements in explaining the asymmetric correlation among various hedge fund styles. They found that after controlling for classic risk factors, an increase in margin requirements leads to a significant increase in the probability of contagion. Others have studied the returns in a regime switching framework. For instance, [BGP08] found that, during high volatility regimes, hedge fund returns have significant exposures to the Large-Small and Credit Spread risk factors. The authors used those two factors as proxies to market liquidity and funding liquidity risk, respectively. They find that the exposure to those two factors largely explains the increase in correlation during high volatility periods. [AB02] also found that the correlation of returns increases during periods of elevated volatility. [BSS08] studied the returns of hedge fund indices and found that there is no contagion between traditional asset classes and hedge fund returns. However they found evidence of contagion across various hedge funds styles. In a follow up paper, [BSS10] shows that adverse shocks to certain funding liquidity and market liquidity measures are associated with a significant increase in hedge fund contagion for certain styles.

During periods of crisis, contagion is not however restricted to hedge funds. [AB08] proposes a new measure for systemic risk based on contagion between financial institutions. The authors show that poor hedge fund returns increase the probability of poor returns of investment banks in the following months. Others have studied the impact of financial institutions leverage on market prices. For instance, [MAE11] develops a capital asset pricing model where a broker-dealer leverage factor enters the pricing kernel. They find that risky asset returns can be largely explained by their covariance to the leverage of broker-dealers. [AS10] shows that collateralized borrowing and lending by financial intermediaries have significant predictive power of market-wide risk, as measured by the implied volatility of the S&P 500 index.

2.1.2 Contributions

Our approach is different from the literature in at least two major ways: 1) we construct a novel and comprehensive index of funding liquidity, and 2) we reproduce three arbitrage strategies with real and daily transactional data to test the effects of funding liquidity.

We construct an index with daily observations, called the Funding Liquidity Stress Index (FLSI), reflecting the funding conditions. The FLSI index is novel in couple of ways. First, its constituents are funding measures representing the various funding channels. They are split into four categories: 1) margin requirements on futures contracts, 2) general collateral repo spreads, 3) LIBOR-based measures, and 4) corporate credit yield spreads. Related studies however have used measures from only one or two of the above four categories. For instance, [BGP08] uses corporate credit spread (category 4), [DN10a] uses margins of futures contracts (category 1), and [BSS10] uses credit spreads (category 4) and TED spreads (category 3). Second, we use principal components analysis to compute the weightings of the various funding measures in the FLSI index. Not unexpectedly, we find that the FLSI index explains 78% of the idiosyncratic risk of banks' stock prices[1].

[1]The banking sector is the most sensitive sector of the economy to funding conditions. Their assets have longer maturity than their liabilities and, as a result, banks need a healthy funding environment to

That strong relationship shows that the FLSI index is a strong measure of the funding conditions.

Second, we reproduce three arbitrage strategies using daily transactional data for the past twelve years. That enables us to 1) obtain *daily* observations of arbitrage spreads, and 2) study the relationship between *actual* arbitrage strategies and funding liquidity. To the best of our knowledge, no other work studied the effects of funding liquidity on actual arbitrage strategies with daily frequency. The majority have used monthly returns of hedge fund indices, while others collected monthly returns of individual hedge funds. Hedge fund returns however are not proxies to fully-hedged arbitrage strategies[2] as many elect to hedge or unhedge certain risk factors on a discretionary basis to improve the returns. To synthetically obtain returns of fully-hedged arbitrage strategies, it is common to regress the returns of hedge funds on various traditional factors and asset classes, and then study the residuals[3]. In this thesis, however, we study actual arbitrage strategies by reproducing them from daily and real transactional data, instead of synthetically creating them from hedge fund returns. Also note that using *daily* observations rather than monthly is very important given the high intra-month volatility of market prices and market conditions[4].

Equipped with the daily observations of the FLSI index and the spreads of three arbitrage strategies, we perform various studies. We show that the FLSI index explains up-to 65% of the variation in arbitrage spreads. We also show that the contagion between the arbitrage strategies is mostly due to their common exposure to the FLSI index.

To further support the theoretical studies on market micro-structure, we perform event-studies surrounding events of changing margin requirements. In particular, we show that margins rise after periods of increasing volatility and decreasing asset prices which supports the "margin setting mechanism" in [BP09] as well as the assumption regarding the use of "backward-looking risk measures" by market participants in [BCG$^+$09]. More importantly, asset prices continue to decrease after margins increase, providing evidence of spiraling effects as predicted by the theoretical model of [BP09]. The opposite is also true. Margins decrease after periods of rising prices and decreasing market volatility. Our observations further support the concepts of "performance-based arbitrage" in [SV97] and "pro-cyclicality" of financing in [Gea09].

roll-over their maturing short term debt [Bru09] [BO10].

[2]See [Aga04] and [FH04].

[3]See [DN10b] and [BSS10].

[4]For instance, the VIX, *"a popular measure of the implied volatility of S&P 500 index options"*, had the following values at month ends in 2008: 39.4 on September 30th, 59.9 on October 31st and 55.3 on November 30th. At a first glance, it seems that the index didn't change much. The intra-month values however where much higher: the VIX reached a value of 80 on October 27th and 80.9 on November 20th.

Throughout this chapter, we perform further literature review and discuss other contributions as we tackle various topics, variables and indices.

2.1.3 Outline

The rest of the chapter is organized into five sections. In Section 2.2, we present and motivate the constituents of our funding liquidity index. The constituents are classified into four categories: 1) margin requirements on futures contracts, 2) general collateral repo spreads, 3) LIBOR-based measures, and 4) corporate credit yield spreads. In Section 2.3, we construct the Funding Liquidity Stress Index (FLSI) using principal component analysis. In Section 2.4, we reproduce and study three arbitrage strategies: 1) Closed-End Funds (CEF) arbitrage, 2) Mergers and Acquisitions (M&A) arbitrage, and 3) on-the-run/off-the-run Treasury arbitrage. In Section 2.5, we study the relationship between the FLSI index and the three arbitrage strategies. In Section 2.6, we conduct event-studies surrounding changes in margin requirements.

2.2 Empirical Funding Measures

Funding liquidity is the ease by which an arbitrageur can borrow to fund his arbitrage positions. There are two types of lending: *secured* and *unsecured*. The dominant form of *unsecured* lending is interbank lending[5]. Another form of *unsecured* lending is corporate credit in the bond market. When lending is *secured*, the borrower posts collateral. For instance, when an arbitrageur purchases a security S_t, he can borrow L_t by posting the security as collateral. The difference between S_t and L_t is called dollar margin, $m_t = S_t - L_t$. The arbitrageur can effectively purchase the security by paying only m_t and posting the security as collateral. The dollar margin m_t is set by the lender. In case of default at time $t+1$, the lender liquidates the security and recovers $S_{t+1} + m_t$. m_t is therefore set so that $m_t > S_t - S_{t+1}$ with a certain level of confidence[6]. The lender updates m_t regularly to adjust for changes in volatility. Two primary examples of *secured* lending are 1) trading on margin, and 2) conducting general collateral repo transactions.

In this section, we present seven *empirical funding measures*. They are classified into four categories reflecting the four forms of lending discussed above: 1) margin requirements

[5] LIBOR is the most popular reference rate associated with this form of lending

[6] For instance, the Chicago Mercantile Exchange (CME) sets margins so they cover 99 percent of the potential price moves.

on futures contracts, 2) general collateral repo spreads, 3) LIBOR-based measures, and 4) corporate credit yield spreads.

2.2.1 Margin Requirements on Futures Contracts

A futures contract is a standardized contract between two parties to exchange a specific asset at a future date for a price determined today. The underlying asset can be a physical asset such as a commodity or a currency, or a referenced item such as an index or a financial instrument. Futures contracts trade on exchanges that perform clearing functions. Those functions aim to reduce the risk of default by either party by asking each of them to post an amount of cash, called an initial margin. As the price of the underlying asset changes, the profits and losses are settled daily by debiting or crediting the margin accounts. The parties are then asked to fund their margin accounts so that they don't fall below a certain level, called maintenance margin. Since the profits and losses are settled on an on-going basis, the asset is exchanged at maturity at the spot price, not at the price specified at contract initiation.

We study the margins of futures contracts that trade on the Chicago Mercantile Exchange (CME). CME publishes the historical margins of certain contracts on their website[7]. We focus on futures contracts on the following assets: 1) the S&P 500 index, 2) the Dow Jones index, 3) the AUS/USD foreign exchange rate, and 4) the NZD/USD foreign exchange rate. CME reports the dollar margin m_t per contract. The dollar margin does not change that often, only few times a year. We are however interested in the percentile margin. We compute the percentile margin as follows: $M_t = \frac{m_t}{S_t \times n}$ where S_t is the spot price of the asset and n is the contract size (i.e., the number of units of the underlying asset). In Figure 2.1 and Figure 2.2, we plot the dollar and percentile margin requirements on the S&P 500 futures contract. In Appendix B, we plot the margins of futures contacts on the remaining three assets (Figure B.1 to Figure B.6). To represent the margin requirements on various assets, we construct an equally weighted index of the percentile margins on three of the four futures contracts. We exclude the margins of the Dow Jones futures contract due to its short history. In Figure 2.3, we plot the index. We note that the custom index increased during most periods of financial crisis. In particular, the areas in yellow correspond to periods of financial stress whereas the areas in grey correspond to official U.S. recessions. For more information, refer to Appendix A. In Table 2.1, we compute the annualized volatilities of the four assets and their mean margin requirements. We note that higher the volatility of the asset, higher is its margin requirement. That is

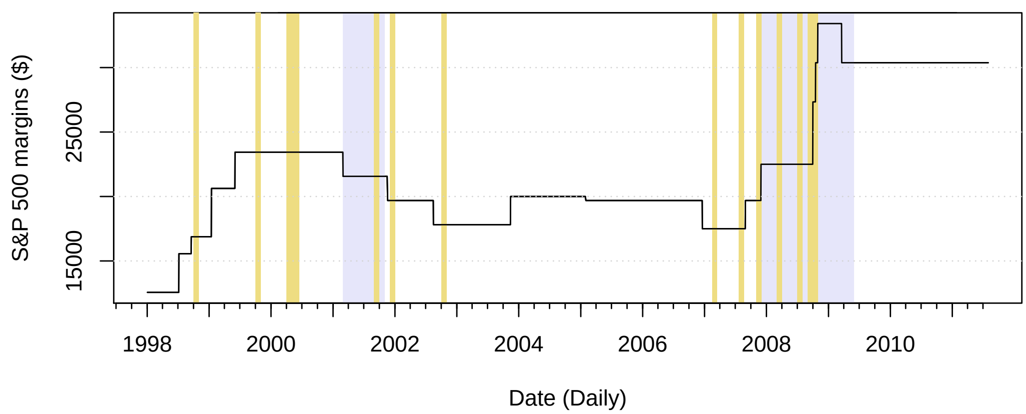
[7] http://www.cmegroup.com/clearing/risk-management/historical-margins.html

Table 2.1: We compute the annualized volatility of the four assets and their mean margin requirements. We note that higher the volatility of the asset, higher is its margin requirement. That is consistent with the hypothesis that lenders use the volatility of an asset to set its margin requirement.

Asset	Annualized Volatility	Average Margin (%)
S&P 500	21.2%	8.3%
Dow Jones	19.4%	8.0%
NZD / USD FX	14.0%	3.5%
AUD / USD FX	12.7%	2.9%

consistent with the hypothesis that lenders use the volatility of the asset to set its margin requirement. We conduct further analysis in Section 2.6 in the form of event-studies.

Figure 2.1: Margin requirement (in dollars) on the S&P 500 futures contract. It is the dollar amount required to hold a position in the contract. Note that the futures contract is on 250 units of the S&P 500 index.



2.2.2 General Collateral Repo Spread

A general collateral repo transaction is another form of secured lending. In a repurchase agreement (repo), a financial institution (the borrower) sells a security and pledges to buy it back from the other party (the lender) at a later date. The security is transferred to the lender and is used as a collateral in case of default. To protect the lender from changes in the market value (i.e., liquidation value) of the security, a "haircut" is applied to the

Figure 2.2: Margin requirement (in percent) on the S&P 500 futures contract. We compute the percentage value as follows: $\frac{Dollar\ margin}{S\&P500\ index \times Contract\ size}$

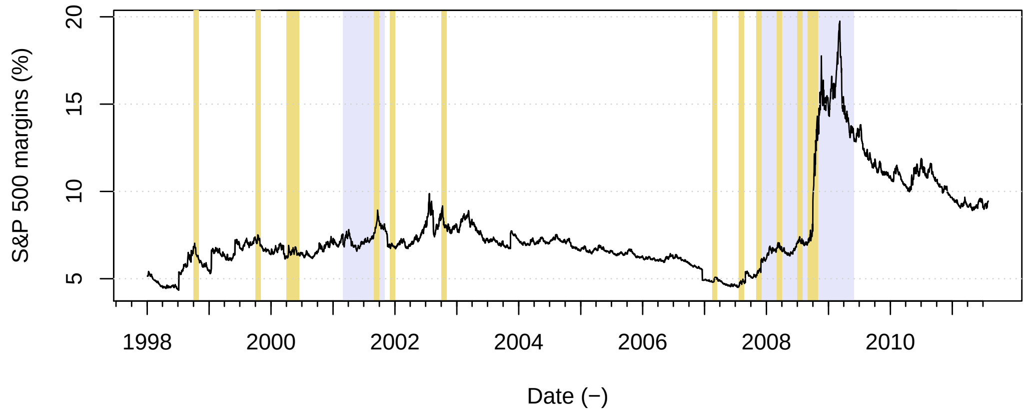
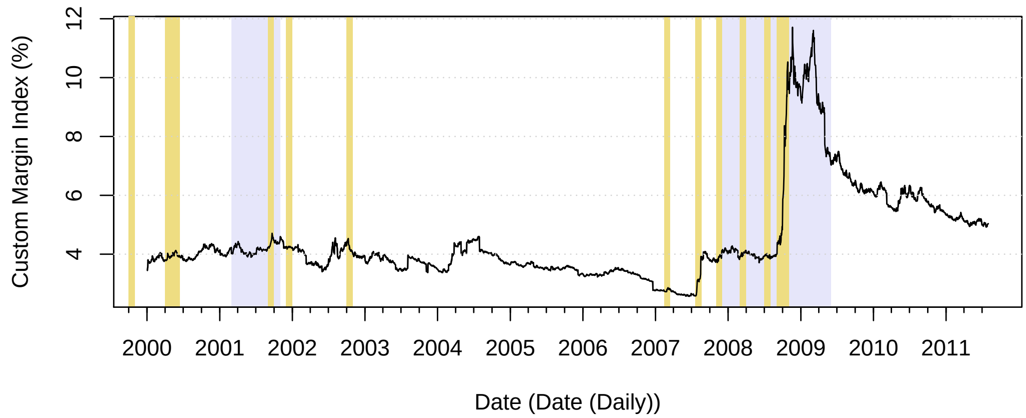


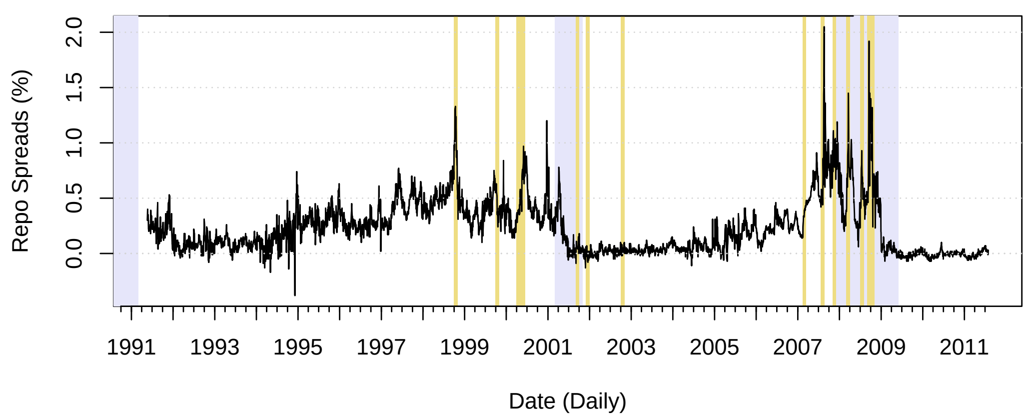
Figure 2.3: Custom Futures Margin Index. We construct an equally weighted index of percentile margin requirements on three futures contracts: 1) the S&P 500 stock index, 2) the AUS/USD foreign exchange, and 3) the NZD/USD foreign exchange



security, valuing it at a price lower than the market value. The security is later purchased by the borrower at a price higher than the original sale price; the difference is effectively the interest rate on the loan, known as the *repo rate*. The Federal Reserve uses repos to make collateralized loans to broker-dealers via auctions over various types of collateral[8] [Kea96]. The typical term of the loan is overnight, but they can exist with terms up to 65 days. Broker-dealers borrow through repos using listed securities as collateral in order to finance their operations (such as buying inventory and funding their customers' margin accounts). [AS10] shows that financial institutions adjust their total leverage through repos and reverse repos.

[AB08] used the difference between the three-month repo rate and the yield on three-month U.S. Treasury Bill as a proxy for short term funding liquidity risk. This funding measure is called the *general collateral repo spread* and is one of the several empirical funding measures used in this chapter to construct our Funding Liquidity Stress Index. The yield on three-month U.S. Treasury Bill is published by the Board of Governors of the Federal Reserve System. Historical values can be obtained from the Federal Reserve Economic Data portal[9]. We obtain the historical values of the three-month General Collateral Repo rate from Bloomberg. In Figure 2.4, we plot the historical values of the spread. We note that it is very sensitive to the level of financial stress in the market as it spikes during most periods of stress. The shaded areas in the figure correspond to periods of stress.

Figure 2.4: General Collateral Repo Spread. It is the difference between the three-month general collateral repo rate and the yield on three-month U.S. Treasury Bill.



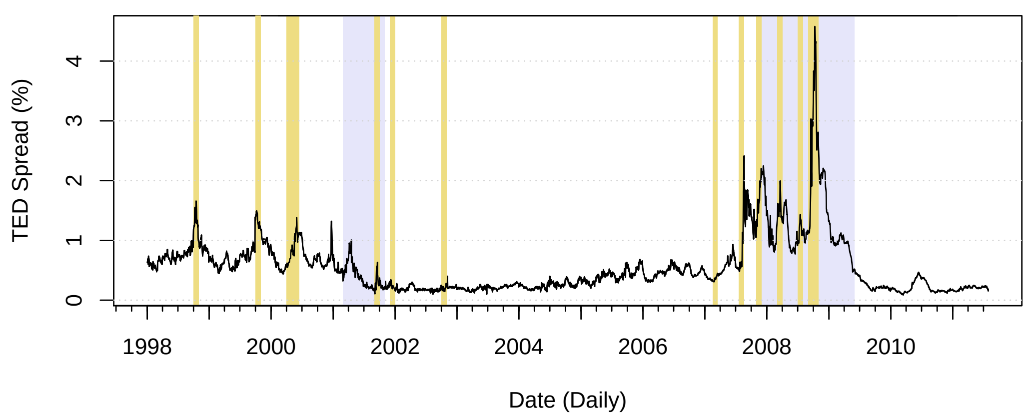
[8] http://www.newyorkfed.org/aboutthefed/fedpoint/fed04.html
[9] http://research.stlouisfed.org/fred2/series/DTB3

2.2.3 Libor-based Measures

To meet reserves requirements, banks borrow unsecured funds from each other in the "*interbank lending market*". The majority of the loans are overnight. The London Interbank Offered Rate (LIBOR) is "*the reference rate at which banks borrow unsecured funds from each other in the London interbank lending market*". LIBOR plays a critical role in the financial system. It is used 1) as a reference rate in many financial instruments[10], 2) to price derivatives on financial securities, and 3) to finance arbitrage operations in major banks.

Since the loans bear some risk, the banks charge each other a rate larger than the yield on Treasury bills. The spread between the three-month LIBOR and the yield on the three-month U.S. Treasury Bill is a widely used measure of the health of the financial system and the state of funding liquidity ([MSW08] [GS00] [CT03]). The measure is called the TED spread. As shown in Figure 2.5, the TED spread spiked during all periods of financial stress, and that is for the following two main reasons. First, banks hoard cash during periods of stress in anticipation to difficulties in obtaining other forms of financing[11] needed to roll-over their short-term debt. Second, an increase in counterparty risk and information asymmetry makes banks more reluctant to lend during periods of stress [MU08].

Figure 2.5: TED Spread. It is the difference between the three-month LIBOR and the yield on the three-month U.S. Treasury Bill. We note that the TED spread spiked during all periods of stress.

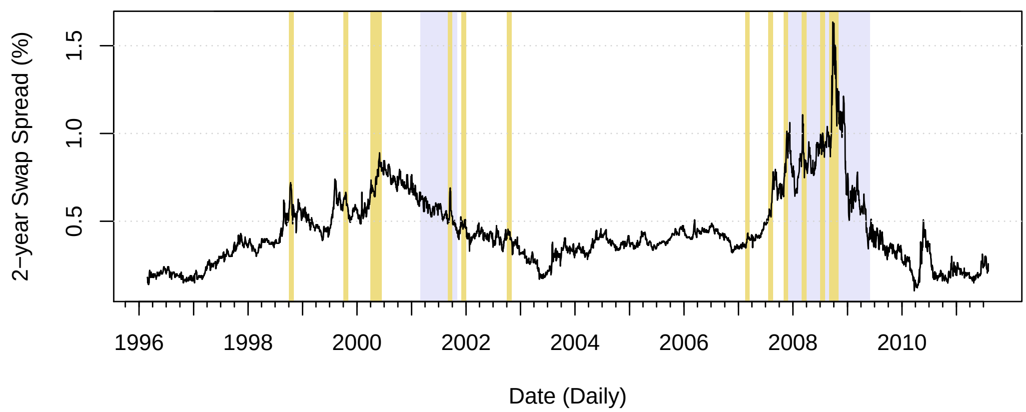


[10]LIBOR is used as a reference rate in many financial instruments such as variable rates mortgages, forward rates agreements, interest rates swaps and futures

[11]Other forms of financing are asset-backed commercial paper, repurchase agreements and securitization.

Another important LIBOR-based funding measure is the two-year swap spread. An interest rate swap is a contract to exchange fixed for floating rate payments. The floating rate is typically the LIBOR while the fixed rate is a specified rate (the spread) on top of the yield of a Treasury security of the same maturity. The spread is positive and captures funding liquidity risk and counterparty risk [Gri02]. In Figure 2.6, we plot the two-year swap spread. Similar to the TED spread, it spiked during most periods of stress as funding and counter-party risk increased.

Figure 2.6: Two-year swap spread.

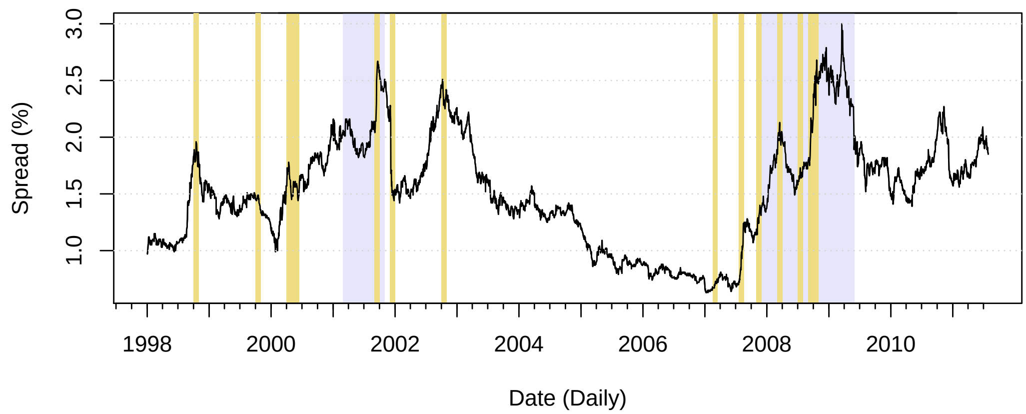


2.2.4 Corporate Credit Yield Spreads

Corporations issue bonds to finance their operations. Credit rating agencies, such as Moody's, Fitch and Standard & Poor, assess the credit worthiness of a bond and assign a rating in the form of letters. The credit rating determines to a large extent the cost to issue debt in the primary market and affects credit yields in the secondary market. In this chapter, we use Moody's credit rating to categorize corporate bonds. Bonds rated below Baa are called junk bonds (or high-yield bonds), while those rated higher are called investment grade bonds. Those rated Aaa (i.e., the highest rating) are called prime bonds.

Bonds issued by the US Treasury and certain major corporations are rated Aaa by Moody's and are considered the safest bonds in the market. The yields of corporate Aaa bonds are however higher than the yield of US Treasury bonds, with similar maturity, for few reasons. First, US Treasury bonds have higher collateral value and can be used to obtain repo loans at favourable rates [Sen11]. Second, US Treasury bonds are more liquid than the highest-rated corporate bonds. As a result, the difference between yields of corporate prime bonds and the yield of US Treasury bonds is positive and increases during periods of stress, as flight-to-liquidity gains traction. In Figure 2.7, we plot the spread between *Moody's Seasoned AAA Corporate Bond Yield*[12] and the yield of ten-year Treasury bonds. We observe that the spread increases during periods of financial stress.

Figure 2.7: Aaa/10-year Treasury Spread. It is the difference between *Moody's Seasoned AAA Corporate Bond Yield* and the yield of ten-year Treasury bonds. We observe that the spread increases during periods of financial stress.

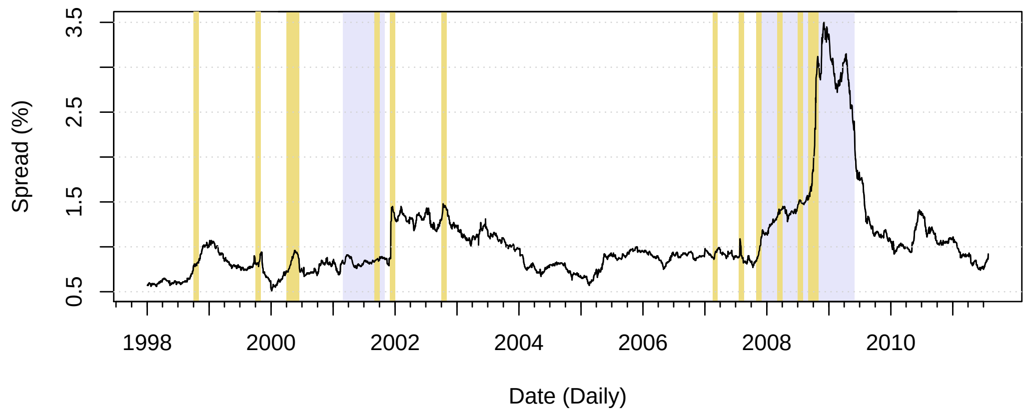


Baa is the lowest credit rating in the investment grade category, while Aaa is the highest. The spread between Baa bonds yields and Aaa bonds yields increase during periods of stress [LMN05] [DNFL11]. The increase is due to 1) a higher perception of risk of Baa bonds, and 2) an increase in information asymmetry as certain bonds are perceived riskier than others in the same Baa category. The issue related to information asymmetry is known as the 'adverse selection problem' or the 'external finance premium' [BG95] [DG08]. In Figure 2.8, we plot the spread between *Moody's Seasoned Baa Corporate Bond Yield*[13] and *Moody's Seasoned AAA Corporate Bond Yield*. Similar to Figure 2.7, the spread increased significantly during the $2007-2009$ financial crisis.

Bonds rated below Baa are called high-yield or junk bonds [HK96] [Mil99]. The difference between the yields of junk bonds and the yields of Baa rated bonds is very sensitive to flight-to-quality [GL99] and flight-to-liquidity [Kwa01]. That is for the following three reasons. First, the difference in credit risk between junk bonds and Baa rated bonds is

[12] Available from the Federal Reserve Economic Data portal.
[13] Available from the Federal Reserve Economic Data portal.

Figure 2.8: Baa/Aaa Corporate Spread. It is the difference between *Moody's Seasoned Baa Corporate Bond Yield* and *Moody's Seasoned AAA Corporate Bond Yield*



higher than the difference between the Baa and Aaa rated bonds. Second, junk bonds have lower liquidity as they are issued in smaller quantities and the majority of institutional investors do not invest in them. Third, due to the low credit rating and sensitivity to economic cycles, investors in junk bonds are subject to adverse selection effects more so than investors in Baa bonds. In Figure 2.9, we plot the spread between *Merrill Lynch US High Yield Effective Yield*[14] and *Moody's Seasoned Baa Corporate Bond Yield*. Similar to the other corporate spreads, it dramatically increased during periods of financial stress. At the peak of the crisis, the spread exceeded 6 standard deviations due to severe flight-to-quality and flight-to-liquidity.

2.3 Funding Liquidity Stress Index

2.3.1 Constructing the Index

In the previous section, we presented seven funding measures grouped into four different categories. Each category captures one aspect or another of the funding liquidity (see Table 2.2). In Table 2.3, we list the seven funding measures. In the middle column, we list the starting date at which each measure became available. In the last two columns, we compute

[14] Also available from the Federal Reserve Economic Data portal `http://research.stlouisfed.org/fred2/series/BAMLH0A0HYM2EY`.

Figure 2.9: Junk Bonds/Baa Corporate Spread. It is the difference between *Merrill Lynch US High Yield Effective Yield* and *Moody's Seasoned Baa Corporate Bond Yield*.

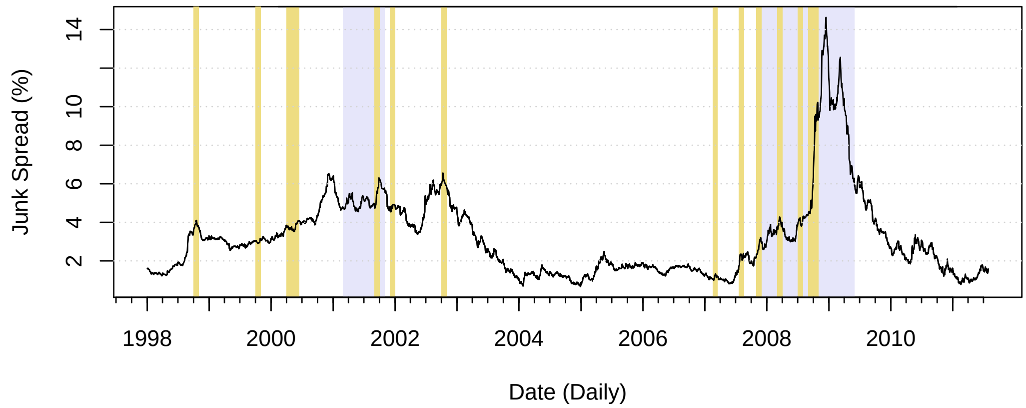


Table 2.2: Four different categories of funding measures.

Category of Funding Measures	**Aspects of Funding Liquidity Captured**
Margins on Futures Contracts	Cost to finance positions in various instruments such as indices, commodities and currencies
Collateral Repo Spreads	Cost of secured lending to broker-dealers
Libor-based Measures	Cost of unsecured lending between major banks
Corporate Credit Yield Spreads	Cost of borrowing by corporations. It also reflects the balance sheet strength of various financial institutions since they are major holders of corporate bonds.

the mean and standard deviations over the period ranging from 2000/01/04 to 2011/07/29 which is the longest period common to all seven measures. In Table 2.4, we compute the correlation matrix over this entire period. We observe that the majority of the correlation coefficients are high and positive. Some of the pair-wise correlation coefficients are however small or even negative. That is a good property as it indicates that the funding measures capture different aspects of the funding liquidity. However we observe that during periods of stress, such as the $2007 - 2009$ financial crisis, all correlation coefficients increase and become positive (see Table 2.5) showing commonality in the worsening of all aspects of funding liquidity.

Our objective is to construct a funding liquidity stress index composed of the seven

Table 2.3: We list the seven Funding Measures. The means and standard deviations are computed over the period from 2000-01-04 to 2011-07-29. It is the longest period common to all measures.

Category	**Funding Measure**	**Date Available**	**Mean**	**Standard Deviation**
Futures Margins	Custom Futures Margin Index	2000-01-04	5.10%	1.60%
Repo Spreads	General Collateral Repo Spread	1991-05-21	0.20%	0.26%
Libor Measures	TED Spread	1985-01-02	0.51%	0.50%
	2-year Swap Spread	1996-02-26	0.47%	0.22%
Corporate Credit Yield Spreads	Aaa / 10-year Spread	1983-01-03	1.55%	0.50%
	Baa / Aaa Spread	1986-01-02	1.11%	0.51%
	Junk / Baa Spread	1991-05-21	3.27%	2.32%

Table 2.4: Correlation matrix of all seven funding measures listed in Table 2.3 over the period ranging from 2000-01-04 to 2011-07-29. It is the longest period common to all measures.

	2-year Swap Spread	TED Spread	Aaa / 10y Spread	Baa / Aaa Spread	Junk / Baa Spread	Gen Repo Spread	Futures Margin Index
2-year Swap Spread	1.00						
TED Spread	0.85	1.00					
Aaa/10y Spread	0.23	0.17	1.00				
Baa/Aaa Spread	0.35	0.46	0.56	1.00			
Junk/Baa Spread	0.51	0.42	0.76	0.79	1.00		
Gen Repo Spread	0.73	0.81	-0.15	0.01	0.07	1.00	
Futures Margin Index	0.14	0.27	0.65	0.80	0.70	-0.18	1.00

Table 2.5: Correlation matrix computed over a shorter time period, the 2007−2009 financial crisis.

	2-year Swap Spread	TED Spread	Aaa / 10y Spread	Baa / Aaa Spread	Junk / Baa Spread	Gen Repo Spread	Futures Margin Index
2-year Swap Spread	1.00						
TED Spread	0.88	1.00					
Aaa/10y Spread	0.88	0.80	1.00				
Baa/Aaa Spread	0.73	0.71	0.85	1.00			
Junk/Baa Spread	0.82	0.83	0.90	0.97	1.00		
Gen Repo Spread	0.52	0.66	0.38	0.12	0.26	1.00	
Futures Margin Index	0.69	0.77	0.80	0.95	0.95	0.21	1.00

funding measures. We use Principal Component Analysis (PCA) [Flu88] to find the most appropriate linear combination of these measures. Many financial stress indices use PCA to weight their variables. Important ones are 1) the St. Louis Feds Financial Stress Index (STLFSI) [SL10], 2) the Kansas City Fed's Financial Stress Index (KCFSI) [HK09], and 3) the Chicago Feds National Financial Conditions Index (NFCI) [BB11] [BB10]. There are two major differences between our funding liquidity stress index and those three indices. First, we focus solely on funding conditions whereas those three financial stress indices cover general financial and economic conditions. Second, our funding liquidity stress index is *daily* whereas the NFCI and the STLFSI are *weekly* indices, and the KCFSI is a *monthly* index. Daily measurements enable us to capture quick changes in funding conditions and to better study the relationship between the index and the various arbitrage strategies. Since most arbitrage strategies and funding measures are mean-reverting, using weekly or monthly frequency prevents us from capturing extreme events that occur in between.

We use PCA to identify the factor that accounts for as much as possible for the variability of our measures. That factor will be used as our funding liquidity stress index. We first standardize our measures by subtracting the mean and then dividing by the standard deviation. The standardized measures will therefore have equal variances and our index won't be biased towards any measure in particular. Next, we compute the eigenvectors using singular value decomposition of the data matrix. The eigenvectors (or sets of coefficients) are used to construct the principal components (or factors), by providing the loadings of the funding measures. We choose the principal component with the highest variance as our funding liquidity stress index. The corresponding eigenvector provides the weights of

Table 2.6: Importance of principal components.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard deviation	1.92	1.49	0.72	0.51	0.40	0.34	0.21
Proportion of Variance	0.53	0.32	0.07	0.04	0.02	0.02	0.01
Cumulative Proportion	0.53	0.85	0.92	0.95	0.98	0.99	1.00

Table 2.7: Coefficients of funding measures.

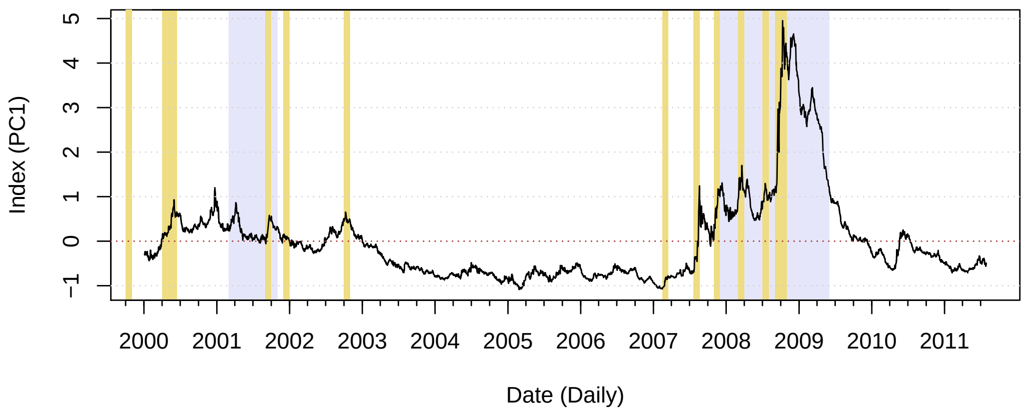
	Funding Liquidity Stress Index (PC1)	PC2
2-year Swap Spread	0.20	0.28
TED Spread	0.19	0.28
Aaa/10y Spread	0.19	−0.21
Baa/Aaa Spread	0.23	−0.14
Junk/Baa Spread	0.24	−0.12
Gen Repo Spread	0.10	0.40
Futures Margin Index	0.20	−0.23

the funding measures. In Table 2.6, we compute the standard deviation of each principal component and the corresponding proportion of total variance explained by each principal component. The principal component PC1 explains 53% of the variability in the data, followed by PC2 which explains 32%. Together, PC1 and PC2 explain 85% of the data. The final step is to standardize PC1 and PC2 so their mean is zero and standard deviation is one. In Table 2.7, we list the resulting weights of the funding measures corresponding to PC1 and PC2. We will call PC1 the *Funding Liquidity Stress Index* (FLSI).

Not unexpectedly, all the coefficients of the Funding Liquidity Stress Index are positive, indicating that any increase in the funding measures (i.e., any deterioration in funding conditions) increases the value of the index. The coefficients range from a low of 0.10 for the *General Collateral Repo Spread* to a high of 0.24 for the *Junk/Baa Corporate Yield Spread*. Since each funding measure is standardized, a one standard deviation move by a funding measure corresponds to a change in the index equal to the variable's coefficient. For instance, a one standard deviation move by the *Junk/Baa Corporate Yield Spread* affects the FLSI index 2.4× more than a one standard deviation move by the *General*

Collateral Repo Rate. We plot the FLSI index in Figure 2.10. The index spiked during all periods of financial crisis as funding conditions deteriorated.

Figure 2.10: The Funding Liquidity Stress Index (FLSI).



The second principal component (PC2), plotted in Figure 2.11, is interesting as well. The coefficients of the seven funding measures, corresponding to PC2, are listed in Table 2.7. The coefficients of the *2-year swap spread*, the *TED spread* and the *general collateral repo spread* (group A) are positive, while those of the *Aaa/10y spread*, the *Baa/Aaa Corp spread*, the *Junk/Baa Corp spread* and the *futures margin index* (group B) are negative. That categorization is not just statistical, but makes sense in finance. Measures in group B are funding measures derived from investable assets such as stocks, corporate bonds and currencies. Those measures therefore reflect the risk premia priced in the underlying assets. So when assets are in a bubble state (i.e., when investors are overly optimistic and risk is mispriced), funding liquidity is artificially deflated. Measures in group A, however, are rates negotiated between two parties and are not derived from investable assets.

In Table 2.8, we calculate the average values of the FLSI index and PC2, over various periods. We observe that the averages are high and positive during periods of stress, and negative during 'normal' periods. The averages are zero over the entire period since the FLSI index and PC2 have been standardized.

2.3.2 Idiosyncratic Volatility of Banks' Stock Prices

The banking sector is the most sensitive sector of the economy to funding conditions. Their assets have longer maturity than their liabilities and, therefore, they rely on a healthy

Figure 2.11: The second principal component (PC2) .

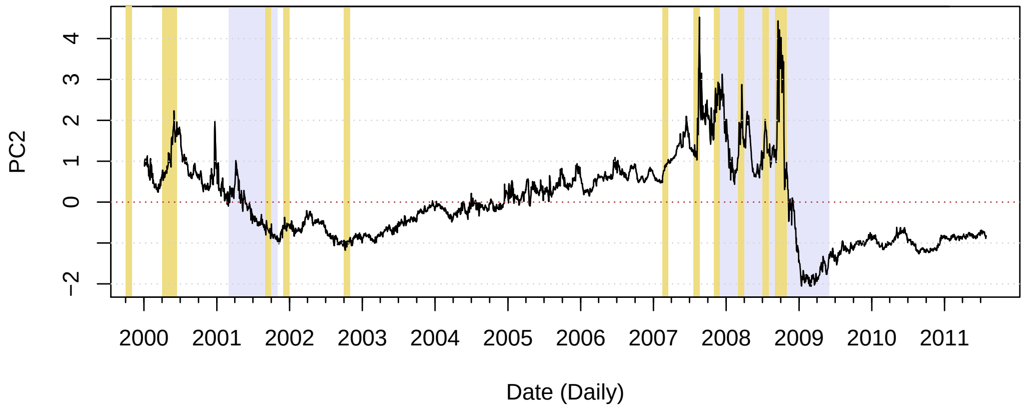


Table 2.8: We calculate the average values of the FLSI index and PC2, over various periods. We observe that the averages are high and positive during periods of stress, and negative during "normal" periods. The averages are zero over the entire period since the FLSI index and PC2 have been standardized.

	Funding Liquidity Stress Index	PC2
07-09 crisis (see Table A.3)	1.17	1.79
98-02 crisis (see Table A.2)	0.35	0.17
Recessions (see Table A.1)	1.44	0.20
"Normal" periods	−0.36	−0.11
All periods	0.00	0.00

funding environment to roll-over their maturing short-term debt [BO10] [Bru09]. If our Funding Liquidity Stress Index is a good measure of funding conditions, then it should have a strong explanatory power for the idiosyncratic volatility (i.e., risk) of banks' stock prices. We define the idiosyncratic volatility as the downside deviation of unexpected returns. Unexpected returns are the returns in excess of what is expected from changes in the general equity market.

The KBW Bank Index is the most widely used index of U.S. banks stocks and is used to conduct our analysis. We use the returns of the S&P 500 index to proxy the returns of the general equity market. The time-varying downside deviation of unexpected returns is computed over a two-month rolling window from 1995/01/01 to 2011/06/01. For instance, to compute the downside deviation on day t, we perform the following four steps: 1) first we regress the daily returns of the KBW Bank index against the daily returns of the S&P 500 over the period from $t - 300$ to $t - 60$ (i.e. the 8 months immediately prior to the two-month window) to obtain the *beta* of the bank index, 2) next we compute the expected returns over the period from $t - 60$ to t (i.e. over the two-month window) by multiplying the daily returns of the S&P 500 index by the *beta*, 3) the *unexpected returns* are then calculated by subtracting the daily expected returns from the daily bank index returns over the period from $t - 60$ to t, and 4) we finally compute the downside deviation of these daily *unexpected returns*. The methodology is the same as in [HK09], [HH09] and [Smi84].

In Figure 2.12, we plot the value of the idiosyncratic volatility of banks' stock prices. We observe a spike in idiosyncratic volatility during the financial crisis of $2007 - 2009$, a period widely known as a banking crisis [Gor08] [Bru09]. In Table 2.9, we regress the idiosyncratic volatility of banks' stock prices against the Funding Liquidity Stress Index (FLSI). The factor loading is positive and R^2 equals 78%. The FLSI index therefore has a strong explanatory power for the idiosyncratic volatility of banks' stock prices. That constitutes another evidence that the custom Funding Liquidity Stress Index is a good proxy to funding conditions.

2.4 Classic Arbitrage Strategies

Next we reproduce three classic arbitrage strategies: 1) closed-end funds arbitrage, 2) mergers and acquisitions arbitrage, and 3) on-the-run/off-the-run Treasury arbitrage.

Figure 2.12: Idiosyncratic Volatility of Banks Stock Prices. We observe a spike during the financial crisis of $2007 - 2009$, a period widely known as a banking crisis

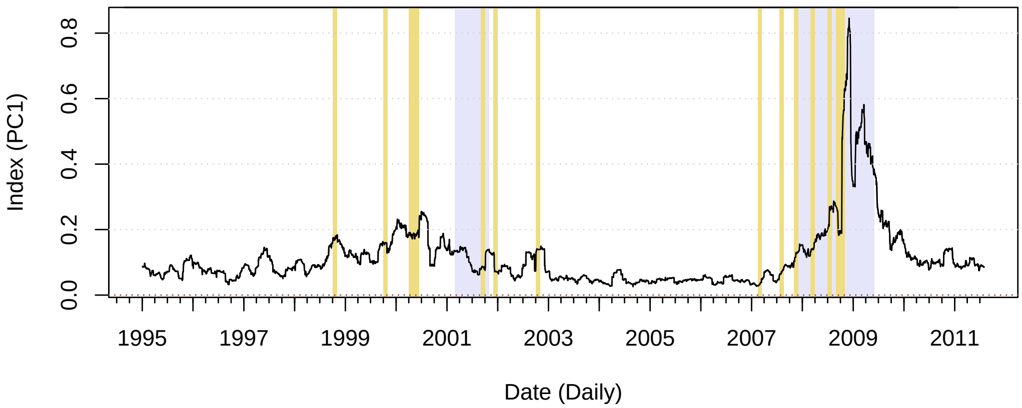


Table 2.9: We regress the idiosyncratic volatility of banks' stock prices against the Funding Liquidity Stress Index (FLSI). The factor loading is positive and R^2 equals 78%. The FLSI index therefore has a strong explanatory power for the idiosyncratic volatility of banks' stock prices.

	(1) idiosyncratic volatility
(Intercept)	0.122***
	(0.001)
FLSI.pc1	0.103***
	(0.001)
R^2	0.783
$adj.R^2$	0.783
N	2823

Standard errors in parentheses

* ($p \leq 0.1$), ** ($p \leq 0.05$), *** ($p \leq 0.01$)

2.4.1 Closed-End Funds Arbitrage

A Closed-End Fund (CEF) is an investment company that raises capital by selling a fixed number of shares through an Initial Public Offering (IPO). The capital is used to purchase an investment portfolio[15]. After the IPO, the shares trade on the secondary market[16] where the market price is affected by the supply and demand of the security. The market price can therefore deviate from its fundamental value known as the Net Asset Value (NAV). The NAV is equal to the value of the investment portfolio divided by the number of shares outstanding. We refer to the deviation of the market price from its NAV as a *spread*. If the CEF trades above the NAV, the spread is positive and we say that it trades at a premium. Conversely, if it trades below the NAV, the spread is negative, and we say that it trades at a discount. According to textbook arbitrage, sophisticated traders can arbitrage these spreads. For instance, if a CEF trades at a discount, they will buy the CEF and short a basket of securities similar to the CEF's underlying investment portfolio[17]. In doing so, they collectively exercise trading pressures that push the market price closer to its NAV.

In practice, however, CEFs trade at various discounts and sometimes at a premium. A large number of studies have been performed to study this phenomenon. [GT] and [Pon96] provide evidence that CEF spreads persist because various costs affect the ability of rational agents to profit from the mispricing. In particular, they show that the spreads vary according to the difficulty in shorting the underlying securities of the NAV. [Mal77] and [LST91] argue that pricing theories based on the fundamentals of the investment portfolio or management fees have had little, if any, ability to explain the mispricing. They state that fluctuations of CEFs' market prices are mostly driven by changing investors sentiments. The vulnerability of CEFs makes them riskier than their underlying investment portfolios, and as a result, they trade at a discount from their NAV.

Our study on CEFs is somewhat complementary to the literature. We don't aim to explain the cross-sectional differences of CEF spreads. Our objective is however to study the changes in CEF spreads through time and under various market conditions. [Mal77] and [LST91] assert that spreads are affected by retail investors sentiment. We provide strong evidence of other factors. We show that the ability of sophisticated investors to conduct arbitrage significantly affects the spreads of CEFs. In particular, we show that 1) funding liquidity has a strong explanatory power for the changes in CEFs spreads, and 2) the spreads are highly correlated with the spreads of other arbitrage strategies.

[15] More information is available on the Security and Exchange Commission (SEC) at http://www.sec.gov/answers/mfclose.htm.

[16] i.e., on a stock exchange such as NYSE and NASDAQ

[17] They place short positions to hedge future changes in the net asset value.

Table 2.10: The number of CEFs in different asset classes.

Asset Class	# CEFs
Debt	282
Equity	146
Many Classes	35
Commodity	2

In other words, deteriorating funding conditions that limits the arbitrageur from using leverage (observation 1) and the reduction of his capital due to losses in other strategies (observation 2) hinder his ability to conduct arbitrage and as a result causes CEF discounts to widen.

CEFs have different investment objectives and invest in different asset classes. There are approximately 600 CEFs listed in the U.S. according to Bloomberg. We remove those that have missing information and those that are chronically illiquid[18]. The universe is reduced to 465 CEFs, representing more than 94% of total assets managed by closed-end funds in the US. In Table 2.10, Table 2.11 and Table 2.12 we show the number of CEFs in different asset classes, different investment objective categories and different size categories, respectively. In Figure 2.13, we plot the number of CEFs indexed in time since January 1st, 1999. The shaded areas correspond to periods with unusual events. In particular, the areas in yellow correspond to periods of stress whereas the areas in grey correspond to official U.S. recessions. For more information, refer to Appendix A. We note that the number of CEFs doubled from 2001 to 2011.

In Figure 2.14 we plot the daily median of the spreads of all outstanding CEFs. The median varied between 0% and negative 30% in the past 12 years. The past 12 years were characterized by two long periods of crisis, each followed by a recovery period. The first period of crisis included the dot-com crash, September 2001 attacks, various corporate accounting scandals, and the 2001 U.S. recession (for a complete list, refer to Table A.2). During that period, the median reached a low of negative 15%. It narrowed back to zero in the following recovery period. The second period of crisis corresponds to the 2007-2009 financial crisis which included the "Great Recession". During that period, the median reached a 12-year low of negative 30%. It narrowed down to zero in the following recovery. We observe the following pattern: CEF spreads widen during periods of stress and contract in the following recovery period. In Section 2.5, we show that funding conditions,

[18] We used the 20-days average daily volume to assess the liquidity of a closed-end fund. In particular, we removed all CEFs that traded less than $12,000$ shares on average in the last 20 business days prior to June 26th, 2011.

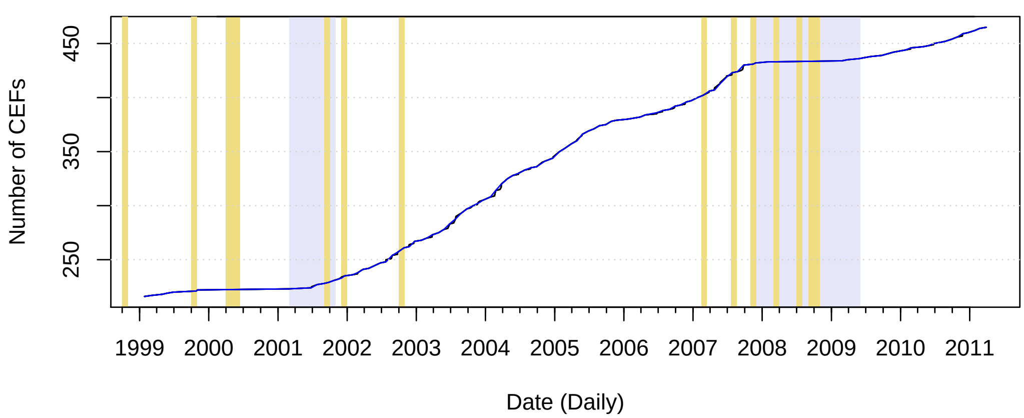
Table 2.11: The number of CEFs in different investment objective categories.

Investment Objective	# CEFs
Municipal	131
Government and Corporate	88
Growth and Income	45
Corporate High Yield	40
Sector Specific	39
Other	39
Country Specific	27
Balanced	26
Income Equity	16
Global Corp Debt	14

Table 2.12: The number of CEFs in different size categories.

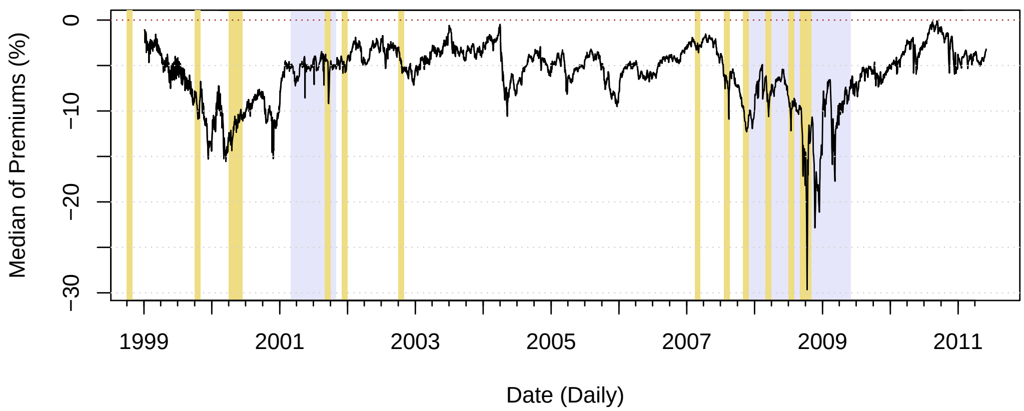
Total Assets (Mln)	# CEFs
$> 1,500$	17
$1,000$ - $1,500$	37
500 - $1,000$	103
250 - 500	140
100 - 250	126
< 100	42

Figure 2.13: The number of CEFs outstanding.



which affects the ability of sophisticated investors to conduct arbitrage, have a significant explanatory power for CEF arbitrage spreads.

Figure 2.14: The median of CEFs premiums.

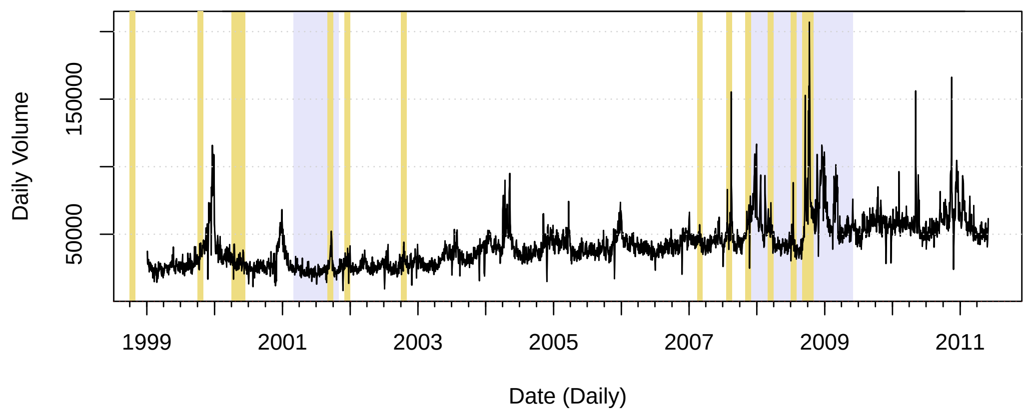


We revisit Figure 2.13 in light of Figure 2.14. The number of CEFs increased every year except two years, 2000 and 2008, in the past 12 years. Those two years correspond to periods of high financial stress and CEFs trading at large discounts. During such periods, investment bankers and CEF sponsors have a hard time raising money for new funds for the following reason. Money is raised in the primary market through an IPO. The newly formed CEF uses the money to pay for IPO expenses and buy the investment portfolio whose value is the new NAV. As a result, the price per share in the primary market (i.e., price at IPO) is larger than the NAV per share. However, in periods of stress, CEFs trade at a large discount to the NAV in the secondary market. That reduces the incentive for investors to buy CEFs in the primary market at a premium.

In Figure 2.15 we plot the median of the daily volumes of CEFs. It is clear that the daily volume is time-varying. We also observe a general trend of increasing daily volumes. It is perhaps the result of increasing liquidity in the market overall, as new trading venues are created and trading commissions drop. More interestingly, we observe spikes in volume during periods of stress, coinciding with sudden widening of the discounts. The coincidence of a sudden increase in volume and a drop in market price provides a clear indication of selling pressures. This phenomenon is perhaps best described by Lasse Pedersen [Ped09]. He provides an analogy between *"The dangers of shouting "fire" in a crowded theatre"* and *"the dangers of rushing to the exit in the financial markets"*.

Arbitrageurs typically use a statistical measure called standard score to evaluate the attractiveness of a particular arbitrage spread. As the name implies, a standard score

Figure 2.15: The median *daily volume* (i.e., number of shares traded) of closed-end funds.



measures the number of standard deviations[19] an observation is above its mean. Arbitrageurs take a long position (a short position) in the arbitrage spread if the standard score is low (high). For instance, a strategy can consist of 1) taking a long position when the standard score falls below -1 and later close the position when it goes to zero, and 2) taking a short position when the standard score rises above $+1$ and later close the position when it goes back to zero[20]. In Figure 2.16, we plot the standard score of the median time-series. We observe that the standard score crosses the interval $[-1, +1]$ many times in the past 12 years, generating arbitrage profits. Arbitrageurs frequently use leverage to boost their returns. They should manage their risk very carefully as arbitrage spreads can widen many standard deviations before narrowing down to zero. For instance, the median of CEFs widened to levels exceeding seven standard deviations during the $2007 - 2009$ financial crisis. In Figure 2.17 we zoom into the $2007 - 2009$ financial crisis. We observe that the standard score dropped temporarily below -1 with unusually high volume during most periods of stress. We hypothesize that closed-end funds suffered from broad sell-off by multi-strategy funds looking to reduce risk and raise cash after incurring losses in other strategies [KL11]. CEF positions were a good source of cash as they were relatively liquid, yet required large margin balances.

[19]The standard deviation used is the historical standard deviation over the whole time period.

[20] Arbitrageurs can change the threshold to -0.5 and $+0.5$ (instead of -1 and $+1$) during periods of low volatility.

Figure 2.16: The *normalized* median of closed-end funds spreads. We normalize the time-series of Figure 2.14 so that the mean is zero and the standard deviation is one. We refer to the normalized spread as standard score. The dashed blue lines correspond to values of +1 and −1 of the standard score.

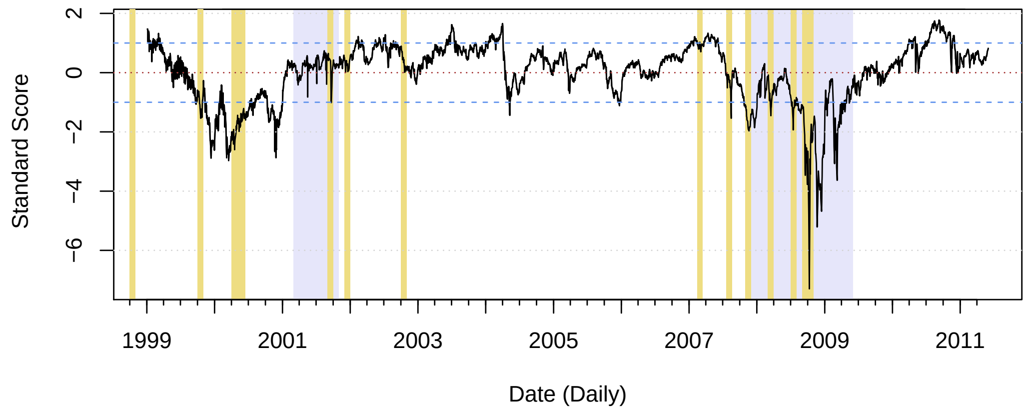
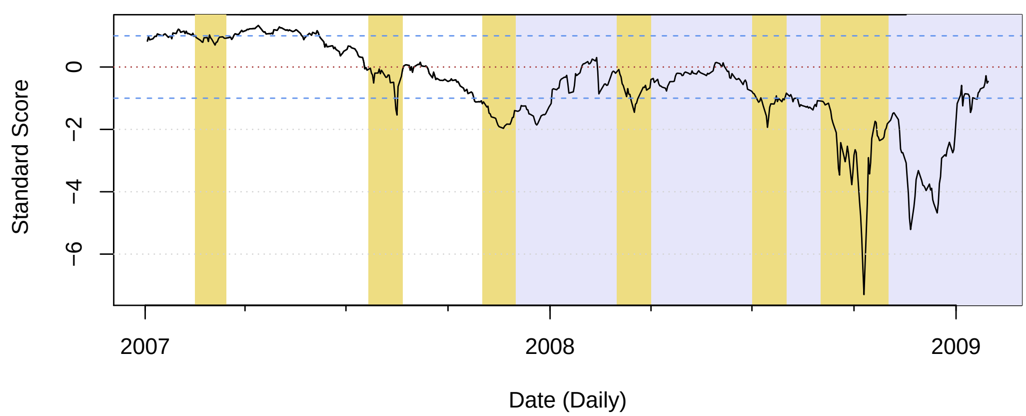


Figure 2.17: A focus on the $2007-2009$ financial crisis. We repeat the plot in Figure 2.16 but zooming over the period ranging from January 2007 to January 2009. Shaded areas correspond to the financial crises listed in Appendix A



2.4.2 Mergers and Acquisitions Arbitrage

Mergers and Acquisitions (M&A) arbitrage is perhaps the most widely used arbitrage strategy. It is commonly known as risk arbitrage. In the event of a merger or an acquisition, a company (the acquirer) announces publicly its intent to purchase another company (the target) at a specific price. The price can be a fixed dollar amount, a certain number of shares of the acquirer, or a combination of both. The arbitrage entails capturing the difference between the price the acquirer has agreed to pay and the price that the target is trading at after the deal is announced[21]. In a "cash" deal, arbitrageurs take a long position in the target and exchange the shares for cash at deal completion. In a "stock" deal, arbitrageurs take simultaneously a long position in the target and a short position in the acquirer. The number of acquirer's shares shorted is equal to the number of shares arbitrageurs expect to receive at deal completion.

Early anecdotal treatment of risk arbitrage was performed by [BM85] and [WP82]. [SR86] and [BR86] examine the relationship between the arbitrage spread and the probability of deal completion. They observe that, under normal conditions, the market properly discriminates between those merger proposals that complete and those that ultimately fail, well in advance of deals completion. In our study, we don't try to explain the relationship between the arbitrage spread and the probability of deal completion. We try however to study the sensitivity of spreads to various market conditions.

Empirical studies on the returns of risk arbitrage report positive annual excess returns. Using a sample of 37 Canadian deals in 1997, [KSoB98] found that the average annualized excess return of a risk arbitrage strategy was 33.9 percent in excess of the TSE 300 index. [DFM92] examined 761 deals in the U.S. between 1971 and 1985. They found that the returns on risk arbitrage in that period of time was exceptionally high as only few arbitrageurs were executing this type of arbitrage. They also show that returns diminish after 1986 as risk arbitrage became more popular. Similar to [SR86] and [BR86], they show that arbitrage spreads were inversely proportional to the probability of deal completion. [Mit01] examines $4,750$ mergers from 1963 to 1998 in the U.S. He finds that risk arbitrage generates 4% annualized excess return after controlling for transaction costs and the non-linear relationship with market returns. [BS02] studies $1,901$ risk arbitrage positions between 1981 and 1996. He reports that a portfolio constructed of those positions has abnormal returns between 0.6% and 0.9% per month.

In what follows, we study $1,417$ risk arbitrage deals between 1997 and 2011. We are not interested in the historical mean return on risk arbitrage or the relationship between

[21] Sometimes investors take speculative long positions in companies that are potential targets. In our study, a merger or an acquisition event is triggered only after the deal is officially announced.

Table 2.13: The number of M&A deals in different size categories.

Deal Size (Mln)	# deals
$> 2,500$	226
$1,000$ - $2,500$	301
500 - $1,000$	269
250 - 500	360
150 - 250	261

arbitrage spreads and the probability of deal completion. We are however interested in 1) the variation of spreads throughout various periods of financial stress, 2) the extent of contagion to/from other arbitrage strategies, and 3) the effects of funding liquidity on the spreads. Clearly, since arbitrageurs have long positions in risk arbitrage, changing spreads affects their profits and losses. If arbitrageurs are highly leveraged, a large widening of the spreads can lead to forced liquidations or even bankruptcies.

We use Thomson Financial SDC database and Bloomberg to construct a list of mergers and acquisitions. We filtered the deals according to the following characteristics:

1. Date Range: 01/01/1997 to 01/01/2011 (14 years)
2. Deal Size: larger than $150 Mln
3. Deal Status: Completed or Terminated (i.e., not currently pending)
4. Country of Target: United States
5. Payment Type: Cash[22]
6. Deal Type: Company Takeover, Tender Offer or Management Buyout
7. Nature of Bid: Friendly or Unsolicited (not Hostile)

$1,255$ deals completed while 162 failed, for a total of $1,417$. In Table 2.13, we show the number of deals in different size categories. As shown in Table 2.14, half the deals take longer than 3 months to either complete or fail. The average is 20 days longer than the median, due to outliers. In Figure 2.18, we plot the number of deals, satisfying our criteria,

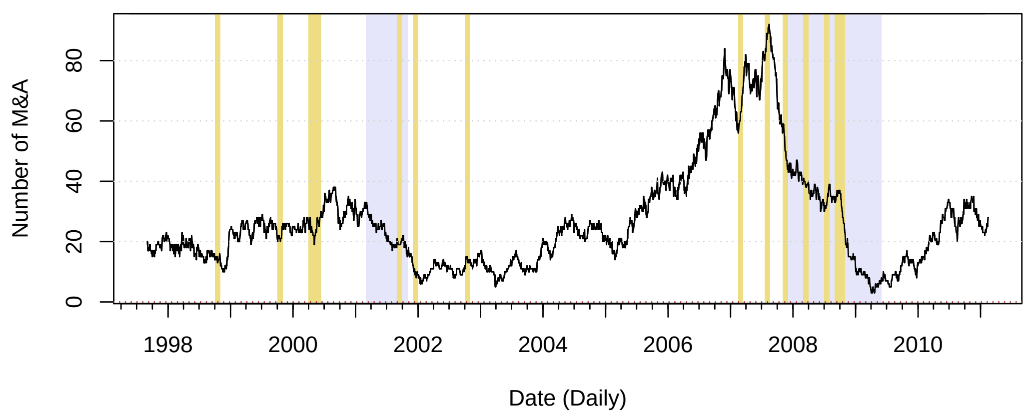
[22]We restricted M&A deals to those that pay only *cash* in order to calculate the spreads accurately. Other forms of M&A deals have a complex pay structure. For instance, some deals offer *Contingent Value Rights* while others have a collar structure for stock exchange.

Table 2.14: Completed vs. terminated M&A deals.

Deal Status	# deals	median lifetime	mean lifetime
Completed	1,255	91 days	109 days
Terminated	162	90 days	121 days
TOTAL	1,417	91 days	111 days

outstanding at any given day. The number varies as deals are completed or terminated, and new ones are announced. The median in the past 14 years was 23. The number is usually higher during periods of improving economic conditions, and lower during periods of stress. A record 93 deals were outstanding just before the downturn started mid-2007. The large number of deals was the result of record low corporate credit yields, loose lending standards by financial institutions and the rise of private equity. The number of deals outstanding reached a low of 3 deals towards the end of the crisis (i.e., early 2009) as all three factors that triggered the boom were muted.

Figure 2.18: The number of deals outstanding



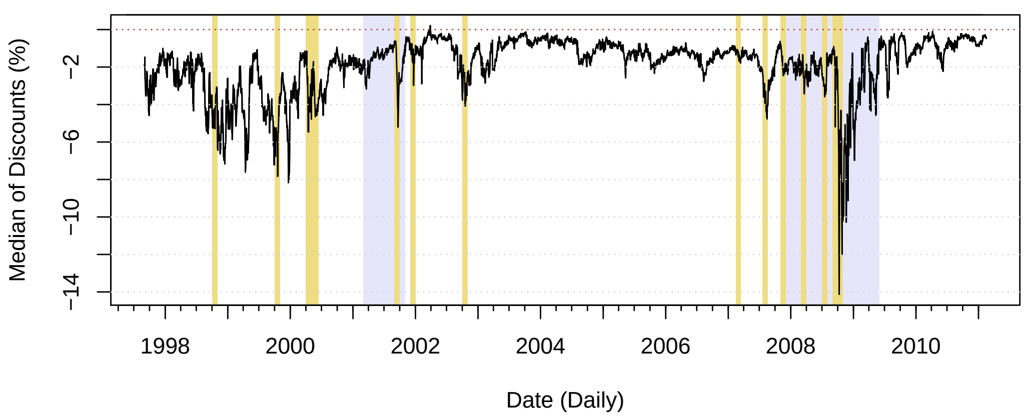
Let p_t be the price of the stock[23] at time t and v the cash amount offered at completion. We compute the arbitrage spread s_t as follows: $s_t = \frac{p_t - v}{p_t}$. If $p_t < v$, the spread is negative and we say that the stock trades at a discount from the offer price. The target company usually trades at a discount to compensate the arbitrageur for 1) the risk that the deal

[23]We identify the common stock of a company being acquired by its 9-digits CUSIP instead of the ticker. We can not use the ticker because it is often reused by another company after the original company is delisted from the exchange. Delisting occurs right after a company is acquired. CUSIP is therefore used to obtain historical stock prices from CRSP and Bloomberg.

might fail, and 2) the risk-free rate. The target company can however trade at a premium if the market expects the acquirer to raise the offer price v.

In Figure 2.19, we plot the daily median of risk arbitrage spreads. We consider only deals that successfully complete (i.e., those that do not fail), representing 89% of our original universe. We do so because our objective is to illustrate the widening of spreads in reaction to liquidity shocks and varying market conditions, rather than to company specific news[24]. In Figure 2.20, we compute the standard score. It is below one standard deviation during most periods of stress[25] and reaches -8 standard deviations at the peak of the $2007-2009$ financial crisis. In Figure 2.21, we plot the median adjusted daily volume[26] of all the target companies. Similar to closed-end funds, the daily volume increases during periods of stress as arbitrageurs are forced to reduce their positions. The combination of unusually high volume and widening arbitrage spreads is evidence of selling pressures that pushes market prices below their fundamentals. In later sections, we show that such forced liquidations are associated with deteriorating funding conditions.

Figure 2.19: The daily median of arbitrage spreads. Note the sharp widening of the median during periods of stress.



[24]Sometimes there is *positive* company specific news that *tightens* the spread of an M&A. We don't have the tools to filter them out. However, by keeping them, we take a conservative stance as we aim to show the extent of spread *widening* during periods of stress.

[25]Recall that periods of stress are represented by yellow regions on the chart

[26]We adjust the daily volume of each company by subtracting its historical median.

Figure 2.20: We normalize the median spread of Figure 2.19. We note that the standard score is below one standard deviation during most periods of stress. Periods of stress are indicated by yellow regions in the chart.

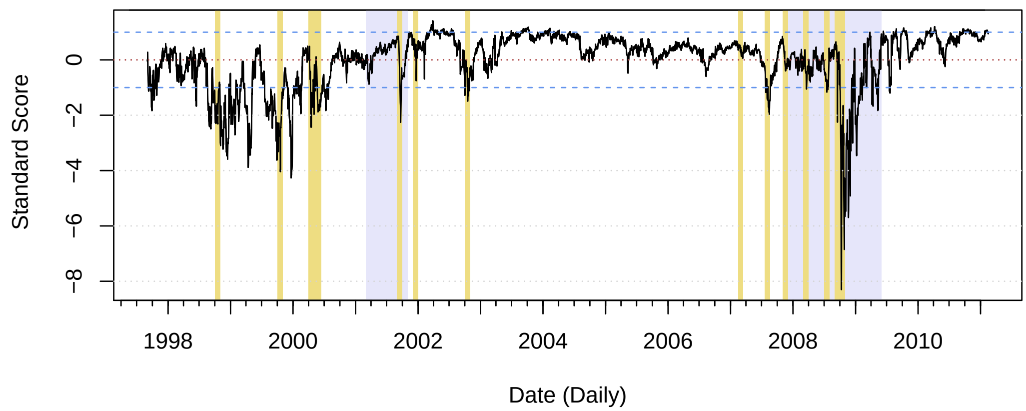
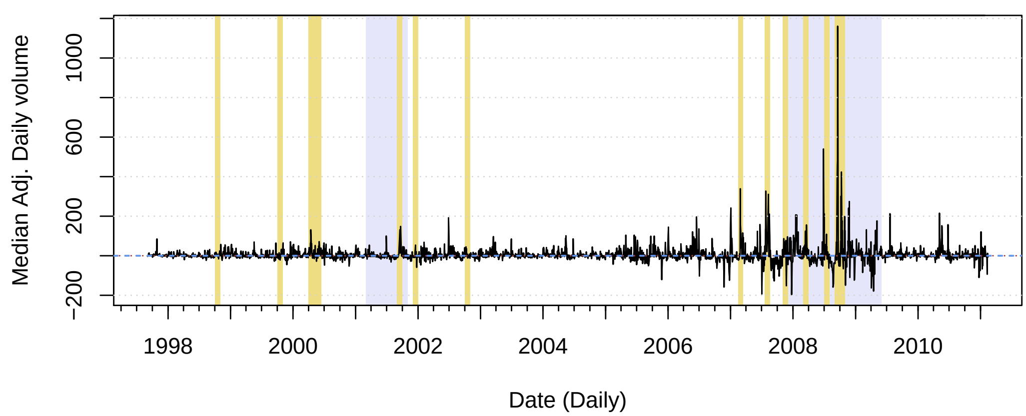


Figure 2.21: We plot the median adjusted daily volume of all target companies. Similar to closed-end funds, the daily volume increases during periods of stress as arbitrageurs are forced to reduce their positions.



2.4.3 Off-the-run/On-the-run Treasury Arbitrage

The US federal government periodically auctions Treasury securities to issue government debt. The most recently issued securities are called on-the-run securities while those issued previously are called off-the-run securities. On-the-run Treasury securities trade at a premium relative to the off-the-run securities for the following reasons. First, on-the-run Treasury securities are more liquid. [SE04] observes that on-the-run issues *"turn over more than fourteen times per week on average and account for* 74 *percent of the total volume in Treasury coupon securities."* In contrast, off-the-run securities have a turnover of just 22 percent. Second, investors can borrow money at a cheaper rate in the repo market when they post on-the-run bonds as collateral (see [VW08] and [FG09]). Third, the bid/ask spread of on-the-run securities is half the bid/ask spread of off-the-run securities [SE04].

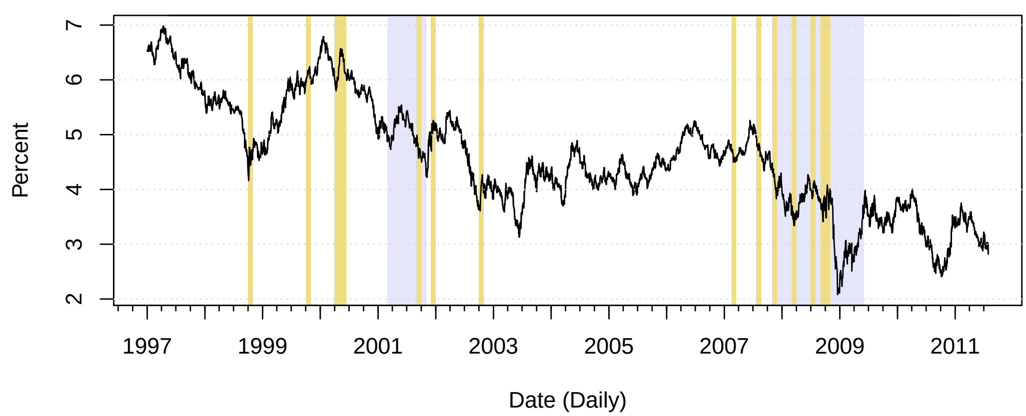
Long Term Capital Management (LTCM) was an active arbitrageur in this space. They bought off-the run securities and shorted on-the-run securities, with similar maturity and coupons, trading at a significant premium[27]. For instance, consider a zero-coupon Treasury bond (A), issued 5 years ago with a maturity of 15 years and a newly issued zero-coupon Treasury bond (B) with a maturity of 10 years. Bond A and bond B have essentially the same payoffs: 1) they don't pay coupons, and 2) they both pay the Par value in 10 years if the U.S. government does not default. Bond A is the off-the-run security while bond B is the on-the run security. If bond B trades at a premium to bond A, an arbitrageur can profit by buying A and shorting B, and wait for the yield differential to disappear. [War92] observes that a portfolio of on-the-run Treasury securities returns 0.55% per annum less than a portfolio of off-the-run securities, with matched duration. Many factors push yields to converge: 1) the on-the-run security becomes off-the-run security once another security with similar maturity is issued, 2) if held to maturity, the two bonds have identical payoffs, and 3) the U.S. Treasury sometimes elects to reduce the government interest expense by buying off-the-run debt and issuing on-the-run debt with lower yield. The yield differential does not however decrease monotonically. During periods of financial stress, spreads widen as investors shun illiquid assets and move into liquid ones, a phenomenon called "flight-to-liquidity". The Federal Reserve monitors the yield differential in addition to other indices to determine whether a "flight-to-liquidity" has occurred and whether the financial system is under stress [FOM98].

In this section, we compute the spread between the on-the-run and off-the-run yields of 10-years U.S. Treasury bonds. The yield of on-the-run Treasury securities is computed

[27] Certain bond indices are required to hold only on-the-run securities and roll their existing positions into new securities when issued. Arbitrageurs like LTCM can be seen as providing liquidity to passive investors like index sponsors.

and officially published by the Federal Reserve in their release entitled "H.15 Selected Interest Rates"[28]. Historical data can be obtained from the Federal Reserve Economic Data portal[29]. The portal is a comprehensive repository of various economic indicator and indices. For the off-the-run yield, we use the 10-year par yield extracted from the off-the-run yield curve computed by [GSW07]. In Figure 2.22, we plot the yield of on-the-run 10-year Treasury bonds. In Figure 2.23 we plot the off-the-run 10-year yield. In Figure 2.24, we plot the yield differential: we subtract the off-the-run yield from the on-the-run yield. The spread varied significantly in the past 15 years. It ranged between -61 bps and -2 bps. "bps" is short for one basis point, i.e., 0.01 percent. We observe that during periods of financial stress, specifically towards the end of 2008 when Lehman Brothers failed and AIG got rescued, the spread widened significantly. In Figure 2.25, we compute the standard score for the period between 1997 and 2011. Spread widening to -61 bps towards the end of 2008 corresponds to a four standard deviations event. We also note that the spread exhibits a mean-reversion process that can lead to profitable arbitrage opportunities. In later sections, we show that spreads widening is associated with worsening funding conditions and contagion from losses in other arbitrage strategies.

Figure 2.22: On-the-run 10 years US Treasury yield.



[28] http://www.federalreserve.gov/releases/h15/
[29] http://research.stlouisfed.org/fred2/series/DGS10

Figure 2.23: Off-the-run 10 years US Treasury yield.

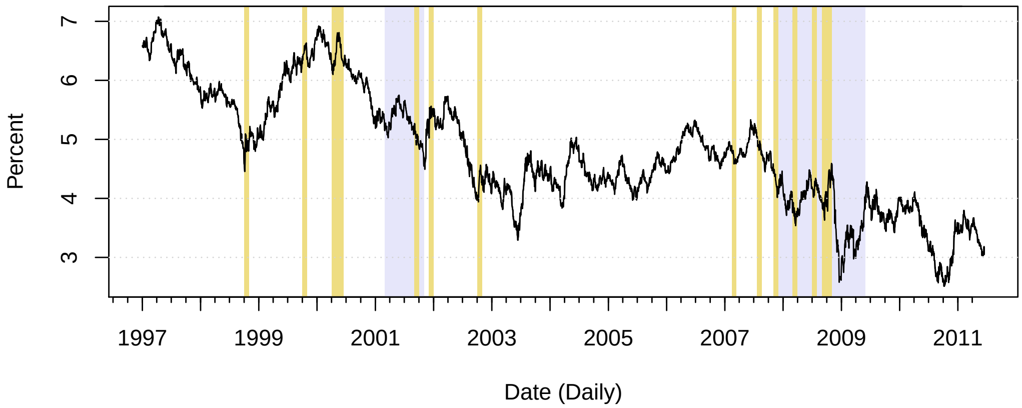


Figure 2.24: The spread between the on-the-run (Figure 2.22) and off-the-run (Figure 2.23) US 10-years Treasury yields. We subtract the off-the-run yield from the on-the-run yied. The negative spread corresponds to a liquidity premium.

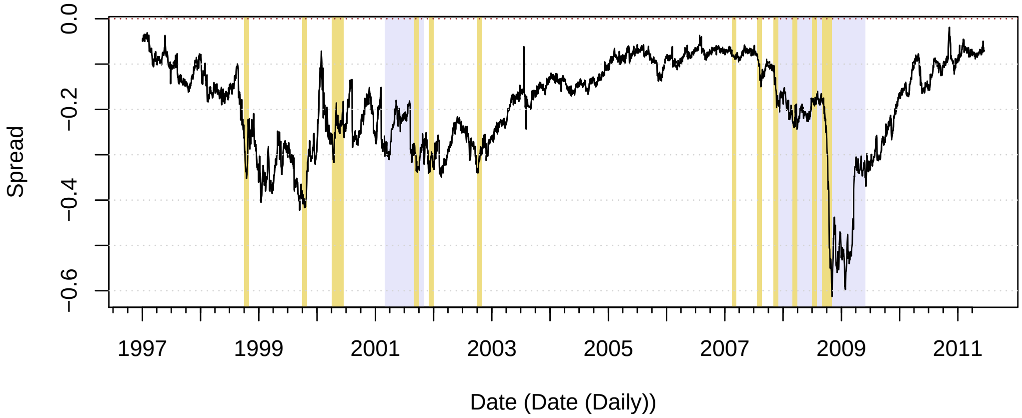
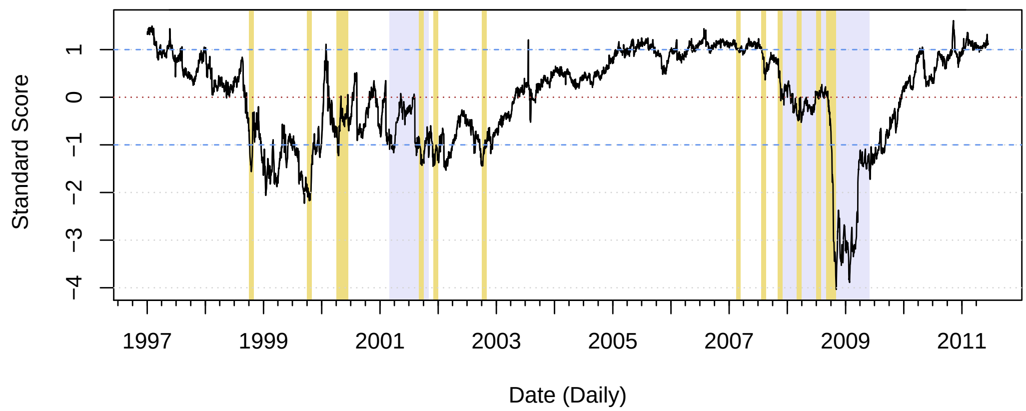


Figure 2.25: We normalize the time-series plotted in Figure 2.24. We subtract the mean and divide by the standard deviation. We make two observations: 1) spread widening towards the end of 2008 corresponds to a four standard deviation event, and 2) the spread exhibits a mean-reversion process that can lead to profitable arbitrage opportunities.



2.5 Relationship Between Funding Liquidity and Arbitrage Spreads

2.5.1 Effects on Arbitrage Spreads

Having constructed the Funding Liquidity Stress Index in Section 2.3 and reproduced the daily arbitrage spreads of three classic arbitrage strategies in Section 2.4, we now study the effects of funding liquidity on arbitrage spreads. In Table 2.15, we perform linear regressions of each arbitrage strategy against the Funding Liquidity Stress Index (FLSI). We find that all three R^2 are high. FLSI explains 48%, 43% and 64% of the variability of CEF spreads, M&A spreads and Treasury spreads, respectively. Funding liquidity is therefore a major risk factor to various arbitrage strategies. Factors loadings are significant and negative, indicating that an increase in funding liquidity stress leads to a widening of arbitrage discounts (i.e., arbitrage spreads become more negative). In Table 2.16, we add an additional factor PC2 and repeat the three linear regressions. All factors loadings are significant and negative. As expected, the explanatory power increases; the three R^2 are now 60%, 52% and 74%. The FLSI index however explains the majority of the variability of the arbitrage spreads. FLSI is more important than PC2, but they are both significant.

Table 2.15: We perform linear regressions against the Funding Liquidity Stress Index (FLSI). We find that all three R2 are high. Funding liquidity is therefore a major risk factor to various arbitrage strategies.

	(1)	(2)	(3)
	CEF arbitrage	M&A arbitrage	Treasury arbitrage
(Intercept)	-5.794***	-1.518***	-0.183***
	(0.044)	(0.018)	(0.001)
FLSI	-2.228***	-0.802***	-0.084***
	(0.044)	(0.018)	(0.001)
R^2	0.477	0.426	0.645
$adj.R^2$	0.477	0.426	0.645
N	2792	2792	2792

Standard errors in parentheses

* (p ≤ 0.1), ** (p ≤ 0.05), *** (p ≤ 0.01)

Table 2.16: We add an additional factor PC2 and repeat the three linear regressions performed in Table 2.15. All factors loadings are significant and negative.

	(1)	(2)	(3)
	CEF arbitrage	M&A arbitrage	Treasury arbitrage
(Intercept)	-5.784***	-1.515***	-0.183***
	(0.039)	(0.016)	(0.001)
FLSI	-2.232***	-0.803***	-0.083***
	(0.039)	(0.016)	(0.001)
PC2	-1.142***	-0.38***	0.033***
	(0.039)	(0.016)	(0.001)
R^2	0.602	0.521	0.743
$adj.R^2$	0.602	0.521	0.743
N	2792	2792	2792

Standard errors in parentheses

* (p ≤ 0.1), ** (p ≤ 0.05), *** (p ≤ 0.01)

2.5.2 Contagion

Next we study whether contagion between arbitrage strategies is mainly attributed to their common exposure to funding liquidity risk factors. [BHN05] defined contagion as "correlation over and above what one would expect from economic fundamentals". The majority of studies have used monthly returns of hedge fund indices to study contagion. Hedge fund returns however are not proxies to fully-hedged arbitrage strategies[30] as many elect to hedge or unhedge certain risk factors on a discretionary basis to improve the returns. To synthetically obtain returns of fully-hedged arbitrage strategies (i.e., to remove the effects of common exposures to fundamentals), it is common to regress the returns of hedge funds on various asset classes and then study the residuals[31]. In this chapter, however, we don't need to synthetically create arbitrage strategies from hedge fund returns or remove the exposure to fundamentals since we reproduce real arbitrage strategies from daily transactional data.

To assess the impact of funding liquidity on contagion, we compute the contagion between the arbitrage spreads twice, before and after regressing against our two funding liquidity risk factors (i.e., the FLSI index and PC2). Essentially, we first check for contagion between the arbitrage spreads and then check for contagion between the residuals of the three regressions performed in Table 2.16. If contagion between the residuals is significantly lower than the contagion between the arbitrage spreads, we conclude that contagion is mainly due to their common exposure to funding liquidity risk factors.

We use three techniques to assess contagion: 1) computing a correlation matrix, 2) computing a "coupling coefficient", and 3) performing a logistic regression on the number of simultaneous threshold exceedances. First we compute the correlation matrix. The correlation matrix of the arbitrage spreads is shown in Table 2.17. Correlations are high, in particular the correlation between the M&A and CEF arbitrage spreads. The correlation matrix of the residuals is shown in Table 2.18. We observe that pair-wise correlations significantly dropped after accounting for the effects of FLSI and PC2 risk factors.

As an alternative to computing correlation coefficients between each pair and then, for instance, computing the average of all pairs, we compute the "coupling coefficient" [SA00]. The coupling coefficient measures the degree of co-movement between all the variables, and is often called the absorption ratio [KLPR10]. It equals the fraction of total variance explained by a particular eigenvector, in a PCA analysis. The eigenvector of interest is the vector with coefficients of same sign across all the variables (i.e., all the variables

[30] See [Aga04] and [FH04].
[31] See [DN10b] and [BSS10].

Table 2.17: The correlation matrix of the arbitrage spreads.

	CEF Spreads	M&A Spreads	Treasury Spreads
CEF Spreads	1.00		
M&A Spreads	0.63	1.00	
Treasury Spreads	0.47	0.47	1.00

Table 2.18: We compute the correlation matrix of the residuals from the regressions performed in Table 2.16. Compared to Table 2.17, we see a significant drop in pair-wise correlations.

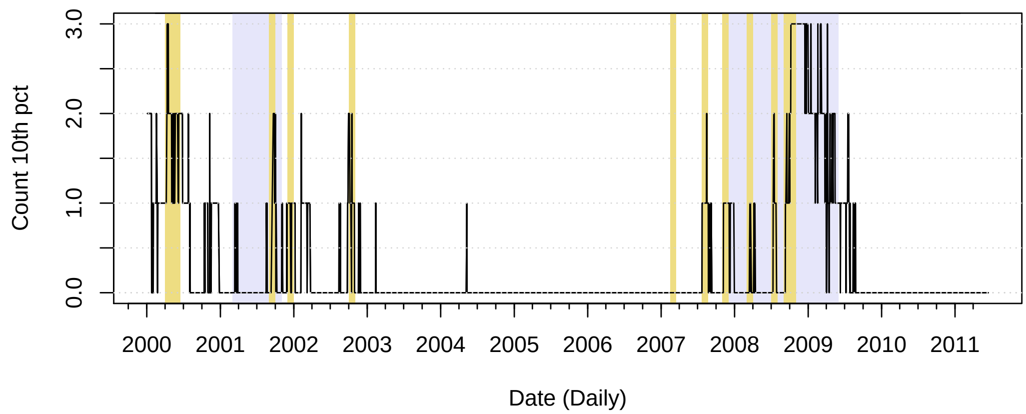
	CEF Residuals	M&A Residuals	Treasury Residuals
CEF Residuals	1.00		
M&A Residuals	0.08	1.00	
Treasury Residuals	0.13	0.17	1.00

moving in the same direction as a single block). Over the period ranging from 2000/01/04 to 2011/07/29, the coupling coefficient of the arbitrage spreads is 70.20%. After removing their common exposure to the funding liquidity risk factors, the coupling coefficient drops to 36.89%.

Last, we perform a logistic regression to test for contagion [BSS10] [BKS03]. This technique checks whether an arbitrage spread is more likely to widen beyond a specific quantile given that other arbitrage spreads widened beyond their corresponding quantiles. We will use the lowest 10^{th} percentile to determine exceedances. Our conclusion holds for other low percentiles as well. Define $I_{k,t}$ as an indicator variable taking the value of 1 if the spread of arbitrage strategy k is below its 10^{th} percentile at time t. In Figure 2.26, we plot the daily count of all three indicators. We note that during financial crises most arbitrage spreads exceeded their 10^{th} percentile at the same time. For instance, during the bankruptcy of Lehman Brothers and the dot-com crash, all three arbitrage spreads widened to levels within their lowest 10^{th} decile. $I_{k,t}$ is the dependent variable in the logistic regression. The independent variable of $I_{k,t}$ is the sum of all *other* indicator variables. In Table 2.19, we show the regression results. We find significant explanatory power in all three regressions. The R^2 ranges from 13.6% to 22.4%. In other words, the likelihood that an arbitrage spread is in its lowest 10^{th} decile increases when other spreads are in their lowest 10^{th} deciles as well. On the other hands, if we apply the logistic regression to the residuals (see Table 2.20), the explanatory power drops significantly. In that case, the R^2

ranges from 0.8% to 2.2%.

Figure 2.26: Counting the number of arbitrage strategies whose spreads exceeded the 10th percentile worst spread. We note that there is a significant number of simultaneous spreads widening beyond the 10th percentile, providing evidence of contagion.



In all three tests for contagion, we find significant contagion between the three arbitrage spreads. However, after removing their common exposure to the funding liquidity risk factors, evidence of contagion drops dramatically.

2.6 Further Evidence from Event-Studies

In this section we provide further evidence to support the theoretical studies on market micro-structure which attempt to explain the relationship between funding liquidity and market prices. This section is primarily motivated by the market micro-structure model developed by [BP09] which we extend in the next chapter. In the model, investors use margins to leverage their positions. Margin requirements are set by uninformed financiers to control counter-party risk. Financiers increase margin requirements if volatility increases, and decrease margin requirements otherwise. An increase in margin requirements however forces highly leveraged investors to reduce their positions to raise cash, exercising undue selling pressures that depress market prices. In addition to presenting empirical evidence supporting the theoretical model developed by [BP09], we also provide empirical evidence supporting various assumptions and predictions of other theoretical models. In particular, our study on the relationship between margins and volatility supports the "margin

Table 2.19: Logistic regression applied to the three arbitrage spreads.

	Spreads		
	$I_{CEF,t}$	$I_{MA,t}$	$I_{Gov,t}$
(Intercept)	0.056***	0.04***	0.053***
	(0.006)	(0.005)	(0.006)
otherCount	0.222***	0.299***	0.236***
	(0.011)	(0.011)	(0.011)
R^2	0.136	0.225	0.152
$adj.R^2$	0.136	0.224	0.152
N	2792	2792	2792

Standard errors in parentheses

* ($p \leq 0.1$), ** ($p \leq 0.05$), *** ($p \leq 0.01$)

Table 2.20: Logistic regression applied to the residuals.

	Residuals		
	$I_{CEF,t}$	$I_{MA,t}$	$I_{Gov,t}$
(Intercept)	0.088***	0.081***	0.08***
	(0.006)	(0.006)	(0.006)
otherCount	0.059***	0.094***	0.103***
	(0.012)	(0.013)	(0.013)
R^2	0.008	0.019	0.022
$adj.R^2$	0.008	0.019	0.022
N	2792	2792	2792

Standard errors in parentheses

* ($p \leq 0.1$), ** ($p \leq 0.05$), *** ($p \leq 0.01$)

setting mechanism" and "margin spiral" concepts in [BP09], as well as the assumption regarding the use of "backward-looking risk measures" by market participants in [BCG$^+$09]. Moreover, our study on the relationship between margins and market returns supports the concepts of "loss spirals" in [BP09], "performance-based arbitrage" in [SV97], and "pro-cyclicality" of financing in [Gea09].

In this section, we illustrate the following: 1) declining market prices coincide with increasing market volatility, 2) margins increase after an increase in market volatility[32], and 3) market prices continue to decline after margins increase. The converse is also true. We perform event studies on three different assets: 1) the S&P 500 index, 2) the Australian dollar / U.S. dollar foreign exchange rate, and 3) the New Zealand dollar / U.S. dollar foreign exchange rate. The event graphs (for instance Figure 2.29) have a vertical line at $T = 0$ corresponding to the date of a margin change, and plots the average path of a variable (typically, volatility or cumulative returns) during a window of time before and after the event. We study events of margin increase and decrease separately.

Futures contracts on each of those three assets trade on the Chicago Mercantile Exchange (CME). CME, like other exchanges, requires investors to post margins to cover a certain percentage of possible price move of a position to reduce counter-party risk. In Figure 2.27, we plot the dollar margin on a S&P 500 Index futures contract. The size of the contract is 250$\times$ S&P 500 Index futures price. From January 1998 to June 2011, CME changed the margin on the S&P 500 Index futures contract 16 times: increasing it 10 times and decreasing it 6 times. In Figure 2.28, we plot the S&P 500 Index implied volatility, proxied by the popular Chicago Board Options Exchange Market Volatility Index, widely known as the VIX index.

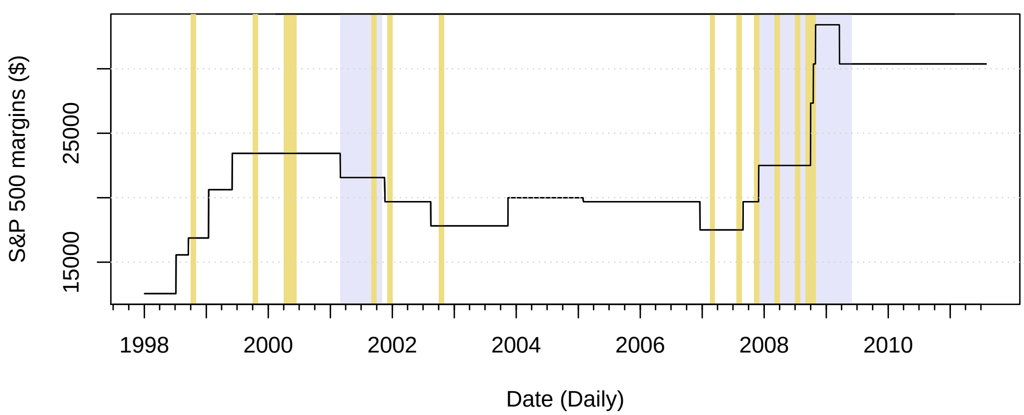
Next we perform two event studies. In Figure 2.29, we plot the average value of the implied volatility of the S&P 500 Index surrounding 10 margin increases. The vertical dashed red line, corresponds to $T = 0$, the date that a margin increase occurred. To the left are the days preceding the event while to the right are the days following the event. We identify a clear trend of increasing volatility prior to a margin increase. The volatility also stays high for almost a month after a margin increase. In Figure 2.30, we plot the average cumulative return of the S&P 500 Index in the period surrounding 10 margin increases. For each one of the 10 instances, the cumulative return starts from a base of 1

[32] This observation is expected since we mentioned earlier that the Chicago Mercantile Exchange (CME) sets margins so they cover 99 percent of potential price moves. It was not clear however whether the changes in margin requirements lead, lag or coincide with changes in market volatility. The event-studies that we perform show that the exchanges are often late in adjusting their margin requirements reflecting the use of backward looking risk measures. We also show that changes in margin requirements exacerbate (i.e., increases further) the volatility of market prices as predicted in [BP09].

at $T = -50$. We note that a margin increase occurs following an average drop of 6%, and, more interestingly, the index would fall another 4% on average after the margin increases. This observation is in accordance with the downward spiraling effect of funding and market liquidity, modelled by [BP09].

We also study decreasing margins events. In Figure 2.31 and Figure 2.32, we plot the average path of the volatility and the average path of the cumulative return, respectively, surrounding periods of decreasing margins. Not unexpectedly, the CME exchange reduced margins on future contracts during periods of decreasing volatility and increasing asset prices.

Figure 2.27: Margin requirement (in dollars) on the S&P 500 Index futures contract. It is the dollar amount required to initiate a position in the contract. Note that the size of the futures contract is 250× the S&P 500 Index futures price.



Next, we perform event studies to the AUD/USD foreign exchange in order to check for consistency across various asset classes. The margins on the AUD/USD foreign exchange futures contract increased 29 times and decreased 30 times during the period from January 2000 to June 2011. In Figure 2.33 and Figure 2.34, we plot the dollar margin requirement and the implied volatility of the AUD/USD foreign exchange, respectively. Comparing these two figures, we note that margins increased during periods of high volatility. In Figure 2.35, we plot the average path of the AUD/USD implied volatility surrounding 29 margin increases. We identify a clear trend of increasing volatility prior to a margin increase. The volatility also stays high for almost a month after a margin increase. In Figure 2.36, we show that the margin increases following a period of declining market prices. It is interesting to note that market prices continue to decrease, possibly due to forced selling following a margin increase. On the other hand, we show that the margin

Figure 2.28: We plot the Chicago Board Options Exchange Market Volatility Index, known as the VIX index. It is widely used as a proxy to the S&P 500 implied volatility.

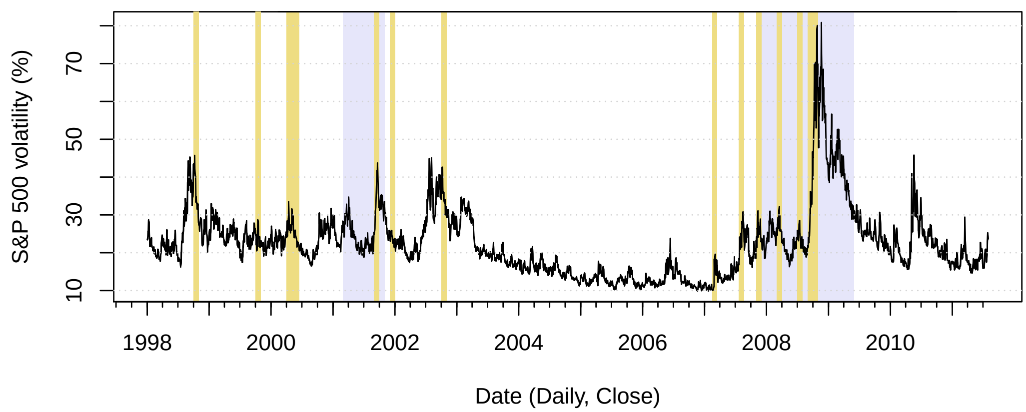


Figure 2.29: Event study of *increasing margins*. We plot the average value of the implied volatility of the S&P 500 Index surrounding 10 margin increases. The vertical dashed red line, corresponds to $T = 0$, the date that a margin increase occurred. To the left are the days preceding the event while to the right are the days following the event. We identify a clear trend of increasing volatility prior to a margin increase. The volatility also stays high for almost a month after a margin increase.

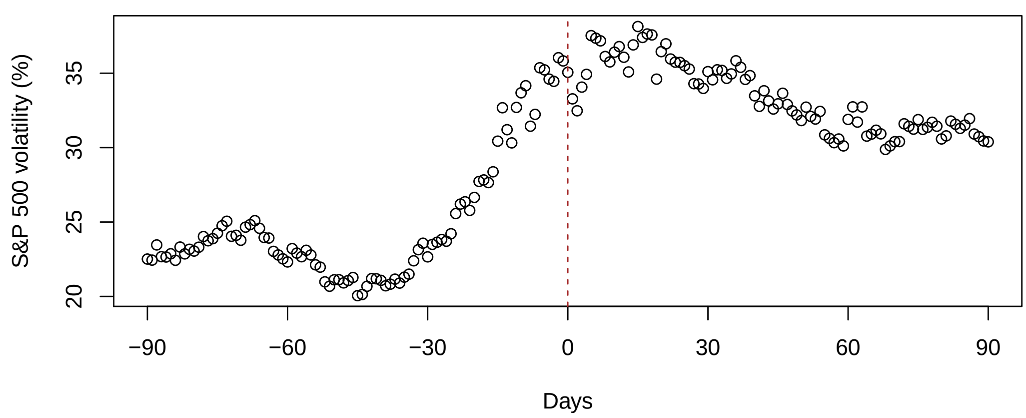


Figure 2.30: Event study of *increasing margins*. We plot the average cumulative return of the S&P 500 Index surrounding 10 margin increases. For each one of the 10 instances, the cumulative return starts from a base of 1 on day $T = -50$. We note that a margin increase occurs following an average drop of 6%, and, more interestingly, the index would fall another 4% on average after the margin increases. This observation is in accordance with the downward spiraling effect of funding and market liquidity, modelled by [BP09].

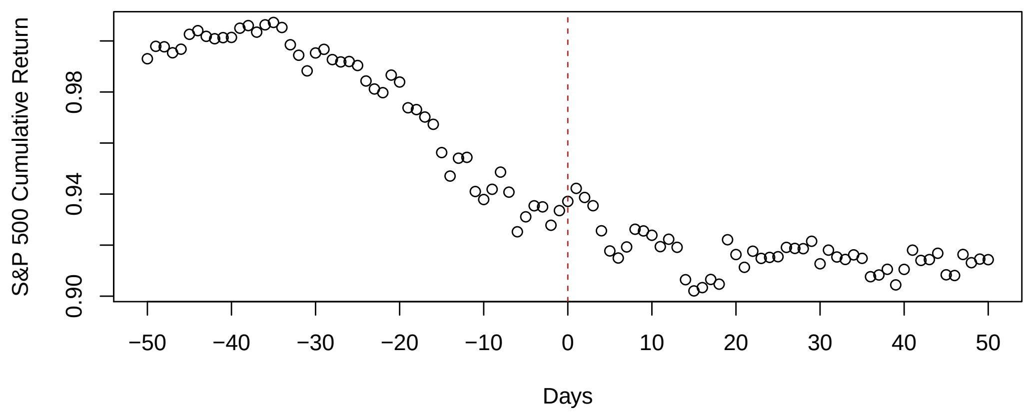


Figure 2.31: Event study of *decreasing margins*. In contrast to Figure 2.29, margins decrease following a period of decreasing volatility.

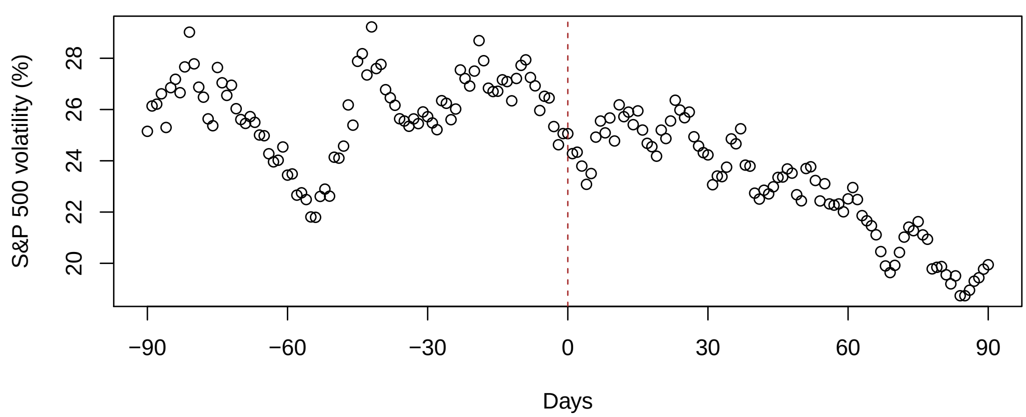
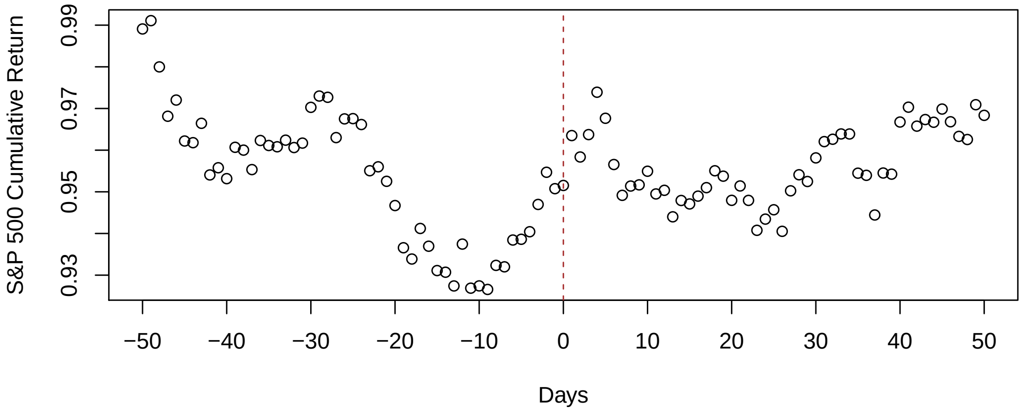
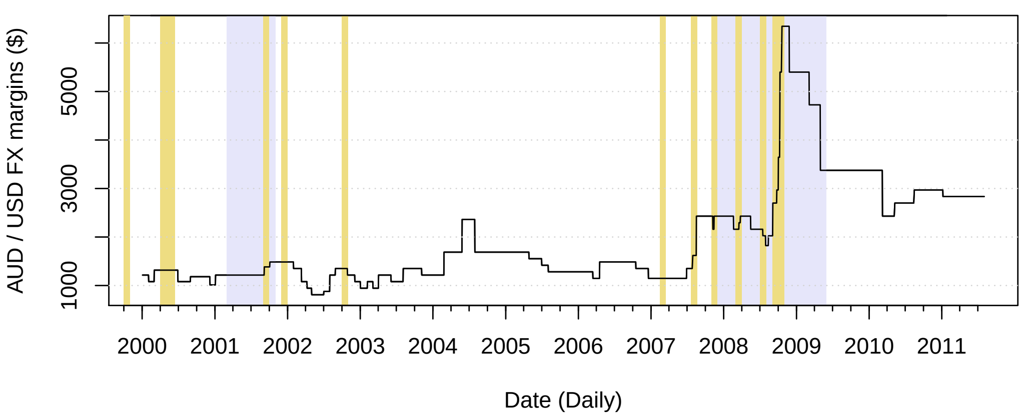


Figure 2.32: Event study of *decreasing margins*. In contrast to Figure 2.30, margins decrease following a period of rising asset prices.



decreases following periods of decreasing volatility and increasing market prices, as shown in Figure 2.37 and Figure 2.38, respectively. In Appendix C, we perform similar event-studies with respect to the NZD/USD foreign exchange. We obtain similar observations.

Figure 2.33: Margin requirement (in Dollar) on the AUD / USD FX futures contract.



The event-studies show strong consistency with respect to changes in margin requirements. In particular, margins increased after periods of rising volatility and decreasing market prices. More importantly, market prices continued to drop after an increase in margins providing evidence of spiraling effects between market and funding liquidity. These

Figure 2.34: We plot the implied volatility of the AUD/USD foreign exchange rate.

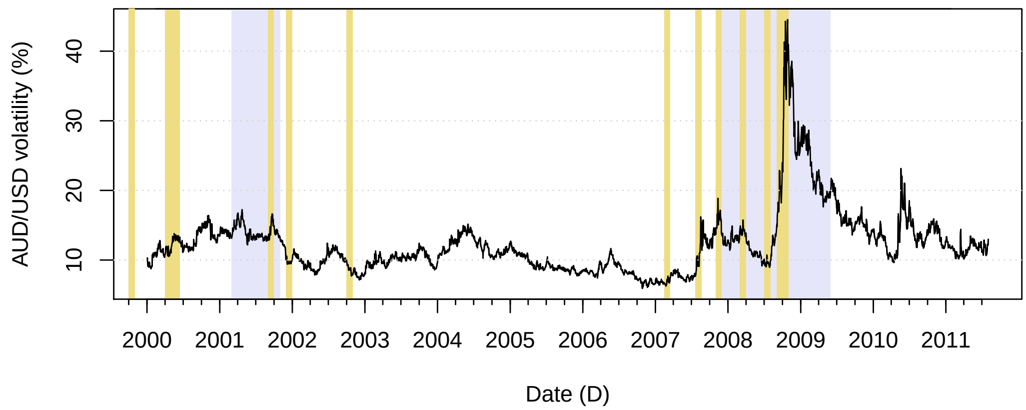


Figure 2.35: Event study of *increasing margins.* We plot the average path of the AUD/USD implied volatility surrounding 29 margin increases. We identify a clear trend of increasing volatility prior to a margin increase. The volatility also stays high for almost a month after a margin increase.

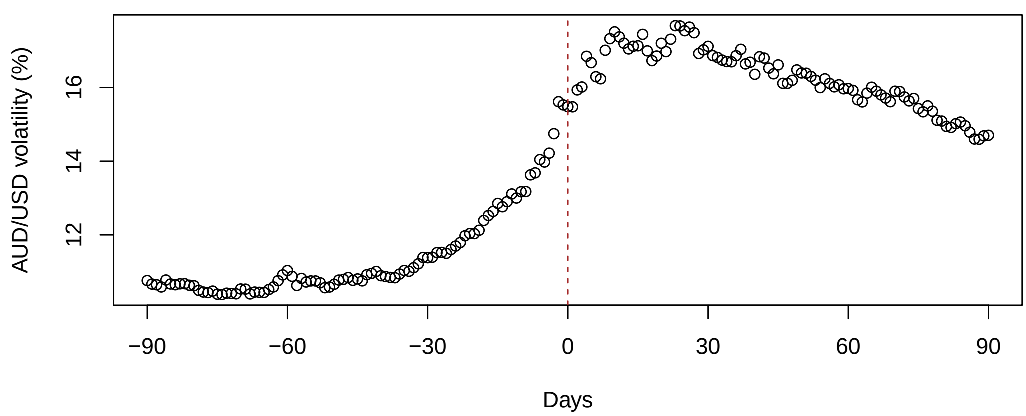


Figure 2.36: Event study of *increasing margins*. We plot the average cumulative return of the AUD/USD foreign exchange in the period surrounding 29 margin increases.

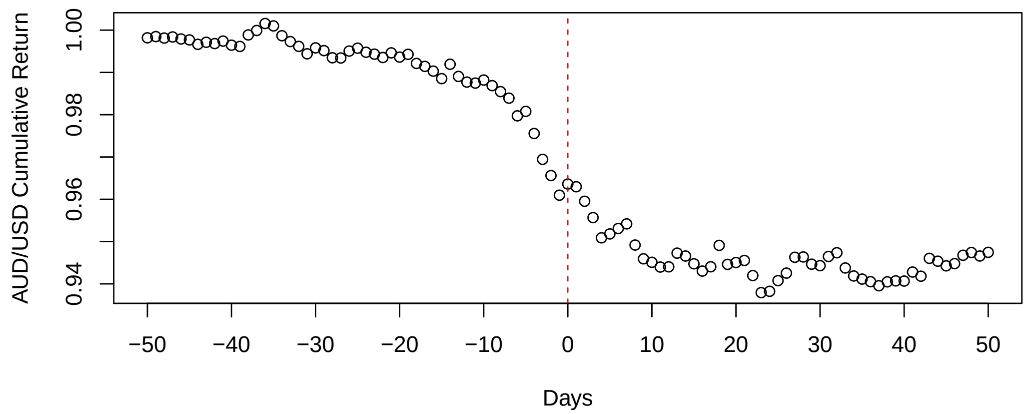


Figure 2.37: Event study of *decreasing margins*. We plot the average value of the AUD/USD implied volatility in the period surrounding 30 margin decreases.

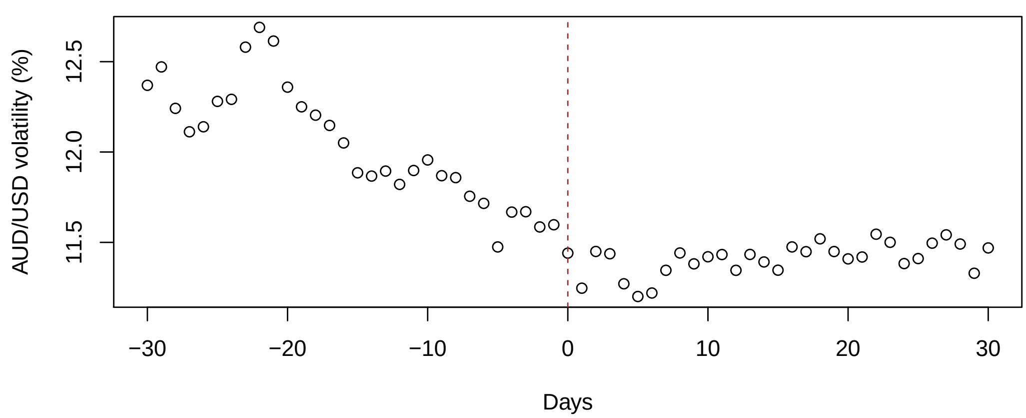
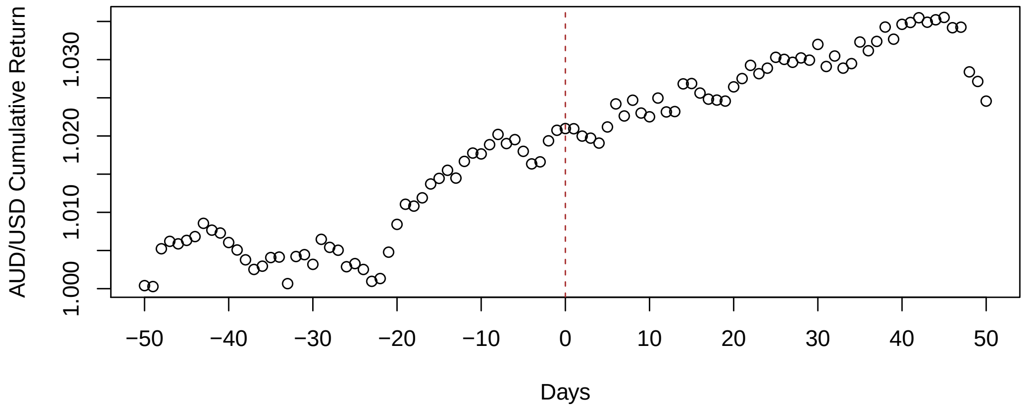


Figure 2.38: Event study of *decreasing margins*. We plot the average cumulative return of the AUD/USD foreign exchange in the period surrounding 30 margin decreases.



observations will become familiar in the next chapter when we talk about "margin setting mechanisms", "margin spirals", and "loss spirals" [BP09].

Chapter 3

Analytical Model

In the previous chapter we provided empirical evidence on the effects of funding liquidity on market prices. In this chapter, we study the mechanism through which funding liquidity affects market prices. Existing theoretical studies have focused on the spiraling effects of funding liquidity and market liquidity when financing constraints bind ("liquidity shock") in a single-period market micro-structure setting [SV97] [BP09]. In this chapter, we extend their model into a multi-period setting and show that arbitrageurs price-in a *liquidity risk premium* that keeps market prices below their fundamental values even when their financing constraints are not binding. The *liquidity risk premium* is further decomposed into 1) a *systemic risk premium* which compensates arbitrageurs for the risk of a "liquidity shock", and 2) a *fundamental risk premium* which compensates them for providing liquidity in a security whose fundamental value is volatile.

3.1 Literature Review

3.1.1 Related Work

The literature on market micro-structure evolved gradually towards an accurate depiction of the functioning of capital markets. In 1988, Grossman and Miller [GM88] proposed a market micro-structure model where market prices are determined by the supply and demand of securities. Their model served as a foundation for future work in the literature. The work by De Long *et al.* [DLSSW90] is perhaps the first formal attempt in the literature to assert that arbitrageurs have short term horizons and are therefore deterred from betting aggressively on mispriced assets. Shleifer and Vishny [SV97] were the first to argue that the aggregate wealth of arbitrageurs affects their ability to conduct arbitrage and keep markets liquid. They further emphasized the importance of the '*agency relationship*" in which external investors use past performance to control the flow of capital to arbitrageurs; a mechanism they call "*performance-based arbitrage*". As a result, arbitrageurs often decide to provide less liquidity by taking smaller positions than what they can afford. Liu and Longstaff [LL04] made similar conclusions but in a different setting. In their model, the price discount is stochastic and follows a Brownian bridge. Gromb and Vayanos [GV02] built on the market micro-structure model developed by [GM88] with the added feature that arbitrageurs are financially constrained, limiting their ability to take on large positions. While others studied market liquidity with margin-constrained arbitrageurs such as [GV02] and [LL04], [BP09] studied "*how market conditions lead to changes in the margin requirement itself, ... as happened in October 1987, and the resulting feedback effects between margins and market conditions as [arbitrageurs] are forced to de-lever*". In particular, [BP09] assumes that margins are set by *"uninformed financiers"* to control their value-at-risk. As a result, market liquidity is *fragile*, and vulnerable to "*loss spirals*" and "*margin spirals*" that push market prices significantly below their fundamental values, creating a *"liquidity shock"*. In what follows, we study in more detail the related works mentioned above.

[GM88] lay the groundwork for future work on market micro-structure. In their model, market prices are determined by the supply and demand of securities. The asynchronous arrival of "natural" investors is bridged by arbitrageurs that charge for providing liquidity. For instance, they charge by offering to buy, from a seller, the security at a discount from its fundamental value. The discount compensates the arbitrageurs for the risk of holding the security until the natural buyers arrive to the market. By transacting now instead of waiting for the natural buyers, the seller mitigates the risk of holding excess shares of the security by selling some of the shares at a price that is not uncertain, but lower on average than what he would expect when the buyers arrive. According to [GM88], the number of arbitrageurs "*will adjust until, in equilibrium, the returns to each from assuming the risk of waiting to trade with the ultimate buyers just balance the [fixed] costs of maintaining a continuous presence in the market. This adjustment determines the equilibrium amount of immediacy provided, i.e., the amount by which price is temporarily depressed by a typical sell order.*"

In [DLSSW90], arbitrageurs have short term horizons and are therefore deterred from betting aggressively on mispriced assets. They worry that "*irrational noise traders with erroneous stochastic beliefs will affect [market] prices*" causing arbitrageurs to exit their arbitrage positions at a loss. Their "*basic model is a stripped down overlapping gener-*

ation model with two-period lived agents [Sam58]". The infinitely extended overlapping generations structure assures that each agent's horizon is short.

In [GM88] and [DLSSW90], it is assumed that the market has "*a very large number of tiny arbitrageurs, each taking an infinitesimal position against the mispricing in a variety of markets. Because their positions are so small, capital constraints are not binding.*" Shleifer and Vishny [SV97] however present a more realistic treatment to arbitrage. "*More commonly, arbitrage is conducted by relatively few professional, highly specialized investors who combine their knowledge with resources of outside investors to take large positions. The fundamental feature of such arbitrage is that brains and resources are separated by an agency relationship. The money comes from investors with only a limited knowledge of individual markets, and is invested by arbitrageurs with highly specialized knowledge of these markets.*" The agency relationship is particularly important considering arbitrage requires capital and the number of specialized arbitrageurs is limited. [SV97] further assumes that outside investors determine the ability of arbitrageurs to invest profitably based on their past performance. As a result, "*the aggregate supply of funds to [arbitrageurs at time t] is an increasing function of arbitrageurs' gross return*" between time $t-1$ and time t. Shleifer and Vichny call this mechanism "*performance-based arbitrage (PBA)*".

The key innovation in [SV97], and later reiterated in different settings by Gromb and Vayanos [GV02], Liu and Longstaff [LL04], and Brunnermeier and Pedersen [BP09], is the negative correlation between capital gains on existing positions and future return opportunities. Profits and losses on existing positions affect arbitrageurs' wealth and therefore affect the ability of financially constrained arbitrageurs to conduct arbitrage. As a result, arbitrageurs often decide to provide less liquidity by taking smaller positions than what they can afford.

Liu and Longstaff [LL04] make similar conclusions to [SV97], using however a different setting. In their model, the equilibrium market price is not determined by the supply and demand for the security. Arbitrageurs are assumed to invest in an arbitrage opportunity with a discount converging to zero at some future point in time. The discount process is modelled as a Brownian-bridge. Since the discount can widen and arbitrageurs have wealth constraints, Liu and Longstaff [LL04] observe that arbitrageurs often do not invest up to their wealth constraints, and as a result, fail to push the discount closer to zero.

Gromb and Vayanos [GV02] build on the market micro-structure model developed by [GM88]. They consider a model in which arbitrageurs exploit discrepancies between the market prices of two identical assets trading in different markets. The existence of two identical assets allows arbitrageurs to hedge the risk of changes in fundamental values. Arbitrageurs use leverage to trade the assets and hold separate margin accounts for each asset (i.e., cross-margining is not possible). Furthermore, arbitrageurs are required to keep in each margin account enough capital to cover the maximum loss that can occur. [GV02] shows that *"if arbitrageurs' wealth is insufficient, they may be unable to eliminate price discrepancies between the risky assets."* [GV02] further provide a qualitative assessment of market prices in a two-period setting where the market is subject to endowment shocks, a form of exogenous shocks similar to the effects of noise traders in [DLSSW90]. In particular, they show that *"if the capital gain on the arbitrage opportunity until the next period is risky, arbitrageurs may choose not to invest up to the financial constraint."*

The model of Brunnermeier and Pedersen [BP09] "*is similar in spirit to [GM88] with the added feature that [arbitrageurs] face the real-world funding constraint.*" In particular, "*uninformed financiers*" increase margin requirements after periods of high volatility to control their value-at-risk. The concept of "*uninformed financiers*" in [BP09] is similar to "*outside investors*" in [SV97] as they both constrain arbitrageurs from taking large positions in attractive arbitrage opportunities after periods of stress. "*Outside investors*" in [SV97] however use past performance, instead of past volatility of market prices, to control the financing of arbitrage positions. Brunnermeier and Pedersen [BP09] demonstrate, in a single-period setting, two important properties of market liquidity: 1) market liquidity is *fragile* as small losses can lead to a large sudden drop in market liquidity, and 2) "*when markets are illiquid, market liquidity is highly sensitive to further changes in funding conditions*" due to two reinforcing liquidity spirals: a "*loss spiral*" and a "*margin spiral*".

The following chain of events illustrates those two liquidity spirals. A fall in market prices reduces the value of the assets relative to the money borrowed, leading to an increase in leverage (the start of the *loss spiral*). At the same time, an increase in volatility during periods of stress erodes the collateral value of the assets leading to an increase in margin requirements (the start of the *margin spiral*). That forces the arbitrageur to cut back on leverage by reducing his positions. That in turn creates a selling pressure on market prices, exacerbating the initial price decline and increasing the volatility further leading to a second round of loss and margin spirals. This vicious cycle causes market prices to deviate significantly from their fundamental values, creating a "*liquidity shock*".

In Table 3.1, we compare the various market micro-structure models discussed in this section. The comparison is based on six factors:

1. *Financing Constraints*: It indicates whether arbitrageurs are limited by financing constraints.

2. *Liquidity Shocks*: It indicates whether *liquidity shocks* are endogenous or exogenous.

3. *Loss Spirals*: It indicates whether loss spirals can occur.

Table 3.1: We compare the various market micro-structure models discussed in this section. The comparison is based on six factors.

Literature	Financing Constraints	Liquidity Shocks	Loss Spirals	Margin Spirals	Liquidity Analysis	Time Periods (* equilibrium)
Grossman and Miller (1988) [GM88]	X	none	X	X	Multi-Period	$t = 0, \dots, T$ $t^* = t$
De Long, et al. (1990) [DLSSW90]	X	exogenous	✓	X	Multi-Period	$t = 0, \dots, T$ $t^* = t$
Shleifer and Vishny (1997) [SV97]	✓	endogenous	X	✓	Single Period	$t = 1, 2, 3$ $t^* = 2$ $t^* = 1$(qualitative)
Gromb and Vayanos (2002) [GV02]	✓	none	X	X	Multi-Period	$t = 0, \dots, T$ $t^* = t$
Gromb and Vayanos (2002) [GV02]	✓	exogenous	X	X	Two Period	$t = 0, 1, 2$ $t^* = 0$(qualitative)
Gromb and Vayanos (2009) [GV09]	✓	endogenous	✓	X	Single Period	$t = 1, 2$ $t^* = 1$
Brunnermeier and Pedersen (2009) [SV97]	✓	endogenous	✓	✓	Single Period	$t = 0, 1, 2, 3$ 2 and 3 are similar $t^* = 1$ $t^* = 0$(qualitative)
Thesis (this Chapter)	✓	endogenous	✓	✓	Multi-Period	$t = 0, \dots, T$ $t^* = t$

4. *Margin Spirals*: It indicates whether margin spirals can occur.

5. *Liquidity Analysis*: It indicates whether it is a single-period or a multi-period analysis of market liquidity.

6. *Time Periods*: We keep the terminology of each paper for ease of reference. We place a star next to the time period where a market liquidity analysis was performed.

It is clear from Table 3.1 that Brunnermeier and Pedersen's model [BP09] is the most comprehensive. They however restrict their study of market liquidity to a single-period. In this chapter, we extend their model to multiple periods and compute, at equilibrium, the discount of market prices from their fundamental values assuming all market participants to be risk-averse. We further decompose the discount into two components: 1) a *fundamental risk premium* and 2) a *systemic risk premium*.

3.1.2 Market Micro-Structure Model

In this section we consolidate the models in [GM88], [GV02] and [BP09].

Economy

Consider an economy with two assets: a risk-free asset called cash, and a risky asset. The risk-free interest rate is normalized to zero. The risky asset trades at times $t = 0, 1, 2, \ldots, T$ and pays a single dividend v at maturity $t = T$

$$v = \bar{v} + \sum_{t=0}^{T} \delta_t \tag{3.1}$$

where $\bar{v}$ is a constant and $\delta_t \sim N(0, \sigma_t)$ is a zero-mean random variable[1] revealed at time t. The δ_t are normally distributed with an ARCH volatility structure:

$$\sigma_t = \overline{\sigma} + \theta \left|\delta_{t-1}\right| \tag{3.2}$$

$\theta \geq 0$ so that shocks δ_t to the fundamental value increase future volatility [BP09]. The fundamental value v_t of the risky asset at time t is the conditional expected value of its final payoff

$$v_t = \mathrm{E}\left[\, v \,\middle|\, v_{t-1}, \delta_t \right] = v_{t-1} + \delta_t \tag{3.3}$$

[1]Intuitively, the δ_t terms can be considered as adjustments announced at time t to the base dividend $\bar{v}$.

Define

$$\Lambda_t = \mathrm{E}\left[\, v \,\middle|\, v_{t-1}, \delta_t \right] - p_t = v_t - p_t \tag{3.4}$$

Λ_t is the discount of the market price p_t from its fundamental value v_t [GV02].

Investors

There are two groups of market participants: *investors* and *arbitrageurs*. They trade assets to maximize their expected utility functions [GM88]. For simplicity we assume that the aggregate effect of each group of *investors* is reduced to a single representative investor. At $t = 0$, the investor is informed that he will receive at $t = T$ "*an endowment of size u in the security, which [at the current price,] is inappropriate in the light of the [investor's] risk preference and information on the risk-return pattern associated with the [asset]*" [GM88][2]. For concreteness in exposition, we assume that u is positive. Since the investor's demand for the asset decreases as u increases, we refer to u as the "supply shock". The investor has initial wealth $W_{inv,0}$ invested in the risk-free asset. At times $t < T$, he chooses to hold y_t shares of the risky asset to maximize his expectation of his utility function $U_\gamma\left(W_{inv,T}\right)$ over final wealth $W_{inv,T}$. The investor is risk-averse[3] and has an exponential utility function

$$U_\gamma\left(W_{inv,s}\right) = \frac{1 - e^{-\gamma W_{inv,s}}}{\gamma} \tag{3.5}$$

where γ is the risk aversion parameter. The wealth of the investor evolves according to:

$$W_{inv,s+1} = W_{inv,s} + \left(y_s + u\right) \cdot \left(p_{s+1} - p_s\right), \qquad \forall s < T \tag{3.6}$$

Arbitrageurs

The investor's demand for immediacy is accommodated by *arbitrageurs* (such as market makers, proprietary traders and hedge funds) who provide market liquidity by acting as intermediaries. For instance, when there is a "supply shock", they offer to buy and hold the security until they find the natural buyer or until the security matures (i.e., pays off its last dividend) [GM88]. For simplicity, we also assume that the aggregate effect of arbitrageurs

[2]This specification of endowments is common across the literature on market micro-structure [GM88] [GV02] [BP09].

[3]Risk aversion of market participants, in particular the exponential utility function, is classic in the literature of market micro-structure [GM88] [DLSSW90] [SV97] [GV02] [LL04]. Brunnermeier and Pedersen [BP09] however assume risk-neutral arbitrageurs for simplicity.

is represented by a single arbitrageur. At times $t < T$, the arbitrageur chooses to hold x_t shares of the risky asset to maximize his expectation of his utility function over final wealth $W_{arb,T}$. The arbitrageur is risk-averse and has an exponential utility function:

$$U_\alpha\left(W_{arb,s}\right) = \frac{1 - e^{-\alpha W_{arb,s}}}{\alpha} \tag{3.7}$$

The arbitrageur is required at all time to have enough capital $W_{arb,s}$ to cover his total margin requirement $m_s\left|x_s\right|$:

$$m_s \cdot |x_s| \leq W_{arb,s} \qquad \forall s < T \tag{3.8}$$

where m_s is the dollar margin required to trade *one share* at time s. The number of shares x_s that the arbitrageur can trade is therefore limited by his wealth $W_{arb,s}$. The wealth of the arbitrageur evolves as follows:

$$W_{arb,s+1} = W_{arb,s} + x_s \cdot (p_{s+1} - p_s)\,, \qquad \forall s < T \tag{3.9}$$

Margin Setting Mechanism

The arbitrageur uses leverage in the form of margin trading. [BP09] proposes a margin setting mechanism where the "*financier sets margins to limit his counterparty credit risk*". In particular, the financier "*ensures that the [total] margin is large enough to cover the position's π-value-at-risk*". Let $\mathcal{G}_s$ be his information set at time s, the financier therefore sets the dollar margin per share, m_t, according to

$$\pi = Pr\left(-\Delta p_{s+1} \ > m_s \left|\mathcal{G}_s\right.\right) \qquad \text{if } x_s > 0 \tag{3.10}$$

and

$$\pi = Pr\left(\ \ \Delta p_{s+1} \ > m_s \left|\mathcal{G}_s\right.\right) \qquad \text{if } x_s < 0 \tag{3.11}$$

In [BP09], the financier is *uninformed* about the fundamental value v_s and the supply shock u, so his information set $\mathcal{G}_s$ is restricted to information on observed prices[4]

[4]Since"*arbitrage is conducted by relatively few professional, highly specialized investors*" [SV97], the financier in general is uninformed about the fundamental value of the security being traded. As a result, he assesses the risk of the security by observing its market price.

$\sigma\{p_0, \ldots, p_s\}$. As a result, he *"interpret[s] price volatility as fundamental volatility"* [BP09] and sets margins according to:

$$\begin{aligned} m_s &= \Phi^{-1}(1-\pi)\,\bar{\sigma} + \Phi^{-1}(1-\pi)\,\theta\,|\Delta p_s| \\ &= \Phi^{-1}(1-\pi)\,\bar{\sigma} + \Phi^{-1}(1-\pi)\,\theta\,|\delta_s - \Delta\Lambda_s| \end{aligned} \tag{3.12}$$

"*Intuitively, since liquidity risk tends to increase price volatility, and since uninformed financiers may interpret price volatility as fundamental volatility, this increases margins. [Equation* (3.12)*] corresponds closely to real-world margin setting, which is primarily based on volatility estimates from past price movements. ... [BP09] denotes the phenomenon that margins can increase as illiquidity rises by destabilizing margins.*"

Equilibrium

Similar to [BP09] and [GV02], we define the "*competitive equilibrium*" of the economy as follows:

Definition 3.1.1 *A competitive equilibrium consists of a price process $\{p_t\}$ such that for all t*

1. *the investor selects y_t to maximize the expectation of his utility function, Equation* (3.5)*, over final wealth subject to Equation* (3.6)*,*
2. *the arbitrageur selects x_t to maximize the expectation of his utility function, Equation* (3.7)*, over final wealth subject to Equations* (3.8) *and* (3.9)*,*
3. *the margin m_t is set according to Equation* (3.12)*, and*
4. *the market clears $x_t + y_t = 0$.*

Note that the conditions (1), (2) and (4) are classic conditions of market equilibrium (see [GM88]). Condition (3) reflects the arbitrageur's funding constraint similar to [BP09]. The equilibrium price p_t at time t is usually solved using dynamic programming, starting from time $t = T$ and working backwards. At time $t = T$, we have the base case $p_T = v_T$ as the asset pays its last dividend[5].

[5] i.e., the asset matures and pays off its fundamental value in full.

If, at any time t, there is more than one solution to the four conditions above, i.e., if the demand function and the supply function intersect at more than one point, we select the price that makes the market most liquid[6]. We therefore add the condition $p_t = \arg_{p \in \mathbb{R}^+} \min |v_t - p|$ such that the equilibrium price p_t 1) satisfies all four conditions in Definition 3.1.1, and 2) is the solution that is the closest to the fundamental value v_t.

3.1.3 Illustration of Equilibrium price at time $T-1$

[BP09] studied the equilibrium market price at time $t = T-1$ in great detail. We repeat some of the illustrations conducted in [BP09] to provide insights into the problem before tackling the general multi-period solution in the next section[7]. [BP09] however assumed that arbitrageurs are risk neutral rather than risk averse; so we will point out the differences. The parameters, used in the following three graphs to illustrate the analytical expressions, are:

$$\begin{aligned} \alpha &= 0.01 & \lambda &= 1.65 & \overline{\sigma} &= 4 \\ \pi &= 0.05 & u &= 15 & \theta &= 0.25 \\ v_{T-1} &= 100 & p_{T-2} &= 102 & W_{arb,T-2} &= 150 \\ \delta_{T-1} &= -6 & x_{T-2} &= 3 & & \end{aligned}$$

At time $T-1$, each market participant makes the decision of how many shares of the asset he or she would like to hold. The desired holding of the asset is a function of the price and is therefore called a “demand function”.

In [BP09], the arbitrageur is *risk-neutral* and his demand function at time $T-1$ is:

$$x_{T-1}(p) = \begin{cases} \frac{W_{arb,T-1}(p)}{m_{T-1}(p)} & \text{if } p < v_{T-1}, \\ \in \left[-\frac{W_{arb,T-1}(p)}{m_{T-1}(p)}, \frac{W_{arb,T-1}(p)}{m_{T-1}(p)}\right] & \text{if } p = v_{T-1}, \\ -\frac{W_{arb,T-1}(p)}{m_{T-1}(p)} & \text{if } p > v_{T-1}. \end{cases} \tag{3.13}$$

[6] We believe that it is the case in practice. To justify our approach, we assume that market participants would like to maximize the sharpe ratio of their investments. We know that at maturity the price will equal the final payoff (i.e., the fundamental value). Since an illiquid price today 1) will increase the volatility, and 2) won't affect the final return on the asset, market participants have the incentive to select the most liquid state since it increases their sharpe ratio.

[7] All the equations in this section are derived later when we solve the general multi-period problem. However, in order to build intuition into the problem and give credit to [BP09], we present here some of the analytical expressions for the period $t = T-1$.

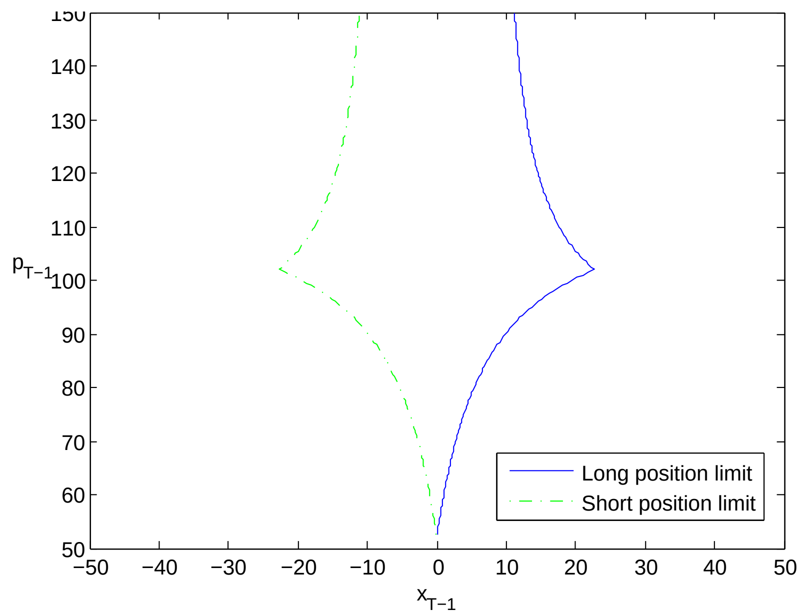


Figure 3.1: Position Limits at time $T-1$. We show the arbitrageur's position limits set according to Equation (3.8). The dotted curve in green shows the limit on a short position while the solid curve in blue shows the limit on a long position. *Uninformed financiers* use changes in the market price as a proxy to changes in the fundamental value to set their margin requirements. Note that $p_{T-2} = 102$. As p_{T-1} deviates from p_{T-2}, the margin m_{T-1} increases reflecting the perceived increase in volatility due to the ARCH effect; hence the shape of a "hyperbolic star" [BP09]. Moreover, note that the "hyperbolic star" is narrower when prices are lower compared to when prices are higher due to profits and losses on the existing position $x_{T-2} = 3$ (See Equation (3.15)).

where

$$m_{T-1}(p) = \Phi^{-1}(1-\pi)\left(\bar{\sigma} + \theta\,|p - p_{T-2}|\right) \tag{3.14}$$

$$W_{arb,T-1}(p) = W_{arb,T-2} + x_{T-2}\,(p - p_{T-2}) \tag{3.15}$$

The arbitrageur's demand function $x_{T-1}(p)$ is constrained by $\frac{W_{arb,T-1}(p)}{m_{T-1}(p)}$ on the long side and by $-\frac{W_{arb,T-1}(p)}{m_{T-1}(p)}$ on the short side. In Figure 3.1, we graph the arbitrageur's long and short position limits which take the shape of a "hyperbolic star"[8]. Those limits reflect his financing constraint given by Equation (3.8). When the price falls below the fundamental value (i.e., $p < v_{T-1}$), the expected profit form holding the asset to the next period becomes positive, and as a result, a *risk-neutral* arbitrageur tries to buy as many shares as possible of the asset. The opposite is true; when the price rises above v_{T-1} a *risk-neutral* arbitrageur tries to short as many shares as possible. In [BP09], $x_{T-1}(p)$ is therefore always on the boundary; that is if $p < v_{T-1}$ it hits the limit on the long position, and if $p > v_{T-1}$ it hits the limit on the short position.

Our model however assumes all market participants *risk-averse*. The demand function when the arbitrageur is *risk-averse* becomes:

$$x_{T-1}(p) = \begin{cases} \frac{W_{arb,T-1}(p)}{m_{T-1}(p)} & \text{if } p < v_{T-1} - \alpha\sigma_T^2 \frac{W_{arb,T-1}(p)}{m_{T-1}(p)}, \\ \frac{v_{T-1}-p}{\alpha\sigma_T^2} & \text{otherwise}, \\ -\frac{W_{arb,T-1}(p)}{m_{T-1}(p)} & \text{if } p > v_{T-1} + \alpha\sigma_T^2 \frac{W_{arb,T-1}(p)}{m_{T-1}(p)}. \end{cases} \tag{3.16}$$

Similar to the risk-neutral arbitrageur, $x_{T-1}(p)$ is bounded by $-\frac{W_{arb,T-1}(p)}{m_{t-1}(p)} \leq x_{T-1}(p) \leq \frac{W_{arb,T-1}(p)}{m_{t-1}(p)}$. However, between those two bounds, the arbitrageur's demand function is downward sloping $x_{T-1}(p) = \frac{v_{T-1}-p}{\alpha\sigma_T^2}$. It is a function of both the price and the volatility of the fundamental value. Since the arbitrageur is risk-averse, he needs the price to be low enough (i.e., the expected profit to be high enough) to compensate for the uncertainty of the final payoff. The red curve in Figure 3.2 corresponds to the demand function of the *risk-averse* arbitrageur. Note that the demand function is downward sloping inside the "hyperbolic star". Rather than hitting the long position limit whenever $p < v_{T-1}$, the risk averse arbitrageur slowly increases the size of his position as the price drops and only hits the position limit when $p < v_{T-1} - \alpha\sigma_T^2 \frac{W_{arb,T-1}(p)}{m_{T-1}(p)}$. If the arbitrageur was risk neutral (as is the case in [BP09]) the red line inside the "hyperbolic star" would be a horizontal line instead of a downward sloping line as is the case in Figure 3.2.

[8]The term "hyperbolic star" is borrowed from [BP09]

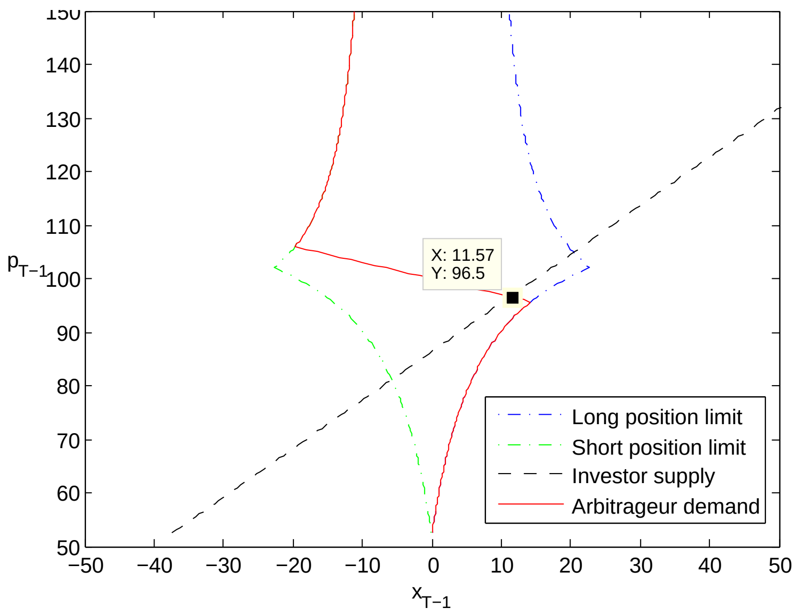


Figure 3.2: Equilibrium price p_{T-1} at time $T-1$. We graph the investor's supply and the arbitrageur's demand functions. The investor's supply function is the upward sloping black dashed curve while the arbitrageur's demand is the curve in red bounded by his position limits shown in Figure 3.1. When the price is higher than the fundamental value, the arbitrageur would like to short the asset and when the price is lower he would like to buy. The equilibrium price, $p_{T-1} = 96.5$, is at the intersection of the supply and demand functions. When $p = 96.5$, $x_{T-1}(p) = -y_{T-1}(p) = 11.57$.

[BP09] derived the analytical expression of the investor's demand function at time $t = T - 1$:

$$y_{T-1}(p) = \frac{v_{T-1} - p}{\gamma \sigma_T^2} - u \tag{3.17}$$

$-y_{T-1}(p)$ is the negative of the demand function and is called the investor's *supply function*. If the market is liquid (i.e., $p = v_{T-1}$), we get $-y_{T-1} = u$ which means that the investor sells all his excess holdings in the asset (i.e., the entirety of his endowment shock u). However, as the expected future profit of holding the asset increases (i.e., $p < v_{T-1}$) the investor sells less than u[9]. The upward sloping black dashed curve in Figure 3.2 corresponds to the investor's supply function.

The equilibrium price p_{T-1} is the value of p where the supply function $-y_{T-1}(p)$ equals the demand function $x_{T-1}(p)$ (i.e., the market clears). The intersection of the black dashed line and the red curve in Figure 3.2 corresponds to the equilibrium market price. If the functions $-y_{T-1}(p)$ and $x_{T-1}(p)$ intersect at two points, then p_{T-1} is the price that minimizes $|v_{T-1} - p|$, i.e., the price that makes the market most liquid by keeping the market price closer to its fundamental value.

Brunnermeier and Pedersen [BP09] further illustrate that market prices are *fragile*, in the sense that a small shock (for instance, a decrease in $W_{arb,T-2}$ or δ_{T-1}) can sometimes trigger a large drop in the equilibrium market price. To illustrate the fragility of market prices, we redraw Figure 3.2 but with $\delta_{T-1} = -18$ and $W_{arb,T-2} = 125$ instead of $\delta_{T-1} = -6$ and $W_{arb,T-2} = 150$. All other parameters are kept the same. The new supply and demand functions are graphed in Figure 3.3. We observe that the equilibrium market price drops significantly as the "hyperbolic star" shrinks. The drop in *market liquidity* is due to the reduced ability of the arbitrageur to purchase shares due to binding financing constraints. In summary, a "liquidity shock" occurs when binding financing constraints cause the equilibrium market price to settle on the "hyperbolic star"[10] leading to a large deviation of the market price from its fundamental value, as shown in Figure 3.3.

Moreover, Brunnermeier and Pedersen [BP09] compute the arbitrageur's demand function x_{T-2} at time $T - 2$. They show that, even if the arbitrageur's financing constraints

[9] To understand the behaviour of the investor, note that the expected profit per share is $\mathrm{E}_{T-1}[p_T - p] = \mathrm{E}_{T-1}[v_T - p] = v_{T-1} - p$. So if $p = v_{T-1}$ (i.e., the market is liquid), the expected profit is zero and, as a result, the risk-averse investor has no incentive to hold onto the risky asset, so he sells all the supply shock u. However, if the market price p drops below the fundamental value v_{T-1}, the expected profit increases and the investor is willing to hold onto some shares which compensates for the riskiness of the asset.

[10] i.e., when $x_{T-1}(p) = -\frac{W_{arb,T-1}(p)}{m_{T-1}(p)}$ or $x_{T-1}(p) = \frac{W_{arb,T-1}(p)}{m_{T-1}(p)}$

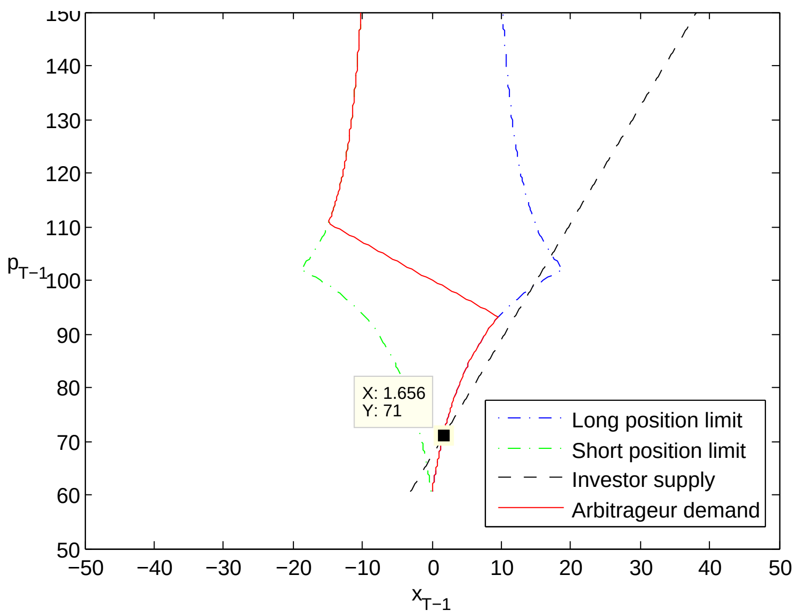


Figure 3.3: Illiquid equilibrium price p_{T-1} at time $T-1$. We redraw the plot shown in Figure 3.2 but with $\delta_{T-1} = -18$ and $W_{arb,T-2} = 125$ instead of $\delta_{T-1} = -6$ and $W_{arb,T-2} = 150$. All other parameters are kept the same. We observe that the market price at *equilibrium* drops significantly as the "hyperbolic star" shrinks.

are not binding, he won't purchase the asset unless it trades at a discount in order to be compensated for future "liquidity shocks". Brunnermeier and Pedersen [BP09] however don't compute the equilibrium price at time $T-2$. In this thesis, we continue their effort and compute the equilibrium market price (i.e., both the arbitrageur's demand and the investor's supply functions) at time $t \leq T-2$. We further decompose the *liquidity risk premium*, i.e., the discount of the market price from its fundamental value, at time t into its two components: the *fundamental risk premium* and the *systemic risk premium*.

3.2 Equilibrium Price p_t : Solution to a Dynamic Optimization Problem

In this section, we express the equilibrium price p_t as the solution to a "dynamic optimization problem". First, we formalize the problem of computing p_t and present a general formulation in a multi-period setting. Then, we derive the corresponding recursive formulation and discuss the numerical computational issues. This section is self-contained. Finance terms are not mentioned, except in Section 3.2.2, as we focus purely on the mathematics of the problem. To ease the transition from the economic model to its general mathematical formulation, we provide the economic intuition and certain derivations in Section 3.2.2.

3.2.1 General Problem Formulation

Objective Function

Let $T \in \mathbb{N}^+$ and $I_t = (\delta_t, v_t, w_{t-1}, p_{t-1}, x_{t-1}) \in \mathbb{R}^5$. Given a state vector I_t, our objective is to compute p_t at time $t \in \mathbb{N}_T$

$$\begin{aligned} p_t : \mathbb{R}^5 &\to \mathbb{R}^+ \\ p_t(I_t) &= \arg_{p \in \mathbb{R}^+} \min |v_t - p| \end{aligned} \tag{3.18a}$$

subject to

$$x_t(p, I_t) + y_t(p, I_t) = 0 \qquad \text{if } t < T \tag{3.18b}$$

Since the constraint in Equation (3.18b) applies only to $t < T$, we have a trivial base case at $t = T$

$$p_T(I_T) = v_T \tag{3.19}$$

Define

$$
\begin{aligned}
x_t : \left(\mathbb{R}^+, \mathbb{R}^5\right) &\to \mathbb{R} && (3.20a)\\
x_t(p, I_t) &= \arg_{x\in\mathbb{S}_{t,I_t}} \min \mathrm{E}\left[\prod_{s=t+1}^{T} k(p_s(I_s), I_s) \,\middle|\, I_t, p_t = p, x_t = x\right] \\
y_t : \left(\mathbb{R}^+, \mathbb{R}^5\right) &\to \mathbb{R} && (3.20b)\\
y_t(p, I_t) &= \arg_{y\in\mathbb{R}} \min \mathrm{E}\left[\prod_{s=t+1}^{T} h(p_s(I_s), I_s) \,\middle|\, I_t, p_t = p, x_t = -y\right]
\end{aligned}
$$

where

$$
\mathbb{S}_{s,I_s} = \{r \in \mathbb{R} : |r| \leq l(p_s(I_s), I_s)\} \quad \forall s < T \tag{3.20c}
$$

Cost functions $k(\cdot,\cdot)$ and $h(\cdot,\cdot)$ are given by Equations (3.23) and (3.24), respectively. The operator $\mathrm{E}[\,\cdot\,|\,\cdot]$ denotes the expectation over future states I_s given the current information set (I_t, p_t, x_t). The state vector evolves according to the p.d.f. $q(I_s|I_{s-1}, p_{s-1}, x_{s-1}) \quad \forall s < T$. For each I_s and $\forall s < T$, computing $p_s(I_s)$ in Equations (3.20a), (3.20b) and (3.20c) corresponds to solving at time s the optimization problem given by Equations (3.18a) and (3.18b). The domain[11] of x in Equation (3.20a) is limited to the set $\mathbb{S}_{t,I_t}$ defined by Equation (3.20c). The function $l(\cdot,\cdot)$ is given by Equation (3.25).

State Vector

The state vector I_s is composed of current values δ_s and v_s, and lagged values w_{s-1}, p_{s-1} and x_{s-1}:

$$
I_s = (\delta_s, v_s, w_{s-1}, p_{s-1}, x_{s-1}) \in \mathbb{R}^5 \tag{3.21}
$$

For all $\forall s < T$, the state vector evolves as follows. Given I_s, we first compute

1. $p_s = p_s(I_s)$, from Equation (3.18a)
2. $x_s = x_s(p_s(I_s), I_s)$, from Equation (3.20a)

Now given (I_s, p_s, x_s):

[11] i.e., the range of $x_t(p, I_t)$

1. generate $\delta_{s+1} \sim N\left(0, \overline{\sigma} + \theta \left|\delta_s\right|\right)$
2. set $v_{s+1} = v_s + \delta_{s+1}$
3. set $w_s = w_{s-1} + x_{s-1}\left(p_s - p_{s-1}\right)$

The evolution can be also expressed using the random function $I_{s+1} = M(I_s, p_s, x_s)$

$$M: \begin{pmatrix} \delta_s \\ v_s \\ w_{s-1} \\ p_{s-1} \\ x_{s-1} \end{pmatrix} \in \mathbb{R}^5, p_s \in \mathbb{R}^+, x_s \in \mathbb{R} \rightarrow \begin{pmatrix} \delta_{s+1} \\ v_s + \delta_{s+1} \\ w_{s-1} + x_{s-1}\left(p_s - p_{s-1}\right) \\ p_s \\ x_s \end{pmatrix} \in \mathbb{R}^5$$

Given (I_s, p_s, x_s), δ_{s+1} is the only random variable involved in the evolution from I_s to I_{s+1}. Its probability distribution is:

$$\delta_{s+1} \sim N\left(0, \overline{\sigma} + \theta \left|\delta_s\right|\right), \qquad \forall s < T \tag{3.22}$$

where $\overline{\sigma}$ and θ are constants.

Some Function Definitions

Below we define the functions used earlier

$$k(p, I_t) = e^{-\alpha x_{t-1}(p - p_{t-1})} \tag{3.23}$$

$$h(p, I_t) = e^{-\gamma(-x_{t-1}+u)(p - p_{t-1})} \tag{3.24}$$

$$l(p, I_t) = \frac{w_{t-1} + x_{t-1}\left(p - p_{t-1}\right)}{\lambda\left(\overline{\sigma} + \theta\left|p - p_{t-1}\right|\right)} \tag{3.25}$$

Note that u, α, γ, $\overline{\sigma}$, θ and λ are constants.

3.2.2 Economic Intuition and some Derivations

The purpose of this section is to explain the connection between the market micro-structure model discussed in Section 3.1.2 and its general mathematical formulation in Section 3.2.1.

Equation (3.18b) corresponds to the market clearing condition of Definition (3.1.1). The equilibrium price is the price p that satisfies $x_t(p, I_t) + y_t(p, I_t) = 0$. If Equation

(3.18b) has more than one solution, we select the one that satisfies Equation (3.18a). In other words, we select the price p that makes the market most liquid, i.e., the price that is closest to the fundamental value v_t.

Equation (3.19) is the base case which corresponds to the final payout. Before time T, the security does not pay any dividend. But at time T, the security pays out its fundamental value as a dividend, and hence $p_T = v_T$.

Equation (3.20b) corresponds to the utility maximization problem of the investor. In what follows, we derive Equation (3.20b) from the market micro-structure model. Recall that the investor's objective is to maximize the expected value of his exponential utility function over final wealth. That is:

$$y_t(p, I_t) = \arg_{y\in\mathbb{R}} \max \mathrm{E}\left[U_\gamma\left(W_{inv,T}\right) \middle| I_t, W_{inv,t-1}, p_t = p, y_t = y\right] \tag{3.26}$$

where

$$U_\gamma\left(W_{inv,T}\right) = \frac{1 - e^{-\gamma W_{inv,T}}}{\gamma} \tag{3.27}$$

$$W_{inv,s} = W_{inv,s-1} + \left(y_{s-1} + u\right) \cdot \left(p_s - p_{s-1}\right), \forall s \leq T \tag{3.28}$$

Equation (3.26) can be written as:

$$y_t(p, I_t) = \arg_{y\in\mathbb{R}} \max \mathrm{E}\left[\frac{1 - e^{-\gamma W_{inv,T}}}{\gamma} \middle| I_t, W_{inv,t-1}, p_t = p, y_t = y\right] \tag{3.29}$$

$$= \arg_{y\in\mathbb{R}} \max \mathrm{E}\left[\frac{1 - e^{-\gamma W_{inv,t-1} - \gamma \sum_{s=t}^{T}(y_{s-1}+u)\cdot(p_s - p_{s-1})}}{\gamma} \middle| I_t, W_{inv,t-1}, p_t = p, y_t = y\right] \tag{3.30}$$

Since $x_s = -y_s \ \forall s < T$, Equation (3.30) can be written as:

$$y_t(p, I_t) = \arg_{y\in\mathbb{R}} \max \mathrm{E}\left[\frac{1 - e^{-\gamma W_{inv,t-1} - \gamma \sum_{s=t}^{T}(-x_{s-1}+u)\cdot(p_s - p_{s-1})}}{\gamma} \middle| I_t, W_{inv,t-1}, p_t = p, x_t = -y\right] \tag{3.31}$$

Further, since $\{x_{t-1}, p_{t-1}\} \subset I_t$, $p_t = p$ and $W_{inv,t-1}$ is known at time t, the equation above is equivalent to

$$y_t(p, I_t) = \arg_{y\in\mathbb{R}} \max -c \cdot \mathrm{E}\left[e^{-\gamma \sum_{s=t+1}^{T}(-x_{s-1}+u)\cdot(p_s - p_{s-1})} \middle| I_t, p_t = p, x_t = -y\right] \tag{3.32}$$

$$= \arg_{y\in\mathbb{R}} \min \mathrm{E}\left[e^{-\gamma \sum_{s=t+1}^{T}(-x_{s-1}+u)\cdot(p_s - p_{s-1})} \middle| I_t, p_t = p, x_t = -y\right] \tag{3.33}$$

where c is a constant and equals to $\frac{e^{-\gamma W_{inv,t-1}-\gamma(-x_{t-1}+u)\cdot(p-p_{t-1})}}{\gamma}$.
Define

$$h(p, I_t) = e^{-\gamma(-x_{t-1}+u)(p-p_{t-1})} \tag{3.34}$$

The utility maximization problem of the investor, $y_t(p, I_t)$, becomes:

$$y_t(p, I_t) \;=\; \arg_{y\in\mathbb{R}} \min \mathrm{E}\left[\prod_{s=t+1}^{T} h(p_s, I_s)\,\middle|\, I_t, p_t = p, x_t = -y\right] \tag{3.35}$$

Note that all p_s in Equation (3.35) are optimized equilibrium market prices that satisfy the Equations (3.18a) and (3.18b). For clarity of exposition, we replace p_s by $p_s(I_s)$ and Equation (3.35) becomes:

$$y_t(p, I_t) \;=\; \arg_{y\in\mathbb{R}} \min \mathrm{E}\left[\prod_{s=t+1}^{T} h(p_s(I_s), I_s)\,\middle|\, I_t, p_t = p, x_t = -y\right] \tag{3.36}$$

which is identical to Equation (3.20b).

Having derived Equation (3.20b), we now derive Equation (3.20a) which is the arbitrageur's demand function $x_t(p, I_t)$. The derivation of $x_t(p, I_t)$ is similar to $y_t(p, I_t)$ except that the range of $x_t(p, I_t)$ is constrained. The constraint on the number of shares is due to the financing constraint, given by Equation (3.8). The arbitrageur should have enough capital to cover his total margin requirement:

$$m_t \cdot |x_t(p, I_t)| \leq w_t \tag{3.37}$$

where

$$w_t = w_{t-1} + x_{t-1}\,(p - p_{t-1}) \tag{3.38}$$

$$m_t = \lambda\,(\overline{\sigma} + \theta\,|p - p_{t-1}|) \tag{3.39}$$

m_t is the dollar margin required to trade *one share* at time t and w_t is the wealth of the arbitrageur at time t. Substituting the variables m_t and w_t by Equations (3.38) and (3.39), respectively, Equation (3.37) becomes:

$$|x_t(p, I_t)| \leq \frac{w_{t-1} + x_{t-1}\,(p - p_{t-1})}{\lambda\,(\overline{\sigma} + \theta\,|p - p_{t-1}|)} \tag{3.40}$$

now set

$$l(p, I_t) = \frac{w_{t-1} + x_{t-1}\,(p - p_{t-1})}{\lambda\,(\overline{\sigma} + \theta\,|p - p_{t-1}|)} \tag{3.41}$$

The constraint on the arbitrageur's position, $x_t(p, I_t)$, can be written as $x_t \in \mathbb{S}_{t,I_t}$ where

$$\mathbb{S}_{t,I_t} = \{r \in \mathbb{R} : |r| \leq l(p_t(I_t), I_t)\} \tag{3.42}$$

which is Equation (3.20c). Proceeding similarly to Equations (3.26)-(3.36) and adding the constraints on the range of $x_t(p, I_t)$, we get:

$$x_t(p, I_t) \;=\; \arg_{x\in\mathbb{S}_{t,I_t}} \min \mathrm{E}\left[\prod_{s=t+1}^{T} k(p_s(I_s), I_s)\,\middle|\, I_t, p_t = p, x_t = x\right] \tag{3.43}$$

where

$$k(p, I_t) = e^{-\alpha x_{t-1}(p - p_{t-1})} \tag{3.44}$$

which is identical to Equation (3.20a).

Finally, Equation (3.22) corresponds to the ARCH volatility structure of the fundamental value according to Equation (3.2), and the constant λ corresponds to $\Phi^{-1}(1-\pi)$ where $\pi = 95\%$.

3.2.3 A Dynamic Optimization Problem

To solve $p_t(I_t)$, we need to solve Equations (3.20a) and (3.20b). The evaluation of each of these two equations requires solving $p_s(I_s)$ for all $t + 1 \leq s < T$ and for infinitely many state vectors I_s. The problem is however more tractable if we express it recursively and use dynamic programming to solve it.

Recursion

Define

$$C_{t,k}(I, p, x) \;=\; \begin{cases} \mathrm{E}\left[\prod_{s=t+1}^{T} k(p_s(I_s), I_s)\,\middle|\, I_t = I, p_t = p, x_t = x\right] & \text{if } t < T \\ 1 & \text{if } t = T \end{cases} \tag{3.45}$$

Equation (3.45) can be written recursively as follows:

$$
\begin{aligned}
&C_{t,k}(I,p,x) \\
&= \mathrm{E}\left[k(p_{t+1},I_{t+1}) \times \prod_{s=t+2}^{T} k(p_s,I_s) \,\middle|\, I_t = I, p_t = p, x_t = x\right] \\
&= \mathrm{E}\left[k(p_{t+1},I_{t+1}) \times \mathrm{E}\left[\prod_{s=t+2}^{T} k(p_s,I_s) \,\middle|\, I_{t+1}, p_{t+1}, x_{t+1}\right] \,\middle|\, I_t = I, p_t = p, x_t = x\right] \\
&= \mathrm{E}\left[k(p_{t+1},I_{t+1}) \times C_{t+1,k}(I_{t+1},p_{t+1},x_{t+1}) \,\middle|\, I_t = I, p_t = p, x_t = x\right] \qquad (3.46)
\end{aligned}
$$

with base case $C_{T,k}(I,p,x) = 1$.

Note that all the indexed variables inside the expectation term are optimized. In particular, $p_j = p_j(I_j) \;\; \forall j < T$ as given by Equation (3.18a) and $x_j = x_j(p, I_j) \;\; \forall j < T$ as given by Equation (3.20a). We can rewrite Equation (3.46) to show explicitly the optimization of the variable x_{t+1} :

$$C_{t,k}(I,p,x) = \mathrm{E}\left[k(p_{t+1},I_{t+1}) \times \min_{x'} C_{t+1,k}(I_{t+1},p_{t+1},x') \,\middle|\, I_t = I, p_t = p, x_t = x\right] \quad (3.47)$$

Now, define

$$C_{t,h}(I,p,y) \;\; = \;\; \begin{cases} \mathrm{E}\left[\prod_{s=t+1}^{T} h(p_s(I_s),I_s) \,\middle|\, I_t = I, p_t = p, x_t = -y\right] & \text{if } \; t < T \\ 1 & \text{if } \; t = T \end{cases} \quad (3.48)$$

Similar to $C_{t,k}(I,p,x)$, $C_{t,h}(I,p,y)$ can be expressed recursively as follows

$$C_{t,h}(I,p,y) = \mathrm{E}\left[h(p_{t+1},I_{t+1}) \times \min_{y'} C_{t+1,h}(I_{t+1},p_{t+1},y') \,\middle|\, I_t = I, p_t = p, x_t = -y\right] \quad (3.49)$$

with base case $C_{T,h}(I,p,y) = 1$.

Putting Things Together

Using the definitions and recursive formulations of $C_{t,k}$ and $C_{t,h}$, our general problem formulation becomes:

$$p_t(I_t) \;\; = \arg_{p \in \mathbb{R}^+} \min |v_t - p| \quad (3.50a)$$

subject to

$$x_t(p, I_t) + y_t(p, I_t) = 0 \qquad \text{if } t < T \tag{3.50b}$$

where

$$\begin{aligned} x_t(p, I_t) &= \arg_{x \in \mathbb{S}_{t,I}} \min C_{t,k}(I, p, x) && (3.51a) \\ y_t(p, I_t) &= \arg_{y \in \mathbb{R}} \min C_{t,h}(I, p, y) && (3.51b) \end{aligned}$$

and

$$C_{t,k}(I, p, x) = \mathrm{E}\left[k(p_{t+1}, I_{t+1}) \times \min_{x'} C_{t+1,k}(I_{t+1}, p_{t+1}, x') \,\middle|\, I_t = I, p_t = p, x_t = x \right] \tag{3.51c}$$

$$C_{t,h}(I, p, y) = \mathrm{E}\left[h(p_{t+1}, I_{t+1}) \times \min_{y'} C_{t+1,h}(I_{t+1}, p_{t+1}, y') \,\middle|\, I_t = I, p_t = p, x_t = -y \right] \tag{3.51d}$$

with a base case at $t = T$

$$\begin{aligned} C_{T,k}(I, p, x) &= 1 && \text{[base case]} \\ C_{T,h}(I, p, y) &= 1 && \text{[base case]} \end{aligned}$$

Compared to our initial problem formulation in Section 3.2.1 , it is relatively easy here to compute $p_t(I_t)$ given the values of the two functions $C_{t,k}(I, p, x)$ and $C_{s,h}(I, p, y)$ evaluated on the entire space $(I, p, x) \in \mathbb{R}^7$ and $(I, p, y) \in \mathbb{R}^7$, respectively.

Computing $C_{t,k}$ and $C_{t,h}$

$C_{s,k}(I, p, x)$ and $C_{s,h}(I, p, y)$ are computed recursively starting at time $s = T$ and going back in time. The base case, at time $s = T$, is:

$$C_{T,k}(I, p, x) = 1 \qquad \text{[base case]} \tag{3.52}$$

$$C_{T,h}(I, p, x) = 1 \qquad \text{[base case]} \tag{3.53}$$

Now, given the functions $C_{s,k}(I,p,x)$ and $C_{s,h}(I,p,y)$ at a time s, $C_{s-1,k}(I,p,x)$ and $C_{s-1,h}(I,p,y)$ are computed as follows:

$$\begin{aligned} &C_{s-1,k}(I,p,x) \\ &\qquad = \mathrm{E}\left[k(p_s(I_s),I_s) \times \min_{x'} C_{s,k}(I_s,p_s(I_s),x') \,\middle|\, I_{s-1}=I, p_{s-1}=p, x_{s-1}=x\right] \\ &C_{s-1,h}(I,p,y) \\ &\qquad = \mathrm{E}\left[h(p_s(I_s),I_s) \times \min_{y'} C_{s,h}(I_s,p_s(I_s),y') \,\middle|\, I_{s-1}=I, p_{s-1}=p, x_{s-1}=-y\right] \end{aligned} \tag{3.54}$$

where

$p_s(I_s) = v_s \quad$ for $s = T$

$p_s(I_s)$ is computed according to Equations (3.50a) and (3.50b) for $s < T$

$h(\cdot,\cdot)$ and $k(\cdot,\cdot)$ functions are given by Equations (3.24) and (3.23)

Generating I_s from $(I_{s-1}, p_{s-1}, x_{s-1})$ is demonstrated in Section 3.2.1

The expectation term $\mathrm{E}[\cdot\,|\,\cdot]$ consists of an integral over the random variable δ_s which enters the function M that maps $(I_{s-1}, p_{s-1}, x_{s-1})$ into I_s. Refer to Section 3.2.1 for more detail on the evolution of the state variable I_s.

3.2.4 Numerical Computational Issues

According to the literature [EW89] [Pak91] [Rus91] [AM07], our optimization problem is an instance of "finite dynamic optimization problems" but with a more complex structure than the problems commonly studied in the literature.

The expectation term $\mathrm{E}[\cdot\,|\,\cdot]$ in $C_{t,k}(I,p,x)$ and $C_{t,h}(I,p,y)$ consists of an integral over the random variable δ_{t+1} which enters the function M that maps $(I_t,p_t,x_t) = (I,p,x)$ into I_{t+1}. Since the integral does not have a closed-form solution, we must use numerical integration. So for each δ_{t+1} drawn from its density function, we generate I_{t+1} and then compute the integrand. That in turn requires computing $p_{t+1}(I_{t+1})$ according to Equations (3.18a) and (3.18b). Here again, there is no analytical solution and we need to solve $x_{t+1}(p^{(i)}_{t+1}, I_{t+1})$ and $y_{t+1}(p^{(i)}_{t+1}, I_{t+1})$, Equations (3.51a) and (3.51a) respectively, at $t+1$ for many values of the pair $\left(p^{(i)}_{t+1}, I_{t+1}\right)$. These two minimization problems, in turn, need to be solved numerically and require computing $C_{t+1,k}(I_{t+1}, p^{(i)}_{t+1}, x^{(j)}_{t+1})$ and $C_{t+1,h}(I_{t+1}, p^{(i)}_{t+1}, y^{(j)}_{t+1})$ for many $x^{(j)}_{t+1}$ and $y^{(j)}_{t+1}$, respectively.

Since the state vector $I_s = \{\delta_s, v_s, w_{s-1}, p_{s-1}, x_{s-1}\} \in \mathbb{R}^5$, price $p_s \in \mathbb{R}$, and the decision variables $(x_s, y_s) \in \mathbb{R}^2$ are continuous, finding the exact solution to our problem is impossible as it would require evaluating $C_{s,k}(I,p,x)$ and $C_{s,h}(I,p,y)$ for an infinite number of combinations of $(I,p,x) \in \mathbb{R}^7$ and $(I,p,y) \in \mathbb{R}^7$, respectively, and for all $t < s \leq T-1$.

Related Work

Three aspects affect the computational burden required to solve such problems [GK01]:

1. The number of (state, price, decision) combinations determines the number of integrals that must be evaluated at each time period. Hence, if the number of state and decision variables is large, the number of integral evaluations may be very large. This is referred to as "curse of dimensionality", a term coined by Richard E. Bellman.

2. The number of stochastic terms in the cost functions (i.e., $k(p,I)$ and $h(p,I)$) and in the law of motion M determines the dimensionality of the integration required to compute the expectation term. Hence, as the number of stochastic terms increases, the computational time required to evaluate each integral increases.

3. The time required to evaluate the integrand affects directly the time required to evaluate each integral and therefore limits the number of random variables realizations that can be practically used to estimate each integral.

Since a base case exists at time $t = T$, "backsolving" through dynamic programming can save a lot of computations while providing a good approximation [SVA98]. Keane and Wolpine (1994) [KW94] propose a technique to numerically solve *finite dynamic optimization problems*. Adapting their technique to our problem, we proceed as follows. For each time period $s \in \{t, t+1, ..., T-1\}$, we compute (and store) $C_{s,k}(I,p,x)$ and $C_{s,h}(I,p,-x)$ at only a subset of the continuous state and decision variables $(I,p,x) \in \mathbb{R}^7$. When we need to evaluate $C_{s,k}(I',p',x')$ at a later time, we retrieve its value if the function was previously evaluated at (I',p',x'), otherwise we interpolate using a regression function. The application of this technique is non-trivial due to the complex evaluation of the integrand in our case.

Regarding the numerical estimation of the expectation term $\mathrm{E}[\,\cdot\,|\,\cdot\,]$ few options are available to obtain good approximations efficiently. Keane and Wolpine (1994) [KW94] suggest using crude frequency simulators. As [GK01] pointed out, crude frequency simulators are highly inefficient if the integrand takes large values with low probabilities (which is the case in our problem). [Rus97] proposes numerically integrating using a "*low discrepancy*" sequence instead of generating a sequence drawn from a random number generator. Importance sampling [GI89] can also be used to increase the efficiency of a Monte Carlo integration. Finally, since the integrals are one-dimensional, we could use numerical quadrature.

Note that the concept of reducing a T-period optimization problem like Equation (3.20a) to a two-period optimization problem like Equations (3.51a) and (3.51c) is known as the *Bellman's principle of optimality* [Bel57].

3.3 Analytical Solution of $p_t(I_t)$ at $t = T - 1$

In this section, we derive the *analytical* solution to $p_{T-1}(I_{T-1})$. Although the *numerical* approximate solution presented in the next section applies to $t = T - 1$ as well, we will use the *analytical* solution whenever we need to compute $p_{T-1}(I_{T-1})$ to improve the accuracy of our numerical results.

$p_{T-1}(I_{T-1})$ can be computed analytically because the functions $x_{T-1}(p, I)$ and $y_{T-1}(p, I)$ have analytical solutions. That is only possible at $t = T - 1$. For ease of exposition, we use numerical values for the constant parameters and for the state variables in I_{T-1} to simplify the derivation of the analytical solution and gain insights into the problem. In appendix D, we provide the analytical expressions without numerical values.

We use the following numerical values for the constant parameters:

$$\alpha = 0.005 \qquad \lambda = 1.65 \qquad \overline{\sigma} = 4$$
$$\gamma = 0.02 \qquad u = 10 \qquad \theta = 0.20$$

and the following numerical values for the state variables in I_{T-1}:

$$v_{T-1} = 100 \qquad p_{T-2} = 102 \qquad w_{T-2} = 150$$
$$\delta_{T-1} = -6 \qquad x_{T-2} = 3$$

Objective

Given I_{T-1}, compute

$$p_{T-1}(I_{T-1}) = \arg_{p \in \mathbb{R}^+} \min |v_{T-1} - p| \tag{3.55}$$

subject to

$$x_{T-1}(p, I_{T-1}) + y_{T-1}(p, I_{T-1}) = 0 \tag{3.56}$$

Solving $y_{T-1}(p, I_{T-1})$

$$y_{T-1}(p, I_{T-1}) \;=\; \arg_{y\in\mathbb{R}} \;\min C_{T-1,h}(I_{T-1}, p, y) \tag{3.57}$$

where

$$\begin{aligned} & C_{T-1,h}(I_{T-1}, p, y) && (3.58)\\ & \quad = \mathrm{E}\left[h(p_T, I_T) \times \min_{y'} C_{T,h}(I_T, p_T, y') \,\middle|\, I_{T-1}, p_{T-1} = p, x_{T-1} = -y\right] \\ C_{T,h}(I, p, y) &= 1 \qquad\qquad [\text{base case}] && (3.59)\\ h(p, I_T) &= e^{-\gamma(-x_{T-1}+u)(p-p_{T-1})} && (3.60)\\ \delta_T &\sim N\left(0, \overline{\sigma} + \theta\left|\delta_{T-1}\right|\right) && (3.61)\end{aligned}$$

By setting $C_{T,h}(I, p, x)$ to 1, we get

$$C_{T-1,h}(I_{T-1}, p, y) = \mathrm{E}\left[e^{-\gamma(-x_{T-1}+u)(p_T-p_{T-1})} \,\middle|\, I_{T-1}, p_{T-1} = p, x_{T-1} = -y\right] \tag{3.62}$$

Now setting $p_T = v_T = v_{T-1} + \delta_T$ (base case) and inserting the numerical values, Equation (3.62) becomes

$$C_{T-1,h}(I_{T-1}, p, y) = \mathrm{E}\left[e^{-0.02(y+10)(100+\delta_T-p)}\right] \;, \quad \delta_T \sim N\left(0, 5.2\right) \tag{3.63}$$

Equation (3.57) can be written as

$$y_{T-1}(p, I_{T-1}) \;=\; \arg_{y\in\mathbb{R}} \;\min \mathrm{E}\left[e^{-0.02(y+10)(100+\delta_T-p)}\right] \tag{3.64}$$

Equation (3.64) is easy to solve as the only random variable is δ_T which is normally distributed $N\left(0, 5.2\right)$. We know that

$$\mathrm{E}\left[e^r\right] = e^{\mathrm{E}(r)+\frac{1}{2}\mathrm{Var}(r)} \qquad \text{if } r \text{ is normally distributed} \tag{3.65}$$

Equation (3.64) can therefore be written as

$$y_{T-1}(p, I_{T-1}) \;=\; \arg_{y\in\mathbb{R}} \;\min e^{\mathrm{E}(r)+\frac{1}{2}\mathrm{Var}(r)} \tag{3.66}$$

where

$$\begin{aligned} r &= -0.02\left(y+10\right)\left(100+\delta_T-p\right) && (3.67)\\ \mathrm{E}\left[r\right] &= -0.02\left(y+10\right)\left(100-p\right) && (3.68)\\ \mathrm{Var}\left[r\right] &= 0.02^2\left(y+10\right)^2\sigma_T^2 && (3.69)\\ &= 0.02^2\left(y+10\right)^2 5.2^2 && (3.70)\end{aligned}$$

Since the exponential function is monotonically increasing, Equation (3.66) is equivalent to

$$y_{T-1}(p, I_{T-1}) \;=\; \arg_{y\in\mathbb{R}}\; \min\left[\mathrm{E}\,(r) + \frac{1}{2}\mathrm{Var}\,(r)\right] \tag{3.71}$$

To find $y_{T-1}(p, I_{T-1})$, we take the derivatives of $\mathrm{E}\,(r) + \frac{1}{2}\mathrm{Var}\,(r)$ with respect to y and set it to 0:

$$-0.02\,(100-p) + 0.02^2\,(y+10)\,5.2^2 = 0 \tag{3.72}$$

The solution is:

$$\begin{aligned} y_{T-1}(p, I_{T-1}) &= \frac{100-p}{0.02 \times 5.2^2} - 10 \\ &= \frac{v_{T-1}-p}{\gamma\sigma_T^2} - u \end{aligned} \tag{3.73}$$

Solving $x_{T-1}(p, I_{T-1})$

We proceed similarly to the previous section.

$$x_{T-1}(p, I_{T-1}) \;=\; \arg_{x\in\mathbb{S}_{T-1,I_{T-1}}}\; \min C_{T-1,k}(I_{T-1}, p, x) \tag{3.74}$$

where

$$C_{T-1,k}(I_{T-1}, p, x) \tag{3.75}$$
$$= \mathrm{E}\left[k(p_T, I_T) \times \min_{x'} C_{T,k}(I_T, p_T, x')\,\middle|\, I_{T-1}, p_{T-1} = p, x_{T-1} = x\right]$$
$$C_{T,k}(I, p, x) = 1 \qquad \text{[base case]} \tag{3.76}$$
$$k(p, I_T) = e^{-\alpha x_{T-1}(p-p_{T-1})} \tag{3.77}$$
$$\delta_T \sim N\left(0, \overline{\sigma} + \theta\,|\delta_{T-1}|\right) \tag{3.78}$$
$$\mathbb{S}_{T-1,I_{T-1}} = \{r \in \mathbb{R} : |r| \le l(p_{T-1}, I_{T-1})\} \tag{3.79}$$
$$l(p, I_{T-1}) = \frac{w_{T-2} + x_{T-2}\,(p - p_{T-2})}{\lambda\,(\overline{\sigma} + \theta\,|p - p_{T-2}|)} \tag{3.80}$$

Replacing the state variables by their numerical values in Equation (3.80), we get:

$$l(p, I_{T-1}) = \frac{150 + 3\,(p-102)}{1.65\,(4 + 0.20\,|p-102|)} \tag{3.81}$$

Proceeding similarly to Equations (3.62) – (3.73) and adding the constraint on x given by Equations (3.79) and (3.81), we get:

$$x_{T-1}(p, I_{T-1}) = \begin{cases} \frac{150+3(p-102)}{1.65(4+0.20|p-102|)} & \text{if } p < 97.69, \\ \frac{100-p}{0.005\times 5.2^2} & \text{otherwise,} \\ -\frac{150+3(p-102)}{1.65(4+0.20|p-102|)} & \text{if } p > 102.98. \end{cases} \tag{3.82}$$

Computing $p_{T-1}(I_{T-1})$

Next, we solve Equations (3.55) and (3.56) using Equations (3.73) and (3.82). The solution is:

$$p_{T-1}(I_{T-1}) = 98.92$$

In Figure 3.4, we plot $x_{T-1}(p, I_{T-1})$ and $-y_{T-1}(p, I_{T-1})$ as a function of p according to Equations (3.73) and (3.82), respectively. The dashed line in black corresponds to $-y_{T-1}$ while the solid curve in red corresponds to x_{T-1}. Their intersection is $p_{T-1}(I_{T-1}) = 98.92$. At $p_{T-1}(I_{T-1}) = 98.92$, $x_{T-1}(p, I_{T-1}) = -y_{T-1}(p, I_{T-1}) = 8.00$. The dotted curve in blue and the dotted curve in green correspond to the upper bound and the lower bound constraints, respectively, on the range of $x_{T-1}(p, I_{T-1})$ as given the Equations (3.79) and (3.81).

3.4 Numerical Estimation of $p_t(I_t)$

An analytical solution to compute $p_t(I_t)$ for $t \leq T-2$ does not exist. In this section, we will solve Equations (3.18a) – (3.20c) numerically and provide a numerical estimate for p_{T-2}, p_{T-3} and p_{T-4}. In Section 3.2.3, we formulated our optimization problem recursively. We discussed the numerical computational issues in Section 3.2.4 and suggested "*backsolving*" using dynamic programming.

Our numerical technique is based on [KW94] and can be summarized as follows. For each time period $s \in \{t, t+1, ..., T-1\}$, we compute and store $C_{s,k}(I, p, x)$ and $C_{s,h}(I, p, -x)$ at only a subset of the continuous state and decision variables $(I, p, x) \in \mathbb{R}^7$. Denote this subset by $\mathbb{R}^7_d \subset \mathbb{R}^7$. When we need to evaluate $C_{s,k}(I', p', x')$ at a later time, we retrieve its value if the function was previously evaluated at (I', p', x'), i.e., if $(I', p', x') \in \mathbb{R}^7_d$, otherwise we interpolate using a regression function. The expectation term $\mathrm{E}[\,\cdot\,|\,\cdot\,]$ in $C_{s,k}(I, p, x)$ and $C_{s,h}(I, p, -x)$ is computed using numerical integration. In particular, we

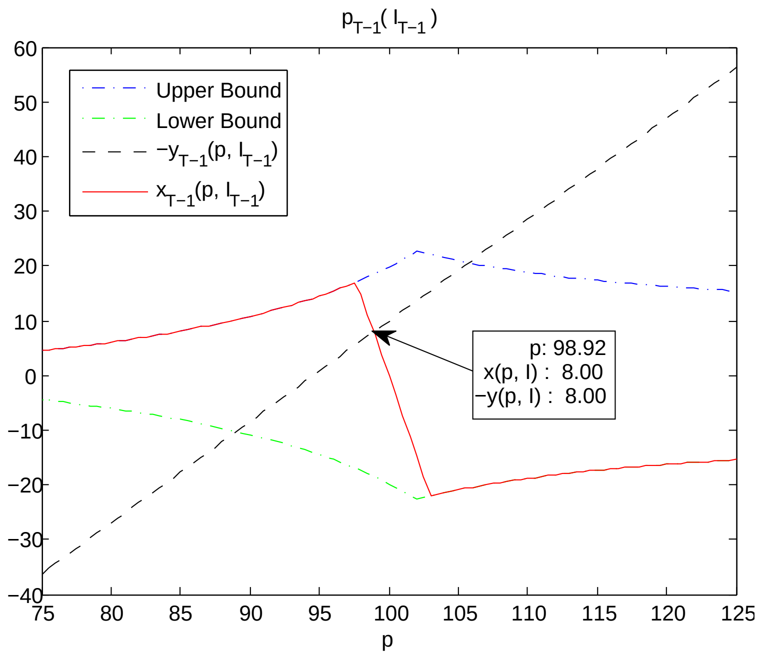


Figure 3.4: $p_{T-1}(I_{T-1})$ is the intersection of $x_{T-1}(p, I_{T-1})$ and $-y_{T-1}(p, I_{T-1})$. The dashed line in black corresponds to $-y_{T-1}$ while the solid curve in red corresponds to x_{T-1}. Their intersection is $p_{T-1}(I_{T-1}) = 98.92$. At $p_{T-1}(I_{T-1}) = 98.92$, $x_{T-1}(p, I_{T-1}) = -y_{T-1}(p, I_{T-1}) = 8.00$. The dotted curve in blue and the dotted curve in green correspond to the upper bound and the lower bound constraints, respectively, on the range of $x_{T-1}(p, I_{T-1})$ as given by Equations (3.79) and (3.81). Note that the axes have changed from previous plots; we flipped the x-axis and the y-axis to illustrate that the demand functions (y-axis) are functions of the price (x-axis).

use quasi-Monte Carlo with Importance Sampling. We compared the accuracy of that technique against the Gauss-Hermite quadrature. Although they were both satisfactory, we obtained more accurate results using the quasi-Monte Carlo method with Importance Sampling due to its ability to sample from the tails and around the point of discontinuity.

3.4.1 Dimension Reduction

The number of state and decision variables affects the number of integrals that must be evaluated and stored, and as a result affects the computational burden involved in solving our dynamic optimization problem. For instance, if we discretise each dimension into d values, the total number of function evaluations that we need to compute and store is $2 \times d^n$ where n is the dimension of (I, p, x). This is referred to as the "curse of dimensionality". It is therefore very important to try to reduce n.

Specific to our problem, we can rewrite the state and decision variables in a more parsimonious way and therefore save on the total number of computations required to obtain an estimate of $p_t(I_t)$. Let

$$p_t(I_t) = v_t - \Lambda_t(I_t^*) \tag{3.83}$$

and

$$I_t^* = (\delta_t, w_{t-1}, \Lambda_{t-1}, x_{t-1}) \in \mathbb{R}^4 \tag{3.84}$$

$I_t^* \in \mathbb{R}^4$ can be derived from $I_t \in \mathbb{R}^5$ by setting $\Lambda_{t-1} = p_{t-1} - v_{t-1}$. According to Equation (3.83), we can compute $p_t(I_t)$ at time t if we have $\Lambda_t(I_t^*)$. In Appendix E, we rewrite our dynamic optimization problem in such a way that $(I^*, \Lambda, x) \in \mathbb{R}^6$ are the only variables that enter the conditioning set in the expectation terms and are the only relevant variables affecting the evaluations of $C_{s,k}$ and $C_{s,h}$ functions. That way, we reduce the number of dimensions n from 7 to 6; we drop the two variables p_s and v_s and add the variable Λ_s. Below we give a quick snapshot on how to compute $\Lambda_s(I_s^*)$. Refer to Appendix E for a complete exposition.

Given $I_s^* \in \mathbb{R}^4$, $\Lambda_s(I_s^*)$ is computed as follows:

$$\Lambda_s(I_s^*) \ = \arg_{\Lambda \in \mathbb{R}} \min |\Lambda| \tag{3.85a}$$

subject to

$$x_s(\Lambda, I_s^*) + y_s(\Lambda, I_s^*) = 0 \qquad \text{if } t \leq s < T \tag{3.85b}$$

where

$$
\begin{aligned}
x_s(\Lambda, I_s^*) &= \arg_{x\in\mathbb{S}_s} \min C_{s,k}(I_s^*, \Lambda, x) && (3.86a)\\
y_s(\Lambda, I_s^*) &= \arg_{y\in\mathbb{R}} \min C_{s,h}(I_s^*, \Lambda, y) && (3.86b)
\end{aligned}
$$

The state vector evolves according to the following:

$$
M: \begin{pmatrix} \delta_s \\ w_{s-1} \\ \Lambda_{s-1} \\ x_{s-1} \end{pmatrix} \in \mathbb{R}^4, \Lambda_s \in \mathbb{R}, x_s \in \mathbb{R} \rightarrow \begin{pmatrix} \delta_{s+1} \\ w_{s-1} + x_{s-1}\left(\delta_s + \Lambda_{s-1} - \Lambda_s\right) \\ \Lambda_s \\ x_s \end{pmatrix} \in \mathbb{R}^4 \tag{3.87}
$$

3.4.2 Numerical Mapping to a Lower Dimension

Similar to [KW94], we compute and store the values of $C_{s,k}(I_s^*, \Lambda, x)$ and $C_{s,h}(I_s^*, \Lambda, -x)$ evaluated on a grid representing only a subset of $(I_s^*, \Lambda, x) \in \mathbb{R}^6$. This step is performed at each time period $s \in \{t, t+1, ..., T-1\}$. For ease of exposition, we will expand the vector $I_s^* = (\delta_s, w_{s-1}, \Lambda_{s-1}, x_{s-1}) \in \mathbb{R}^4$ in our treatment below. The dimension of the grid, as it stands right now, is 6 consisting of the following vector $(I_s^*, \Lambda, x) = (\delta_s, w_{s-1}, \Lambda_{s-1}, x_{s-1}, \Lambda, x)$ which forms the conditioning set of the expectation term in $C_{s,k}$ and $C_{s,h}$.

Not all possible combinations of $(I_s^*, \Lambda, x) \in \mathbb{R}^6$ need to be considered when designing the grid. For instance, $C_{s,k}(\delta_s, w_{s-1} = 150, \Lambda_{s-1} = 5, x_{s-1} = 10, \Lambda, x)$ and $C_{s,k}(\delta_s, w_{s-1} = 140, \Lambda_{s-1} = 6, x_{s-1} = 10, \Lambda, x)$ are identical. The vector $(w_{s-1}, x_{s-1}, \Lambda_{s-1}) \subset (I_s^*, \Lambda, x) \in \mathbb{R}^6$ is used just to compute w_s in the random function M (see Equation (E.8)). In particular, the vector appears in the expression $w_s = w_{s-1} + x_{s-1}\left(\delta_s + \Lambda_{s-1} - \Lambda\right)$ and, as a result, the value of w_s is identical for many values of $(w_{s-1}, x_{s-1}, \Lambda_{s-1})$, given δ_s and Λ.

We therefore only need to evaluate the functions $C_{s,k}$ and $C_{s,h}$ for the various combinations of $\mathrm{U_s} = (w_s, \delta_s, \Lambda, x) \in \mathbb{R}^4$. For instance, if we need to evaluate the function $C_{s,k}$ at $(I_s^{*\prime}, \Lambda', x') \in \mathbb{R}^6$, we first compute the intermediate vector $\mathrm{U_s}' = (w_s', \delta_s', \Lambda', x') \in \mathbb{R}^4$ and then retrieve the value of $C_{s,k}$ corresponding to the vector $\mathrm{U_s}'$. The mapping from I_s^* to $\mathrm{U_s}$ effectively reduced the dimension of our numerical problem further by two. The numerical solution needs to deal with just four dimensions $(w_s, \delta_s, \Lambda, x) \in \mathbb{R}^4$, down from seven initially.

3.4.3 Grid Discretization

In the previous two sections, we successfully reduced the dimension of the parameter vector of the functions $C_{s,k}$ and $C_{s,h}$ from $n = 7$ to $n = 4$. Since the variables are continuous, the parameter vector must be discretized and the functions $C_{s,k}$ and $C_{s,h}$ evaluated at a subset of the infinitely many combinations of $(w_s, \delta_s, \Lambda, x) \in \mathbb{R}^4$. Discretizing each dimension consists of selecting 1) the range to focus on, and 2) the number of observations in that range. We select the range of each dimension by inspecting the problem, using economic intuition and performing simple trials. For instance, if the value of the constant u is positive (i.e., there is selling pressure), then Λ must be positive (i.e., the price trades below its fundamental value). We therefore narrow down the discretization range of Λ to $\mathbb{R}^+$ instead of $\mathbb{R}$. Furthermore, since the price can not be negative, then Λ can not be larger than v_s. For all those reasons, we discretize the variable Λ over the range $[0, 100]$. We discretize w_s, the wealth of the arbitrageur, in the range $[0, 350]$. His wealth restricts the size of his position x_s in the security. We performed simulations and found that if his wealth exceeds 325, there are effectively no constraints on the position size relative to the amounts he would like to hold. As a result, the values of the functions $C_{s,k}$ and $C_{s,h}$ when $w_s \in [325, +\infty]$ are practically identical to the values when $w_s = 325$. It is therefore sufficient to discretize the variable w_s in the range $[0, 350]$. To select the range of x, we use economic reasoning. Since the constant u is positive (i.e., investors are forced to sell) the market clears only if the arbitrageurs take the other side of the trade and buy the asset. Hence $x > 0$. Furthermore, investors won't sell more than u shares (the value of the constant u is 10), hence $x \in [0, 10]$. To discretise δ_s, we use a low-discrepancy sequence with importance sampling, as explained in the next section.

We determine the number of discretizations, d, within each range, by assessing the sensitivity of the solution to the variable being discretized. For instance, if the solution is continuous and almost linear with respect to a particular variable, we can use a coarse discretization along that dimension. On the other hand, if the solution is discontinuous and non-linear with respect to a particular variable, we increase the number of observation in the range.

3.4.4 Numerical Integration

As mentioned earlier, the expectation term $\mathrm{E}[\,\cdot\,|\,\cdot]$ in $C_{s,k}(I_s^*, \Lambda, x)$ and $C_{s,h}(I_s^*, \Lambda, x)$ consists of an integral over the random variable δ_{s+1} which enters the function M that maps (I_s^*, Λ_s, x_s) into I_{s+1}^*. Since the integral does not have a closed-form solution, we compute it numerically. So for each δ_{s+1} drawn from its density function, we generate I_{s+1}^* and evaluate the integrand. To select an appropriate sampling technique for δ_{s+1}, we should take into account the following. First, evaluating the integrand is computationally intensive as it requires solving $\Lambda_{s+1}(I^*_{s+1})$ each time we generate a new δ_{s+1}. Second, certain realizations of δ_{s+1} that occur with low probability affects the integrand significantly more than others. Third, the integrand has one point of large discontinuity with respect to δ_{s+1} (a characteristic of market liquidity called *liquidity fragility* [BP09]). The point at which the discontinuity occurs is different for different values of the parameter vector(I^*_s, Λ, x).

There are many ways to generate the random variable δ_{s+1} whose distribution is given by Equation (3.22). Given the three considerations mentioned above, Importance Sampling is perhaps the most suitable technique as it enables us to obtain accurate results with fewer draws of δ_{s+1} by sampling more frequently from "important" areas of the distribution. Importance Sampling is a classic variance reduction technique. The work by Glynn and Iglehart (1989) [GI89] provides a good overview. To improve the efficiency of the Importance Sampling technique, we use a low-discrepancy sequence. The advantage of using a low-discrepancy sequence instead of a sample generated randomly is the fast convergence rate and systematic improvement in results as we increase the sample size. In Appendix F, we compare the Gauss-Hermite quadrature method against the Quasi-Monte Carlo with Importance Sampling. Both numerical integration methods were satisfactory. However since the integrand takes on large values with small probability (at the tails) and the existence of a point of discontinuity when $t < T-1$, we find that the quasi-Monte Carlo with Importance Sampling is more appropriate in our particular case.

3.4.5 Numerical Results

We compute the value of $p_t(I_t)$ at five different time periods, $t \in \{T-4, T-3, T-2, T-1, T\}$. We use the following numerical values for the state variables in I_t:

$$
\begin{array}{lll}
v_t = 100 & p_{t-1} = 102 & w_{t-1} = 150 \\
\delta_t = -6 & x_{t-1} = -y_{t-1} = 3 &
\end{array}
$$

The values of the constants are:

$$
\begin{array}{lll}
\alpha = 0.005 & \lambda = 1.65 & \overline{\sigma} = 4 \\
\gamma = 0.02 & u = 10 & \theta = 0.20
\end{array}
$$

In what follows, we remove references to the state vector I_t as we use the numerical values shown above. At time $t = T$, we have the trivial base case $p_T = v_T \in I_T$. Hence, $p_T = 100$.

Table 3.2: Numerical estimates to $p_t(I_t)$ and $x_t(p_t(I_t), I_t)$ with $w_{t-1} = 150$.

	$t = T$	$t = T - 1$	$t = T - 2$	$t = T - 3$	$t = T - 4$
$p_t(I_t)$	100	98.92	97.3	93.15	89.01
$x_t(p_t(I_t), I_t)$	-	8.0	6.8	5.0	3.10
$\Lambda_t(I_t)$	0	1.08	2.80	6.83	10.99

Table 3.3: In this set of simulations, we consider loose constraints to the range of the function $x_t(p, I_t)$ by setting $w_{t-1} = 250$ instead of $w_{t-1} = 150$.

	$t = T$	$t = T - 1$	$t = T - 2$	$t = T - 3$	$t = T - 4$
$p_t(I_t)$	100	98.92	97.86	96.05	93.48
$x_t(p_t(I_t), I_t)$	-	8.0	7.93	7.00	6.11
$\Lambda_t(I_t)$	0	1.08	2.14	3.95	6.52

At time $t = T - 1$, we use the analytical solution of p_{T-1} that we derive in Section 3.3. We get $p_{T-1} = 98.92$ and $x_{T-1}(p_{T-1}) = -y_{T-1}(p_{T-1}) = 8$. The liquidity discount at time $t = T - 1$ is $\Lambda_{T-1} = v_{T-1} - p_{T-1} = 1.08$. At time $t \leq T - 2$, we can only obtain a numerical estimate of $p_t(I_t)$ and $x_t(p_t(I_t), I_t)$. In Table 3.2, we provide the estimated numerical values of $p_t(I_t)$, $x_t(p_t(I_t), I_t)$ and $\Lambda_t(I_t)$ at $t \in \{T - 4, T - 3, T - 2\}$. We observe that $p_t(I_t)$ and $x_t(p_t(I_t), I_t)$ decrease as the time horizon $T - t$ increases.

We also perform additional simulations for two different values of $w_{t-1} \in I_t$. First, Table 3.3 shows the values of $p_t(I_t)$ and $x_t(p_t(I_t), I_t)$ when $w_{t-1} = 250$. The increase in w_{t-1} has the effect of loosening the constraint on the range of the demand function $x_s(p, I_s)$ for all $t \leq s$. More specifically, the constraint is imposed by the expression $x \in \mathbb{S}_{t,I_t}$ in Equation (3.20a). Visually, the constraint on $x_s(p, I_s)$ is represented by a "hyperbolic star" in Figure 3.1. In a finance context, this is equivalent to loosening the financing constraints on the arbitrageur's positions.

Next we estimate the values of $p_t(I_t)$ and $x_t(p_t(I_t), I_t)$ when $w_{t-1} = 400$. The values are shown in Table 3.4. The case $w_{t-1} = 400$ is representative to all scenarios where $w_{t-1} > 400$ as they all lead to similar results. In all those cases, the constraint on $x_s(p, I_s)$ for all $t \leq s$ is very loose, and hence almost non-existent.

Typically, when the constraint is very tight, the intersection of $x_s(p, I_s)$ and $-y_s(p, I_s)$ occurs on the "*hyperbolic star*" and, as a result, the values of $p_s(I_s)$ and $x_s(p, I_s)$ drop significantly. On the other hand, when the constraints are loose, as is the case when $w_{t-1} = 400$, the intersection occurs inside the "hyperbolic star". For illustration, refer to Figures 3.2 and 3.3.

Table 3.4: In this set of simulations, we consider no constraints on $x_t(p, I_t)$ by setting $w_{t-1} = 400$.

	$t = T$	$t = T-1$	$t = T-2$	$t = T-3$	$t = T-4$
$p_t(I_t)$	100	98.92	97.92	96.90	95.74
$x_t(p_t(I_t), I_t)$	-	8.0	8.0	8.0	8.0
$\Lambda_t(I_t)$	0	1.08	2.08	3.10	4.26

3.5 Summary

In Table 3.2, Table 3.3 and Table 3.4 we estimate numerically the equilibrium market price p_t, the number of shares traded $x_t(p_t)$ and the discount Λ_t of the market price from its fundamental value v_t. The difference between the three tables is the value of the input parameter w_{t-1}, representing the initial wealth of the arbitrageur. The discount Λ_t is the expected profit that compensates the arbitrageur for the risks of holding the publicly traded asset. As discussed earlier, there are two major risks of holding the asset: 1) the risk of changes in the fundamental value driven by the uncertainty in the final payout, and 2) the risk of *liquidity shocks* triggered by margin calls forcing the arbitrageur to sell at a loss before the final payout. The first risk, called *fundamental risk*, is mainly associated with the volatility of the fundamental value v_t while the second risk, called *systemic risk*, is associated with the fragility of the market micro-structure specifically the leverage of market participants and the margin setting mechanism. The *liquidity risk premium* Λ_t can therefore be decomposed into a *fundamental risk premium* and a *systemic risk premium*, corresponding to each of the two risks discussed above.

We can decompose the value of the liquidity risk premium into its two components if we can build a hypothetical scenario where one of the two risk components is absent. For instance, this can be accomplished by considering a scenario where the arbitrageur has a large initial wealth so that his leverage is too low and as a result, margin calls and *liquidity shocks* never happen. The discount Λ'_t in this hypothetical case corresponds solely to the *fundamental risk premium* as it bears only the risk of changes in the final payout. We can then compute the *systemic risk premium* in our original case by taking $\Lambda_t - \Lambda'_t$.

The values of $\Lambda_t(I_t)$ in Table 3.4 correspond to Λ'_t since $w_{t-1} = 400$. $w_{t-1} = 400$ is the numerical proxy to large arbitrageur wealth. Increasing his wealth further will not change the results since beyond this point the financing constraints are effectively non-existent. In Table 3.5 we summarize our results. The *liquidity risk premium* in Table 3.5 corresponds to $\Lambda_t(I_t)$ in Table 3.2 and the *fundamental risk premium* corresponds to $\Lambda_t(I_t)$ in Table 3.4. The *systemic risk premium* is the difference between the two. We observe that all

Table 3.5: The liquidity risk premium at each time period is decomposed into a *fundamental risk premium* and a *systemic risk premium*, for the case where $w_{t-1} = 150$.

	$t = T$	$t = T - 1$	$t = T - 2$	$t = T - 3$	$t = T - 4$
Liquidity Risk premium	0	1.08	2.80	6.83	10.99
Fundamental Risk premium	0	1.08	2.08	3.10	4.26
Systemic Risk Premium	0	0	0.72	3.73	6.73

Table 3.6: The liquidity risk premium at each time period is decomposed into a *fundamental risk premium* and a *systemic risk premium*, for the case where $w_{t-1} = 250$.

	$t = T$	$t = T - 1$	$t = T - 2$	$t = T - 3$	$t = T - 4$
Liquidity Risk premium	0	1.08	2.14	3.95	6.52
Fundamental Risk premium	0	1.08	2.08	3.10	4.26
Systemic Risk Premium	0	0	0.06	0.85	2.26

three risk premiums increase as the time-to-maturity increases (i.e., as $T - t$ increases). Clearly, the volatility of final payoff increases as time-to-maturity increases and the risk of a liquidity shock before the final payoff increases as well. Note that at time $t = T - 1$, the *systemic risk premium* is zero since there is no risk of liquidity shocks in the next period as the arbitrageur is not selling the asset to anyone, but instead receiving a cash dividend equal to the fundamental value v_T. On the other hand, the fundamental risk premium is non-zero since the value of the cash dividend is uncertain.

In Table 3.6, we compute the risk premiums when the initial wealth of the arbitrageur is higher, $w_{t-1} = 250$. In this case, the financing constraints are more loose and the arbitrageur is able to provide a better liquidity to the seller by purchasing a larger quantity of the asset at a price closer to the fundamental value. Comparing the risk premiums in Table 3.6 to those in Table 3.5, we observe that the *liquidity risk premium* is smaller due to a smaller *systemic risk premium*. Since the financing constraints are looser, the risk of liquidity shocks prior to the final payoff is lower. The *fundamental risk premium* however doesn't change since it depends on the volatility of the fundamental value v_t, not the wealth of the arbitrageur.

Chapter 4

Implications for Risk Management

In Chapter 2, we provided empirical evidence on the effects of funding liquidity on arbitrage spreads, and in Chapter 3 we illustrated the mechanism through which that happens. In this chapter, we discuss the implications for risk management. From a risk management perspective, it is very important to understand the dynamic nature of the covariances and expected returns since they constitute major inputs to any portfolio construction process or other protocols such as risk budgeting and exposure constraints at a portfolio level.

In this chapter, we take a two-step approach. First, we fit a Markov regime-switching model and show that it identifies two regimes: 1) a "good regime" characterized by positive returns, moderate volatilities and low correlations, and 2) a "bad regime" characterized by negative returns, high volatilities and high correlations. We find that the FLSI index has a strong explanatory power for the probability of being in a "bad regime". Therefore, a rising FLSI indicates an increasing probability of entering a regime with high volatility and high correlations.

Second, we decompose the changes in market prices into two components: a change in the fundamental value and a change in the liquidity discount. We find that the effects of funding liquidity on the liquidity discount accounts for the majority of the increase in the volatility and correlations during periods of stress.

4.1 Regime Switching

In this section, we fit a two-states Markov regime switching model on the returns of two different asset classes represented by: 1) a currency carry trade index and 2) the S&P

500 index. By studying two fundamentally uncorrelated asset classes, we gain insights on the dynamic nature of correlation. We further show that the Markov model switches between two regimes which we label "good regime" and "bad regime". The "bad regime" is characterized by periods of negative returns, high volatilities and high correlation. It also captures most of the financial crises and recessions that occurred in the last two decades. Regressing the probability of being in a "bad regime" against the FLSI index, we observe that the factor loading is positive and significant. It indicates that "bad regimes" coincide with periods of worsening funding liquidity.

4.1.1 Currency Carry Trade Index

A currency carry trade is a trading strategy consisting of selling low interest rate currencies and investing the proceeds in high interest rate currencies. The low-yielding currencies are called "funding currencies", while the high-yielding currencies are called "investment currencies". The total return of a currency carry trade is composed of 1) the interest rate differential, and 2) the change in foreign exchange rates. In theory, if a country raises its interest rate, it attracts foreign capital leading to an immediate appreciation of its currency. A currency *depreciation* should follow such that the "uncovered interest rate parity" (UIP) holds. In practice however, money flows into the country slowly [MPP07] and, as a result, the currency *appreciates* gradually. The gradual appreciation of the high-yielding currency is exactly the opposite of what is expected in theory. This phenomenon is known as the "forward premium puzzle". [FT90], [Lew95] and [Eng96] conducted empirical studies, reporting positive mean returns on carry trades. They essentially provide empirical evidence of the "forward premium puzzle".

Other studies have attempted to explain why carry trades are profitable on average. [BNP09] argues that the positive average return on currency carry trades compensates investors for "endogenous" crash risk. The carry trade is essentially a crowded trade[1] [PL11] given the simplicity to construct such trade. The returns of crowded trades have large negative skewness as leveraged investors are often forced to liquidate their positions at the same time, creating selling pressures on market prices [Ped09].

[FG08a] argues that carry trades have positive average returns because countries of the high-yielding currencies are more exposed to global economic disasters, i.e., the carry trade is sensitive to "exogenous" crash risk. According to the market microstructure model developed by [BP09], endogenous shocks are often triggered by exogenous shocks to arbitrageurs' capital, and hence, the explanations provided by [BNP09] and [FG08a] are

[1] A crowded trade is an *"investment strategy that [is very] similar across hedge funds"*.

Table 4.1: We compute 1) the annualized mean, 2) annualized standard deviation, and 3) skewness of the carry trade index returns. We consider the period from 1999 to 2011 and we compute those statistics for two exclusive sub-periods: 'normal' and 'crisis'. The 'crisis' sub-period covers two recessions (refer to Table A.1) as well as other periods of financial stress listed in Table A.2 and Table A.3. Not unexpectedly, the returns of the carry trade index have higher volatility, higher skewness and lower mean return during the 'crisis' sub-period.

Period	# of days	Mean (annual)	Volatility (annual)	Skewness
Normal	4,856 days	5.78%	9.4%	-0.09
Crisis	951 days	-2.29%	17.07%	-0.21

reinforcing. Moreover, [BH07] and [BGP07] argue that investors' general preference for positive skewness, leads to higher expected returns for strategies with negative skewness such as carry trades.

The most popular carry trade consists of selecting the high-yielding Australian dollar (AUD) and New Zealand dollar (NZD) as the "investment currencies" [BNP09] and the low-yielding U.S. dollar (USD) as the "funding currency". We plot the foreign exchange rates of the AUD and NZD with respect to the USD in Figure 4.1 and Figure 4.2, respectively. In Figure 4.3, we plot the currency carry trade index consisting of selling the USD and investing the proceeds equally in AUD and NZD. The index has a base of 100 on 1989/02/28. In Figure 4.4 we plot the returns of the index. It is essentially the total return of the carry trade: interest rate differential + changes in foreign exchange rates. In Figure 4.5, we normalize the returns over the period from 1999 to 2011 by subtracting the mean and dividing by the standard deviation. We observe that at the peak of the crisis (i.e., around the Lehman Brothers bankruptcy), the carry trade incurred a record loss exceeding -8 standard deviations. Moreover, during most periods of stress losses exceeded -2 standard deviations.

In Table 4.1, we compute 1) the annualized mean, 2) annualized standard deviation, and 3) skewness of the carry trade index returns. We consider the period from 1999 to 2011 and we compute those statistics for two exclusive sub-periods: 'normal' and 'crisis'. The 'crisis' sub-period covers two recessions (refer to Table A.1) as well as other periods of financial stress listed in Table A.2 and Table A.3. Not unexpectedly, the returns of the carry trade index have higher volatility, higher skewness and lower mean during the 'crisis' sub-period.

Figure 4.1: AUD/USD exchange rate.

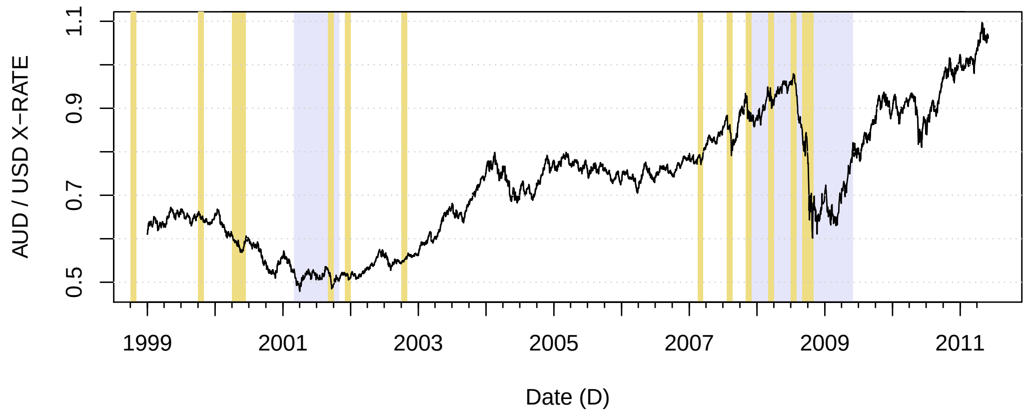


Figure 4.2: NZD/USD exchange rate.

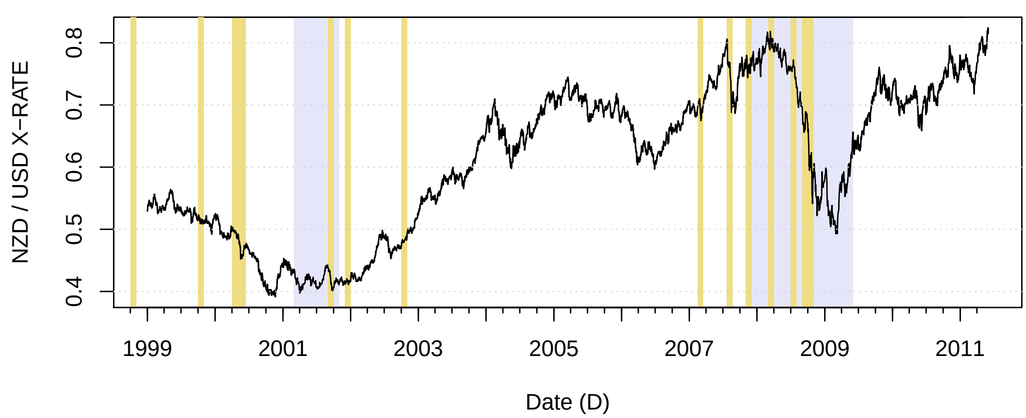


Figure 4.3: Currency Carry Trade Index. Long: AUD & NZD, Short: USD. It is the cumulative total return of the carry trade starting from a base of 100 in 1989/02/28.

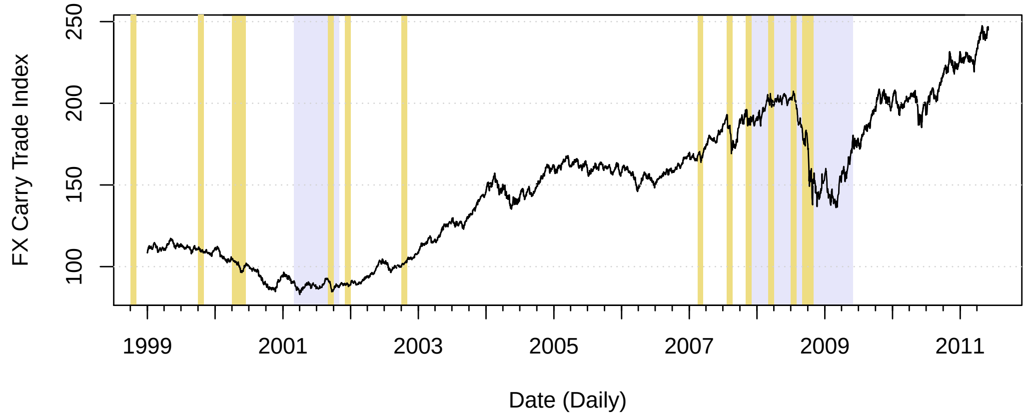


Figure 4.4: Currency Carry Trade Index Returns. Long: AUD & NZD, Short: USD. The return is essentially the total return of the carry trade: interest rate differential + changes in foreign exchange rates.

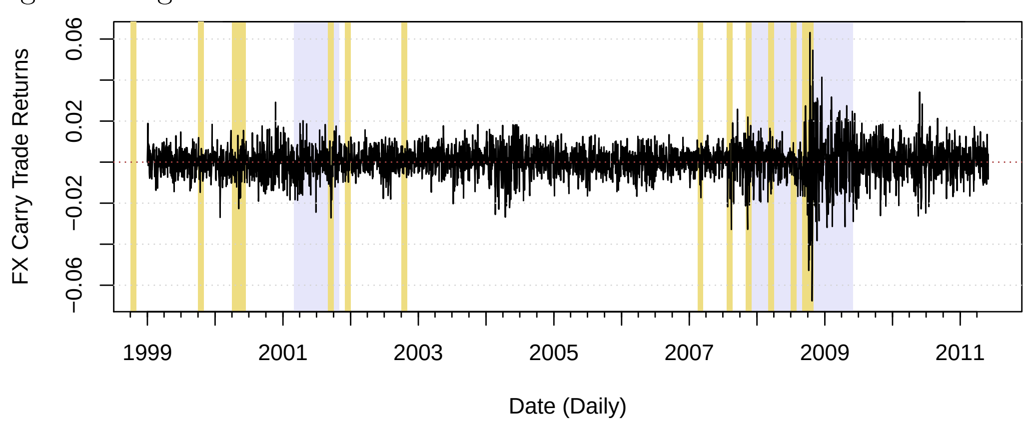
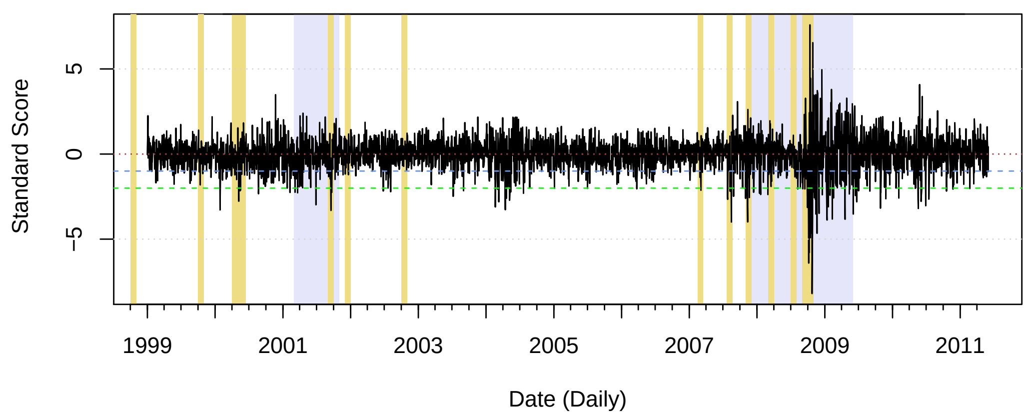


Figure 4.5: We normalise the time-series in Figure 4.4 over the period ranging from 1999 to 2011. During the period of stress, losses on the carry trade exceeded -8 standard deviation. The blue and green dashed lines are the -1 and -2 standard deviations, respectively. We observe that, during most periods of stress, losses on the carry trade exceeded -2 standard deviations. After the periods of stress (see also Figure 4.3), losses were reversed with daily gains exceeding $+2$ standard deviations. It is therefore important that arbitrageurs have staying power to survive through the crises and participate in the recovery. Those that had high leverage, were most likely forced to liquidate their positions at a loss and hence couldn't participate fully in the recovery.



4.1.2 Regime Switching and FLSI

Next we calibrate a two-state Markov regime switching model on two time-series: 1) the weekly returns of our currency carry trade index, and 2) the weekly returns of the S&P 500 index. Regime switching regression models date back to [Qua58]. [GQ73] introduced a regime switching model in which the latent state variable follows a Markov chain. Such models are called Markov switching models. In [Ham89], Hamilton proposed a numerical technique to compute the parameters. His technique is however limited to relatively small systems since it is computationally expensive to maximize numerically a coarse likelihood surface with respect to many unknown parameters. In [Ham90], he introduced an alternative model to estimate the parameters. He proposed an EM algorithm for obtaining the maximum likelihood estimates, a technique that we use in this section.

We consider a Markov switching lognormal model, similar to [Har01], where $\rho_t = 1, 2$ determines the regime in the period $(t, t+1)$ and S_t is the index value at time t. In particular, we have:

$$log\frac{S_{t+1}}{S_t} \mid \rho_t \sim N\left(\mu_{\rho_t}, \Omega_{\rho_t}\right) \tag{4.1}$$

The transition matrix P denotes the probabilities of moving regimes, that is,

$$p_{i,j} = Pr\left[\rho_{t+1} = j \mid \rho_t = i\right] \qquad i = 1, 2 \quad \text{and} \quad j = 1, 2 \tag{4.2}$$

The following are the parameters that we need to estimate:

$$\Theta = \{\mu_1, \mu_2, \Omega_1, \Omega_2, p_{1,1}, p_{2,2}\} \tag{4.3}$$

We obtain the following estimates (applied to weekly returns, i.e, not annualized) using the EM algorithm [Ham90]:

$$\begin{aligned} &p_{1,1} = 0.96 && p_{2,2} = 0.89 \\ &\Omega_1 = \begin{bmatrix} 1.08\ bps & 0.11\ bps \\ 0.11\ bps & 2.57\ bps \end{bmatrix} && \Omega_2 = \begin{bmatrix} 5.64\ bps & 4.16\ bps \\ 4.16\ bps & 13.20\ bps \end{bmatrix} \\ &\mu_1 = \begin{bmatrix} 0.14\% \\ 0.30\% \end{bmatrix} && \mu_2 = \begin{bmatrix} -0.06\% \\ -0.23\% \end{bmatrix} \end{aligned} \tag{4.4}$$

In regime 1, the annualized mean and volatility of the *currency carry trade index* returns are 7.28% and 7.49%, respectively. That compares with a mean of −3.2% and a volatility of

17.12%, in regime 2. In regime 1, the annualized mean and volatility of the *S&P* 500 *index* returns are 15.60% and 11.56%, respectively. That compares with a mean of −11.96% and a volatility of 26.20%, in regime 2. The correlation coefficient between the currency carry trade and the S&P 500 returns jumped from 0.06 in regime 1 to 0.48 in regime 2. Compared to regime 1, returns in regime 2 have lower means, higher volatilities and higher correlations. We therefore label regime 2 as the "bad regime". We plot in Figure 4.6 the probability of being in the "bad regime" (i.e., regime 2). We see that during periods of financial stress and during recessions, the probability of being in the "bad regime" is high. In Table 4.2, we regress the probability of being in the "bad regime" (i.e., the time series in Figure 4.6) against the Funding Liquidity Stress Index (FLSI) and PC2. Not unexpectedly, we find that the loading coefficient of FLSI is positive and significant at the 1 percentile level, while the loading coefficient of PC2 is not significant. The adjusted R^2 of the regression is high, 33.9%, indicating that "bad regimes" coincide with periods of deteriorating funding conditions.

Figure 4.6: We fit a regime switching model to the returns of our Currency Carry Trade Index and the S&P 500 Index. We plot the implied probability of being in regime 2 (labelled "bad regime"). The areas in yellow and grey correspond to periods of financial crisis and recessions, respectively. We note the high probability of being in the "bad regime" during those periods.

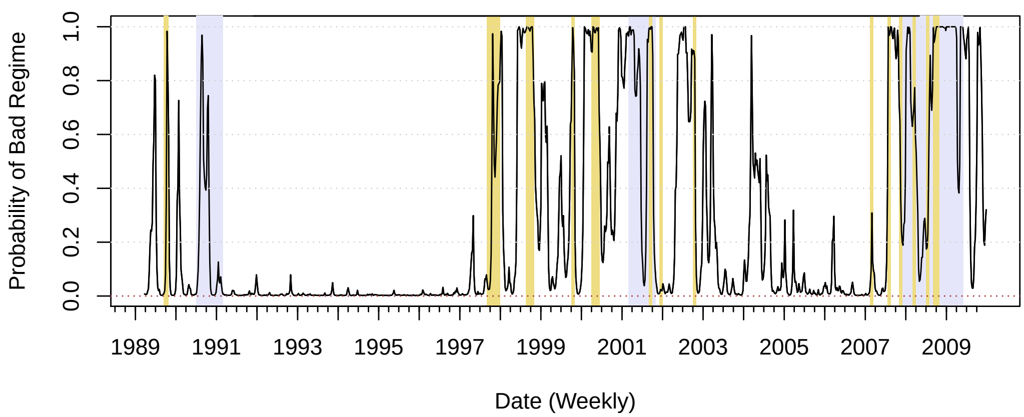


Table 4.2: We regress the probability of being in the "bad regime" against the Funding Liquidity Stress Index (FLSI) and PC2. FLSI is significant at the 1 percentile level while PC2 is not significant. The adjusted R^2 of the regression is high, 33.9%, indicating that "bad regimes" coincide with periods of deteriorating funding conditions.

	Probability of being in the "bad regime"
(Intercept)	0.387***
	(0.014)
FLSI	0.231***
	(0.013)
PC2	-0.015
	(0.014)
R^2	0.339
$adj.R^2$	0.337
N	575

Standard errors in parentheses
* (p $\leq$ 0.1), ** (p $\leq$ 0.05), *** (p $\leq$ 0.01)

4.2 Liquidity Discounts and the Measurement of Risk

Next, we decompose the changes in market prices into two components: a change in the fundamental value and a change in the liquidity discount. We show that the effects of funding liquidity on the liquidity discount accounts for the majority of the increase in the volatility and correlations during periods of stress. The arbitrage spreads studied in Chapter 2 are very good examples of liquidity discounts and will be used to illustrate the contribution of liquidity discounts to the measurement of risk of market prices.

Let p_t be the market price of a security and v_t its fundamental value. Fundamental values are observable for certain types of securities while unobservable for others. For instance, the fundamental value of a closed-end fund is its Net Asset Value (NAV), the fundamental value of a company being acquired is the offered cash value, and the fundamental value of a bond is the present value of future cash flows. On the other hand, the fundamental value of a commodity is harder to compute as it is mainly dependent on the economic supply and demand which are hard to measure in practice. The fundamental value of a common stock is typically computed through a dividend discount model.

4.2.1 Decomposing the Returns of Market Prices

The market price can be written as:

$$p_t = v_t - \Lambda_t \tag{4.5}$$

where Λ_t is the price discount from its fundamental value. We shall refer to Λ_t as the liquidity discount. When $\Lambda_t > 0$, we say that the security trades at a discount. When $\Lambda_t < 0$, we say that it trades at a premium. Let $r_{p,t+1}$, $r_{v,t+1}$ and $r_{\Lambda,t+1}$ be the percentage change (i.e., return), from t to $t+1$, of the market price, fundamental value and liquidity discount, respectively. For instance,

$$r_{p,t+1} = \frac{p_{t+1} - p_t}{p_t} = \frac{\Delta p_{t+1}}{p_t} \tag{4.6}$$

Using Equation (4.5) and Equation (4.6), the return of the market price can be written as:

$$\begin{aligned} r_{p,t+1} &= r_{v,t+1} \cdot \frac{v_t}{p_t} - r_{\Lambda,t+1} \cdot \frac{\Lambda_t}{p_t} && (4.7) \\ &= r_{v,t+1} \cdot w_t + r_{\Lambda,t+1} \cdot (1 - w_t) && (4.8) \end{aligned}$$

where $w_t = \frac{v_t}{p_t}$. The return of the market price is therefore a weighted average of 1) the percentile change of its fundamental value and 2) the percentile change of the liquidity discount. These two components behave differently and dictate the behaviour of the market price. The fundamental value can be a stochastic process reflecting the state of the economy and/or the prospects of a firm. The liquidity discount however follows a mean-reverting process reflecting the funding liquidity conditions. The behaviour of the liquidity discount was illustrated by the arbitrage spreads in previous sections.

To understand Equation (4.8), suppose that the security trades at a discount. That is $p_t < v_t$, $w_t = \frac{v_t}{p_t} > 1$, and $(1 - w_t) < 0$. Then, a positive change of the fundamental value *increases* the price return, whereas a positive change of the discount (i.e., widening of the spread) *decreases* the return.

4.2.2 Decomposing the Expected Return

From Equation (4.8), the expected return of the security at time t is:

$$\mathrm{E}_t\left[r_{p,t+1}\right] = \mathrm{E}_t\left[r_{v,t+1}\right] \cdot w_t + \mathrm{E}_t\left[r_{\Lambda,t+1}\right] \cdot (1 - w_t) \tag{4.9}$$

When the security trades at a discount, $\frac{v_t}{p_t} = w_t > 1$, the expected return is essentially a leveraged exposure to the expected percentile change of the fundamental value, in addition to the expected percentile change of the liquidity discount.

For illustration, we'll consider a closed-end fund, the Western Asset High Income Fund (ticker: HIF). On October 22$^{\text{nd}}$ 2008, the fund distributed a monthly dividend of \$0.07 and was trading at \$4.50. That is equivalent to an annual yield of 18.67% on a publicly traded fund of investment grade bonds. However, the value of the underlying bonds (i.e., the net asset value of the fund) was \$7.49 which implies a yield of 11.21% on the fundamental value. An investor therefore could earn a yield of 18.67% by buying the fund at its market price instead of buying the underlying bonds yielding only 11.21% [2]. In this case, the change in fundamental value is

$$r_{v,t+1} = \frac{\delta_{t+1} + v_{t+1} - v_t}{v_t} = \frac{\delta_{t+1} + \Delta v_{t+1}}{v_t} \tag{4.10}$$

where δ_t is the dividend distributed by the bonds at time $t+1$. Filling the values of the variables in Equation (4.9) and using Equation (4.10), we obtain the expected return of the closed-end fund:

$$\begin{aligned}
\mathrm{E}_t\left[r_{p,t+1}\right] &= \frac{7.49}{4.50} \cdot \mathrm{E}_t\left[r_{v,t+1}\right] + \left(1 - \frac{7.49}{4.50}\right) \cdot \mathrm{E}_t\left[r_{\Lambda,t+1}\right] && (4.11)\\
&= 1.66 \cdot \mathrm{E}\left[\frac{\delta_{t+1}}{7.49} + \frac{\Delta v_{t+1}}{7.49}\right] - 0.66 \cdot \mathrm{E}_t\left[\frac{\Lambda_{t+1} - 2.99}{2.99}\right] && (4.12)\\
&= 1.66 \cdot 11.21\% + 1.66 \cdot \mathrm{E}\left[\frac{\Delta v_{t+1}}{7.49}\right] + 0.66 - 0.66 \cdot \mathrm{E}_t\left[\frac{\Lambda_{t+1}}{2.99}\right] && (4.13)
\end{aligned}$$

The return on the fund is composed of three components: 1) a leveraged yield, 2) a leveraged price return on the NAV, and 3) a return associate with the future change of the liquidity term (the last two terms of Equation (4.13)). The last term goes to zero if the market price of the closed-end fund converges to its fundamental value ($\Lambda_{t+1} \to 0$), which is typically the case in the long run. Hedging the changes in the fundamental value (i.e., the second term) creates a very attractive arbitrage opportunity for long term investors: they can earn 66% on their investment as the market price convergences to its fundamental value and will earn an additional $1.66 \times 11.21\% = 18.67\%$ each year from dividend distributions while waiting for the price to converge. That opportunity in particular existed in the aftermath of the bankruptcy of Lehman Brothers. In January 2009, the market price of HIF converged to its fundamental value, yielding a minimum

[2] The two alternatives have the same long-term risk, that is, the risk of default by the underlying bonds.

gain of 66% on a diversified portfolio of investment grade bonds. Arbitrage opportunities of such amplitude exist mostly during periods of stress.

4.2.3 Decomposing the Variance

The variance of $r_{p,t+1}$ can be written as:

$$\begin{aligned}\operatorname{var}(r_{p,t+1}) = w_t^2 \cdot \operatorname{var}(r_{v,t+1}) + (1-w_t)^2 \cdot \operatorname{var}(r_{\Lambda,t+1}) \\ + 2 \cdot (1-w_t) \cdot w_t \cdot \operatorname{covar}(r_{v,t+1}, r_{\Lambda,t+1})\end{aligned} \tag{4.14}$$

When the market price is equal to the fundamental value, we have $w_t = 1$ and as a result, $\operatorname{var}(r_{p,t+1}) = \operatorname{var}(r_{v,t+1})$. However, during periods of stress, $w_t > 1$ and the variance of the market price depends on three components: 1) the variance of the fundamental value, 2) the variance of the liquidity discount, and 3) the covariance between the two.

The first component, the variance of the fundamental value, is very asset specific. It depends on the economy and the prospects of the asset. For instance, if it is the common stock of a company, then the variance of the fundamental value depends on the cyclical nature of the business, its balance sheet leverage, and its competitive positioning in the industry, among other factors. The treatment of this term is beyond the scope of this work.

The second component is the variance of the liquidity discount. Since the liquidity discount is the deviation of the market price from its fundamental value, arbitrage spreads are very good examples. In Table 4.3, we compute the variance of the arbitrage spreads of the following strategies: 1) Closed-End Funds (see Figure 2.14), 2) Mergers & Acquisitions (see Figure 2.19), and 3) on-the-run/off-the-run Treasuries (see Figure 2.24). The variance is computed, from January 2000 to June 2011, over two exclusive periods: 1) a "favourable" period, and 2) a"non-favourable" period. The "non-favourable" period consists of all the days when the Funding Liquidity Stress Index (FLSI) was above its median. The remaining half of the time is the "favourable" period. In Table 4.3, we observe that the variance of each arbitrage spread over the "non-favourable" period is many times larger than its variance over the "favourable" period. The variance of the liquidity discount therefore contributes significantly to the total variance of the market price during the "non-favourable" period.

The third component in Equation (4.14) is the covariance between the fundamental value and the liquidity discount. In Table 4.4, we compute the covariance between the fundamental value and the liquidity discount for a sample of closed-end funds[3]. For each

[3] We select randomly a couple of closed-end funds from each "investment objective" category listed in Table 2.10. The sample obtained is representative of the wide variety of closed-end funds across various investment objective categories.

Table 4.3: We compute the variance of the arbitrage spreads of the following strategies: 1) Closed-End Funds (see Figure 2.14), 2) Mergers & Acquisitions (see Figure 2.19), and 3) on-the-run/off-the-run Treasuries (see Figure 2.24). The variance is computed during "non-favourable" and "favourable" periods, from January 2000 to June 2011. We observe that the variance of each arbitrage spread over the "non-favourable" period is many times larger than its variance over the "favourable" period.

	CEF Spreads	M&A Spreads	Government Spreads
"favorable" Period	4.08	0.43	0.02
"non-favorable" Period	13.19	2.24	0.09

closed-end fund, we compute the covariance during two periods: the two months surrounding the bankruptcy of Lehman Brother ("stress" period) and the same two months but two years earlier ("normal period"). As shown in Table 4.4, the covariance is significantly negative during the "stress" period for *all* the closed-end funds in our sample. During the "normal" period however, the amplitudes of the covariances are smaller and the signs vary. We perform a $t-$test of the null hypothesis that the covariance observed in the "normal" period is drawn from a normal distribution with mean zero and unknown variance, against the alternative that the mean is not zero. We obtain a pvalue of 0.90, and therefore can not reject the hypothesis. As a result, since both $(1-w_t)\cdot w_t$ and $\mathrm{covar}\left(r_{v,t+1}, r_{\Lambda,t+1}\right)$ are negative and have large magnitudes during periods of stress, the covariance term in Equation 4.14 contributes significantly more to the variance of market prices during periods of stress than during "normal" periods.

4.2.4 Decomposing the Covariance

The covariance between the returns of security a and security b can be decomposed as follows:

$$\begin{aligned}\mathrm{cov}\left(r^a_{p,t}, r^b_{p,t}\right) = {} & w^a_t \cdot w^b_t \cdot \mathrm{cov}\left(r^a_{v,t}, r^b_{v,t}\right) + w^a_t\left(1-w^b_t\right)\mathrm{cov}\left(r^a_{v,t}, r^b_{\Lambda,t}\right) \\ & + \left(1-w^a_t\right) w^b_t \cdot \mathrm{cov}\left(r^a_{\Lambda,t}, r^b_{v,t}\right) + \left(1-w^a_t\right)\left(1-w^b_t\right)\mathrm{cov}\left(r^a_{\Lambda,t}, r^b_{\Lambda,t}\right)\end{aligned} \tag{4.15}$$

The first term, $w^a_t \cdot w^b_t \cdot \mathrm{cov}\left(r^a_{v,t}, r^b_{v,t}\right)$, is the covariance between the fundamental values of the two securities. If both price discounts are zero, such that $w^a_t = 1$ and $w^b_t = 1$, the covariance of the market prices of the two securities is equal to the covariance of their

Table 4.4: We compute the covariance between the fundamental value and the liquidity discount for a representative sample of closed-end funds. For each closed-end fund, we compute the covariance during two periods: the two months surrounding the bankruptcy of Lehman Brother ("stress" period) and the same two months but two years earlier ("normal period").

Ticker	"Stress" period 11/08/2008 to 11/10/2008	"normal" period 11/08/2006 to 11/10/2006	Ticker	"Stress" period 11/08/2008 to 11/10/2008	"normal" period 11/08/2006 to 11/10/2006
HIF	−35.4%	6.5%	EEF	−23.5%	0.1%
CIK	−32.7%	−18.0%	EOI	−42.2%	−4.5%
ADF	−41.1%	25.1%	EOS	−30.0%	−19.9%
KMM	−23.8%	2.0%	EVG	−25.4%	−4.3%
ERC	−24.6%	−4.3%	GFY	−32.2%	1.3%
NKT	−28.2%	1.3%	HIO	−31.6%	−11.9%
ARK	−24.7%	9.5%	JOC	−27.4%	−17%
AVK	−55.2%	2.6%	KHI	−16.9%	28.5%
CYE	−21.3%	4.2%	NOX	−53.3%	−14.3%
NKT	−20.8%	23.2%	RDR	−9.1%	−2.2%

fundamental values. The covariance of fundamentals depends on the type of assets or the nature of the businesses (if the securities are common stocks).

The last term, $(1-w_t^a)\left(1-w_t^b\right)\operatorname{cov}\left(r^a_{\Lambda,t}, r^b_{\Lambda,t}\right)$, is the covariance between the liquidity discounts of the two securities. This term is primarily driven by funding liquidity conditions. During periods of stress, liquidity discounts become highly correlated as leveraged investors are often forced to reduce their positions to adjust their leverage, and as a result, liquidity discounts of the various assets held in their portfolios widen [BP09]. Fundamentally uncorrelated assets therefore become correlated due to the co-movement of their liquidity discounts. For instance, the covariance between the arbitrage spreads of the Mergers & Acquisitions and Closed-End Funds strategies jumps to 3.2 during "non-favourable" periods[4] compared to 0.65 during "favourable" periods (a five-fold increase!). The covari-

[4]Similar to the definition in Table 4.3, "Non-favourable" periods consist of the days from January 2000 to June 2010 when the Funding Liquidity Stress Index (FLSI) was above its median. The "favourable" periods are composed of the remaining days. The paste 10 years are therefore divided into two exclusive periods depending on whether the funding conditions are favourable or not, that is the FLSI is below or above its median, respectively.

ance of the liquidity discounts is therefore a significant contributor to the covariance of market prices when funding conditions are non-favourable.

Chapter 5

Conclusion

"Despite the cries of newspapers to lower the interest rates, the Fed would sometimes do much better to attend to the economy-wide leverage and leave the interest rate alone". - John Geanakoplos [Gea09].

Arbitrageurs are active market participants and most arbitrage is conducted with leverage. Funding liquidity, the ease by which arbitrageurs can borrow to fund their positions, is a critical indicator of the health of the financial system. In Chapter 2, we studied the effects of funding liquidity on arbitrage strategies. We provided empirical evidence to various theoretical work on market micro-structure models, which we generalize in Chapter 3, that attempt to explain the vulnerability of arbitrage to funding liquidity [SV97] [GV09] [BP09]. We constructed a novel Funding Liquidity Stress Index (FLSI), with daily values, composed of various measures of funding conditions. The measures were grouped into four categories: 1) margin requirements on futures contracts, 2) general collateral repo spreads, 3) LIBOR-based measures, and 4) corporate credit yield spreads. Interestingly, the index explains 78% of the idiosyncratic volatility of banks stock prices. That strong relationship is not unexpected as the banking sector is very dependent on a healthy funding environment to roll-over its maturing debts. We also reproduced three actual arbitrage strategies: 1) closed-end funds arbitrage, 2) mergers & acquisitions arbitrage, and 3) on-the-run/off-the-run Treasury arbitrage. We find that the FLSI index has a strong explanatory power to the changes of arbitrage spreads. It is also the main source of contagion between them. Next, we performed event-studies surrounding events of changing margin requirements. In particular, we showed that margins rise after periods of increasing volatility and decreasing asset prices which supports the "margin setting mechanism" in [BP09] as well as the assumption regarding the use of "backward-looking risk measures" by market participants in [BCG$^+$09]. More importantly, asset prices continue to decrease after margins increase, providing evidence of spiralling effects as predicted by the theoretical model of [BP09]. The opposite is also true. Margins decrease after periods of rising prices and decreasing market volatility. Our observations further support the concepts of "performance-based arbitrage" in [SV97] and "pro-cyclicality" of financing in [Gea09].

Next, we studied the *liquidity risk premium* in a market micro-structure framework where market prices are determined by the supply and demand of securities. We extend the model developed by Brunnermeier and Pedersen [BP09] to multiple periods and generalize their work by considering all market participants to be risk-averse. Computing the liquidity risk premium consists of solving an optimization problem. We show that it is an instance of "finite dynamic optimization problems" to which we derive a recursive formulation. We use the recursion to compute numerical values of the liquidity risk premium at various points in time. We further decompose the liquidity risk premium into two components: 1) a *fundamental risk premium* and 2) a *systemic risk premium*. The fundamental risk premium compensates market participants for providing liquidity in a security whose fundamental value is volatile, while the systemic risk premium compensates them for taking positions in a market that is vulnerable to liquidity shocks. The first component is therefore related to the nature of the security while the second component is related to the fragility of the market micro-structure (such as leverage of market participants and margin setting mechanisms).

Last, we discussed the implications for risk management. In particular, we showed that an increasing Funding Liquidity Stress Index indicates an increasing probability of entering a regime with high volatility and high correlations. We further explained that the liquidity component of market prices accounts for the majority of the increase in volatility and correlations.

Liquidity discounts play a very important role in risk management. Although securities can belong to various asset classes, they all have in common the exposure of their liquidity discounts to the prevalent funding conditions. During non-crisis periods, the changes of the fundamentals outweigh the changes of the liquidity discounts. Models backtested over non-crisis periods will therefore falsely assume that fundamentals govern the behavior of market prices at all time. As a result, the models will underestimate the skewness, co-variance and serial correlation of market prices during periods of stress in which liquidity discounts dominate changes in market prices. During a speech[1] made on November 12, 2008, the Vice Chairman of the federal reserve stated: *"...good economic performance provided skewed data and bred complacency: House prices could only go up; income interruptions and problems servicing debt were likely to be short lived; financial markets would*

[1] http://www.federalreserve.gov/newsevents/speech/kohn20081112a.htm

always be liquid. Models based on theory and estimated with data from the 1990s and early 2000s fit well–too well. Complacency, in turn, contributed to the unwillingness of many financial market participants to enhance their risk-management systems sufficiently to take full account of the new (perhaps unknown) risks they were taking on. . . . Finally, many market participants also had inadequate liquidity backstops, apparently because they (wrongly) assumed that markets would be sufficiently liquid to smoothly adjust risk profiles to new developments in markets and the broader economy." In his speech, the Vice Chairman essentially blamed models calibrated over "good" periods for the lack of capital cushion in financial institutions to withstand a large systemic risk. A dynamic risk management approach is perhaps more appropriate. For instance, risk systems could use a financial stress indicator, similar to the Funding Liquidity Stress Index that we developed, to assess the probability of being in a crisis (or a non-crisis) state. Models would then be calibrated and tested over those past periods similar to the present time.

APPENDICES

Appendix A

List of Financial Events

In the United States, recessions are officially determined and announced by the *National Bureay of Economic Research*. We list the dates of the last six recessions in Table A.1. A complete list is available at http://www.nber.org/cycles.html. In the past two decades, the United States witnessed two recessions and various periods of financial stress. We divide those periods into two groups: 1) the $1998-2002$ crisis, and 2) the $2007-2009$ crisis. We list the periods of financial stress corresponding to the $1998-2002$ crisis in Table A.2 and those corresponding to the $2007-2009$ crisis in Table A.3. Throughout this work, whenever we plot a time-series, we represent recessions by grey areas and periods of financial stress by yellow areas. Our reader can refer to Table A.2 and Table A.3 to identify the event for any particular yellow area.

Table A.1: Official dates of recessions in the United States. Recessions are announced by the *National Bureay of Economic Research*. They are represented by grey areas in the plots throughout this work. Source: http://www.nber.org/cycles.html

Start Date	**End Date**	**Duration in Months**
2007-12-01	2009-06-01	18
2001-03-01	2001-11-01	8
1990-07-01	1991-03-01	8
1981-07-01	1982-11-01	16
1980-01-01	1980-07-01	6
1973-11-01	1975-03-01	16

Table A.2: Periods of financial stress corresponding to the 1998-2002 crisis

Start Date	**End Date**	**Event**
2002-10-01	2002-11-01	Accounting Scandals
2001-12-01	2002-01-01	Enron Bankruptcy
2001-09-01	2001-10-01	September 11 2001
2000-04-01	2000-06-15	tech collapse
1999-10-01	1999-11-01	Y2K
1998-10-01	1998-11-01	Russia - LTCM

Table A.3: Periods of financial stress corresponding to the 2007-2009 crisis

Start Date	**End Date**	**Event**
2008-10-01	2008-11-01	Political debate over TARP
2008-09-01	2008-10-01	Lehman Brothers failure - AIG rescue
2008-07-01	2008-08-01	Indymac failure - Fannie and Freddie woes
2008-03-01	2008-04-01	Bear Stearns collapse
2007-11-01	2007-12-01	bank mortgage writeoffs - monoline troubles
2007-07-21	2007-08-21	BNP freezes redemptions - Bear Stearns HF losses
2007-02-15	2007-03-15	Chinese correction

Appendix B

Margins on three Different Futures Contracts

We plot the dollar and percentile margins on three different futures contracts: 1) the Dow Jones index, 2) the AUD/USD foreign exchange rate, and 3) the NZD/USD foreign exchange rate. The margins on the S&P 500 futures contract were plotted previously in Figure 2.1 and Figure 2.2.

Figure B.1: Margin requirement (in dollars) on the Dow Jones futures contract. It is the dollar amount required to initiate a position in the contract. Note that the futures contract is on 10 units of the S&P 500 index.

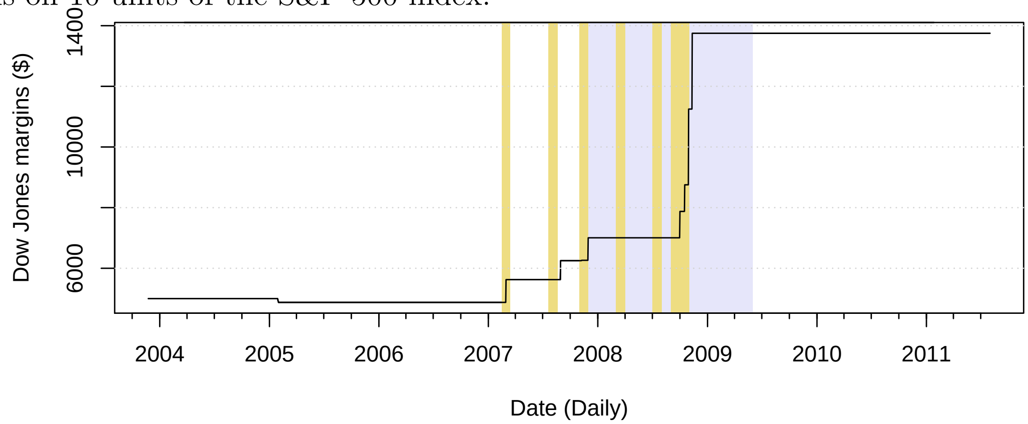


Figure B.2: Margin requirement (in percent) on the Dow Jones futures contract.

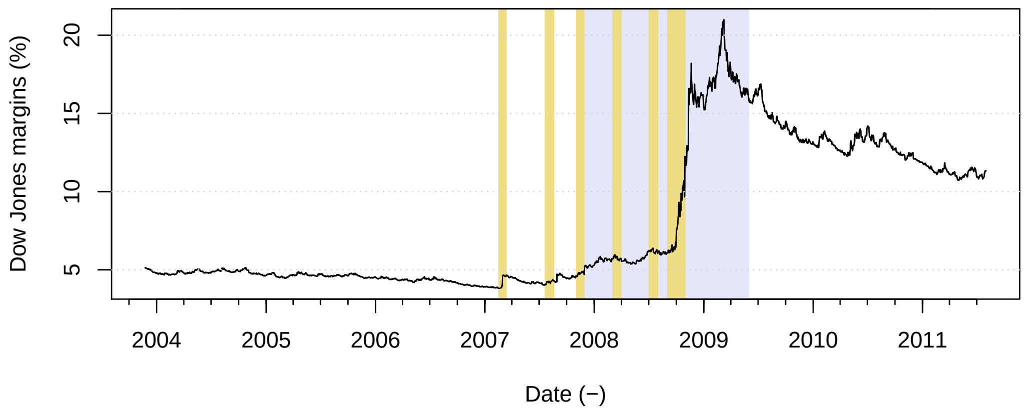


Figure B.3: Margin requirement (in Dollars) on the AUD / USD FX futures contract.

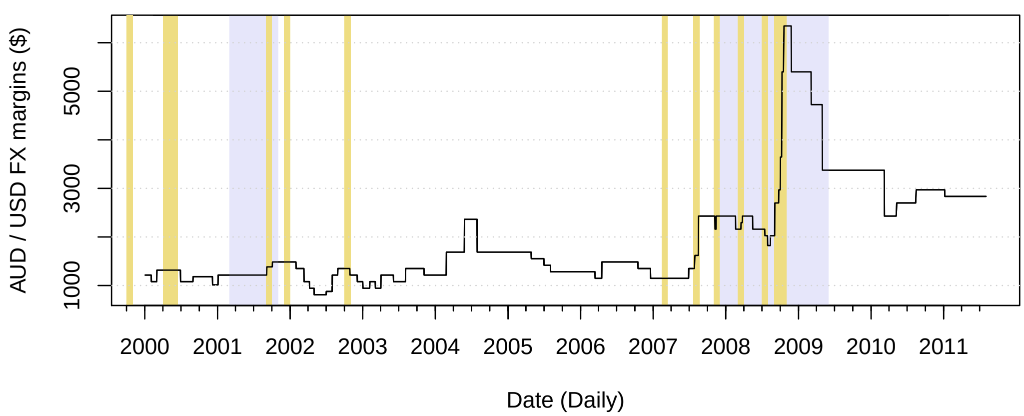


Figure B.4: Margin requirement (in percent) on the AUD / USD FX futures contract.

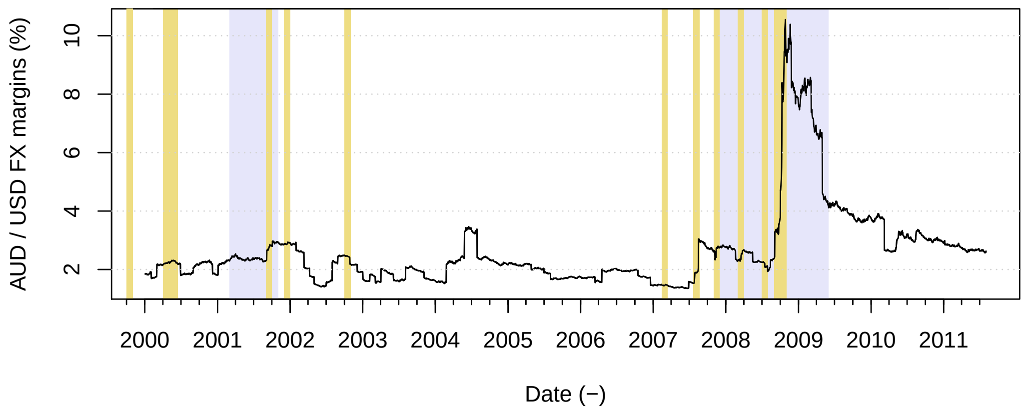


Figure B.5: Margin requirement (in Dollars) on the NZD / USD FX futures contract.

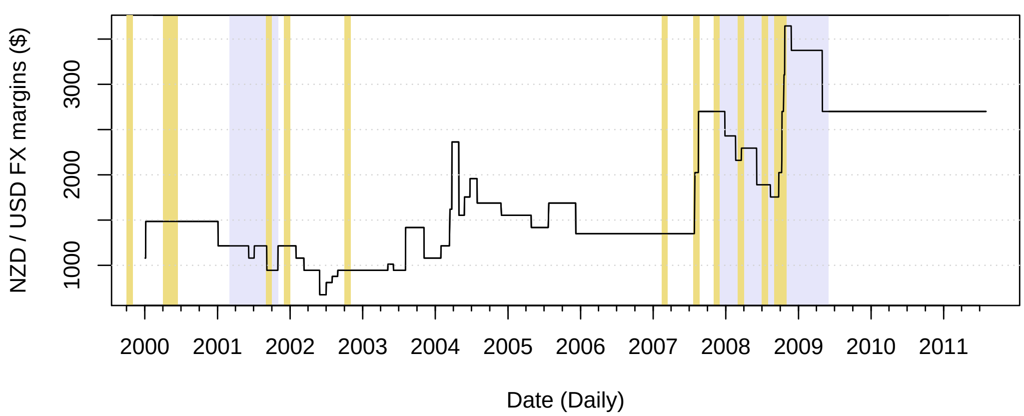
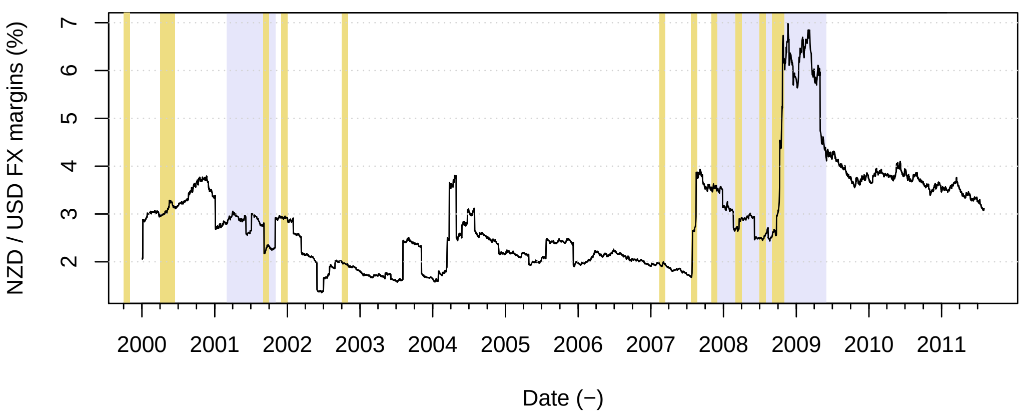


Figure B.6: Margin requirement (in percent) on the NZD / USD FX futures contract.



Appendix C

Event-Studies (cont.)

We perform event studies on margins requirements for the NZD/USD foreign exchange futures contract. We performed similar studies on margin requirements for the S&P 500 Index and the AUD/USD foreign exchange futures contracts in Section 2.6. There are 23 margin increases and 19 margin decreases during the period from January 2000 to June 2011. In Figure C.1 and Figure C.2, we plot the dollar margin requirement and the implied volatility of the NZD/USD foreign exchange, respectively. Comparing those two figures, we note that margin increases coincided with periods of high volatility. In Figure C.3, we plot the average value of the NZD/USD implied volatility surrounding 23 margin increases. We identify a clear trend of increasing volatility prior to a margin increase. The volatility also stays high for almost a month after a margin increase. In Figure C.4, we show that the margin increases following a period of declining market prices. It is interesting to note that market prices continue to decrease, possibly due to forced selling following margin increases. On the other hand, we show that the margin decreases following a period of increasing market prices (see Figure C.5).

Figure C.1: Margin requirement (in Dollar) on the NZD/USD foreign exchange futures contract. Comparing with Figure C.2, we note that the margin increases coincided with periods of high volatility.

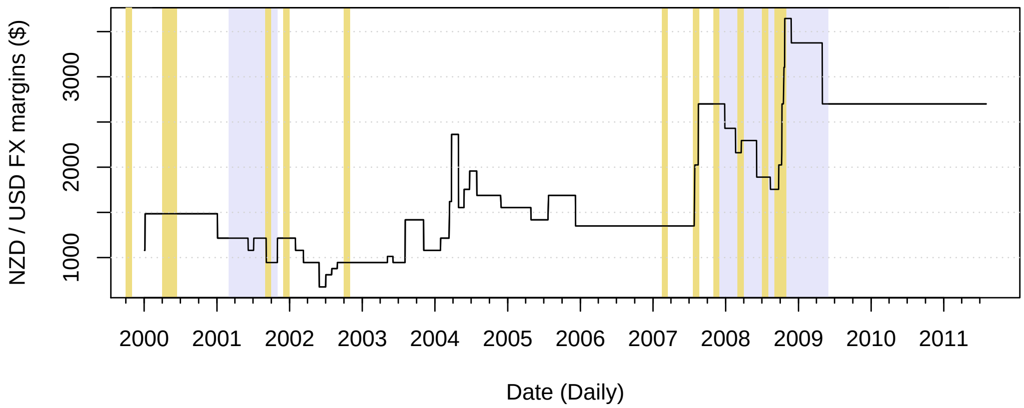


Figure C.2: We plot the implied volatility of the NZD/USD foreign exchange rate.

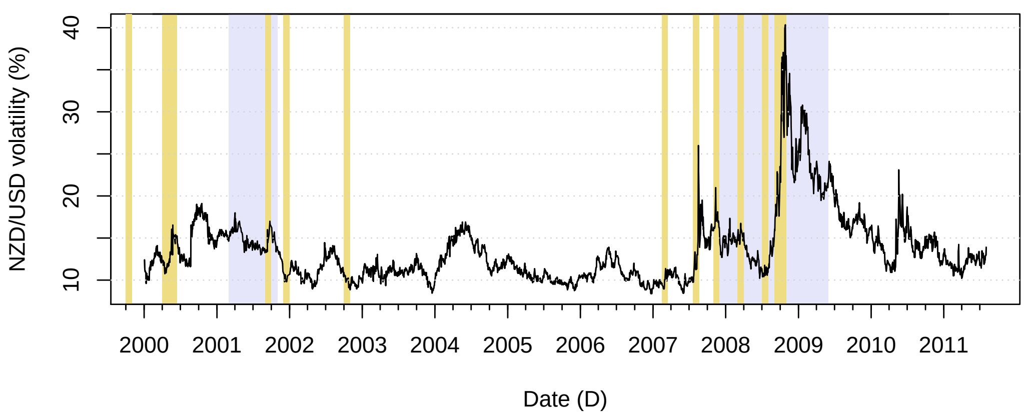


Figure C.3: Event study of *increasing margins*. We plot the average value of the NZD/USD implied volatility surrounding 23 margin increases. We identify a clear trend of increasing volatility prior to a margin increase. The volatility also stays high for almost a month after a margin increase.

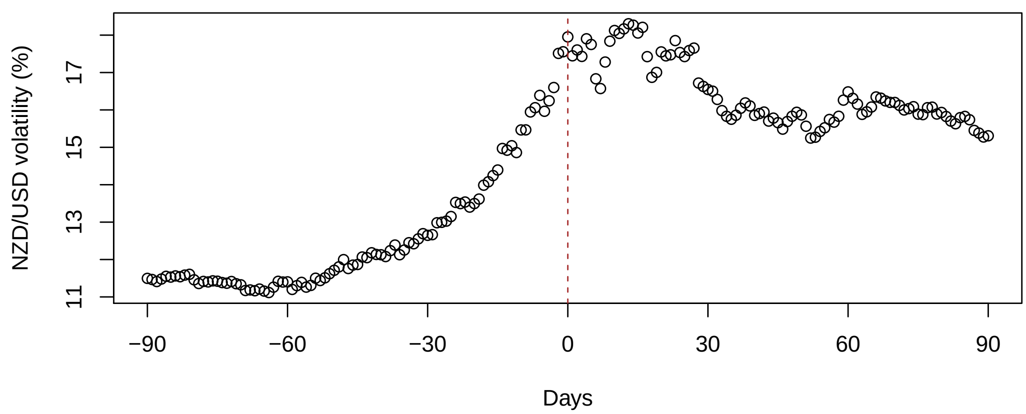


Figure C.4: Event study of *increasing margins*. We plot the average cumulative return of the NZD/USD foreign exchange surrounding 23 margin increases.

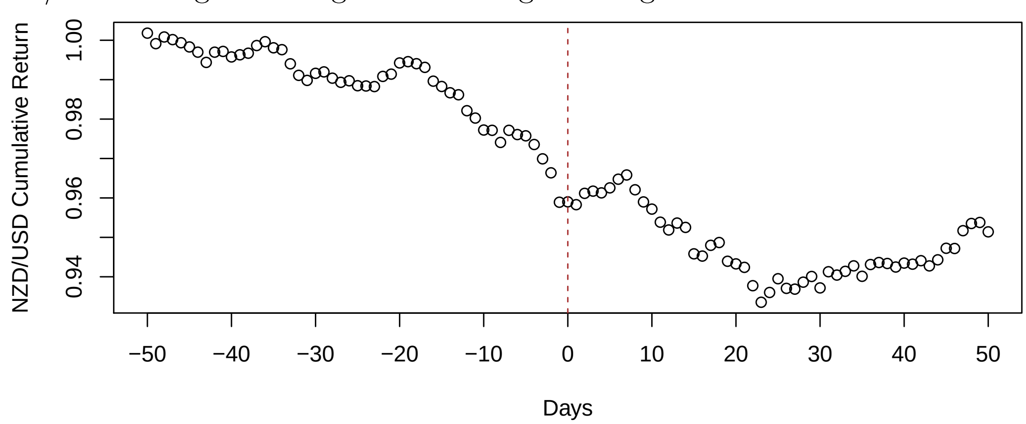
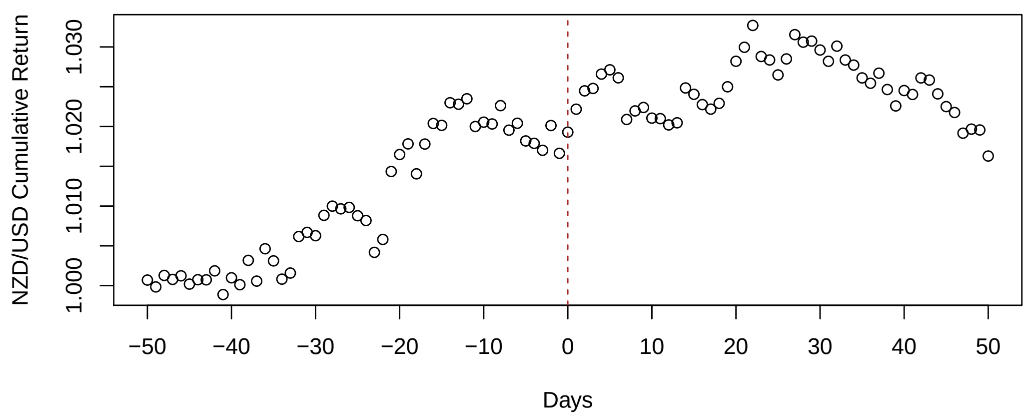


Figure C.5: Event study of *decreasing margins*. We plot the average cumulative return of the NZD/USD foreign exchange surrounding 19 margin decreases.



Appendix D

General Analytical Solution to $p_{T-1}(I_{T-1})$

We re-write the solution in terms of the constant parameters and time-varying state variables included in I_{T-1}. Equation (3.73) becomes

$$y_{T-1}(p, I_{T-1}) = \frac{v_{T-1} - p}{\gamma\left(\overline{\sigma} + \theta\left|\delta_{T-1}\right|\right)^2} - u \tag{D.1}$$

where $\delta_{T-1}, v_{T-1} \in I_{T-1}$and Equation (3.82) becomes

$$x_{T-1}(p, I_{T-1}) \;\; = \begin{cases} \frac{w_{T-2}+x_{T-2}(p-p_{T-2})}{\lambda(\overline{\sigma}+\theta|p-p_{T-2}|)} & \text{if } p < bound_1, \\ \frac{v_{T-1}-p}{\alpha(\overline{\sigma}+\theta|\delta_{T-1}|)^2} & \text{otherwise}, \\ -\frac{w_{T-2}+x_{T-2}(p-p_{T-2})}{\lambda(\overline{\sigma}+\theta|p-p_{T-2}|)} & \text{if } p > bound_2. \end{cases} \tag{D.2}$$

where

$$\begin{aligned}
bound_1 &= \max\left\{\frac{-b_1+\sqrt{b_1^2-4a_1c_1}}{2a_1}, \frac{-b_2+\sqrt{b_2^2-4a_2c_2}}{2a_2} \times \mathrm{I}_{isreal(\frac{-b_2+\sqrt{b_2^2-4a_2c_2}}{2a_2})}\right\} \\
a_1 &= -\lambda\theta \\
b_1 &= \alpha\sigma_T^2 x_{T-2} + \lambda\left(\theta v_{T-1} + \overline{\sigma} + \theta p_{T-2}\right) \\
c_1 &= \alpha\sigma_T^2\left(w_{T-2} - x_{T-2}p_{T-2}\right) - v_{T-1}\lambda\left(\overline{\sigma} + \theta p_{T-2}\right) \\
a_2 &= \lambda\theta \\
b_2 &= \alpha\sigma_T^2 x_{T-2} + \lambda\left(-\theta v_{T-1} + \overline{\sigma} - \theta p_{T-2}\right) \\
c_2 &= \alpha\sigma_T^2\left(w_{T-2} - x_{T-2}p_{T-2}\right) + v_{T-1}\lambda\left(\theta p_{T-2} - \overline{\sigma}\right)
\end{aligned}$$

and

$$bound_2 = \min\left\{\frac{-b_3-\sqrt{b_3^2-4a_3c_3}}{2a_3}, \frac{-b_4-\sqrt{b_4^2-4a_4c_4}}{2a_4}\times \mathrm{I}_{isreal(\frac{-b_4-\sqrt{b_4^2-4a_4c_4}}{2a_4})}\right\}$$

$$\begin{aligned}
a_3 &= -\lambda\theta \\
b_3 &= \alpha\sigma_T^2 x_{T-2} + \lambda\left(-\theta v_{T-1} + \overline{\sigma} + \theta p_{T-2}\right) \\
c_3 &= \alpha\sigma_T^2\left(w_{T-2} - x_{T-2}p_{T-2}\right) + v_{T-1}\lambda\left(\overline{\sigma} - \theta p_{T-2}\right) \\
a_4 &= \lambda\theta \\
b_4 &= \alpha\sigma_T^2 x_{T-2} + \lambda\left(-\theta v_{T-1} - \overline{\sigma} - \theta p_{T-2}\right) \\
c_4 &= \alpha\sigma_T^2\left(w_{T-2} - x_{T-2}p_{T-2}\right) + v_{T-1}\lambda\left(\theta p_{T-2} + \overline{\sigma}\right)
\end{aligned}$$

The solution to Equation (3.55) is to compute the intersections of Equations (D.1) and (D.2) and select the one that is closest to v_{T-1}. It is a bit complicated to express the final step analytically. However, for most realistic cases (i.e., parameters with economic sense), the solution is given by Algorithm (1).

Algorithm 1 Pseudo-code to solve Equation (3.55)

if $bound_1 \leq v_{T-1} - \frac{u\alpha\gamma(\overline{\sigma}+\theta|\delta_{T-1}|)^2}{\alpha+\gamma} \leq bound_2$ **then**
  $p_{T-1}(I_{T-1}) \leftarrow v_{T-1} - \frac{u\alpha\gamma(\overline{\sigma}+\theta|\delta_{T-1}|)^2}{\alpha+\gamma}$
else if $\frac{-b_5-\sqrt{b_5^2-4a_5c_5}}{2a_5} \leq bound_1$ **then**
  $p_{T-1}(I_{T-1}) \leftarrow \frac{-b_5-\sqrt{b_5^2-4a_5c_5}}{2a_5}$
else
  $p_{T-1}(I_{T-1}) \leftarrow \frac{-b_6+\sqrt{b_6^2-4a_6c_6}}{2a_6}$
end if

The new parameters of Algorithm (1) are:

$$\begin{aligned}
a_5 &= -\lambda\theta \\
b_5 &= \gamma\sigma_T^2 x_{T-2} + \lambda\left(-\theta u\gamma\sigma_T^2 + \theta v_{T-1} - \overline{\sigma} + \theta p_{T-2}\right) \\
c_5 &= \gamma\sigma_T^2\left(w_{T-2} - x_{T-2}p_{T-2} + u\lambda\left(-\overline{\sigma} + \theta p_{T-2}\right)\right) + v_{T-1}\lambda\left(\overline{\sigma} - \theta p_{T-2}\right) \\
a_6 &= -\lambda\theta \\
b_6 &= -\gamma\sigma_T^2 x_{T-2} + \lambda\left(-\theta u\gamma\sigma_T^2 + \theta v_{T-1} + \overline{\sigma} + \theta p_{T-2}\right) \\
c_6 &= \gamma\sigma_T^2\left(-w_{T-2} + x_{T-2}p_{T-2} + u\lambda\left(\overline{\sigma} + \theta p_{T-2}\right)\right) - v_{T-1}\lambda\left(\overline{\sigma} - \theta p_{T-2}\right)
\end{aligned}$$

Appendix E

Dimension Reduction

In this appendix, we rewrite our dynamic optimization problem in such a way that $(I^*, \Lambda, x) \in \mathbb{R}^6$ are the only variables that enter the conditioning set in the expectation terms and are the only relevant variables affecting the evaluations of $C_{s,k}$ and $C_{s,h}$ functions. That way, we reduce the number of dimensions n from 7 to 6; we drop the two variables p_s and v_s and add the variable Λ_s.

Define

$$I_s^* = (\delta_s, w_{s-1}, \Lambda_{s-1}, x_{s-1}) \in \mathbb{R}^4 \tag{E.1}$$

For all $t \leq s \leq T$, $\Lambda_s(I_s^*)$ is obtained as follows:

$$\begin{aligned} \Lambda_s : \mathbb{R}^4 &\to \mathbb{R} \\ \Lambda_s(I_s^*) &= \arg_{\Lambda \in \mathbb{R}} \min |\Lambda| \end{aligned} \tag{E.2a}$$

subject to

$$x_s(\Lambda, I_s^*) + y_s(\Lambda, I_s^*) = 0 \qquad \text{if } t \leq s < T \tag{E.2b}$$

At $t = T$,

$$\Lambda_T(I_T) = 0 \qquad \textbf{[base case]} \tag{E.3}$$

where

$$x_s(\Lambda, I_s^*) = \arg_{x \in \mathbb{S}_s} \min C_{s,k}(I_s^*, \Lambda, x) \tag{E.4a}$$

$$y_s(\Lambda, I_s^*) = \arg_{y \in \mathbb{R}} \min C_{s,h}(I_s^*, \Lambda, y) \tag{E.4b}$$

$$\mathbb{S}_{s,I_s^*} = \{r \in \mathbb{R} : |r| \leq l(\Lambda(I_s^*), I_s^*)\} \quad \forall s < T \tag{E.4c}$$

and

$$
\begin{aligned}
k(\Lambda, I_s^*) &= e^{-\alpha x_{s-1}(\delta_s + \Lambda_{s-1} - \Lambda)} && \text{(E.5)} \\
h(\Lambda, I_s^*) &= e^{-\gamma(-x_{s-1}+\mathbf{u})(\delta_s + \Lambda_{s-1} - \Lambda)} && \text{(E.6)} \\
l(\Lambda, I_s^*) &= \frac{w_{s-1} + x_{s-1}\left(\delta_s + \Lambda_{s-1} - \Lambda\right)}{\lambda\left(\overline{\sigma} + \theta\left|\delta_s + \Lambda_{s-1} - \Lambda\right|\right)} && \text{(E.7)} \\
C_{T,k}(I_T^*, \Lambda, x) &= 1 \\
C_{s,k}^*(I_s^*, \Lambda, x) &= \mathrm{E}\left[k(\Lambda_{s+1}, I_{s+1}^*) \times \min_{x'} C_{s+1,k}(I_{s+1}^*, \Lambda_{s+1}, x') \,\middle|\, I_s^*, \Lambda_s = \Lambda, x_s = x\right] \\
C_{T,h}(I_T^*, \Lambda, y) &= 1 \\
C_{s,h}(I_s^*, \Lambda, y) &= \mathrm{E}\left[h(\Lambda_{s+1}, I_{s+1}^*) \times \min_{y'} C_{s+1,h}(I_{s+1}^*, \Lambda_{s+1}, y') \,\middle|\, I_s^*, \Lambda_s = \Lambda, x_s = -y\right]
\end{aligned}
$$

The evolution of the state vector can be expressed using the random function $I_{s+1}^* = M(I_s^*, \Lambda_s, x_s) \quad \forall s < T$.

$$
M: \begin{pmatrix} \delta_s \\ w_{s-1} \\ \Lambda_{s-1} \\ x_{s-1} \end{pmatrix} \in \mathbb{R}^4, \Lambda_s \in \mathbb{R}, x_s \in \mathbb{R} \rightarrow \begin{pmatrix} \delta_{s+1} \\ w_{s-1} + x_{s-1}\left(\delta_s + \Lambda_{s-1} - \Lambda_s\right) \\ \Lambda_s \\ x_s \end{pmatrix} \in \mathbb{R}^4 \qquad \text{(E.8)}
$$

Appendix F

Numerical Integration

In this appendix, we compare two numerical integration methods to select the one that suits our particular case better.

Objective

The objective is to numerically estimate the value of $\mathrm{E}[f(X)]$. It is equivalent to computing the following integration:

$$\mathrm{E}[f(X)] = \int_{-\infty}^{+\infty} f(x) \cdot h_X(x)\, dx \tag{F.1}$$

where $h_X(x)$ is the density function of X. In particular, we consider $X \sim N(\mu, \sigma^2)$. The corresponding density function is

$$h_{X;\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{F.2}$$

Equation (F.1) becomes

$$\mathrm{E}[f(X)] = \int_{-\infty}^{+\infty} f(x) \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx \tag{F.3}$$

Gauss-Hermite Quadrature Method

Gauss-Hermite quadrature method is used to approximate the value of integrals of the form:

$$\int_{-\infty}^{+\infty} g(x) \cdot e^{-x^2}\, dx \approx \sum_{i=1}^{n} w_i \cdot g(x_i) \tag{F.4}$$

where x_i are the roots of the Hermite polynomial $H_n(x) \quad (i = 1, 2, 3..., n)$ and $w_i = \frac{2^{n-1} n! \sqrt{pi}}{n^2 [H_{n-1}(x_i)]^2}$ are the weights.

To estimate the integral in Equation (F.3) using the Gauss-Hermite quadrature method, we need to change the variable. Let $y = \frac{x-\mu}{\sqrt{2}\sigma}$, Equation (F.3) becomes:

$$\begin{aligned} \mathrm{E}\left[f(X)\right] &= \int_{-\infty}^{+\infty} f(\sqrt{2}\sigma y + \mu) \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-y^2}\ \sqrt{2}\sigma dy \\ &= \int_{-\infty}^{+\infty} g(y) \cdot e^{-y^2}\ dy \\ &\approx \sum_{i=1}^{n} w_i \cdot g(x_i) \end{aligned} \tag{F.5}$$

where

$$\begin{aligned} g(x) &= f(\sqrt{2}\sigma x +\ mu) \cdot \frac{1}{\sqrt{\pi}} \\ x_i &\text{ are the roots of } H_n(x) \\ w_i &= \frac{2^{n-1} n! \sqrt{pi}}{n^2 \left[H_{n-1}(x_i)\right]^2} \end{aligned} \tag{F.6}$$

The Gauss-Hermite quadrature method, like any other Gauss quadrature method, will only produce accurate results if the function $g(x)$ can be closely approximated by a polynomial function. Furthermore, the method is not appropriate for functions with discontinuities. We will show later that our function $g(x)$ is exponential and therefore can not be approximately well with a finite order polynomial function. In certain situations, $g(x)$ has a point of discontinuity making it even more difficult to obtain accurate results with the Gauss-Hermite quadrature method. The method does however yield good approximations, but we were able to obtain more accurate results using quasi-Monte Carlo with Importance Sampling.

Quasi-Monte Carlo with Importance Sampling

If we take a sample of X, $(x_1, x_2, ..., x_n)$, and compute the mean of $f(x)$ over the sample, we obtain the Monte Carlo estimate of $\mathrm{E}[f(X)]$:

$$\mathrm{E}[f(X)] \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i) \tag{F.7}$$

To obtain a sample $(x_1, x_2, ..., x_n)$ that is normally distributed, we first generate a sample $(u_1, u_2, ..., u_n)$ distributed uniformly in the interval $(0, 1)$ and then compute $x_i = \Phi(u_i)$ where $\Phi(\cdot)$ is the inverse cumulative distribution function (CDF) of X.

Instead of generating the sample $(u_1, u_2, ..., u_n)$ randomly using a random number generator, we can construct the sample using the rectangle rule where the n points are chosen as $u_i = \frac{i}{n+1}$ in the interval $(0, 1)$. The rectangle rule is a simple low-discrepancy sequence. The advantage of using a low-discrepancy sequence instead of a sample generated randomly is that, as we increase the size of the sample, we obtain a fast convergence rate and we improve the results more systematically.

If certain values of X have more impact on the function $f(x)$ than others, sampling poorly from the distribution of X can lead to bad estimates of $\mathrm{E}[f(X)]$. One way to reduce the estimation error is to increase the sample size. However, in certain situations, computing $f(x)$ is computationally intensive and it is therefore important to sample efficiently from the distribution of X.

Importance Sampling can achieve high efficiency because it does not sample the random variable from its original distribution, but uses a more appropriate sampling distribution that samples more frequently from "important" regions. $f(x)$ is then weighted accordingly to correct for the use of a sampling distribution different from the original distribution. Recall that $h_X(x)$ is the original density function. Let $q(x)$ be the sampling density function. We can write $\mathrm{E}[f(X)]$ as follows:

$$\begin{aligned}
\mathrm{E}[f(X)] &= \int_{-\infty}^{+\infty} f(x) \cdot h_X(x)\, dx \\
&= \int_{-\infty}^{+\infty} \frac{f(x) \cdot h_X(x)}{q(x)} \cdot q(x)\, dx && \text{(F.8)} \\
&= \mathrm{E}_q\left[\frac{f(x) h_X(x)}{q(x)}\right] && \text{(F.9)}
\end{aligned}$$

where E_q denotes expectation with respect to the density function q. The quasi-Monte Carlo estimator becomes

$$\mathrm{E}\left[f(X)\right] \approx \frac{1}{n}\sum_{i=1}^{n} f(y_i)\cdot\frac{h_X(y_i)}{q(y_i)} \qquad y \sim q(y) \tag{F.10}$$

$$\approx \sum_{i=1}^{n} w_i \cdot f(y_i) \tag{F.11}$$

where

$$w_i = \frac{h_X(y_i)}{n \cdot q(y_i)} \tag{F.12}$$

Note the similarity between Equation (F.12) and Equation (F.5). Equation (F.5) estimates $\mathrm{E}\left[f(X)\right]$ using Gauss-Hermite quadrature of order n while Equation (F.12) estimates $\mathrm{E}\left[f(X)\right]$ using quasi-Monte Carlo with Importance Sampling. The flexibility to select the sampling distribution that suits our problem best is a major advantage of Importance Sampling. In the next couple of sections, we compare these two methods to see which one is more suitable to our problem. It is however very important to use the same n (i.e., Gauss-Hermite of order n for the first method, and a quasi-Monte Carlo with sample size n for the second method) because we need to keep the number of function evaluations $f(x)$ to a minimum.

Example: Investor's Demand Function at Time $t = T-1$

We compute numerically the investor's demand function at time $t = T - 1$ using various methods. The demand function is defined as follow: at each price p the investor selects the number of shares y_{T-1} that he would like to hold. Since the demand function of the investor at time $T - 1$ has an analytical solution, we can assess the accuracy of the various numerical methods. At $t = T - 1$, the problem is simple:

$$y_{T-1}(p) \;=\; \arg_{y\in\mathbb{R}}\, \min \mathrm{E}\left[e^{-0.02(y+10)(100+\delta_T - p)}\middle|p, y\right] \tag{F.13}$$

where

$$\delta_T \sim N\left(0, 5.2\right) \tag{F.14}$$

Table F.1: Gauss-Hermite Quadrature Numerical Integration. We compute the mean squared difference between the analytical and numerically estimated functions.

	$n = 8$	$n = 10$	$n = 12$	$n = 16$	$n = 20$
Mean Squared of Differences	63.04	55.98	49.65	37.76	25.43

The analytical solution is easy to derive since δ_T, which is normally distributed $N(0, 5.2)$, is the only random variable in the expectation term and we are conditioning on the other two variables. The analytical solution is:

$$y_{T-1}(p) = \frac{100 - p}{0.02 \times 5.2^2} - 10 \tag{F.15}$$

There are many numerical methods to estimate the expectation term. Gauss-Hermite quadrature, presented in Section F, is one method. Quasi-Monte Carlo with Importance Sampling, presented in Section F, is another. Since we know the analytical solution, we can compute the deviation of the numerical estimates from y_{T-1}. With respect to the entire demand function $y_{T-1}(p)$, we proceed as follows: 1) we discretize p, 2) at each price p, we compute the squared difference between the analytical solution and the numerical estimate, and 3) we take the square root of the mean of those squared differences. In particular, we discretize p in the interval $[55, 145]$ in increments of 0.25 (i.e., 360 points in total). We compute the deviation as follows:

$$\text{Mean Squared of Differences} = \sum_{i=1}^{360} \left(\tilde{y}_i - y_{T-1}(p_i)\right)^2 \tag{F.16}$$

where $\tilde{y}_i$ is the numerical estimate of $y_{T-1}(p_i)$.

Comparing the Two Numerical Methods

In Figure F.1, the dashed red line corresponds to the analytical solution of the demand function. The solid lines correspond to numerical estimates of the demand function using the Gauss-Hermite quadrature method. We estimated the demand function five times, using different orders of the Hermite polynonial each time. As we increased the order, going from $n = 8$ to $n = 20$, the accuracy of our numerical estimates improved.

In Figure F.2, the solid lines correspond to numerical estimates using quasi-Monte Carlo with Importance Sampling. Each expectation is numerically estimated using a sample size

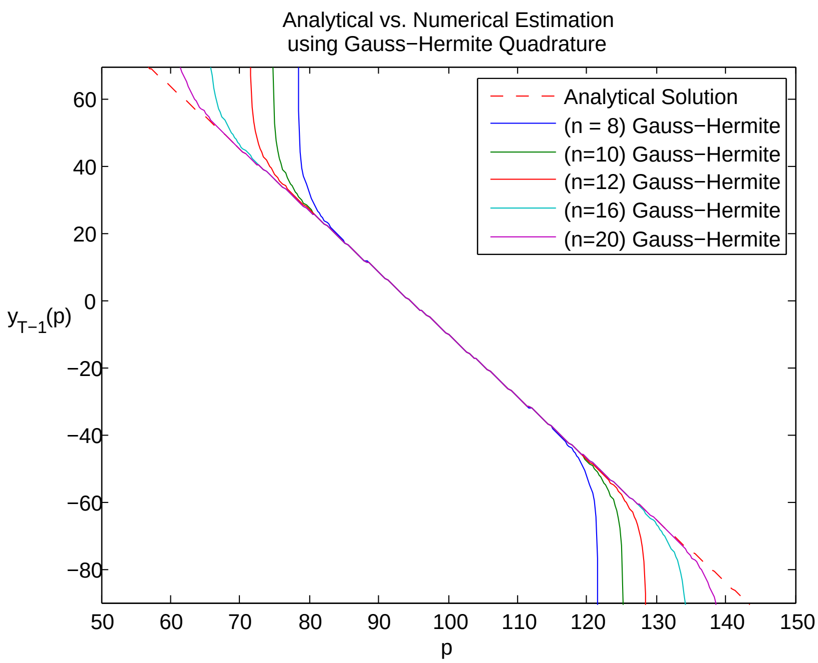


Figure F.1: Gauss-Hermite Quadrature Numerical Integration. We compute the investor's demand function. At each price p, we compute the demand $y_{T-1}(p)$ according to Equation (F.13). The expectation term in Equation (F.13) is numerically estimated using Gauss-Hermite quadrature. We observe that as we increase the order of the Hermite polynomial, going from $n = 8$ to $n = 20$, we improve the accuracy of our numerical estimates.

Table F.2: Importance Sampling quasi-Monte Carlo method. We compute the mean squared difference between the analytical and numerically estimated functions.

	2×	3×	4×	5×	6×
Mean Squared of Differences	71.24	54.58	37.17	13.61	0.18

of 20 observations. The various numerical estimates in Figure F.2 correspond to using different sampling distributions. The sampling distributions are similar to the original distribution but with a standard deviation a multiple of the original standard deviation, σ. We consider five cases ranging from a sampling standard deviation of 2σ to 6σ. We observe that using a sampling standard deviation of 5σ or 6σ, we can obtain very close numerical estimates. By selecting a sampling distribution with a larger standard deviation than the original distribution, we are able to sample better from the tails where the integrand $f(x)$ takes large values.

In Figure F.3, we compare the two numerical methods using in each case a sample size of 20. That is, each requires computing[1] the integrand $f(x)$ 20 times. In particular, we compare

1. the Gauss-Hermite quadrature of order $n = 20$, and
2. quasi-Monte Carlo with Importance Sampling using a sample size of $n = 20$ and a sampling distribution with 5σ.

We observe that the Importance Sampling method leads to more accurate results because it samples better from the tails of the original distribution, which is important since the integrand takes large values at the tails. Note that in certain situations in our general problem, the integrand has a point of discontinuity. Importance Sampling can be particularly useful in that situation since we can adjust the mean of the sampling distribution around the point of discontinuity and sample more from that area. Since the integrand is lognormal at $t = T - 1$, there might be a better quadrature technique in this case. However, we are more concerned about $t < T - 1$ when the integrand is not lognormal and its distribution is unknown. Therefore a more flexible technique such as quasi-Monte Carlo with Importance Sampling might be more appropriate.

[1] At times $t < T - 1$, computing the integrand becomes computationally intensive. We therefore need to obtain a numerical estimate with as few function evaluations $f(x)$ as possible.

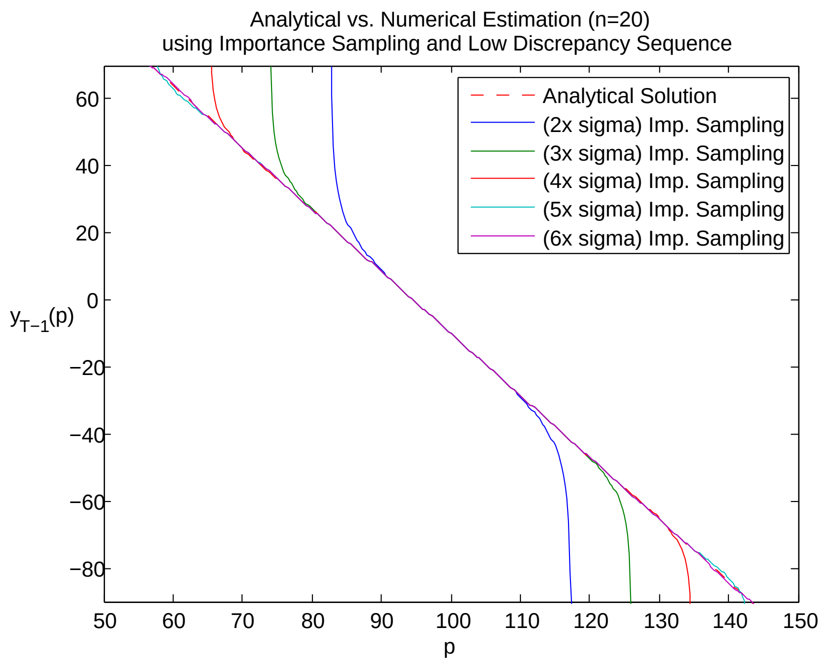


Figure F.2: Importance Sampling quasi-Monte Carlo method. We compute the investor's demand function. At each price p, we compute the demand $y_{T-1}(p)$ according to Equation (F.13). The expectation term in Equation (F.13) is numerically estimated using quasi-Monte Carlo with Importance Sampling. Each expectation is numerically estimated using a sample size of 20 observations. The sampling distributions are similar to the original distribution but with a standard deviation a multiple of the original standard deviation, σ. We consider five cases ranging from a sampling standard deviation of 2σ to 6σ. We observe that using a sampling standard deviation of 5σ or 6σ, we can obtain very close numerical estimates.

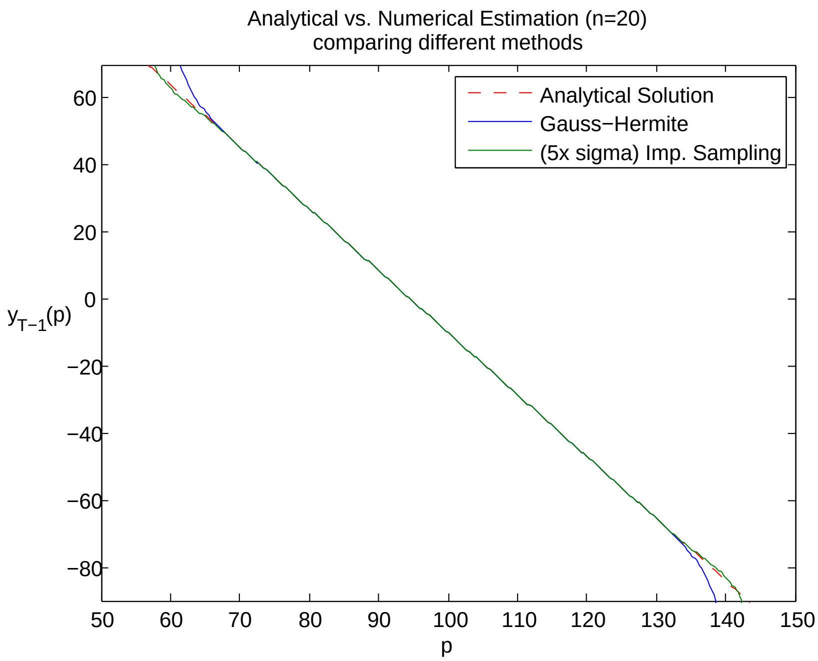


Figure F.3: We compare the Gauss-Hermite quadrature of order $n = 20$ against quasi-Monte Carlo with Importance Sampling on a sample size of $n = 20$ and sampling standard deviation of 5σ. Both methods require an equal amount of function evaluation, $f(x)$. We observe that the Importance Sampling method leads to more accurate results.

References

[AB02] Andrew Ang and Geert Bekaert, *International asset allocation with regime shifts*, Review of Financial Studies **15** (2002), no. 4, 1137–1187.

[AB08] Tobias Adrian and Markus Brunnermeier, *Covar*, Tech. report, 2008.

[Aga04] Vikas Agarwal, *Risks and portfolio decisions involving hedge funds*, Review of Financial Studies **17** (2004), no. 1, 63–98.

[AM07] Victor Aguirregabiria and Pedro Mira, *Dynamic discrete choice structural models: A survey*, Working Papers tecipa-297, University of Toronto, Department of Economics, July 2007.

[AS10] Tobias Adrian and Hyun Song Shin, *Liquidity and leverage*, Journal of Financial Intermediation **19** (2010), no. 3, 418–437.

[BB10] Scott Brave and R. Andrew Butters, *Gathering insights on the forest from the trees: a new metric for financial conditions*, Tech. report, 2010.

[BB11] ———, *Monitoring financial stability: a financial conditions index approach*, Economic Perspectives (2011), no. Q I, 22–43.

[BCG+09] Markus Brunnermeier, Andrew Crockett, Charles Goodhart, Avi Persaud, and Hyun Shin, *The fundamental principles of financial regulation*, International Center for Monetary and Banking Studies Centre for Economic Policy Research, Geneva London, 2009.

[Bel57] Richard E. Bellman, *Dynamic programming*, Princeton University Press, Princeton, NJ, 1957.

[BG95] Ben Bernanke and Mark Gertler, *Inside the black box: The credit channel of monetary policy transmission*, Journal of Economic Perspectives **9** (1995), no. 4, 27–48.

[BGP07] Markus Brunnermeier, Christian Gollier, and Jonathan A. Parker, *Optimal beliefs, asset prices, and the preference for skewed returns*, American Economic Review **97** (2007), no. 2, 159–165.

[BGP08] Monica Billio, Mila Getmansky, and Loriana Pelizzon, *Crises and hedge fund risk*, Yale School of Management Working Papers amz2561, Yale School of Management, May 2008.

[BH07] Nicholas Barberis and Ming Huang, *Stocks as lotteries: The implications of probability weighting for security prices*, NBER Working Papers 12936, National Bureau of Economic Research, Inc, February 2007.

[BHN05] Geert Bekaert, Campbell R. Harvey, and Angela Ng, *Market integration and contagion*, Journal of Business **78** (2005), no. 1, 39–70.

[BKS03] Kee-Hong Bae, G. Andrew Karolyi, and Rene Stulz, *A new approach to measuring financial contagion*, Review of Financial Studies **16** (2003), no. 3, 717–763.

[BM85] I.F. Boesky and J.G. Madrick, *Merger mania: arbitrage, wall street's best kept money-making secret*, Holt, Rinehart and Winston, 1985.

[BNP09] Markus Brunnermeier, Stefan Nagel, and Lasse Pedersen, *Carry trades and currency crashes*, NBER Macroeconomics Annual 2008, Volume 23, NBER Chapters, National Bureau of Economic Research, Inc, Summer 2009, pp. 313–347.

[BO10] Markus Brunnermeier and Martin Oehmke, *The maturity rat race*, NBER Working Papers 16607, National Bureau of Economic Research, Inc, December 2010.

[BP09] Markus Brunnermeier and Lasse Pedersen, *Market liquidity and funding liquidity*, Review of Financial Studies **22** (2009), no. 6, 2201–2238.

[BR86] Keith Brown and Michael Raymond, *Risk arbitrage and the prediction of successful corporate takeovers*, Financial Management **15** (1986), 54–63.

[Bru09] Markus Brunnermeier, *Deciphering the liquidity and credit crunch 2007-2008*, Journal of Economic Perspectives **23** (2009), no. 1, 77–100.

[BS02] Malcolm Baker and Serkan Savasoglu, *Limited arbitrage in mergers and acquisitions*, Journal of Financial Economics **64** (2002), no. 1, 91–115.

[BS09] Markus Brunnermeier and Yuliy Sannikov, *A macroeconomic model with a financial sector.*, Working paper, Princeton University, 2009.

[BSS08] Nicole Boyson, Christof Stahel, and Rene Stulz, *Is there hedge fund contagion*, Working Papers 08-2, University of Pennsylvania, Wharton School, Weiss Center, March 2008.

[BSS10] Nicole M. Boyson, Christof W. Stahel, and Rene Stulz, *Hedge fund contagion and liquidity shocks*, Journal of Finance **65** (2010), no. 5, 1789–1816.

[CT03] John Y. Campbell and Glen B. Taksler, *Equity volatility and corporate bond yields*, Journal of Finance **58** (2003), no. 6, 2321–2350.

[DFM92] William P Dukes, Cheryl J Frohlich, and Christopher K Ma, *Risk arbitrage in tender offers*, no. 4, 47–55.

[DG08] Ferre De Graeve, *The external finance premium and the macroeconomy: US post-WWII evidence*, Journal of Economic Dynamics and Control **32** (2008), no. 11, 3415–3440.

[DLSSW90] J Bradford De Long, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, *Noise trader risk in financial markets*, Journal of Political Economy **98** (1990), no. 4, 703–38.

[DN10a] Evan Dudley and Mahendrarajah Nimalendran, *Funding Risk and Hedge Fund Returns*, SSRN eLibrary (2010).

[DN10b] ———, *Margins and Hedge Fund Contagion*, SSRN eLibrary (2010).

[DNFL11] Jens Dick-Nielsen, Peter Feldhtter, and David Lando, *Corporate Bond Liquidity Before and After the Onset of the Subprime Crisis*, SSRN eLibrary (2011).

[Eng96] Charles Engel, *The forward discount anomaly and the risk premium: A survey of recent evidence*, Journal of Empirical Finance **3** (1996), no. 2, 123–192.

[EW89] Zvi Eckstein and Kenneth I. Wolpin, *The specification and estimation of dynamic stochastic discrete choice models: A survey*, Journal of Human Resources **24** (1989), no. 4, 562–598.

[FF93] Eugene F. Fama and Kenneth R. French, *Common risk factors in the returns on stocks and bonds*, Journal of Financial Economics **33** (1993), no. 1, 3–56.

[FG08a] Emmanuel Farhi and Xavier Gabaix, *Rare disasters and exchange rates*, NBER Working Papers 13805, National Bureau of Economic Research, Inc, February 2008.

[FG08b] Ana Fostel and John Geanakoplos, *Leverage cycles and the anxious economy*, American Economic Review **98** (2008), no. 4, 1211–44.

[FG09] Jean-Sebastien Fontaine and Rene Garcia, *Bond liquidity premia*, Working papers, Bank of Canada, 2009.

[FH04] William (Bill) Fung and David A. Hsieh, *Hedge Fund Benchmarks: A Risk-Based Approach*, Financial Analysts Journal **60** (2004), no. 5, 65–80.

[Flu88] B. Flury, *Common principal components and related multivariate models*, Wiley series in probability and mathematical statistics, Wiley, 1988.

[FOM98] FOMC, *Minutes of the federal open market committee*, Federal Reserve, October 15 1998, http://www.federalreserve.gov/monetarypolicy/files/FOMC19981015confcall.pdf.

[FT90] Kenneth Froot and Richard Thaler, *Anomalies: Foreign exchange*, The Journal of Economic Perspectives **4** (1990), no. 3, 179–192.

[Gea09] John Geanakoplos, *The leverage cycle*, Cowles Foundation Discussion Papers 1715, Cowles Foundation for Research in Economics, Yale University, July 2009.

[GI89] Peter W. Glynn and Donald L. Iglehart, *Importance sampling for stochastic simulations*, Management Science **35** (1989), no. 11, 1367–1392.

[GK01] John Geweke and Michael Keane, *Computationally intensive methods for integration in econometrics*, Handbook of Econometrics (J.J. Heckman and E.E. Leamer, eds.), Handbook of Econometrics, vol. 5, Elsevier, 2001, pp. 3463–3568.

[GL99] Mark Gertler and Cara Lown, *The information in the high-yield bond spread for the business cycle: Evidence and some implications*, Oxford Review of Economic Policy **15** (1999), no. 3, 132–150.

[GM88] Sanford J Grossman and Merton H Miller, *Liquidity and market structure*, Journal of Finance **43** (1988), no. 3, 617–37.

[Gor08] Gary B. Gorton, *The panic of 2007*, Proceedings Payments System Research Conferences (2008), 131–262.

[GQ73] Stephen M. Goldfeld and Richard E. Quandt, *A markov model for switching regressions*, Journal of Econometrics **1** (1973), no. 1, 3–15.

[Gri02] Mark Grinblatt, *An analytic solution for interest rate swap spreads*, Yale School of Management Working Papers ysm39, Yale School of Management, January 2002.

[GS00] Anurag Gupta and Marti G. Subrahmanyam, *An empirical examination of the convexity bias in the pricing of interest rate swaps*, Journal of Financial Economics **55** (2000), no. 2, 239–279.

[GSW07] Refet S. Gurkaynak, Brian Sack, and Jonathan H. Wright, *The U.S. treasury yield curve: 1961 to the present*, Journal of Monetary Economics **54** (2007), no. 8, 2291–2304.

[GT] Gordon Gemmill and Dylan C. Thomas, *Noise Trading, Costly Arbitrage, and Asset Prices: Evidence from Closed-end Funds*, The Journal of Finance, Vol. 57, pp. 2571-2594, 2002.

[GV02] Denis Gromb and Dimitri Vayanos, *Equilibrium and welfare in markets with financially constrained arbitrageurs*, Journal of Financial Economics **66** (2002), no. 2-3, 361–407.

[GV09] ———, *Leverage and liquidity dry-ups: a framework and policy implications*, Working paper, INSEAD, 2009.

[Ham89] James D Hamilton, *A new approach to the economic analysis of nonstationary time series and the business cycle*, Econometrica **57** (1989), no. 2, 357–384.

[Ham90] James D. Hamilton, *Analysis of time series subject to changes in regime*, Journal of Econometrics **45** (1990), no. 1-2, 39–70.

[Har01] Mary R Hardy, *A regime-switching model of long-term stock returns*, North American Actuarial Journal **5** (2001), no. 2, 4153.

[HH09] Mamiza Haq and Richard Heaney, *European bank equity risk: 1995-2006*, Journal of International Financial Markets, Institutions and Money **19** (2009), no. 2, 274–288.

[HK96] Jean Helwege and Paul Kleiman, *Understanding aggregate default rates of high yield bonds*, Current Issues in Economics and Finance (1996), no. May.

[HK09] Craig S. Hakkio and William R. Keeton, *Financial stress: what is it, how can it be measured, and why does it matter?*, Economic Review (2009), 5–50.

[Kea96] Frank Keane, *Repo rate patterns for new treasury notes*, Current Issues in Economics and Finance (1996).

[KL11] Amir E. Khandani and Andrew W. Lo, *What happened to the quants in august 2007? evidence from factors and transactions data*, Journal of Financial Markets **14** (2011), no. 1, 1–46.

[KLPR10] Mark Kritzman, Yuanzhen Li, Sebastien Page, and Roberto Rigobon, *Principal Components as a Measure of Systemic Risk*, SSRN eLibrary (2010).

[KSoB98] G.A. Karolyi, J. Shannon, and Richard Ivey School of Business, *Where's the risk in risk arbitrage?*, Working Paper Series, University of Western Ontario, 1998.

[KW94] Michael P. Keane and Kenneth I. Wolpin, *The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte carlo evidence*, Tech. report, Federal Reserve Bank of Minneapolis, 1994.

[Kwa01] Simon Kwan, *Rising junk bond yields: liquidity or credit concerns?*, FRBSF Economic Letter (2001).

[Lew95] Karen K. Lewis, *Puzzles in international financial markets*, Handbook of International Economics (G. M. Grossman and K. Rogoff, eds.), Handbook of International Economics, vol. 3, Elsevier, 1995, pp. 1913–1971.

[LL04] Jun Liu and Francis Longstaff, *Losing money on arbitrage: Optimal dynamic portfolio choice in markets with arbitrage opportunities*, Review of Financial Studies **17** (2004), no. 3, 611–641.

[LMN05] Francis A. Longstaff, Sanjay Mithal, and Eric Neis, *Corporate yield spreads: Default risk or liquidity? new evidence from the credit default swap market*, Journal of Finance **60** (2005), 2213–2253.

[LST91] Charles M C Lee, Andrei Shleifer, and Richard H Thaler, *Investor sentiment and the closed-end fund puzzle*, Journal of Finance **46** (1991), no. 1, 75–109.

[MAE11] Tyler Muir, Tobias Adrian, and Erkko Etula, *Financial Intermediaries and the Cross-Section of Asset Returns*, SSRN eLibrary (2011).

[Mal77] Burton G Malkiel, *The valuation of closed-end investment-company shares*, Journal of Finance **32** (1977), no. 3, 847–59.

[Mil99] Michael Milken, *Prosperity and social capital*, The Wall Street Journal, June 23 1999, http://www.npac.syr.edu/projects/cpsedu/summer98summary/examples/hpf/hpf.html.

[Mit01] Mark Mitchell, *Characteristics of risk and return in risk arbitrage*, Journal of Finance **56** (2001), no. 6, 2135–2175.

[MPP07] Mark Mitchell, Lasse Pedersen, and Todd Pulvino, *Slow moving capital*, American Economic Review **97** (2007), no. 2, 215–220.

[MSW08] James McAndrews, Asani Sarkar, and Zhenyu Wang, *The effect of the Term Auction Facility on the London Inter-Bank Offered Rate*, Tech. report, 2008.

[MU08] Franois-Louis Michaud and Christian Upper, *What drives interbank rates? evidence from the libor panel*, BIS Quarterly Review (2008).

[Pak91] Ariel Pakes, *Dynamic structural models: Problems and prospects. mixed continuous discrete controls and market interactions*, Cowles Foundation Discussion Papers 984, Cowles Foundation for Research in Economics, Yale University, July 1991.

[Ped09] Lasse Pedersen, *When everyone runs for the exit*, International Journal of Central Banking **5** (2009), no. 4, 177–199.

[PL11] Momtchil Pojarliev and Richard M Levich, *Detecting crowded trades in currency funds*, Financial Analysts Journal **67** (2011), no. 1, 1–14.

[Pon96] Jeffrey Pontiff, *Costly arbitrage: Evidence from closed-end funds*, The Quarterly Journal of Economics **111** (1996), no. 4, 1135–51.

[Qua58] Richard E. Quandt, *The estimation of the parameters of a linear regression system obeying two separate regimes*, Journal of the American Statistical Association **53** (1958), no. 284, 873–880.

[Rus91] John Rust, *Estimation of dynamic structural models: Problems and prospects part i: Discrete decision processes*, Working papers, Wisconsin Madison - Social Systems, 1991.

[Rus97] John Rust, *A comparison of policy iteration methods for solving continuous-state, infinite-horizon markovian decision problems using random, quasi-random, and deterministic discretizations*, Computational Economics 9704001, EconWPA, April 1997.

[SA00] Kevin P. Scherer and Marco Avellaneda, *All For One... One For All? A Principal Component Analysis of the Latin American Brady Bond Debt from 1994 to 2000*, SSRN eLibrary (2000).

[Sam58] Paul A. Samuelson, *An exact consumption-loan model of interest with or without the social contrivance of money*, Journal of Political Economy **66** (1958), 467.

[SE04] Brian Sack and Robert Elsasser, *Treasury inflation-indexed debt: a review of the U.S. experience*, Economic Policy Review (2004), no. May, 47–63.

[Sen11] Henny Sender, *Banks lose out as Washington rigs the rules*, The Financial Times, June 24 2011, http://www.ft.com/intl/cms/s/0/edd9f86a-9e68-11e0-8e61-00144feabdc0.html#axzz1QK9uh0AO.

[SL10] Federal Reserve Bank of St. Louis, *Minutes of the federal open market committee*, St. Louis Fed Research, January 2010, http://research.stlouisfed.org/publications/net/NETJan2010Appendix.pdf.

[Smi84] Michael Smirlock, *An analysis of bank risk and deposit rate ceilings : Evidence from the capital markets*, Journal of Monetary Economics **13** (1984), no. 2, 195–210.

[SR86] William Samuelson and Leonard Rosenthal, *Price movements as indicators of tender offer success*, Journal of Finance **41** (1986), no. 2, 481–99.

[SV97] Andrei Shleifer and Robert W Vishny, *The limits of arbitrage*, Journal of Finance **52** (1997), no. 1, 35–55.

[SVA98] Manuel S. Santos and Jesus Vigo-Aguiar, *Analysis of a numerical dynamic programming algorithm applied to economic models*, Econometrica **66** (1998), no. 2, 409–426.

[VW08] Dimitri Vayanos and Pierre-Olivier Weill, *A search-based theory of the on-the-run phenomenon*, Journal of Finance **63** (2008), no. 3, 1361–1398.

[War92] Arthur Warga, *Bond returns, liquidity, and missing data*, Journal of Financial and Quantitative Analysis **27** (1992), no. 04, 605–617.

[WP82] Guy P. Wyser-Pratte, *Risk arbitrage ii*, Salomon Brothers Center for the Study of Financial Institutions, Graduate School of Business Administration, New York University, [New York, N.Y.], 1982.