Three Essays on Nonlinear Limits to Arbitrage

Jingzhi Chen

Doctor of Philosophy

University of York

Economics

September 2017

Abstract

Arbitrage costs and funding constraints are two major frictions that limit arbitrage. Arbitrage costs, such as learning costs, transaction costs and holding costs, render arbitrageurs unwilling to take on positions, whilst funding constraints, including equity and leverage constraints, reduce their ability to obtain funding and thus correct mispricings. This thesis contains theoretical studies and empirical applications of these two frictions, both individually and jointly. First we investigate the combined impact of arbitrage costs and funding constraints on the arbitrage activity, where we reveal the nonlinearity of limits to arbitrage: when funding constraint is not binding, arbitrage costs serve as the dominating friction, and thus the arbitrage activity increases with mispricing; however, in extreme situations where funding constraints become binding and establish dominance over arbitrage costs, the arbitrage activity tends to decline with larger mispricing. Second we narrow our focus on the time-varying leverage constraint, and construct a funding liquidity measure via the efficacy of arbitrage. The measure ex ante identifies four periods of binding funding constraint: the collapse of dot-com bubble, the financial crisis in 2007-2008 and the two debt ceiling crises in 2011 and 2013, which supports the slow-moving capital hypothesis. The measure also predicts the market volatility and the volatility risk premia, and the predictive power is most prominent during the period of binding funding constraint, which confirms the presence of amplification. Third, we provide further investigation on the holding costs, i.e. fundamental and sentiment risk, and reveal their distinctive influences on the arbitrage activity. It offers an unified approach to examine the level of the respective risk exposure, which is then applied to investigate the value premium anomaly. We find supporting evidence for the behavioral explanation, such that higher sentiment risk exposure deters the arbitrage activity and earns a higher return.

Contents

List of Figures

List of Tables

Acknowledgements

Foremost, I would like to thank my supervisor, Professor Yongcheol Shin for his continuous support and guidance. Without our inspiring discussion, I would have never completed this work. I would like to also thank my TAP members, Dr Michael Thornton and Dr Adam Golinski, for their helps and comments during the TAP meeting, and Professor Charlie X. Cai from University of Liverpool for providing data to support the analyses in Chapter 2. I acknowledge the Department of Economics and Related Studies, University of York, for providing the departmental PhD studentship to financially support this work. I am deeply grateful to my parents, Zuoqiang and Lihuan, my beloved partner, Jia, and all my friends in the UK and China for bringing endless supports and plenty of interesting things in my life aside from "limits to arbitrage".

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Jingzhi Chen

Chapter 1

Introduction

"Noise makes financial markets possible, but also makes them imperfect."–Fischer Black.

"Markets can stay irrational longer than you can stay solvent"–J.M.Keynes.

Modern financial economic theory is built upon the the Efficient Market Hypothesis (EMH), which assumes that agents in the financial market are fully rational. As stated in Barberis and Thaler (2005, BT hereafter), rationality rests on two arguments. First, after receiving new information, agents update their beliefs correctly following the Bayes law. Second, given their beliefs, agents make decisions correctly, in the sense that they are consistent with the expected utility framework. Although there are irrational agents trading on noises (Black, 1986), noises are either claimed to be canceled out in the equilibrium (Bagehot, 1972), or immediately offset by the force of arbitrage (Friedman, 1953). Therefore, EMH posits that rational agents, who understand the Bayes' law and have acceptable preference, ensure that asset prices will be equal to their fundamental value with all available information (Malkiel and Fama, 1970). Hence, an asset's fundamental value is the discounted sum of expected future cash flows.

However, there has been abundant empirical evidence, contradicting to the implications of EMH, prompting an alternative approach to understand the financial market, which is called "behavioral finance." In broad terms, behavioral finance suggests to model and analyse the consequences under weaker assumptions about investor rationality and the force of arbitrage. More specifically, it has two building blocks: psychology and limits of arbitrage. First, psychology refers to the form of agents' irrationality, either because of mistaken beliefs or preferences. Mistaken beliefs arise when agents form their expectations with cognitive biases such as overconfidence and optimism. Preference stems from the way agents evaluate the risky opportunity that is contradicted to the expected utility framework. See BT for an excellent survey on the form of psychology. Second, unlike the argument of Friedman (1953), the limits of arbitrage illustrates the impediments that prevent rational arbitrageurs from driving asset prices towards fundamentals, so that the irrationality can have significant and long-lived impact on asset prices.

In this thesis, we present three theoretical and empirical studies on (nonlinear) limits to arbitrage, and attempt to address the following issues: what are the limitations and constraints that deter arbitrage activity; how they prevent rational arbitrageurs from bringing prices towards fundamentals; and how to identify their effects on deterring arbitrage activity.

According to EMH, arbitrage is an investment strategy that offers a free lunch, i.e. riskless profits at zero cost, whilst behavioral finance argues that it can be costly and limited, rendering it unattractive to rational agents. Here we distinguish the arbitrage frictions into two categories based on their effects on arbitrage: arbitrage costs and funding constraints. Arbitrage costs occur when arbitrage opportunities are observed, and when arbitrage strategies are implemented. First, the learning cost can be significant as mispricing opportunities are difficult to be virtually detected; As highlighted by Shiller et al. (1984) and Summers (1986), although noise is substantially large to cause a persistent mispricing, it do not generate the form of predictability in returns. This is why less sophisticated individual investors do not intervene to exploit mispricings. The next cost occurs during transaction, such as commissions, liquidity cost and short-selling cost.[1] Finally, the holding costs, that incurred when the position remains open, are argued to be the most important costs borne by arbitrageurs (Pontiff, 2006). In particular, fundamental risk is the most obvious risk when arbitrageurs attempt to exploit a mispricing opportunity, such that uncertainty in future fundamental value can lead to losses in arbitrageurs' position (Shleifer and Summers, 1990). Although the majority can be hedged by shorting a substitute asset, it cannot remove all fundamental risk as perfect substitute is rarely available. Another risk is known as the noise trader risk, introduced by De Long et al. (1990). Noise trader risks stem from the the future noise trader demand shocks, which is

[1] See Clinton (1988) and D' Avolio (2002) for studies of transaction costs, Cehn, Stanzl and Watanabe (2002) for price impact costs, Ofek, Richardson and Whitelaw (2004) for short sales restriction, Abreu and Brunnermeier (2002) on synchronization risk.

unrelated to the asset fundamental. Due to the uncertain noise, mispricing can intensify prior to dividend payoff, leading to substantial losses to arbitrageurs in the short-term. Moreover, arbitrageurs are subject to periodic evaluation and funding withdrawal after poor performance. Therefore they tend to have short horizons, and care about the short-term resale prices before dividends are realized. Overall, higher arbitrage costs reduce arbitrageurs' willingness to implement strategies that exploit mispricing. Given these costs, small mispricings may not be arbitraged away since arbitrageurs require certain compensations to outweigh the arbitrage costs.

Funding constraints, as another category of arbitrage frictions, tend to deter arbitrageurs' ability to exploit mispricings even when they are willing to do so. Literature related to slow-moving capital suggests that arbitrage activity requires capital, and the amount of capital available to arbitrageurs determines their ability to eliminate mispricing and provide the liquidity to other investors.[2] Shleifer and Vishny (1997) suggest that arbitrageurs are sophisticated investors, who are able to identify mispricing opportunity and attract capital from outside (less sophisticated) investors. Consider the hedge funds, which are often regarded as the arbitrageurs in the real world. Their capital structure is composed by the equity capital and leverage capital (Bunnermeier and Pedersen, 2009; Ang et al., 2011). Equity is the long-term capital supplied by the outside investors, who can also withdraw their capital. It implies that equity is not always locked into the firm indefinitely. Shleifer and Vishny (1997) argue that outside investors withdraw funding early in response to previous poor performance, which may force arbitrageurs to even liquidate their position, especially when the funding constraint is binding. A loss spiral then arises, such that short-term losses will trigger funding withdrawal, which forces arbitrageurs to liquidate and pushes the price further away, leading to even larger losses. Insecure leverage capital can also be raised on the liability side through the repo markets, prime brokers and derivatives.[3] The leverage capital can be limited since financiers set leverage (margin) requirements to control their value-at-risk. Bunnermeier and Pedersen (2009, BP hereafter) show that leverage constraint can give rise to amplification when leverage requirement increases with the market illiquidity, e.g. the margin spiral. After an initial loss, arbitrageurs are less capable to exploit mispricing and provide liquidity. Illiquidity then leads to higher leverage requirement, which further tightens

[2]See also Mitchell, Pulvino and Pedersen (2007), Brunnermeier and Pedersen (2009), Duffie (2010) and Mitchell and Pulvino (2012)

[3]See Bunnermeier and Pedersen (2009), Ang et al. (2011) and Fung and Hsieh (2013) for more discussion about hedge funds and their operation.

the capital supplied to arbitrageurs. These two amplification mechanisms (loss and margin spirals) are essential to understand the sudden market crash that is attributed to a moderate triggering event (Brunnermeier and Oehmke, 2013).

The limits to arbitrage determines to what extent irrationality can affect the aggregate stock market, the cross-section of average return and corporate finance decisions. This issue has been put on spotlight during the 2007-2009 financial crisis, as mispricings are substantial and persistent in the global financial markets.[4] Understanding what limits arbitrage is therefore of great importance to investors who aim to profit from mispricings, and to regulators who intends to improve the market efficiency.

In Chapter 2 we address this issue and propose to identify the limits to arbitrage through examining the conflicting impacts of arbitrage costs and funding constraints on arbitrage activity. In the presence of arbitrage costs, arbitrage activity tends to increase with the size of mispricing as it reflects the higher cost-adjusted return, which we call the positive capital allocation effect. Consider, on the other hand, the constraints on arbitrage capital. Arbitrageurs are forced to be rather passive on larger mispricing since it leads to worse funding condition, which is denoted as the negative funding constraint effect. Our theoretical framework extends SV, and describes the combined consequences of these two categories of arbitrage impediments: when funding constraint is not binding, arbitrage costs are the dominating friction over funding constraints, and the positive capital allocation effect appears; when funding constraint becomes binding, funding constraints exhibit dominance, and thus arbitrage activity tend to drop with mispricing; the overall arbitrage activity displays an inverse U-shape against the size of mispricing error. We then carry out an empirical investigation by applying a state-dependent, Generalized Error Correction Model (GECM) to test our theoretical predictions in the S&P 500 index spot and future markets. The empirical evidence is generally supportive of such nonlinear limits to arbitrage. Furthermore, our study makes the contributions to the literature on slow-moving capital by identifying the periods when the negative funding constraint effect dominates, which are coincide with the market crashes in 1987, 1998, 2000 and 2007-2008 (Mitchell, Pulvino and Pedersen, 2007; Mitchell and Pulvino, 2012).

In Chapter 3 we focus on investigating the time-varying leverage constraints,

[4] In the credit markets, Garleanu and Pedersen (2011) and Bai and Collin-Dufresne (2013) document large CDS-bond basis. In currency markets, Coffey, Hrung and Sarkar (2009) find the violations to covered interest parity (CIP). In treasury market, Fontaine and Garcia (2012) also document large price deviation between bonds with identical cash-flow.

which is also known as funding liquidity, i.e. the ease with which arbitrageurs can obtain capital. As demonstrated by SV and BP, the binding funding constraint may lead to loss and margin spirals, in which a small shock could be amplified to large spillovers across the financial market. Although the implications behind the funding illiquidity are well-understood,[5] one main issue is that they fail to identify the periods exactly when the funding constraint is binding and amplification mechanisms are at work. We extend the theoretical framework in Chapter 2 by allowing for an endogenous leverage setting. We define the funding liquidity as the marginal leverage raised by arbitrageurs and the arbitrage efficacy as the marginal mispricing correction achieved by arbitrageurs with respect to one more unit of mispricing. We find that when the funding constraint is not binding, arbitrageurs are able to raise sufficient capital in order to cope with larger mispricings. In this situation arbitrage is effective as larger mispricing induces the higher mispricing correction. However, during the periods when the funding constraint becomes binding, arbitrageurs fail to obtain sufficient capital, which renders arbitrage ineffective. The model thus posits that funding liquidity affects the efficacy of arbitrage, the sign of which is able to identify binding funding constraint. To empirically investigate the validity of this measure, we suggest to estimate the arbitrage efficacy by regressing the daily difference of dynamic mispricing correction on the daily variations of mispricing error. Next, we apply this methodology to the S&P 500 index spot and e-mini future markets over the period, 1999-2015, and obtain evidence in favour of the theoretical predictions as follows: First, the funding liquidity measure, proxied by arbitrage efficacy, is closely related to other popular measures such as TED spread, VIX index of implied volatility and the dividend yield of S&P 500 index. More importantly, using this measure, we can successfully identify four periods when the funding constraint is binding, i.e. the burst of dot-com bubble in 2000, the global financial crisis in 2007-2008, and the debt ceiling crises in 2011 and 2013. Second, we find that changes in the funding liquidity measures can predict the market volatility, especially the volatility risk premia. Furthermore, the predictability is mostly prominent during the periods when arbitrage is inefficacious, which is consistent with the presence of amplification effect under the binding funding constraint (SV and BP). As the proposed funding liquidity measure not only captures the funding condition among arbitrageurs, but also identifies the period of binding funding constraint, it offers an useful tool for

[5] See Fontaine and Garcia (2011), Garleanu and Pedersen (2011) Nagel (2012) and Frazzini and Pedersen (2014).

regulators to monitor the market and guide the timing of policy making.

In Chapter 4 we aim to redress the ongoing debate on the value premium, known as one of the anomalies that cannot be explained by the Capital Asset Pricing Model. Fama and French (1992, 1993, 1996) claim, under the rational point of view, that the value stocks outperform the growth stocks due to the higher exposure of fundamental risk. However, behavioral finance argues that value stocks are those out-of-favor stocks with a higher sentiment risk. We propose to analyse the level of fundamental or sentiment risk exposure within an unified model through the distinctive impacts of the respective risk exposures on arbitrage activity. In particular, we extend the theoretical framework in Chapter 2 and allow for risk averse and heterogeneous arbitrageurs. The model suggests that higher fundamental (sentiment) risk tends to deter the initial mispricing correction and increases (reduces) subsequent noise momentum. To carry out the empirical investigation, we apply a two-stage estimation methodology to the daily S&P value and growth spot and future indices over the period, 1999-2014. In the first stage, we model the joint distribution of HML returns (portfolios that long in value and short in growth) in spot and future as a regime-switching process. We can identify the three distinctive regimes as follow: the regime with value premium corresponds to the crisis period of 2000 and 2008 while the regime with value discount to the bear markets. Finally, the bull markets do not display any anomaly. In the second stage, the two-period generalized ECM is applied to each regime and estimate the implied arbitrage activity. We find that the implied arbitrage activity in value (growth) stocks tend to be limited by higher exposure to sentiment risk under the value premium (discount) regime, which provides support for the behavioral-based view. The chapter is the first attempt to analyse the impact of fundamental (sentiment) risk exposure on arbitrage activity, whilst the empirical methodology and results add further evidence to understand the ongoing debate on value premium anomaly.

The rest of the thesis is organised as follows. Chapter 2 investigate the two categories of limits to arbitrage: arbitrage costs and funding constraint, and develops their combined effects on arbitrage activity. Chapter 3 designs a framework to measure the time-varying funding liquidity and identify periods under which the markets suffer from severe funding constraints. In Chapter 4, we reinvestigate the ongoing debate on the value premium by deriving the impacts of the fundamental and sentiment risk exposures on arbitrage activity. Chapter 5 summarises the main conclusions of the thesis, and suggests the possible venues for future research extensions.

Chapter 2

Nonlinear Limits to Arbitrage

2.1 Introduction

Arbitrageurs' agressive search for arbitrage opportunities ensure that mispricing is short-lived. However, arbitrage is costly and risky in practice, and the presence of arbitrage frictions prevent arbitrageurs from making full mispricing corrections, leading to market anomalies and resource misallocations (Gromb and Vayanos 2010). Therefore, measuring the level of arbitrage activity and, more importantly, understanding what limits arbitrage is of great ongoing research interest to both market participants and regulators. In previous literature, mispricing level is often used as an important variable in arbitrageur decision making (Dwyer et al. 1996, Balke and Fomby 1997, Martens et al. 1998, Tse 2001, Tao and Green 2009, Theissen 2011, Gyntelberg et al. 2016). We study the limits of arbitrage by examining how arbitrageurs respond to different levels of mispricing opportunity.

There are two distinct and countervailing views of what limits arbitrage and causes persistent mispricing: arbitrage costs and funding constraints. On the one hand, prior literature, such as Roll et al. (2007), Bai and Collin-Dufresne (2013) and Gyntelberg et al. (2016), suggests that arbitrage costs (e.g., illiquidity, transaction costs and risk), are responsible for the persistence of mispricing. Arbitrageurs are willing to exploit the mispricing only when it exceeds a certain threshold that reflects the cost of conducting the arbitrage trade. Moreover, this suggests that arbitrage activity will increase with the size of mispricing, since larger mispricing provides a higher cost/risk-adjusted return. We call this the positive capital allocation effect with respect to the size of mispricing.

On the other hand, various studies document the importance of funding constraints in limiting arbitrage. They suggest that severe and prolonged mispricing during times of market turmoil results from arbitrageur funding constraints (Mitchell, Pulvino and Pedersen 2007, Oehmke 2009, Acharya et al. 2009, Duffie 2010, Garleanu and Pedersen 2011, Mitchell and Pulvino 2012). Furthermore, the theoretical work of Brunnermeier and Pedersen (2009) argues that larger mispricing can exaggerate financier expectations of future volatility, which tightens the funding constraint. Under this view, larger mispricing tends to make funding constraints more binding, and further deters mispricing correction by arbitrageurs. We call this the negative funding constraint effect with respect to the size of mispricing.

The two sources of limits to arbitrage drive opposite conclusions as to how arbitrageurs respond to mispricing. While the former view has been evidenced through threshold models (Martens et al. 1998, Tse 2001, Tao and Green 2009, Gyntelberg et al. 2016), the combined effect of arbitrage costs and funding constraints on arbitrageur response to different magnitudes of mispricing is not well understood in the literature, theoretically or empirically. Our paper seeks to address this knowledge deficit.

We extend the seminal work by Shleifer and Vishny (1997; henceforth SV) and show that a combination of arbitrage cost and funding constraint explanations of the limits to arbitrage, produces a nonlinear relationship between the size of mispricing and arbitrage activity. When the mispricing error is small, funding constraints tend to be loose, the positive capital allocation effect dominates and arbitrage activity intensifies with the size of mispricing. In contrast, extremely large mispricing makes funding constraints become binding. In this circumstance, the negative funding constraint effect becomes the dominant driver of the limits to arbitrage, such that increases in the size of mispricing induce a lower relative level of arbitrage activity. Intuitively, when funding constraints are binding, arbitrageurs are forced to adopt the full investment strategy. As the size of mispricing enlarges, arbitrageur capital in relation to the size of mispricing becomes smaller.

Empirically, to study the nonlinear dynamic, we adopt a Markov switching extension of the Cai, Faff and Shin (2017; henceforth CFS) generalized error correction model[1]. Applying this model to the S&P 500 Index spot and futures markets over

[1]To capture such multi-period arbitrage activity, CFS develop a generalized error correction model (GECM) and estimate both the initial mispricing correction and the subsequent noise momentum parameters, where the latter are designed to measure the persistence of the uncorrected pricing errors. Applying the model to a wide range of international spot–futures market pairs, CFS document pervasive evidence of noise momentum around the world. In this unified theoretical framework, a higher

the period 1986–2015, we find strong evidence of regime-dependent nonlinear limits to arbitrage. In particular, we are able to identify three distinct regimes: a normal market state with a small mispricing error and low volatility, a transition market state with a medium mispricing error and medium volatility, and an extreme market state with a large mispricing error and high volatility. We observe a relatively low mispricing correction in the normal state, but a dramatic increase during the transition state. This suggests that arbitrage activity tends to intensify with the size of mispricing error when the mispricing level increases from low to medium. By contrast, the mispricing correction is lowest during the extreme state. This suggests that when mispricing increases from a medium to a high level, funding constraints become binding and arbitrageurs are unable to raise external funds when the arbitrage opportunity is at its best. Overall, arbitrage activity thus displays an inverse U-shape against the magnitude of mispricing errors, which is consistent with our model prediction.

To verify that our regime estimation captures variation in funding constraints, we examine the potential linkages between the three hidden market states and various observable measures of funding conditions, illiquidity and risk[2]. Our analyses show that arbitrage funding constraints increase monotonically from normal to extreme states. The data also document the flight-to-quality/safety phenomenon such that fund flows into passive index funds decrease from normal to transition states but increase from transition to extreme states. Overall, extreme states capture a period of considerable market stress. From the arbitrageur perspective, the extreme state presents a 'cocktail' of good and bad phenomena. On the positive side, it entails large mispricing errors and higher valuation uncertainty – thus presenting arbitrageurs with more profitable opportunities to exploit. On the negative side, however, arbitrageurs will tend to face a higher cost of capital and higher transaction costs, which lower net profit.

Our study differs from other studies of the funding constraint effect on the limits of arbitrage and contributes to the literature in several ways. First, it offers an approach for determining the relationship between funding constraints and arbitrage dynamics. The effect of funding constraints has been studied in the literature through the proxies of arbitrage activity (Cielinska et al. 2017), the size of arbi-

---

initial mispricing correction and a lower mispricing persistence induce a faster overall speed of adjustment.

[2]For funding conditions we use hedge and mutual fund flows, growth rate of total financial assets, financial sector leverage and broker-dealer leverage; for illiquidity we use the Amihud (2002) illiquidity measure of the spot index, Treasury security-based funding illiquidity and TED spread; and for arbitrage risk we use the idiosyncratic risk of the index constituents.

trage violations (Fontaine and Garcia 2012, Garleanu and Pedersen 2011, Frazzini and Pedersen 2014, Akbas et al. 2016) and market liquidity (Nagel 2012, Schuster and Uhrig-Homburg 2015). This paper is the first attempt to study the funding constraint effect via its impact on arbitrageur error correction with respect to mispricing. Duffie (2010) suggests that price reversal (measured by error correction) reveals the arbitrage frictions borne by arbitrageurs, which provides insights regarding the level of arbitrage activity. Furthermore, studying funding constraints in an error correction context enables us to integrate both arbitrage cost and funding constraint explanations of the limits to arbitrage in a unified framework and to identify the time-varying interplay between the two types of limit. Specifically, we show that whether or not funding constraints are binding generates two different predictions regarding the marginal change in arbitrageur response to the change in the size of mispricing. Such modeling unifies our understanding of normal- and extreme-time arbitrage activity and demonstrates nonlinear limits to arbitrage. In normal market conditions, what limits arbitrage activity is the trade-offs between risk and return; in extreme market conditions, scarcity of funding limits arbitrage activity the most, which is consistent with slow-moving capital. So relatedly, this shows that relying solely on the (linear) arbitrage cost explanation will likely lead to flawed predictions about the limits to arbitrage.

Second, extant literature has been devoted to identifying the source of limits to arbitrage that result in large and persistent mispricing empirically. Gallagher and Taylor (2001) and Tse (2001) find supporting evidence for the positive capital allocation hypothesis in index arbitrage using the smooth transition autoregressive (STAR) model. Coffey, Hrung and Sarkar (2009) and Griffoli and Ranaldo (2011) investigate the time-series linkage between persistent mispricing and measurements of arbitrage risk or funding constraints. Bai and Collin-Dufresne (2013) rely on the cross-sectional variation in different asset classes, and find that arbitrage cost is the main explanation of this variation. Tao and Green (2009) and Gyntelberg et al. (2016) analyze the time-varying width of the no-arbitrage threshold through the threshold error correction model (ECM) to reveal the arbitrage frictions that deter arbitrage activity. While we too empirically apply the nonlinear ECM, our paper also advances the literature by providing theoretical connections to the empirical model. We contribute to the empirical study of arbitrage by developing an empirical model having a meaningful range of parameters with clear theoretical backing, in describing the arbitrage and price dynamics. Our study strengthens the interpretation of such a nonlinear model.

Finally, our study offers a general way of capturing arbitrage activity in the market and analyzes the effect of funding constraints through a state-dependent ECM backed by theoretical analysis. Direct measures of arbitrage activity are difficult to achieve, as this would require explicit identification of arbitrageur trades and post-trade transaction-level data, such as trade repository data; such data is not generally available, other than to regulators. Our estimation of arbitrage activity uses only publicly available data. In addition, there is a fast-growing literature that documents the significant impact of fund flow in creating mispricings, such as short-term momentum and long-term reversal (Frazzini and Lamont 2008, Luo 2012, Akbas et al. 2015). Our empirical results show that, although we study equity index arbitrage, the regimes identified by our nonlinear model are positively associated with a wide range of traditional measures of funding constraints. In other words, estimating such a model provides an insight into the time-series variation of arbitrage funding constraints beyond the index arbitrage activity. Identifying the conditions and periods under which funding constraints are binding is pivotal to enhancing our understanding of limits to arbitrage.

The remainder of our paper is organized as follows. In Section 2 we present the theoretical framework which builds on an important extension of Shleifer and Vishny (1997), and develop the main propositions and predictions. In Section 3 we develop a general empirical framework designed to best capture the various predictions derived from our theoretical framework. In Section 4 we outline a specific empirical application based on the linkage between S&P 500 Index futures and spot markets, and we present and discuss our empirical results. In Section 5 we make concluding remarks. Mathematical proofs are collected in the Appendix.

2.2 Theory and Predictions

2.2.1 The model

We begin with an introduction to a range of basic concepts in line with the SV setup. There is one asset in unit supply with fundamental value V, and three types of market participant: noise traders, arbitrageurs and fund investors, trading in three periods, $t = 1, 2, 3$. Noise traders arrive in period t with shocks S_t, which represents the extent to which noise traders in aggregate under-value the asset price relative to its fundamental value, V. In particular, S_1 is observable to arbitrageurs; S_2 is allowed to be stochastic, taking a value of 0 (the 'good' state pertains) with probability $1-q$, or $S_2 = S_2^* > S_1$ (the 'bad' state pertains) with probability q; $S_3 = 0$, such that price converges to fundamental value in period 3, $P_3 = V$.

After noticing the mispricing opportunity in period 1, arbitrageurs accumulate funding resources, F_1, from fund investors and determine the fraction of funding, $0 \leq \beta_1 \leq 1$, to invest in the asset. Hence market clearing implies that the price of the asset in period 1 is given by $P_1 = V - S_1 + D_1$ with $D_1 = \beta_1 F_1$. Notice that a general funding constraint is always imposed in the SV model, such that $F_1 < S_1$, so that arbitrageurs are unable to fully correct the mispricing error in period 1. Funding resources in period 2, F_2, are determined endogenously by past performance, such that $F_2 = F_1\left[1+\alpha\beta_1\left(\frac{P_2}{P_1}-1\right)\right]$, where $P_2 = V$ in the good state, or $P_2 = P_2^* = V - S_2^* + F_2$ [3]in the bad state; α captures the sensitivity of fund flows to past performance, which is assumed to follow the stability condition, such that

$$1 \leq \alpha < (V - S_1 + F_1)/(S_2^* - S_1 + F_1).$$

It ensures that fund investors are not so overly sensitive that arbitrageurs will lose all their funding in period 2.

2.2.2 The role of funding constraint

Under this model setup, where arbitrageurs actively choose their investment strategy, β_1, so as to maximize their wealth at period 3, F_3, SV obtain the following first order condition:

$$(1-q)\left(\frac{V}{P_1}-1\right)+q\left(\frac{P_2^*}{P_1}-1\right)\frac{V}{P_2^*} \geq 0. \tag{2.1}$$

To understand the role of funding constraints it is more convenient to rewrite Eq. (2.1) as

$$(1-q)(R_1)-q(R_2-R_1) \geq 0, \tag{2.2}$$

where $R_1 = \frac{V}{P_1}-1$, $R_2 = \frac{V}{P_2^*}-1$.

Intuitively the LHS of Eq. (2.2) measures the net marginal return, MR, of invest-

[3]Arbitrageurs make no investment ($\beta_2 = 0$) when $S_2 = 0$ and $P_2 = V$, but full investment ($\beta_2 = 1$) when $S_2 = S_2^*$ and $P_2 = P_2^*$, since price will reliably converge in period 3.

ing in period 1[4]. When MR is larger than zero, arbitrageurs will adjust their investments, β_1, in period 1 as much as they can[5]: on the one hand, if there is no funding constraint, arbitrageurs will continue increasing their investment so that it eventually drives MR towards zero; as the equality holds, arbitrageurs will stop increasing their investment in period 1 and save the rest of the funding for future investment. Thus we observe the partial investment strategy employed by the arbitrageurs[6].

On the other hand, in the presence of a binding funding constraint, the inequality in Eq. (2.1) suggests that arbitrageurs cannot take further action to reduce MR once they exhaust their available funding. Thus they fail to obtain $MR = 0$ and the inequality holds. Under this circumstance, arbitrageurs are known to adopt the full investment strategy in the SV model. Importantly, we show that they do so because their funding constraint is binding, since if there were more funds available, arbitrageurs would have invested them in the initial error correction. Comparing this with the partial investment strategy where arbitrageurs limit their initial arbitrage activity strategically, funding constraints deter their arbitrage activity under the full investment strategy.

2.2.3 Arbitrage activity

To study arbitrage activity under different strategies, we first define our measurements of arbitrage activity. To this end, CFS introduce the concept of the initial mispricing correction and the subsequent mispricing persistence (called "noise momentum") in the framework of the two-period generalized ECM (GECM). The inclu-

---

[4]In particular, R_1 measures the marginal return on one dollar of investment in period 1 (the initial investment) when price recovers in period 2 (the good state pertains, i.e., $S_2 = 0$). $R_2 - R_1$ is the marginal opportunity loss on one dollar of investment in period 1 when mispricing deepens further in period 2 (the bad state pertains, i.e., $S_2 = S_2^* > S_1$). This is the amount that the arbitrageurs have to give up as they have already invested in period 1, and so that same dollar cannot be invested in period 2 when a better arbitrage opportunity arises because of the further deteriorating price. When taking into consideration the probability of each market state through the parameter q, the LHS of Eq. (2.2) measures the net marginal return of investing in period 1.

[5]Arbitrageurs' choice of investment would have an impact on the relative cost and benefit of investing in period 1. Specifically, an increase in D_1 would reduce R_1 and therefore also the level of MR.

[6]Furthermore, it is only when the equality holds that it is a partial-investment equilibrium solution. Suppose the opportunity gain of investing in period 2 when the bad state pertains is higher than the gain of investing in period 1 when the good state pertains ($MR < 0$); then arbitrageurs would save more money to invest in period 2, which would increase F_2 through both the increase in funding allocation and the reduction in the performance-based funding effect. Therefore, the attractiveness of investing in period 2 would graduallly reduce until the opportunity gain became equal to the return on investing in period 1.

sion of noise momentum provides an advantageous framework for analyzing the asset pricing dynamics and overall arbitrage process. CFS study 26 index future relationships around the globe and show that the traditional one-period ECM is misleading in the presence of noise momentum. Following CFS, we characterize arbitrage activity by initial mispricing correction and subsequent noise momentum. The initial mispricing correction is defined as

$$K = \frac{D_1}{S_1} = \frac{\beta_1 F_1}{S_1}, \tag{2.3}$$

which is designed to capture the proportion of mispricing correction achieved by arbitrageurs in period 1, while the subsequent noise momentum is defined as

$$\Lambda = \frac{V - P_2}{V - P_1} = \frac{V - P_2}{S_1 - D_1}, \tag{2.4}$$

capturing the degree of mispricing error persistence into the next period. In order to connect these two empirical measures with our theoretical analyses, we express the two parameters as the expectation with respect to q in period 1, such that

$$\kappa = E_q(K) = \frac{\hat{\beta}_1 F_1}{S_1},\ \lambda = E_q(\Lambda) = q\frac{V - P_2^*}{V - P_1}, \tag{2.5}$$

where $\hat{\beta}_1$ is the equilibrium investment strategy informed by the first-order condition (FOC), Eq. (2.2). Eq.(2.5) implies that both κ and λ are below unity[7] when rational arbitrageurs choose the equilibrium investment strategy to engage in the arbitrage opportunity.

Notice that the initial mispricing correction, κ, is the product of $\hat{\beta}_1$ and the ratio F_1/S_1. The equilibrium investment strategy $\hat{\beta}_1$ captures the strategic response of arbitrageurs to the risky arbitrage opportunity: whether to invest in period 1 or to avoid the potential loss and invest in period 2. It thus represents the willingness of arbitrageurs to engage in arbitrage activity, and we refer to it as the capital allocation

[7] We rewrite the first-order condition, Eq. (2.2) as

$$\frac{V - P_1}{P_1} \geq q\left(\frac{V - P_2^*}{P_2^*}\right),$$

so it is easily seen that

$$\lambda = q\frac{V - P_2^*}{V - P_1} \leq \frac{P_2^*}{P_1} < 1.$$

effect. The term F_1/S_1 is the ratio of available funding over mispricing shock, which captures the arbitrageur's relative funding condition. Therefore, the initial correction reflects the two important impediments to arbitrage activity: arbitrage cost/risk and funding constraints.

While the mispricing correction parameter measures the immediate arbitrage effect, the noise momentum parameter captures the subsequent price recovery. Eq. (2.5) shows that the noise momentum, λ, captures both the probability, q, of deepening error in the subsequent period, and the degree of deepening mispricing error (i.e., $\frac{V-P_2^*}{V-P_1}$). In the SV model, the probability, q, of deepening error in the subsequent period reflects the fact that the arbitrage opportunity is risky. There exists a threshold point q^*, such that when $q < q^*$, the probability that the mispricing error deepens is relatively low, and arbitrageurs will be more likely to fully invest at period 1. Alternatively, when $q > q*$ (i.e., the probability of deepening misperceptions is 'critically' high), arbitrageurs will defer some of their investment. However, both q and q^* are unobservables in practice. Eq. (2.5) shows that λ captures the important information regarding the probability q, such that higher noise momentum indicates higher expected misperceptions in the future, which results in higher uncertainty in the pattern of price recovery.

2.2.4 Mispricing and arbitrage activity

We are most interested in establishing the impacts of initial mispricing error, S_1 on the arbitrage activity parameters: κ and λ, and the main results are summarized in Proposition 1 (Proofs are provided in the appendix).

Proposition 1. *Consider the model setup from Section 2.2.1 and 2.2.3, and the equilibria from Eq. (2.2). Under the stability condition,* $1 < \alpha < (V - S_1 + F_1)/(S_2^* - S_1 + F_1)$*, the impacts of mispricing error on arbitrage activity are derived as follows:*

$$\frac{\partial \kappa}{\partial S_1} \left\{ \begin{array}{ll} = 0 & for\ \hat{\beta}_1 = 0 \\ > 0 & for\ 0 < \hat{\beta}_1 < 1 \\ < 0 & for\ \hat{\beta}_1 = 1 \end{array} \right\}, \quad \frac{\partial \lambda}{\partial S_1} < 0.$$

Furthermore, we have:

$$\left| \frac{\partial \lambda}{\partial S_1} \right|_{0<\hat{\beta}_1<1} < \left| \frac{\partial \lambda}{\partial S_1} \right|_{\hat{\beta}_1=1}.$$

The proposition is intuitive. Consider first the capital allocation effect: as the mispricing error S_1 becomes higher, arbitrageurs are willing to allocate more resources to correct the mispricing (i.e., $d\hat{\beta}_1/dS_1 > 0$) in the partial investment equilibrium, which enhances the mispricing correction. Consider second the funding constraint effect: the deeper mispricing would intensify the general funding constraint, measured by the arbitrageur funding condition relative to the size of error (i.e., a lower F_1/S_1), which would tend to deter the mispricing correction. Proposition 1 shows that the positive capital allocation effect dominates the negative funding constraint effect under the partial investment equilibrium (i.e., $\partial\kappa/\partial S_1 > 0$). This is consistent with the earlier studies showing that larger mispricing errors induce a greater mispricing correction and faster speed of adjustment[8]. On the contrary, when the funding constraint binds, the initial mispricing correction is determined mainly by the funding availability, since the change in mispricing will have no effect on the strategy (i.e., $\partial\beta_1/\partial S_1 = 0$). As S_1 grows, arbitrageurs are forced to face a relatively deteriorating funding condition and disengage in arbitrage activity (i.e., $\partial\kappa/\partial S_1 < 0$). This confirms that arbitrageurs would be able to strategically allocate their funds in different periods only if the funding constraint is not binding.

Furthermore, Proposition 1 has additional important implications about the impacts on the noise momentum parameter. First, λ is always negatively related to S_1, such that deeper mispricing reduces the expected noise persistence. Second, the negative relationship between λ and S_1 is not monotonic between the partial and full investment equilibria. Here, changes in S_1 affect λ, mainly on the relative degree of deepening mispricing, but rather independent from probability q. In the partial investment equilibrium, as S_1 rises, the negative impact of S_1 on λ is relatively small since the increase in S_1 renders both P_1 and P_2 less efficient (both P_2^* and P_1 tend to drop with deeper mispricing at similar speed). By contrast, in the full investment equilibrium, λ declines more sharply with S_1, since the relative size of deepening mispricing becomes much smaller (P_1 drops faster while P_2^* starts to increase; thus the difference $P_1 - P_2^*$ becomes smaller)[9].

Proposition 1 also reveals implications for the overall speed of price adjustment. Note that the speed of adjustment is positively associated with κ, but negatively with

[8]Empirical evidence is documented under the threshold ECM model (Dwyer et al. 1996), Martens, Kofman and Vorst 1998) and the smooth transition model (Gallagher and Taylor 2001, Tse 2001).

[9]In the partial investment equilibrium, higher S_1 leads to less efficient pricing in period 1, P_1 since arbitrageurs ability to bear against mispricing is limited. Moreover, higher S_1 leads to less efficient pricing in period 2, P_2^*, since arbitrageurs tend to lose more funding after augmenting their investment in period 1. However, the latter case differs in the full investment equilibrium. The scarcity of arbitrage funding deters arbitrageurs initial investment, which also prevents them from losing too much in period 2. This relatively improves the pricing efficiency in period 2.

λ. In the partial investment equilibrium, κ rises and λ marginally falls simultaneously with S_1. Thus, in this regime, the overall speed of adjustment improves with S_1. However, due to the binding funding constraint, κ starts to drop with S_1 while λ keeps falling sharply. In this case, the impact of S_1 on the overall speed of adjustment is uncertain and empirically determined.

Proposition 1 clearly demonstrates that the mispricing correction, κ, does not always have a positive association with the magnitude of mispricing error, as suggested by the earlier studies focusing only on arbitrage cost. This highlights the importance of taking into account the limits to arbitrage related to funding constraints, which indicate a rather negative association in the full investment equilibrium. We continue by investigating the link between the size of mispricing and the probability of entering full investment equilibrium. In particular, we examine the determinants of q^*, the threshold parameter that drives the arbitrageur's choice of strategy. (Proofs are provided in Appendix A).

Proposition 2. *Consider the model setup from Section 2.2.1 and 2.2.3, and the equilibria from Eq. (2.2). Under the stability condition,* $1 < \alpha < (V - S_1 + F_1)/(S_2^* - S_1 + F_1)$*, then the higher is the initial noise trader shock,* S_1*, the higher is the threshold probability* q^* *of taking full investment strategy, i.e.* $\frac{\partial q^*}{\partial S_1} > 0$.

Proposition 2 extends the SV model and shows how q^* is affected by mispricing error. Notice that q^* is derived at the point where arbitrageurs are indifferent as to strategy (i.e., $D_1 = F_1$ under the partial investment strategy). Thus q^* can be rewritten as[10]

$$q^* = \frac{R_1}{R_2} \text{ if } D_1 = F_1,$$

where $R_1 = \frac{V}{P_1} - 1$, $R_2 = \frac{V}{P_2^*} - 1$. Intuitively, as S_1 increases, the threshold probability that arbitrageurs adopt the full investment strategy increases[11]. While q^* is determined by the model parameters and known to arbitrageurs, q is an exogenous variable that is estimated by arbitrageurs. Given a uniform distribution of q, arbi-

[10]From a modeling perspective, q^* is determined as a complex function of the parameter set $\{V, S_1, S_2^*, F_1, \alpha\}$. So we have

$$q^* = \frac{(S_1 - F_1)(V + F_1 - S_2^* - \alpha F_1)}{V(S_2^* - S_1) + (S_1 - F_1)(V + F_1 - S_2^* - \alpha F_1)}.$$

[11]As the initial mispricing deepens, the arbitrage return if the good state pertains, R_1, increases. In the meantime, the return if the bad state pertains, R_2, decreases relatively, which indicates a smaller opportunity loss, $R_2 - R_1$, for adopting the full investment strategy.

trageurs are more likely to be fully invested since their threshold probability q^* is higher.

When combining the above two propositions, we find an inverse U-shaped relationship between mispricing error and initial arbitrage activity. We illustrate this nonlinearity by means of a numerical analysis. Figure 2.1 illustrates how arbitrageurs' initial mispricing correction and subsequent noise momentum innovates with respect to the mispricing error, following the numerical example in the SV paper. Let $V = 1$, $S_2^* = 0.4$, $q = 0.3$, $\alpha = 1.2$ and $F_1 = 0.2$. In the left panel, we show that as S_1 increases from 0.2 to 0.4, the impact of this innovation on mispricing correction is nonlinear, conditional on the partial and full investment equilibria. In particular, the initial mispricing correction (solid line) displays an inverse U-shaped relation with respect to mispricing error. In the partial investment equilibrium where the threshold q^* lies below the probability q, both the capital allocation and funding constraint effects contribute to the mispricing correction, while only the negative funding constraint effect remains in the full investment equilibrium. The right panel plots the variation in noise momentum (solid line), which is relatively high and remains stable in the partial investment equilibrium, but drops sharply after the threshold q^* exceeds the probability q in the full investment equilibrium.

Overall, the numerical analysis portrayed in Figure 2.1 shows that when the mispricing error is small, funding constraint is less binding and therefore arbitrageurs adopt the partial investment strategy; conventional arbitrage costs, such as transaction and holding costs, tend to be the dominant factors limiting arbitrage, suggesting that arbitrageur correction rises mostly with mispricing error during these low to moderate mispricing periods. As mispricing deepens further, the funding constraint is likely to be binding due to margin\haircut increases and industry-wide fire sales. Therefore arbitrageurs' initial error correction is limited as mispricing deepens during the extreme mispricing period.

2.2.5 Empirical predictions

According to our theoretical framework, we consider three regimes with different magnitude of mispricing error (small, medium and large), and denote the arbitrage activity κ^r and λ^r with $r \in (s, m, l)$[12]. We summarize three main predictions derived from our theoretical propositions as follows:

[12]We later interpret these states in the context of market condition as normal, transition and extreme period.

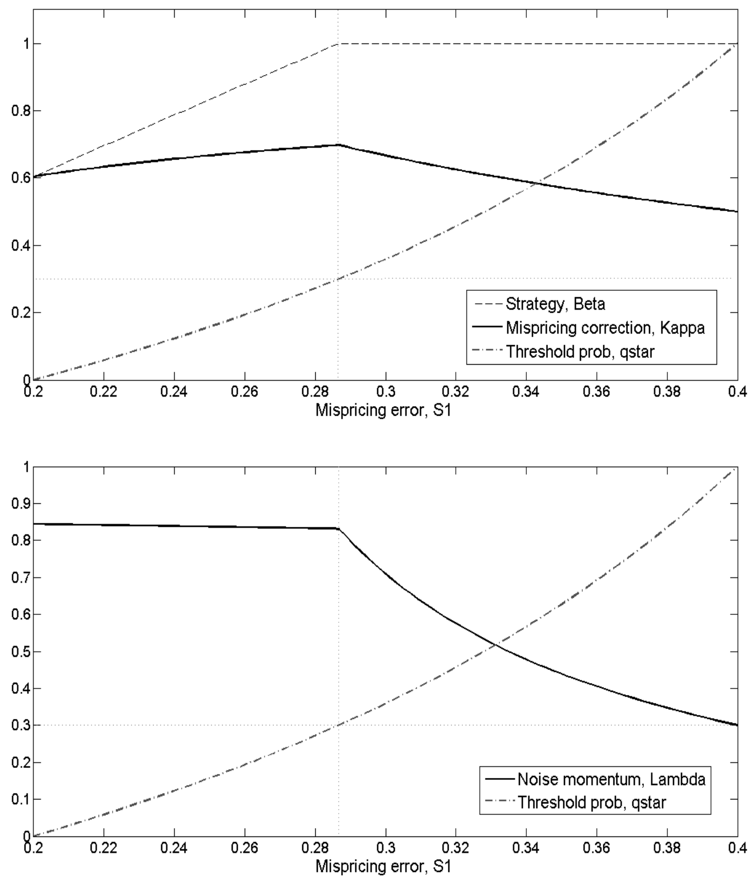

Figure 2.1: The strategic response effect on mispricing correction and noise momentum

Figure 1 shows the nonlinear impact of mispricing errors on mispricing correction (top) and noise momentum (bottom). It follows the numerical examples from the SV paper, such that $V = 1$, $S_2^* = 0.4$, $q = 0.3$, $F_1 = 0.2$, $\alpha = 1.2$ and initial mispricing S_1 increases from 0.1 to 0.4. The top figure also plots the optimal investment strategy ($\hat{\beta}_1$) and the threshold probability q^* of deepening mispricing, while the bottom figure provides additional plots of the threshold q^*. Both figures highlight the nonlinearity in the arbitrageurs' activity under two path to equilibrium.

Prediction 1: The U-shaped initial arbitrage activity. (i) $\kappa^m > \kappa^s$; an initial mispricing correction rises with the size of mispricing error. (ii) $\kappa^l < \kappa^m$; In the presence of the binding funding constraint, further rise in mispricing error will induce the slower mispricing correction.

Prediction 2: The regime-dependent noise momentum. (i) λ^s is relatively large, suggesting that mispricing tends to persist in period 2 as the initial mipricing error is too small. (ii) $\lambda^m \approx \lambda^s$; the difference between λ^s and λ^m is relatively negligible under the partial investment strategy. (iii) λ^l is significantly smaller than λ^s and λ^m.

Prediction 3: The overall speed of adjustment (SOA) tends to be faster with the larger mispricing error under the partial investment strategy when the funding constraint is not binding. On the contrary the impact of mispricing error on SOA will be uncertain when the funding constraint is binding.

2.3 Empirical Applications

2.3.1 The State-Dependent Markov Switching Generalized Error Correction Model (MS-GECM)

To empirically examine the validity of the main hypotheses regarding the limits of arbitrage and its nonlinear impacts on the asset pricing dynamics developed in the previous section, we apply our study to the S&P 500 Index futures market. We first consider the two-period GECM advanced by CFS:

$$\Delta f_t = \kappa z_{t-1} + \lambda\,(1+\kappa)\,z_{t-2} + \delta\Delta f_t^* + \gamma\Delta f_{t-1} + u,\; u_t \sim iid(0,\,\sigma_\mu^2), \qquad (2.6)$$

where f_t is the (observed) market price, f_t^* is the fundamental value for the asset, $z_t\,(= f_t - f_t^*)$ is the pricing error that is the short-term deviation of price from its fundamental value, Δ is the first-difference operator, and u_t is the zero-mean idiosyncratic error term with zero mean and finite variance σ_u^2, whilst $\kappa, \lambda, \delta, \gamma$ are the parameters of interest. The distinguishing feature of the GECM is that we can simultaneously capture the multi-period (complex) arbitrage activity by accommodating both initial mispricing correction through κ, and noise momentum through $\lambda\,(1+\kappa)\,z_{t-2}$, with λ measuring the strength of noise momentum and $(1+\kappa)\,z_{t-2}$ representing the unarbitraged component of the pricing errors from the previous period. Moreover, δ measures the impact of the momentum effect with respect to the contemporaneous fundamental changes while γ presents the short-run momentum effect. Finally, notice that the overall speed of convergence to equilibrium is determined jointly by κ and λ, namely $\kappa + \lambda(1+\kappa)$, implying that the standard one-period ECM is likely to be biased and misleading in the case where $\lambda \neq 0$.

Our key theoretical predictions suggest that arbitrage activity is fundamentally nonlinear, crucially depending on the magnitude of mispricing errors, as described in Section 2. By construction the linear model cannot test our hypotheses because it imposes (potentially invalid) symmetry restrictions and is thus likely to yield misleading results. Accordingly, in our empirical application, we choose to embed the GECM within the Markov switching model popularized by Hamilton (1989). In particular, we consider a three-regime empirical setup, which is compatible with our theoretical model with two alternative paths to equilibrium:

$$\Delta f_t = \alpha_{R_j} + \kappa_{R_j}\hat{z}_{t-1} + \lambda^*_{R_j}\hat{z}_{t-2} + \delta_{R_j}\Delta f^*_t + \gamma_{R_j}\Delta f_{t-1} + u_{tR_j},\, u_{tR_j} \sim iid\left(0, \Sigma_{R_j}\right), \tag{2.7}$$

where f_t is the natural log of the futures contract price, f^*_t is the natural log of the fundamental value, and $\{\alpha_{R_j}, \kappa_{R_j}, \lambda^*_{R_j}, \delta_{R_j}, \gamma_{R_j}\}$ are regime-dependent parameters, with $\sum_{R_j}$ being the regime-dependent covariance of the residuals. The pricing error, $\hat{z}_t$, is estimated from the following long-run equation[13]:

$$f_t = \mu + \theta f^*_t + z_t \tag{2.8}$$

The regime-specific noise momentum coefficient, λ_{R_j} can be obtained from $\lambda^*_{R_j} = \lambda_{R_j}\left(1+\kappa_{R_j}\right)$.

We will estimate the MS-GECM with three regimes where R_j is a scalar geometric ergodic Markov chain with a 3-dimensional-regime space, having the following transition matrix:

$$\begin{bmatrix} P_{11} & P_{21} & P_{31} \\ P_{12} & P_{22} & P_{32} \\ P_{13} & P_{23} & P_{33} \end{bmatrix},$$

where $P_{ij} = Pr\left(R_i \mid R_j\right)$ is the transition probability from State j to State i.

[13] According to the cost-of-carry model, the theoretical value of θ is 1, which is strongly supported by our empirical analysis.

	Mean	Median	Minimum	Maximum	Std Dev	AutoCor
Δs	0.028	0.058	-22.833	10.957	1.169	-0.039
Δf	0.028	0.062	-33.700	17.749	1.262	-0.040
$f-s$	0.243	0.196	-11.027	2.958	0.541	0.813
$f-f^*$	0.055	0.056	-11.451	2.767	0.294	0.420
r	3.380	3.790	0.000	9.100	2.493	0.999
q	2.275	2.080	1.070	4.100	0.718	0.999

Table 2.1: Basic descriptive statistics

Table 2.1 reports the the descriptive statistics for all variables. The sample used is the daily series of the S&P 500 index and its futures contract covering the period June 4, 1986 to December 3, 2015. There are a total of 7,442 observations. Δs (Δf) is the first difference of log spot (futures) price. $f^*_{t,T} = s_t + (r_t - q_t)\,\tau_t$ where s_t is the log spot price of the index, r_t is the annualized risk-free (3 month T-bill) interest rate on an investment for the period , q_t is the annualized dividend yield on the index, and τ_t is the time to maturity. All numbers are recorded in percentage point terms.

2.3.2 The data

We study the daily S&P 500 Index spot and futures contracts between June 1986 and December 2015[14]. All data are sourced from Datastream. Our proxy for the risk-free interest rate is the US 3-month Treasury bill (T-bill) rate. Divided yields on the indices are also collected. A continuous series of the nearest-term futures contracts is constructed. These series switch to the next nearest contract on the first day of the expiry month for the nearest term contract. We use a full set of daily price information for every contract to ensure correct matching of the date to maturity with the continued futures price series. Table 2.1 reports the descriptive statistics for all variables (measured in percentage terms).

As expected, the movements of the spot and futures prices closely mimic each other. The average price changes are of the same magnitude while the volatilities are higher in the futures contracts. The average basis (the log difference between futures and spot prices) is 24 basis points. After applying the cost-of-carry model, the difference between the futures price and the fair estimate ($f-f^*$) is 5.5 basis points.

[14]We did not use the early data from the period 1982–1986, since the estimated mispricing errors are more than double on average during this early period, compared to the period over 1986–2015. The index futures contracts were first introduced in 1982, where the market had higher transaction costs and lacked index arbitrage. Thus larger mispricing errors occurred. Errors then became more stable after 1986, and fluctuated with major market events. See also Figure *3.1* in Appendix B for a plot of the moving-average mispricing error through time.

2.3.3 Main empirical results

The MS-GECM estimation results are reported in Table 2.3.3 with three (smoothed) regime probabilities plotted in Figure 2.2. We first discuss the stylized feature of three distinct regimes from Figure 2.2, which we call States 1, 2 and 3, respectively. State 1 is the dominant market state, with the smallest mispricing error (measured as the absolute value of the deviations: $\mid z_{t-1} \mid = 0.103$) and volatility ($\Sigma = 0.113$) in Panel A and the highest ergodic probability (53%) in Panel C. It covers three major bull markets during 1992–1995, 2003–2007 and 2012–2015, and thus we call it the normal market state. State 2 covers 44% of the sample, with mispricing error (0.207) and volatility (0.239) at twice the levels seen in State 1. This state corresponds to the periods 1986–1991, 1996–2002 2009 and 2011, which are mostly the transition periods between bull and bear markets. We refer to this period as the transition market state. Finally, State 3 is characterized by extremely large mispricing error (0.774) and volatility (1.077)[15]. It covers only 2.4% of the sample, and coincides mostly with the stressed episodes that are captured in our sample period including the stock market crash in 1987, the Russian financial crisis in 1998, the market meltdown in 2001 and the global financial crisis in 2008. We call this the extreme market state.

Moreover, State 1 is most persistent with 98% transition probability, followed by State 2 with 97% and State 3 with 86% (see Panel C in Table 2.3.3), which suggests that State 1 (3) is the most (least) 'sticky'. The transition probabilities between States 1 and 3 in either direction are nil, confirming that State 2 is indeed the transition market state. The three distinct market states identified by the MS-GECM are mostly consistent with the different historic episodes we have observed during the whole sample period.

Linking the findings in Table 2.3.3 to our key predictions, we find that the estimated mispricing correction parameters, κ_{R_j}, are all negative and significantly less than unity in the absolute sense, in all three regimes. State 2 displays the fastest initial correction (82%), followed by State 1 (70%) and State 3 (60%). Comparing the difference between κ coefficients across the three different market conditions, it appears that arbitrageurs play a bigger role in bringing the price back to its fundamental value when switching from States 1 to 2, with the difference (–12%) being statistically significant. By contrast, the coefficient differential between States 2 and 3 becomes significantly positive (22.4%), suggesting that arbitrage activity is rather

[15] Notice that the average mispricing error over the whole sample period is 0.161.

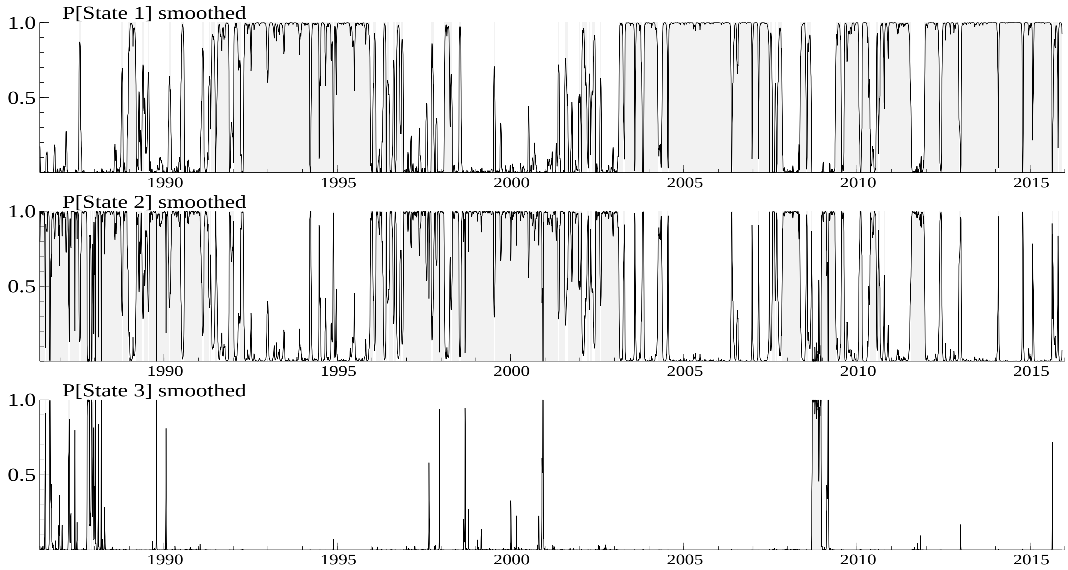

Figure 2.2: The smoothed regime probabilities: the three regime Markov Switching Generalized Error Correction Model

Figure 2.2 plots the smoothed regime probability of being in State 1, 2 and 3. State 1 consists of the years 1991-1996, 2003-2007 and 2012-2015, State 2 consists of the years 1986-1991 and 1996-2002, while State 3 are found in the years of 1987, 1997, 1998 and also 2007-2008.

Panel A: Estimation Results										
	State 1		State 2		State 3		State 2-1		State 3-2	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
α	-0.005***	-2.51	0.016***	3.34	0.058	0.66	0.022***	4.15	0.042	0.64
δ	0.991***	332.	1.024***	259.	1.142***	48.3	0.031***	6.36	0.118***	4.93
γ	-0.004	-1.38	0.015***	4.18	0.092***	3.58	0.018***	4.11	0.078***	2.98
κ	-0.699***	-36.7	-0.819***	-41.5	-0.596***	-6.56	-0.120***	-4.38	0.224**	2.41
λ^*	0.113***	12.8	0.167***	9.24	0.113	1.29	-0.079***	-3.00	-0.053	-0.59
Σ	0.113***	59.1	0.239***	43.2	1.077***	13.2	0.126***	21.4	0.838***	10.2
SOA	0.453***	19.1	0.652***	27.6	0.482***	8.73	-0.199***	-5.98	0.171***	2.84
$\mid z_{t-1} \mid$	0.103		0.207		0.774					
Panel B. Recovered Coefficients										
ω	-0.008**	-2.67	0.024***	5.94	0.142***	5.99	0.032***	6.35	0.118***	4.93
π	-0.474	-1.24	-0.630***	-3.42	-0.652***	-3.22	-0.156	-0.36	-0.022	-0.09
λ	0.817***	8.96	0.922***	5.86	0.279	1.03	0.107	0.58	-0.647***	-4.92
Panel C. Matrix of Markovian Transition Probabilities										
	State 1_{t-1}		State 2_{t-1}		State 3_{t-1}					
State1_t	0.980		0.024		0.000					
State2_t	0.019		0.968		0.143					
State3_t	0.000		0.008		0.857					
Ergodic	0.531		0.443		0.024					

Table 2.2: Estimation of the Markov-Switching Generalized Error Correction Model

Table 2.3.3 reports the estimation of the Markov Switching Generalized Error Correction Model. The sample used is the daily series of the S&P 500 index and its futures contract covering the period June 4, 1986 to December 3, 2015. There are a total of 7,442 observations, of which 4,070 , 3,222 and 148 fall into State 1, State 2 and State 3. Specifically, Panel A reports the estimation results for:

$$\Delta f_t = \alpha_{R_j} + \kappa_{R_j}\hat{z}_{t-1} + \lambda^*_{R_j}\hat{z}_{t-2} + \delta_{R_j}\Delta f^*_t + \gamma_{R_j}\Delta f_{t-1} + \mu_{tR_j}$$

where $\hat{z}_t$ is estimated from equation 2.8, $\{\alpha_{R_j}, \delta_{R_j}, \gamma_{R_j}, \kappa_{R_j}, \lambda^*_{R_j}\}$ are state-dependent coefficients with the covariance of the residuals (Σ_{R_j}) taking different values across the two states. Panel B reports the recovered coefficients. Specifically, ω_{R_j} is recovered by $\delta_{R_j} - 1$; $\pi_{R_j} = -\gamma_{R_j}/\omega_{R_j}$ and $\lambda_{R_j} = \lambda^*_{R_j}/(1+\kappa_{R_j})$. The final two columns in Panels A and B report the difference in estimated coefficients and associated t-statistic between States 1 and 2. For non-linear combinations of the coefficients, a delta method is applied to obtain the variance of the recovered coefficients and their differences. Panel C reports the transition and ergodic probabilities. ***, ** and * indicate significance at 1%, 5% and 10% levels, respectively.

limited even though the mispricing error in State 3 is 3.7 times higher than in State 2. These findings provide strong support for Prediction 1 that initial arbitrage activity follows the inverse U-shaped pattern with respect to the size of mispricing errors.

Next, we turn to the noise momentum coefficients, λ_{R_j}, reported in Panel B. Notably, we find that mispricing persistence is significant and relatively high during normal and transition market states, implying that the unarbitraged error coming from the previous period is highly persistent, respectively at 82% and 92% in the current trading period. By contrast, noise momentum is substantially smaller (28%) during the extreme market state, though the coefficient is not statistically significant. The difference in λ coefficients between States 2 and 1 is insignificant and negligible, whilst the difference between States 3 and 2 is significant and negative (–65%). Overall, this finding provides some support for Prediction 2, on the regime-dependent differences in mispricing persistence. Especially, it highlights that λ is another important parameter in characterizing the overall speed of adjustment process.

Combining both mispricing correction and noise momentum coefficients, we find that overall speeds of adjustment (given by $\kappa + \lambda(1+\kappa)$) are 45%, 65% and 60% respectively, for the normal, transition and extreme regimes. The overall speed of adjustment becomes significantly faster when the market switches from States 1 to 2, mainly due to κ (i.e., the increment of mispricing correction). The combination of initial correction and size of mispricing suggests that the capital allocation effect is the main driver behind the different adjustment speed between these two states. By contrast, when we move from States 2 to 3, we observe that both κ and λ play a role in changing the overall speeds of adjustment. This evidence prompts new insights into the cause(s) of a prolonged error correction process, which in previous literature is often explained by the presence of transaction costs (Bai and Collin-Dufresne 2013, Gyntelberg et al. 2016, Roll et al. 2007). We show that arbitrageur funding constraints play an important role in delaying price convergence through their initial error correction. Contrary to common perception, we find noise momentum is lower in extreme market conditions. In other words, arbitrageurs expect mispricing to be reduced rather than enlarged, which is supported by the empirical evidence of significant negative feedback trading during the period. This finding suggests that funding constraints are the main driver behind the outcome that overall arbitrage activity is significantly deterred in State 3.

Furthermore, our finding confirms that arbitrageurs tend to adopt the partial investment strategy during normal and transition market states, which are characterized by relatively high κ and λ values, whereas they are more likely to follow the full investment strategy during the extreme market state, which is characterized by relatively low κ and λ. In particular, our empirical findings during the extreme state are consistent with the slow-moving capital hypothesis in the literature (Mitchell, Pulvino and Pedersen 2007, Mitchell and Pulvino 2012, Schuster and Uhrig-Homburg 2015).

We also observe a range of notable findings for the other parameters in our model. First, the intercepts, α_{R_j}, in States 1 and 2 are statistically significant. A large positive intercept is found in State 2, while a negligible intercept is found in State 1 – the positive sign indicating a regime in which the futures price is more bullish than the spot price, other things equal. As such, this suggests that the transition state coincides with periods in which the futures market is more bullish than the spot market. On average, during this regime, there is a 16 basis-point daily return in the futures market, regardless of the spot market movement. However, such returns are accompanied by larger risks, as reflected by the variance of the futures return. Such a pricing difference between the two markets under different market conditions is not directly considered in previous theories, and thus adds a new result to the literature. Second, Table 2 shows that the contemporaneous market reaction coefficient, $w_{R_j} = \delta_{R_j} - 1$, is statistically different from zero, but small and negative in the normal state, while it is highly significant and positive in the transition and extreme states. In the case of the latter, it suggests that for a 1 percentage-point change (up or down) in the fundamental value there will be a 0.024 or 0.142 percentage-point price movement, respectively, in the futures market in the same direction (i.e., a 2.4 or 14 basis-point overreaction respectively). Third, Table 2 also shows that while in the normal market state there is no feedback trading (i.e., we cannot reject $\pi = 0$), there is large, negative and significant feedback trading in transition and extreme market conditions. Such negative feedback trading is consistent with the high volatility observed in these market conditions.

2.3.4 Linking the hidden states to observable

The advantage of using a Markov-Switching model is the ability to estimate the likelihood of being in a given latent state, which can then be examined for potential linkages to various observable economic factors. It offers an opportunity to better understand and characterize what the states are really capturing. Particularly, we are interested in how well it classify the states according to the funding conditions.

	State 1		State 2		State 3	
Variables	Mean	Median	Mean	Median	Mean	Median
Fund Flows						
Hedge Fund Flows	10.22	7.92	5.43	4.99	1.83	0.14
Active Fund Flows	130.72	135.69	74.04	78.42	40.40	51.07
Index Fund Flows	9.49	8.66	6.81	6.29	9.13	9.10
Capital Structure						
Growth Rate of Total Financial Asset	0.02	0.03	0.05	0.04	-0.01	-0.03
Financial Sector Leverage	7.89	3.60	12.09	5.21	30.94	38.87
Broker-Dealers' Leverage Factor	43.59	42.21	39.54	39.68	61.81	59.38
Liquidity						
Aumihud Illiquidity of SPX 500	28.29	2.76	85.59	14.55	71.26	3.78
Treasury Security-based Funding Illiquidity	-0.21	-0.23	0.16	0.25	1.81	1.96
TED Spread	0.00	0.00	0.01	0.01	0.03	0.03
Volatility						
VIX	15.95	15.22	24.24	23.30	52.35	55.31
Idiosyncratic risk	0.03	0.02	0.03	0.03	0.05	0.05

Table 2.3: Linking the hidden states with observables

This table reports the mean and median statistics of monthly funding and liquidity measures in the three States. Definition of the variables are given in Appendix XX. Months are to given States by finding the which state has large number of days being the dominant state (largest probability). We have 189, 160 and 6 months of observations for States 1, 2 and 3 respectively.

Table 2.3 reports the mean and median statistics of monthly funding and liquidity measures in the three States. Months are assigned to a given States by finding which state has the largest number of days being the dominant state (the largest probability among the three States). We have 189, 160 and 6 months of observations for States 1, 2 and 3 respectively. Definition of the variables are given in Appendix 2.B. We include the variables of aggregated hedge and mutual fund flows, capital structure and stock, bond market liquidity and volatlity.[16]

Fund flows statistics suggests that funds available for arbitrageurs (hedge and active mutual funds) drop as volatility increases from States 1 to 3. The decreasing funding confirms that funding constraints are more likely to be binding in the extreme regime when inflows to hedge fund and active funds are at their lowest. By contrast, passive index received larger inflows during this same extreme period which suggests that funding constraints are partly due to relocation of funds within equity focused funds (flight to quality/liquidity). The results of growth rate of the total financial asset and leverage factors points to a similar story that funding constraint is most binding in States 3. The financial asset growth is at its lowest (in fact negative) while financial sector leverage and broker-dealer leverage are at its highest. The differences between States 1 and 2 are small relative to their differences from State 3, which justifies State 3 to be the most extreme state of the three.

For market liquidity measures, Amihud illiquidity measure suggests that it plays little role in affecting arbitrageurs' decision between different states. In general, we expect that illiquidity of spot market would deter arbitrage. We observe the contrary considering States 1 and 2. Although the spot market is more illiquid in States 2 we observe larger error correction. Rather market illiquidity is more likely to be affected by funding constraints (Brunnermeier and Pedersen, 2009), such that more binding funding constraints in States 2 and 3 exhibit severer market illiquidity. The bond illiquidity and the TED spread as a measure of funding liquidity risk (Fontaine and Garcia, 2011) confirm that funding liquidity risk is at its highest in State 3. The difference between States 2 and 3 is much higher than that between States 1 and 3.

Finally, the VIX index confirms that the volatility increases from States 1 to 3 with State 3 having a value of more than double of the State2. The idiosyncratic risk of index constituents is also highest in State 3 and indifferent between States 1 and 2. Higher volatility in State 3 (systematic or idiosyncratic), instead of deterring the

[16]I acknowledge the help from Professor Charlie X. Cai from University of Liverpool for providing the data in this section.

initial reaction of arbitrage, tends to make it more difficult for arbitrageurs to raise funding and thus correct mispricing, as suggested by Brunnermeier and Pedersen (2009).

Overall, the observables confirm that our regime estimation captures the variations of funding constraint and therefore supports our hypothesis that funding constraint is an important driver of the variations in arbitrage.

2.4 Conclusions

We develop a unified approach to generate our main theoretical predictions regarding the effects of both capital allocation and funding constraints on the limits to arbitrage. Building on the seminal work by Shleifer and Vishny (1997; SV above), we analyze the nonlinear impacts of mispricing on arbitrage activity. To replicate the situation where arbitrageurs attempt to exploit price inefficiency while simultaneously facing market frictions and funding constraints, we model arbitrageurs as facing a trade-off between making investments now and waiting for larger arbitrage opportunities in the subsequent period. To capture such multi-period arbitrage activity, we follow the framework advanced by Cai, Faff and Shin (2017; elsewhere CFS) and investigate the impacts of mispricing on both the initial mispricing correction and the subsequent noise momentum, where the latter is designed to measure the persistence of uncorrected pricing errors.

We investigate these issues under two alternative paths to equilibrium: in the normal state, where funding constraints are not binding, arbitrageurs are more likely to adopt the partial investment strategy and thus larger mispricing will induce faster arbitrage activity; and in the extreme state, characterized by extremely large mispricing error, arbitrageurs are faced with severely binding funding constraints and are more likely to adopt the full investment strategy and funding constraints becomes the dominant factors of limits to arbitrage, and thus larger mispricing deters the arbitrageurs' initial mispricing correction. In sum, our theory suggests that overall arbitrage activity does not rise linearly with mispricing error. Rather, the relationship tends to be regime-dependent, and overall arbitrage activity displays an inverse U-shape against the size of mispricing error. Our approach is thus able to indicate whether it is arbitrage cost/risk or funding constraint that predominantly limits arbitrage activity.

To test the empirical validity of our theoretical predictions, we extend the general error correction model by CFS to the state-dependent Markov switching model. Applying this model with three regimes to the S&P 500 Index spot and futures markets over the period 1986–2015, we find strong evidence in favor of regime-dependent nonlinear limits to arbitrage. Furthermore, we can identify the stress periods of binding funding constraint as the years 1987, 1998, 2001 and 2008. In particular, our study provides both theoretical and empirical evidence that is consistent with the slow-moving capital hypothesis documented in the literature (Mitchell, Pulvino and Pedersen 2007, Brunnermeier and Pedersen 2009).

The potential applications of this approach go well beyond that developed in the current paper. For example, our approach could be applied to explore the short-term dynamics associated with fundamental long-run co-integrating relationships (e.g., the price-dividend relationship) and the pricing dynamics between segmented markets for single assets (e.g., cross-listing and commodity contracts in different markets). As our paper highlights the fruitful results of studying the limits to arbitrage via arbitrage activity, another important extension would be to analyze the cross-sectional effects of specific arbitrage impediments to arbitrage activity, such as fundamental risk and non-fundamental risk. We commend these and other meaningful extensions to the future research agenda.

Appendix 2.A The Size of The Mispricing Error over Time

Figure 2.3 plots the moving average of mispricing error in absolute value, covering the full sample period from 1982 to 2015. It is easily seen that mispricing error is large and volatile in early period before 1986. It fluctuates above 0.25, and can reach as high as 2.25 in extreme. At the time when index future contracts are first introduced in 1982, the market is characterized with high transaction costs, low number of participating arbitrageurs and low level of available arbitrage capital. Therefore larger mispricing error tend to occurs during the early periods. Over time knowledge diffused and entry barriers and implementation costs dropped. Mispricing becomes more stable after 1986, and comoves with major market events. It tend to stay below 0.25 at most of the sample period, while only exceed at a few extreme market events, such as the 1987 market crash and 2007-2009 global financial crisis.

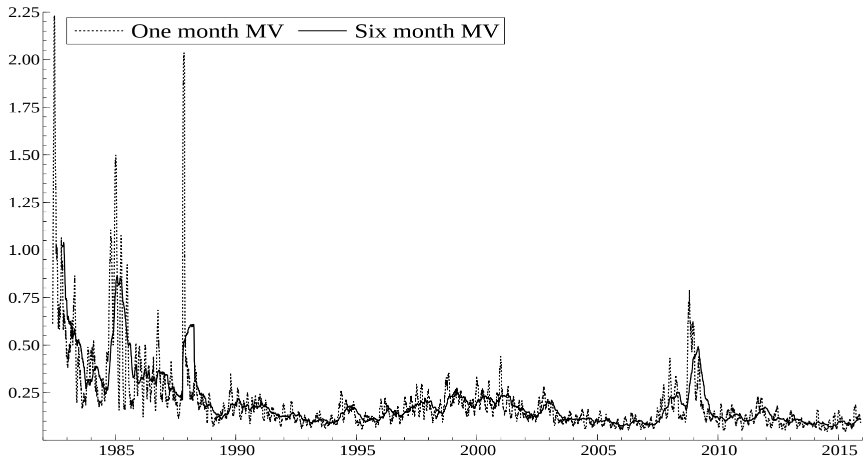

Figure 2.3: The plot of the moving-average mispricing error over 1982-2015
The figure plots the moving average of the absolute value of the pricing error, $\hat{z}_t$, estimated from the long-run equation:

$$f_t = \mu + \theta f_t^* + z_t,$$

where f_t and f_t^* is the spot price of S&P 500 future contract and the fundamental price implied by the cost of carry model, respectively. The dotted line plots the one-month moving average mispricing error, while the solid line plots the six-month moving average mispricing error.

Appendix 2.B Definitions of the Observable Variable

In this appendix, we introduce a numbers of variables as the funding and market liquidity measures.

1. Aggregate Hedge fund flows, Aggregate index fund flows, Aggregate mutual fund flow
   Ang, Gorovyy, Inwegen (2011) find that hedge fund leverage decrease prior to the leverage of financial intermediaries. They also find that funding cost and fund return volatility can negatively predict fund leverage, while the market value of hedge funds positively predict fund leverage. Empirically, the aggregate hedge fund flow(past three-month flows) is positively predicting the gross leverage and long-only leverage. We follow Ang , Gorovyy, Inwegen (2011) to construct aggregate actively-managed US mutual fund flows , aggregate passively-managed US index fund flows , aggregate US hedge fund flows . Under the US mutual fund universe of Morningstar Direct, we select the mutual funds with the indicator ‘index funds’ stating ‘No’ and the indicator

'oldest share class' stating 'Yes' to remove identical funds with different share class, the Global broad category stating "Equity " or "Allocation", and remain 5152 funds as our mutual fund sample. We select the index funds with the indicator 'index funds' stating 'No' and the indicator 'oldest share class' stating 'Yes' to remove identical funds with different share class, the Global broad category stating "Equity " or "Allocation", and remain 291 funds as our index fund sample. Under the global hedge fund universe of Morningstar Direct, we select the indicator 'Domicile' stating 'United States' and remain 1451 fund as our hedge funds sample. All the return and total net asset data are download on a monthly basis from Jan 1976 to June 2016. The monthly mutual fund/index fund/hedge fund flows are constructed as follow:

$$flow_{i.t} = \frac{TNA_{i,t}}{TNA_{i,t-3}} - (1 + r_{i,t-2})(1 + r_{i,t-1})(1 + r_{i,t})$$

$$Aggreflow_t = \sum_{i=1}^{k} flow_{i,t}$$

where $TNA_{i,t}$ is the total net asset of fund i in quarter t, $r_{i,t}$ is the total return of fund i in quarter t from Morningstar Direct database.

2. Idiosyncratic risk

Ang, Hodrick, Xing, and Zhang (2006) find that stock with high idiosyncratic risk relative to Fama and French 3 factor model (1993) have low average return. Neither aggregative volatility risk, size, book-to-market, momentum, and liquidity can explain this phenomenon. We follow Ang, Hodrick, Xing, and Zhang (2006), first obtain the daily stock return and the SP500 index return from CRSP, then obtain the daily risk –free return from the website. Fama and French Data Library. We calculate the idiosyncratic risk of each stock in each month and then take the mean of the monthly aggregate idiosyncratic risk of all US stocks. The idiosyncratic risk is constructed as follow,

$$Ret_{i,t} - Ret_{riskfree,t} = \alpha_t + \beta_t \left(Ret_{spx500,t} - Ret_{riskfree}\right) + \varepsilon_{i,t}$$

$$idiorisk_{i,t} = Std\left(\varepsilon_{i,t=1}, \varepsilon_{i,t=2}, \varepsilon_{i,t=3}, \ldots\ldots \varepsilon_{i,t=end\,day\,of\,month}\right)$$

where $Ret_{i,t}$ is the daily return of stock on day i at month t, $Ret_{riskfree,t}$ is daily return of the 30-day T Bill return from Fama-French website and $\varepsilon_{i,t=1}$ is the residual obtained from the regression in day 1.

3. Amihud (2002) Illiquidity measure

   Amihud finds (2002) that illiquidity positively predicts cross-sectional stock excess return, especially small-cap stocks. We follow Amihud (2002) to construct the illiquidity factor of stocks as follow:

$$ILLQ_{i,y} = \left(\frac{1}{D_{i,y}}\right) \sum_{k=0}^{D_{i,y}} \frac{\mid R_{i,y,d} \mid}{VOLD_{i,v,y,d}}$$

   where $D_{i,y}$ is the number of days in year y that are available of stock i, $R_{i,y,d}$ is the daily return on day d in year y of stock i, $VOLD_{i,v,y,d}$ is the trading volume in dollars in day d of stock i.

4. Institutional ownerships

   Ali, Hwang and Trombley (2013) argue that large institutional ownership will reduce the predictive power of B/M ratio on stock return. It implies that large institution ownerships might increase arbitrage cost.

$$insown_{i,t} = \sum_{k=0}^{n} \frac{insshares_{i,j,y}}{outssares_{i,t}}$$

   where $insshares_{i,j,y}$ is the shares of stock i held by institution j at month t and $outssares_{i,t}$ is outstanding shares of stock i at month t.

5. Financial asset growth

   Adrian and Shin (2010) find that the growth of financial asset measures the increase of aggregate liquidity, the fast growth of asset will increase their "surplus capital", they seek to expand this captial and search for borrowers. Then Aggregate liquidity comes from urging to people borrow the money, though they might have no ability to repay. The asset growth is constructed as follow:

$$assetgrowth_{i,t} = \frac{asset_{i,t}}{asset_{i,t-1}} - 1$$

   where $asset_{i,t}$ is the total financial asset of industry i at quarter t.

6. Broker-dealer leverage

   Ang, Gorovyy and Van Inwegen (2011) argue that hedge fund leverage is counter-cyclical to the market leverage of listed financial intermediaries. In

2007 mid, hedge fund leverage decrease prior to start of 2007-mid financial crisis, when listed investment banks and finance sector continues to increase. The leverage is constructed as follow:

$$Leverage_t^{BD} = \frac{Total\,Financial\,Asset_t^{BD}}{Total\,Financial\,Asset_t^{BD} - Total\,Liabilities_t^{BD}}$$

where $Total\,Financial\,Asset_t^{BD}$ is the aggregate quarterly total financial of security broker-dealers and $Total\,Liabilities_t^{BD}$ is the aggregate quarterly total financial liability of security broker-dealers reported.

7. Treasury security-based funding liquidity
   Fontaine and Garcia (2011) document that the liquidity premium also shares a funding component with risk premia in another market. High liquidity indicates high cost in money supplying. The data is directly obtained from Jean-Sebastien Fontaine's website.

8. TED spread
   The TED spread is the three-month Eurodollar deposits yield (LIBOR) subtracted by three-month US T-bills. Both LIBOR and T-bills yields are monthly data downloaded from the FRED data library. The Ted is constructed as follow:

$$TED_t = Yield_{EU,t} - Yield_{US,t}$$

   where $Yield_{EU,t}$ is the yield of the three-month Eurodollar deposits yield (LIBOR) and $Yield_{US,t}$ is the yeield of three-month US T-bills.

Appendix 2.C Proofs

In this Appendix we provide the proofs to the main propositions derived in Section 2.2. We first derive the stability condition for sensitivity α to ensure that equity in period 2 is nonnegative after mispricing deepens in the second period. We denote the equity in period 2 when mispricing deepens by F_2^*, which can be written as

$$F_2^* = F_1 \left[1 + \alpha\beta_1 \left(\frac{P_2^*}{P_1} - 1 \right) \right] \tag{2.9}$$

where $P_2^* = V - S_2^* + F_2^*$ and $P_1 = V - S_1 + \beta_1 F_1$. We simplify (2.9) as

$$F_2^* = F_1 + F_1\left[\frac{a\beta_1 F_1(1-\beta_1) - a\beta_1(S_2^* - S_1)}{V - S_1 - (a-1)\beta_1 F_1}\right]. \tag{2.10}$$

After rearrangement, it is easily seen that the inequality $F_2^* > 0$ holds when

$$\alpha < \frac{V - S_1 + \beta_1 F_1}{\beta_1\left(S_2^* - S_1 + \beta_1 F_1\right)}.$$

As β_1 can take value between 0 and 1, we rearrange the condition in terms of V:

$$V > (S_1 - \beta_1 F_1) + a\beta_1(S_2^* - S_1 + \beta_1 F_1).$$

For different value of β_1, the maximum of RHS is obtained by either $\beta_1 = 0$ or $\beta_1 = 1$. Thus it is easily rewritten as

$$V > max\{(S_1 - F_1) + a(S_2^* - S_1 + F_1), S_1\}.$$

which can be simplified as $V > (S_1 - F_1) + a(S_2^* - S_1 + F_1)$, since we always have $V > S_1$. Thus the stability condition is

$$\alpha < \frac{V - S_1 + F_1}{S_2^* - S_1 + F_1}. \tag{2.11}$$

We note that the stability condition is not a strong restriction, and it holds in most general cases. Since the fundamental value is assumed to be a lot larger than the shocks, i.e. $V \gg S_2^*$, thus even for the extreme value of $\alpha \gg 1$, the stability condition still holds and guarantee $F_2^* > 0$. However, the case with $\alpha \gg 1$ rarely happens in real world, as financiers will not give such award (punishment) to arbitrageurs who perform well (badly).

Second, we provide two Lemmas, which will be used in the proof of main propositions.

Lemma 1. *Consider the model setup from Section 2.2.1 and 2.2.3, and the equilibria from Eq. (2.2). Suppose that arbitrageurs adopt the full investment strategy,* $\beta_1 = 1$. *Under the stability condition,* $1 < \alpha < (V - S_1 + F_1)/(S_2^* - S_1 + F_1)$, *then as* S_1 *rises, the fund in period 2,* F_2^* *increases.*

Proof. From Eq. (2.10), we obtain the partial derivative of F_2^* with respect to S_1:

$$\frac{\partial F_2^*}{\partial S_1} = a\beta_1 F_1 \frac{V - S_2^* + (1 - a\beta_1) F_1}{(V - S_1 - (a-1)\beta_1 F_1)^2}.$$

Replacing $\beta_1 = 1$, we need to verify that whether the numerator $V - S_2^* + (1-a) F_1$ is positive or not. Using the stability condition, $V > aS_2^* + (a-1)(F_1 - S_1)$ and subtracting S_2^* from both side of inequality, then we have:

$$V - S_2^* > (a-1)(S_2^* + F_1 - S_1) > (a-1) F_1, \tag{2.12}$$

which shows that $V - S_2^* + (1-a) F_1 > 0$. **QED**

Lemma 2. *Consider the model setup from Section 2.2.1 and 2.2.3, and the equilibria from Eq. (2.2). Suppose that arbitrageurs adopt the partial investment strategy* $\left(0 < \hat{\beta}_1 < 1\right)$. *Under the stability condition,* $1 < \alpha < (V - S_1 + F_1)/(S_2^* - S_1 + F_1)$, *then the sign of* $\frac{\partial P_1}{\partial S_1}$ *is the same as that of* $\frac{\partial P_2^*}{\partial S_1}$.

Proof. We rewrite (2.2) as

$$P_2^* = \frac{qVP_1}{V - (1-q)P_1}$$

Taking the first differentiation with respect to S_1, then

$$\frac{\partial P_2^*}{\partial S_1} = qV\left[\frac{\theta\,(V - (1-q)P_1) + \theta\,(1-q)P_1}{(V - (1-q)P_1)^2}\right] = \frac{\theta q V^2}{(V - (1-q)P_1)^2}$$

where $\theta = \frac{\partial P_1}{\partial S_1}$, which clearly shows that sign of $\frac{\partial P_{2^*}}{\partial S_1}$ is the same as that of $\frac{\partial P_1}{\partial S_1}$. **QED**

Now, we provide the proof for Proposition 1:

Proof. First, we consider the partial investment strategy with $0 < \hat{\beta}_1 < 1$. Before we show $\frac{\partial \kappa}{\partial S_1} > 0$, we reveal the derivations of $\frac{\partial \hat{\beta}_1}{\partial S_1}$, which contains some tedious calculation, and reach a number of properties.

The equilibrium strategy $\hat{\beta}_1$ can be obtained from the first order condition, such that

$$(1-q)\left(\frac{V}{P_1} - 1\right) + q\left(\frac{P_2}{P_1} - 1\right)\frac{V}{P_2} = 0 \tag{2.13}$$

and it is expressed as the follows

$$\hat{\beta}_1 = \frac{n_1 - n_3}{2a(1-q)F_1}$$

$$
\begin{aligned}
n_1 &= V + (1-q)\left(F_1 + aS_1 - S_2^*\right) \\
n_2 &= V + (1-q)\left(F_1 - aS_1 - S_2^*\right) \\
n_3 &= \sqrt{(n_2)^2 + 4aVq(1-q)\left(S_2^* - F_1\right)}
\end{aligned}
$$

The partial derivative of $\hat{\beta}_1$ with respect to S_1 can be expressed as

$$
\frac{\partial \hat{\beta}_1}{\partial S_1} = \frac{1}{2F_1}\left(1 - \frac{n_2}{n_3}\right) > 0. \tag{2.14}
$$

The inequality holds since $n_3 > n_2$ under the assumption that $F_1 < S_1 < S_2^b$. This confirms the positive capital allocation effect. In addition, the partial derivative of P_1 with respect to S_1 can be expressed as

$$
\frac{\partial P_1}{\partial S_1} = \frac{1}{2}\left(1 - \frac{n_2}{n_3}\right) - 1 < 0, \tag{2.15}
$$

where the inequality holds since $n_3 > n_2$ under the assumption that $F_1 < S_1 < S_2^b$.

Now consider the partial differentiation of $\kappa = \frac{\hat{\beta}_1 F_1}{S_1}$ with respect to S_1 given by

$$
\frac{\partial \kappa}{\partial S_1} = F_1\left(\frac{\frac{\partial \hat{\beta}_1}{\partial S_1}S_1 - \hat{\beta}_1}{S_1^2}\right),
$$

and we derive the condition that ensures the numerator $\frac{\partial \hat{\beta}_1}{\partial S_1}S_1 - \hat{\beta}_1$ is positive with Eq. (2.14):

$$
V > (1-q)\left\{\left(S_2^* - F_1\right) + a\left[\frac{\left(S_1 - \beta_1 F_1\right)^2}{\beta_1 F_1} + \beta_1 F_1\right]\right\} \tag{2.16}
$$

The RHS in Eq. (2.16) takes the smallest value when $\beta_1 = 1$, and the condition becomes

$$
\begin{aligned}
V &> \left(S_2^* - F_1\right) + a\left[\frac{\left(S_1 - F_1\right)^2}{F_1} + F_1\right] \\
&= S_2^* - F_1 - 2\alpha\left(S_1 - F_1\right) + \alpha\frac{S_1^2}{F_1} \qquad (2.17)
\end{aligned}
$$

Given a moderate setting of α that accords with the stability condition and the assumption that $V \gg S_1, S_2^*, F_1$, the condition in Eq. (2.17) can easily hold true,[17] and $\frac{\partial \kappa}{\partial S_1} > 0$.

Furthermore, under the partial investment, we rearrange the FOC in Eq. (2.13), and express λ as

$$\lambda = q\frac{V - P_2^*}{V - P_1} = \frac{P_2^*}{P_1} = q + P_2^*\frac{1-q}{V} \tag{2.18}$$

Together with Lemma 2 and Eq. (2.15), we have $\frac{\partial P_2^*}{\partial S_1} < 0$. Therefore, it is easily seen from Eq. (2.18) that as S_1 rises, λ falls. This proves $\frac{\partial \kappa}{\partial S_1} > 0$ and $\frac{\partial \Lambda}{\partial S_1} < 0$ for partial investment equilibrium.

Next, we consider the full investment, $\hat{\beta}_1 = 1$, in which case

$$\kappa = \frac{\hat{\beta}_1 F_1}{S_1} = \frac{F_1}{S_1}.$$

Then, it is easily seen that $\frac{\partial \kappa}{\partial S_1} < 0$, given F_1. Furthermore, we rewrite Eq. (2.5):

$$\lambda = q\frac{V - P_2^*}{V - P_1} = q\frac{S_2^* - F_2^*}{S_1 - F_1}$$

From Lemma 1, we have $\frac{\partial F_2^*}{\partial S_1} > 0$, which clearly shows that $\frac{\partial \lambda}{\partial S_1} < 0$. **QED**

Finally, we provide the proof of our Proposition 2:

Proof. From a modeling perspective, q^* is determined as a complex function of the parameter set, $\{V, S_1, S_2^*, F_1, \alpha\}$. q^* is derived at the point where arbitrageurs are indifferent between taking either strategy, i.e. $D_1 = F_1$ under partial investment strategy. We have

$$q^* = \frac{(S_1 - F_1)(V + F_1 - S_2^* - \alpha F_1)}{V\left(S_2^* - S_1\right) + (S_1 - F_1)\left(V + F_1 - S_2^* - \alpha F_1\right)} = \frac{1}{m+1}$$

where $m = V\,(S_2^* - S_1)\,/\,(S_1 - F_1)\,(V + F_1 - S_2^* - \alpha F_1)$.

[17] We note that there are extreme cases where Eq. (2.17) can be violated even when the stability condition holds, especially when sensitivity is extremely large, $\alpha \gg 1$. The reason is that model setup does not have further assumption about the rational range of value of sensitivity α, except for the stability condition. As we noted in the derivation of stability condition, some α that satisfy the stability condition can be rather extreme and impossible to occur in real world. Thus it is a rather weak assumption which cannot provide more rational restriction on sensitivity α.

First, it can be easily seen that $0 < q^{*} < 1$, since we have

$$S_1 - F_1 > 0, S_2^{*} - S_1 > 0, V - S_2^{*} + (1-\alpha) F_1 > 0 \tag{2.19}$$

from the model assumption and the implication from stability condition in Eq. (2.12).

Next, taking the partial derivation of m w.r.t S_1 and F_1 respectively, we have S_{2}^{*}-F_{1}-2\alpha\left(S_{1}-F_{1}\right)+\alpha\frac{S_{1}^{2}}{F_{1}}

$$\frac{\partial m}{\partial S_1} = \frac{V\left(F_1 - S_2^{*}\right)}{\left(S_1 - F_1\right)^2 \left(V + F_1 - S_2^{*} - \alpha F_1\right)} < 0$$

$$\frac{\partial m}{\partial F_1} = \frac{V\left(S_2^{*} - S_1\right)\left(V - S_2^{*} + (1-\alpha) F_1 + (a-1)\left(S_1 - F_1\right)\right)}{\left(S_1 - F_1\right)^2 \left(V + F_1 - S_2^{*} - \alpha F_1\right)^2} > 0$$

The inequality holds for Eq. (2.19). Thus, we have $\frac{\partial q^{*}}{\partial S_1} > 0$ and $\frac{\partial q^{*}}{\partial F_1} < 0$. **Q.E.D**

Chapter 3

Funding Liquidity and The Efficacy of Arbitrage

3.1 Introduction

The financial crisis during 2007-2009 and the subsequent great recession have put spotlight on the financial intermediaries, and the role of their financial health on both asset pricing and the whole economy. Among other intermediaries, hedge funds are often recognized as the sophisticated and rational arbitrageurs (called arbs in short hereafter), as they largely employ quantitative modeling in making investment decision and heavily use leverage in supporting their daily management. The relatively high use of leverage can enhance the hedge funds' ability to capitalize on mispricing opportunities and reduce pricing anomalies, which is referred to as "smart money" by Akbas et al. (2015). However, it also expose them to higher funding risks, such that as the funding problem propagates to the hedge funds, they will find it more difficult to manage their leverage and react to the transient discrepancy in market prices.[1] Therefore, the arbs' ability to obtain leverage funding, i.e. funding liquidity, plays an important role in asset pricing, mispricing correction and market efficiency.

The importance is rather nonlinear, depending on whether the tightening funding constraint is binding or not. When it is binding, a self-reinforcing process can occur and induce the amplification mechanism, such that even a modest trigger can result in large spillovers across the financial system. The most recent global financial crisis in 2007 provides a notable example, such that it is triggered by the collapse

[1] See literature on slow-moving capital: Mitchell, Pulvino and Pedersen (2007) Duffie (2010) Garleanu and Pedersen (2011) and Mitchell and Pulvino (2012).

of the subprime mortgage market, which constitutes only around 4% of the overall mortgage market. Shleifer and Vishny (1997) suggest the loss spiral, e.g. the arbs with poor performance may face funding withdrawal from the lenders and are forced to liquidate their position when funding constraint is binding, which leads to further price distortions and losses. Brunnermeier and Pedersen (2009) illustrate the margin spiral in extreme circumstance, such that losses in capital reduce the traders' ability to provide market liquidity, which will raise margin requirement and further jeopardize the traders' funding condition. Mitchell, Pedersen, and Pulvino (2007) and Mitchell and Pulvino (2012) provide a number of empirical evidence of extremely large pricing anomalies caused by slow-moving capital during the extreme periods.

Understanding and identifying when amplification will occur is vital for market participants and policy makers during the financial crisis (Brunnermeier and Oehmke, 2013). Extant literature has successfully measured the funding (il)liquidity through the available arbitrage capital (Comerton-Forde et al., 2008; Adrian and Shin, 2010; Akbas et al., 2016) and asset pricing effect (Fontaine and Garcia, 2011; Garleanu and Pedersen, 2011; Nagel, 2012; Frazzini and Pedersen, 2014). However, a key question they fail to address is how to identify when funding illiquidity becomes so severe that the amplification will occur, i.e. how to identify when the funding constraint becomes binding. Our paper attempts to fill this gap.

To do so, we narrow our focus on the arbitrage activity capitalizing on the mispricing opportunity, following Chapter 2 of this thesis. We have suggested in Chapter 2 that the arb's efforts to bear against mispricing error, i.e. mispricing correction, reveals the arbitrage impediments they face: arbitrage costs and funding constraints. To further understand the role of funding constraint, we augment the model of Shleifer and Vishny (1997, SV hereafter) and Stein (2009, Stein hereafter) by allowing the endogenous leverage constraint set by outside financiers. In our model, mispricings are generated by noise traders; Arbs enter the market to smooth price fluctuation and correct the mispricing error in the use of leverage; Leverage debt can be financed from outside financiers, who set a maximum leverage constraint to protect their own capital from the arb's potential insolvency. We derive the competitive equilibrium of the model where arbs capitalize on mispricings subject to the leverage constraint, and explore the implications on the arb's ability to raise leverage capital, i.e. funding liquidity, and to conduct arbitrage activity, i.e. arbitrage efficacy.

We evaluate the arb's funding liquidity by the marginal leverage debt raised by the arbs in order to bear against additional mispricings. Leverage is denoted as (de)stabilizing when the marginal leverage is (negative) positive, such that arbs are able to raise (less) more leverage debt with mispricings. Moreover we consider the efficacy of arbitrage, which is defined by the marginal percentage of mispricing correction achieved by the arbs against additional mispricings. Arbitrage is (in)efficacious when the marginal correction is (negative) positive, such that arbs are able to achieve higher mispricing correction with larger error. While funding liquidity and arbitrage efficacy are often viewed by the return-based definition in the literature, such that funding liquidity is defined by the shadow cost of capital (Brunnermeier and Pedersen, 2009; Garleanu and Pedersen, 2011) and arbitrage efficacy by the size of arbitrage violation (Akbas et al. 2016), our paper first captures them via the arbitrage activity and the follow-up analyses are proved fruitful.

Our model suggests that (i) when leverage constraint is loose, leverage is stabilizing, such that they are able to raise more funding with higher mispricing error. Sufficient funding enable them to achieve higher mispricing correction, i.e. arbitrage is efficacious. This is consistent with the existing view of risky arbitrage, such that arbitrage force tend to intensify with larger mispricing which represents a higher cost-adjusted return. (ii) However, extremely large mispricing makes leverage constraint binds, where arbs are forced to take the maximum leverage limit set by financiers. The arbs retain a positive but far lower marginal leverage when financiers are more informed and willing to offer a higher leverage limit as mispricing increases. However, destabilizing leverage may occur when financiers become less informed, and tend to interpret the initial mispricing as higher future uncertainty. This exaggerates their estimates of the future mispricings, and renders them unwilling to increase leverage supply. Instead of correcting mispricing, the arbs are forced to deleverage and cause more intense mispricing. (iii) Due to the binding leverage constraint, arbitrage becomes inefficacious regardless of the level of informativeness, such that arbs fail to achieve a higher percentage of correction. (iv) In the presence of stabilizing leverage and efficacious arbitrage, arbs are able to effectively absorb varying selling pressure and smooth volatile price, which provides liquidity to the market and prevents the asset trading at distressed price. Therefore, the expected market illiquidity and price volatility is insensitive to the innovations in mispricing errors. However, destabilizing leverage and inefficacious arbitrage can result in significant asset pricing effects, such that small changes in mispricing can trigger large grow in the expected market illiquidity and price volatility.

Our model thus links the arb's funding liquidity to the efficacy of arbitrage. More importantly, the sign of marginal correction, indicating whether arbitrage is effective or not, can be viewed as an early signal of the binding leverage constraint. It offers a great tool for policy makers to not only monitor the funding liquidity condition in the financial sector, but also guide the timing of public and monetary policy making.

After deriving the predictions arising from the possible equilibrium, we design an empirical strategy to capture the arbitrage efficacy (marginal correction) by the first-difference estimator: regressing daily difference of mispricing corrections on daily difference of mispricing errors, where mispricing correction can be estimated from the Generalized Error Correction Model.[2] We then apply the strategy to the index arbitrage between S&P500 index and E-mini future from September 1997 (the earliest possible time for E-mini future) to June 2015, and obtain the time-varying implied arbitrage efficacy, denoted as AE. The index and E-mini future arbitrage offers the following advantages. First, the index arbitrage can be easily verified with the cost of carry model; Second, the E-mini S&P500 future contract are one of the most traded future contracts in US, and contains a large number of hedge funds who aim to capitalize on the mispricing opportunity with low transaction costs, high market liquidity and low short-selling constraint; Third, the sample period over 1997-2015 accords with the boom of the hedge fund industry and some major liquidity events, such as the recent financial crisis in 2007.

Figure 3.1 illustrates some key findings in our empirical application, where we plot the implied arbitrage efficacy (solid line, left axis) along with the VIX index of implied volatility in S&P 500 index options (dashed line, right axis). We find that (v) the four major periods when the implied arbitrage efficacy was negative match the major liquidity turmoils during the sample period: the bust of dot-com bubble in 2000-2003, the financial crisis in 2007-2009, the debt ceiling and sovereign debt crisis in 2010-2011 and the debt ceiling crisis in 2013. For example, AE dropped sharply below zero in June 2007 and stayed negative during the financial crisis in 2007-2009. It rebounded back to zero only after the first round of quantitative easing was implemented by Federal Reserve. AE stayed positive during some notable exemptions, like the Flash Crash on May 2010, which was later proved to be a temporary freeze of liquidity. (vi) The effectiveness of Fed's monetary policy and lending facilities on funding liquidity provision was varying. While the first round of Quantitative Easing announced by Federal Reserve had significant improvement on the implied arbitrage efficacy, the second and the third rounds of QE leaded to a declin-

[2]See Cai, Faff and Shin (2015).

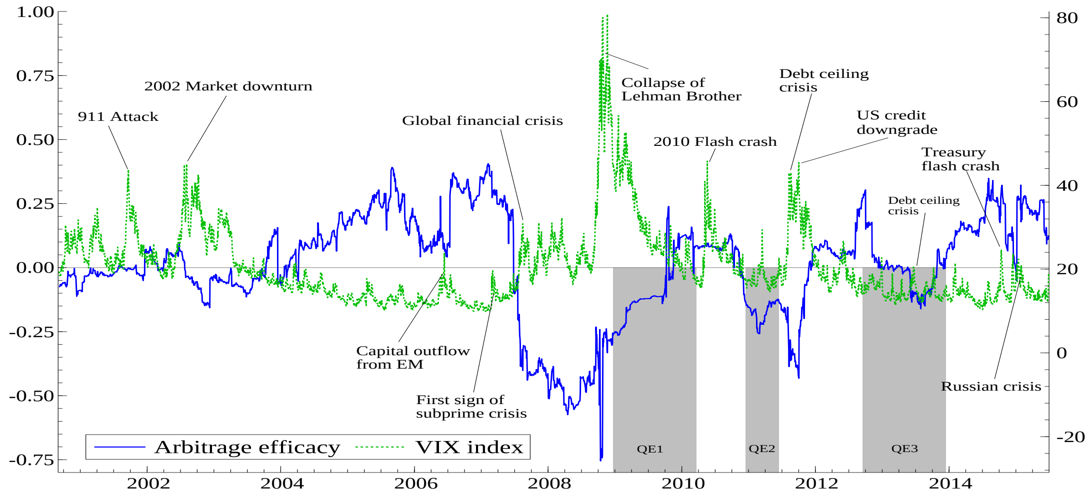

Figure 3.1: The plot of the implied arbitrage efficacy and VIX index, September 7, 2000 to June 30, 2015

The figure plots the time series of arbitrage efficacy (left axis) implied by the arbitrage relationship between S&P 500 index and E-mini future, and the VIX index (right axis) of implied volatility in S&P 500 index options. Major financial market events that causes large jump on VIX index are pointed out in the figure, while the shaded periods mark the three rounds of Quantitative Easing announced by Federal Reserve.

ing trend in *AE* due to the worries about budget limit.

To further verify the validity of *AE*, we document (vii) the significant link between the implied arbitrage efficacy *AE* and other broad measures of funding liquidity, e.g. the TED spread, the VIX index of implied volatility and the dividend yield of S&P 500 index. We also find that TED spread tend to be the driver of innovations in *AE* when arbitrage is inefficacious, i.e. funding constraint is binding; VIX index, on the other hand, becomes the dominating explanatory variable when *AE* is positive. This supports the nonlinear limits to arbitrage hypothesis in Chapter 2 that arbitrage risk (funding constraint) is the dominating arbitrage frictions during the period of the loose (binding) funding constraint.

We further illustrate the consequence of the implied arbitrage efficacy on asset pricing. (viii) On aggregate, we find evidence that changes in *AE* are significant predictors for daily innovations of the market volatility, measured by the VIX index of implied volatility in S&P 500 index option, and the volatility risk premia, measured by the difference between the implied and realized volatility of S&P 500. A drop in the lagged *AE* leads to increments in the market volatility and volatility risk premia.

More importantly, (ix) the predictive power is nonlinear conditional on the sign of AE. It is most prominent during the period of inefficacious arbitrage, i.e. binding leverage constraint, while negligible during the period of positive AE. This result verifies the amplification effect attributed to funding liquidity under binding leverage constraint, which is consistent with the slow-moving capital literature.

Our theoretical analysis contributes to the growing literature on funding liquidity as a limit to arbitrage. First, we extend the model of SV and Stein (2009) by allowing endogenous leverage setting, where financiers tend to use past information to determine the leverage constraint. SV and Stein study funding/leverage constraint as a limit to arbitrage and its impact on asset pricing and market efficiency. SV highlight that financiers tend to withdraw funds from the arbs with poor performance, which enhances the downward pressure on asset price, while Stein illustrates that leverage constraint may associated with fire-sale externalities, which also leads to price crashes. However, we suggest that less informed financiers may misinterpret past mispricing as future uncertainty, which renders arbs to deleverage and further intensifies selling pressure. Unlike Brunnermeier and Pedersen (2009), who also model the financiers' informational role, we set the informational level of financiers to be continuous, rather than binary.

Second, we introduce the theoretical definitions of funding liquidity and arbitrage efficacy, which are rather vague in current literature. Funding liquidity is often captured by the shadow cost of capital (Brunnermeier and Pedersen, 2009; Garleanu and Pedersen, 2011), and arbitrage efficacy by the size of persistent mispricings. For example, Vayanos and Weill (2006), Garleanu and Pedersen (2011) and Fontaine and Garcia (2011) link the size of arbitrage violation to arbitrage costs and funding constraints. Brunnermeier and Pedersen (2009) illustrate the interaction between an asset's market liquidity, measured by the price deviation from fundamentals, to the traders' funding liquidity, measured by the shadow cost of capital.[3] Rather, Duffie (2010) tend to focus on the association of asset price dynamics and the financial frictions faced by the arbs, such that sufficient arbitrage capital can speed up the mean reversion and price convergence. Similarly, we propose to define the funding liquidity via the arbs strategy: the marginal leverage with respect to the size of mispricing, which captures the arb's ability to raise leverage funding, and the arbitrage efficacy as the marginal error correction achieved by the arbs, which captures the arb's ability

[3]Several other papers, such as He and Krishnamurthy (2012, 2013), Acharya et al. (2009) and Gromb and Vayanos (2010), provide theoretical evidence of funding constraint as a limit to arbitrage.

to bear against mispricing.

Our paper also makes the following contributions to the empirical literature on funding liquidity. First, we introduce a distinct measurement for funding liquidity through the efficacy of arbitrage. Extant literature has divided into two types of measurements, where the first type is directly based on stated interest rates. The Treasury-Eurodollar (TED) spread is one of the most used, which captures the difference between the three-month LIBOR rate and the three-month T-Bill yield, a proxy for the overall funding costs faced by market intermediaries. Garleanu and Pedersen (2011) suggest another measure of a similar kind: the spread between LIBOR and the repo rate. Drehmann and Nikolaou (2013) measure the funding liquidity by the bank's aggressive bidding, which reflects how much a bank would pay to gain liquidity, at central bank auctions during 2005-2008. However, the state interest rate measures are often accused for underestimation (Aragon and Strahan, 2012) and subject to manipulation (Gandhi et al., 2015).

Our measure of the implied arbitrage efficacy is closer related to those indirect studies of funding liquidity as a friction of arbitrage, liquidity provision or market making. Comerton-Forde et al. (2008) and Adrian and Shin (2010) assess the intermediaries' funding liquidity by directly investigating the balance sheet data. Comerton-Forde et al. look into the position of NYSE specialist (the major liquidity provider in the market), while Adrian and Shin explore the repurchase agreement (repo) in the balance sheet of the financial intermediaries, which tend to be the primary tool of short-term borrowing and lending for investment banks and hedge funds. Akbas et al. (2015, 2016) proxy the available arbitrage capital by the capital flow to hedge funds who conduct arbitrage.[4]

While these papers tend to focus on the availability of arbitrage capital, others, like Garleanu and Pedersen (2011), Fontaine and Garcia (2011), Nagel (2012) and Frazzini and Pedersen (2014), pay attention to the size of arbitrage violation due to funding illiquidity. Garleanu and Pedersen argue that funding illiquidity give rise to the price deviation between securities with identical cash-flow but different margin requirements. Therefore, they show that the mispricings between credit default swap and the corresponding corporate bond are associated with funding illiquidity. Fontaine and Garcia measure the funding liquidity in U.S. treasury market through the price deviation between bonds with different ages but similar cash flows. Nagel

[4]Chordia et al. (2005) and Fleckensitein et al. (2014) use the flows into bond, equity funds and hedge funds as the measure for funding condition in financial intermediaries.

constructs a proxy for the return of liquidity provision in the equity market (NYSE, AMEX and Nasdaq stocks) by a reversal strategy that buy (sell) stocks with poor (good) past performance, and documents significant relationship with funding costs, funding supply and other broad measures.

Second, the empirical findings of this paper provide complementary evidence for the asset pricing consequence of funding liquidity. On the aggregate level, funding liquidity predicts the risk appetite, proxied by VIX index of financial intermediaries (Adrian and Shin, 2010); It is also predicted by the VIX index (Nagel, 2012), since higher market volatility enhances the scarcity of capital (Brunnermeier and Pedersen, 2009); Funding liquidity is associated with the return premia in fix-income securities, i.e. increase in funding liquidity results in lower excess bond returns (Fontaine and Garcia, 2011). We follow Adrian and Shin (2010) to forecast the VIX index and volatility risk premia with the implied arbitrage efficacy, but in a higher frequency (daily).

Third, it sheds lights on the nonlinear (amplification) effect attributed to the liquidity spirals, suggested by Brunnermeier and Pedersen (2009). Studies such as Comerton-Forde et al. (2008), Schuster and Uhrig-Homburg (2015)[5] and Drehmann and Nikolaou (2013) provide evidence for the nonlinear pricing impact during the period of the binding funding constraint. These works document that funding illiquidity leads to future market illiquidity, and the effect is most sensitive when funding constraint is binding. While we are able to ex ante identify the periods of binding funding constraint by the sign of implied arbitrage efficacy, and find statistically significant empirical evidence of the nonlinear consequence, Comerton-Forde et al. select the threshold for binding funding constraint at the 25th percentile of their measurement exogenously; Schuster and Uhrig-Homburg determine the nonlinear relationship endogenously within a regime-switching model; Drehmann and Nikolaou allocate the stress regime of binding funding constraint by the ex post events, such as the 2007-2008 financial crisis.

The paper proceeds as follows. In section 3.2, we extend the theoretical framework of Shleifer and Vishny (1997) and Stein (2009) and derive a number of testable predictions. In section 3.3, an empirical design is introduced to best capture the arbitrage efficacy from the theoretical work, as well as the application on S&P500 index

[5]Schuster and Uhrig-Homburg (2015), in their empirical analysis, use three measures for funding liquidity: market volatility, TED spread and dividend yield, which capture the frictions and scarcity of intermediaries' capital.

and E-mini future arbitrage. Section 3.4 summarizes the empirical results, including the time-varying implied arbitrage efficacy, the linkage with other measures of funding liquidity and the nonlinear asset pricing consequence. Finally section 3.5 concludes.

3.2 The Model

3.2.1 Market structure

We consider the market structure similar to Shleifer and Vishny (1997), where an asset with a fundamental value, V, trades for three periods, $t = 1, 2, 3$. At period 1, noise trader arrive with a pessimistic shock of size, s_1 pushes the asset price away from fundamental. Then the arbs attempt to enter the arbitrage trade to correct the mispricing error, and prevent the asset trading at distressed price. Denoting the arb's total (arbitrage) fund in period 1 as f_1, we derive the market clearing price by

$$P_1 = V - s_1 + f_1.$$

There exist two different market states at period 2. Under a bad state, noise intensifies such that $s_2^b > s_1$ with a probability, $q > 0$, in which case the price becomes:

$$P_2^b = V - s_2^b + f_2^b,$$

where f_2^b is the total funds available in period 2 under bad state. Under a good state with a probability, $1 - q$, noise disappears (i.e. $s_2^g = 0$), and the asset price thus converges towards fundamental:

$$P_2^g = V.$$

There is no investment required under good state. Finally, at period 3, price is assumed to recover:

$$P_3 = V$$

We maintain the assumption that the arbs are risk-neutral. Further, we follow Stein (2009) and allow the arbs to employ leverage to exploit the mispricing opportunity as most hedge funds seek arbitrage capital from outside financiers to support their operation in practice. Specifically, the arbs hold an equity, f_1^e and borrow funds through short-term debt,[6] f_1^d. Thus, the total arbitrage fund available at period 1 becomes the sum of equity and debt:

$$f_1 = f_1^e + f_1^d.$$

If the arbs can access to external capital without any friction, they are able to always eliminate any mispricing and guarantee the law of one price. In practice, however, they are faced with several funding constraints on equity and leverage debt. To accommodate the financial constraints observed in the real world,[7] we introduce the assumptions of the equity and leverage constraint borne by the arbs.

First, given that the (long-term) equity providers do not withdraw their funds early (period 2), we assume that the equity is constrained by

$$f_1^e < s_2^b, \tag{3.1}$$

such that equity supply cannot be guaranteed to cover the potentially deepening noise shock in the future.[8] For a sufficient equity supply, i.e. $f_1^e \geq s_2^b$, arbs are able to fully correct the mispricing without leverage and enforce the law of one price by investing, $f_1 = s_1$ at period 1, and $f_2 = s_2^b$ at period 2. Therefore, by imposing the equity constraint, we can focus on the situation where mispricing errors cannot be corrected with equity funding, and where leverage is adopted to conduct arbitrage.

Next, we allow the arbs to strategically determine the level of short term debt, f_1^d, and thus the level of leverage.[9] We assume that f_1^d must be repaid in full at period 2.

[6]Without loss of generality we assume zero interest rate

[7]Hedge funds' capital consists of equity capital, and leverage capital. Equity is the long-term capital supplied by the investors, who can withdraw their capital, so equity is not always locked into the firm indefinitely. Thus, in order to maintain and to protect funding, hedge funds impose initial lock-up periods and redemption periods prior to withdrawal. Other arrangements such as side pocket, gate limits and withdrawal suspensions are also employed. Hedge funds can also raise insecure leverage capital on liability side. The main source of leverage for hedge funds are (1) collateralized borrowing financed through repo market; (2) collateralized borrowing financed by the hedge fund's prime broker; (3) implicit leverage using derivatives, either exchange traded or over the counter. Leverage plays a central role in hedge fund management. Hedge funds use leverage to take advantage of mispricing opportunities by buying the underpriced and shorting the overpriced. Hedge funds also manipulate leverage to respond to changing investment opportunity set.

[8]Our setting of equity supply is different from that in Stein (2009). While the equity supply in Stein's model can be infinite but with a capital cost per dollar, our model impose a constraint on the size of equity, but equity is free to arbs. However it derives similar results, such that an equity constraint or a positive cost of capital will prevent arbs from fully mispricing correction. Under the equity constraint imposed in our model, arbs are induced to use leverage to exploit mispricings.

[9]Leverage is often defined as the ratio of total asset to equity:

This implies that the arbs can only invest their equity (i.e., $f_2 = f_2^e + f_2^d$ and $f_2^d = 0$) under bad state. We now introduce the upper and lower bounds of f_1^d, denoted D_U and D_L such that

$$D_L \leq f_1^d \leq D_U \tag{3.2}$$

The lower bound is given as $D_L = -f_1^d$, indicating that the arbs can lend their equity in full to other arbs. The upper bound D_U imposes the leverage constraint above which arbs are not able to raise the leverage fund. It is often set by the outside financiers who intend to control their value-at-risk.

3.2.2 Optimization problems

Hedge fund managers make optimal leverage decisions as a function of the investment strategies, the risk-return trade-offs and the cost of leverage, all subject to the leverage constraint imposed by external investors. Similarly, the arbs in our model ex ante manipulate their leverage in response to the risky arbitrage opportunity subject to the equity and leverage constraint. They face a simple trade-off: arbs are induced to raise as much short-term debt as they can to invest in period 1, and thus to exploit the positive return when price converge towards fundamental value. On the other hand, arbs may take a cautious leverage position in order to capitalize on a better opportunity in period 2 if bad state occurs. The arbs maximize their expected total wealth at the end of the period under perfect competition, which is given by

$$E\left(f_3^e\right) = (1-q) f_2^g + q \frac{V}{P_2^b} f_2^b. \tag{3.3}$$

where the equity available in period 2 under good and bad states, denoted f_2^g and f_2^b, can be expressed as

$$f_2^g = f_1^e + f_1 \left(\frac{V}{P_1} - 1\right) \text{ and } f_2^b = f_1^e + f_1 \left(\frac{P_2^b}{P_1} - 1\right). \tag{3.4}$$

$$L = \frac{Total Asset}{Equity} = \left\{ \begin{array}{cc} 1 + \frac{f_1^d}{f_1^e} & \text{for } f_1^d \geq 0 \\ 1 & \text{for } f_1^d < 0 \end{array} \right\},$$

which is determined by the size of the short-term debt, f_1^d.

Maximizing the expected total wealth in period 3 given by Eq.(3.3) subject to the leverage constraint, (3.2), we derive the optimal strategy of the short-term debt, f_1^d by the first order conditions:

$$\begin{Bmatrix} f_1^d = D_L & \text{for } R^1 < R^2 \\ D_L \leq f_1^d < D_U & \text{for } R^1 = R^2 \\ f_1^d = D_U & \text{for } R^1 > R^2 \end{Bmatrix} \tag{3.5}$$

where $R^1 = \frac{V}{P_1} - 1$ is the return of investing in period 1 and holding to price convergence, and $R^2 = q\left(\frac{V}{P_2^b} - 1\right)$ represents the expected return of investing in period 2. It is clear from Eq.(3.5) that the arbs can select one of the three equilibrium leverage strategies (see also Stein, 2009). For $R^1 < R^2$, the optimal decision is not to enter the market at period 1 (the waiting strategy with $f_1 = 0$), since waiting for the future opportunity provides a higher expected return. For $R^1 > R^2$, the arbs opt to borrow as much as they can to exploit the return of investing in period 1 (the max-leverage strategy with $f_1^d = D_U$), but subject to the binding funding constraint. Only when the two returns are indifferent, $R^1 = R^2$, the cautious investment strategy becomes optimal with $D_L \leq f_1^d < D_U$. The cautious strategy contains two sub-strategies: the dry powder strategy and the partial-leverage strategy. Arbs with the dry powder strategy are unlevered and invest only a part of their equity, i.e. $D_L \leq f_1^d < 0$, while arbs with the partial-leverage strategy are willing to be partially levered with $0 \leq f_1^d < D_U$.

As leverage is not actively used under the waiting and dry powder strategy, our further investigation, without loss of generality, will focus on the partial- and max-leverage strategy. While the optimal leverage decision for the partial-leverage strategy can be easily found under first order condition by $R^1 = R^2$, that for the max-leverage strategy is strongly determined by the upper leverage limit, D_U, as we will discuss next.

3.2.3 Leverage setting

It is important to understand how the financiers set the upper leverage limit, D_U, as arbs will raise the short-term leverage up to D_U to exploit the arbitrage opportunity under max-leverage strategy. In practice, such leverage setting depends critically on the financiers own predictions about future price movements. Brunnermeier and Pedersen (2009) assume that the estimated future price volatility consists of fundamental and liquidity volatility; Informed financiers are able to distinguish the two different types of volatility and obtain the correct estimation of the fundamental volatility, while uninformed financiers only acknowledge the total price volatility and thus exaggerate the future estimation of volatility.

In our model future price movement is associated with the adverse noise shock under bad state at period 2, s_2^b. We first assume that outside financiers acknowledge the arbitrage strategy at period 1, but may be less informed than the arbs about the future, such that they have to predict s_2^b conditional on the available information.[10] Denoting the financier's estimate as $\tilde{s}_2^b$, financiers are able to predict the future distressed price $\tilde{P}_2^b$ and the arbitrage equity $\tilde{f}_2^b$ in bad state, such that

$$\tilde{P}_2^b = V - \tilde{s}_2^b + \tilde{f}_2^b \tag{3.6}$$

$$\tilde{f}_2^b = f_1^e - \left(f_1^e + f_1^d\right)\left(\frac{\tilde{P}_2^b}{P_1} - 1\right) \tag{3.7}$$

Secondly, in order to determine the level of leverage constraint, we assume that competitive financiers set the rate of return as the riskless rate (zero in our model). In other words, financiers must ensure that the potential loss under a bad state must be covered by the arb's equity. This no-default condition at period 2 can be presented as: $\tilde{f}_2^b \geq 0$, such that financiers set limits on leverage debt D_U to protect themselves against the adverse future price movements.[11] Therefore the upper leverage limit D_U

[10]In Shleifer and Vishny (1997), outside investors are assumed to be blank about the arbitrage strategy due to the required specialized knowledge and the opacity within hedge funds. Thus investors update their belief from the arbs' past performance. In our model, however, financiers acknowledge the arbitrage strategy in period 1, i.e. long the under-priced asset. This is because short-term leverage debt are mostly financed through repo market, where arbitrageurs can be financiers, or the other way around. It is also captured in our model when arbitrageurs adopt the waiting or dry-powder strategy. Thus it is plausible that financiers have the information about the arbitrage strategies.

[11]The leverage setting in our model is similar to the margin setting in other theoretical paper, such as Brunnermeier and Pedersen (2009) and Gromb and Vayanos (2010). Margin or haircut is defined as the difference between the asset price and the collateral value, which must be financed through arbitrageur's own capital. Leverage, on the other hand, captures the ratio of the total asset to equity, that is, asset price over margin. Thus higher leverage limit implies lower margin requirement, and thus more capital offering to arbs.

Let's consider an example in the real world. To trade derivatives like futures and options, a hedge fund trades through a clearing broker. The exchange requires margins from the broker, and the margin is set to make the exchange almost immune to losses. Hence riskier contracts have larger margins. The broker, in turn, typically passes the margin requirement on to the hedge fund. The estimation about risk from the exchange or the broker will reflects on the margin requirement on the hedge fund.

is derived by the solution to

$$\tilde{f}_2^b = f_1^e - (f_1^e + D_U)\left(\frac{V - \tilde{s}_2^b}{P_1} - 1\right) = 0. \tag{3.8}$$

It implies that the upper leverage limit D_U is determined by financiers estimates about future shock, $\tilde{s}_2^b$, and the arbs equity capital, f_1^e.

The key question is how financiers gather their predictions about the future mispricings. Thus thirdly we represent the misperception as the difference between $\tilde{s}_2^b$ and s_2^b, and assume:

$$\tilde{s}_2^b - s_2^b = G(s_1) > 0. \tag{3.9}$$

Intuitively, financiers tend to exaggerate the future shock so as to protect their funding,[12] and estimate the future shock based on the past information, i.e. the initial mispricing error s_1.[13] On one hand, a higher initial error s_1 implies that a potentially higher arbitrage return will be realized when the price converges to fundamental. On the other hand, initial error may be misinterpreted as a higher uncertainty about the future, such that it exaggerates the estimation about future shock. The G function specifically captures the magnitude of the misinterpretation about s_1.

As the subsequent analysis does not depend on the convexity of the function G, we focus on a linear one:

$$G(s_1) = \tau s_1. \tag{3.10}$$

The parameter τ measures the level of informativeness among financiers, which is assumed to satisfy (see Appendix for proof):

$$0 \leq \tau \leq \frac{V - s_2^b}{s_1}. \tag{3.11}$$

In one extreme, $\tau = 0$ means that all financiers are fully informed such that they are free from any misinterpretation of s_1. As τ grows, financiers is described as less informed as they find it difficult to gather information about the market, which is likely to occur during the market turmoil. They tend to suffer from severe misinter-

[12] For those whose estimate is lower than s_2^b, they keep losing their short-term loans to insolvent arbs in the bad state.

[13] Misinterpretation among less informed financiers are common when adopting estimation from past data. See also Brunnermeier and Pedersen (2009) and Ang et al. (2011) for more detail about leverage setting and funding constraint in the real world.

pretation about s_1, which exaggerate their fear about the future uncertainty. In the extreme when all financiers become completely uninformed, $\tau = \frac{V - s_2^b}{s_1}$, they ban any short-term lending to arbs,[14] i.e. $D_U = 0$. While Brunnermeier and Pedersen (2009) consider only two types of financiers, informed and uninformed, to set their margin requirements, we allow the level of informativeness to be measured continuously by τ and derive more implications about the financiers informational role. We summarize the properties about the leverage constraint D_U as follows (see Appendix for proof).

Proposition 3. *Consider the model with market structure in Section 3.2.1 and the leverage setting in Section 3.2.3. The upper leverage constraint, D_U is positively related to the arbitrageurs' equity holding, f_1^e, and negatively related to the financiers' informational level, τ.*

The proposition implies that 1. arbs with higher equity holding are offered with a higher leverage constraint and thus able to borrow more short-term debts without reaching the upper bound; 2. more informed financiers with a lower misperception about future mispricing shock will provide a higher leverage constraint. Thus fully informed financiers with $\tau = 0$ will set the upper leverage limit, D_U as the highest.

3.2.4 Funding liquidity and arbitrage efficacy

3.2.4.1 Definitions

After setting up the model, we start to explore the arb's funding liquidity and the arbitrage efficacy, and the linkage between them. Unlike existing studies,[15] we allow the arbs to adjust their leverage position with respect to the prevailing circumstances, e.g. various size of mispricing caused by noise traders, and therefore define the funding liquidity as the arb's ability to raise leverage debt in order to bear against

[14]For any $\tau > \frac{V - s_2^b}{s_1}$, the leverage constraint D_U becomes negative, which attracts no arbitrageurs.

[15]The funding liquidity is loosely defined in the literature. Brunnermeier and Pedersen (2009) and Garleanu and Pedersen (2011) suggest to measure it by the marginal value of an extra dollar used, i.e. the shadow cost of capital. In Brunnermeier and Pedersen (2009), the shadow cost of capital is nil when leverage constraint is loose, since there is no arbitrage return as price recovers to fundamental. But, it becomes positive when the leverage constraint becomes binding. Chapter 2 define funding liquidity by the ratio of available funding to arbitrage return, which reflects the funding condition with respect to the size of potential investment opportunity. However, the capital of the intermediaries is exogenously given in these studies.

mispricing error:

$$\ell = \frac{\partial f_1^d}{\partial s_1} \leq 1. \tag{3.12}$$

Specifically, ℓ measures the marginal funds financed by the arbs to bear against one more unit of mispricing error. When ℓ is positive, arbs can lever up to exploit a larger mispricing and thus smooth the price fluctuation. We call this case "the stabilizing leverage". On the other hand, when arbs are faced with the binding leverage constraints, i.e. $f_1^d = D_U$, they might have to deleverage their position, leading to the situation where ℓ becomes even negative. We call this "the destabilizing leverage". The concept of (de)stabilizing leverage is consistent with the (de)stabilizing margin in Brunnermeier and Pedersen (2009).[16]

We also follow Cai et al. (2015) and define the initial mispricing correction as κ:

$$\kappa = \frac{f_1}{s_1} = \frac{f_1^e + f_1^d}{s_1}. \tag{3.13}$$

κ captures the percentage of mispricing correction achieved by arbs at period 1, which depends upon arbs' capital, f_1^e and f_1^d, and the size of mispricing, s_1. In one extreme, $\kappa = 0$, suggesting that arbs decide to select the waiting strategy, $f_1^d = -f_1^e$ and $f_1 = 0$. In another extreme $\kappa = 1$,[17] implying that arbs can raise sufficient funding to achieve full correction, such that $f_1^e + f_1^d = s_1$.

To this end, we define the arbitrage efficacy as the arb's ability to eliminate mispricing errors:

$$\alpha = \frac{\partial \kappa}{\partial s_1} \tag{3.14}$$

α measures the marginal mispricing correction by the arbs against one more unit of mispricing error. Its sign dictates whether arbitrage is effective or limited. For $\alpha > 0$, arbitrage is efficacious, implying that the arbs can achieve higher mispricing corrections with larger mispricing. By contrast arbitrage can be inefficacious, suggesting that larger mispricing error is followed rather by a lower mispricing correction.

Finally, to investigate the impact of funding liquidity on market liquidity and volatility, we define two other parameters: the market illiquidity (or mispricing errors that persists after arbitrage activity in period 1) by $\phi = E_1(P_3 - P_1) = s_1 -$

[16]Stabilising margins imply that reduced margins can be followed by the higher market illiquidity. In this case, arbs can raise more funds to exploit the potentially larger returns.

[17]In practice, mispricing correction can be captured by the Generalized Error Correction model, as in Cai et al. (2015). However, it is almost rare to observe the full error correction in empirical applications due to the existence of arbitrage frictions.

$f_1 > 0$, and the price volatility, following Hombert and Thesmar (2014), as $\sigma = E_1\left(\mid \frac{P_2-P_1}{P_1} \mid + \mid \frac{P_3-P_2}{P_2} \mid\right)$. We will investigate how sensitive are these two market-level parameters with respect to noise trader shock s_1, i.e. $\frac{\partial \phi}{\partial s_1}$ and $\frac{\partial \sigma}{\partial s_1}$, under different level of funding liquidity.

3.2.4.2 Linkages

Let ℓ^j, α^j, ϕ^j and σ^j be the funding liquidity, arbitrage efficacy, market illiquidity and price volatility associated with different leverage strategies, $j \in (p, m)$, where the superscripts p and m indicate the partial-leverage strategy and the max-leverage strategy, respectively. As the parameters do not have a simple expressions, we provide the link among the short-term leverage f_1^d, the funding liquidity ℓ^j and the arbitrage efficacy α^j as follows (see Appendix for proof and detailed derivations of f_1^d, ℓ^j and α^j):

Proposition 4. *Consider the model with market structure in Section 3.2.1 and the leverage setting in Section 3.2.3.*

*(i) (**Stabilizing leverage and efficacious arbitrage**) Under the partial-leverage strategy with $D_L \leq f_1^d < D_U$, we have: $0 < \ell^p < 1$ and $\alpha^p > 0$. As a result, higher mispricing slightly raises the market illiquidity $\frac{\partial \phi^p}{\partial s_1} > 0$ and price volatility $\frac{\partial \sigma^p}{\partial s_1} > 0$.*

*(ii) (**Stabilizing leverage but inefficacious arbitrage**) There exist a informational threshold $0 < \tau^* < 1$, such that under the max-leverage strategy with $f_1^d = D_U$ and $0 \leq \tau^- \leq \tau^*$, we have: $0 \leq \ell^{m,\tau^-} < \ell^p$, and $\alpha^{m,\tau^-} < 0$. Arbitrage fails, and larger mispricing leads to massive grow in the market illiquidity, $\frac{\partial \phi^{m,\tau^-}}{\partial s_1} > \frac{\partial \phi^p}{\partial s_1}$, and price volatility, $\frac{\partial \sigma^{m,\tau^-}}{\partial s_1} > \frac{\partial \sigma^p}{\partial s_1}$.*

*(iii) (**Destabilizing leverage and inefficacious arbitrage**) Under the max-leverage strategy with $f_1^d = D_U$ and $\tau^+ > \tau^*$, we have: $\ell^{m,\tau^+} < 0$ and $\alpha^{m,\tau^+} < \alpha^{m,\tau^-}$. The amplification is more intense, such that $\frac{\partial \phi^{m,\tau^+}}{\partial s_1} > \frac{\partial \phi^{m,\tau^-}}{\partial s_1}$, and $\frac{\partial \sigma^{m,\tau^+}}{\partial s_1} > \frac{\partial \sigma^{m,\tau^-}}{\partial s_1}$.*

The proposition is intuitive. Arbs with partial-leverage strategy will actively adjust their leverage in response to mispricings. As s_1 rises, the expected return of investing in period 1 is relatively higher, and arbs are able to borrow sufficiently more leverage capital from outside financiers and achieve higher mispricing corrections. We call this the stabilizing leverage and efficacious arbitrage. The increment in mispricing can be absorbed by the arbs with more leverage debt, which smooths the price fluctuation, thus the market illiquidity and price volatility is insensitive to changes in mispricing.

Proposition 4 further reveals the complex situations under the binding leverage constraint. When the max-leverage strategy is adopted, the arbs' funding liquidity is associated with the leverage constraint, which is crucially dependent on the informational level τ. The informational threshold[18] that differentiate stabilizing and destabilizing leverage is derived as (see Appendix for proof):

$$\tau^* = \frac{V - s_2^b + V - \sqrt{\left(s_2^b\right)^2 + 4\left(V - s_2^b\right) f_1^e}}{2\left(V - f_1^e\right)}. \tag{3.15}$$

In particular, informed financiers with $\tau \leq \tau^*$ are still willing to lend more capital to the arbs because they realize that expected arbitrage return is relatively higher with mispricings. Thus, leverage still rises with mispricing errors (stabilizing leverage). Even through leverage is stabilizing, the arb's funding liquidity is deteriorated, namely $\ell^{m,\tau^-} < \ell^p$, implying that they can raise less marginal funding against additional mispricing error. More importantly, arbitrage becomes inefficacious even with informed financiers, because the arbs fail to raise sufficient leverage fund and achieve higher mispricing correction, i.e. $\alpha^{m,\tau^-} < 0$. Due to the failure of arbitrage, additional mispricing can induce sharp rises in the market illiquidity and price volatility.

During the periods of market turmoil, financiers may find it more difficult to obtain information about the market, i.e. $\tau^+ > \tau^*$, and tend to heavily misinterpret the higher mispricing error as the higher volatility in the future. As a result, leverage shrinks even as mispricing rises, i.e. $\ell^{m,\tau^+} < 0$ (destabilizing leverage), which further dampen the efficacy of arbitrage, $\alpha^{m,\tau^+} < \alpha^{m,\tau^-}$. These results are broadly consistent with the theoretical prediction by Brunnermeier and Pedersen (2009) and the empirical evidence in favor of the pro-cyclical leverage documented in Adrian and Shin (2010) and Gorton and Metrick (2012). Moreover, as the arbs are forced to deleverage, which pushes price further away from the fundamental, they do not correct mispricing anomalies but instead enlarge them. It consists with literature on the limits of arbitrage that emphasizes the role of financial institutions and agency frictions on asset prices (Shleifer and Vishny, 1997; Gromb and Vayanos, 2010).

Proposition 4 also reveals some important implications from the model. First, the arbs' funding liquidity ℓ reflects the arbitrage frictions borne by them: the risk exposure and the leverage constraint. Notice that the arbs' funding liquidity is at its

[18] τ^* lies between 0 and 1, and τ^* increase in f_1^e and decrease in s_2^b. Intuitively, the more equity capital in the arbs or the less noise shock in the future, the more likely financiers are able to impose stabilizing leverage.

highest, $\ell = 1$, under the special case where arbitrage is riskless even in a short run ($q = 0$, 1) and leverage is not constrained ($D_U \to \infty$).[19] However, under the presence of risky arbitrage, e.g. $0 < q < 1$, funding liquidity is deteriorated, $\ell^p < 1$ ((i) in Proposition 4), as the arbs become more cautious in the use of leverage. Furthermore, when leverage constraint becomes binding, $f_1^d = D_U$, funding liquidity is determined by the leverage constraint set by financiers, which is further tightened, $\ell^m < \ell^p$ ((ii) and (iii) in Proposition 4). Therefore, the model posits that the risk exposure of the underlying arbitrage strategy (leverage constraint) tend to be the driving force of the arb's funding liquidity ℓ, under partial-leverage (max-leverage) strategy.

Second, it highlights the importance of leverage constraint on the efficacy of arbitrage. Without the restriction on leverage, we always have stabilizing leverage and efficacious arbitrage as shown in (i) of Proposition 4. This implication has intensively studied by the extent literature, which tend to focus on various arbitrage cost that deter the arbitrage activity. In particular, it is empirically captured by the threshold effect within the error correction model, such that arbitrage force tends to strengthen as the mispricing error exceeds a certain threshold, which reflects the costs borne by the arbs (Tse, 2001; Gyntelberg et al., 2016). However, the model suggests that the positive relation between arbitrage activity and the size of error can reverse when the leverage constraint binds. It provides further support to the work in Chapter 2, which point out the possible negative relation under the binding funding constraint.

Third, it is a nontrivial and challenging issue to identify when the leverage constraint binds. Adrian and Shin (2010) and Ang et al. (2011) propose to directly gather the information about changes in leverage position out of an analysis of balance sheets of financial intermediaries. However, this is based on the quarterly (or even longer) reports of the major hedge funds and investment banks, in which data is difficult to collect and poor in quality due to the opaqueness of the industry. The return-based measures, like Fontaine and Garcia (2011) and others, successively capture the innovation in funding liquidity, but cannot identify when the leverage constraint becomes binding. By introducing the analysis of arbitrage efficacy, we find that the arbitrage efficacy not only links to the funding liquidity, inefficacious arbitrage also identifies the binding leverage constraint, even when financiers are in-

[19] In the special case of riskless arbitrage ($q = 1$) and unlimited leverage ($D_U \to \infty$), the bad state must occur in period 2. Then we must have $P_1 = P_2^b$ from the first order condition (Eq. (3.5)), and thus the optimal leverage is $f_1^d = s_1 - s_2^b$. For another case of riskless arbitrage ($q = 0$), where only good state occurs in period 2, we have $P_1 = V$ from the first order condition, and thus the optimal funding is $f_1^d = s_1 - f_1^e$. Under these two special cases, it is easily seen that arbs achieve the highest funding liquidity, $\ell = 1$. See Appendix for proof.

formed. Proposition 4 thus illustrates the most crucial feature the arbitrage efficacy can offer, while existing measurements such as the size of arbitrage violation and the state interest rate fail to draw.

3.2.5 A numerical example

We use a numerical example to illustrate the arb's funding liquidity, arbitrage efficacy and their link. Let the fundamental value of the asset be, $V = 1$, probability of bad state to occur be, $q = 0.1$, the deepening noise shock in bad state be, $s_2^b = 0.4$. Arbs have a size of equity holding, $f_1^e = 0.1$, and we allow the initial noise shock s_1 varying from 0 to0.35. We consider two cases with different level of informativeness. The benchmark case is plotted on the top panel of Figure 3.2.5, with the level of informativeness, $\tau = 0$. It illustrates the arbs' choice of short-term leverage debt f_1^d and mispricing correction κ_1 achieved after period 1 with different size of initial shocks, s_1, the slopes of which reflects their funding liquidity and arbitrage efficacy. The waiting strategy is shown in the first block on the left, where arbs make a full short-term loan. Arbs enter the market from the dry powder strategy, where they are unlevered and safe some equity for future investment. Leverage is used by arbs in both partial- and max-leverage strategy. The arbs' funding liquidity and arbitrage efficacy (the slopes of the leverage debt and mispricing correction) is clearly positive and higher under the dry powder and partial-leverage strategy, as suggested in (i) of Proposition 4. Notice that without the assumption on the upper leverage constraint set by financiers, we will always be in the dry-powder or partial-leverage strategy, where leverage is stabilizing and arbitrage is effective. However, as shown in the top plot where arbs adopt the max-leverage strategy, they are forced to follow the upper leverage constraint (dash line). As a result, funding liquidity is dampened but remain positive (stabilizing leverage) while arbitrage efficacy becomes negative (inefficacious arbitrage). This is captured in (ii) of Proposition 4.

The middle panel on Figure 3.2.5 shows the arbs' leverage debt and mispricing correction with a informational level, $\tau = 1$. Arbs are under a similar situation as the former case when adopting partial-leverage strategy. As they enter max-leverage strategy, leverage becomes destabilizing such that arbs have to deleverage with mispricings. Comparing with top panel, arbitrage efficacy also exhibits a sharper decline. The results reflect on (iii) of Proposition 4. The bottom plot in Figure 3.2.5 illustrates market illiquidity and price volatility when the leverage constraint binds. Under the partial-leverage strategy, arbitrage is effective and able to smooth price fluctuation,

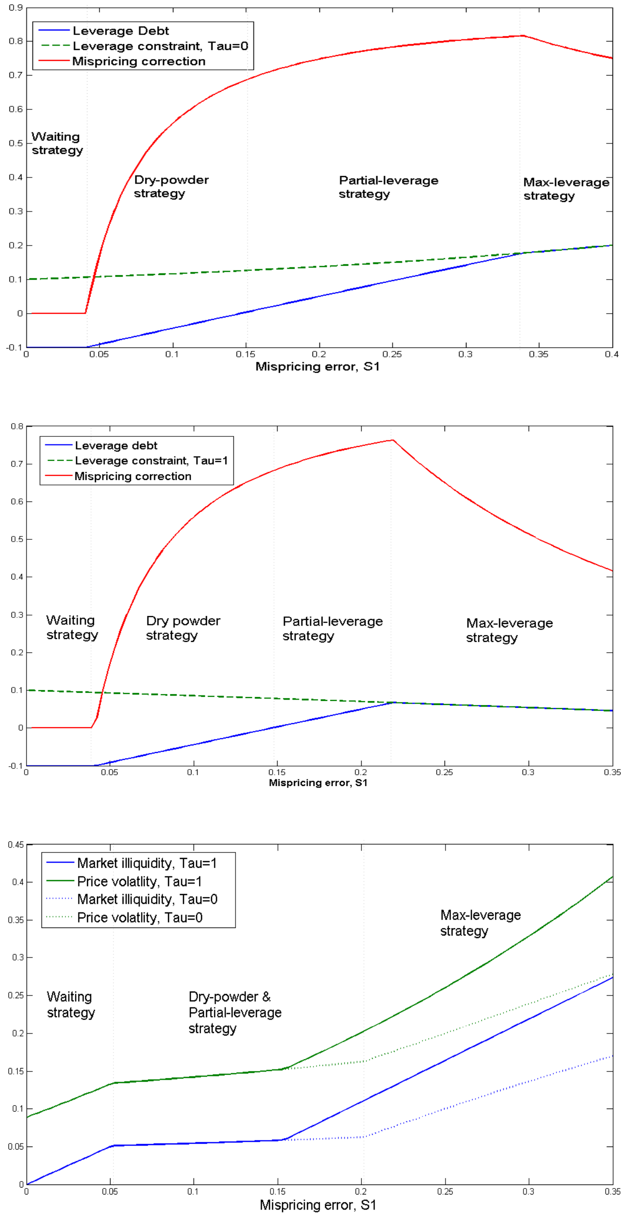

Figure 3.2: Leverage debt, leverage constraint and mispricing correction

The top and middle figures show the leverage funding financed (bottom solid line), mispricing correction conducted by the arbs (top solid line) and the leverage setting implemented by financiers (dashed line) against varying shocks, while the bottom figure plots the expected market illiquidity (bottom solid/dashed line) and price volatility (top solid/dashed line) against increasing mispricing error, s_1, under different leverage strategy. The asset has a fundamental value of $V = 1$, probability of bad state be, $q = 0.1$, the size of equity be, $f_1^e = 0.05$, deepening liquidity shock in bad state be, $s_2^b = 0.4$, and the initial liquidity shock s_1 varying from 0 to 0.35. In the top left plot, financiers are fully informed, such that the level of informativeness, $\tau = 0$, while the top right plot has a positive level of informativeness, $\tau = 1$. In all plots, the arb's leverage strategy are classified as waiting, dry powder, partial-leverage and max-leverage strategy.

therefore market illiquidity and price volatility are insensitive to mispricing shock. However, as arbitrage becomes inefficacious, both parameters rise sharply with mispricing, and the effect is most prominent with less informed financiers (solid line with higher τ).

3.3 The Empirical Design

It would be ideal to empirically capture the arbs' funding liquidity defined in our model. However to identify the innovation in leverage position with respect to the arbitrage opportunity is always challenging. After establishing the link between the arbs' funding liquidity and the arbitrage efficacy, we propose an empirical design to capture arbitrage efficacy in practice. In this section, we first introduce the strategy to capture an asset's mispricing error and the arbs' mispricing correction. Second, we illustrate the design to capture the arbitrage efficacy as a measure of funding liquidity. Finally, we describe the underlying data that are applied to the strategy for our empirical tests.

3.3.1 Measuring errors and corrections

In order to capture the arbitrage efficacy defined in our theoretical framework, we first introduce the empirical strategy to capture the mispricing errors and the arbs' mispricing correction. Estimating the mispricing error of an asset requires the value of both its spot and fundamental value. Suppose p_t and p_t^* is the natural log of the asset's spot and fundamental price at time t, then z_t, the mispricing error, can be estimated from the long-run equation:

$$p_t = \mu + \theta p_t^* + z_t. \tag{3.16}$$

Under the assumption of frictionless arbitrage, there is no price deviation from the no-arbitrage relation. Even when asset price is drifted away from the fundamental value, the arbs will take place and correct the mispricing error in no time. In practice, however, one tends to observe persistent price deviation from the fundamental value, which indicates that the market is far from frictionless and arbitrage is limited.

In order to capture the arbitrage activity, Cai et al. (2015) introduce a two-period

GECM to capture the mispricing correction achieved by arbs, such that:

$$\Delta p_t = \kappa z_{t-1} + \lambda^* z_{t-2} + \delta \Delta p_t^* + \gamma \Delta p_{t-1} + \mu_t, \ \mu_t \sim iid(0, \sigma_\mu^2) \tag{3.17}$$

where p_t and p_t^* is the natural log of the asset spot and fundamental price at time t. The lagged one error correction term, κ then captures the initial mispricing correction achieved by arbs during the sample period, while $\lambda^* = \lambda\,(1+\kappa)$ represents the percentage of unarbitraged error persisting to the next period. The interpretation of the error correction term as the force of arbitrage has been widely documented in the literature (Dwyer et al., 1996; Martens, Kofman and Vorst, 1998; Tse, 2001). Chapter 2 further consider the combined effect of arbitrage cost and funding constraint as the limits to arbitrage, and suggest that the mean reversion process can be nonlinear. Therefore the arbs' mispricing correction is time variant, such that it changes accordingly with the market circumstances and mispricing errors.

3.3.2 Measuring the arbitrage efficacy

Notice that the mispricing correction κ estimated in Eq. (3.17) is a static measure, which only looks at the average correction within a period of time, while the dynamic mispricing error z_t changes in a daily basis. To obtain the dynamic measure of mispricing correction κ_t, we apply a rolling-window time-series regression of Eq. (3.17) and assign the estimated κ_t to the ending date of the window. Specifically, we apply for a fixed window of 500 days. Therefore each dynamic κ_t for date t is estimated over the past 500 day window, correcting a series of mispricing errors z_i, $i = t-500, \ldots t-1$. Before we continue to estimate the arbitrage efficacy, the empirical data requires some adjustments as outilers are eliminated, i.e. when the estimated $\Delta\kappa_t$ or Δz_{t-1} is at the highest or lowest 1% tails of the distribution.

We aim to capture the arbitrage efficacy as the marginal correction: the ratio of changes in daily mispricing correction, $\Delta\kappa_t = \kappa_t - \kappa_{t-1}$ to the changes in mispricing errors Δz_t. By definition, the daily difference $\Delta\kappa_t$ is affected by a series of daily difference of mispricing errors, $\Delta z_i = z_i - z_{i-1}$ for $i = t-500, \ldots t-1$. Instead of taking the whole series of daily difference of mispricing error into account, we focus on the impact of daily changes of error at the nearest time, $\Delta z_{t-1} = z_{t-1} - z_{t-2}$.

We estimate the arbitrage efficacy, denoted as *AE* by the first-difference estimator: regressing the daily difference of dynamic mispricing correction on the daily variations of mispricing error they observed at the nearest time,[20] such that

$$\kappa_t - \kappa_{t-1} = AE_{t-n,t}\,(z_{t-1} - z_{t-2}) + \varepsilon_t,\ \ \varepsilon_t \sim iid(0, \sigma_\varepsilon^2) \tag{3.18}$$

and

$$\begin{aligned} AE_{t-n,t} &= \sum_{t-n}^{t} (\kappa_t - \kappa_{t-1})\,(z_{t-1} - z_{t-2}) \,/\, \sum_{t-n}^{t} (z_{t-1} - z_{t-2})^2 \\ &= \sum_{t-n}^{t} \Delta\kappa_t \Delta z_{t-1} / \sum_{t-n}^{t} \Delta z_{t-1}^2 \end{aligned} \tag{3.19}$$

where n represents the size of sample period, $\Delta\kappa_t$ and Δz_t represents the daily changes in mispricing correction and mispricing error, respectively. The slope coefficient $AE_{t-n,t}$, captures the average arbitrage efficacy over the period from $t-n$ to t. It is a symmetric measure, such that it can be negative if $\Delta\kappa_t$ and Δz_{t-1} has different sign. It indicates that funding liquidity is tight and capital constraint is binding during the sample period. The methodology has three merits. First, arbitrage efficacy may also depend on other arbitrage costs, such as transaction costs, which tend to be less time-variate. The first-difference estimator mitigates the noise of these omitted time invariant variables; Second, by taking the first-difference estimator with a large n, the estimator is more likely to wipes out the temporary variation and reflect the fundamental of the funding liquidity with more accuracy. Third, the methodology, with a recognizable arbitrage relationship, can be applied to various assets, sectors, and countries, to evaluate funding liquidity in broader scope.

For the purpose of tracking the financial market in response to the funding liquidity, data with a higher frequency is more in favor. In contrast to the static approach to estimate $AE_{t-n,t}$, the dynamic version can be obtained on a rolling window basis. In particular, we take a large window $n = 250$,[21] and the dynamic arbitrage efficacy is assigned to the ending date of the window as AE_t, representing the implied arbitrage efficacy during the past 250 days.[22]

[20] We get similar results if we include a constant term.

[21] The choice of window is considered to be large enough to reduce the noise of daily fluctuation, but also small enough to retain the important information about the fundamental in daily innovation. While the size of the window can affect the estimation results, robustness check has been done using a window of 120 and 500 days to estimate both the daily series of κ_t and AE_t, where the results are mostly similar and significant. Due to the scale of this paper, the results available upon request from the authors.

[22] We summarize the process of the empirical design to capture the arbitrage efficacy in the Ap-

	Mean	Median	Minimum	Maximum	Std Dev
Δs	0.016	0.056	-9.469	10.957	1.263
Δp	0.017	0.065	-10.399	13.197	1.288
$p-s$	0.115	0.001	-1.689	2.030	0.516
$p-p^*$	0.074	0.058	-2.201	1.792	0.213
r	2.01	1.42	-0.02	6.24	2.077
q	1.83	1.86	1.07	3.37	0.387

Table 3.1: Basic descriptive statistic

The table reports the the descriptive statistics for all variables. The sample used is the daily series of the S&P 500 index and its E-mini futures contract covering the period September 4, 1997 to June 30, 2015. Δs (Δp) is the first difference of log spot (futures) price. The log fundamental value is computed as $p^*_{t,T} = s_t + (r_t - q_t)\,\tau_t$ where r_t is the annualized risk-free (3 month T-bill) interest rate on an investment for the period , and q_t is the annualized dividend yield on the index. All numbers are recorded in percentage point terms.

3.3.3 The data

Our empirical study applies to the arbitrage relation between S&P 500 index and E-mini S&P 500 future. It provides several advantages. First, the index-future relationship provides a way to estimate the fundamental value of the future contract, and thus the mispricing errors observed by the arbs. Specifically, we infer the fundamental value of the E-mini contract from the cost of carry model, which is indicated in the following relationship to hold in equilibrium:

$$p^*_{t,T} = s_t + (r_t - q_t)\,\tau_t, \tag{3.20}$$

where $p^*_{t,T}$ is the natural log of the fundamental price of E-mini future contract with a maturity date T implied from cost of carry model; s_t is the log spot price of the S&P 500 index; r_t and q_t is the risk-free interest rate and dividend yield of the asset, respectively; $\tau_t = T - t$ is the time to maturity. Then, the mispricing error and mispricing correction in Eq. (3.16) and (3.17) can be estimated with the log spot price $p_{t,T}$ and fundamental price $p^*_{t,T}$ of the E-mini future contract. For completeness, we collect our proxies for risk-free interest rate: the US three-month T-bill rate, and dividend yields on the S&P 500 index. All data are sourced from DataStream. We provide the summary statistics of the daily S&P 500 index and E-mini S&P 500 future data in Table 3.1.

Second, the E-mini future is one of the most traded future contracts, which contains large numbers of financial intermediaries, such as hedge fund and investment

pendix.

banks, that will exploit any arbitrage opportunity in the market. In particular, the E-mini S&P 500 future is traded on Chicago Mercantile Exchange (CME) electronic platform, which is accessible to off-floor traders. Domowitz and Steil (1999) argue that electronic markets tend to offer liquidity at lower cost than the floor-traded standard one. E-mini future also offers a smaller size of contract and a longer trading hours, which attracts traders with modest capital, such as high frequency traders and market makers. Hasbrouck (2003) and Kurov and Lasser (2004) suggest that E-mini contracts with small denomination have a higher efficacy in the price discovery process for the S&P 500 index, than the standard contracts. Thus the arbitrage relationship between S&P 500 index and E-mini future is more likely to reflect the broad arbitrage efficacy in the US stock market.

Third, the E-mini future has been trading since September 16, 1997, which include the periods of rapid growth in hedge fund industry[23] and some noticeable market events, like the burst of dot-com bubble, the recent financial crisis 2007-2008, the European sovereign debt crisis and the Flash Crash in 2010. It helps to verify the validity of the funding liquidity measure in different market circumstances, especially the extreme ones.

Overall, we end up with 3654 observations of AE_t, from September 7, 2000 to June 30, 2015, and the main results and the tests of theoretical hypotheses are presented in the next section in detail.

3.4 Main Results

3.4.1 The implied arbitrage efficacy

Figure 3.3 shows the arbitrage efficacy (solid line) implied by the arbitrage relationship between S&P 500 index and E-mini future in a 250-day window. At first glance, we see that the implied arbitrage efficacy AE varies through time; the daily series of AE fluctuates gently in the early stage of the sample period from 2000 to 2003, varying between ± 0.1; It then displays a noticeably constant increasing trend until reaching its all time peak at early 2007. It is after early 2007 that the variation in AE began to be remarkably volatile, and even drop to its bottom, at around -0.75, during

[23]According to the data of Fung and Hsieh (2013) gathering from BarclayHedge, HFR, Lipper-Tass and Hedgefund.net, the hedge fund industry starts to grow dramatically after 2000. The number of funds and the asset under management (AUM) are more than five times larger in 2010 than that in 2000.

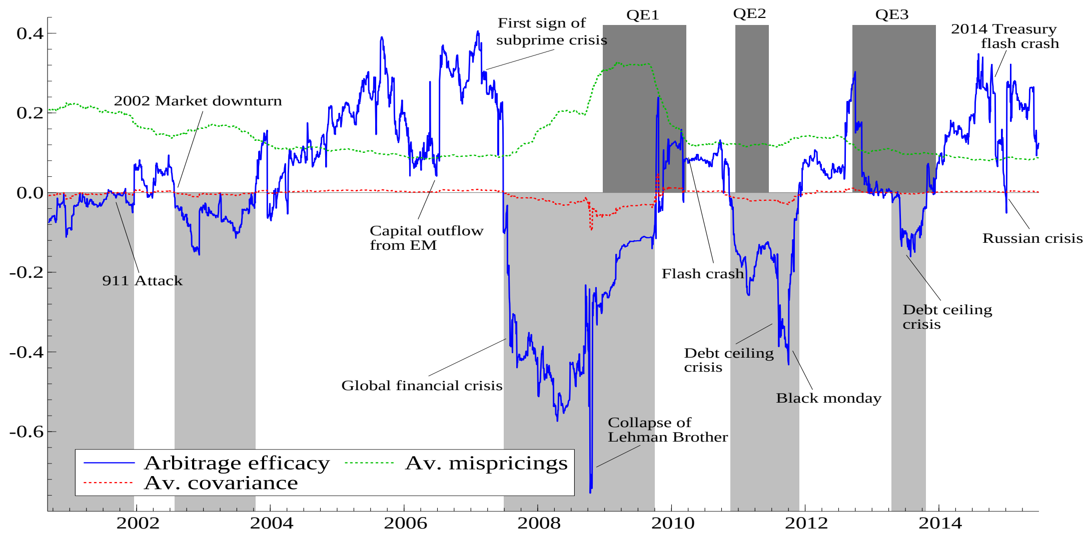

Figure 3.3: The plot of the implied arbitrage efficacy, the average mispricing error and the average covariance, September 7, 2000 to June 30, 2015

The figure shows the implied arbitrage efficacy (solid line), the average mispricing errors in absolute value (top dashed line) and the average covariance between $\Delta\kappa_t$ and Δz_{t-1} (bottom dashed line), implied by the arbitrage relationship between S&P 500 index and E-mini future, and computed through a 250-days rolling window. The implied arbitrage efficacy is calculated by regressing the difference in mispricing corrections on the variation in mispricings; Mispricing error is calculated as the absolute value of the error term in the long run relationship in Eq. (3.16); The average mispricing error in absolute value reflects the denominator of AE, while the average covariance represents the numerator of AE. The periods of the three round of QE are shaded on the top, while the periods of inefficacious arbitrage are shaded in the bottom. Some major financial market events are pointed out in the figure. The data sample is from September 7, 2000 to June 30, 2015 at a daily frequency.

the collapse of Lehman Brother. After the Global financial crisis in 2007-2008, AE again shows a constant increasing trend till the end of our sample period, despite two noticeable declines in 2011 and 2013 due to the debt ceiling crisis. Moreover, there are four significant periods of time where the sign of AE is negative, i.e. arbitrage is inefficacious, which are coincide with most of the market crashes and turmoils during our sample periods. The major events include: the market downturn in 2000-2003, the Global Financial Crisis in 2007-2008, the Debt Ceiling crisis in 2011 and the Debt Ceiling crisis in 2013.

The trajectory of AE also reflects the effectiveness of the monetary policy announced by Federal Reserve for liquidity expansion, which may help improve funding liquidity condition and the efficacy of arbitrage. In particular, it is after November 2008 that AE witnessed a constantly rapid increment due to the introduction of the first round of quantitative easing (QE). However, as Fed continued to implement the second of QE (QE2) in November 2010 and the third round of QE (QE3) in September 2012, the impact and effectiveness of QE2 and QE3 seem to diverge. While the implied arbitrage efficacy continues to drop after the announcement of QE2, it witnesses a sharp increment and stay positive after QE3.

The average mispricing errors (top dashed line) and the average covariances between correction and error (bottom dashed line) observed in a 250-day window are plotted along side with arbitrage efficacy in Figure 3.3. We notice that the sharp decline in arbitrage efficacy, such as Global Financial Crisis in 2007-2008, is due to both a rapid growth in the size of mispricings and fall in the covariance between the mispricing correction and the mispricing error. It also shows that innovation in the size of mispricing errors are negatively associated with the arbitrage efficacy; In particular, mispricing error continues to decline from 2000 to early 2007, whilst the arbitrage efficacy are increasing and the efficacious arbitrage are dominating. However, during the 2007-08 financial crisis, the size of mispricing error displays an apparent grow, where the arbitrage efficacy witnesses a sharp drop and becomes inefficacious.

3.4.2 Linkage to other funding liquidity measures

To further verify the link between the funding liquidity and the implied arbitrage efficacy AE , we test the linkage of AE with three other variables that have been applied in the literature to measure the funding liquidity. One typical measure is the TED spread, e.g. the spread between the three-month risky LIBOR rate and the three-month risk free T-Bill yield. It measures the cost of funding among the financial intermediaries. The VIX index of implied volatility in S&P 500 index options is also applied as a measure of funding liquidity in Ang et al. (2011) and Schuster and Uhrig-Homburg (2015). The VIX index is calculated from the S&P 500 index options, as one of the factor to determine the option price.[24] It reflects the market forecast of the aggregate financial market volatility, i.e. higher VIX means traders are expecting that the market is more likely to fluctuate sharply in the near future. Hence the VIX index are often used in forecasting the market volatility, and the impact of major financial market events reflects on the spikes of VIX index. The dividend yield of S&P 500 index, regarded as a measure of required returns, is also proxying for funding illiquidity (Garleanu and Pedersen, 2011), such that a high dividend yield relates to a poor state of economy where the funding condition is worse. In figure 3.4, we plot the time series of the arbitrage efficacy, TED spread, VIX index and dividend yield of S&P 500 index, respectively. At first glance, the three broad variables of funding illiquidity are negatively associated with the implied arbitrage efficacy, such that lower arbitrage efficacy, especially during the period of inefficacious arbitrage, coincides with higher TED spread, VIX index and dividend yield.

We test the correlation more formally in Panel A of Table 3.2, where we report the time-series regressions of the implied arbitrage efficacy (AE) on the TED spread, the VIX index and the dividend yield of S&P 500 index, both individually and jointly. For the full sample period from 2000 to 2015, the implied arbitrage efficacy loads significantly on all three measures of funding illiquidity. The slope coefficients are significantly negative, which supports our model predictions that arbitrage efficacy captures the information of funding liquidity. Both the TED spread and the VIX index have the highest R^2, just over 30%, while that for the dividend yield of S&P 500 index is only 6%. The TED spread jointly with the VIX index explains over 40% of the variation in AE with significant negative sign while the dividend yield as another explanatory variable does not improve the explanatory power. Therefore, on aggregate, the arbitrage efficacy is closely linked to the cost of funding and the stock market risk.

We also run the regression conditional on (in)efficacious arbitrage (the sign of the implied arbitrage efficacy). Column 1 to 4 in Panel B of Table 3.2 report the conditional results on the period of efficacious arbitrage. We notice that only the coefficient of the VIX index remains significantly negative, while that of the TED

[24] See CBOE for more detail of construction of the index.

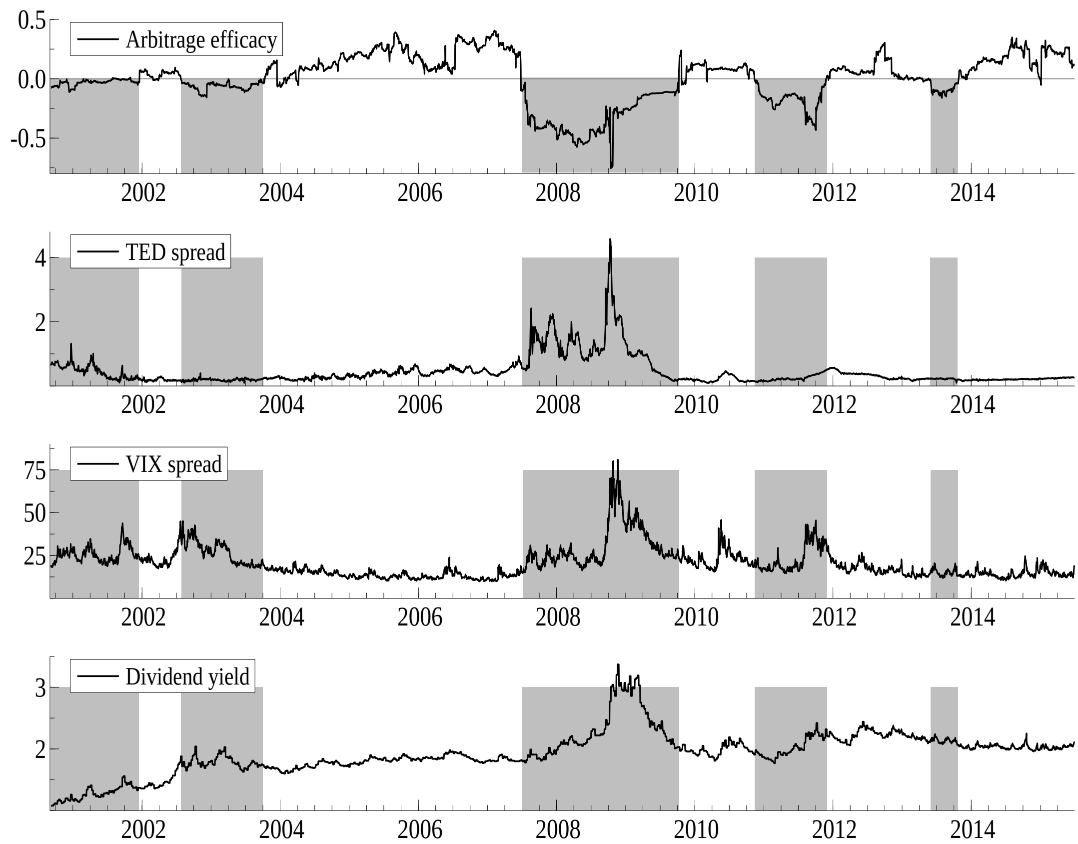

Figure 3.4: The plot of the implied arbitrage efficacy, VIX index, TED spread and dividend yield, September 7, 2000 to June 30, 2015

The figure plots the time series of arbitrage efficacy, VIX index (proxying for market risk), TED spread (proxying for funding cost) and Dividend yield (proxying for risk premia) of of S&P 500. The periods with inefficacious arbitrage are shaded. Daily series of VIX index, TED spread and Dividend yield of S&P 500 index are collected from Datastream.

spread becomes positive and that of the dividend yield is no longer significant. The VIX index tend to be the dominating determinants among the three variables, as it has the highest R^2 of 23%, while that of the TED spread drops to only 13%. On the other hand, the conditional results on the periods of inefficacious arbitrage are shown in the last four column of Panel B, which changes completely. We see that coefficients of all three variables are significantly negative. The R^2 for the TED spread becomes the highest, around 47%, which tend to be the dominating explanatory variable for AE. But the R^2 for the VIX index declines to only 4%. Notice that conditional on inefficacious arbitrage, the dividend yield of S&P 500 index gathers a high degree of explanatory power, comparing to the results from the full sample and the sub-period of efficacious arbitrage.

The results show that the implied arbitrage efficacy are mainly explained by the VIX index during the period of efficacious arbitrage, but by the TED spread during the period of inefficacious arbitrage, which is consistent with the nonlinear limits to arbitrage predictions in Chapter 2 and our modified model. Therefore, the overall market risk tend to be the first concern for the arbs to obtain funding during the good times, while the extremely high funding cost becomes the dominating factor during the bad time.[25]

3.4.3 Forecasting the market risk

Our model predicts that the arbitrage efficacy forecasts the future market volatility, such that lack of funding liquidity reduces the arbs' ability to bear against larger exogenous shocks, which induces higher price volatility. The relation tend to be more prominent during the periods where the leverage constraint is binding (inefficacious arbitrage) due to the amplification effect. To examine these predictions, we test whether the past arbitrage efficacy is able to predict the future market volatility. We have the VIX index as the ex ante risk-neutral expectation of the future market volatility, which is forward-looking over the next 30 days. In order to further

[25] Also notice that the TED spread becomes positively associated with arbitrage efficacy when arbitrage is efficacious, which seems to be counter-intuitive. For instance, from 2004 to early 2007 where arbitrage is efficacious and market liquidity tend to be ample, the TED spread, although at its historically low level, displays a slightly increasing trend. It also appears on the periods of 2012-2015, where the market seems to be liquid. Therefore, simply focusing on the trend of the TED spread might be misleading. Although a sufficiently high magnitude of increment in the size and volatility of the TED spread could indicate the funding problem, such as the spikes during the financial crisis, how to ex ante distinguish those from the normal time innovations of the TED spread or other return-based measures is proved to be challenging.

understanding the forecasting mechanism, we follow Bollerslev, Tauchen and Zhou (2009), and decompose the VIX index of implied volatility into two components: the realized volatility of S&P 500 index and the price of risk of market volatility (the volatility risk premia). We first compute the ex post realized volatility for S&P 500 index over the past one month,[26] assuming a zero mean, such that

$$RV_t = \sqrt{\frac{252 * \sum_{i=0}^{N-1} \left(\frac{P_{t-i}}{P_{t-i-1}} - 1\right)^2}{N}} * 100 \tag{3.21}$$

where RV_t is the realized volatility level at time t, P_t is the price return index level of S&P 500 index on day t, and N is the look back period, which is 21 days for our case. Then the volatility risk premia is computed as the difference between the ex ante VIX index of implied volatility over the next one month and the ex post realized volatility of S&P 500 index over the past one month, i.e.

$$VRP_t = VIX_t - RV_t. \tag{3.22}$$

Bollerslev, Tauchen and Zhou (2009) document that volatility risk premia are the dominating predictor of the excess return of S&P 500, among other predictors, such as P/E ratio, default spread and consumption-wealth ratio.

Figure 3.5 plots the daily time series of the VIX index of implied volatility, the realized volatility of S&P 500 index and their difference as the volatility risk premia. We see that the VIX index and the realized volatility of S&P 500 co-move, and the volatility risk premia is mostly positive through the sample period.

[26] In Bolloerslev, Tauchen and Zhou (2009), the realized volatility is constructed from high-frequency intraday data, while in our paper and the work of Adrian and Shin (2010), it is computed on a daily basis.

Panel A		AggregateAE			
		(1)	(2)	(3)	(4)
TED	Coef	-0.243***			-0.166***
	T-stat	-27.10			-15.54
VIX	Coef		-0.012***		-0.008***
	T-stat		-37.44		-21.53
DY	Coef			-0.147***	0.012**
	T-stat			-18.55	2.26
Cst	Coef	0.114***	0.257***	0.291***	0.224***
	T-stat	26.81	36.90	19.46	17.60
	R^2	0.3089	0.3019	0.064	0.4071

Panel B		Conditional on $AE > 0$				Conditional on $AE < 0$			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TED	Coef	0.260***			0.189***	-0.175***			-0.152***
	T-stat	17.36			13.90	-22.64			-19.40
VIX	Coef		-0.009***		-0.008***		-0.003***		
	T-stat		-23.86		-21.73		-9.03		
DY	Coef			-0.012				-0.163***	-0.073***
	T-stat			-1.11				-26.00	-12.47
Cst	Coef	0.066***	0.301***	0.170***	0.222***	-0.056***	-0.075***	0.154***	0.072***
	T-stat	14.59	42.32	8.08	25.50	-13.36	-7.28	14.27	6.97
	R^2	0.1312	0.2312	0.0007	0.2967	0.4719	0.046	0.2153	0.5079

Table 3.2: Linkage between the implied arbitrage efficacy and other measures of funding illiquidity

The table reports regressions of the implied arbitrage efficacy (AE) on the TED spread, the VIX index and the Dividend yield of S&P 500 index, from September 7, 2000 to June 30, 2015. We run the regression on each individual variable and the joint set of variables, covering the full sample period (Panel A) and conditionally on the sign of the arbitrage efficacy (Panel B). All t-statistic are computed with Newey-West standard errors. ***, ** and * indicate significance at 1%, 5% and 10% levels, respectively.

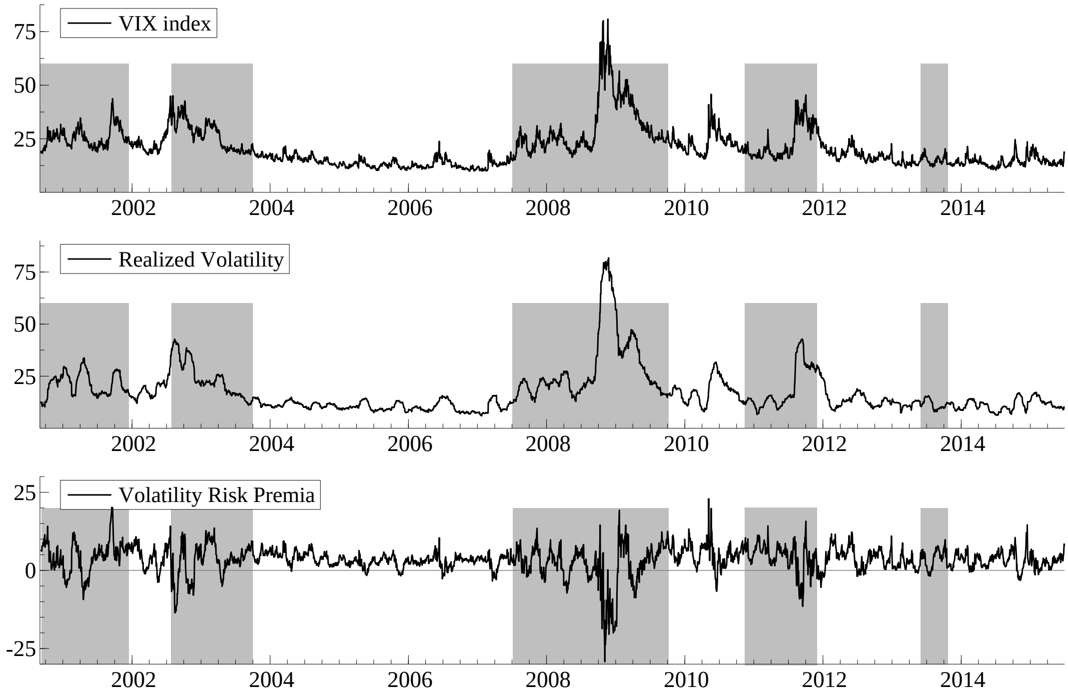

Figure 3.5: The plot of VIX index, realized volatility of S&P 500 index and the volatility risk premia, September 7, 2000 to June 30, 2015

The figure plots the time series of VIX index, Realized volatility of S&P 500 index and the volatility risk premia, covering the period from September 7, 2000 to June 30, 2015. The periods of inefficacious arbitrage are shaded.

Forecast by		Daily VIX (Change)					
		Aggregate		$AE > 0$		$AE < 0$	
		(1)	(2)	(3)	(4)	(5)	(6)
Lag 1 Daily VIX (change)	Coef	-0.123***	-0.127***	-0.089***	-0.089***	-0.129***	-0.140***
	t-stat	-4.50	-4.50	-3.38	-3.37	-3.56	-3.73
Lag 2 Daily VIX (change)	Coef	-0.119***	-0.120***	-0.102***	-0.102***	-0.126**	-0.127***
	t-stat	-3.08	-3.30	-2.90	-2.91	-2.45	-2.73
Lag 3 Daily VIX (change)	Coef	-0.045	-0.041	-0.055*	-0.054*	-0.038	-0.030
	t-stat	-1.44	-1.31	-1.86	-1.83	-0.91	-0.71
Lag 4 Daily VIX (change)	Coef	-0.105**	-0.102**	-0.061	-0.062	-0.114*	-0.106**
	t-stat	-2.20	-2.39	-1.49	-1.50	-1.81	-2.01
Lag 1 Daily AE (change)	Coef		-6.112**		-0.769		-9.979***
	t-stat		-2.20		-0.45		-2.58
Lag 2 Daily AE (change)	Coef		-4.352		-0.099		-8.396*
	t-stat		-1.34		-0.07		-1.80
Lag 3 Daily AE (change)	Coef		-4.810*		-1.311		-7.512*
	t-stat		-1.75		-0.81		-1.88
Constant	Coef	-0.000	0.000	0.005	0.005	-0.005	-0.013
	t-stat	-0.01	0.02	0.37	0.42	-0.25	-0.59
R-square		0.0327	0.0441	0.0191	0.0188	0.0369	0.0636

Table 3.3: The Granger causality test of VIX index on lagged value of Arbitrage efficacy

The table reports the forecasting regressions of the daily difference of VIX index on the lagged VIX index and the lagged arbitrage efficacy, respectively, in the full sample period and conditionally on the sign of arbitrage efficacy. We consider the choice of lag-order up to five (a week), and the Akaike (AIC) and Schwartz (SIC) information criteria are applied to select the preferred specifications. Data are daily from September 7, 2000 to June 30, 2015. All t-statistic are computed with Newey-West standard errors. ***, ** and * indicate significance at 1%, 5% and 10% levels, respectively.

Forecast by		Daily RV/VRP (Change)							
		Aggregate				$AE > 0$		$AE < 0$	
		RV	RV	VRP	VRP	RV	VRP	RV	VRP
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Lag 1 Daily VIX (change)	Coef	0.042***	0.039***	-0.166***	-0.166***	0.034**	-0.123***	0.035*	-0.176***
	t-stat	2.75	2.59	-5.81	-5.59	2.17	-3.71	1.75	-4.49
Lag 2 Daily VIX (change)	Coef	0.056***	0.058***	-0.175***	-0.178***	0.059***	-0.161***	0.057***	-0.185***
	t-stat	3.76	4.00	-4.87	-4.94	3.70	-4.06	2.90	-3.88
Lag 3 Daily VIX (change)	Coef	0.035**	0.035***	-0.080**	-0.076**	0.024**	-0.079**	0.040**	-0.071
	t-stat	2.52	2.62	-2.37	-2.32	2.37	-2.53	2.24	-1.60
Lag 4 Daily VIX (change)	Coef	0.031**	0.031**	-0.136**	-0.133***	0.039***	-0.101**	0.025	-0.132**
	t-stat	2.27	2.35	-2.53	-2.77	3.26	-2.15	1.47	-2.20
Lag 1 Daily AE (change)	Coef		1.061		-7.174**	-0.717	-0.051	2.107	-12.086**
	t-stat		0.83		-2.04	-1.10	-0.02	1.11	-2.50
Lag 2 Daily AE (change)	Coef		-2.972**		-1.379	-0.764	0.664	-4.462**	-3.933
	t-stat		-2.14		-0.67	-1.27	0.47	-2.18	-1.36
Lag 3 Daily AE (change)	Coef		-0.764		-5.575	0.416	-1.728	0.639	-8.152
	t-stat		-0.64		-1.58	0.78	-1.11	0.33	-1.50
Constant	Coef	0.000	0.000	-0.000	0.000	-0.003	0.009	0.003	-0.026
	t-stat	0.00	0.01	-0.01	0.00	-0.42	0.63	0.23	-0.69
R-square		0.0231	0.0293	0.0567	0.0672	0.0254	0.0404	0.0354	0.0842

Table 3.4: The Granger causality test of Realized Volatility index and Volatility Risk Premium on lagged value of VIX index and Arbitrage efficacy

The table reports the forecasting regressions of the daily difference of realized volatility (RV) of S&P 500, and that of the volatility risk premia (VRP), on the predictor variables: the VIX index and the implied arbitrage efficacy, respectively, in the full sample period and conditionally on the sign of arbitrage efficacy. The realized volatility of S&P 500 is computed as return volatility over the past one month, (Eq. 3.21) and the volatility risk premia is the difference between the implied and realized volatility (Eq. 3.22). Data are daily from September 7, 2000 to June 30, 2015. All t-statistic are computed with Newey-West standard errors. ***, ** and * indicate significance at 1%, 5% and 10% levels, respectively.

Table 3.3 reports the forecasting regression for changes in VIX on the lagged predictor variables: the VIX index itself and the implied arbitrage efficacy, in the full sample period (September 7, 2000 to June 30, 2015), the sub-period of efficacious arbitrage and the sub-period of inefficacious arbitrage. Since we use daily data, it is not obvious how quickly shocks to the implied arbitrage efficacy will transfer to market volatility. Therefore we allow for a choice of lag order up to five, and choose the preferred lag order based on the Akaike (AIC) and Schwartz (SIC) information criteria. On aggregate, the forecasting results are negative and significant at the 5% level, and the forecasting R^2 increases from 3.27% with only the lagged VIX index, to 4.41% with both the lagged VIX index and the implied arbitrage efficacy. It means that reduction in arbitrage efficacy is associated with an increasing financial market volatility in the future. In the conditional forecasting regressions, depending on (in)efficacious arbitrage, we can see in column (3) to (6) of Table 3.3 that the the degree of predictability is mostly prominent when arbitrage is inefficacious, e.g. binding leverage constraint, with the largest t-statistic and maximum R^2 of 6.36%. We continue by running the forecasting regression for both the realized volatility of S&P 500 and the volatility risk premia on the lagged predictor variables. We see that in Table 3.4, the volatility risk premia is being predicted by the implied arbitrage capacity with higher R^2, rather than the realized volatility of S&P 500, and the degree of predictability is much higher during the time of inefficacious arbitrage. Therefore, innovation in the arbitrage efficacy predicts the price of volatility risk, instead of the actual realized volatility itself.

These results accord with our model predictions: the efficacy of arbitrage predicts the future market volatility, and more importantly, the nonlinear effect of arbitrage efficacy on market volatility verifies the amplification effect during the period of the binding leverage constraint, which is consistent with Brunnermeier and Pedersen (2009). Also arbitrage efficacy predicts the price of volatility risk, which is an important risk factor in predicting stock market returns. The result therefore highlights the asset pricing consequence of funding liquidity, especially when funding constraint becomes binding.

3.5 Conclusion

In this paper we study the arbs' funding liquidity and the efficacy of arbitrage, and their linkage. We narrow our focus on how arbs exploit the mispricings caused by noise traders, when subjecting to leverage constraint. We define the arb's funding liquidity as their ability to raise leverage debt, which is captured by the marginal leverage with respect to mispricing, and the arbitrage efficacy as their ability to bear against mispricing, which is captured by the marginal correction achieved by the arbs. The model implies that the arbs' funding liquidity affects the efficacy of arbitrage, and more importantly the binding leverage constraint leads to inefficacious arbitrage and the amplification effect, such that market illiquidity and price volatility are extremely sensitive to noise shocks. Overall, we propose to capture the funding liquidity by the arbitrage efficacy, the sign of which identifies whether leverage constraint is binding or not.

We empirically estimate the implied arbitrage efficacy from the index arbitrage relationship between S&P 500 index and E-mini future, and find statistically significant evidence that the implied arbitrage efficacy is related to other broad measure of funding liquidity and significantly predicts the future market volatility and the volatility risk premia. More importantly, the sign of the implied arbitrage efficacy identifies the binding leverage constraint, such that the periods of inefficacious arbitrage exhibit strong amplification effects and coincide with the episodes of liquidity crises within the sample period. The measure of implied arbitrage efficacy thus provides vital and helpful tool for policy maker and regulators to evaluate the funding condition among the financial intermediaries and the potential existence of amplification due to the binding funding constraint.

Our work provides a number of direction that future researches might address. First, since the implied arbitrage efficacy can be estimated in various arbitrage relationship across different markets or countries, it would be interesting to extend the measure to several markets or countries. In doing so, one might be able to address the spillover and contagion effect in funding liquidity especially during the crisis period. Second, Brunnermeier and Pedersen (2009) suggest that funding illiquidity among arbitrageurs leads to the phenomenon of flight to liquidity, flight to quality, and commonality in liquidity. Hence, the implied arbitrage efficacy can be a good tool to distinguish the sample into two regimes: the period of loose and binding funding constraint, and empirically examine these hypotheses. Third, our paper only focuses on the initial mispricing correction, while the pattern of subsequent price recovery also reflects the impediments faced by arbs, as suggested by Duffie (2010). Combining both immediate and subsequent pricing dynamics might generate more fruitful results. Fourth, the amplification effect under binding funding constraint can lead to long-lasting consequences, which has been explored in the macroeconomics literature. It would be interesting to empirically verify the impact of funding liquidity in the financial sector on overall economics, especially during the period of inefficacious arbitrage.

Appendix 3.A The Empirical Design

Let sp_t be natural log of the spot price of the S&P 500 index; p_t and p_t^* be the natural log of the spot and fundamental price of the E-mini future contract respectively; r_t and q_t be the risk-free interest rate and dividend yield of the asset, respectively; $\tau_t = T - t$ be the time to maturity.

1. The fundamental value of the E-mini future contract can be implied by the cost of carry model:

$$p_t^* = sp_t + (r_t - q_t)\,\tau_t,$$

2. The dynamic mispricing error, z_t is then computed from the long-run equilibrium:

$$p_t = \mu + \theta p_t^* + z_t,$$

which is later used in the Error Correction model for computing mispricing correction.

3. The dynamic mispricing correction, κ_t is calculated by the Error Correction model in a rolling window basis, such that in a rolling window of 500 days, we run the regression:

$$\Delta p_t = \kappa z_{t-1} + \lambda^* z_{t-2} + \delta \Delta p_t^* + \gamma \Delta p_{t-1} + \mu_t,\ \mu_t \sim iid(0, \sigma_\mu^2).$$

where Δ is the difference operator. After assign each κ_t to the ending date of the window, we have the dynamic version of mispricing correction, κ_t.

4. The data of daily difference in mispricing correction and error, $\Delta\kappa_t$ and Δz_{t-1} is applied to compute funding liquidity measure. We first eliminate the outilers, i.e. time t when the estimated $\Delta\kappa_t$ or Δz_{t-1} is at the highest or lowest 1% tails of the distribution.

5. Funding liquidity measure is then computed by regressing the daily difference in mispricing correction on the changes in mispricing error. Similarly, the dynamic version of funding liquidity measure is computed in a rolling window of 250 days,

such that

$$AE_t = \sum_{t-250}^{t} \Delta\kappa_t \Delta z_{t-1} / \sum_{t-250}^{t} \Delta z_{t-1}^2.$$

Finally, we assign each AE_t to the ending date of window, i.e. time t, and results in the dynamic funding liquidity measure.

Appendix 3.B The Implied Arbitrage Efficacy and The Real World Events

We highlight the five market events during the sample periods, and associate them with the innovations in the implied arbitrage efficacy.

First, the market downturn in 2000-2003. The bear market began in 2000 after the burst of Dot-com bubble and finally reached the bottom in later 2002. AE, during bear market, fluctuated around zero, which provided a weak signal of financial instability. The 911 attack affected the market heavily as the occurrence of high volatility, but it did not seem to affect the funding liquidity on average.

Second, the Global Financial Crisis in 2007-2008. Prior to the crisis, the implied arbitrage efficacy, AE reached its peak on February 2007 at around 0.4 and started to drop ever since. AE experienced a sharp decline and sent the warning sign of a possible liquidity problem on June 2007. On exactly 27th of June, AE became negative whilst on the very next day, the Federal Open Market Committee (FOMC) voted to maintain the federal funds rate. Soon after then, the market began to be hit by news of liquidation, such as Bear Sterns was forced to liquidate two hedge funds that invested in mortgage-backed securities in July, and news of bankruptcy, such as the American Home Mortgage Investment Corporation in August and the Northern Rock in September. AE continued to drop, despite a number of liquidity expansion policy announced by Federal Reserve, until March 2008. There existed a sharp decline after the Lehman bankruptcy, since hedge funds using Lehman as a prime broker found it difficult to obtain capital and face a decline in the funding liquidity.

Third, the Flash Crash in 2010. The Flash Crash is one of the two market crash captured in the sample period along with a positive AE, around 0.1. The Flash Crash occurred on May 6, 2010, which started at 2:30 p.m., market indices collapsed and bottomed at 2:45 p.m. with up to 9% decline, and soon rebounded rapidly before 3:00 p.m.. After the Flash Crash, the implied arbitrage efficacy remained stable at round

0.1, and does not affected by the event. Overall, the result shows that the market is able to cope with a large unexpected shock when arbitrage capital is sufficient.

Fourth, the Debt Ceiling crisis and The Black Monday in 2011. The Debt ceiling crisis referred to the debate regarding the maximum borrowing that the US government is allowed to undertake. The contention was resolved before exhaustion by the Budget Control Act of 2011. The very next week on August 8, 2011 witnessed a large U.S. and global stock markets slide due to the credit rating downgrade of U.S. sovereign debt by Standard and Poor. The negative sign of *AE* warns the fragile financial condition more than 8 months ago. After a short reverse in May, it continued to drop till September 2011 and bounced back to zero after only two months. With sufficient liquidity supply, the S&P 500 index soon recovered its losses in the crash at the end of 2011.

Fifth, the Debt Ceiling crisis in 2013. The crisis began in January 2013 where the US government had reached the debt ceiling set in 2011, and ended in October after the enact of the Continuing Appropriations Act 2014. The trajectory of *AE* reflected the worries of the market participants. It was close to zero at the beginning of 2013, and became significantly negative after April. Soon after the crisis was resolved, *AE* rebounded in November 2011.

The trajectory of *AE* also reflects the effectiveness of the monetary policy announced by Federal Reserve in terms of liquidity expansion, which may help improve the efficacy of arbitrage. In reaction to the major market crashes since the market crash in 1987, Federal Reserve undertook a number of monetary policy to influence the liquidity condition in the market. Conventional tools include: open-market operations, setting the federal funds rate and reserve requirements. In general, *AE* tend to recover above zero after market turmoils, and we summarize the major recoveries in the arbitrage efficacy as follows.

At the start of the liquidity crisis in August 2007, Federal Reserve seek to ease liquidity constraints by providing short-term funding liquidity to banks and lowering the federal funds rate and primary credit rate. However, the traditional central bank tools did not appear to sound, as suggested by the innovation in *AE*. After the intervention of Federal Reserve since August, the rate of decline in *AE* became more gentle, comparing to the sharpness before August. However, it did not stop declining until May, 2008. After realizing the failure of the traditional moves, Federal Reserve carried out a number of new lending creations, such as the Term Auction Facility (TAF) in December 2007, and the Primary Dealer Credit Facility (PDCF) and Term

Securities Lending Facility (TSLF) March 2008. Despite the slightly rebound in AE since May, 2008, it was after November 2008 that AE witnessed a constantly rapid increment due to the introduction of the first round of QE (quantitative easing). Within the procedure of QE1, AE turned positive in November 2009, showing a massive improvement in funding liquidity.

Fed continued to implement the second of QE (QE2) in November 2010 and the third round of QE (QE3) in September 2012. According to the trajectories of AE, the impact and effectiveness of QE2 and QE3 seemed to diverge due to the worries of Debt Ceiling crisis. While the implied arbitrage efficacy continued to drop after the announcement of QE2, it witnessed a sharp increment and stayed positive after QE3.

Appendix 3.C Proofs

The derivation of main parameters, f_1^d and D_U, is straightforward but complicated. Thus, in this appendix, we provide the main analytic results derived using MATLAB12. We first provide the derivations for Proposition 3.

Proof. To derive the upper leverage limit D_U, we impose the following condition that the arbitrageurs' equity in period 2 evaluated at $\tilde{s}_2^b$ (financiers' estimate of s_2^b) is equal to 0, namely

$$\tilde{f}_2^b = f_1^e + (f_1^e + D_U)\left(\frac{\tilde{P}_2^b}{P_1} - 1\right) = 0 \tag{3.23}$$

where

$$P_1 = V - s_1 + f_1^e + f_1^d,\ \tilde{P}_2^b = V - \tilde{s}_2^b + \tilde{f}_2^b,\ \tilde{s}_2^b = \tau s_1 + s_2^b, \tag{3.24}$$

Using Eq.(3.23) and (3.24), we find that financiers set D_U by

$$D_U = \frac{1}{2}(m_1 - m_0)$$

where

$$m_0 = \left(f_1^e + s_2^b - (1-\tau)s_1\right)$$
$$m_1 = \sqrt{(m_0)^2 + 4f_1^e\left(V - s_1\tau - s_2^b\right)}$$

Under the condition that $\left(V - s_1\tau - s_2^b\right) \geq 0$, D_U is nonnegative. Therefore, we obtain the restriction on τ in Eq. (3.11):

$$0 \leq \tau \leq \frac{V - s_2^b}{s_1}. \tag{3.25}$$

To derive the relation between D_U, f_1^e and τ, we first write the partial derivative of D_U w.r.t f_1^e as

$$\frac{\partial D_U}{\partial f_1^e} = \frac{1}{2}\left(\frac{m_2}{m_1} - 1\right)$$

where

$$m_2 = 2V - \left(s_1 + s_1\tau + s_2^b - f_1^e\right).$$

Under the restriction of τ in Eq. (3.25), we find that $m_2 \geq m_1$, since

$$(m_2)^2 - (m_1)^2 = 4\,(V - s_1)\left(V - s_1\tau - s_2^b\right) \geq 0.$$

Therefore, we have $\frac{\partial D_U}{\partial f_1^e} \geq 0$. The equality holds when $D_U = 0$.

Next, we consider the partial derivative of D_U w.r.t τ given by

$$\begin{aligned}\frac{\partial D_U}{\partial \tau} &= -\frac{1}{2}\left(s_1 + \frac{(1-\tau)\,s_1^2 + s_1\left(f_1^e - s_2^b\right)}{m_1}\right)\\ &= -\frac{1}{2}\left(\frac{s_1\left(2f_1^e + m_1 - m_0\right)}{m_1}\right)\\ &< 0\end{aligned}$$

The inequality holds since the numerator is always positive because $f_1^e > 0$ and under Eq.(3.25), we must have $m_1 - m_0 > 0$. Therefore we have $\frac{\partial D_U}{\partial \tau} < 0$. **Q.E.D.**

Now we provide the proofs for Proposition 4.

Proof. We consider all possible strategies: the waiting strategy, the cautious strategy (we focus on the partial-leverage strategy) and the max-leverage strategy. Under the waiting strategy, the results in follow trivially for $f_1^d = D_L = -f_1^e$, $\ell^w = 0$ *and* $\alpha^w = 0$.

Under the partial-leverage strategy described in Proposition 4 (i), we can derive an optimal f_1^d by solving the first order condition in Eq.(3.5):

$$\frac{V}{P_1} - 1 = q\left(\frac{V}{P_2^b} - 1\right)$$

where

$$P_1 = V - s_1 + f_1, P_2^b = V - s_2^b + f_2.$$

For the special case of riskless arbitrage, $q = 1$, where the bad state must occur in period 2, then we must have $P_1 = P_2^b$ from the first order condition, and thus the optimal funding is $f_1^d = s_1 - s_2^b$. For another extreme, $q = 0$, where only good state occurs in period 2, we have $P_1 = V$ from the first order condition, and thus the optimal funding is $f_1^d = s_1 - f_1^e$. In both cases, it is easily seen that the funding liquidity is at its highest: $\ell^p = 1$.

For $0 < q < 1$, the optimal leverage fund is given by:

$$f_1^d = \frac{n_0 - n_2}{2(1-q)} \tag{3.26}$$

where

$$\begin{aligned} n_0 &= V - (1-q)\left(f_1^e + s_2^b - s_1\right) \\ n_1 &= V - (1-q)\left(s_1 + s_2^b - f_1^e\right) \\ n_2 &= \sqrt{(n_1)^2 + 4Vq(1-q)\left(s_2^b - f_1^e\right)} \end{aligned}$$

We use Eq.(3.12) and evaluate the funding liquidity ℓ^p under the partial-investment strategy by

$$\ell^p = \frac{\partial f_1^d}{\partial s_1} = \frac{(1-q) - (1-q)n_1/n_2}{2(1-q)} = \frac{1}{2}\left(1 - \frac{n_1}{n_2}\right) \tag{3.27}$$

Since $n_2 >| n_1 |$and $-1 <| \frac{n_1}{n_2} |< 1$, we have $0 < \ell^p < 1$. Therefore, risky arbitrage with $0 < q < 1$ dampens the arbs' willingness to raise capital.

Next, we use Eq.(3.14) and (3.26), and write the arbitrage efficacy under the partial-investment strategy as

$$\alpha^p = \frac{\partial \kappa_1}{\partial s_1} = \frac{\ell^p - \kappa_1}{s_1} = \frac{\ell^p}{s_1} - \frac{1}{s_1^2}\left(\frac{n_3 - n_2}{2(1-q)}\right) \tag{3.28}$$

where

$$n_3 = V - (1-q)\left(s_2^b - s_1 - f_1^e\right).$$

To show that $\alpha^p > 0$, we consider the worse case possible: the lowest α^p with the initial shock as large as s_2^b, i.e. $s_1 \to s_2^b$. Then α^p can be expressed as,

$$\begin{aligned}\lim_{s_1 \to s_2^b} \alpha^p &= \frac{s_2^b (1-q)\left(\frac{n_4}{n_5} - 1\right) + n_5 - n_4}{2(1-q)\left(s_2^b\right)^2} \\ &= \frac{n_5 - n_4}{2(1-q)\left(s_2^b\right)^2}\left(1 - \frac{s_2^b (1-q)}{n_5}\right)\end{aligned}$$

where

$$\begin{aligned}n_4 &= V - (1-q)\left(2s_2^b - f_1^e\right) \\ n_5 &= \sqrt{(n_4)^2 + 4q(1-q)V\left(s_2^b - f_1^e\right)}\end{aligned}$$

Therefore to ensure that $\lim_{s_1 \to s_2^b} \alpha^p > 0$, we must have

$$n_5 > s_2^b (1-q) \tag{3.29}$$

The right hand side in Eq. (3.29) reach its largest when $q \to 0$, i.e. $RHS = s_2^b$, while the left hand side is at its lowest when $q \to 0$, i.e. $LHS = V - 2s_2^b + f_1^e$. Thus the inequality in Eq. (3.29) holds conditional on

$$V > 3s_2^b - f_1^e.$$

Since $V \gg s_2^b, f_1^e$, the condition can be generally satisfied. Thus the inequality $\lim_{s_1 \to s_2^b} \alpha^p > 0$ holds, and we have $\alpha^p > 0$.

Now we prove the results under max-leverage strategy described in Proposition 4 (ii) and (iii). We first notice that $f_1^d = D_U$ under the max-leverage strategy. Hence, the funding liquidity is different from that under partial-leverage strategy, and is now given by

$$\ell^m = \frac{\partial D_U}{\partial s_1} = \frac{1}{2}\left(\frac{(1-\tau)m_1 - m_3}{m_1}\right)$$

where

$$m_0 = \left(f_1^e + s_2^b - (1-\tau)\, s_1\right)$$

$$m_1 = \sqrt{(m_0)^2 + 4 f_1^e \left(V - s_1 \tau - s_2^b\right)}$$

$$m_3 = f_1^e (1+\tau) + (1-\tau)\left(s_2^b + s_1 \tau - s_1\right).$$

For ℓ^m to be positive, we must have $(1-\tau)\, m_1 - m_3 > 0$, and thus following should hold:

$$m_4 = (1-\tau)^2 (m_1)^2 - (m_3)^2 = V(\tau-1)^2 + \left(s_2^b - f_1^e\right)(\tau - 1) - f_1^e > 0$$

m_4 is a quadratic equation with $\tau - 1$. It is easily seen that when $\tau = 0$, we have $m_4 = V - s_2^b > 0$ and $\ell^m > 0$. For $\tau = 1$, then we have $m_4 = -f_1^e < 0$ and $\ell^m < 0$. This shows that the threshold τ^* that determines the sign of ℓ^m must lie between 0 and 1, which is derived by setting $m_4 = 0$, such that

$$\tau^* = \frac{V - s_2^b + V - \sqrt{\left(s_2^b\right)^2 + 4\left(V - s_2^b\right) f_1^e}}{2\left(V - f_1^e\right)}.$$

Therefore, the result implies that ℓ^m is negatively correlated to the level of informativeness, τ, e.g. higher τ leads to lower ℓ^m; the sign of ℓ^m is determined by τ, such that $\ell^m > 0$ for $0 < \tau < \tau^*$ and $\ell^m < 0$ for $\tau > \tau^*$.

Second, we compare between $\ell^{m,\tau=0}$ (consider the highest ℓ^m with $\tau = 0$) and ℓ^p. Showing $\ell^{m,\tau=0} - \ell^p$ directly require some tedious mathematics, we thus apply an alternative method by comparing the value of $D_U^{\tau=0}$ and f_1^d in two extreme cases where $s_1 = 0$ and $s_1 = s_2^b$. For $s_1 = 0$, we have

$$D_U^{\tau=0} = \frac{1}{2}\left(\sqrt{\left(f_1^e + s_2^b\right)^2 + 4 f_1^e \left(V - s_2^b\right)} - \left(f_1^e + s_2^b\right)\right) > 0$$

and

$$f_1^d = \frac{1}{2(1-q)}\left(\left(V - (1-q)\left(f_1^e + s_2^b\right)\right) - \sqrt{\left[V - (1-q)\left(s_2^b - f_1^e\right)\right]^2 + 4Vq(1-q)\left(s_2^b - f_1^e\right)}\right).$$

We find that $f_1^d < 0$ after we compare the terms in the bracket, such that

$$\begin{aligned}
&\left(V-(1-q)\left(f_1^e+s_2^b\right)\right)^2-\left[V-(1-q)\left(s_2^b-f_1^e\right)\right]^2-4Vq(1-q)\left(s_2^b-f_1^e\right)\\
=&-4(1-q)f_1^e\left(V-s_2^b+qs_2^b\right)-4Vq(1-q)\left(s_2^b-f_1^e\right)\\
<&0.
\end{aligned}$$

Thus, we have $D_U^{\tau=0} > f_1^d$ under $s_1 = 0$. For another extreme $s_1 = s_2^b$, we also compare

$$D_U^{\tau=0} = \frac{1}{2}\left(\sqrt{\left(f_1^e\right)^2+4f_1^e\left(V-s_2^b\right)}-f_1^e\right)$$

where in the bracket we take the square of each term and obtain the following after rearrangement:

$$\left(f_1^e\right)^2+4f_1^e\left(V-s_2^b\right)-\left(f_1^e\right)^2=4f_1^e\left(V-s_2^b\right) \tag{3.30}$$

The equilibrium funding f_1^d can be written as

$$f_1^d=\frac{1}{2}\left[\frac{\left(V-(1-q)f_1^e\right)-\sqrt{\left[V-(1-q)\left(2s_2^b-f_1^e\right)\right]^2+4Vq(1-q)\left(s_2^b-f_1^e\right)}}{(1-q)}\right]$$

After we take the square of each term in the bracket, we have:

$$\begin{aligned}
&\left(\frac{V-(1-q)f_1^e}{(1-q)}\right)^2-\frac{\left[V-(1-q)\left(2s_2^b-f_1^e\right)\right]^2-4Vq(1-q)\left(s_2^b-f_1^e\right)}{(1-q)^2}\\
=&\frac{4(1-q)\left(s_2^b-f_1^e\right)\left(V-s_2^b+qs_2^b\right)-4Vq(1-q)\left(s_2^b-f_1^e\right)}{(1-q)^2}\\
=&4\left(s_2^b-f_1^e\right)\left(V-s_2^b\right)
\end{aligned} \tag{3.31}$$

After we compare the results in the bracket, i.e. Eq. (3.30) and (3.31), it is easily seen that when $f_1^e \lesssim \frac{1}{2}s_2^b$ (approximated), we have $f_1^d > D_U^{\tau=0}$ under $s_1 = s_2^b$. The result indicates that f_1^d and $D_U^{\tau=0}$ will interact at $0 < s_1 < s_2^b$, which means the slope of $D_U^{\tau=0}$ and f_1^d must satisfy $\ell^m < \ell^p$. On the other hand, for a higher $f_1^e \gtrsim \frac{1}{2}s_2^b$, we always have $f_1^d < D_U^{\tau=0}$, which means f_1^d and $D_U^{\tau=0}$ will not interact. Intuitively, arbs will never face the binding leverage constraint when their equity is large enough.

Third, we write the arbitrage efficacy under the max-leverage strategy by

$$\alpha^m = \frac{\ell^m - \kappa_1}{s_1}$$

and consider the case with $\tau = 0$ where funding liquidity ℓ^m is the highest. Then, we have

$$\alpha^{m,\tau=0} = \frac{1}{2s_1^2}\left(f_1^e - s_1 + s_2^b + m_5\right) - \frac{1}{2s_1}\left(\frac{f_1^e - s_1 + s_2^b}{m_5} - 1\right)$$

where

$$m_5 = \sqrt{\left(f_1^e - s_1 + s_2^b\right)^2 + 4f_1^e\left(V - s_2^b\right)}.$$

For $\alpha^{m,\tau=0}$ to be negative, we must have the following condition after rearrangement:

$$\begin{aligned} s_1 &< \frac{1}{s_2^b}\left(V\left(s_2^b + f_1^e\right) - \sqrt{V\left(V - s_2^b\right)}\left(s_2^b - f_1^e\right)\right) \\ &< \frac{2Vf_1^e}{s_2^b}. \end{aligned}$$

It can be simplified as

$$V > \frac{s_1}{2f_1^e}s_2^b,$$

It is easily satisfied in most cases since $V \gg s_1, s_2^b$. Therefore, we always have $\alpha^m < 0$ for $\tau = 0$. **Q.E.D.**

Appendix 3.D Robustness Check

In this section, we provide the robustness check with arbitrage efficacy estimated using different rolling windows. In particular, we estimate the implied arbitrage efficacy in Eq. (3.19) using 120-day and 500-day rolling windows. Figure 3.6 plots the daily series of the implied arbitrage efficacy obtained with 120-, 250- and 500-day rolling window. Table 3.5, 3.6 and 3.7 report the forecasting results of the market risk variables on the lagged arbitrage efficacy estimated using 120- and 500-day rolling window, where the predictability and the nonlinear consequence conditional on the sign of arbitrage efficacy are documented. The overall results are in line with that documented in Section 3.4.2, which are regressed by arbitrage efficacy using 250-day rolling window.

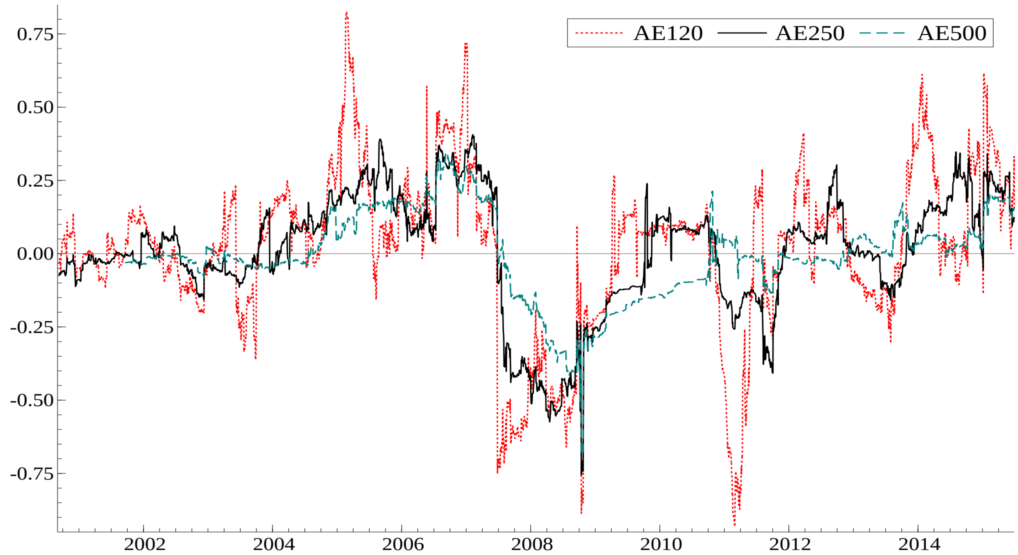

Figure 3.6: The plot of implied arbitrage efficacy estimated using different rolling windows, September 7, 2000 to June 30, 2015

The figure shows the implied arbitrage efficacy estimated using 120-day rolling window (dotted line), 250-day rolling window (solid line) and 500-day rolling window (dashed line). Data for arbitrage efficacy that estimated using 120- and 250-day rolling windows are daily from September 7, 2000 to June 30, 2015, while that estimated using 500-day rolling window are daily from September 18, 2001 to June 30, 2015.

Forecast by		VIX (change)					
		Agg	$AE > 0$	$AE < 0$	Agg	$AE > 0$	$AE < 0$
		(1)	(2)	(3)	(4)	(5)	(6)
Lagged VIX (change)		...					
Lag 1 AE (change)	Coef	-2.335*	-0.853	-3.689*	-9.764*	-0.497	-13.380**
	t-stat	-1.73	-0.88	-1.65	-1.91	-0.17	-2.14
Lag 2 AE (change)	Coef	-2.111	-0.245	-3.817	-6.961	2.516	-11.684
	t-stat	-1.34	-0.31	-1.35	-1.15	0.86	-1.53
Lag 3 AE (change)	Coef	-1.829	-0.367	-3.389*	-8.075*	-1.634	-11.087*
	t-stat	-1.32	-0.45	-1.39	-1.76	-0.58	-1.94
Constant	Coef	0.000	-0.011	0.004	-0.007	0.015	-0.033
	t-stat	0.00	-0.78	0.22	-0.28	0.14	-1.40
R-square		0.0392	0.0172	0.0592	0.0487	0.0126	0.0613

Table 3.5: The Granger causality test of VIX index on lagged value of arbitrage efficacy (120-day and 500-day rolling window)

The table reports the forecasting regressions of the daily difference of VIX index on the lagged VIX index and the lagged arbitrage efficacy that estimated using a 120-day (Column 1 to 3) and 500-day rolling window (Column 4 to 6). The results are documented in the full sample period and conditionally on the sign of arbitrage efficacy. We consider the choice of lag-order up to five (a week), and the Akaike (AIC) and Schwartz (SIC) information criteria are applied to select the preferred specifications. Data for arbitrage efficacy that estimated using 120-day rolling window are daily from September 7, 2000 to June 30, 2015, while that estimated using 500-day rolling window are daily from September 18, 2001 to June 30, 2015. All t-statistic are computed with Newey-West standard errors. ***, ** and * indicate significance at 1%, 5% and 10% levels, respectively.

Forecast by		Aggregate		$AE > 0$		$AE < 0$	
		RV	VRP	RV	VRP	RV	VRP
		(1)	(2)	(3)	(4)	(5)	(6)
Lagged VIX (change)		...					
Lag 1 AE (change)	Coef	0.479	-2.815*	-0.660*	-0.193	1.927**	-5.617**
	t-stat	0.73	-1.72	-1.91	-0.18	2.28	-2.03
Lag 2 AE (change)	Coef	-0.914	-1.196	0.235	-0.481	-2.055	-1.761
	t-stat	-1.11	-1.22	0.70	0.62	-1.47	-0.99
Lag 3 AE (change)	Coef	0.348	-2.178	-0.167	-0.199	0.777	-4.167
	t-stat	0.59	-1.23	-0.52	-0.25	0.71	-1.30
Constant	Coef	0.000	-0.000	-0.005	-0.006	0.006	-0.001
	t-stat	0.00	-0.00	-0.57	-0.40	0.44	-0.05
R-square		0.0248	0.0627	0.0132	0.0291	0.0356	0.0853

Table 3.6: The Granger causality test of Realized Volatility index and Volatility Risk Premium on lagged value of VIX index and Arbitrage efficacy (120-day rolling window)

The table reports the forecasting regressions of the daily difference of realized volatility (RV) of S&P 500, and that of the volatility risk premia (VRP), on the predictor variables: the VIX index and the implied arbitrage efficacy (estimated using 120-day rolling window), respectively, in the full sample period and conditionally on the sign of arbitrage efficacy. Data are daily from September 7, 2000 to June 30, 2015. All t-statistic are computed with Newey-West standard errors. ***, ** and * indicate significance at 1%, 5% and 10% levels, respectively.

Forecast by 500		Aggregate		$AE > 0$		$AE < 0$	
		RV	VRP	RV	VRP	RV	VRP
		(1)	(2)	(3)	(4)	(5)	(6)
Lagged VIX (change)		...					
Lag 1 AE (change)	Coef	2.848	-12.612**	-0.415	-0.082	3.495	-16.876**
	t-stat	1.45	-1.96	-0.41	-0.02	1.36	-2.14
Lag 2 AE (change)	Coef	-5.689**	-1.271	-1.651	4.167	-7.188**	-4.495
	t-stat	-2.20	-0.32	-1.11	1.27	-2.02	-0.98
Lag 3 AE (change)	Coef	1.599	-9.674*	-0.112	-1.522	1.702	-12.790*
	t-stat	0.71	-1.64	-0.05	-0.49	0.57	1.65
Constant	Coef	-0.003	-0.004	0.000	0.014	-0.003	-0.030
	t-stat	-0.20	-0.14	0.08	1.23	-0.20	-1.15
R-square		0.0336	0.0735	0.0152	0.0276	0.0400	0.0854

Table 3.7: The Granger causality test of Realized Volatility index and Volatility Risk Premium on lagged value of VIX index and Arbitrage efficacy (500-day rolling window)

The table reports the forecasting regressions of the daily difference of realized volatility (RV) of S&P 500, and that of the volatility risk premia (VRP), on the predictor variables: the VIX index and the implied arbitrage efficacy (estimated using 500-day rolling window), respectively, in the full sample period and conditionally on the sign of arbitrage efficacy. Data are daily from September 18, 2001 to June 30, 2015. All t-statistic are computed with Newey-West standard errors. ***, ** and * indicate significance at 1%, 5% and 10% levels, respectively.

Chapter 4

Revisiting the Value Premium Anomaly: Fundamental or Sentiment Risk?

4.1 Introduction

Value stocks tend to outperform growth stocks. Such value premium has been investigated extensively by two main approaches: the fundamental-based and sentiment-based theories. On one hand, Fama and French (1992, 1993, 1996) claim that both stocks are correctly priced, and value stocks are fundamentally riskier. As value stocks tend to be those companies under financial distress and potential bankruptcy, their higher average returns represent the compensations for the higher fundamental cash-flow risk. On the other hand, Lakonishok, Shleifer and Vishny (LSV 1994) suggest that value stocks are mispriced and relatively cheaper due to investor sentiment. In particular, irrational investors tend to be overly optimistic about the future growth prospect of growth stocks due to their past good earnings, but excessively pessimistic about value stocks. Barberis and Shleifer (2003) and Barberis, Shleifer and Wurgler (2005) suggest alternatively that due to systematic preference, value stocks are actually those out-of-favor stocks, while growth stocks are glamour stocks that are in favor by investors. Due to the limits of arbitrage (Shleifer and Vishny, 1997; Gromb and Vayanos, 2010), the correlated trading of irrational investors cannot be offset by the rational arbitrageurs, and thus results in a systematic price impact. Therefore, value stocks are more likely to have higher sentiment risk exposure, which is not priced in classical asset pricing theory.[1]

A substantial body of empirical studies have attempted to examine which view, the fundamental- or the sentiment-based one, provides a more appropriate explanation behind the debate on the value premium anomaly, though most empirical studies have been conducted under quite different frameworks. This is mainly because that the two views are implied by two potentially conflicting frameworks: the fundamental-based view follows the efficient market hypothesis whereas the sentiment-based one accords with the behavioral finance framework. In this regard, Campbell and Vuolteenaho (2004) have carefully addressed this issue by decomposing the CAPM beta into two components: the one related to cash-flow fundamental (fundamental risk) and the other related to the market discount rates (sentiment risk). They find that value stocks tend to have the higher fundamental-related betas, providing the support for the fundamental-based view on the value premium. However, the validity and the robustness of this approach depend crucially on the identification of the level of fundamental or sentiment risk exposures across value and growth stocks, which is often criticized in the literature (Daniel and Titmen, 2005; Lewellen et al., 2006; Phalippou, 2007).

To the best of our knowledge, however, there are no rigorous studies that have successfully tested the validity of the fundamental-based view against the sentiment-based one on the value anomaly under a common empirical framework. In this paper we aim to fill this gap by examining the distinct impacts of the fundamental and sentiment risk on arbitrage activity. Literature in limits to arbitrage suggests that fundamental risk and sentiment risk are two sources of arbitrage frictions that prevent arbitrage force from bringing price towards fundamental. Fundamental risk matters as reasonable and close substitutes for the underlying asset and short-selling opportunity are rarely available. Thus risk averse arbitrageurs who take on the arbitrage trade will be exposed to the unhedged fundamental risk, such that the realiza-

[1]Baker and Wurgler (2006) argue that classical asset pricing theory does not price investor sentiment. In particular, competitive and rational investors, who diversify to optimize returns in their portfolio, induce an equilibrium in which the asset price is informed by the present value of the expected future cash flows and the expected returns depends only on the systematic risk exposure. Even if irrational investors might misprice an asset, it will be offset by the force of arbitrage and thus have impact on asset price. However, arbitrage is far from costless and riskless in practice. Stocks that are costly and risky to arbitrage thus are more sensitive to the shifts in investor sentiment, which results in persistent mispricing and excessive price volatility. As noted in Lee, Shleifer and Thaler (1991) "*like fundamental risk, noise trader risk arising from the stochastic investor sentiment will be priced in equilibrium. As a result, assets subject to noise trader risk will earn a higher expected return than assets not subject to such risk. Relative to their fundamental value, these assets will be underpriced*".

tion of dividends may be better (worse) than expected. Sentiment risk stems from the future noise trader demand shocks, which may push price further away from fundamental and induce potential losses to arbitrageurs in a short run. Moreover, arbitrageurs are constrained by the agency frictions, e.g. periodic evaluations and funding withdrawal after poor performance, which shorten their investment horizons and expose them to the sentiment risk. Therefore higher fundamental or sentiment risk deters arbitrageurs willingness to conduct mispricing correction (De Long et al., 1990; Shleifer and Summers, 1990; Shleifer and Vishny, 1997; Mitchell, Pulvino and Stafford, 2002; Gromb and Vayanos, 2010).

Clearly, fundamental risk comes from the uncertainty of the future dividends when they are realized, while sentiment risk arises due to the unpredictability of interim resale price prior to the dividend realization. An open question is how they affect the arbitrage activity that corrects mispricing, similarly or differently. In order to address, we extend the work of Shleifer and Vishny (1997, SV hereafter) and Gromb and Vayanos (2010, GV hereafter), and develop the integrated model in which an asset is traded in a market with noise traders and fully rational risk-averse arbitrageurs. The asset is mispriced by noise traders in the short run, but will recover to fundamental value in the long-term. Rational arbitrageurs attempt to exploit the mispricing subject to both fundamental and sentiment risk. Arbitrageurs in our model are far from homogeneous, as we follow Bushee (1998, 2001)[2] and allow two heterogeneous arbitrageurs, denoted by the transient and dedicated arbitrageurs. Transient arbitrageurs are investors with the short investment horizon, high portfolio turnover and high trading frequency, and they tend to be concerned about the short-term sentiment risk exposure, which limits their willingness to bet against the noise traders. Dedicated arbitrageurs are those who have the long holding period, less diversified portfolio and low turnover, and therefore they are affected by the fundamental risk rather by the short-term sentiment risk.[3]

[2]Arbitrage is often performed by specialized but heterogeneous institutional investors. Bushee (2001) classify institutional investors into three groups: transient, quasi-indexer and dedicated, based on their portfolio turnover, diversification, and momentum trading.

[3]Barberis and Thaler (2005) conclude the conditions that make arbitrage costly and risky. When a mispriced asset does not have a perfect substitute, arbitrage is deterred by fundamental risk if (i) arbitrageurs are risk averse and (ii) fundamental risk is systematic. Even if a perfect substitute exists, sentiment risk remains a concern to arbitrageurs if (i) arbitrageurs are risk averse and have short horizons and (ii) the sentiment risk is systematic. Condition (i) guarantees that mispricing is not wiped out by a single large arbitrageur, and condition (ii) ensures that it is not wiped out by large number of small arbitrageurs. In the presence of learning costs and transaction costs, condition (ii) may not be necessary, as it is too costly for small investors.

Our theoretical framework suggests that the expected arbitrage return is jointly determined by the investors sentiment that induces the initial mispricing, the funding constraint faced by the transient arbitrageurs, the fundamental risk component that arises from asset payoffs and the sentiment risk component that stems from the future investors sentiment. More importantly, the impacts of fundamental and sentiment risk on arbitrage return are channeled through the arbitrage activity towards those risks. To this end, we follow the recent paper of Cai et al. (2015, CFS hereafter)[4] to capture the arbitrage activity by the initial mispricing correction and the subsequent noise momentum coefficients. The initial mispricing correction measures the proportion of mispricing corrected by arbitrageurs in the initial period while noise momentum captures the degree of persistence of unarbitraged errors in the subsequent period.

We derive a number of theoretical predictions on the arbitrage activity, the validity of which can be tested in the subsequent empirical application. The first prediction obtained under higher fundamental risk is that dedicated arbitrageurs tend to reduce their initial and subsequent investments in exploiting the mispricing. Arbitrage activity is thus limited, representing as a lower initial mispricing correction, a higher noise momentum and a slower speed of price adjustment. The second prediction is achieved under higher sentiment risk, where transient arbitrageurs refuse to bet against noise trader risks initially, but save funding for the next period when noise shock intensifies. As a result, the initial mispricing correction is deterred, but subsequent noise momentum is also reduced since transient arbitrageurs have more funding to deal with the future mispricing; the combined impact on speed of adjustment is rather uncertain. Overall, higher fundamental and sentiment risk tend to deter the initial mispricing correction, but have opposite impacts on the subsequent noise momentum.

In the empirical application, we consider a value and a growth portfolio, i.e. the S&P 500 value and growth index. The S&P 500 value (growth) index is a market-capitalization-weighted index, consisting of those stocks within the S&P 500 index

[4]To capture such multi-period arbitrage activity, CFS develop a generalized error correction model (GECM) and estimate both the initial mispricing correction and the subsequent noise momentum parameters where the latter is designed to measure persistence of the uncorrected pricing errors. Applying it to a wide range of international spot-futures market pairs, CFS documents pervasive evidence of noise momentum around the world. In this unified theoretical framework, a higher initial mispricing correction and a lower mispricing persistence induce a faster overall speed of adjustment. Further, Chapter 2 investigate the relation between arbitrage activity and the size of mispricing error in the time series, which have strong implication on the nonlinearity of limits to arbitrage.

and NASDAQ that have strong value (growth) characteristics. Across the value and growth index, we attempt to analyze the arbitrage activity implied by the index-future arbitrage relation. The fundamental of index future can be inferred by the spot price and the cost of carry, consisting the dividend yield on the index and a risk-free interest rate. Text book arbitrage implies that rational investors actively exploit the mispricing between the implied fundamental and the future spot, and guarantee the law of one price. However, the spot-future arbitrage is far from costless and riskless in practice. First, in order to exploit the spot-future mispricing, arbitrageurs are challenged to replicate all the component stocks in S&P value and growth index with appropriate weights. The difficulty in obtaining the perfect substitute imposes the fundamental risk on arbitrageurs, since it cannot be fully hedged.[5] Second, hedge funds, as commonly believed to be rational arbitrageurs, tend to use leverage to support their operation, and thus face potential funding constraints and agency frictions (SV and Fung and Hsieh, 2013). These impediments force hedge funds to concern about the short-term sentiment risk before price convergence. As documented in Chen, Han and Pan (2014), hedge fund returns are associated with their exposure to sentiment risk. As a result, sentiment risk is also a major concern for the arbitrageurs who exploiting the spot-future arbitrage relation.

To examine which prediction is valid empirically, we employ the two-stage methodology as follows. First, to uncover the time variation of the value effect, we follow Guidolin and Timmermann (2008) and apply the regime-switching VAR models to the HML returns obtained from the value and growth index and futures. This produces three regimes with distinctive and significant HML returns as follows: the value premium, the value discount, and the no-anomaly. The regime without anomaly has a HML returns close to zero, corresponds to the highly persistent bull markets with low volatility over 2004-2008 and 2010-2014. Value premium regime appears in the market turmoils with the highest volatility, e.g. the market downturn in 2000, 2002 and 2008. Finally, a regime with significant value discount is likely to be the persistent bear market over 2000-2003 and 2008-2010, which is also quite volatile.

Once we identify the three regimes by the HML returns, we apply the two-period generalized Error Correction Model[6] to each regime and evaluate the arbitrage ac-

[5] Richie, Daigler and Gleason (2008) and Marshall, Nguyen and Visaltanachoti (2013) argue that trading S&P 500 future against the replicated S&P 500 index is costly. First, it involves large transaction cost to replicate the index, and liquidity risk to implement arbitrage strategy; Second, as stocks in the index do not trade at the same time, the replicated index price is often stale;

[6] The methodology is introduced in CFS.

tivity across value and growth. We find that the arbitrage activity implied by the value (growth) stocks tend to limited by higher exposure to sentiment risk, since we observe lower mispricing correction and noise momentum under the value premium (discount) regime whilst the arbitrage activity is similar under the no-anomaly regime. Moreover, the overall speed of adjustment, determined jointly by the mispricing correction and the noise momentum coefficients, are more or less similar across value and growth in each regime, indicating that the impacts of sentiment risk on overall speed of adjustment are generally unimportant. Overall, these empirical evidence suggests that the period of value premium (discount) is characterized with higher sentiment risk within value (growth), which provides strong support for the sentiment-based view on the value premium anomaly.

To this end, our paper contributes to the literature by providing the first theoretical framework to reveal the distinctive impact of fundamental and sentiment risk on the arbitrage activity. In addition, the empirical evidence from the S&P 500 value and growth index have important implications behind the ongoing debate of value premium anomaly, suggesting that the value premium or discount associated with the large-cap stocks is mainly driven by the investor sentiment. Nevertheless, our empirical study is based not on the stock portfolios that are sorted by book-to-market ratio, but rather on the large-cap S&P500 value and growth indices. The arbitrage activity is not implied by the entire stock market, but rather by the index spot-future arbitrage. These important limitations require some cautious interpretations of our results. In additional, further studies will be warranted to investigate the possibility that both the fundamental- and sentiment-based theories will jointly explain the value premium anomaly.

The rest of this chapter is organized as follows. Section 4.2 reviews the related literature, while an integrated model of limits to arbitrage is introduced in Section 4.3. Section 4.4 develops the main theoretical predictions under the fundamental/sentiment-based view. Section 4.5 shows the two-stage methodology for empirical application and presents the main results. Section 4.6 concludes. All the proofs are relegated in the Mathematical Appendix.

4.2 Literature Review

4.2.1 Tests of the fundamental-based View

The fundamental-based explanation is derived under the Efficient Market Hypothesis. In an efficient market without fractions, economists have long evaluated the risk of a given stock by its beta, i.e. the sensitivity of a stock's return to the return on the market as a whole. Thus fundamental-based explanation predicts that value stocks with higher expected returns should have higher betas. However, the empirical results are mixed. Fama and French (1992) and others show that the CAPM of Shape (1964) and Lintner (1965) cannot account for the value premium, i.e. value stocks with higher expected returns yet do not have higher betas. LSV report that superior return in value stocks are not accompanied by notable risks, measured by beta and volatility. Rather, Fama and French (1996) find that value betas are higher than growth betas from 1926 to 1963, but fail to reach the same conclusion from 1963 to 2004. Moreover, they argue that the book-to-market ratio is proxy for the firm's financial distress risk that is not captured by the single beta. The anomaly tend to disappear when adding size and value factors, i.e. the SMB and HML factors in the CAPM model.

Another implication from the fundamental-based view is further investigated by Zhang (2005), who suggest that due to the counter-cyclical price of risk, the return of value-minus-growth strategies is high (low) in bad (good) times when the expected premium for risk is high (low). Early studies by LSV find that performance in value are higher than that in growth during both good times and bad times. Debondt and Thaler (1987) and Chopra et al. (1992) also find the similar evidence that cannot be explained by fundamental-view alone. Rather, Petkova and Zhang (2005) revisit this issue and find that value-minus-growth betas tend to co-move positively with the expected market risk premium, which supports the fundamental view. Choi (2013) also shows that the beta and leverage condition of value companies are more sensitive to economic conditions than growth companies; thus the beta in value sharply increases during market downturns, which is consistent with the fundamental-based view.

4.2.2 Tests of the sentiment-based View

Researchers in behavioral finance introduce an alternative explanation for the value premium anomaly: the sentiment-based explanation, which rests on two necessary conditions. First, stocks are mispriced due to correlated sentiments. The aggregate trading of noise traders, instead of canceling each other, generates the non-fundamental demand shock to push asset price away from fundamental. Second, there must be limits to the ability and willingness of rational arbitrageurs to offset the mispricings, the mispricings thus persist. There are a large number of theoretical and empirical literature in behavioral finance attempting to verify these conditions and the relation with the value premium anomaly.

For the theoretical base of the first condition, LSV introduce the extrapolation view, such that investors tend to be overly optimistic about firms with good past performance, while overly pessimistic to those with poor past performance. Barberis and Shleifer (2003) introduce the category view, such that investors tend to group assets into categories such as small-cap stocks, value stocks, etc. Noise traders might allocate more funds in growth stocks, whilst withdrawing from value stocks. As a result, value (growth) stocks co-move even regardless of their cash flow fundamentals. Empirical studies have provided supportive evidence. Barberis, Shleifer and Wurgler (2005) test the category view in explaining the index effect, i.e. index inclusion or exclusion, and find positive results that verifies the sentiment-based explanation. Kumar (2009) also provides empirical evidence for the category view, such that individual investors tend to systematically shift their preferences across certain portfolios, like value versus growth, and such behaviors are not induced by the changes in future cash flow or macroeconomic variables. Empirical evidence also documents that sentiment risk is priced in the equilibrium. Baker and Wurgler (2006, 2007) show that sentiment risk is systemic, and sentiment is more likely to be priced among stocks that suffer from severe arbitrage difficulty, such as small stocks, high volatility shocks and distressed stocks. Barber, Odean and Zhu (2009) find that tradings in individual investors are correlated and persistent, and have significant systematic pricing impact. Beer, Watfa and Zouaoui (2011) also document the systematic sentiment risk premium in the financial market, such that portfolio returns are associated with its exposure to sentiment risk.

Much more work on sentiment-based explanations has been done to examine the limits to arbitrage, such that arbitrageurs require higher premium in exploiting the arbitrage opportunity with larger frictions, which results in larger value premium. In this regard, a number of literature tests the relation between the limits of arbitrage and the magnitude of value premium. Ali et al. (2003) find that value premium is greater for stocks with higher idiosyncratic return volatility, higher transaction costs and lower investor sophistication. Griffin and Lemmon (2002) find that stocks with low analyst coverage exhibit value premium. Phalippou (2004) uses institutional ownership as a indicator of the level of arbitrage cost, and shows that stocks that held by institutional investors do not exhibit any significant value premium. Nagel (2005) claims that value premium is most prominent among stocks with low institutional ownership, where short-selling constraints tend to bind. Agarwal and Wang (2007) find that value premium disappear after controlling for transaction costs.

4.2.3 Joint tests

Instead of testing the explanations of value premium individually, our extension on the study of arbitrage activity allows us to identify the specific risk exposure cross value and growth stocks, thus test the fundamental-based view against the sentiment-based one. A substantial body of theoretical and empirical work have also attempted to do so. Theoretically, it is difficult to test one view against the other in absence of an integrated framework. The work of Daniel, Hirshleifer and Subrahmanyam (2001a) first offer an explicit theoretical model in which asset returns are jointly predicted by the CAPM beta and the current mispricing. In their model, investors receive information about the asset's systematic and firm-specific idiosyncratic factors. Mispricing arises due to the overconfident investors about the factors, and rational and risk-averse investors then enter the market to exploit the pricing error. However, as mispricing is induced by information signal possessed by traders, it is unobservable in reality.

Empirical methodologies and evidence are mixed. Daniel and Titman (1997) deny the fundamental-based view by showing that return of value/growth stocks does not associated with the factor model. Rather they find that value stocks tend to have similar characteristics and co-vary with one another despite of being distressed or not. Daniel et al. (2001b) extend the former work to the Japanese stock market, which again rejects the fundamental-based explanation. Chui et al. (2012) test the two explanations jointly on a country level, i.e. the relation between value premium and investors risk aversion (fundamental-based) or stock market development (sentiment-based). They find that value premium is higher in countries where investors tend to be more risk averse, but fail to find any relation with market development. These results support the fundamental-based explanation. The beta decomposition of Campbell and Vuolteenaho (2004) provides a direct method that are able to determine whether the factors that predict future returns are related to fundamentals or investors sentiment. Specifically, they extend the Sharpe-Lintner CAPM, and decompose the single beta into two components: a cash-flow beta and a discount-rate beta. They find that value stocks tend to have higher beta that related to cash-flow fundamentals than growth stocks. Moreover, Campbell, Plok and Vuolteenaho (2009) further construct the proxies for news about cash-flow fundamentals and sentiment, and show that the systematic risk of value and growth stocks are mainly determined by cash-flow fundamentals, which further verifies fundamental-based view. Santos and Veronesi (2005) also find that dividends in value tend to comove more with the macro-economy than those of growth, and thus value stocks are exposed to higher cash flow risk rather than sentiment risk.

However, the various adaptations of the CAPM, such as the Fama and French (1993) three factor model and beta decomposition in Campbell and Vuolteenaho (2004), are often questioned by its validity and robustness. Daniel and Titmen (2005) argue that the reason why the statistical tests fail to reject the models is because the tests have lack of power to reject the models. Lewellen et al. (2006) also accuse for the statistical tests, such that high R^2 or low pricing error is not sufficiently good indicators to evaluate the models. Phalippou (2007) tests the validity of various adaptations of the CAPM by reassessing the robustness of the results on different time periods and sets of data. The author finds significantly large pricing errors after sorting stocks on book-to-market ratio and institutional ownership, and thus questions the validity of these asset pricing model. Our paper attempts to avoid this problem by applying an alternative but unified approach to distinguish fundamental and sentiment risk via their distinctive impacts on the arbitrage activity.

4.3 The Model

4.3.1 Market structure

Our analysis builds on an extended model of SV and GV. We consider a risky asset in the market, trading in period $t = 1, 2$ and pays off in period $t = 3$. The riskless rate is exogenous and set to zero. For tractability, payoffs in period 3, d_3 are assumed to be normal with mean, $\bar{d}$, and standard deviation, δ_d. The fundamental risk δ_d thus stems from the uncertainty of future payoffs. There are two types of agents in the market, noise traders and rational arbitrageurs. The noise traders tend to push the asset price away from its fundamental, while rational arbitrageurs observe the price discrepancy and enter the market to offset the mispricing.

4.3.1.1 Noise traders

Noise traders do not know the expected pay off value $\bar{d}$ in period 3. Suppose that (i) noise traders experience pessimistic shocks, S_t,[7] before final payoffs are realized. In period 1, S_1 is known to arbitrageurs, while S_2 is a normally distributed random variable with mean, $\bar{S_2}$, and standard deviation, δ_n. The expected noise shocks in period 2 is assumed to intensify, such that $\bar{S_2} > S_1$. (ii) In period 3, shocks become zero, i.e. $S_3 = 0$, as the asset pays off dividends (iii) S_t are independent of asset payoffs, d_3. To this end, the sentiment risk, δ_n, in our model arises due to the noise trader shock that might affect asset price in period 2, which is independent from the fundamental. In absence of sentiment, the price at period 1 and 2 equal to the asset's fundamental, i.e. the expected payoff $\bar{d}$.

4.3.1.2 Arbitrageurs

Rational arbitrageurs enter the market to exploit the price discrepancies at different trading period, or in other words, intertemporal arbitrage. Unlike SV and GV, arbitrageurs in our model are far from homogeneous, as we distinguish between the transient and dedicated arbitrageurs (denoted with arbitrageur types $j \in (T, D)$). Both types of arbitrageurs are risk averse, and maximize the expected CARA utility with risk tolerance, γ^j:

$$E\left(U\left(f_t^j\right) \mid \mathscr{F}_t^j\right) = E\left[-exp\left\{-\frac{1}{\gamma^j} f_t^j\right\} \mid \mathscr{F}_t^j\right] \tag{4.1}$$

where $\mathscr{F}_t^j$ is the price information available to arbitrageurs j in period t and f_t^j is the funding accumulated by type j arbitrageurs from external investors in period t in order to conduct the arbitrage trade. The initial funding f_1^j is exogenously given while f_2^j and f_3^j are endogenously determined, such that the funding in period $t+1$

[7] Results are symmetric under optimistic noise shocks.

can be expressed by:[8]

$$f_{t+1}^{j} = f_t^j + \beta_t^j f_t^j \left(\frac{P_{t+1}}{P_t} - 1 \right). \tag{4.2}$$

where P_t is the price at period t and β_t^j is the optimal position for type j arbitrageurs. We allow both types of arbitrageurs to long or short the asset, but they have to fully collateralize their position, i.e. $-1 \leq \beta_t^j \leq 1$. The market clearing condition in period t is that demand for the asset must equal to the unit supply, from which the asset price can be derived as:

$$P_t = \bar{d} - S_t + \beta_t^T f_t^T + \beta_t^D f_t^D. \tag{4.3}$$

The arbitrageurs are further distinguished as follows. First, transient arbitrageurs do not have long-term access to capital, i.e. they must close their position in period $t+1$ after they enter the market in period t, and re-raise the capital in period $t+1$. Theoretically, there is no guarantee that transient arbitrageurs will be able to fully re-raise the amount of f_2^T implied by Eq. (4.2). But without loss of generality, we consider only the case where they fully re-raise the amount of f_2^T. Due to the short-term access to capital, transient arbitrageurs has short investment horizon and are exposed to sentiment risk. Second, transient arbitrageurs tend to overlook the long-term fundamental risk since they must close their position in a short run. As a result, their strategy in period 2 when mispricing is not fully corrected is making the full investment, i.e. $\beta_2^T = 1$. Third, transient arbitrageurs are assumed to face funding constraint, $\bar{S}_2 > f_1^T$.[9] In contrast, dedicated arbitrageurs are endowed with

[8] In SV, the performance-based arbitrage (PBA) is introduced, such that outside investors might augment or withdraw funding from arbitrageurs based on the previous performance. It is represented by the PBA sensitivity, α, such that the funding in period $t+1$ can be expressed by

$$f_{t+1}^{j} = f_t^j + \alpha \beta_t^j f_t^j \left(\frac{P_{t+1}}{P_t} - 1 \right),$$

where α is normally greater than 1, indicating that if arbitrageurs had a bad track record, i.e. $P_{t+1} < P_t$, outside investors will withdraw funding. With a larger PBA sensitivity, arbitrageurs tend to reduce their initial investment. The risk aversion assumption on arbitrageurs in our model will act the same as they face a larger α; they would require higher premium to bear risk. For simplicity, we set $\alpha = 1$ to avoid some tedious calculations in what follows.

[9] Notice that we impose the assumption on the transient arbitrageurs' funding, while there is no restriction on that of the dedicated arbitrageurs. Since dedicated arbitrageurs are risk averse and require a certain compensation for bearing the risk, thus cannot fully correct the mispricing in period 1. Transient arbitrageurs, although are also risk averse, but tend to overlook the fundamental risk in period 3, thus are expected to fully correct mispricing in period 1 if they do not face this funding constraints. However, as we are interested in the cases where mispricing persists in a short run.

sufficient long-term finance, thus they are able to hold the position and realize the arbitrage return when price finally converges. Therefore dedicated arbitrageurs make investment decision in period 1 with unrestricted funding, f_1^D, and hold to the final period 3, which expose them to fundamental risk. Finally, we assume that the investment of transient arbitrageurs in period 1 can be observed by dedicated arbitrageurs, and vice versa.

The price information received by the two types of arbitrageurs can thus be summarized as follows. The information set for dedicated arbitrageurs in period 1, $\mathscr{F}_1^D$ can be described as $\left\{\bar{d}, S_1, \gamma^D, \delta_d, \beta_t^T f_t^T, P_1\right\}$, while that for transient arbitrageurs in period 1, $\mathscr{F}_1^T$ can be described as $\left\{\bar{d}, S_1, \gamma^T, \beta_t^D f_t^D, \bar{S_2}, \delta_n, P_1\right\}$.

4.3.2 Optimization problems

We first derive the optimal strategy for dedicated arbitrageurs. Since dedicated arbitrageurs determine their overall investment strategy in period 1 and liquidate only when price converges in period 3, and by the use of normality distribution in d_3, they will maximize (See Appendix for proof):

$$E\left(f_3^D \mid \mathscr{F}_1^D\right) - \frac{1}{2}\gamma^D Var\left(f_3^D \mid \mathscr{F}_1^D\right),$$

subject to the restriction on collateralization, i.e. $-1 \leq \beta_1^D \leq 1$. We can derive the solution for β_1^D as (See Appendix for proof):

$$\beta_1^D = \frac{P_1\left(\bar{d} - P_1\right)}{f_1^D \gamma^D \delta_d^2} \approx \frac{\bar{d}\left(S_1 - \beta_1^T f_1^T\right)}{f_1^D\left(\bar{d} + \gamma^D \delta_d^2\right)} < 1, \, otherwise \, \beta_1^D = 1 \tag{4.4}$$

The result is intuitive. First β_1^D is always non-negative since $\bar{d} - P_1 \geq 0$, thus dedicated arbitrageurs will not short the asset. Second $\beta_1^D = 0$ when $\bar{d} - P_1 = 0$, i.e. dedicated arbitrageurs stop investing when price is equal to fundamental. Third, dedicated arbitrageur's position are affected by the initial shock, S_1, the position of transient arbitrageurs, $\beta_1^T f_1^T$, the funding in hand, f_1^D, the fundamental risk, δ_d and their risk tolerance, γ^L. Fourth, the total investment $\beta_1^D f_1^D$ is independent from funding f_1^D when $\beta_1^D < 1$, which means that the dedicated arbitrageur's ability to raise funding does not affect their investment unless they are fully invested, $\beta_1^D = 1$. Full investment is more likely to occur when f_1^D is relatively small, such that long-term

Imposing this simple assumption assures that in period 2 transient arbitrageurs cannot fully correct the mispricing.

funding is scarce for dedicated arbitrageurs, and the ability to raise funding will directly affect arbitrageurs' investment.

Second, we derive the optimal strategy for transient arbitrageurs. Transient arbitrageurs care about the resale price in period 2 since they have to close their position in this period. Therefore, they aim to maximize the utility in period 2, i.e. $E\left(U\left(f_2^T\right) \mid \mathscr{F}_1^T\right)$, subject to the funding constraint, $\bar{S_2} > f_1^T$, and the restriction on collateralization, i.e. $-1 \leq \beta_1^T \leq 1$, from which we obtain the following result (See Appendix for proof):

$$\beta_1^T = \frac{\left(\bar{d} - S_1\right)\left(f_1^T + S_1 - \bar{S_2}\right)}{f_1^T\left(\bar{d} - S_1 + \gamma^T \delta_n^2\right)} > -1, otherwise\, \beta_1^T = -1. \tag{4.5}$$

We describe a few implications. First, $\beta_1^T < 1$, such that, transient arbitrageurs never make full investment to long the asset due to the expectation that noise shock will intensify in period 2, $\bar{S_2} > S_1$. Second, the equilibrium position is determined by the initial noise trader shock, S_1, the expectation of future shock, $\bar{S_2}$, the sentiment risk, δ_n, their own risk tolerance γ^T and the funding in hand f_1^T. Notice that the position of transient arbitrageurs is determined independently from dedicated arbitrageurs, because dedicated arbitrageurs do not withdraw or augment their investment in period 2 when transient arbitrageurs close their position. Third, transient arbitrageurs might short the asset. β_1^T is positive (long) when $\bar{S_2} < f_1^T + S_1$, i.e. the future noise is relatively small, and is negative (short) when $\bar{S_2} > f_1^T + S_1$, such that expected future noise is too large. When $-1 < \beta_1^T < 0$, price deviation is rather amplified by transient arbitrageurs, and dedicated arbitrageurs become the only rational agents to exploit the price discrepancy.

4.3.3 The expected arbitrage return

After deriving the optimal strategy in Eq. (4.4) and (4.5), we obtain the expected arbitrage return over the whole period as the expected payoff minus the equilibrium price,

$$E\left(R\right) = \bar{d} - P_1 = S_1 - \left(\beta_1^T f_1^T + \beta_1^D f_1^D\right) \tag{4.6}$$

where β_1^D and β_1^T satisfy Eq. (4.4) and (4.5).

Consider first the special case with partially-invested transient arbitrageurs only.[10]

[10] Transient arbitrageurs only make full investment when they short asset, i.e. $\beta_1^T = -1$. Under this case, the expected return is obtained as $E\left(R\right) = S_1 + f_1^T$, which is independent from the sentiment risk, δ_n.

Then the expected arbitrage return can be written as (See proof in Appendix):

$$E^T(R) = S_1 - \beta_1^T f_1^T \approx \left(S_1 + f_1^T - \bar{S}_2\right) R^n + \left(\bar{S}_2 - f_1^T\right) \tag{4.7}$$

Eq. (4.7) shows that the expected return contains three components: the initial shock, S_1, the funding constraint, $\bar{S}_2 - f_1^T$, and the sentiment risk premium,

$$R^n = \frac{\gamma^T \delta_n^2}{\bar{d} + \gamma^T \delta_n^2}.$$

The sentiment risk premium lies between 0 and 1. It is zero when there is no sentiment risk, $\delta_n = 0$, thus arbitrageurs would not require any premium. It becomes one when arbitrageurs are extremely risk averse, $\gamma^T \to \infty$, and ask for the greatest compensation for bearing the risk. In absence of the sentiment risk, i.e. $R^n = 0$, the arbitrage return equals to the funding constraint, $\bar{S}_2 - f_1^T$.[11] In absence of the transient arbitrageurs (equivalent to extreme risk aversion, i.e. $R^n = 1$), the asset return becomes S_1, which means no mispricing has been corrected by the arbitrageurs. Overall, the arbitrage return is higher when there is larger noise trader demand shocks (larger S_1); when the trade is exposed to higher sentiment risk (larger δ_n); when transient arbitrageurs are more risk averse (higher γ^T) and more limited in obtaining capital (larger $\bar{S}_2 - f_1^T$).

Next consider another special case with partially-invested dedicated arbitrageurs only.[12] According to Eq. (4.4), we rewrite the expected return as (See proof in Appendix):

$$E^D(R) = S_1 - \beta_1^D f_1^D = S_1 R^d \tag{4.8}$$

The arbitrage return compose of two components, the initial noise trader demand shock, S_1, and fundamental risk premium required by dedicated arbitrageurs,

$$R^d = \frac{\gamma^D \delta_d^2}{\bar{d} + \gamma^D \delta_d^2}.$$

In absence of the fundamental risk, i.e. $\delta_d = 0$ and $R^d = 0$, the expected return becomes zero, since dedicated arbitrageurs with sufficient funding supply that can

[11] This is mainly due to the assumption of funding constraint, $f_1^T < \bar{S}_2$, on transient arbitrageurs, which makes sure that mispricing is not fully corrected in period 1. Without this restriction, there exist a $f_1^T > \bar{S}_2$ that ensures $E^T(R) = 0$ and $P_1 = \bar{d}$.

[12] For the case with fully-invested dedicated arbitrageurs, the arbitrage return can be easily written as $E^D(R) = S_1 - f_1^D$, which is independent from the fundamental risk, δ_d.

fully eliminate the mispricing in the initial period and ensure $P_1 = \bar{d}$. In absence of the dedicated arbitrageurs, $R^d = 1$, the asset return becomes S_1. Intuitively, the expected return is higher with larger noise shock, more uncertainty in payoffs (larger δ_d) and more risk aversion among the dedicated arbitrageurs (higher γ^D).

Consider the final case with a mixture of transient and dedicated arbitrageurs, the expected arbitrage return can be written as (See proof in Appendix):

$$E^{D,T}(R) = \left[\left(S_1 + f_1^T - \bar{S}_2\right) R^n + \left(\bar{S}_2 - f_1^T\right)\right] R^d \tag{4.9}$$

In general, the expected arbitrage return is determined by the fundamental and sentiment risk premia, the noise trader sentiment and the funding constraint among arbitrageurs. More importantly, the impacts of fundamental and sentiment risk are channeled through the arbitrageurs' strategies, β_1^T and β_1^D, or in other words, the arbitrage activity, which we will further demonstrate in the next section.

4.4 The Impact of Fundamental and Sentiment Risk

4.4.1 Defining the arbitrage activity

We first define the measurements for the arbitrage activity, as the strategies, β_1^T and β_1^D are unobservable in practice. The recent work of CFS extends the SV model and introduce two measures to capture the arbitrage activity. We follow CFS and define the mispricing correction, K and the noise momentum, Λ as

$$K = \frac{\beta_t^T f_t^T + \beta_t^D f_t^D}{S_1}, \ \Lambda = \frac{\bar{d} - P_2}{\bar{d} - P_1}, \tag{4.10}$$

where K captures the proportion of initial mispricing correction achieved by both group of arbitrageurs in period 1 and Λ captures the persistence of the unarbitraged errors to the next period. In order to better connecting these two empirical measures with our theoretical analyses, we express these two parameters as the expectation with respect to the information $\mathscr{F}_1^T, \mathscr{F}_1^D$ in period 1, such that

$$\kappa = E\left(K \mid \mathscr{F}_1^T, \mathscr{F}_1^D\right) = \frac{\beta_t^T f_t^T + \beta_t^D f_t^D}{S_1}, \ \lambda = E_1\left(\Lambda \mid \mathscr{F}_1^T, \mathscr{F}_1^D\right) = \frac{\bar{d} - E\left(P_2 \mid \mathscr{F}_1^T, \mathscr{F}_1^D\right)}{\bar{d} - P_1}, \tag{4.11}$$

In particular, the noise momentum parameter is able to capture the information whether the transient arbitrageurs long or short the asset in the equilibrium. To provide a better understanding, we note that transient arbitrageurs make the long/short decisions based on the expected price in period 2 as follows (See Appendix for proof):

$$\left\{ \begin{array}{ll} E\left(P_2 \mid \mathscr{F}_1^T\right) < P_1, & \text{for } -1 < \beta_1^T < 0 \\ E\left(P_2 \mid \mathscr{F}_1^T\right) > P_1, & \text{for } 0 < \beta_1^T < 1 \end{array} \right\}. \tag{4.12}$$

It indicates that transient arbitrageurs short the asset only when expected price in period 2 is lower than that in period 1. Eq. (4.12) can be rewritten in terms of λ, such that

$$\left\{ \begin{array}{ll} \lambda > 1, & \text{for } -1 < \beta_1^T < 0 \\ 0 < \lambda < 1, & \text{for } 0 < \beta_1^T < 1 \end{array} \right\} \tag{4.13}$$

However, empirical evidence never observe $\lambda > 1$ on average, according to CFS who focus on international data and the Chapter 2 which bases on different market circumstance in S&P 500 index.[13] Even in the extreme circumstances allocated in Chapter 2, we find that λ is below unity. Our empirical study later applied to the Value and Growth index and future also fails to observe that $\lambda > 1$ in all cases, which indicates that transient arbitrageurs are mostly rational and exploiting the mispricing opportunity, i.e. $\beta_1^T > 0$. As a result, we continue in further discussions under the assumption that

$$\bar{S_2} < f_1^T + S_1 \tag{4.14}$$

to ensure $\beta_1^T > 0$, while the implications of $\beta_1^T < 0$ will be included in the Appendix. Notice that together with the assumption on funding constraint, the restriction on f_1^T can be summed up as

$$\bar{S_2} - S_1 < f_1^T < \bar{S_2}. \tag{4.15}$$

Overall, these two measures allow us to study the impact of fundamental and sentiment risk on the arbitrage activity and generate several testable predictions under fundamental- or sentiment-based view of the value premium anomaly in the following analysis.

[13]In Chapter 2, we conclude that λ tend to be less than unity on average, while our model is able to capture $\lambda > 1$. The difference is resulted from the model setting. In Chapter 2, period 2 noise shock is rather binary, e.g. it becomes either zero with probability q or $S_2 > S_1$ with probability $1-q$. Thus it does not give the arbitrageurs incentive to short sell in period 1, and thus $\lambda < 1$. However, in our model period 2 noise shock is stochastic with mean $\bar{S_2}$. For a large $\bar{S_2}$, it's possible for transient arbitrageurs to adopt short-selling in period 1, which is indicated by $\lambda > 1$.

4.4.2 The impact of sentiment risk

We conclude the impact of sentiment risk in the following proposition (See proof in Appendix).

Proposition 5. *Consider the 3-period model setup from section 4.3.1 and 4.3.2. Under the assumption that* $\bar{S}_2 - S_1 < f_1^T < \bar{S}_2$*, we find that (i)* $\frac{\partial \beta_1^T}{\partial \delta_n} < 0$ *and* $\frac{\partial \beta_1^D}{\partial \delta_n} > 0$*; (ii)* $\frac{\partial \kappa}{\partial \delta_n} < 0$ *and* $\frac{\partial \lambda}{\partial \delta_n} < 0$*.*

Proposition 5 shows the *direct* effect of sentiment risk on transient arbitrageurs and the *indirect* effect on dedicated arbitrageurs. As the higher sentiment risk reduces the demand of transient arbitrageur, it causes a decline in the asset price and the initial mispricing correction, which means that return for dedicated arbitrage is potentially higher. As a result, dedicated arbitrageurs are tempted to augment their initial investment, which on the other hand, improves the initial pricing efficiency and the mispricing correction. Since the former effect is always stronger in aggregate, sentiment risk tends to deter the initial correction achieved by the arbitrageurs, $\frac{\partial \kappa}{\partial \delta_n} < 0$. The impact on noise momentum is also dominated by the transient arbitrageurs. As they reduce their position in the first period, they are saving more funding in order to invest in the next period, which relatively improves the pricing efficiency. Thus sentiment risk reduces future noise persistence, $\frac{\partial \lambda}{\partial \delta_n} < 0$.

CFS show that the overall speed of adjustment is positively associated with the initial mispricing correction κ, but negatively with the subsequent noise momentum λ. According to Proposition 5, the overall impact of sentiment risk is rather uncertain on the overall speed of adjustment, since higher sentiment risk tend to impedes initial mispricing correction κ and also reduces further noise momentum λ.

4.4.3 The impact of fundamental risk

We turn our focus on the impact of fundamental risk, which can be summarized as follows (See proof in Appendix).

Proposition 6. *Consider the 3-period model setup from section 4.3.1 and 4.3.2. Under the assumption that* $\bar{S}_2 - S_1 < f_1^T < \bar{S}_2$*, we find that (i)* $\frac{\partial \beta_1^T}{\partial \delta_d} = 0$ *and* $\frac{\partial \beta_1^D}{\partial \delta_d} < 0$*; (ii)* $\frac{\partial \kappa}{\partial \delta_d} < 0$ *and* $\frac{\partial \lambda}{\partial \delta_d} > 0$*.*

The impact on the mispricing correction is not surprising, since dedicated arbitrageurs will reduce their demand on the asset, which results in a lower initial mispricing correction. The impact on noise momentum is rather positive, comparing to that of sentiment risk. The reason is that: as dedicated arbitrageurs reduce their position in the first period due to higher fundamental risk, their position in period 2 is also reduced proportionally since they do not alter their strategy unless price converges. As a result, the pricing efficiency in period 2 is not improved, but further dampened, which leads to the higher noise momentum. According to Proposition 6, the impact of fundamental risk decelerate the overall speed of adjustment, since higher fundamental risk not only deters the initial mispricing correction, but also intensifies the future noise momentum. For complement, we further illustrate the results by a numerical analysis in the appendix.

4.4.4 Testable predictions

For the purpose of our empirical application on the value premium anomaly, we consider two sections of stocks, value and growth (denoted as V and G), that only vary in the level of fundamental/sentiment risk exposure, and summarize the theoretical predictions of the fundamental- and the sentiment-based view on the value premium anomaly under three possible scenarios, namely the no-anomaly, the value premium and the value discount.

First of all, according to the funding constraint assumptions on transient arbitrageurs , i.e. $\bar{S_2} > f_1^T > \bar{S_2} - S_1$, we obtain the first prediction on the noise momentum parameter.

Prediction 1. The noise momentum satisfies: $0 < \lambda^V, \lambda^G < 1$.

Next, we consider the value premium regime. On one hand, the sentiment-based explanation implies that value stocks are exposed to higher sentiment risk, $\delta_n^V > \delta_n^G$, which makes them more difficult to arbitrage. On the other hand, the fundamental-based explanation implies that value are fundamental riskier, $\delta_d^V > \delta_d^G$, and thus arbitrageurs require higher risk premium for bearing the risk. Thus Proposition 5 and 6 predicts the follows:

Prediction 2 under the value premium regime: (i) Under the sentiment-based assumption that sentiment risk associated with the value stocks is higher than that with the growth stocks, i.e. $\delta_n^V > \delta_n^G$, our model predicts that $\kappa^V < \kappa^G$ and $\lambda^V < \lambda^G$, suggesting that higher sentiment risk in value stocks will induce lower initial mispricing correction and subsequent noise momentum. (ii) Under the fundamental-based assumption that the fundamental risk associated with the value stocks is higher than that with the growth stocks, i.e. $\delta_d^V > \delta_d^G$, our prediction is that $\kappa^V < \kappa^G$ and $\lambda^V > \lambda^G$, suggesting that higher fundamental risk in value stocks will deter initial mispricing correction and raise subsequent noise momentum.

We also consider the scenario of value discount in which case it is easily seen that we obtain the opposite predictions on the arbitrage activity as follows:

Prediction 3 under the value discount regime: (i) Under the sentiment-based assumption that sentiment risk associated with the growth stocks is higher than that with the value stocks, i.e. $\delta_n^V < \delta_n^G$, we obtain the prediction that $\kappa^V > \kappa^G$ and$\lambda^V > \lambda^G$. (ii) Under the fundamental-based assumption that the fundamental risk associated with the growth stocks is higher than that with the value stocks, i.e. $\delta_d^V < \delta_d^G$, our prediction is that $\kappa^V > \kappa^G$, $\lambda^V < \lambda^G$.

For complement the implication on the equivalency regime is obtained:

Prediction 4 under the no-anomaly regime: Under both the sentiment-based and the fundamental-based view, the sentiment and fundamental risks associated with value and growth stocks are the same, i.e. $\delta_n^V = \delta_n^G$ ($\delta_d^V = \delta_d^G$). Our prediction is that both the initial mispricing correction and the subsequent noise persistence parameters are equivalent for value and growth stocks, i.e. $\kappa^V = \kappa^G$, $\lambda^V = \lambda^G$.

Finally we combine the above and yield the last prediction on the overall speed of adjustment:

Prediction 5 on the overall speed of adjustment: (i) Under the value premium regime, the sentiment-based view on the overall speed of adjustments of value and growth stocks is generally uncertain while the fundamental-based view suggests that the overall speed of adjustment of value stocks is slower than that of growth asset. (ii) Under the value discount regime, the sentiment-based view on the overall speed of adjustments of value and growth stocks is still uncertain while the fundamental-based view suggests that the overall speed of adjustment of value stocks is faster than that of growth stocks. (iii) Under the no-anomaly regime, the overall speed of adjustments of both value and growth stocks are predicted to be the same.

4.5 An Empirical Application

To investigate the arbitrage activity among the value and growth stocks, we consider the arbitrage relationship between the value (growth) index spot and future prices. According to the law of one price, the price of future contract should be equal to the fundamental value that accounts for the cost of carry. In practice, however, the price deviations from this no-arbitrage relation are not arbitraged away immediately due to transaction cost, arbitrage risk, liquidity, short sale restriction and funding constraint

(Richie et al., 2008; Bertone, Paeglis and Ravi, 2015). While the transaction costs to implement the arbitrage strategy and the funding constraints of the arbitrageurs tend to be indifferent across value and growth, we give prominence to the arbitrage risk: the fundamental and sentiment risk.

In order to perform the risk-free index arbitrage, one must replicate the portfolio of the index component stocks with appropriate weights, which is costly and challenging to execute. Even after the introduction of Exchange-Traded Fund (ETF),[14] Richie et al. (2008) find that mispricing still exists and highly associated with volatility. Bertone, Paeglis and Ravi (2015) also find that volatility in the index constituents and ETF is associated with price deviation between DJIA index portfolio and the ETF, and the impact of volatility shocks on the deviation will last for several hours. Therefore fundamental risk cannot be fully hedged. Moreover, classical index spot-future arbitrage assumes that the position is held until expiration, which takes at least three months. During the holding period, price deviation may intensify due to investor sentiment and cause net losses, even insolvency in extreme circumstance, to arbitrageurs. A notable example is the collapse of Long-Term Capital Management, which fails to bet against the mispricings in bonds and derivatives as mispricings, instead of narrow down, widen. Sentiment risk thus is imposed to arbitrageurs (De Long et al., 1990; SV). Therefore, arbitrageurs, who attempt to implement the index arbitrage in practice, face both the fundamental risk stems from the index constituents and the sentiment risk arises from investor sentiment.

4.5.1 The data

Our datatset covers the period from 7 April 1999 to 5 December 2014 (a total of 4093 observations) for the S&P 500 value and growth indices and the S&P 500 value and growth future indices, which are collected from DataStream. The S&P 500 value (growth) index is a market-capitalization-weighted index, mainly composing of those stocks within the S&P 500 index and NASDAQ[15] that have a strong value (growth) characteristic. Thus, the indices capture mainly large capitalization value and growth

[14]Index ETF replicates the index portfolio as closely as possible, can be bought and sold daily with low transaction cost and can be traded by smaller investors. Therefore, ETF is a less expensive way to implement index arbitrage, which reduces the potential risk exposure, liquidity problems, short-selling costs and transaction costs.

[15]For example, on January 30 2015, the S&P 500 value index consists of 55 stocks from NASDAQ and 308 stocks from NYSE whereas S&P 500 growth index includes 93 from NASDAQ and 229 from NYSE.

stocks. In particular, value index is constructed by stocks with significant value factors: book to market ratio, earning to market ratio and sales to market ratio, whilst growth index consists those with strong growth factors: three-year earnings (sales) per share growth rate and three-year internal growth rate, and momentum. The indices will be rebalanced once a year in December.[16]

Denote s_t and f_t as the logged spot and future prices. We construct the fundamental, f_t^* using the cost-of-carry formula,

$$f_t^* = s_t + (\pi_t - q_t)\,\tau_t, \tag{4.16}$$

where q_t is the annualized dividend yield (proxy by the S&P 500 dividend yield) and π_t is the risk-free interest rate (proxy by the US three month T-bill rate) and τ_t is the time to maturity. Table 4.1 presents the descriptive statistics. Notice that the differences between logged futures and spot prices are on average -0.008 and 0.141 for value and growth indices, while the differences between the futures price and the fundamental drop to -0.015 and 0.122 after accounting for the cost of carry. Clearly, the value future tend to be undervalued compare to the spot price while the growth future is more bullish on average.

To construct the HML return that captures the value effect, we compute the return associated with the long-short portfolio, i.e. long in the value index (future) and short in the growth index (future). Denote HML_f and HML_s as the daily HML return of the future and spot index,[17] which are constructed as

$$HML_f = \left(f_t^V - f_{t-1}^V\right) - \left(f_t^G - f_{t-1}^G\right) \text{ and } HML_s = \left(s_t^V - s_{t-1}^V\right) - \left(s_t^G - s_{t-1}^G\right)$$

where $f_t^V\left(f_t^G\right)$ is the logged value (growth) index future price and $s_t^V\left(s_t^G\right)$ is the logged value (growth) index spot price. Both HML_f and HML_s are zero-investment portfolios.

Over the full sample period, the HML returns of index and index future are slightly positive at 0.0019 and 0.0019 percent per day (around 0.04% per month), but both returns are not statistically significant with t-statistics, 0.181 and 0.206, respectively. This finding is generally consistent with the extant literature that value

[16]Please refer to the S&P Dow Jones Indices website for more information about index construction.

[17]For complement, we also consider the monthly HML portfolios in future and spot index. Details of the robustness are reported in the appendix.

	Mean	Median	Minimum	Maximum	Std Dev
	Panel A: Value index				
Δf_t	0.012	0	-11.01	10.60	1.25
Δs_t	0.012	0.018	-10.28	10.56	1.30
Δf_t^*	0.012	0.020	-10.27	10.56	1.30
$f_t - s_t$	-0.008	-0.079	-3.08	5.95	0.55
$f_t - f_t^*$	-0.019	-0.053	-3.67	5.52	0.42
	Panel B: Growth index				
Δf_t	0.011	0	-11.53	10.24	1.22
Δs_t	0.010	0.029	-10.02	12.06	1.27
Δf_t^*	0.011	0.033	-10.03	12.08	1.28
$f_t - s_t$	0.141	0.025	-4.77	5.13	0.57
$f_t - f_t^*$	0.122	0.059	-5.67	4.58	0.44
	Panel C: Value minus Growth				
$HMLf$	0.001	0	-6.15	6.89	0.67
$HMLs$	0.001	0	-4.81	4.36	0.59
	Panel D: Market				
q_t	1.82	1.80	1.11	3.60	0.43
π_t	1.96	1.21	0	6.24	2.00

Table 4.1: Basic descriptive statistics

Table 4.1 reports the descriptive statistics. First two panels include the results of value and growth stocks. s_t is the daily log index spot price, Δf_t (Δf_t^*) is the difference of log future (fundamental) price and $f_t - f_t^*$ is the difference between future and fundamental. Panel C captures the return of long value and short growth in both index and index future market, while panel D captures the market statistics. q_t and π_t are the annualized dividend yield and risk-free interest rate. HML_f and HML_s are daily HML return of future and spot index, which are constructed as

$$HML_f = \left(f_t^V - f_{t-1}^V\right) - \left(f_t^G - f_{t-1}^G\right) \text{ and } HML_s = \left(s_t^V - s_{t-1}^V\right) - \left(s_t^G - s_{t-1}^G\right)$$

where f_t^V $\left(f_t^G\right)$ is the logged value (growth) index future price and s_t^V $\left(s_t^G\right)$ is the logged value (growth) index spot price. All numbers are recorded in percentage point terms.

premium tend to vanish in large-cap stocks. Chan, Karceski and Lakonishok (2002) find that value stocks in Russell 1000 index (large cap) does not significantly outperform growth stocks. Loughran (1997) claims that the value premium has been mainly observed for small stocks, though he also documented evidence for a weak value premium among large US firms from 1963 to 1995. Fama and French (2012) report no evidence in favor of the value premium (with mean, 0.10% per month and t statistic, 0.49) in large stocks in the North America region.

4.5.2 Methodology

In order to track the time-variations in the HML returns with possible regimes, we first model the joint distribution of the HML returns in spot and future as a regime-switching process. In particular, we follow Guidolin and Timmermann (2008), and apply the Markov-Switching Vector Autoregression (MS-VAR) model to the vector of HML returns:

$$r_t = \mu_{S_t} + \sum_{j=1}^{q} A_{j,S_t} r_{t-j} + \varepsilon_t,\ \varepsilon_t \sim iid\left(0, \Omega_{S_t}\right). \tag{4.17}$$

Here q is the lag-order, $r_t = (r_{1t}, r_{2t}, \ldots, r_{nt})^{'}$ is an $n \times 1$ vector of returns, S_t are the regime variables, $\mu_{S_t} = (\mu_{1S_t}, \ldots, \mu_{nS_t})$ is an $n \times 1$ vector of mean returns in regime S_t, A_{j,S_t} is an $n \times n$ matrix of the autoregressive terms at lag j in regime S_t and $\varepsilon_t = (\varepsilon_{1t} \ldots \varepsilon_{nt})$ is the vector of return innovations that are assumed to be normally distributed with zero mean and state-specific covariance matrix Ω_{S_t}. The return series r_t comprise future HML and index HML returns, i.e. $HMLf$ and $HMLs$, and thus $n = 2$. The unobserved regime-switching variable, S_t, are assumed to be governed by the $m \times m$ transition probability matrix $\mathbf{P}$ with element p_{ji} defined as

$$p_{ji} = \mathbf{Pr}\left(S_t = i \mid S_{t-1} = j\right),\ i, j = 1, \ldots, m \tag{4.18}$$

p_{ji} is the transition probability from state j to i. Each regime is the realization of a first order Markov chain with constant transition probabilities. The model thus allows the HML returns to vary across regimes, which induces various implications in the cross section of return and systematic risk borne by the investors. For example, knowing that the current state is characterized as value premium, value stocks tend to be exposed to either higher fundamental or sentiment risk, which will deter the arbitrage activity in value stocks.

Second, we assign the specific regime to each day by the largest smoothed state probability among all. In each regime, we aim to capture the implied arbitrage activity across value and growth that exploits the price discrepancy between index spot and future. Specifically, we follow CFS to capture the arbitrage activity: the mispricing correction and the noise momentum, by applying the two period GECM, which is given as:

$$\Delta f_t = \alpha + \kappa \hat{z_{t-1}} + \lambda^* \hat{z_{t-2}} + \delta \Delta f_t^* + \gamma \Delta f_{t-1} + \mu_t,\, \mu_t \sim iid\left(0,\, \sigma_\mu^2\right) \tag{4.19}$$

where Δ is the first difference operator, f_t is the natural log of the future contract price, f_t^* is the natural log of fundamental price implied by the cost of carry model in Eq. (4.16) and $\hat{z}_t$ is the long-run mispricing error. The pricing error $\hat{z}_t$ is estimated by the long-run equation:

$$f_t = \mu + \theta f_t^* + z_t. \tag{4.20}$$

In particular, we can simultaneously capture the arbitrage activity by accommodating both 'the mispricing correction' through κ, and 'the noise momentum' through $\lambda\,(1+\kappa)\,z_{t-2}$, in which λ measures the strength of noise momentum and $(1+\kappa)\,z_{t-2}$ represents the unarbitraged component of the pricing errors from the previous period.[18]

4.5.3 Empirical results

4.5.3.1 The time variation in value effect

In order to avoid mis-specification in the MS-VAR model, the work of Guidolin and Timmermann (2008) offers guidance to determine the number of regimes and lag-orders for the process. We conduct the specification analysis as follows. First, we consider a range of reasonable value for the number of regimes up to six, $m = 1, 2, 3, 4, 6$ and the lag-order up to three, $q = 0, 1, 2, 3$. Second, the Akaike (AIC) and Schwartz (SIC) information criteria are applied to select among the regime specifications. The preferred specification for the MS-VAR model has three-regime

[18]The model in (4.19) also accommodates the dynamics of price overreaction or underreaction with respect to fundamental changes through the contemporaneous reaction coefficient, δ, as well as the short-run momentum effects through the coefficient, γ.

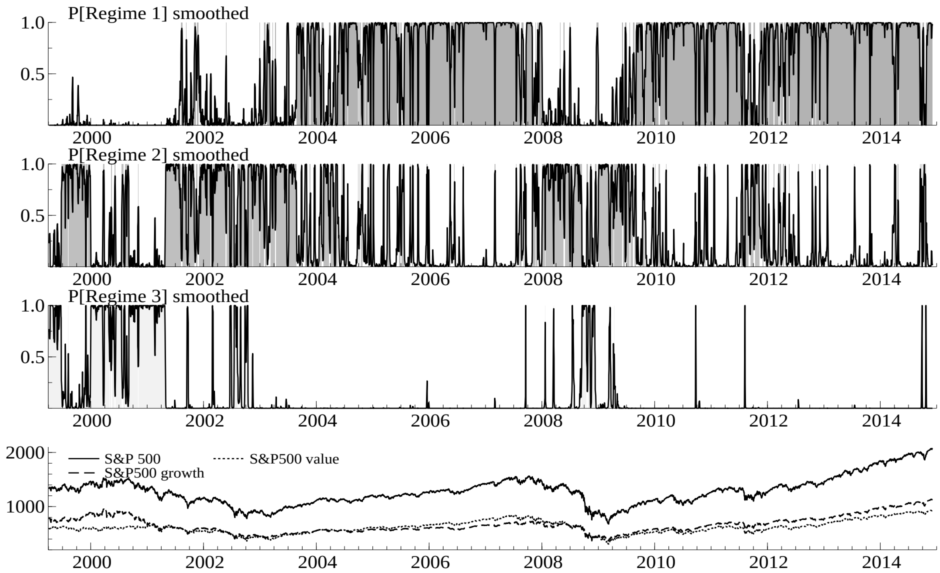

Figure 4.1: The smoothed regime probabilities: the three-regime Markov Switching VAR model for return of $HMLf$ and $HMLs$

The figure plots the smoothed regime probability of regime 1-3 from the regression of three-regime Markov-switching VAR in Eq. (4.17) for returns on series $HMLf$ and $HMLs$, along with the figure of the market performance of S&P 500 at the same time. The sample period is 07/04/1999 - 05/12/2014.

$(m = 3)$[19] but no autoregressive terms $(q = 0)$, and we present the estimation results in Table 4.2 and Figure 4.1. The zero lag-order (no autoregressive terms) is a result of lack of serial correlation in the daily return series.

Regime 1 represents the period with no-anomaly as HML_f and HML_s returns are indifferent from zero. This regime has a duration of over 22 days and consists half of the observations. Figure 4.1 shows that regime 1 mainly captures two persistent bull markets, 2004-2008 and 2010-2014. The market return in regime 1 is the highest (0.060) while the volatility is the lowest compare to other regimes. Overall regime 1 is characterized as the highly persistent, low-volatility bull state with no value premium anomaly.

Regime 2 is a moderately persistent state with significant value discount, where both HML_f and HML_s are significantly negative (-0.037 and -0.034 respectively). It contains 34% of the observation and has a duration of 11 days. The market return

[19]Both criteria suggest the preferred specification is four regimes, $m = 4$. However, the results capture two regimes with no anomaly, i.e. the bull market in 2004-2008 and 2010-2014. In both regimes without anomaly, the HML returns and volatilities are rather indifferent. Note that there is a trade off between over-parameterization and the economic interpretation. Since Guidolin and Timmermann (2008) suggest that AIC tends to suffer from over-parameterization, we select the second best, i.e. three regime, $m = 3$, with a clear economic intuition behind the results.

	Regime 1		Regime 2		Regime 3	
	Panel A: The three-regime model					
	1. the HML returns					
	Mean	T-value	Mean	T-value	Mean	T-value
HML_f	-0.000	-0.06	-0.037**	-2.11	0.122*	1.70
HML_s	-0.001	-0.23	-0.034**	-2.04	0.120**	1.99
obs	2216		1389		487	
Dur	21.48		10.29		13.81	
	2. the variance matrices					
	HML_f	HML_s	HML_f	HML_s	HML_f	HML_s
HML_f	0.058		0.362		2.456	
HML_s	0.050	0.057	0.313	0.337	1.597	1.698
	3. linkage to market conditions					
r_{sp}	0.060		-0.053		-0.025	
Stdev	0.820		1.364		2.196	
TED	0.346		0.558		0.865	
VIX	17.03		24.76		28.91	
	Panel B: Transition Probability					
Regime 1	0.9435		0.0861		0.0036	
Regime 2	0.0553		0.8862		0.0799	
Regime 3	0.0011		0.0275		0.9164	
Ergodic	0.5341		0.3450		0.1210	

Table 4.2: Estimation of the three-regime Markov-Switching VAR with no autoregressive term

This table reports the estimation of the vector Markov switching model. The sample period is 01/04/1999 - 05/12/2014, a total 4093 observations. First table in Panel A reports the estimated results of three-regime switching VAR model in (4.17) with no autoregression terms for returns of HML future and HML index returns:

$$r_t = \mu_{S_t} + \varepsilon_t,\ \varepsilon_t \sim iid\,(0, \Omega_{S_t})$$

where the return series $r_t = (r_{1t}, r_{2t})^{'}$ is an 2×1 vector of returns, $r_{1t} = HML_{f,t}$ and $r_{2t} = HML_{s,t}$ are the daily return of long value index (future) and short growth index (future), S_t are the regime variables, $\mu_{S_t} = (\mu_{1S_t}, \mu_{2S_t})$ is an 2×1 vector of mean returns in regime S_t and $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})$ is the vector of return innovations that are assumed to be normally distributed with zero mean and state-specific covariance matrix Ω_{S_t}. Obs and Dur report the number of observation and duration in each regime. The second table in Panel A reports the diagonals of the variance matrices, while the third table shows the average level of market return in excess of risk-free rate, return volatility, TED spread and VIX index in each regimes. Panel B reports the transition and ergodic probabilites. ***, ** and * indicate significance at 1%, 5% and 10% levels, respectively.

in regime 2 (-0.053) is the lowest, but has relatively high volatility. Together with Figure 4.1, regime 2 is regarded as the bear state, capturing the market downturn over 2001-2003 and the global financial crisis over 2007-2009, and the recovering state, i.e. the transition period between bull and bear markets, such as the year of 2004 and 2010. Thus regime 2 represents a moderately volatile bear market with significant value discount.

Finally regime 3 is a moderately persistent state, which emerges significantly high value premium. The HML returns, HML_f and HML_s, are significantly positive (0.122 and 0.120 respectively) with the highest volatility. Although it consist only 10% of the observations, it has a longer duration (14 days) than regime 2, as it is more sticky with a higher transition probability (91.6%). Regime 3 mainly captures the market crashes in the sample period, such as the collapse of internet bubble in 2000, the financial crisis in 2008, the Russian crisis in the end of 2014. According to Panel B in table 4.2, the transition probability between regime 1 and 3 are rather low and close to zero, which implies that the market cannot switch from bull states to market crash in direct. Overall the result illustrates a highly volatile crash state with value premium in regime 3.

Some of our results in Table 4.2 are contradicted to GT, who study the monthly HML return, controlled for size effect, from 1927 to 2001. GT document a no-anomaly (value premium) state in the persistent bear (bull) market, while our study reports value discount (no-anomaly). It seems that results emerging from our data tend to document relatively smaller value effect than GT's, and the results are robust after we apply the monthly return of HML portfolio (See appendix for the robustness check). The outcomes are likely to result from the size effect, as our data only covers the large-cap HML portfolios, while GT have controlled for the size effect by including both small- and large-cap stocks. Value premium tend to be significantly higher in small-cap stocks, as Fama and French (2012) report a significant 0.56% monthly small-cap HML return and an insignificant 0.10% monthly large-cap HML return. Therefore, HML returns emerged from our data are smaller than those in GT. What's worth noticing is that both GT and our paper capture the large value premium state in the extreme circumstance, such as the Great Depression in the 30s, the burst of dot-com bubble in 1999-2001 and the global financial crisis in 2008.

Although GT does not further investigate the reason of such time variation in value effect, it is natural to ask whether the results in table 4.2 accords with the fundamental-based or the sentiment-based view. On one hand, LSV argue that under the fundamental-based view, value stocks tend to underperform during bad times where marginal utility of wealth is high. However, Zhang (2005) find that due to the counter-cyclical price of risk, value stocks tend to outperform relative to growth stocks during the bear state when the expected premium for risk is high. Petkova and Zhang (2005) measure the economic condition by the expected market risk premium, and find that value tend to have relatively higher (lower) beta than growth during the period of higher (lower) expected risk premium, which verifies the fundamental-based view. The results in table 4.2 partly verify the argument, since HML returns tend to be lower in the bear market than that in the bull market. But the fundamental-based view cannot explain why value stocks largely outperform during the market crashes. The sentiment-based view, on the other hand, suggests that the time variation in limits to arbitrage may be responsible for results. As document in table 4.2, the bull market is characterized with low volatility and ample liquidity, proxy by the VIX index and the TED spread,[20] arbitrage is thus more effective. Any irrational biased can be eliminated immediately, and thus no anomaly appears. As arbitrage becomes less effective during the bear and crash state (relatively high TED spread and VIX index), anomalies occur and are statistically significant. However, there is lack of theoretical argument showing why irrational investors tend to in favor of value (growth) stock during the bear (crash) state.

4.5.3.2 Value effect and the arbitrage activity

We next study the implied arbitrage activity across the value and growth stocks in the three regimes. Panel A and C in Table 4.3 report the result of the two-period GECM of value and growth index respectively, while Table 4.4 report the results of coefficient differences between value and growth. In general, the results reveal the following findings across three regimes. First, the regime 1 with no anomaly has the lowest volatility σ among all in both value and growth index, while the regime with value premium has the highest. Second, the absolute value of pricing errors $\mid z_{t-1} \mid$ are lowest in the regime with no anomaly, and slightly higher in the value discount regime. Pricing error in the value premium regime are almost twice of that in Regime 1 and 2. Third, the estimated mispricing correction κ are negative and the noise

[20]TED spread , e.g. the spread between the three-month risky LIBOR rate and the three-month risk free T-Bill yield, measures the cost of funding for arbitrageurs. VIX index of implied volatility in S&P 500 index options reflects the market forecast of the aggregate financial market volatility, i.e. higher VIX means traders are expecting that the market is more likely to fluctuate sharply in the near future.

momentum λ are positive, and both are less than unity, which is consistent with the predictions of CFS and Chapter 2. More importantly, $\lambda < 1$ indicates that transient arbitrageurs are on average exploiting the mispricing opportunity, i.e. $0 < \beta_1^T < 1$, which confirms our **Prediction 1.** Fourth, as the volatility and the initial pricing errors $\mid z_{t-1} \mid$ increase from regime 1 to 3, the mispricing correction κ rises. This is consistent with the risky and costly arbitrage hypothesis, which predicts that the arbitrage force tend to increase with the degree of mispricing (Gallagher and Taylor, 2001).

Our primary focus is on the comparison of the implied arbitrage activity (in absolute level) between the value and growth index. In regime 1, the no-anomaly regime, value index has a slightly greater κ than growth index (0.386 and 0.366 respectively) with a difference of 2%, but the difference is not significant (we cannot reject the difference is equal to zero, as shown in table 4.4). Similarly, the noise momentum in both value and growth are indifferent (0.295 and 0.291 respectively). Consistent with our theoretical prediction, both fundamental- and sentiment-based explanations predict that value and growth tend to display similar κ and λ when they have the same returns (**Prediction 4.**). It indicates that arbitrageurs who trade value and growth face indifferent cash-flow fundamentals and sentiment risk as reflected by their activities.

Comparing κ and λ across value and growth in regime 2 where the value discount arises, we find that they are both higher in value than in growth. In particular, the mispricing correction κ is 56.9% in value but only 53.0% in growth, while the noise momentum λ is 58.0% in value but only 49.3% in growth. The differences in κ and λ between value and growth are 3.8% and 8.7%, respectively, which are significantly larger than those in regime 1 and are statistically significant. This finding strongly coincides with the prediction under sentiment-based explanation (**Prediction 3.** (i)) that due to higher sentiment risk in growth index, arbitrageurs, who attempt to exploit the price discrepancy between growth index and future, are rather limited in the initial correction, a lower κ, but are able to reduce the subsequent noise momentum, a lower λ.

Regime 3 is the state with the significant value premium, where value tend to experience a lower mispricing correction and noise momentum. The differences between mispricing correction and noise momentum across value and growth are the largest among the three regimes, 3.9% and 33.3% respectively, and are both significant statistically. The implied arbitrage activity also accords with the sentiment-based view in **Prediction 2.** (i), which implies the arbitrage activity is mainly ex-

	Regime 1			Regime 2			Regime 3		
	Mean		T-stat	Mean		T-stat	Mean		T-stat
	Panel A: Value index								
α	-0.005	***	-2.69	0.002		0.84	0.014	***	3.55
κ	-0.386	***	-28.4	-0.569	***	-40.0	-0.638	***	-42.4
λ^*	0.181	***	17.2	0.250	***	18.0	0.143	***	9.10
δ	0.893	***	286.	0.937	***	317.	0.910	***	177.
γ	0.038	***	10.6	0.011	***	3.69	0.012	**	2.36
σ	0.128			0.158			0.253		
	Panel B: Recoverd Coefficients								
λ	0.295	***	11.1	0.580	***	12.8	0.397	***	7.26
w	-0.106	***	-33.1	-0.062	***	-21.2	-0.089	***	-17.4
π	0.365	***	10.3	0.180	***	3.59	0.144	**	2.39
SOA	0.205	***	19.7	0.318	***	23.1	0.494	***	21.1
$\mid z_{t-1} \mid$	0.207			0.243			0.592		
	Panel C: Growth index								
α	-0.006	***	-3.11	0.004	*	1.65	0.012	***	3.03
κ	-0.366	***	-27.9	-0.530	***	-37.1	-0.677	***	-46.6
λ^*	0.184	***	14.5	0.231	***	16.6	0.235	***	15.5
δ	0.890	***	278.	0.919	***	287.	0.891	***	171.
γ	0.044	***	12.3	0.010	***	3.03	0.030	***	5.53
σ	0.122			0.158			0.271		
	Panel D: Recoverd Coefficients								
λ	0.291	***	17.9	0.493	***	19.9	0.730	***	16.5
w	-0.109	***	-34.2	-0.080	***	-25.1	-0.108	***	-20.7
π	0.402	***	11.5	0.127	***	2.98	0.286	***	5.26
SOA	0.181	***	18.9	0.298	***	22.4	0.442	***	27.9
$\mid z_{t-1} \mid$	0.219			0.255			0.615		

Table 4.3: Estimation of the two-period Generalized Error Correction Model on value/growth Index and index future across three regimes

This table reports the two-period GECM for value and growth index in the three regimes that identified in table 4.2. Regimes in each day are identified by the largest smoothed probability among the three regimes. Panel A and C reports the results of the two-period GECM in (4.19) in each regime across value and growth index, such that

$$\Delta f_t = \alpha + \kappa \hat{z_{t-1}} + \lambda^* \hat{z_{t-2}} + \delta \Delta f_t^* + \gamma \Delta f_{t-1} + \mu_t$$

where κ measure the initial mispricing correction, $\lambda = \lambda^*/(1+\kappa)$ captures noise momentum, f_t and f_t^* are spot and fundamental price for the future contract, $\hat{z}_t$ is the mispricing error estimated from equation (4.20), μ_t is the zero-mean idiosyncratic error term with zero mean and finite variance, σ_u^2. Panel B and D reports the recovered coefficient in the model, where SOA represents the overall speed of adjustment, $\kappa + \lambda\,(1+\kappa)$; w is recovered by $\delta - 1$ and $\pi = -\gamma/w$; $\mid z_{t-1} \mid$ is the absolute value of $\hat{z}_t$, constructed as the measure of mispricing error. Notice that To obtain the variance of the recovered coefficients, a delta method is applied. ***, **, * indicate significance at 1%, 5% and 10% levels, respectively.

	Regime 1			Regime 2			Regime 3		
	Mean		T-stat	Mean		T-stat	Mean		T-stat
				Value minus Growth					
α	0.000		0.20	-0.002		-0.57	-0.011	*	-1.97
κ	-0.020		-1.05	-0.038	*	-1.92	0.039	*	1.88
λ^*	-0.003		-0.19	0.018		0.95	-0.091	***	-4.19
λ	0.003		0.15	0.087	**	2.37	-0.333	***	-4.73
δ	0.003		0.68	0.017	***	4.10	0.018	***	2.54
γ	-0.005		-1.02	0.001		0.23	-0.018	**	-2.31
π	-0.037		-0.74	0.053		0.80	-0.142	*	-1.75
SOA	0.023	*	1.67	0.019		1.04	0.052	*	1.85

Table 4.4: The differences in coefficients between value and growth across three regimes

This table reports the difference in coefficients between value and growth in three regimes identified in table 4.3. A delta method is applied to obtain the variance of the differences. ***, **, * indicate significance at 1%, 5% and 10% levels, respectively.

plained by the higher sentiment risk exposure in value stocks.

The overall speed of adjustment, characterizing by $\kappa + \lambda\,(1+\kappa)$ are slightly faster in value than in growth among all three regimes: 20.5% and 18.1% in Regime 1; 31.8% and 29.8% in Regime 2; 49.4% and 44.2% in Regime 3, respectively. However, the differences are small, 2.3%, 1.9% and 5.2%, and insignificant. The result is consistent with the sentiment-based explanation in **Prediction 5.**, since sentiment risk induces two competing effect on speed of adjustment and the aggregate effect might be canceled out.

To this end, the implied arbitrage activity reported in our paper accords with the sentiment-based view in all three regimes, and suggest that the higher exposure to sentiment risk in large-cap value (growth) stocks deter the arbitrage activity toward mispricing, and earns a higher expected return as known as the value (growth) premium.

4.6 Conclusion

This paper studies the distinctive impact of fundamental and sentiment risk on the arbitrage activity, and distinguishes the two explanations behind the value premium anomaly: the fundamental-based view, which attributes it to cash-flow fundamental risk, and the sentiment-based view, which suggests that stocks are priced by investors sentiment. We first design an integrated model in which heterogeneous arbitrageurs exploit the price discrepancies at multiple trading periods subject to both fundamental and sentiment risk. The model highlights the distinctive arbitrage activity in response to fundamental and sentiment risk. We refer to the work of CFS for two measures of the arbitrage activity, i.e. initial mispricing correction and subsequent noise momentum. The model predicts that: 1, higher fundamental risk deters initial mispricing correction and enlarges subsequent noise momentum; 2, higher sentiment risk also deters initial mispricing correction but reduces subsequent noise momentum.

For the purpose of our empirical work, we apply a two-stage methodology on the S&P 500 value and growth index and index future. First, we identify three regimes with distinctive HML returns, using a Markov-Switching VAR model. Value premium tends to occur during the extreme period of market crashes, while value discount is likely to appear in the persistent bear market. The persistent bull market, during 2004-2008 and 2010-2014, captures most of the regime with no anomaly. In the second stage, we apply the two-period generalized ECM to each regime and capture the arbitrage activity that conducts the value and growth index-arbitrage. The results show that the arbitrage activity in value tend to display a lower mispricing correction and noise momentum when it outperforms, which accords with predictions under higher sentiment risk.

Our theoretical and empirical results produce the following implications. First, our results indicate that sentiment risk is the main determinant of large-cap value premium anomaly, such that the supreme return in value stocks are attributed to the relatively higher exposure to sentiment risk. However, like Lakonishok et al. (1994) and Daniel et al. (2001a), our theory and empirical results do not dispute the possibility that both fundamental- and sentiment-based theories will jointly explain the value premium anomaly. Fundamental risk can explain the total asset return, and might be able to explain some of the value anomaly as well. Second, we highlight the analysis via the arbitrage activity. Arbitrageurs strategically adjust their investment in response to different types and level of arbitrage risk, which ensures that an arbitrage opportunity with higher risk exposure earns higher returns. The implied arbitrage activity can help understand and identify the level of fundamental and sentiment risk exposure. Future studies can be extended to test several other pricing anomalies, such as the small-cap and momentum effect, and other cross-sectional mispricings. It would also be interesting to extend our approach to the model of Daniel et al. (2001a), which include more discussion about asset pricing with the traditional CAPM beta and the price-related misvaluation. Third, our paper strongly supports the work of CFS for the importance of the noise momentum in the market. Although the initial impacts of fundamental and sentiment risk (impact on κ) are similar, but their impacts on the subsequent noise momentum show significant differences. It reveals important information of how arbitrageurs' effort on exploiting mispricing can affect the subsequent price convergence.

Appendix 4.A The Role of Initial Mispricing

We first provide the determinants of asset price for complementary (See Proofs in the appendix).

Proposition 7. *Consider the 3-period model setup from section 4.3.1 and 4.3.2. Under the assumption that $\bar{S_2} - S_1 < f_1^T < \bar{S_2}$, we find that $\frac{\partial P_1}{\partial S_1} < 0$, $\frac{\partial P_1}{\partial f_1^T} > 0$, $\frac{\partial P_1}{\partial f_1^L} \geq 0$,* $\frac{\partial P_1}{\partial \delta_d} < 0$ and $\frac{\partial P_1}{\partial \delta_n} < 0$.

Proposition 7 highlights a number of properties on the equilibrium price. First, larger noise trader demand shocks have larger price impact, which is consistent with the original SV model. Second, given a noise trader shock S_1, it is straightforward that the equilibrium price is more efficient when arbitrageurs receive more funding (f_1^T and f_1^L). Third, the price impact is also larger when arbitrageurs have higher risk tolerance (r^j) and when the asset has higher fundamental (δ_d) or sentiment (δ_n) risk because arbitrageurs require more compensations to bear the risks.

Notice that the initial noise trader shock, determine the initial price displacement, is one of the important information for transient and dedicated arbitrageurs. CFS and Chapter 2 extend the SV model and derive several implications on the arbitrage activity in varying mispricings. We follow these two works and reveal a similar role of initial mispricing in our extended model, which is concluded more formally in the following proposition (See Proofs in the appendix).

Proposition 8. *Consider the 3-period model setup from section 4.3.1 and 4.3.2. Under the assumption that $\bar{S_2} > f_1^T > \bar{S_2} - S_1$, we find that (i) $\frac{\partial \beta_1^T}{\partial S_1} > 0$, $\frac{\partial \beta_1^D}{\partial S_1} > 0$; (ii) $\frac{\partial \kappa}{\partial S_1} > 0$ and $\frac{\partial \lambda}{\partial S_1} < 0$.*

Intuitively, both groups of arbitrageurs will invest more when S_1 rises. Because of the independent distributed noise trader shocks, a higher S_1 indicates a larger price deviation and thus a larger expected long-term profit when price converges to fundamental, which induces arbitrageurs to invest more. Together with Proposition 7, we find that although arbitrageurs react positively to the noise trader shocks, but fail to improve the pricing efficiency. This indicates the limited ability of arbitrageurs to bear against a non-fundamental demand shock.

We highlight that the equilibrium holding of dedicated arbitrageurs is influenced not only by the initial noise S_1 directly, but also by the indirect effect of transient arbitrageurs $\beta_1^T f_1^T$. On one hand, dedicated arbitrageurs will enhance their investment with S_1 as imposed in Eq. (4.4). On the other hand, higher S_1 indicates a higher investment of transient arbitrageurs, which reduces the price deviation and the willingness of dedicated arbitrageurs to invest. These two effects are canceling each other, but the former is always higher. Thus in aggregate, dedicated arbitrageurs are still positively reacting to initial noise. Compare to the sensitivity of the transient arbitrageurs, we always find that

$$\frac{\partial \beta_1^T}{\partial S_1} > \frac{\partial \beta_1^D}{\partial S_1}. \tag{4.21}$$

It indicates that transient arbitrageurs are more sensitive to changes in current market circumstances.

Appendix 4.B A Numerical Example

In this section we provide a numerical example to illustrate the main propositions in this chapter. We let the expected payoff of the asset be, $\bar{d} = 1$, the risk tolerance of transient and dedicated arbitrageurs be, $\gamma^T = 4$ and $\gamma^D = 4$, the arbitrage capital raised by transient and dedicated arbitrageurs in period 1 be, $f_1^T = 0.2$ and $f_1^D = 0.2$, the initial noise trader shock be, $S_1 = 0.3$, and the expected noise trader shock in period 2 be, $\bar{S}_2 = 0.4$. We allow the fundamental risk δ_d varies from 0.35 to 0.8, while the sentiment risk δ_n ranges from 0.2 to 0.6. The top two panels in Figure 4.2 describe the impacts of sentiment risk, δ_n, on the strategies of the arbitrageurs and the arbitrage activity captured by the mispricing correction and the noise momentum. It is easily seen that higher sentiment risk reduce the holding of transient arbitrageurs, while dedicated arbitrageurs who aim at the long-term return tend to step in. The overall impact results in a lower mispricing correction and a declining noise persistence in the future. The bottom two panels in Figure 4.2 display, on the other hand, the impacts of fundamental risk, δ_d, on the strategies and the activities of the arbitrageurs. In contrast, higher fundamental risk only brings down the position of dedicated arbitrageurs, but independent from that of transient arbitrageurs. The combined effect shows that initial mispricing correction tend to drop while noise momentum is increasing with fundamental risk.

Appendix 4.C Robustness Check

We turn our focus on the monthly HML return for a robustness check, since most HML portfolios are created to hold for a month, a year, and even 5 years. Denote HML_{fm} and HML_{sm} as the monthly HML return of future and spot index,[21] which are constructed as

$$HML_{fm} = \left(f_t^V - f_{t-22}^V\right) - \left(f_t^G - f_{t-22}^G\right) \text{ and } HML_{sm} = \left(s_t^V - s_{t-22}^V\right) - \left(s_t^G - s_{t-22}^G\right)$$

where $f_t^V \left(f_t^G\right)$ is the logged value (growth) index future price and $s_t^V \left(s_t^G\right)$ is the logged value (growth) index spot price, 22 days is the approximated trading days in a month.

Table 4.5 reports the three-regime Markov-Switching VAR for the joint distribution of monthly return series, HML_{fm} and HML_{sm}, Figure 4.3 plots the regime probabilities for the MS-VAR and Table 4.6 reports the implied arbitrage activity in each regime across value and growth index. In particular, regime 1 is the no-anomaly regime with insignificant HML returns. The arbitrage activity in this regime across value and growth are indifferent. Regime 2 is characterized as the bear market state with significant value discount. During this period of time, the arbitrage activity in growth stocks tend to be affected by higher exposure to sentiment risk, since the mispricing correction is 44% (51%) and the noise momentum is 36% (52%) in growth (value). Regime 3 is the crash state, which has the significant value premium, 3.1% per month on average. The arbitrage activity cross value and growth is again explained by the sentiment-based predictions, such that both mispricing correction and noise momentum are much lower in value stocks. The results are mostly consistent with our early results that generated from the daily return series, which provides further support to the sentiment-based explanation of the value premium anomaly.

[21] We also consider the portfolio holding for a month and generate a month HML returns in future and spot index. Detail of the robustness check are reported in the appendix.

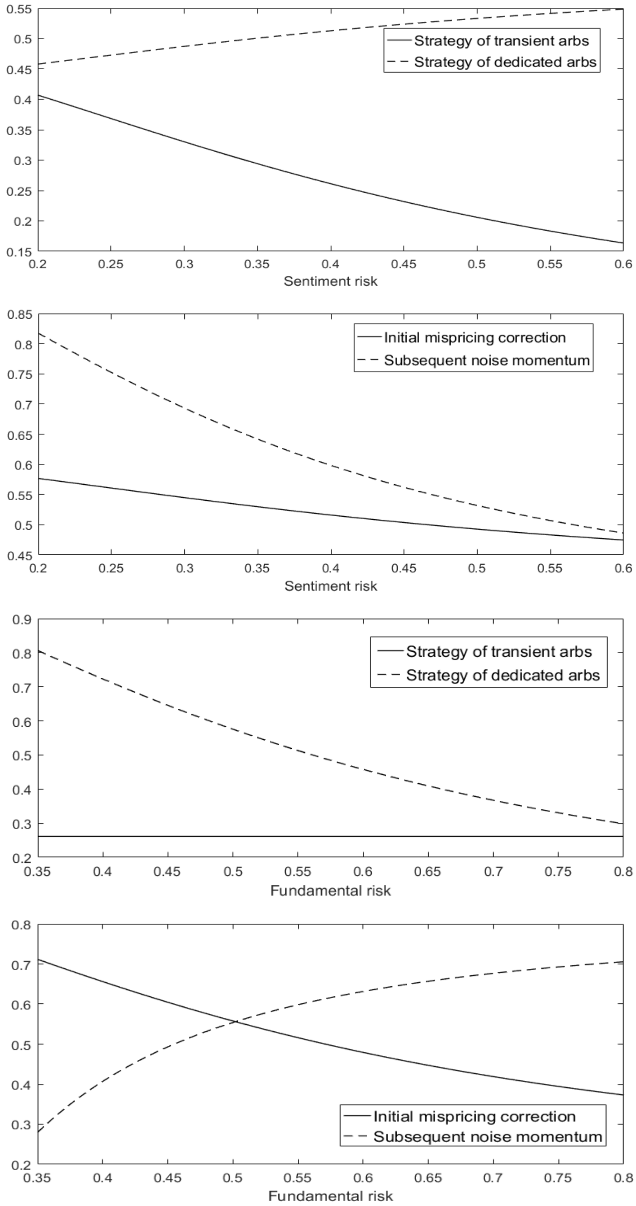

Figure 4.2: Arbitrageurs strategies, mispricing correction and noise momentum with respect to sentiment and fundamental risk

The figure plots the impact of sentiment and fundamental risk on the strategies of transient and dedicated arbitrageurs, and the arbitrage activity captured by initial mispricing correction and subsequent noise momentum. We consider an asset with expected payoff $\bar{d} = 1$; transient and dedicated arbitrageurs with risk tolerance, $\gamma^T = 4$ and $\gamma^D = 4$, and available arbitrage capital in period 1, $f_1^T = 0.2$ and $f_1^D = 0.2$; noise traders with initial shock, $S_1 = 0.3$, and the expected shock in period 2, $\bar{S_2} = 0.4$. We allow fundamental risk δ_d varies from 0.35 to 0.8, while sentiment risk δ_n ranges from 0.2 to 0.6. The top two panels plot the impact of sentiment risk while the bottom two display that of fundamental risk.

	Regime 1		Regime 2		Regime 3	
	Panel A: The three-regime model					
	1. the HML returns					
	Mean	T-value	Mean	T-value	Mean	T-value
HML_{fm}	0.015	0.03	-3.169***	-36.8	3.169***	25.1
HML_{sm}	0.018	0.56	-3.290***	-37.3	3.132***	26.5
obs	2273		861		936	
	2. the variance matrices					
	HML_{fm}	HML_{sm}	HML_{fm}	HML_{sm}	HML_{fm}	HML_{sm}
HML_{fm}	0.812		4.202		7.818	
HML_{sm}	0.770	0.774	4.032	4.046	7.260	7.295
	Panel B: Transition Probability					
Regime 1	0.9652		0.0432		0.0450	
Regime 2	0.0167		0.9479		0.0078	
Regime 3	0.0180		0.0088		0.9470	
Ergodic	0.5585		0.2115		0.2300	

Table 4.5: Estimation of the three-regime Markov-Switching VAR with no autoregressive term (monthly)

This table reports the estimation of the Markov switching vector autoregressive model. The sample period is 03/05/1999 - 05/12/2014. First table in Panel A reports the estimated results of three-regime switching VAR model in (4.17) with no autoregression terms for returns of HML future and HML index returns:

$$r_t = \mu_{S_t} + \varepsilon_t,\ \varepsilon_t \sim iid\,(0, \Omega_{S_t})$$

where the return series $r_t = (r_{1t}, r_{2t})'$ is an 2×1 vector of returns, $r_{1t} = HML_{fm,t}$ and $r_{2t} = HML_{sm,t}$ are the monthly return of long value index (future) and short growth index (future), S_t are the regime variables, $\mu_{S_t} = (\mu_{1S_t}, \mu_{2S_t})$ is an 2×1 vector of mean returns in regime S_t and $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})$ is the vector of return innovations that are assumed to be normally distributed with zero mean and state-specific covariance matrix Ω_{S_t}. The second table in Panel A reports the diagonals of the variance matrices. Panel B reports the transition and ergodic probabilites. ***, ** and * indicate significance at 1%, 5% and 10% levels, respectively.

	Regime 1			Regime 2			Regime 3		
	Mean		T-stat	Mean		T-stat	Mean		T-stat
	Panel A: Value index								
α	-0.005	**	-1.90	0.000		0.17	0.014	***	3.55
κ	-0.475	***	-34.4	-0.505	***	-34.2	-0.664	***	-45.0
λ^*	0.232	***	17.2	0.256	***	17.5	0.103	***	6.82
λ	0.442	***	12.7	0.519	***	12.4	0.307	***	5.99
δ	0.907	***	264.	0.954	***	344.	0.881	***	166.
γ	0.021	***	5.85	0.021	***	7.30	0.006		1.20
σ	0.149			0.148			0.243		
	Panel B: Recoverd Coefficients								
SOA	0.243	***	21.5	0.248	***	20.9	0.561	***	32.8
$\mid z_{t-1} \mid$	0.220			0.305			0.340		
	Panel C: Growth index								
α	-0.004	**	-1.85	0.002		1.00	0.008	***	2.14
κ	-0.434	***	-31.2	-0.440	***	-30.1	-0.717	***	-49.9
λ^*	0.222	***	16.2	0.202	***	14.0	0.210	***	14.2
λ	0.393	***	12.4	0.360	***	10.9	0.746	***	10.0
δ	0.902	***	259.	0.927	***	306.	0.869	***	161.
γ	0.029	***	7.90	0.016	***	4.89	0.026	***	4.70
σ	0.146			0.148			0.261		
	Panel D: Recoverd Coefficients								
SOA	0.211	***	20.0	0.238	***	20.8	0.507	***	30.2
$\mid z_{t-1} \mid$	0.231			0.316			0.364		

Table 4.6: Estimation of the two-period Generalized Error Correction Model on value/growth Index and index future across three regimes (monthly)

This table reports the two-period GECM for value and growth index in the three regimes that identified in table 4.5. Regimes in each day are identified by the largest smoothed probability among the three regimes computed from the month return series. Panel A and C reports the results of the two-period GECM in (4.19) in each regime across value and growth index, such that

$$\Delta f_t = \alpha + \kappa \hat{z_{t-1}} + \lambda^* \hat{z_{t-2}} + \delta \Delta f_t^* + \gamma \Delta f_{t-1} + \mu_t$$

where κ measure the initial mispricing correction, $\lambda = \lambda^*/(1+\kappa)$ captures noise momentum, f_t and f_t^* are spot and fundamental price for the future contract, $\hat{z}_t$ is the mispricing error estimated from equation (3.16), μ_t is the zero-mean idiosyncratic error term with zero mean and finite variance, σ_u^2. Panel B and D reports the recovered coefficient in the model, where SOA represents the overall speed of adjustment, $\kappa + \lambda\,(1+\kappa)$; $\mid z_{t-1} \mid$ is the absolute value of $\hat{z}_t$, constructed as the measure of mispricing error. Notice that To obtain the variance of the recovered coefficients, a delta method is applied. ***, **, * indicate significance at 1%, 5% and 10% levels, respectively.

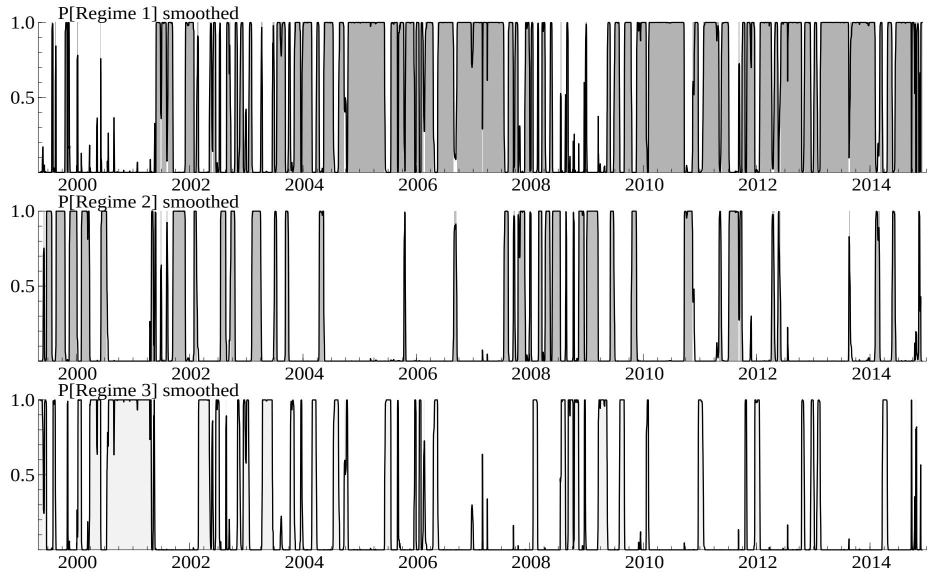

Figure 4.3: The smoothed regime probabilities: the three-regime Markov Switching VAR model for return of $HMLf$ and $HMLs$ (monthly)

The figure plots the smoothed regime probability of regime 1-3 from the regression of three-regime Markov-switching VAR in Eq. (4.17) for monthly returns on series HML_{fm} and HML_{sm}. The sample period is 03/05/1999 - 05/12/2014.

Appendix 4.D Proofs

4.D.1 Optimal position for dedicated arbitrageurs

We known that dedicated arbitrageurs does not alter their holdings in period 2:

$$\beta_2^D f_2^D = \beta_1^D f_1^D \frac{P_2}{P_1}, \tag{4.22}$$

It means that the number of share dedicated arbitrageurs hold is unchanged, i.e. $\frac{\beta_2^D f_2^D}{P_2} = \frac{\beta_1^D f_1^D}{P_1}$. The total funding in period 3 for dedicated arbitrageurs can be expressed as

$$f_3^D = f_1^D + \beta_1^D f_1^D \left(\frac{d_3}{P_1} - 1\right).$$

Since the fundamental d_3 is normally distributed, maximizing the value of $E\left(U\left(f_3^D\right) \mid \mathscr{F}_1^D\right)$ is equivalent to maximizing the mean-variance objective: $E\left(f_3^D \mid \mathscr{F}_1^D\right) - \frac{1}{2}\gamma_a Var\left(f_3^D \mid \mathscr{F}_1^D\right)$. Using the normally distributed asset payoffs, $d_3 \sim N\left(\bar{d}, \delta_d\right)$, then the expectation and variance of funding in period 3 can be expressed as

$$E\left(f_3^D \mid \mathscr{F}_1^D\right) = f_1^D + \beta_1^D f_1^D \left(\frac{\bar{d}}{P_1} - 1\right)$$

$$Var\left(f_3^D \mid \mathscr{F}_1^D\right) = \left(\frac{\beta_1^D f_1^D \delta_d}{P_1}\right)^2$$

Finally, the first order condition in terms of β_1^D leads to the optimal position:

$$\beta_1^D = \frac{P_1\left(\bar{d} - P_1\right)}{f_1^D \gamma^D \delta_d^2}.$$

After substituting $P_1 = \bar{d} - S_1 + \beta_1^T f_1^T + \beta_1^D f_1^D$ into the optimal position, we have

$$\beta_1^D = \frac{2\left(S_1 - \beta_1^T f_1^T\right) - \left(\bar{d} + \gamma^D \delta_d^2\right) + \sqrt{\left(\bar{d} + \gamma^D \delta_d^2\right)^2 - 4\left(S_1 - \beta_1^T f_1^T\right)\gamma^D \delta_d^2}}{2 f_1^D}$$

To provide better understanding of the optimal position, we approximate the square root term in terms of $S_1 - \beta_1^T f_1^T$ (The persistent mispricing error is rather small after the arbitrage trade, thus we treat the second moment $\left(S_1 - \beta_1^T f_1^T\right)^2$ as 0). The square root term can be represented as the sum of the first two terms of its Taylor series:

$$\left(\bar{d} + \gamma^D \delta_d^2\right) - \left(S_1 - \beta_1^T f_1^T\right) \frac{2\gamma^D \delta_d^2}{\bar{d} + \gamma^D \delta_d^2}$$

and the optimal position becomes

$$\beta_1^D \approx \frac{\bar{d}\left(S_1 - \beta_1^T f_1^T\right)}{f_1^D\left(\bar{d} + \gamma^D \delta_d^2\right)}. \tag{4.23}$$

4.D.2 Optimal position for transient arbitrageurs

For transient arbitrageurs, the price in period 2 is expressed as:

$$P_2 = \bar{d} - S_2 + f_2^T + \beta_2^D f_2^D,$$

since transient arbitrageurs expect that they are fully invested in period 2 and they tend to ignore the long run fundamental risk in period 1. Then together with Eq.

(4.22), it can be rewrite as

$$P_2 = \left(\bar{d} - S_2 + f_2^T\right) \frac{P_1}{P_1 - \beta_1^D f_1^D} \tag{4.24}$$

From Eq. (4.2), one can rewrite f_2^T as

$$f_2^T = f_1^T + \beta_1^T f_1^T \left(\frac{P_2}{P_1} - 1\right). \tag{4.25}$$

After combining Eq. (4.24) and (4.25) above we can rearrange and have:

$$P_2 = \frac{\left(\bar{d} - S_2 + f_1^T - \beta_1^T f_1^T\right) P_1}{P_1 - \beta_1^D f_1^D - \beta_1^T f_1^T}, \tag{4.26}$$

Using the normality assumption on S_2, we can write the expected mean and variance of f_2^T conditional on $\mathscr{F}_1^T$ as

$$E\left(f_2^T \mid \mathscr{F}_1^T\right) = f_1^T + \beta_1^T f_1^T \left(\frac{\bar{d} - P_1 - \bar{S_2} + f_1^T + \beta_1^D f_1^D}{P_1 - \beta_1^D f_1^D - \beta_1^T f_1^T}\right)$$

and

$$Var\left(f_2^T \mid \mathscr{F}_1^T\right) = \left(\frac{\beta_1^T f_1^T \delta_n}{P_1 - \beta_1^D f_1^D - \beta_1^T f_1^T}\right)^2$$

With a similar technique, the optimal position in period 1 for transient arbitrageurs is derived after maximize the expected utility $E\left(U\left(f_2^T\right) \mid \mathscr{F}_1^T\right)$ in period 2, which leads to:

$$\bar{\beta_1^T} = \frac{\left(P_1 - \beta_1^T f_1^T\right)\left(P_1 - \beta_1^T f_1^T - \left(\bar{d} - S + f_1^T\right)\right)}{f_1^T \left(P_1 - \beta_1^T f_1^T - \left(\bar{d} - \bar{S_2} + f_1^T\right) - \gamma^T \delta_n^2\right)}.$$

As P_1 is informed by Eq. (4.3), one can rearrange and simplify the optimal position of transient arbitrageurs:

$$\beta_1^T = \frac{\left(\bar{d} - S_1\right)}{\left(\bar{d} - S_1 + \gamma^T \delta_n^2\right)} \frac{\left(f_1^T + S_1 - \bar{S_2}\right)}{f_1^T}. \tag{4.27}$$

4.D.3 The expected arbitrage return

The expected return with partially-invested dedicated arbitrageurs can be expressed with β_1^D satisfying Eq. (4.23) and $f_1^T = 0$:

$$\begin{aligned} E^D(R) &= S_1 - \beta_1^D f_1^D \\ &= S_1 - S_1 \frac{\bar{d}}{\left(\bar{d} + \gamma^D \delta_d^2\right)} \\ &= S_1 \frac{\gamma^D \delta_d^2}{\bar{d} + \gamma^D \delta_d^2}. \end{aligned}$$

The expected return with partially-invested transient arbitrageurs is expressed with β_1^T satisfying Eq. (4.27) and $f_1^D = 0$:

$$\begin{aligned} E^T(R) &= S_1 - \beta_1^T f_1^T \\ &= S_1 - \frac{\left(\bar{d} - S_1\right)}{\left(\bar{d} - S_1 + \gamma^T \delta_n^2\right)} \left(f_1^T + S_1 - \bar{S_2}\right) \\ &= S_1 - \left(1 - \frac{\gamma^T \delta_n^2}{\bar{d} - S_1 + \gamma^T \delta_n^2}\right) \left(f_1^T + S_1 - \bar{S_2}\right) \\ &= \left(\bar{S_2} - f_1^T\right) + \left(S_1 + f_1^T - \bar{S_2}\right) \frac{\gamma^T \delta_n^2}{\bar{d} - S_1 + \gamma^T \delta_n^2} \\ &\approx \left(\bar{S_2} - f_1^T\right) + \left(S_1 + f_1^T - \bar{S_2}\right) \frac{\gamma^T \delta_n^2}{\bar{d} + \gamma^T \delta_n^2}. \end{aligned} \tag{4.28}$$

The expected return with both partially-invested dedicated and transient arbitrageurs can be expressed as:

$$\begin{aligned} E^{D,T}(R) &= S_1 - \beta_1^T f_1^T - \beta_1^D f_1^D \\ &= S_1 - \beta_1^T f_1^T - \frac{\bar{d}\left(S_1 - \beta_1^T f_1^T\right)}{\left(\bar{d} + \alpha \gamma^D \delta_d^2\right)} \\ &= \left(S_1 - \beta_1^T f_1^T\right) \frac{\gamma^D \delta_d^2}{\bar{d} + \gamma^D \delta_d^2} \\ &= \left[\left(\bar{S_2} - f_1^T\right) + \left(S_1 + f_1^T - \bar{S_2}\right) \frac{\gamma^T \delta_n^2}{\bar{d} + \gamma^T \delta_n^2}\right] \frac{\gamma^D \delta_d^2}{\bar{d} + \gamma^D \delta_d^2} \end{aligned} \tag{4.29}$$

The equilibrium price in period 1 can be written as $P_1 = \bar{d} - E(R)$, where $E(R)$ takes the value at Eq. (4.29). Then, results in Proposition 7 can be easily obtained under the assumption that $S_1 + f_1^T - \bar{S_2} > 0$.

4.D.4 The role of initial mispricings

We now provide the proof for Proposition 8, i.e. the role of initial mispricing.

Proof. To see $\frac{\partial \beta_1^T}{\partial S_1} > 0$, we write the partial derivative of β_1^T on S_1 as

$$\begin{aligned}
\frac{\partial \beta_1^T}{\partial S_1} &= \frac{1}{f_1^T} - \frac{\gamma^T \delta_n^2 \left(\bar{d} + \gamma^T \delta_n^2 - \bar{S_2} + f_1^T\right)}{f_1^T \left(\bar{d} - S_1 + \gamma^T \delta_n^2\right)^2} \\
&= \frac{2\gamma^T \delta_n^2 \left(\bar{d} - S_1\right) + \gamma^T \delta_n^2 \left(\bar{d} - \bar{S_2}\right) + f_1^T \gamma^T \delta_n^2 + 2\left(\gamma^T \delta_n^2\right)^2 + \left(\bar{d} - S_1\right)^2}{f_1^T \left(\bar{d} - S_1 + \gamma^T \delta_n^2\right)^2} \\
&> 0.
\end{aligned}$$

The inequality holds since the numerator is positive with $\bar{d} - S_1 > 0$ and $\bar{d} - \bar{S_2} > 0$.

Similarly, to see $\frac{\partial \beta_1^D}{\partial S_1} > 0$, we write the partial derivative of β_1^D on S_1 as

$$\begin{aligned}
\frac{\partial \beta_1^D}{\partial S_1} &= \frac{\bar{d}}{f_1^D \left(\bar{d} + \alpha \gamma^D \delta_d^2\right)} \frac{\gamma^T \delta_n^2 \left(\bar{d} - \bar{S_2} + \gamma^T \delta_n^2 + f_1^T\right)}{\left(\bar{d} - S_1 + \gamma^T \delta_n^2\right)^2} \\
&> 0.
\end{aligned}$$

The inequality holds since $\bar{d} - \bar{S_2} > 0$.

The mispricing correction K can be simplified using Eq. (4.29) with partially invested arbitrageurs ($\beta_t^T, \beta_t^D < 1$), such that:

$$\begin{aligned}
\kappa &= \frac{\beta_t^T f_t^T + \beta_t^D f_t^D}{S_1} \\
&= 1 - \left[\frac{\left(\bar{S_2} - f_1^T\right)}{S_1} + \frac{\left(S_1 + f_1^T - \bar{S_2}\right)}{S_1} R^n\right] R^d \\
&= 1 - \left[R^n + \left(1 - R^n\right) \frac{\left(\bar{S_2} - f_1^T\right)}{S_1}\right] R^d
\end{aligned} \tag{4.30}$$

It can be easily seen that κ increases with S_1.

We rewrite the noise momentum λ as

$$\lambda = \frac{\bar{d} - E\left(P_2 \mid \mathscr{F}_1^T, \mathscr{F}_1^D\right)}{\bar{d} - P_1} \tag{4.31}$$

Using Eq. (4.26) and (4.28), we can simplify the expected price in period 2:

$$E\left(P_2 \mid \mathscr{F}_1^T, \mathscr{F}_1^D\right) = \frac{\left(\bar{d} - \bar{S_2} + f_1^T - \beta_1^T f_1^T\right) P_1}{\bar{d} - S_1}$$
$$= \left(1 + \frac{\left(f_1^T + S_1 - \bar{S_2}\right)}{\bar{d} - S_1} R^n\right) P_1. \tag{4.32}$$

Using Eq. (4.29), we can write

$$P_1 = \bar{d} - \left[\left(\bar{S_2} - f_1^T\right) + \left(S_1 + f_1^T - \bar{S_2}\right) R^n\right] R^d. \tag{4.33}$$

Since $\frac{\left(f_1^T + S_1 - \bar{S_2}\right)}{\bar{d} - S_1}$ is increasing with S_1 whilst P_1 is decreasing with S_1 (Proposition 7), the numerator in Eq. (4.31) will increase slower than the denominator in response to increment in S_1. Thus Λ is decreasing with S_1, i.e. $\frac{\partial \Lambda}{\partial S_1} < 0$. **QED**

4.D.5 The impact of fundamental and sentiment risk

First, we derive the proof for Proposition 5, i.e. the impact of sentiment risk.

Proof. In Proposition 5, $\frac{\partial \beta_1^T}{\partial \delta_n} < 0$ and $\frac{\partial \beta_1^D}{\partial \delta_n} > 0$ can be easily seen from the expression of β_1^T and β_1^D in Eq. (4.27) and (4.23). Consider the partially-invested arbitrageurs only, $\beta_t^T, \beta_t^D < 1$, κ can be written using Eq. (4.30):

$$\kappa = 1 - \left[\frac{\left(\bar{S_2} - f_1^T\right)}{S_1} + \left(\frac{S_1 + f_1^T - \bar{S_2}}{S_1}\right) R^n\right] R^d \tag{4.34}$$

and under the assumption that $S_1 + f_1^T - \bar{S_2} > 0$, we can easily find that $\frac{\partial K}{\partial \delta_n} < 0$.

Using Eq. (4.32), $E\left(P_2 \mid \mathscr{F}_1^T, \mathscr{F}_1^D\right)$ can be further simplified as

$$E\left(P_2 \mid \mathscr{F}_1^T, \mathscr{F}_1^D\right) = \left(1 + G\left(\delta_n\right)\right) P_1$$

where the G function: $G\left(\delta_n\right) = \frac{\left(f_1^T + S_1 - \bar{S_2}\right)}{\bar{d} - S_1} R^n$, is positive and increasing with δ_n. Thus we can write the noise momentum parameter as

$$\lambda = \frac{\bar{d} - \left(1 + G\left(\delta_n\right)\right) P_1}{\bar{d} - P_1}. \tag{4.35}$$

Notice that as δ_n increases, P_1 will drop (Proposition 7). The numerator will increase slower than the denominator in response to increment in δ_n, since

$$\begin{aligned}
\frac{\partial\left(1+G\left(\delta_n\right)\right)P_1}{\partial\delta_n} &= \frac{\partial P_1}{\partial\delta_n} - G\left(\delta_n\right)\left(S_1+f_1^T-\bar{S_2}\right)R^d\frac{\partial R^n}{\partial\delta_n} + P_1\frac{\left(f_1^T+S_1-\bar{S_2}\right)}{\bar{d}-S_1}\frac{\partial R^n}{\partial\delta_n} \\
&= \frac{\partial P_1}{\partial\delta_n} + \frac{\partial R^n}{\partial\delta_n}\frac{\left(f_1^T+S_1-\bar{S_2}\right)}{\bar{d}-S_1}\left(P_1-\left(f_1^T+S_1-\bar{S_2}\right)R^nR^d\right) \\
&> \frac{\partial P_1}{\partial\delta_n} + \frac{\partial R^n}{\partial\delta_n}\frac{\left(f_1^T+S_1-\bar{S_2}\right)}{\bar{d}-S_1}\left(\bar{d}-\left(f_1^T+S_1-\bar{S_2}\right)\right) \qquad (4.36) \\
&> \frac{\partial P_1}{\partial\delta_n}
\end{aligned}$$

Notice that $\frac{\partial P_1}{\partial\delta_n} < 0$. The inequality holds since we have $\bar{d} > f_1^T$, $S_1 < \bar{S_2}$ and $f_1^T + S_1 - \bar{S_2} > 0$. As a result, we have $\frac{\partial\lambda}{\partial\delta_n} < 0$. **QED**

Next, we show the prove of Proposition 6 and 6, i.e. the impact of fundamental risk.

Proof. In Proposition 6, $\frac{\partial\beta_1^T}{\partial\delta_d} = 0$ and $\frac{\partial\beta_1^D}{\partial\delta_d} < 0$ can be easily seen from the expression of β_1^T and β_1^D in Eq. (4.27) and (4.23). Moreover, the result of $\frac{\partial\kappa}{\partial\delta_d} < 0$ can be easily derived under Eq. (4.34). Hence, we focus on the derivation of $\frac{\partial\lambda}{\partial\delta_d}$. Consider Eq. (4.35) again. As δ_d increases, P_1 will drop (Proposition 7). The numerator in Eq. (4.35) will increase faster than the denominator in response to increment in δ_d since $G(\delta_n) > 0$, i.e.

$$\frac{\partial\left(1+G\left(\delta_n\right)\right)P_1}{\partial\delta_d} < \frac{\partial P_1}{\partial\delta_d}.$$

Notice that $\frac{\partial P_1}{\partial\delta_d} < 0$. As a result, we have $\frac{\partial\lambda}{\partial\delta_d} > 0$. **QED**

Finally we make two important remarks. First, together with the proof of Proposition 5 and 6, we can find that for higher sentiment risk, transient arbitrageurs tend to reduce their initial investment, which allow them to allocate more funding in the second period. As a result, the pricing efficiency is relatively improving in period 2, such that the expected price in period 2 $E\left(P_2 \mid \mathscr{F}_1^T, \mathscr{F}_1^D\right) = \left(1+G\left(\delta_n\right)\right)P_1$ is decreasing **slower** than the price in period 1 P_1 with higher sentiment risk, since $G(\delta_n)$ is increasing with δ_n. Therefore, the noise momentum decreases with non-fundamental risk. In contrast, higher fundamental risk will force dedicated arbitrage to reduce their position, which means they are endowed with relatively less funding in period 2. As a result, the pricing efficiency is dampened in period 2, such that the expected price in period 2 $E\left(P_2 \mid \mathscr{F}_1^T, \mathscr{F}_1^D\right) = \left(1+G\left(\delta_n\right)\right)P_1$ is decreasing **faster** than the price in period 1 P_1 with higher fundamental risk, since $1+G(\delta_n)$ is independent of δ_d and is larger than 1. Hence, the noise momentum will grow with larger fundamental risk.

Second, we obtain the above results under the assumption that $S_1 + f_1^T - \bar{S_2} > 0$. In contrast, we will obtain the following results if $S_1 + f_1^T - \bar{S_2} < 0$ holds, such that transient arbitrageurs become noise traders and push price further away from the fundamental. First consider the impact of sentiment and fundamental risk on κ. It is easily seen from Eq. (4.34) that $\frac{\partial \kappa}{\partial \delta_n} > 0$ and $\frac{\partial \kappa}{\partial \delta_d} < 0$ since $S_1 + f_1^T - \bar{S_2} < 0$. Second, the impacts on λ are changed, such that $\frac{\partial \lambda}{\partial \delta_n} > 0$ and $\frac{\partial \lambda}{\partial \delta_d} < 0$, as indicated by Eq. (4.36) and $G(\delta_n) < 0$. This is mainly due to the fact that when transient arbitrageurs are shorting the asset, such that $\beta_1^T < 0$ and $\frac{\partial \beta_1^T}{\partial \delta_n} > 0$. For a higher sentiment risk, there will be less shorting from transient arbitrageurs, which thus improves the initial mispricing correction.

Chapter 5

Concluding Remarks

Behavioral finance suggests a more plausible understanding to the financial market in which mispricing is long-lived and arbitrage is limited, such that rational arbitrageurs fail to drive asset prices to the fundamental values implied by the traditional asset pricing model. In Chapter 2, we propose to distinguish the two types of frictions faced by the arbitrageurs: the arbitrage costs/risks that render them unwilling to undertake arbitrage position, and the funding constraints that make them unable to conduct arbitrage activity even when they are willing to. In particular, we study the process of arbitrage using a model in which arbitrageurs face with both risk and funding constraint, and find that the arbitrage activity towards mispricings are rather nonlinear depending on the dominance of the arbitrage impediments: arbitrage activity tend to increase (decrease) with mispricings when arbitrage cost (funding constraint) establishes dominance. The model thus posits that the overall arbitrage activity displays an inverse U-shape against the size of mispricing errors. To test the theoretical predictions, we provide an empirical application on the S&P 500 index arbitrage by extending the general error-correction model of CFS to the State-dependent Markov-Switching model. We document strong evidence of such nonlinear limits to arbitrage, and capture the periods when the funding constraint is the dominating force as the years of 1987, 1998, 2000 and 2007-08. This chapter adds further evidence on the slow-moving capital literature both theoretically and empirically.

This chapter raises some questions that required further researches in the future. First, the empirical methodology that captures the arbitrage activity is based on the estimation of an asset's fundamental value. This value is often computed as the net present value of the future cash flows at an appropriate discounted rate. While we take advantage of the spot-future relationship to compute the fundamental value using the cost of carry model, yet it is challenging to estimate for an individual asset. Thus the methodology cannot be extended to those assets where their fundamental values are not available or difficult to estimate. Second, the identification of the binding funding constraint is rather ex post, which cannot be detected by investors or regulators in an ex ante point of view. In order to address this issue, we further construct a measure for funding liquidity in Chapter 3 that ex ante identifies the period of the binding funding constraint. Third, this chapter shows that arbitrage frictions can be better understood via their impacts on the arbitrage activity. While we only study the general impact of arbitrage costs and funding constraints so far, more specific analyses can be done: how different types of holding costs, i.e. fundamental and non-fundamental risk, can affect the arbitrage activity; how different types of funding constraints, i.e. equity and leverage constraint, can affect the arbitrage activity. In Chapter 4, we study the former issue to distinguish the impact of fundamental and sentiment (non-fundamental) risk on the arbitrage activity.

In Chapter 3 we study funding liquidity in the time series via the efficacy of arbitrage. In a more general model where the arbitrageurs' funding liquidity is determined by the endogenous leverage constraint set by financiers, we show how funding liquidity affects arbitrage efficacy under two paths to equilibrium, i.e. the loose and binding funding constraint. To capture the essence of the arbitrage efficacy defined in the model, we design an empirical methodology and estimate the implied arbitrage efficacy in the US stock market. Our measure relates to other funding liquidity measures, and more importantly, it identifies the periods of the binding funding constraint and the amplification effect attributed to funding liquidity.

This new measurement provides a number of direction that future researches might carry on. First, the measure can be estimated by the arbitrage relationship across different markets or countries, for example, the covered interest parity in the currency market and the CDS-bond relationship in the credit market. By extending to multiple markets or countries, one might be able to analyze the spillover and contagion effect in funding liquidity, especially during the crises period. Second, Brunnermeier and Pedersen (2009) suggest that the phenomenon of flight to liquidity, flight to quality, and commonality in liquidity are attributed to the funding illiquidity among arbitrageurs, especially when the funding constraint is binding. Empirical tests of these hypotheses can thus be adopted with the arbitrage efficacy measure. Third, literature on macroeconomics has long explored the long-lasting effects of the amplification mechanisms. Bernanke and Gertler (1989) argue that leveraged entrepreneurs are sensitive to the funding condition, and low funding liquidity renders higher probability of default, lower overall economic activity and profits. It would be interesting to empirically verify the impact of funding liquidity in the financial sector on overall economy, especially during the periods of inefficacious arbitrage.

Last, we discuss the specific impact of fundamental and sentiment risk exposure on the arbitrage activity with an application on the value premium anomaly. We first modify the model in Chapter 2 by allowing risk averse and heterogeneous arbitrageurs, who concern about the fundamental and sentiment risk. We then show that fundamental and sentiment risk have different effects on the arbitrage activity, captured by the initial mispricing correction and the subsequent noise momentum, which allows us to empirically identify the level of the respective risk exposure. Our empirical evidence from S&P 500 value and growth index reveals that arbitrage in value tend to be limited by higher sentiment risk during the periods of value premium, which strongly supports the behavioral point of view. However, our theory and empirical results do not dispute the possibility that both the fundamental- and sentiment-based theories will jointly explain the value premium anomaly. The potential applications of this approach can go well beyond in testing other cross-sectional anomalies, such as the small-cap effect, short-run momentum effect and long-run reversal effect.

References

Acharya, V. V., Amihud, Y., Bharath, S. T., 2013. Liquidity risk of corporate bond returns: conditional approach. Journal of Financial Economics 110 (2), 358–386.

Acharya, V. V., Shin, H. S., Yorulmazer, T., 2009. A theory of slow-moving capital and contagion. CEPR Discussion Paper No. DP7147.

Adrian, T., Moench, E., Shin, H. S., 2010. Financial intermediation, asset prices and macroeconomic dynamics. FRB of New York Staff Report No. 422.

Adrian, T., Shin, H. S., 2010. Liquidity and leverage. Journal of Financial Intermediation 19 (3), 418–437.

Adrian, T., Shin, H. S., 2011. Financial intermediary balance sheet management. Annual Review of Financial Economics 3 (1), 289–307.

Agarwal, V., Wang, L., 2007. Transaction costs and value premium. Tech. rep., CFR Working Paper.

Akbas, F., Armstrong, W. J., Sorescu, S., Subrahmanyam, A., 2015. Smart money, dumb money, and capital market anomalies. Journal of Financial Economics 118 (2), 355–382.

Akbas, F., Armstrong, W. J., Sorescu, S., Subrahmanyam, A., 2016. Capital market efficiency and arbitrage efficacy. Journal of Financial and Quantitative Analysis 51 (02), 387–413.

Ali, A., Hwang, L.-S., Trombley, M. A., 2003. Arbitrage risk and the book-to-market anomaly. Journal of Financial Economics 69 (2), 355–373.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5 (1), 31–56.

Ang, A., Gorovyy, S., Van Inwegen, G. B., 2011. Hedge fund leverage. Journal of Financial Economics 102 (1), 102–126.

Aragon, G. O., Strahan, P. E., 2012. Hedge funds as liquidity providers: Evidence from the lehman bankruptcy. Journal of Financial Economics 103 (3), 570–587.

Bagehot, W., 1971. The only game in town. Financial Analysts Journal 27 (2), 12–14.

Bai, J., Collin-Dufresne, P., 2013. The cds-bond basis. AFA 2013 San Diego Meetings Paper.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. The Journal of Finance 61 (4), 1645–1680.

Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. The Journal of Economic Perspectives 21 (2), 129–151.

Balke, N. S., Fomby, T. B., 1997. Threshold cointegration. International Economic Review, 627–645.

Barber, B. M., Odean, T., Zhu, N., 2009. Systematic noise. Journal of Financial Markets 12 (4), 547–569.

Barberis, N., Shleifer, A., 2003. Style investing. Journal of Financial Economics 68 (2), 161–199.

Barberis, N., Shleifer, A., Wurgler, J., 2005. Comovement. Journal of Financial Economics 75 (2), 283–317.

Barberis, N., Thaler, R., 2003. A survey of behavioral finance. Handbook of the Economics of Finance 1, 1053–1128.

Beer, F., Watfa, M., Zouaoui, M., 2012. Is sentiment risk priced by stock market? Journal of Applied Business Research 28 (4), 683.

Bernanke, B., Gertler, M., 1989. Agency costs, net worth, and business fluctuations. The American Economic Review, 14–31.

Bertone, S., Paeglis, I., Ravi, R., 2015. (how) has the market become more efficient? Journal of Banking & Finance 54, 72–86.

Black, F., 1986. Noise. The Journal of Finance 41 (3), 528–543.

Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected stock returns and variance risk premia. Review of Financial Studies 22 (11), 4463–4492.

Brunnermeier, M. K., Oehmke, M., 2013. Bubbles, financial crises and systemic risk. Handbook of the Economics of Finance 1, 1222–1281.

Brunnermeier, M. K., Pedersen, L. H., 2008. Market liquidity and funding liquidity. The Review of Financial Studies 22 (6), 2201–2238.

Bushee, B. J., 1998. The influence of institutional investors on myopic r&d investment behavior. Accounting Review, 305–333.

Bushee, B. J., 2001. Do institutional investors prefer near-term earnings over long-run value? Contemporary Accounting Research 18 (2), 207–246.

Cai, C. X., Faff, R., Shin, Y., 2015. Noise momentum around the world. Abacus.

Cai, C. X., McGuinness, P. B., Zhang, Q., 2011. The pricing dynamics of cross-listed securities: The case of chinese a-and h-shares. Journal of Banking & Finance 35 (8), 2123–2136.

Campbell, J. Y., Polk, C., Vuolteenaho, T., 2009. Growth or glamour? fundamentals and systematic risk in stock returns. The Review of Financial Studies 23 (1), 305–344.

Campbell, J. Y., Vuolteenaho, T., 2004. Bad beta, good beta. The American Economic Review 94 (5), 1249–1275.

Chen, Y., Han, B., Pan, J., 2014. Sentiment risk and hedge fund returns. Working Paper, Texas A&m Univ., Univ. of Toronto, and Univ. of Utah.

Choi, J., 2013. What drives the value premium? the role of asset risk and leverage. The Review of Financial Studies 26 (11), 2845–2875.

Chui, A. C., Titman, S., Wei, K., Xie, F., 2012. Explaining the value premium around the world: risk or mispricing? Social Science Research Network.

Cielinska, O., Joseph, A., Shreyas, U. P., Tanner, J., Vasios, M., 2017. Gauging market dynamics using trade repository data: The case of the swiss franc de-pegging. Financial Stability Paper No.41.

Cochrane, J. H., 2011. Presidential address: discount rates. The Journal of Finance 66 (4), 1047–1108.

Coffey, N., Hrung, W. B., Sarkar, A., 2009. Capital constraints, counterparty risk, and deviations from covered interest rate parity. FRB of New York Report No. 393.

Comerton-Forde, C., Hendershott, T., Jones, C. M., Moulton, P. C., Seasholes, M. S., 2010. Time variation in liquidity: the role of market-maker inventories and revenues. The Journal of Finance 65 (1), 295–331.

Constantinides, G. M., Harris, M., Stulz, R., 2013. Handbook of the Economics of Finance. Vol. 2. Elsevier.

Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross sectional variation in stock returns. The Journal of Finance 52 (1), 1–33.

Daniel, K., Titman, S., et al., 2012. Testing factor-model explanations of market anomalies. Critical Finance Review 1 (1), 103–139.

Daniel, K. D., Hirshleifer, D., Subrahmanyam, A., 2001a. Overconfidence, arbitrage, and equilibrium asset pricing. The Journal of Finance 56 (3), 921–965.

Daniel, K. D., Titman, S., Wei, J., 2001b. Cross-sectional variation in common stock returns in japan. Journal of Finance 56 (2), 743–766.

De Long, J. B., Shleifer, A., Summers, L. H., Waldmann, R. J., 1990a. Noise trader risk in financial markets. Journal of Political Economy, 703–738.

De Long, J. B., Shleifer, A., Summers, L. H., Waldmann, R. J., 1990b. Positive feedback investment strategies and destabilizing rational speculation. the Journal of Finance 45 (2), 379–395.

Domowitz, I., Steil, B., 1999. Automation, trading costs, and the structure of the securities trading industry. Brookings-Wharton papers on Financial Services 2, 33–92.

Drehmann, M., Nikolaou, K., 2013. Funding liquidity risk: definition and measurement. Journal of Banking & Finance 37 (7), 2173–2182.

Duffie, D., 2010. Presidential address: asset price dynamics with slow-moving capital. The Journal of Finance 65 (4), 1237–1267.

Duffie, D., Gârleanu, N., Pedersen, L. H., 2007. Valuation in over-the-counter markets. Review of Financial Studies 20 (6), 1865–1900.

Dwyer, G. P., Locke, P., Yu, W., 1996. Index arbitrage and nonlinear dynamics between the s&p 500 futures and cash. Review of Financial Studies 9 (1), 301–332.

Fama, E. F., French, K. R., 1992. The cross-section of expected stock returns. The Journal of Finance 47 (2), 427–465.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33 (1), 3–56.

Fama, E. F., French, K. R., 1996. Multifactor explanations of asset pricing anomalies. The Journal of Finance 51 (1), 55–84.

Fama, E. F., French, K. R., 2006. The value premium and the capm. The Journal of Finance 61 (5), 2163–2185.

Fama, E. F., French, K. R., 2012. Size, value, and momentum in international stock returns. Journal of Financial Economics 105 (3), 457–472.

Fontaine, J.-S., Garcia, R., 2011. Bond liquidity premia. Review of Financial Studies 25 (4), 1207–1254.

Frazzini, A., Lamont, O. A., 2008. Dumb money: mutual fund flows and the cross-section of stock returns. Journal of Financial Economics 88 (2), 299–322.

Frazzini, A., Pedersen, L. H., 2014. Betting against beta. Journal of Financial Economics 111 (1), 1–25.

Friedman, M., 1953. Essays in positive economics. University of Chicago Press.

Fung, W., Hsieh, D. A., 2013. Hedge funds. Handbook of the Economics of Finance 1, 1063–1124.

Gallagher, L. A., Taylor, M. P., 2001. Risky arbitrage, limits of arbitrage, and nonlinear adjustment in the dividend-price ratio. Economic Inquiry 39 (4), 524–536.

Garleanu, N., Pedersen, L. H., 2011. Margin-based asset pricing and deviations from the law of one price. The Review of Financial Studies 24 (6), 1980–2022.

Ghandi, P., Golez, B., Jackwerth, J., Plazzi, A., 2015. Libor manipulation: Cui bono? Working Paper.

Gorton, G., Metrick, A., 2012. Securitized banking and the run on repo. Journal of Financial Economics 104 (3), 425–451.

Griffin, J. M., Lemmon, M. L., 2002. Book-to-market equity, distress risk, and stock returns. The Journal of Finance 57 (5), 2317–2336.

Gromb, D., Vayanos, D., 2002. Equilibrium and welfare in markets with financially constrained arbitrageurs. Journal of Financial Economics 66 (2), 361–407.

Gromb, D., Vayanos, D., 2010. Limits of arbitrage. The Annual Review of Financial Economics 2, 251–75.

Guidolin, M., Timmermann, A., 2008. Size and value anomalies under regime shifts. Journal of Financial Econometrics 6 (1), 1–48.

Gyntelberg, J., Hoerdahl, P., Ters, K., Urban, J., 2017. Arbitrage costs and the persistent non-zero cds-bond basis: Evidence from intraday euro area sovereign debt markets. BIS Working Paper No. 631.

Hamilton, J. D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica: Journal of the Econometric Society, 357–384.

Hasbrouck, J., 1991. Measuring the information content of stock trades. The Journal of Finance 46 (1), 179–207.

Hasbrouck, J., 2003. Intraday price formation in us equity index markets. The Journal of Finance 58 (6), 2375–2400.

He, Z., Krishnamurthy, A., 2013. Intermediary asset pricing. The American Economic Review 103 (2), 732–770.

Hendershott, T., Seasholes, M. S., 2007. Market maker inventories and stock prices. American Economic Review 97, 210–214.

Hombert, J., Thesmar, D., 2014. Overcoming limits of arbitrage: Theory and evidence. Journal of Financial Economics 111 (1), 26–44.

Hull, J., Treepongkaruna, S., Colwell, D., Heaney, R., Pitt, D., 2013. Fundamentals of futures and options markets. Pearson Higher Education AU.

Kumar, A., 2009. Dynamic style preferences of individual investors and stock returns. Journal of Financial and Quantitative Analysis 44 (3), 607–640.

Kurov, A., Lasser, D. J., 2004. Price dynamics in the regular and e-mini futures markets. Journal of Financial and Quantitative Analysis 39 (2), 365–384.

Lakonishok, J., Shleifer, A., Vishny, R. W., 1994. Contrarian investment, extrapolation, and risk. The Journal of Finance 49 (5), 1541–1578.

Lee, C., Shleifer, A., Thaler, R. H., 1991. Investor sentiment and the closed-end fund puzzle. The Journal of Finance 46 (1), 75–109.

Lewellen, J., Nagel, S., 2006. The conditional capm does not explain asset-pricing anomalies. Journal of financial economics 82 (2), 289–314.

Lewellen, J., Nagel, S., Shanken, J., 2010. A skeptical appraisal of asset pricing tests. Journal of Financial economics 96 (2), 175–194.

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. The Review of Economics and Statistics, 13–37.

Lou, D., 2012. A flow-based explanation for return predictability. The Review of Financial Studies 25 (12), 3457–3489.

MacKinlay, A. C., Ramaswamy, K., 1988. Index-futures arbitrage and the behavior of stock index futures prices. Review of Financial Studies 1 (2), 137–158.

Malkiel, B. G., Fama, E. F., 1970. Efficient capital markets: A review of theory and empirical work. The journal of Finance 25 (2), 383–417.

Mancini-Griffoli, T., Ranaldo, A., 2011. Limits to arbitrage during the crisis: funding liquidity constraints and covered interest parity. Swiss National Bank Working Paper.

Marshall, B. R., Nguyen, N. H., Visaltanachoti, N., 2013. Etf arbitrage: Intraday evidence. Journal of Banking & Finance 37 (9), 3486–3498.

Martens, M., Kofman, P., Vorst, T. C., 1998. A threshold error-correction model for intraday futures and index returns. Journal of Applied Econometrics, 245–263.

McLean, R. D., 2010. Idiosyncratic risk, long-term reversal, and momentum. Journal of Financial and Quantitative Analysis 45 (4), 883–906.

Merton, R. C., 1973. An intertemporal capital asset pricing model. Econometrica: Journal of the Econometric Society, 867–887.

Michael, P., Nobay, A. R., Peel, D. A., 1997. Transactions costs and nonlinear adjustment in real exchange rates: an empirical investigation. Journal of Political Economy 105 (4), 862–879.

Mitchell, M., Pedersen, L. H., Pulvino, T., 2007. Slow moving capital. American Economic Review 97 (2), 215–220.

Mitchell, M., Pulvino, T., 2012. Arbitrage crashes and the speed of capital. Journal of Financial Economics 104 (3), 469–490.

Nagel, S., 2005. Short sales, institutional investors and the cross-section of stock returns. Journal of Financial Economics 78 (2), 277–309.

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

Petkova, R., Zhang, L., 2005. Is value riskier than growth? Journal of Financial Economics 78 (1), 187–202.

Phalippou, L., 2004. Institutional ownership and the value premium. Tech. rep., EFA.

Phalippou, L., 2007. Can risk-based theories explain the value premium? Review of Finance 11 (2), 143–166.

Pontiff, J., 2006. Costly arbitrage and the myth of idiosyncratic risk. Journal of Accounting and Economics 42 (1), 35–52.

Ritter, J. R., 2003. Behavioral finance. Pacific-Basin finance journal 11 (4), 429–437.

Roll, R., Schwartz, E., Subrahmanyam, A., 2007. Liquidity and the law of one price: the case of the futures-cash basis. The Journal of Finance 62 (5), 2201–2234.

Santos, T., Veronesi, P., 2005. Cash-flow risk, discount risk, and the value premium. NBER Working Paper.

Sarno, L., Valente, G., 2000. The cost of carry model and regime shifts in stock index futures markets: An empirical investigation. Journal of Futures Markets 20 (7), 603–624.

Schuster, P., Uhrig-Homburg, M., 2015. Limits to arbitrage and the term structure of bond illiquidity premiums. Journal of Banking & Finance 57, 143–159.

Scruggs, J. T., 2007. Noise trader risk: Evidence from the siamese twins. Journal of Financial Markets 10 (1), 76–105.

Sharpe, W. F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk*. The Journal of Finance 19 (3), 425–442.

Shiller, R. J., Fischer, S., Friedman, B. M., 1984. Stock prices and social dynamics. Brookings papers on economic activity 1984 (2), 457–510.

Shleifer, A., Summers, L. H., 1990. The noise trader approach to finance. The Journal of Economic Perspectives 4 (2), 19–33.

Shleifer, A., Vishny, R., 2011. Fire sales in finance and macroeconomics. The Journal of Economic Perspectives 25 (1), 29–48.

Shleifer, A., Vishny, R. W., 1997. The limits of arbitrage. The Journal of Finance 52 (1), 35–55.

Simutin, M., 2013. Cash holdings and mutual fund performance. Review of Finance 18 (4), 1425–1464.

Stein, J. C., 2009. Presidential address: sophisticated investors and market efficiency. The Journal of Finance 64 (4), 1517–1548.

Summers, L. H., 1986. Does the stock market rationally reflect fundamental values? The Journal of Finance 41 (3), 591–601.

Tao, J., Green, C. J., 2013. Transactions costs, index arbitrage and non-linear dynamics between ftse100 spot and futures: a threshold cointegration analysis. International Journal of Finance & Economics 18 (2), 175–187.

Thaler, R. H., 2005. Advances in behavioral finance. Vol. 2. Princeton University Press.

Theissen, E., 2012. Price discovery in spot and futures markets: a reconsideration. The European Journal of Finance 18 (10), 969–987.

Tse, Y., 2001. Index arbitrage with heterogeneous investors: A smooth transition error correction analysis. Journal of banking & finance 25 (10), 1829–1855.

Vayanos, D., Weill, P.-o., 2008. A search-based theory of the on-the-run phenomenon. The Journal of Finance 63 (3), 1361–1398.

Zhang, L., 2005. The value premium. The Journal of Finance 60 (1), 67–103.